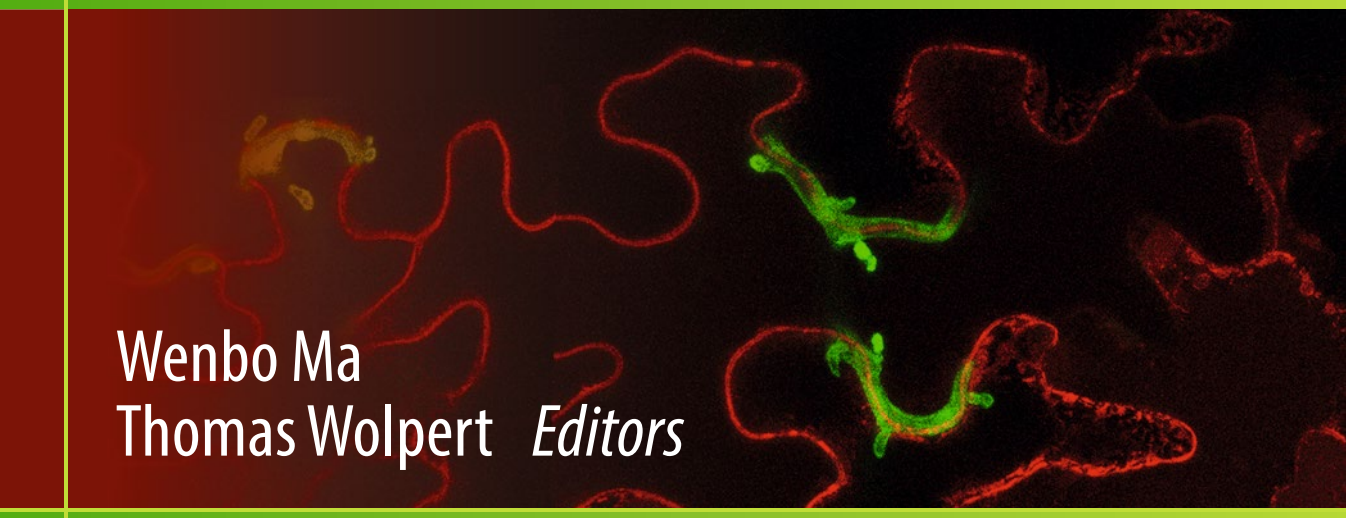
Methods in Molecular Biology 1848

Springer Protocols



Wenbo Ma
Thomas Wolpert *Editors*

Plant Pathogenic Fungi and Oomycetes

Methods and Protocols

Humana Press

METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:
http://www.springer.com/series/7651

Plant Pathogenic Fungi and Oomycetes

Methods and Protocols

Edited by

Wenbo Ma

Department of Microbiology and Plant Pathology, University of California Riverside , Riverside, CA, USA

Thomas Wolpert

Oregon State University , Corvallis, OR, USA

Humana Press

Editors
Wenbo Ma
Department of Microbiology
and Plant Pathology
University of California Riverside
Riverside, CA, USA

Thomas Wolpert
Oregon State University
Corvallis, OR, USA

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-8723-8 ISBN 978-1-4939-8724-5 (eBook)
https://doi.org/10.1007/978-1-4939-8724-5

Library of Congress Control Number: 2018950066

© Springer Science+Business Media, LLC, part of Springer Nature 2018
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Humana Press imprint is published by the registered company Springer Science+Business Media, LLC part of Springer Nature.
The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

Preface

Filamentous eukaryotic pathogens, including fungi and oomycetes, are notorious for causing destructive diseases on plants. Historically, these pathogens have had tremendous impacts on natural and managed ecosystems and continue to threaten agriculture and forestry. They exert significant limitations to food and fiber production and cause damage to natural and man-made landscapes. Deciphering the virulence mechanisms employed by fungal/oomycete pathogens and conditions that influence their ability to cause disease is critically important for our future well-being. Such discovery has considerable economic, social, and environmental relevance. This book attempts to inform the readers of techniques applicable to interrogating a broad range of phenomena that impact the virulence of filamentous eukaryotic pathogens.

The methods described in this book provide a diversity of approaches that range from ecosystem to in vitro and wet-lab to computational analyses. Chapters 1–4 present methods developed to investigate multipartite interactions. These interactions have tremendous potential to impact virulence and plant disease. In the first two chapters, methods for the evaluation of filamentous organisms as hosts of endosymbionts (viruses and bacteria) are provided. In the next two chapters, techniques are described for examining the metabolic and communal contribution of filamentous endophytes to the plant host environment.

The ubiquitous role of small RNAs in gene regulation is becoming increasingly apparent. Emerging evidence suggests that small RNAs play an important role in plant-pathogen interactions. Included among these functions is the regulation of both pathogen virulence and host defense. Chapters 5 and 6 provide methods for investigating small RNAs in fungi and plants. Another critical regulatory mechanism of protein function involved in various aspects of growth, development, and virulence is post-translational modification, especially phosphorylation. Chapters 7 and 8 provide techniques for evaluating protein phosphorylation using either genome-wide analysis or targeted approaches.

A comprehensive understanding of virulence demands a detailed visual examination and monitoring of the interactions of filamentous pathogens with their plant hosts including the function, growth, and development of the pathogen. Chapters 9 and 10 provide protocols for pursuing these analyses using current cell biology approaches and confocal microscopy. Chapters 11 and 12 present distinct approaches for using the host to interrogate virulence functions with the ultimate goal of developing disease resistance. The first chapter utilizes transient expression of pathogen effectors in wild plant species to identify naturally occurring resistant germplasm whereas the second chapter describes a host-induced gene silencing (HIGS) approach, which exploits the potential of the aforementioned small RNAs and the potential cross-kingdom trafficking of small RNAs from plants to silence pathogen genes.

Characterization of virulence invariably requires genetic analyses, which are significantly strengthened by fast-evolving sequencing technologies. Computational analyses are becoming more and more in demand with the rapid generation of sequencing data. Chapters 13, 14, and 15 provide bioinformatics protocols for genomic analyses including fungi/oomycete genome assembly, transcriptome analysis, and specific sequencing

procedures designed to identify genes involved in virulence. Finally, confirmation of the contribution of genes identified as having a potential virulence function inevitably requires their genetic manipulation. The final three chapters describe a variety of approaches to generate mutants for achieving this objective.

The editors would like to thank all the authors for their contribution to this book. It is hoped that the inclusion of techniques developed through this wide range of perspectives can encourage a comprehensive understanding of these important filamentous pathogens. The editors invite you to exploit these methods.

Riverside, CA, USA — *Wenbo Ma*
Corvallis, OR, USA — *Thomas Wolpert*

Contents

Preface . . . . . *v*
Contributors . . . . . *ix*

1 Quantifying Re-association of a Facultative Endohyphal Bacterium with a Filamentous Fungus 1
David A. Baltrus, Joseph Spraker, and A. Elizabeth Arnold

2 Characterizing Mycoviruses 13
Mahtab Peyambari and Marilyn J. Roossinck

3 Analysis of Secondary Metabolites from Plant Endophytic Fungi 25
Jiajia Liu and Gang Liu

4 Protocols for Investigating the Leaf Mycobiome Using High-Throughput DNA Sequencing. 39
Shawn P. Brown, Devin R. Leopold, and Posy E. Busby

5 Characterizing Small RNAs in Filamentous Fungi Using the Rice Blast Fungus, *Magnaporthe oryzae*, as an Example 53
Vidhyavathi Raman, Blake C. Meyers, Ralph A. Dean, and Nicole M. Donofrio

6 Plant Small RNAs Responsive to Fungal Pathogen Infection 67
Yun Jin and Hui-Shan Guo

7 Sequential Phosphopeptide Enrichment for Phosphoproteome Analysis of Filamentous Fungi: A Test Case Using *Magnaporthe oryzae* 81
Yeonyee Oh, William L. Franck, and Ralph A. Dean

8 Assays for MAP Kinase Activation in *Magnaporthe oryzae* and Other Plant Pathogenic Fungi 93
Xue Zhang, Zhuyun Bian, and Jin-Rong Xu

9 Visualizing the Movement of *Magnaporthe oryzae* Effector Proteins in Rice Cells During Infection 103
Kiersun Jones and Chang Hyun Khang

10 Illuminating *Phytophthora* Biology with Fluorescent Protein Tags. 119
Audrey M. V. Ah-Fong, Meenakshi Kagda, and Howard S. Judelson

11 Methods for Transient Gene Expression in Wild Relatives of Potato. 131
Sidrat Abdullah and Dennis Halterman

12 Host-Induced Gene Silencing (HIGS) for Elucidating *Puccinia* Gene Function in Wheat. 139
Chuntao Yin and Scot H. Hulbert

13 From Short Reads to Chromosome-Scale Genome Assemblies 151
Kyle Fletcher and Richard Michelmore

14 BLASTmap: A Shiny-Based Application to Visualize BLAST Results as Interactive Heat Maps and a Tool to Design Gene-Specific Baits for Bespoke Target Enrichment Sequencing 199
Katie Baker, Gordon Stephen, Shona Strachan, Miles Armstrong, and Ingo Hein

15 A Computational Protocol to Analyze Metatranscriptomic Data Capturing Fungal–Host Interactions 207
Yong Zhang, Li Guo, and Li-Jun Ma

16 Application of the Cre/*lox* System to Construct Auxotrophic Markers for Quantitative Genetic Analyses in *Fusarium graminearum* 235
Lanelle R. Connolly, Allyson A. Erlendson, Corinne M. Fargo, Kendra K. Jackson, Morgan M. G. Pelker, Jacob W. Mazzola, Mark S. Geisler, and Michael Freitag

17 Protocol of *Phytophthora capsici* Transformation Using the CRISPR-Cas9 System 265
Zhiwen Wang, Brett M. Tyler, and Xili Liu

18 Generating Gene Silenced Mutants in *Phytophthora sojae* 275
Yao Zhao, Yufei Li, Min Qiu, Wenbo Ma, and Yuanchao Wang

Index 287

Contributors

Sidrat Abdullah • *Department of Plant Pathology, University of Wisconsin-Madison, Madison, WI, USA*

Audrey M. V. Ah-Fong • *Department of Microbiology and Plant Pathology, University of California, Riverside, CA, USA*

Miles Armstrong • *The James Hutton Institute, Cell and Molecular Sciences, Dundee, UK*

A. Elizabeth Arnold • *School of Plant Sciences, University of Arizona, Tucson, AZ, USA; Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA*

Katie Baker • *Synpromics, Edinburgh, UK*

David A. Baltrus • *School of Plant Sciences, University of Arizona, Tucson, AZ, USA; School of Animal and Comparative Biomedical Sciences, University of Arizona, Tucson, AZ, USA*

Zhuyun Bian • *Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN, USA*

Shawn P. Brown • *Department of Biological Sciences, University of Memphis, Memphis, TN, USA*

Posy E. Busby • *Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, USA*

Lanelle R. Connolly • *Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR, USA*

Ralph A. Dean • *Fungal Genomics Lab, Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC, USA*

Nicole M. Donofrio • *Department of Plant and Soil Sciences, University of Delaware, Newark, DE, USA*

Allyson A. Erlendson • *Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR, USA*

Corinne M. Fargo • *Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR, USA*

Kyle Fletcher • *The Genome Center, Genome and Biomedical Sciences Facility, University of California, Davis, CA, USA*

William L. Franck • *Fungal Genomics Lab, Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC, USA*

Michael Freitag • *Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR, USA*

Mark S. Geisler • *Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR, USA*

Hui-Shan Guo • *State Key Laboratory of Plant Genomics, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China; College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China*

Li Guo • *School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an, China*
Dennis Halterman • *US Department of Agriculture, Agricultural Research Service, Madison, WI, USA*
Ingo Hein • *The James Hutton Institute, Cell and Molecular Sciences, Dundee, UK; Division of Plant Sciences at the James Hutton Institute, School of Life Sciences, University of Dundee, Dundee, UK*
Scot H. Hulbert • *Department of Plant Pathology, Washington State University, Pullman, WA, USA*
Kendra K. Jackson • *Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR, USA*
Yun Jin • *State Key Laboratory of Plant Genomics, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China*
Kiersun Jones • *Department of Plant Biology, University of Georgia, Athens, GA, USA*
Howard S. Judelson • *Department of Microbiology and Plant Pathology, University of California, Riverside, CA, USA*
Meenakshi Kagda • *Department of Microbiology and Plant Pathology, University of California, Riverside, CA, USA*
Chang Hyun Khang • *Department of Plant Biology, University of Georgia, Athens, GA, USA*
Devin R. Leopold • *Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, USA*
Yufei Li • *Department of Plant Pathology, Nanjing Agricultural University, Nanjing, China*
Gang Liu • *State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China; University of Chinese Academy of Sciences, Beijing, China*
Jiajia Liu • *State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China; University of Chinese Academy of Sciences, Beijing, China*
Xili Liu • *Department of Plant Pathology, China Agricultural University, Beijing, China*
Li-Jun Ma • *Department of Biochemistry and Molecular Biology, University of Massachusetts Amherst, Amherst, MA, USA*
Wenbo Ma • *Department of Microbiology and Plant Pathology, University of California, Riverside, Riverside, CA, USA*
Jacob W. Mazzola • *Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR, USA*
Blake C. Meyers • *Division of Plant Sciences, Danforth Center, St. Louis, MO, USA*
Richard Michelmore • *The Genome Center, Genome and Biomedical Sciences Facility, University of California, Davis, CA, USA*
Yeonyee Oh • *Fungal Genomics Lab, Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC, USA*
Morgan M. G. Pelker • *Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR, USA*
Mahtab Peyambari • *Center for Infectious Disease Dynamics, Millennium Science Complex, Pennsylvania State University, University Park, PA, USA*
Min Qiu • *Department of Plant Pathology, Nanjing Agricultural University, Nanjing, China*
Vidhyavathi Raman • *Noble Research Institute, Ardmore, OK, USA*

Marilyn J. Roossinck • *Center for Infectious Disease Dynamics, Millennium Science Complex, Pennsylvania State University, University Park, PA, USA*
Joseph Spraker • *School of Plant Sciences, University of Arizona, Tucson, AZ, USA*
Gordon Stephen • *The James Hutton Institute, Information and Computational Sciences, Dundee, UK*
Shona Strachan • *The James Hutton Institute, Cell and Molecular Sciences, Dundee, UK; University of St Andrews, College Gate, St Andrews, UK*
Brett M. Tyler • *Department of Botany and Plant Pathology, Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR, USA*
Yuanchao Wang • *Department of Plant Pathology, Nanjing Agricultural University, Nanjing, China*
Zhiwen Wang • *Department of Plant Pathology, China Agricultural University, Beijing, China*
Jin-Rong Xu • *Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN, USA*
Chuntao Yin • *Department of Plant Pathology, Washington State University, Pullman, WA, USA*
Xue Zhang • *Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN, USA*
Yong Zhang • *Department of Biochemistry and Molecular Biology, University of Massachusetts Amherst, Amherst, MA, USA*
Yao Zhao • *Department of Plant Pathology, Nanjing Agricultural University, Nanjing, China; Department of Microbiology and Plant Pathology, University of California, Riverside, Riverside, CA, USA*



Chapter 1

Quantifying Re-association of a Facultative Endohyphal Bacterium with a Filamentous Fungus

David A. Baltrus, Joseph Spraker, and A. Elizabeth Arnold

Abstract

We present here a method to quantify reassociation between facultative endohyphal bacteria and filamentous fungal hosts. Our method takes advantage of the capabilities of fungal cell walls to selectively protect internal bacteria from gentamicin treatment, an assay adapted from studies of internalized bacterial pathogens in cell culture. We report the efficacy of gentamicin to kill planktonic bacteria treated during fungal coculture, and also describe and characterize a sampling scheme to recover and quantify culturable bacteria from the growing edge of fungal mycelium in vitro. This assay enables qualitative and quantitative tests of reassociation capabilities for facultative endohyphal bacteria with host fungi and provides a means to investigate the genetic basis for these associations in a repeatable way.

Key words Endohyphal bacteria, Endofungal bacteria, Ascomycete, Reassociation

1 Introduction

Interactions between bacteria and fungi are well documented throughout the natural world. From a basis in Fleming's observations regarding antibiotic production by fungi to records of mycophagy by diverse bacterial species, a rich history of studies highlights the physiological and ecological interplay that shapes bacterial and fungal traits [1–5]. Over the past three decades it has become clear that some bacteria can function as facultative and/or obligate bacterial symbionts within fungal hyphae and spores, and that this interaction modulates fungal traits in specific contexts [5–8]. Commonly known as endohyphal bacteria (EHB; also, endobacteria, endofungal bacteria, bacterial endosymbionts of fungi; reviewed by [5]), EHB are now recognized as common among fungi in clades such as the Mucoromycota, Ascomycota, and Basidiomycota. Some of these EHB are vertically transmitted, but the greatest diversity occurs among horizontally transmitted and often facultative symbionts representing the Proteobacteria,

Wenbo Ma and Thomas Wolpert (eds.), *Plant Pathogenic Fungi and Oomycetes: Methods and Protocols*, Methods in Molecular Biology, vol. 1848, https://doi.org/10.1007/978-1-4939-8724-5_1,
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Firmicutes, and Bacteroidetes [9, 10]. Realizations about the ubiquity of intricate inter-microbe symbioses have motivated the development and optimization of assays to qualify and quantify the ability of bacteria to colonize fungal structures (e.g., [11, 12]), with the main goals being a deeper understanding of the genetic basis of interactions, quantification of their phenotypic importance, and delimitation of host ranges.

Here, we describe an assay to quantify colonization of filamentous, septate fungi (here represented by a foliar endophyte in the Ascomycota) by diverse, facultative bacteria (here represented by a species of *Luteibacter* in the Proteobacteria). We further present data that demonstrate the sensitivity of this assay to several experimental variables. In the assay, bacteria and fungi are grown separately and then mixed as a coculture in minimal media to facilitate associations [12]. After 3 days, gentamicin is added to kill free-living bacteria, a technique typically used to select for bacteria invading eukaryotic cells in tissue culture [13]. As with previous assays based on tissue culture, we find that endohyphal bacteria that successfully enter fungal hyphae are protected from this killing. Fungal hyphae then are transferred to water agar and hyphae are allowed to grow from the central mass. After hyphal growth is apparent, usually >3 days later, samples of hyphae are cored to sample for living bacteria.

A current version of this protocol can be found on the open source platform protocols.io: dx.doi.org/10.17504/protocols.io.jq2cmye. We will update this protocol when necessary on protocols.io, so that the current working version may differ from that described below.

2 Materials

2.1 Media

1. Lysogeny Broth (1 L).

   Mix:

   10 g Tryptone.

   5 g Yeast Extract.

   10 g NaCl.

   Adjust to 1 L with distilled H_2O.

   Sterilized by autoclaving.

2. M9 Salts (1 L).

   Mix:

   64 g Na_2HPO_4-$7H_2O$.

   5 g KH_2PO_4.

   2.5 g NaCl.

   5.0 g NH_4Cl.

Stir until dissolved.

Adjust to 1 L with distilled H_2O.

Sterilize by autoclaving.

3. M9 Media with Glucose (1 L).

   Mix:

   ~700 mL of distilled H_2O.

   200 mL M9 Salts.

   2 mL of 1 M $MgSO_4$.

   20 mL of 20% glucose.

   100 μL of $CaCl_2$.

   Adjust to 1 L with distilled H_2O.

   Sterilize by filtration.

4. Water Agar.

   Mix:

   15 g Agar.

   Adjust to 1 L with distilled H_2O.

2.2 Focal Fungal and Bacterial Strains

For assays in this chapter we used *Petalotiopsis* sp. 9143 as a host fungus (Sordariomycetes, Pezizomycotina, Ascomycota). This strain was isolated as an endophyte from healthy foliage of *Platycladis orientalis* (Cupressaceae) in Durham, NC, and is available through the University of Arizona Mycological Herbarium on request from AE Arnold [10].

A rifampicin-resistant isolate of *Luteibacter* sp. 9143, DBL564, was used as a focal bacterial strain for all experiments. This strain originated from a single rifampicin resistant colony from subcultures of the original isolate that emerged from *Pestalotiopsis* sp. 9143 during laboratory growth [12]. The strain is considered to have been a "natural" symbiont of the fungus, and aspects of its biology have been previously described [5, 10, 12] (*see* **Note 1**).

3 Methods

3.1 Preparation of Fungal and Bacterial Strains

Four to seven days before the start of experiment, transfer a small sample of fungal hyphae to 3 mL of M9 medium supplemented with 0.4% glucose. Alternatively, cultures can be initiated from spores. Place on a rotating shaker at 200 rpm and at 27 °C. One day before starting the experiment, inoculate a culture of LB (Lysogeny broth, or other appropriate rich media, supplemented with antibiotics if necessary) with the bacterial strain of interest for reassociation. Place on a rotating shaker at 200 rpm and at 27 °C (*see* **Notes 2** and **3**).

3.2 Preparation of Fungi for Coculture

Transfer mycelium from liquid culture to a 2 mL Eppendorf tube containing 4 × 3 mm glass beads and 500 μL of M9 media with 0.4% glucose. Run 2 × 20 s cycles of shaking using a MP Fastprep (or equivalent) machine. Add an additional 500 μL of 10 mM $MgCl_2$ to the tube.

3.3 Fungal Culture Initiation

To 20 mL of M9 media with 0.4% glucose, add 50 μL of fungal inoculum prepared in **step 2**. Mix. Aliquot 2 mL from this master mix to test tubes and place on a rotating shaker at 27 °C and 200 rpm.

3.4 Preparation of Bacteria for Coculture

Pellet overnight culture (typically ~8000 × *g* for 2 min). Pour off supernatant and resuspend pellet in 1 mL of 10 mM $MgCl_2$ to wash. Centrifuge, decant, and resuspend twice more to wash bacterial cells. Finally, resuspend washed cells in 1 mL of 10 mM $MgCl_2$.

3.5 Addition of Bacteria to Fungal Cultures

Bacteria can be added to the master mix mentioned in **step 3**, at the same time as the fungi. In this case, add 50 μL of bacteria (as prepared in **step 4** above) for every 10 mL of M9 media in the master mix. We have found that bacteria can also be added at various points (over the course of days) after the fungal cultures are initiated. In this case, prepare a new bacterial inoculum master mix as per **step 4** and adding 50 μL of bacteria for every 10 mL of M9 media with 0.4% glucose. Add 2 mL of this master mix to the growing fungal cultures. Four days after bacteria and fungi are mixed, add M9 media with 0.4% glucose containing 40 ng/μL gentamicin in a 1:1 volume with the coculture. Gentamicin treatment is not explicitly necessary, but helps to remove extrahyphal bacteria by killing those not established within fungal hyphae. As a negative control, add the same amount of M9 media with 0.4% glucose (w/o gentamicin) to another set of cocultures.

3.6 Transfer to Agar Plate

Transfer the mass of fungal hyphae from the coculture tube to a silica DNA spin column as one would find in a plasmid preparation kit. Add 750 μL of 10 mM $MgCl_2$ to the spin column and spin in centrifuge (typically 1 min at 12,000 × *g*). Discard flow-through and add 750 μL of 10 mM $MgCl_2$ to the spin column and spin in centrifuge (as above). Transfer the fungal hyphal mass to a 1.5% (w/v) water agar plate and seal with Parafilm. Incubate at 27 °C and allow for growth (usually takes at least 4 days).

3.7 Harvesting of Bacteria

Use a cork borer to sample hyphae from the water agar plate, selecting newly grown hyphae outside of the originally transferred sample. Visually inspect the plate for bacterial growth, and sample only if bacterial growth is not apparent. Collect hyphal sample and place in a 2 mL Eppendorf tube containing glass beads and 500 μL of 10 mM $MgCl_2$. Run 2 × 20s cycles of shaking using a MP

Fastprep (or equivalent) machine. Add an additional 500 μL 10 mM $MgCl_2$ to the tube. Plate a dilution series of each sample on LB plates containing 100 ng/mL cyclohexamide and place plates in a 27 °C incubator for at least 2 days or until growth is visible. Count colonies, typically on the order of 10^5 to 10^6 for effective reassociations (*see* Fig. 3). It is possible to add additional antibiotics or other amendments to select for specific bacteria or prevent fungal growth (*see* **Note 4**).

3.8 Gentamicin Treatment Eliminates Culturable Extrahyphal Bacteria

To test the efficacy of the gentamicin treatment in killing bacteria outside of fungal hyphae, we conducted bacterial counts of coculture supernatants before (i.e., 3 days after initiation of coculture) and 1 day after (i.e., 4 days after initiation of coculture) treatment of cultures with gentamicin. Cultures were initiated as per **steps 1** through **5** above. Gentamicin was added to fresh M9 media, which was then added to the growing 2 mL cultures at a 1:1 volume, to a final concentration of 20 ng/mL in M9 media. As a negative control, 2 mL of M9 media without antibiotic supplementation was added to additional growing cultures. For comparison, population counts were also conducted when the aforementioned cocultures were initiated. Each experimental group consisted of four replicate cultures per assay, initiated from the same fungal and bacterial inoculum, and assays were repeated twice independently. Counts were conducted twice independently. Cells were counted after dilution plating on LB media supplemented with rifampicin (to a final concentration of 50 ng/mL) and cyclohexamide (100 ng/mL).

In both assays, bacterial population sizes in all cultures increased by 100-fold over the course of the experiment, and reached the same approximate densities by day three (Fig. 1). Although cell counts remained stable from day 3 to day 4 after the addition of M9 media to cultures, the addition of gentamicin (to a final concentration of approximately 20 ng/mL) caused a dramatic decline in culturable cells in the supernatant. No viable cells were apparent in the coculture supernatant in either replicate assay (the theoretical limit of detection was 100 CFU/mL). Although it is possible that bacterial cells entered into a viable but nonculturable state, it is clear that addition of gentamicin to the cultures dramatically altered the planktonic and extrahyphal bacterial populations under these assay conditions.

3.9 Coculturing Leads to Endohyphal Bacteria

After coculture in M9 media, each of the fungal masses from the experiments shown in Fig. 1 were rinsed with 10 mM $MgCl_2$ and placed on individual 1.5% water agar plates as per **step 6** above. Plates were wrapped in Parafilm and incubated at 27 °C for up to 7 days. Beginning on day 4, one core (6 mm in diameter) was sampled each day from the growing edge of the mycelium (Fig. 2). This core was placed into a 2 mL Eppendorf tube containing 500 μL of 10 mM $MgCl_2$ and 3 glass beads, macerated in 2 × 20 s

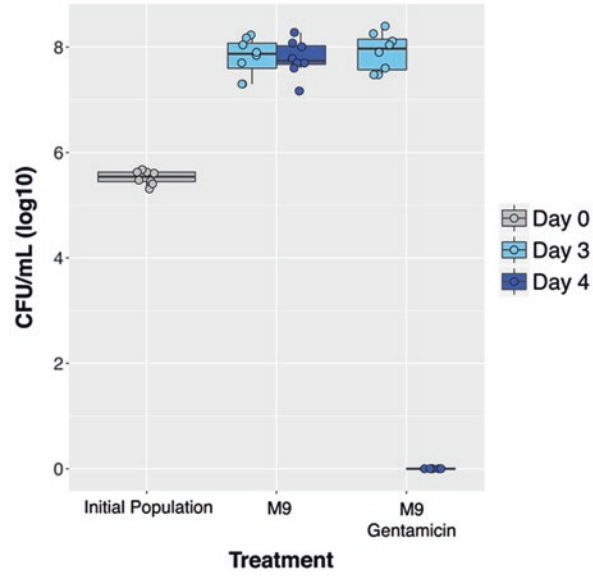


Fig. 1 Gentamicin treatment kills planktonic extrahyphal bacteria culturable bacterial populations were enumerated by dilution plating from supernatants of cocultures treated with either M9 supplemented with gentamicin (to a final concentration of 20 ng/μL) or M9 without antibiotic supplementation. Samples were taken at the start of coculture (initial population); prior to antibiotic addition (Day 3); and 1 day after antibiotic addition (Day 4). Two independent assays were carried out, with four replicates per assay. There was no change in bacterial population sizes after the addition of M9 alone, but addition of gentamicin eliminated culturable bacteria from the supernatant in all cultures

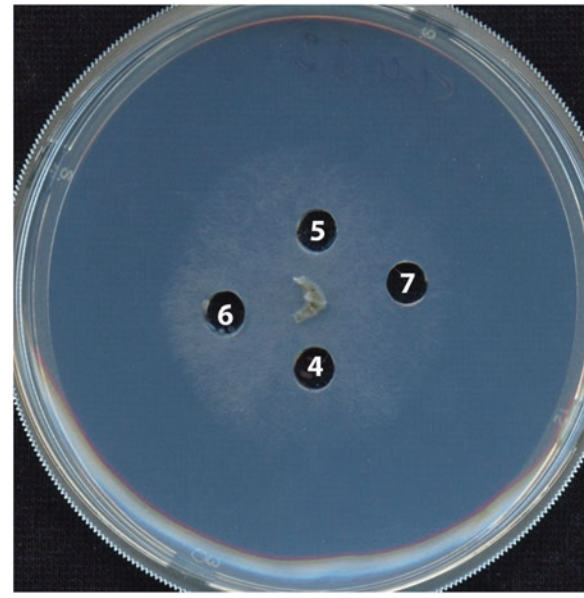


Fig. 2 Sampling scheme for bacterial counts. A representative plate with cores taken after 4, 5, 6, and 7 days of growth after liquid culture is shown

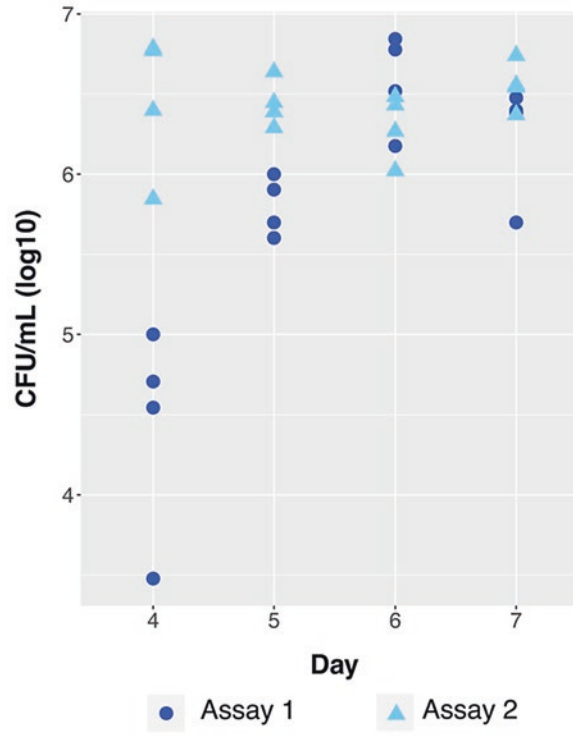


Fig. 3 Bacterial population sizes associated with fungal hyphae after coculture. Fungal masses were plated onto 1.5% water agar after 4 days of coculture with bacteria in M9 media without antibiotic supplementation. One core per plate was sampled after 4, 5, 6, and 7 days of fungal growth on the agar plates. Two independent assays with four replicates each were carried out

cycles, and brought up to a final volume of 1 mL through addition of 500 μL of 10 mM $MgCl_2$. Bacteria were enumerated by dilution plating on LB agar plates supplemented with 100 ng/mL cycloheximide and 50 ng/mL rifampicin.

Population counts of bacteria associated with fungal hyphae can be quite high after 3 days of growth (~10^6 per core), and remain fairly constant from days 4 through 7 (Fig. 3). While this is a representative set of data, bacterial population counts from assay 1 increased with longer incubation times to a limit, and can be quite variable from day to day and from plate to plate (see outlier day 7 sample).

3.10 Gentamicin Treatment During Liquid Coculture Does Not Affect Bacterial Counts Associated with Growing Hyphae on Agar Plates

We sampled bacterial population sizes from fungal tissue after 7 days of growth on 1.5% water agar, for two independent reassociation experiments and for two different treatments. In one treatment, M9 media supplemented with gentamicin was added to liquid cocultures after 3 days of growth. In the second treatment, M9 media alone was added to the cocultures and did not affect overall bacterial population counts. Fungal masses from each of these cocultures were rinsed and plated on water agar, and

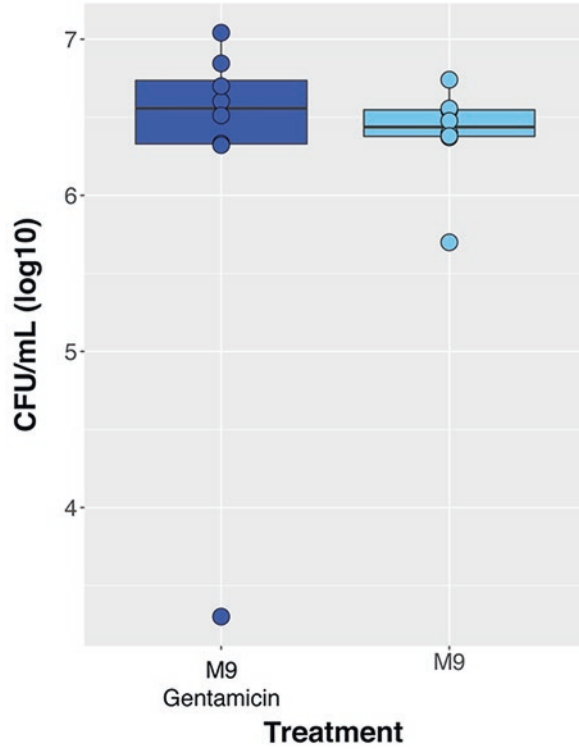


Fig. 4 Gentamicin treatment of cocultures does not affect bacterial counts after plating. Cocultures treated with either M9 without antibiotic supplementation or M9 supplemented with gentamicin (to a final concentration of 20 ng/mL) were plated onto 1.5% water agar plates and cores were taken after 7 days of growth. Two independent assays were carried out, with four replicates per assay. Although gentamicin treatment eliminated planktonic bacteria within these experiments (***see*** Fig. 1), there is no overall difference in colony counts from fungal tissue, as sampled by coring after 7 days of growth on agar ($p = 0.674$)

incubated for 7 days. At this point, one core was taken from the growing hyphal edge and sampled for viable bacteria. Although gentamicin treatment eliminated all viable bacteria within the coculture supernatant across all replicates (Fig. 1), we observed no difference in viable bacteria sampled from hyphae after 7 days of growth on agar plates (Fig. 4).

4 Notes

1. We have developed this assay for use with hyphal inocula of *Pestalotiopsis* sp. 9143 and various *Luteibacter* strains. We have tested this assay with a handful of additional fungi and bacteria, and have successfully found reassociation, but the efficacy of this protocol across other strains and species remains unclear.

2. Although we are unsure of how temperature will affect reassociation generally, we note that some environmental bacteria (i.e., the *Leutibacter* used herein) as well as some fungi do not grow well (or at all) at higher temperatures and may grow more slowly (if at all) at lower temperatures.
3. We have found that it is possible for fungal spore preps to be used for initial inoculation during reassociation assays, but results may be more variable.
4. We typically use rifampicin resistant versions of bacterial strains to insure identification of strains after reassociation, but have found that other antibiotic resistances also work for the purposes of selecting for individual bacteria (i.e., streptomycin, kanamycin, naldixic axid). We always include cyclohexamide in dilution plates after reassociation to avoid overgrowth of fungi on media before bacterial growth appears. To identify any natural symbionts of fungi, it can also be useful to plate on media without additional antibiotics.

5 Conclusions

Searches for the genes that establish and govern interactions between bacteria and fungi will be greatly aided by the development of repeatable assays to establish and then quantify association success, cell number, and symbiotic affiliation. Such assays also will be crucial in testing for quantitative and qualitative differences in association and the delimitation of host ranges across bacterial and fungal isolates.

In this chapter we have demonstrated that liquid coculture of a representative fungus in the Ascomycota (*Pestalotiopsis* sp. 9143) and bacterial endosymbiont representing the Proteobacteria (*Luteibacter* sp. 9143) in M9 media enables reestablishment of a close association, and that population sizes of endohyphal bacteria can be measured repeatedly by taking hyphal cores after 4–7 days of growth on water agar. We also note that in each of the above assays, where we began each with confirmed "axenic" fungi, we did not recover any bacteria from cultures that were not inoculated with bacteria. Furthermore, we have presented evidence that the gentamicin protection assay, pioneered for in vitro selection of intracellular bacterial parasites during tissue culture [13], is a viable method for selecting bacteria that have potentially established endohyphal relationships within fungal hosts. We have further shown that, so long as the leading edge of the growing hyphae is sampled, there is no difference in endohyphal bacterial counts between days 5–7 after fungal growth on a plate, although our data also suggest an increase of bacterial population sizes before this point (Fig. 3). Thus, our current best practice for sampling involves

a gentamicin treatment from day 3 to day 4 of liquid cocultures, plating on water agar at day 4, and sampling bacteria from the growing edge of hyphae ca. 7 days after plating on water agar.

We have not quantified the percent of endohyphal, attached, or free-living bacteria during sampling of the agar core, as our methods do not distinguish these readily. Microscopy from our labs has demonstrated that endohyphal bacteria can be found from the expanding edge of the hyphae during growth on water agar (Spraker, unpublished data), but *Luteibacter* sp. 9143 cannot grow on water agar alone. Our experiments in liquid culture suggest that the gentamicin treatment can selectively kill extrahyphal bacteria before plating and that all subsequently sampled bacteria from these gentamicin treated cultures likely were endohyphal at the time of plating on water agar.

Although gentamicin treatment appears to select for endohyphal bacteria during reassociation between *Pestalotiopsis* and *Luteibacter*, as suggested herein, it is possible that fungal species are differentially tolerant to or have different protection capabilities from gentamicin. Thus we suggest that efficacy of this treatment should be checked when investigating reassociation with different fungal species. It is also important to take note that gentamicin protection will not be effective for bacterial strains that are already resistant to this antibiotic, as is the case with many strains labelled with commonly used Tn7 localized fluorophores or with some transposons used in the production of mutant libraries [12, 14].

In sum, we have presented an assay that can be used to score the ability of bacterial strains to reassociate with fungal hyphae during liquid culture. This assay can be used to determine the host ranges and abilities of a variety of different bacterial strains to reassociate with *Pestalotiopsis* sp. 9143 and potentially the competence of other types of fungi to host bacteria. Our protocol will also be useful in the design and implementation of experiments like TnSeq, to determine the genetic basis of association between endohyphal bacteria and fungi.

Acknowledgments

Financial support for this project was provided by the National Science Foundation (NSF IOS-1354219 to D. A. B., A. E. A., and Rachel E. Gallery). We thank Kayla Arendt, Sarah Araldi-Brondolo, Kevin Hockett, Rachel Gallery, and Justin Shaffer (University of Arizona) for helpful discussion and technical assistance.

References

1. Scherlach K, Graupner K, Hertweck C (2013) Molecular bacteria-fungi interactions: effects on environment, food, and medicine. Annu Rev Microbiol 67:375–397
2. Bonfante P, Anca I-A (2009) Plants, mycorrhizal fungi, and bacteria: a network of interactions. Annu Rev Microbiol 63:363–383
3. Frey-Klett P, Burlinson P, Deveau A, Barret M, Tarkka M, Sarniguet A (2011) Bacterial-fungal interactions: hyphens between agricultural, clinical, environmental, and food microbiologists. Microbiol Mol Biol Rev 75(4):583–609
4. Kobayashi DY, Crouch JA (2009) Bacterial/fungal interactions: from pathogens to mutualistic endosymbionts. Annu Rev Phytopathol 47:63–82
5. Araldi-Brondolo SJ, Spraker J, Shaffer JP, Woytenko EH, Baltrus DA, Gallery RE, Arnold AE (2017) Bacterial endosymbionts: master modulators of fungal phenotypes. Microbiol Spec 5(5). https://doi.org/10.1128/microbiolspec.FUNK-0056-2016
6. Salvioli A, Ghignone S, Novero M, Navazio L, Venice F, Bagnaresi P, Bonfante P (2016) Symbiosis with an endobacterium increases the fitness of a mycorrhizal fungus, raising its bioenergetic potential. ISME J 10(1):130–144
7. Shaffer JP, U'Ren JM, Gallery RE, Baltrus DA, Arnold AE (2017) An endohyphal bacterium (chitinophaga, bacteroidetes) alters carbon source use by fusarium keratoplasticum (f. Solani species complex, nectriaceae). Front Microbiol 8(350). https://doi.org/10.3389/fmicb.2017.00350
8. Lackner G, Hertweck C (2011) Impact of endofungal bacteria on infection biology, food safety, and drug development. PLoS Pathog 7:e1002096
9. Shaffer JP, Sarmiento C, Zalamea PC, Gallery RE, Davis AS, Baltrus DA, Arnold AE (2016) Diversity, specificity, and phylogenetic relationships of endohyphal bacteria in fungi that inhabit tropical seeds and leaves. Front Ecol Evol 4:116
10. Hoffman MT, Arnold AE (2010) Diverse bacteria inhabit living hyphae of phylogenetically diverse fungal endophytes. Appl Environ Microbiol 76:4063–4075
11. Moebius N, Üzüm Z, Dijksterhuis J, Lackner G, Hertwick C (2014) Active invasion of bacteria into living fungal cells. eLife 3:e03007
12. Arendt KR, Hockett KL, Araldi-Brondolo SJ, Baltrus DA, Arnold AE (2016) Isolation of endohyphal bacteria from foliar ascomycota and in vitro establishment of their symbiotic associations. Appl Environ Microbiol 82(10):2943–2949
13. Edwards AM, Massey RC (2011) Invasion of human cells by a bacterial pathogen. J Vis Exp. https://doi.org/10.3791/2693
14. Liberati NT, Urbach JM, Miyata S, Lee DG, Drenkard E, Wu G, Villanueva J, Wei T, Ausubel FM (2006) An ordered, nonredundant library of *Pseudomonas Aeruginosa* strain PA14 transposon insertion mutants. Proc Nat Acad USA 103(8):2833–2838



Chapter 2

Characterizing Mycoviruses

Mahtab Peyambari and Marilyn J. Roossinck

Abstract

A variety of methods for the detection and characterization of fungal viruses are available. For many years, serological and biological assays were used for virus detection. Today, more sensitive methods like polymerase chain reaction, together with sequencing, are widely used to study viruses. Extracting double-stranded (ds) RNA can be a useful approach to detect and study mycoviruses from fungal tissues, as dsRNAs accumulate in infected cells as copies of viral genomes or as replicative intermediates of single-stranded RNA genomes. Here we present a basic protocol for growing fungal strains and isolating dsRNA using cellulose chromatography, followed by molecular diagnostic methods including cDNA synthesis, sequencing, and determination of 5′ ends by primer ligation.

Key words Nucleic acid, Phenol, RNase, DNase, RACE

1 Introduction

Mycoviruses or fungal viruses are common in all major taxa of fungi including *Ascomycota*, *Basidiomycota*, and *Deuteromycota* [1]. Mycoviruses are often associated with latent infection in their hosts, with some exceptions [2–4]. In some cases mycoviruses can impact higher order biological interaction, such as plant thermal tolerance conferred by a mycovirus-infected endophytic fungus [5], or by reducing the pathology of plant-infecting fungi, such as the hypovirus of the chestnut blight fungus *Cryphonectria parasitica* [6]. There is little evidence of natural vector transmission, but mycoviruses are transmitted vertically by spores or horizontally via hyphal anastomosis [7].

Purification of viral particles requires different methods for each virus, usually determined empirically. Protocols have been published for some mycoviruses, but there is no universal method. However, most mycoviruses have either double-stranded (ds) RNA or single-stranded (ss) RNA genomes [8]. Since ssRNA viruses produce a dsRNA structure known as replicative intermediate during their replication, a simple and rapid method to detect RNA viruses (ss/ds RNAs)

Wenbo Ma and Thomas Wolpert (eds.), *Plant Pathogenic Fungi and Oomycetes: Methods and Protocols*, Methods in Molecular Biology, vol. 1848, https://doi.org/10.1007/978-1-4939-8724-5_2,
© Springer Science+Business Media, LLC, part of Springer Nature 2018

in fungi is dsRNA enrichment of total nucleic acids. Large dsRNA molecules are only generated by viruses, so the presence of dsRNA is strong evidence of viral infection [9]. Unique properties of dsRNA molecules including resistance to RNases under high ionic strength conditions, binding to cellulose in ethanol-containing buffers, and relatively high stability compared with ssRNA, make the use of dsRNA practical for the detection and identification of mycoviruses. CF11 cellulose chromatography has been the most useful method to extract dsRNA. The method, first used to isolate dsRNA of phage R17 from *Escherichia coli*, is based on the affinity of cellulose powder for nucleic acids and the specific adsorption of dsRNA at ethanol concentration from 15 to 18% [10]. Here, we present methods to extract dsRNA from fungal tissue using Whatman CF11 cellulose chromatography as described by Dodds et al. [11] with some modifications (Fig. 1), and alternative methods for using type 101 cellulose, which is more readily available. We also include methods to confirm the identity of dsRNAs by resistance to deoxyribonuclease and ribonuclease digestion, characterize mycovirus genomes using random primed RT-PCR (Fig. 2), and determine 5′ ends by primer ligation (Fig. 3).

2 Materials

All buffers are prepared in purified water (18 MΩ-cm), and stored at room temperature unless indicated otherwise.

2.1 Cultivation and Preparation of the Fungus

1. Difco potato dextrose broth (PDB): 24 g/L. Autoclaved in aliquots. For potato dextrose agar (PDA), 15 g/L agar is added before autoclaving. PDA is a general-purpose solid medium used for culturing fungi. It is often used to isolate, grow, and maintain fungi (*see* **Notes 1** and **2**).
2. Shaking incubator, to grow biomass.
3. Buchner funnel, vacuum, and Whatman #1 filter paper to filter mycelia.

2.2 Enrichment of dsRNA

1. Extraction buffer: 0.1 M NaCl, 50 mM Tris–HCl pH 8, 1 mM EDTA pH 8, 1% SDS, 25 mM glycine, 0.1% 2-mercaptoethanol (*see* **Note 3**).
2. Application buffer: 0.1 M NaCl, 50 mM Tris–HCl pH 8, 0.5 mM EDTA pH 8, 25 mM glycine, 16.5% ethanol (*see* **Note 4**).
3. Elution buffer: 0.1 M NaCl, 50 mM Tris–HCl pH 8, 0.5 mM EDTA pH 8.
4. Absolute ethanol.
5. NAE buffer: 0.3 M NaOAc, 0.1 mM EDTA.
6. Liquid nitrogen.

7. Mortar and pestle.
8. 10 mL BioRad Econocolumn/
9. Phenol-chloroform (1:1, weight:weight) saturated with TE (10 mM Tris–HCl (pH 8.0) 0.1 mM EDTA).
10. 70% ethanol in water.
11. 3 M NaOAc (pH 5.5)
12. Whatman CF11 cellulose or Sigma cellulose Type 101 (*see* **Note 5**).
13. 10 × TBE buffer: 89 mM Tris base, 89 mM boric acid, 2 mM EDTA (*see* **Note 6**).
14. Agarose.

2.3 Enzyme Digestion

1. RQ1 RNase-free DNase (Promega).
2. RNase A (10 mg/mL, Sigma, boiled for 10 min and then stored at −20 °C).
3. Pronase E (100 mg/mL, Sigma, stored at −20 °C).

2.4 Reverse Transcription of dsRNA

1. Tagged random primer (*see* **Note 7**).
2. Superscript IV (Invitrogen).
3. Heat block set to 50 °C.
4. 35% guanidine HCl (dissolved in water).
5. Qiagen PCR purification kit.

2.5 PCR Amplification of Viral cDNA and Cloning

1. 10× PCR buffer (100 mM Tris pH 8.3, 500 mM KCl, 15 mM $MgCl_2$).
2. Tag primer (*see* **Note 8**).
3. Taq DNA polymerase (New England Biolabs).
4. PCR machine and appropriate tubes.
5. pGEM-T Easy vector system (Promega) or similar cloning system.

2.6 Determination of 5′ Ends by Primer Ligation (Primer Ligation RACE)

1. 50% Polyethylene glycol (PEG), made fresh.
2. T4 RNA ligase 2 (New England Biolabs).
3. 5′-phosphorylated primer (*see* **Note 9**).
4. Linear acrylamide (Ambion).
5. 2.5 M NH_4OAc (pH 5.5).
6. Complementary primer (*see* **Note 9**).
7. AMV reverse transcriptase (New England Biolabs).
8. Internal primer based on viral sequences.

3 Methods

3.1 Cultivation and Preparation of Fungus for dsRNA Isolation

1. The fungus is cultured on a standard petri dish containing solid medium, and grown at an appropriate temperature (*see* **Note 10**).
2. Five mycelial agar discs from the growing edge of the fungal culture on the agar plate are inoculated to 150 mL broth, and grown at an appropriate temperature with shaking at about 150 rpm, for 3–7 days (*see* **Note 11**).
3. The biomass is harvested by filtration. The liquid culture is poured into the Buchner funnel supporting Whatman #1 filter paper under vacuum. The mycelia are washed with water and, after removing excess moisture by further application of vacuum, are frozen at −20°C overnight, and then lyophilized prior to extraction. Lyophilized mycelia can be kept at −20°C until the dsRNA extraction.

3.2 The dsRNA Isolation (Fig. 1)

1. 200 mg of lyophilized mycelia is ground in liquid nitrogen to a fine powder using a mortar and pestle. The pulverized tissue is transferred to a 50 mL tube containing 10 mL extraction buffer and 10 mL phenol:chloroform (1:1, w:w) saturated with TE buffer (*see* **Note 12**).
2. The sample is mixed by rigorous shaking for 10 min, followed by centrifugation at top speed in a clinical centrifuge.
3. The aqueous phase, containing total nucleic acids, is transferred to a fresh 50 mL tube. Avoid the interface.
4. Ten milliliter of phenol:chloroform are added and **steps 2** and **3** repeated. These steps should be repeated a third time if the aqueous phase is still cloudy and the interface is still significant.

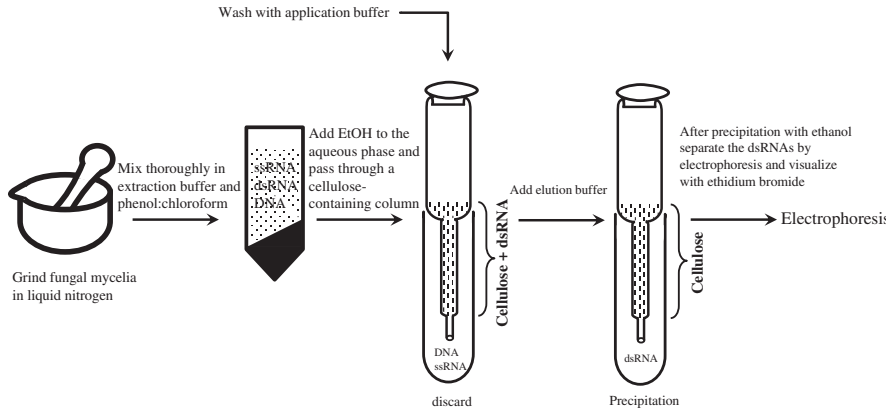


Fig. 1 Schematic of dsRNA isolation from fungal mycelia

5. The final aqueous phase, avoiding any remaining interface, is transferred to a 12 mL Falcon tube. The volume is measured using the graduations on the tube, and room temperature absolute ethanol is added to a final concentration of 16.5% (multiply volume by 0.198) followed by gentle but thorough mixing.
6. Approximately 100 mg CF11 cellulose (Whatman) is placed into a 10 mL BioRad Econocolumn (*see* **Note 13**). This will provide enough cellulose to fill the narrow portion of the column.
7. The extracted nucleic acid from **step 5** is added to the column, the top and bottom are closed, and mixed thoroughly. The top and bottom are removed, and the column is placed in the Falcon tube and centrifuged at low speed for 30 s (*see* **Note 14**). The dsRNA is bound to the cellulose while the aqueous phase contains ssRNA and DNA. The aqueous phase is discarded or retained for other studies requiring DNA and ssRNA.
8. The charged column, with bound dsRNA, is washed with the Application buffer to remove unbound impurities. The bottom of the column is sealed and the column is filled with the Application buffer, the top is closed, and contents mixed thoroughly. The ethanol will build some pressure in the column during mixing so care must be taken to secure both closures of the column. Spin as in **step 7**. The column is washed three or four additional times with the Application buffer.
9. The dsRNA is eluted from the column with 4.5 mL of the Elution buffer (No ethanol!). Buffer is added to the column and mixed, as in **step 7**. The column is placed over a clean 15 mL Corex tube, or equivalent tube that can be used in a high speed centrifuge (*see* **step 11**), and is spun for 1 min at top speed (about 250 × *g*) in the clinical centrifuge.
10. The eluate is retained, 0.5 mL of 3 M NaOAc (pH 5.5) and 10 mL of cold absolute ethanol are added, and the tube is mixed gently. RNA is precipitated by incubation at −20 °C overnight (*see* **Note 15**).
11. The tube is centrifuged for 30 min, at 11,000 × *g* at 4 °C. The supernatant is discarded and the tube is inverted on a fresh Kimwipes to completely drain the pellet. Drying the pellet is finished by leaving the tube right side up. The pellet should be thoroughly dried before resuspension, and this can be sped up by using a vacuum. The pellet is dissolved in 0.5 mL NAE buffer and transferred to a fresh 1.5 mL microcentrifuge tube. Two volumes of 100% ethanol (1 mL) are added, followed by gentle mixing and incubation at −20 °C overnight, or at −80 °C for 30 min (*see* **Note 15**).
12. The tube is centrifuged in a microcentrifuge for 10 min at top speed (e.g., 16,000 × *g*). The supernatant is discarded and the

pellet is washed by vortexing for 30 s with 0.5 mL of 70% ethanol followed by centrifugation for 2 min as above. The supernatant is poured off and the pellet is washed with 0.5 mL 100% ethanol and centrifuge for 2 min as above.

13. The supernatant is thoroughly drained off and the pellet is dried thoroughly at room temperature or by using a centrifugal vacuum. The final pellet is dissolved in 40 μL of 0.1 mM EDTA. The dsRNA is stored at −20 °C for further steps.
14. The dsRNA is checked on a 1.2% agarose gel in 0.5 × TBE by standard electrophoresis methods (*see* **Note 16**).

3.3 Enzyme Digestion

To confirm that the nucleic acid bound and eluted from the cellulose is dsRNA, digestions are performed as follows.

1. For DNA digestion, samples are incubated with 1 unit per 10 μL sample of RQ1 ribonuclease-free DNase at 37 °C for 30 min. One microliter of R1 DNase stop solution is added to terminate the reaction followed by incubation at 65°C for 10 min to inactivate the DNase. This mixture can be loaded directly onto an agarose gel for electrophoresis (*see* **Note 17**).
2. For RNase digestion, 5 μL of dsRNA (about 50 ng) is incubated at room temperature for 30 min with 10 μg of the boiled RNase A in high (0.3 M) and low (0.03 M) NaCl concentrations. DsRNAs are resistant to RNase A at higher salt concentrations. The enzyme is inactivated by incubation with 10 μg of Pronase E for 30 min, followed by extraction with phenol:chloroform, and ethanol precipitation as described above. If the volume is small, add sufficient 0.1 mM EDTA prior to phenol extraction to bring the volume to at least 50 μL. The final pellet is resuspended in 20 μL of 0.1 mM EDTA before loading on an agarose gel.

3.4 Reverse Transcription of dsRNA (Fig. 2)

The method for reverse transcription described here is based on one published previously [12].

1. In a 1.5 mL microcentrifuge tube, 1 μL TE (10 mM Tris pH 8, 10 mM EDTA), 2 μL tagged random primer (Table 1) (20 μM), 2 μg dsRNA, and H_2O are mixed for a total volume of 12 μL. The mixture is placed in boiling water for 2 min to denature the dsRNA.
2. The tube is removed from boiling water and immediately chilled on ice for 2 min.
3. 8 μL of RT mixture [4 μL of 5× superscript IV buffer, 2 μL DTT, 1 μL dNTPs (10 mM), and 1 μL Superscript IV Reverse Transcriptase] is added to the tube. The tube is incubated on ice for 15 min.

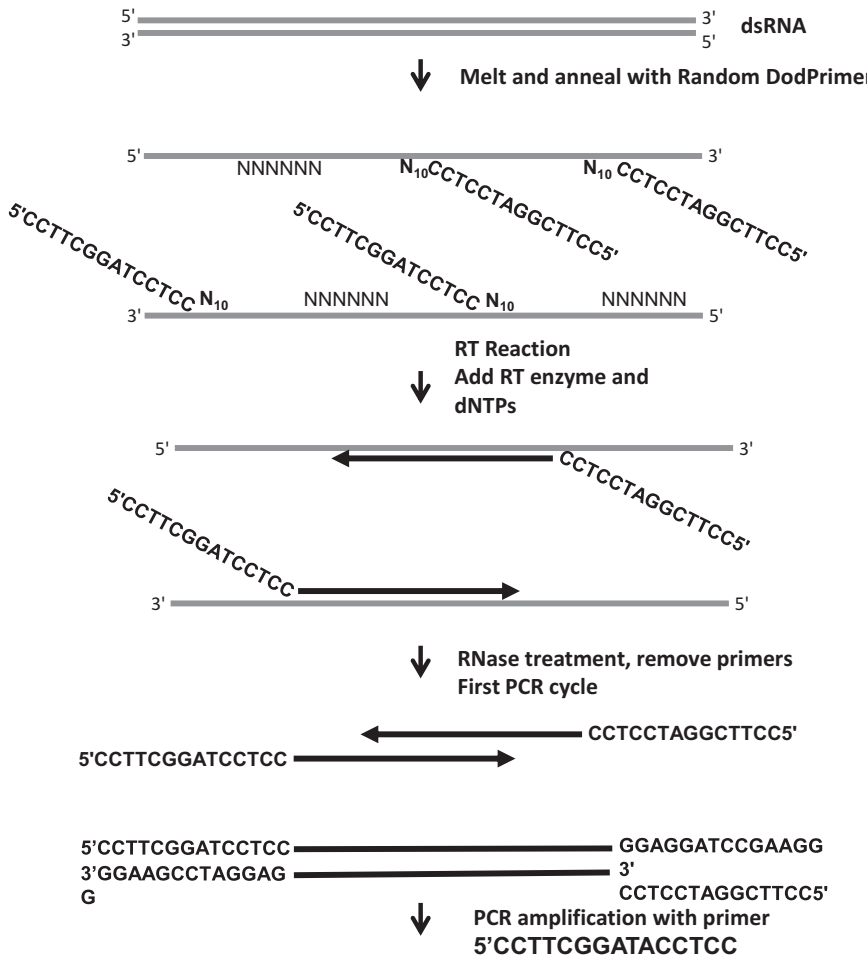


Fig. 2 Schematic outline of generation of cDNA from dsRNA using a random priming method. RNA is shown as a gray line; DNA is shown as a black line

4. The tube is transferred to a 50 °C heating block for 10 min to generate cDNA from viral RNA (*see* **Note 18**).
5. The tube is removed from the heating block, cooled to room temperature and 1 μL boiled RNase A is added. The sample is incubated at room temperature for 15 min.
6. To release the primers from the cDNA, the tube is placed at 85 °C for 2 min, and then immediately 5 volumes of PB buffer (100 μL) is added to 1 volume of cDNA reaction from the QIAquick PCR purification kit (*see* **Note 19**). All of the liquid from the tube is transferred to the Qiagen column. The column is centrifuged for 1 min at top speed of a microcentrifuge.

Table 1
Primers used for RT and PCR for analyzing dsRNA

Primer purpose	Sequence options	Primer example
cDNA synthesis	Random decamer with linker[a]	CCTTCGGATCCTCCNNNNNNNNNN
PCR amplification-cloning[b]	Linker	CCTTCGGATCCTCC
PCR amplification-multiplexing[c]	Tagged linker[d]	XXXXCCTTCGGATCCTCC
Primer ligation	Phosporylated linker	5′P-GGAGGATCCGAATTCAGG
Primer ligation amplification	Complementary primer	CCTGAATTCGGATCCTCC

[a]Linker portion contains restriction enzyme sites
[b]For cDNA that you plan to clone
[c]For cDNA that you plan to sequence directly by high-throughput sequencing
[d]Primer contains a unique 4 nt tag (XXXX) as a barcode. For Illumina sequencing this can be at the 5' end because additional sequences are added in the library preparation process

7. The column is washed with 750 μL of 35% Guanidine HCl. The liquid, which has spun through the column, is discarded.
8. The column is washed by adding 750 μL of the PE buffer from the Qiagen kit and is centrifuged for 1 min at top speed. The liquid is removed, and the column is placed back in the tube and centrifuged for 1 additional min.
9. The column is placed in a clean 1.5 mL tube. To elute the DNA, 30 μL of 0.1 × EB (dilute this from the Qiagen kit) is added, and the column is centrifuged for 1 min at top speed. The cDNA can be used immediately for PCR or stored at −20 °C.

3.5 PCR Amplification of Viral cDNA and Cloning (See Note 20) (Fig. 2)

1. Make a mix for PCR reaction: 25 μL H_2O, 5 μL 10× PCR buffer, 5 μL dNTPs (2 mM), 5 μL Tag primer (20 μM), 5 μL Taq DNA polymerase (2 units/50 μL PCR).
2. 5 μL of Qiagen prepared cDNA is added to PCR reaction in a final volume 50 μL.
3. We typically perform PCR in an Idaho technology machine in capillary PCR tubes under the following parameters: 1 cycle of 94 °C for 1 min, 65 °C for 0 s, 72 °C for 45 s, Slope = 9, followed by 40 cycles of 94 °C for 0 s, 45 °C for 0 s, 72 °C for 30 s, Slope = 5, and final hold of 72 °C for 5 min, and 37 °C for 5 min. However, this protocol can be adapted to any PCR machine. The PCR product is purified using a DNA purification kit such as a Cycle pure kit (Omega, Bio-Tek) or a Qiagen kit.
4. The purified amplified cDNA is cloned in the pGEM-T Easy vector system using T4 DNA ligase according to manufacturer's instructions, and sequenced with a Sanger sequencer (*see* **Note 21**).

3.6 Determination of 5′ Ends by Primer Ligation (See Note 22) (Fig. 3)

The 5′ ends of the viral RNAs can be determined by standard Rapid Amplification of cDNA ends (RACE) [13], but this is sometimes difficult with dsRNA, so an alternative method is given here. Note that with dsRNA 3' RACE is not required as each end can be determined as 5′ ends.

1. A dsRNA primer ligation reaction mixture (total volume of 20 μL) is set up as follows: 5 μL dsRNA, 10 μL of 50% PEG (made fresh within a week), 1 μL 5′-phosphorylated primer (50 μM), 2 μL of 10× T4 RNA ligase 2 buffer, 1 μL T4 RNA ligase 2 (10 units/μL), and H_2O up to 20 μL. The reaction is incubated at 4 °C overnight, or at room temperature for 1 h.
2. To inactivate the ligase, the reaction is heated to 65 °C for 10 min.
3. Add 1 μL linear acrylamide (Ambion) (*see* **Note 23**) to the reaction, followed by precipitation with ethanol overnight at −20 °C by adding 50 μL of 2.5 M NH_4OAc and 100 μL of cold 100% ethanol.
4. The precipitate is centrifuged for 10 min, at 16,200 × *g*. The pellet is washed with cold 70% ethanol, 30% of 0.25 M NH_4OAc as described above for **step 12** of Subheading 3.2, followed by a final wash with cold absolute ethanol. After thorough drying, the final pellet is resuspended in 20 μL of 0.1 mM EDTA.

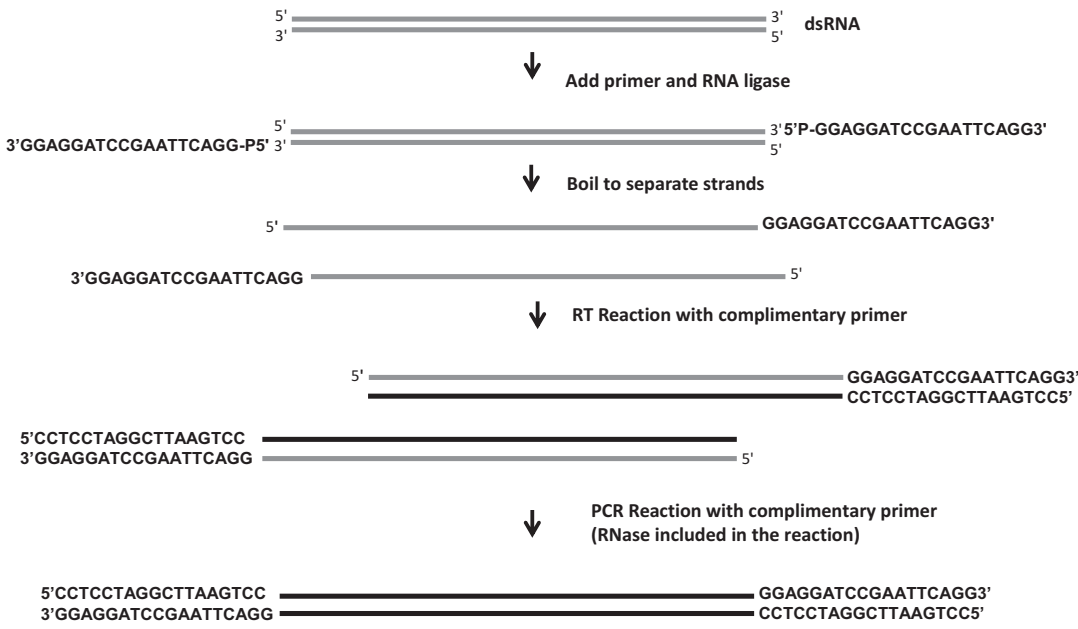


Fig. 3 Schematic outline of the primer ligation method for converting dsRNA to cDNA. RNA is shown as a gray line; DNA is shown as a black line

5. An RT reaction is done as described above (Subheading 3.4) with the following initial components: use 1 μL complementary primer (50 μM), 1 μL random hexamer (50 μM), 1 μL TE, 6 μL H_2O, and 1 μL of primer ligation reaction. Boil for 2 min, then follow **steps 2–4** of Subheading 3.4. It is not necessary to remove the primers.
6. Follow the PCR protocol described above, using the complementary primer and an internal primer based on the viral genome. The PCR products are cloned into a convenient vector, and at least ten clones are analyzed by Sanger sequence analysis.

4 Notes

1. 0.1× PDA is usually used for long-term storage of fungi.
2. A wide range of media are used for growing fungi. All fungi require several specific elements for growth and reproduction. Media should be appropriate for the growth requirements of the particular fungus of study. Most fungi thrive on PDA, but this can be too rich for some species.
3. The extraction buffer contains sufficient detergent to denature proteins and release nucleic acids. The elevated pH and moderate salt concentration (0.1 M) help to reduce ribonuclease activity. 2-mercaptoethanol prevents oxidation and should be added just before use.
4. Ethanol should be added to the Application buffer just before use to prevent reduction of the concentration due to evaporation.
5. CF11 cellulose is difficult to buy; Whatman requires ordering a large quantity to justify producing it. We have used Sigma type 101 cellulose with equal success, but this requires converting the protocol to batch chromatography rather than a column (*see* **Note 13**).
6. TBE is usually made as a 10 × stock solution, and diluted before use.
7. Random primers are a mixture of oligonucleotides representing all possible sequences. Random primers tagged by a specific sequence at the 5′ end are commonly used to introduce primer sequences to facilitate PCR amplification (*see* Table 1). The tagged random decamer 5′-CCT TCG GAT CCT CCN_{10}-3′ is used in our lab. Random hexamers can be purchased or made by DNase digestion of salmon sperm DNA until the average size of the oligos is six nucleotides, as determined by chromatography.
8. PCR is performed using a specific primer matching all the nucleotides of the 5′ tag sequence (5′- CCT TCG GAT CCT CC-3′).

9. In our lab the primer 5′-PO4 –GGAGGATCCGAATTCA GG-3′ is used for ligation and 5′-CCTGAATTCGGA TCCTCC-3′ used as complementary primer for doing RT-PCR (*see* Table 1 for detailed examples of primers).
10. Traditionally, fungi are purified by single spore or hyphal tip isolation and cultured on solid media. Many viruses can be lost during this process, especially when grown on solid media. Single spore and hyphal tip isolation are common strategies to remove viruses from fungal cultures. When studying viruses, it is preferable to use culture grown on minimal solid media.
11. Appropriate temperature and time for growth will vary among fungal species. Some species prefer room temperature (around 25 °C), but others require lower or higher temperatures. Some species grow rapidly (3–7 days), whereas some grow slowly (as long as 4–5 weeks).
12. Phenol, an organic protein denaturant, and chloroform are used to complete protein removal after separation of phases by centrifugation.
13. The method described here can be modified for Sigma Type 101 cellulose, by using batch chromatography instead of the column method described here. In the modified method, **steps 1** to **5** are the same as described for the CF11 protocol, but extracted nucleic acid from **step 5** is thoroughly mixed with Sigma cellulose (~100 mg) in a Falcon tube and centrifuged at low speed (~250 × *g*) for 2 min. The supernatant is poured off rapidly, and 10 mL of the Application buffer is thoroughly mixed with cellulose, followed by centrifugation as above to pellet the cellulose. The supernatant is discarded, and the washing step is repeated for a total four washes. To elute dsRNA, 4.5 mL of elution buffer is added to the pelleted cellulose, followed by thorough resuspension and centrifugation at ~1000 × *g*, and the supernatant is poured off to a clean 15 mL Corex tube. **Steps 10–14** are followed as for the CF11 protocol.
14. Except for the final step, the lowest practical centrifuge time and speed should be used to facilitate liquid passage through the column or to pellet the cellulose in the batch method.
15. Precipitation can be done at −80 °C for at least 30 min, but better results can be achieved by overnight precipitation at −20 °C.
16. It is best to remove potential RNase from gel electrophoresis equipment by soaking the gel rig in 0.1 M NaOH for 1 h, followed by thorough rinsing in water. The NaOH should be made fresh at least every few weeks.
17. The dsRNA may be contaminated with low levels of RNase. Using RQ1 DNase reaction buffer will add Mg to the reaction, which is not necessary for DNase activity, and may cause degradation and smearing of RNA on gel.

18. Superscript IV is a highly processive enzyme that can rapidly catalyze cDNA synthesis to generate long, full-length cDNA fragments. Most other commercially available RTs require longer time to synthesize cDNA. If using a different enzyme follow the manufacturer's recommendations, but increase the temperature to keep the dsRNA from reannealing.
19. E.Z.N.A. Cycle pure kit (Omega, Bio-Tek) can be used as an alternative for Qiagen PCR purification kit.
20. The PCR protocol given here is for a light-cycling machine. The protocol should be modified for the specific machine that is used by the lab. The first cycle should be modified to use a higher annealing temperature and a greater slope than is standard for PCR, followed by normal PCR cycles.
21. cDNA can also be used for deep sequencing. Using primers that incorporate a 4 nt barcode can allow multiplexing of at least 96 samples per Illumina library (*see* Table 1 for primer suggestions).
22. This method can also be used for initial cDNA and amplification of full length dsRNA molecules, as outlined in Fig. 3.
23. Linear acrylamide can be added to enhance the efficiency of ethanol precipitation whenever small amounts of nucleic acids are used.

References

1. Ghabrial SA, Castón JR, Jiang D, Nibert ML, Suzuki N (2015) 50-plus years of fungal viruses. Virology 479-480:356–368
2. Nuss DL (2005) Hypovirulence: mycoviruses at the fungal-plant interface. Nat Rev Microbiol 3:632–642
3. Chiba S, Salaipeth L, Lin U-H, Sasaki A, Kanematsu S, Suzuki N (2009) A novel bipartite double-stranded RNA mycovirus from the white root rot fungus *Rosellinia necatrix*: molecular and biological characterization, taxonomic considerations, and potential for biological control. J Virol 83(24):12801–12812
4. Ahn I-P, Lee Y-H (2001) A viral double-stranded RNA upregulates the fungal virulence of *Nectria radicicola*. Molec Plant-Microbe Interac 14(4):496–507
5. Márquez LM, Redman RS, Rodriguez RJ, Roossinck MJ (2007) A virus in a fungus in a plant --three way symbiosis required for thermal tolerance. Sci 315:513–515
6. Dawe AL, Nuss DL (2013) Hypovirus molecular biology: from Koch's postulates to host self-recognition genes that restrict virus transmission. Adv Virus Res 86:110–147
7. Pearson MN, Beever RE, Boine B, Arthur K (2009) Mycoviruses of filamentous fungi and their relevance to plant pathology. Molec Plant Path 10(1):115–128
8. Ghabrial SA (1998) Origin, adaptation and evolutionary pathways of fungal viruses. Virus Genes 16:119–131
9. Zabalgogeazcoa I, Benito EP, Ciudad AG, Criado BG, Eslava AP (1998) Double-stranded RNA and virus-like particles in the grass endophyte *Epichloë festucae*. Mycol Res 102(8):914–918
10. Franklin RM (1966) Purification and properties of the replicative intermediate of the RNA bacteriophage R17. Proc Natl Acad Sci U S A 55(6):1504–1511
11. Dodds JA, Morris TJ, Jordan RL (1984) Plant viral double-stranded RNA. Annu Rev Phytopathol 22:151–168
12. Roossinck MJ, Saha P, Wiley GB, Quan J, White JD, Lai H, Chavarría F, Shen G, Roe BA (2010) Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. Molec Ecol 19(S1):81–88
13. Schaefer BC (1995) Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. Anal Biochem 227:255–273



Chapter 3

Analysis of Secondary Metabolites from Plant Endophytic Fungi

Jiajia Liu and Gang Liu

Abstract

Filamentous fungi produce a wide variety of secondary metabolites, which are recognized as a rich source of biomolecules with potential medicinal applications. According to chemical structures and biosynthetic pathways, secondary metabolites are mainly divided into four groups, including polyketide, nonribosomal peptides, alkaloids, and terpenes. Owing to their special living environment, plant endophytic fungi potentially produce many bioactive metabolites with diverse structural features. Some of these metabolites have been demonstrated to hold pharmaceutical and ecological significance. In order to identify bioactive compounds, different methods have been developed to isolate secondary metabolites from plant endophytic fungi. In this chapter, we describe the procedures for fungal fermentation, metabolite extraction, and analysis using high-performance liquid chromatography (HPLC). The endophytic fungus *Pestalotiopsis fici*, which was isolated from the healthy branches of *Camellia sinensis* (Theaceae), produces 88 secondary metabolites including 70 new natural products. Combining 74 secondary metabolite clusters revealed by genome sequencing analysis, *P. fici* is used as an example for analysis of secondary metabolites. The procedures could also be adopted for other fungi after small modifications.

Key words Endophytic fungi, Secondary metabolites, Extraction, Preparation

1 Introduction

Fungal secondary metabolites display a variety of structural types and have been recognized as a rich source of natural products with potential medicinal applications [1]. Based on structure and biosynthetic pathway, fungal secondary metabolites are mainly classified as polyketides, nonribosomal peptides, alkaloids, or terpenes. Among them, polyketides contain the highest level of variety [2–7]. Plant endophytic fungi reside in the internal tissues of living plants without causing any apparent disease [8]. Plant endophytes occupy the tissues of their hosts that grow in various environments, especially unusual environments. In order to adapt to the unique living environment, endophytes generally produce a multitude of

Wenbo Ma and Thomas Wolpert (eds.), *Plant Pathogenic Fungi and Oomycetes: Methods and Protocols*, Methods in Molecular Biology, vol. 1848, https://doi.org/10.1007/978-1-4939-8724-5_3,
© Springer Science+Business Media, LLC, part of Springer Nature 2018

bioactive metabolites with diverse structural features [9]. Some of these metabolites display significant pharmaceutical and ecological properties. For example, the anticancer drug paclitaxel that was initially isolated from the pacific yew tree is produced by the endophytic fungi *Taxomyces andreanae* [10] and *Pestalotiopsis microspora* [11]. Also, the cold season fungal endophyte of grass, *Epichloe festucae,* produces a loline alkaloid that can increase the resistance of grasses to various environmental stresses [12].

Usually genes involved in the production of secondary metabolites are clustered in the fungal genome. With the development of gene sequencing technologies, there is accumulating evidence that many secondary metabolic gene clusters exist in fungal genomes and that only a few have been associated with the biosynthesis of known metabolites [13–15]. Therefore, the potential of endophytic fungi for producing bioactive metabolites is immense. *Pestalotiopsis fici* (CGMCC3.15140) was isolated from the healthy branches of *Camellia sinensis* in Hangzhou, China. Through chemical investigations, 88 secondary metabolites including 70 new natural products have been identified [16, 17]. Among them, chloropupukeananin with a highly functionalized tricyclo-[4.3.1.$0^{3,7}$]-decane skeleton shows significant antimicrobial, antitumor, and anti-HIV activities [18, 19]. Therefore, this strain has been regarded to possess a high diversity of natural products. Furthermore, 74 secondary metabolite gene clusters have been identified in the genome of *P.fici* [16, 17]. Sequence analysis of these clusters revealed a large number of core enzymes that are linked to secondary metabolite synthesis: 31 polyketide synthases (PKSs) including four putative PKS-like enzymes, 27 nonribosomal peptide synthases (NRPSs) including 15 putative NRPS-like enzymes, five hybrids of PKS-NRPS, five dimethylallyl tryptophan synthases, 15 terpenoid synthases, seven terpenoid cyclases, and seven fatty-acid synthases. However, only some of polyketides and terpenoids have been identified so far, indicating this strain still have greater potential for mining novel bioactivity natural products. All of these above mentioned 88 compounds were isolated and identified based on a general procedure for metabolite extraction and isolation from filamentous fungi (Fig. 1).

Techniques for secondary metabolite analysis of filamentous fungi have been evolving for many years. In principle, the procedure to isolate secondary metabolites from fungi is different from their isolation from plants or animals, especially when executed at an industrial scale. Nowadays, the preparation of secondary metabolite extracts is less laborious and time-consuming due to widely used chromatographic techniques and other advanced analytical methods. In this chapter, we use *P. fici* as an example to describe the procedures for secondary metabolite analysis of filamentous fungi starting from fungal preservation, to fermentation, and finally analysis by HPLC.

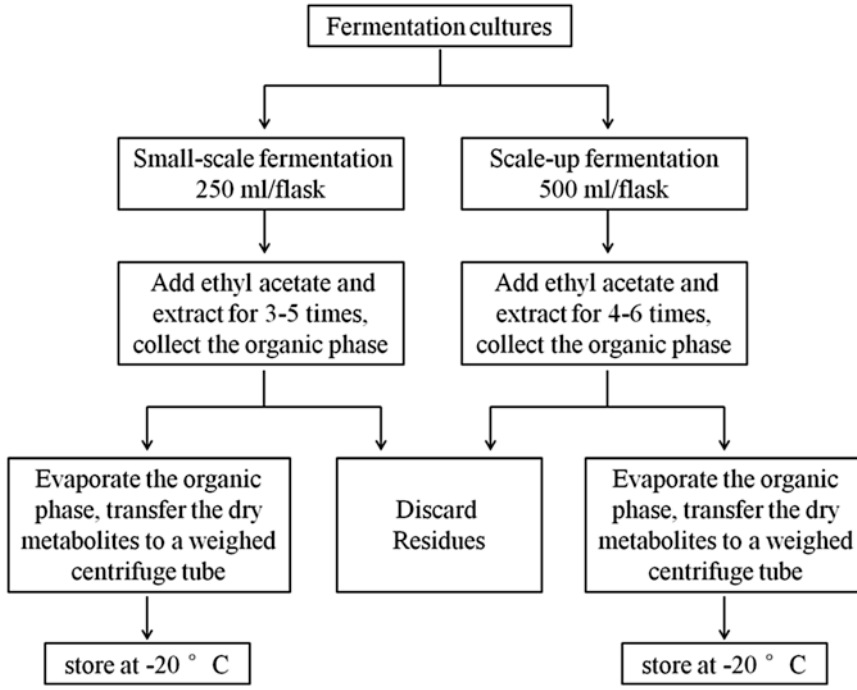


Fig. 1 Flowchart for the extraction of secondary metabolites from plant endophytic fungi

2 Materials

2.1 Media for Preservation and Cultivation of P. fici

Prepare all media using double-distilled water.

1. Sterile glycerol solution: 15% glycerol (w/v) in double-distilled H_2O, sterilized by autoclaving.
2. Potato dextrose agar (PDA): After boiling 200 g sliced washed peeled potato in 1 L double-distilled water for 30 min, strain the broth through a piece of gauze. Then, add 2% glucose (w/v) and 1.6% agar (w/v), and sterilize by autoclaving.
3. Seed medium: 0.4% glucose (w/v), 1% malt extract (w/v), and 0.4% yeast extract (w/v).
4. Potato dextrose broth (PDB): After boiling 200 g washed, peeled, and sliced potato in 1 L double-distilled water for 30 min, strain the broth through a piece of gauze. Then, add 2% glucose (w/v) and sterilize by autoclaving.
5. Rice medium: 30 g white rice is soaked in 50 mL of double-distilled water for small-scale fermentation or 100 g white rice is soaked in 120 mL of double-distilled water for scale-up fermentation. Rice medium is used after sterilization by autoclaving.

6. Mueller-Hinton broth (MHB): 2 g beef extract powder, 1.5 g soluble starch, 17.5 g acid-hydrolyzed casein, add water to 1 L, sterilize by autoclaving.
7. Mueller-Hinton agar (MHA): 2 g beef extract powder, 1.5 g soluble starch, 17.5 g acid-hydrolyzed casein, and 17 g agar, add water to 1 L, sterilize by autoclaving.
8. 15% sterile glycerol: 15 mL glycerol, add double-distilled water to 100 mL, sterilize three times at an interval of 12 h.

2.2 Extraction for P. fici Cultures

1. Reagent: Pure ethyl acetate (EtOAc) is the extraction liquid; anhydrous methanol is the organic solvent for metabolites.
2. Instruments: rotary evaporator for dryness of the metabolites; ultrasonic cleaner for disruption of the mycelia; Termovap Sample Concentrator for dryness of the metabolites.

2.3 Analysis and Preparation of Fungal Metabolites

Sephadex LH-20 column chromatography (200–300 mesh); Silica gel vacuum liquid chromatography (VLC).

2.3.1 Preparation of Metabolites

2.3.2 Analysis of Fungal Metabolites

1. High performance liquid chromatography (HPLC) system (Agilent 1200 Series).
2. Stationary phase for analysis: RP-C18 column (e.g., Agilent ZORBAX, 9.4 × 250 mm, 5 μm).
3. Mobile phase:
   (a) Solvent A: double-distilled water.
   (b) Solvent B: anhydrous methanol.
   (c) Other reagents: CH_2Cl_2; methyl cyanide (MeCN).

2.4 Antimicrobial Activity Assays

1. Reagents: dimethylsulfoxide (DMSO) for dissolving the test samples and positive control drugs; ampicillin and carbendazim as the positive control drugs for antibacterial and antifungal assays respectively; Alamar blue as the indicator of cellular growth.
2. Instruments: microplate reader (TECAN) for detection of the cellular growth; 96-well clear plates for the growth of bacteria and fungi.

3 Methods

3.1 P. fici Preservation

1. Prepare three plates of solid PDA media.
2. Get 1×10^7 fungal spores from the stock and spread them on each plate using an inoculation loop.

3. For preparing fresh spores, incubate the plates at 25–28 °C for 7 days.
4. Wash each plate with 2 mL sterile water thoroughly, filter the spore suspension through miracloth (Millipore) and collect the spores into sterile microfuge tubes and centrifuge for 2 min at 12,000 × *g*. Resuspend the spore pellet with 1 mL sterile water and count the number of spores. Dilute the spore suspension at a concentration of 1×10^8/mL.
5. Centrifuge for 2 min at 12,000 × *g*, resuspend the spores in 300 μL of 15% sterile glycerol (w/v) at a concentration of 1×10^8/mL in sterile microfuge tubes and store them at −80 °C as fungal spore stock solution for long-term storage (*see* **Note 1**).

3.2 Small-Scale Fermentation of P. fici

3.2.1 Prepare the Seed Culture

1. Prepare three plates of solid PDA media.
2. Spread 100 μL of fungal spore stock solution over each plate, and incubate the plates at 25–28 °C for 7 days (*see* **Note 2**).
3. The fungal mycelia on the medium surface are cut into small pieces (0.5 squares each piece) under aseptic conditions; transfer 10–15 agar plugs into a 250 mL Erlenmeyer flask containing 100 mL of seed medium, and incubate at 25–28 °C on a rotary shaker (150–220 rpm) for 5–7 days as the fungal seed culture (*see* **Note 3**).

3.2.2 Small-Scale Fermentation

Small-scale fermentation can be used for both liquid-state and solid-state fermentation.

1. For liquid-state fermentation, prepare a sterile 250 mL flask containing 100 mL of PDB.
2. Transfer 5 mL of the fungal seed culture into the flask containing 100 mL of PDB and incubate at 25–28 °C on a rotary shaker (150–220 rpm) for 7–10 days (*see* **Note 4**).
3. For solid-state fermentation, prepare a sterile 250 mL flask with rice medium (30 g rice in 50 mL double-distilled water) (*see* **Note 5**).
4. Transfer 5 mL of the fungal seed culture and spread over the surface of the rice medium per flask and incubate the flask at 25–28 °C for 30 days (*see* **Note 4**).

3.3 Extraction of Secondary Metabolites

A diagram for the extraction procedure is shown in Fig. 2.

1. After 30 days of fermentation, chop the solid-state fermentation cultures into small pieces (*see* **Note 6**).
2. Add 100 mL of ethyl acetate into the chopped solid-state fermentation culture or liquid-state fermentation culture. Soak the culture in ethyl acetate at room temperature for 24 h (*see* **Note 7**).

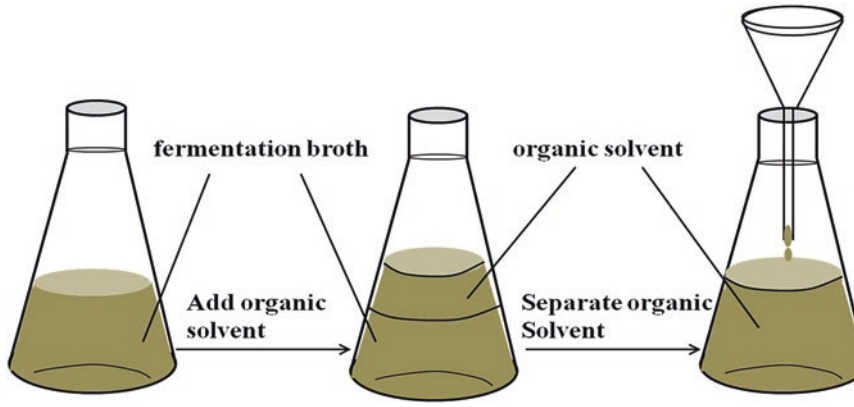


Fig. 2 Extraction of endophytic fungal secondary metabolites

3. Disrupt mycelia by submerging the culture in ethyl acetate and performing sonic disruption using an ultrasonic cleaner for 30 min, at 100 W power.
4. Let the mixture stand until the organic phase and aqueous phase completely separate; this usually takes about 6–8 h (*see* **Note 8**).
5. Recover the organic phase (upper phase). Discard the "debris" in the solid state culture. Add 100 mL of ethyl acetate to re-extract for 3–5 times and pool all extracts together (*see* **Note 9**).
6. Evaporate the organic phase at a temperature ≤40 °C (the room temperature is a little lower for the dryness of sample) in a rotary evaporator, till 2–3 mL of metabolites is left. Transfer it into another weighed microfuge tubes and evaporate with a gentle stream of nitrogen gas using Termovap Sample Concentrator to obtain the dry metabolites (*see* **Note 10**).
7. Determine the weight of the dried sample and store it at −20 °C. Generally, we can harvest about 100 mg sample from this protocol.

3.4 Establishment of the Chromatographic Fingerprint for Small-Scale Fermentation

This method applies to small-scale solid and liquid fermentation. The chromatographic fingerprint from small-scale fermentations is illustrated (Fig. 3).

1. Dissolve 10 mg of the sample from small-scale fermentation in 1 mL of anhydrous methanol.
2. Centrifuge at 12,000 × *g* for 10 min, transfer the supernatant to an autosampler tube for HPLC analysis with the HPLC analysis column (RP-C18 column, Agilent ZORBAX, 9.4 × 250 mm, 5 μm).

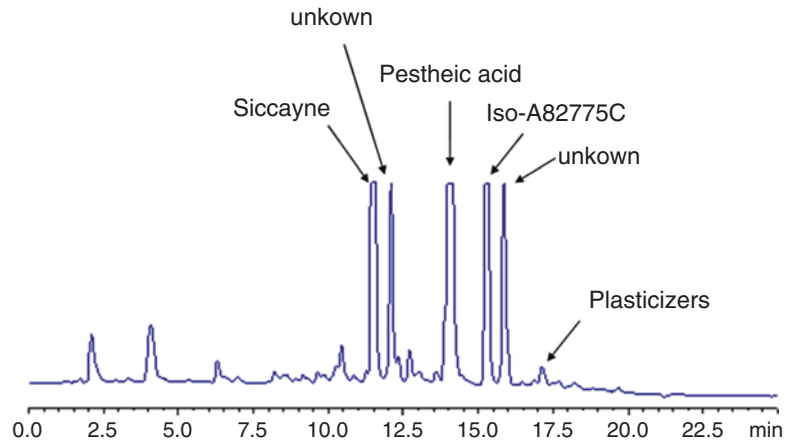


Fig. 3 The chromatographic fingerprint of secondary metabolites from *P. fici* monitoring at 254 nm after 30 days of fermentation in rice medium. Pestheic acid, the main polyketide product from the plant endophytic fungus *P. fici*; Iso-A82775C, a new stereoisomer of the epoxycyclohexane metabolite A82775C; Siccayne, an exocyclic isopropenylallene; Plasticizers, impurities from plastic containers; Unknown,compounds which are not identified

3. Equilibrate the HPLC analysis column with 20% methanol (a 4:1 mixture of solvent A and B). The equilibration time is 2–5 min.
4. Use a linear gradient of 20–100% methanol–H_2O for 18 min, with a flow rate of 1 mL/min, and then wash the column with 100% methanol for 7 min.
5. Detect metabolites using different UV absorption wavelengths, monitoring at 210, 230, 254, and 280 nm.
6. Analyze the peak of different UV absorption wavelengths and establish the chromatographic fingerprint. For example, the chromatographic fingerprint of secondary metabolites of *P. fici* is detected at 254 nm; the different peak times represent different compounds (Fig. 3).

3.5 Preparation of Secondary Metabolites from Scale-Up Fermentation

3.5.1 Scale-Up Fermentation

1. Prepare five PDA plates.
2. Inoculate 100 μL of *P. fici* spore stock solution onto each PDA plate, and incubate the plates at 25–28 °C for 7 days.
3. Cut the agar plugs into small pieces (0.5 squares each piece) under aseptic conditions; transfer 15–20 pieces to a sterile 500 mL flask containing 200 mL of PDB, and incubate at 25–28 °C on a rotary shaker (150–220 rpm) for 5–7 days (*see* **Note 3**), to obtain the seed culture.

4. For liquid-state fermentation, prepare 20 sterile 500 mL flasks, each containing 200 mL of PDB. Transfer 10 mL of the seed culture into each flask and incubate at 25–28 °C on a rotary shaker (150–220 rpm) for 7–10 days (*see* **Notes 4** and **11**).
5. For solid-state fermentation, prepare 20 sterile 500 mL flasks, each containing solidified rice medium prepared with 100 g rice in 120 mL double-distilled H_2O. Transfer 10 mL of fungal seed culture into each flask and spread it over the surface of the rice medium, and incubate at 25–28 °C for 30 days (*see* **Notes 4** and **12**).

3.5.2 Extraction of Secondary Metabolites

1. For liquid fermentation, add 200 mL ethyl acetate to each flask.
2. For solid-state fermentation, chop the fermentation culture into small pieces. Add 200 mL ethyl acetate (make sure that ethyl acetate submerges the mycelia) to each flask (*see* **Note 13**).
3. Soak the culture in ethyl acetate for 24 h. Then disrupt mycelia with an ultrasonic cleaner for 30 min and let the mixture stand for 6–8 h until the organic phase and aqueous phase completely separate. Recover the organic phase (upper phase). Add 100 mL of ethyl acetate to re-extract for 4–5 times and pool all extracts together (*see* **Note 8** and **9**).
4. Evaporate the organic phase under reduced pressure (using a rotary evaporator) at a temperature below 40 °C till 8–10 mL organic phase is left. Transfer it into another weighed microfuge tube and evaporate with a gentle stream of nitrogen gas using Termovap Sample Concentrator to get the dry metabolites and determine the weight of the sample. Generally, about 10 g sample can be obtained from this protocol. Redissolve the dry metabolites with 8–10 mL ethyl acetate and transfer it to microfuge tubes to obtain a crude extract (*see* **Note 10**).
5. The extract is fractionated by silica gel vacuum liquid chromatography (VLC) using petroleum ether–EtOAc gradient elution [19].
6. The fraction (about 180 mg) eluted with 30% EtOAc is separated by Sephadex LH-20 column chromatography (CC) (200–300 mesh) using 1:1 CH_2Cl_2–MeOH as eluent (*see* **Note 14**).
7. Each of the resulting subfractions is combined and further separated by a preparative column (RP-C18 column, Agilent ZORBAX, 9.4 × 250 mm, 5 μm; 50% MeCN in H_2O for 2 min, followed by 50–55% MeCN for 25 min; 2 mL/min) to obtain the pure metabolite.
8. Store the pure metabolite for further analyses by liquid chromatography–mass spectrometry (LC/MS) and nuclear mag-

netic resonance (NMR) to identify the accurate structure of the metabolite (*see* **Note 15**).

3.6 Biological Activity Assays

Antibacterial and antifungal assays are conducted in triplicates following the recommendations by National Center for Clinical Laboratory Standards (NCCLS).

3.6.1 Antibacterial Activity Assays

1. The human pathogenetic gram-positive bacteria *Enterococcus faecium* (CGMCC 1.2025) and *Enterococcus faecails* (CGMCC 1.2135) were obtained from the China General Microbial Culture Collection (CGMCC). The IC50/MIC of the positive control drug against some bacteria is listed in Table 1 (*see* **Note 16**).
2. The preserved bacteria are diluted and grown on Mueller-Hinton agar (MHA) at 37 °C for 24 h to get a single colony.
3. Isolate one bacterial clone and incubate in 5 mL of MHB medium at 37 °C for 24 h. The final cell concentration is diluted to 1×10^6 cells/mL.
4. Make serial dilutions of the test samples (20 mg/mL as stock solution in DMSO) in MHB to attain final concentrations of 200 μg/mL, 100 μg/mL, 50 μg/mL, and 25 μg/mL (*see* **Note 17**).
5. The positive control drug ampicillin (4 mg/mL as a stock solution in H_2O) is first diluted to 40 μg/mL of double-distilled water and then a serial dilution to attain final concentrations of 40 μg/mL, 20 μg/mL, 10 μg/mL, 5 μg/mL, and 2.5 μg/mL in double-distilled water is made.
6. The test samples, the positive control drug, and the bacteria suspension are transferred into the 96-well clear plates in triplicate, achieving a final volume of 100 μL.
7. For antibacterial assay, the absorbance at 595 nm is measured using a microplate reader after incubation at 37 °C for 12–24 h using ampicillin as the positive control (*see* **Note 18**).
8. Inhibition activity of the metabolite is calculated by plotting against the concentration to extrapolate the IC_{50} (half maximal inhibitory concentration). The IC_{50}/MIC of the positive control drug against some bacteria is listed in Table 1.

3.6.2 Antifungal Activity Assays

1. The plant pathogenic fungus *Fusarium oxysporum* (CGMCC 3.2830) was obtained from the China General Microbial Culture Collection (CGMCC) (*see* **Note 16**).
2. The preserved fungus is diluted and grown on PDA at 28 °C for 48 h to get a single colony.
3. Isolate one fungal clone and incubate in 5 mL of PDB medium at 28 °C for 96 h. Harvest spores with spore filters. The final spore concentration is 1×10^4 spores/mL.

Table 1
IC_{50}/MIC of various positive control drugs for some bacteria and fungi

	Name	No.	Abbreviation	Positive control drugs	IC_{50}/MIC(μg/mL)
Bacteria	*Enterococcus faecium*	CGMCC 1.2025	EFm	Daptomycin	0.5–1
	Enterococcus faecalis	CGMCC 1.2135	EFs	Daptomycin	0.25–4
	Staphylococcus aureus COL	CGMCC 1.2465	SA	Ampicillin	0.021 ± 0.003/0.16
	Bacillus subtilis	ATCC 6633	BS	Gentamicin	0.615 ± 0.011/1.25
	Streptococcus pneumoniae	CGMCC 1.1692	SP	Ampicillin	0.380 ± 0.029/10.00
	Escherichia coli	CGMCC 1.2430	EC	Gentamicin	1.621 ± 0.114/2.5
	Pseudomonas aeruginosa 2033E	CGMCC 1.2387	PA	Gentamicin	1.643 ± 0.025/5.00
Fungi	*Aspergillus fumigatus*	CGMCC 3.5835	*A. fumigatus*	Ketoconazole Amphotericin B	1.169 ± 0.095/2.5 2.996 ± 0.225/5.0
	Aspergillus flavus	CGMCC 3.0951	*A. flavus*	Ketoconazole Amphotericin B	0.244 ± 0.019/0.625 0.889 ± 0.041/2.5
	Fusarium oxysporum	CGMCC 3.2830	FO	Ketoconazole Carbendazim	9.913 ± 0.134/ 0.439 ± 0.008/0.625
	Fusarium moniliforme	CGMCC 3.2835	FM	Ketoconazole Carbendazim	4.092 ± 0.078/ 0.398 ± 0.009/0.625
	Gibbberella zeae	CGMCC 3.2873	GZ	Ketoconazole Carbendazim	6.025 ± 1.235/ 0.578 ± 0.007/2.50
	Alternaria longipes	CGMCC 3.2875	AL	Ketoconazole	0.108 ± 0.005/0.625
	Pyricularia oryzae	CGMCC 3.3283	PO	Ketoconazole	2.321 ± 0.037/5.00
	Verticillium albo-atrum	CGMCC 3.4306	VA	Ketoconazole Carbendazim	15.054 ± 2.387/20.00 0.470 ± 0.009/0.625
	Fusarium culmorum	CGMCC 3.4595	FC	Ketoconazole	13.971 ± 0.245/20.00
	Fusarium nivale	CGMCC 3.4600	FN	Ketoconazole Carbendazim	0.63 ± 0.13/20.00 0.490 ± 0.034/0.625

Table 2
Antifungal assays

	Blank control	Negative control	Positive control	Sample
PDB	89 µL	0 µL	0 µL	0 µL
Suspensions of spores	0	89 µL	89 µL	89 µL
Samples	0	0	0	1 µL
Positive control drugs	0	0	1 µL	0
DMSO	1 µL	1 µL	0	0
Alamar blue	10 µL	10 µL	10 µL	10 µL

4. Make serial dilutions of the test samples (20 mg/mL as stock solution in DMSO) in MHB or PDB to attain final concentrations of 200 µg/mL, 100 µg/mL, 50 µg/mL, and 25 µg/mL (*see* **Note 17**).
5. The positive control drug carbendazim (4 mg/mL as a stock solution in DMSO) is first diluted to 40 µg/mL with DMSO and then make a serial dilution with DMSO to attain final concentrations of 40 µg/mL, 20 µg/mL, 10 µg/mL, 5 µg/mL, and 2.5 µg/mL.
6. The test samples, the positive control drug, and the fungal spore suspension are transferred into the 96-well clear plates in triplicate, achieving a final volume of 100 µL. A detailed setup of the antifungal assay is described in Table 2. Alamar blue (10 µL of 10% solution) is added to each well as an indicator of the cellular growth.
7. For antifungal assay, the fluorescence intensity is measured using a microplate reader at 544/590 nm after incubation at 28 °C for 36–72 h (*see* **Note 18).**
8. Inhibition activity of the metabolite is calculated by plotting against the concentration to extrapolate the IC_{50}. The IC_{50}/MIC of the positive control drug against some fungi is listed in Table 1.

4 Notes

1. The volume of 15% sterile glycerol (w/v) to be used is dependent on the amount of spore pellet, usually 100 µL of glycerol for a plate. At −80 °C, the fungal spores can be kept alive for 2–3 years. For activating the fungal strains, a single colony is generally isolated and used for subsequent experiments.

2. For different fungi, the incubation time can be prolonged or shortened from day 5 to day 10 since the sporulation time of different fungi varies. For most fungi, a 7-day incubation is commonly used for sporulation. Since the growth rates of different fungi are different and the secondary metabolites could be produced either at the late stage of exponential growth phase or at the stationary growth phase, the harvest time should be regulated according to the production of target metabolites. For most filamentous fungi, a 7-day harvest is commonly used for secondary metabolite extraction. However, longer incubation, up to 20 days, is used in some cases.
3. For slow-growing fungi, the incubated amount of agar plugs can be increased to 20–25 agar plugs. Generally, fungal growth reaches the stationary growth phase after 48 h of fermentation in liquid medium.
4. The fungal growth status should be monitored to ensure that there is no contamination during fermentation. Contamination can be identified by observation under a microscope. Generally, fungal growth reaches the stationary growth phase after 48 h of fermentation in liquid medium.
5. The rice medium consists of 30 g rice and 50 mL of double-distilled H_2O for small-scale fermentation of *P. fici*. The rice medium solidifies after sterilization by autoclaving.
6. In solid-state fermentation the mycelia and rice medium are combined after 30 days of fermentation. For better extraction the solid-state cultures should be chopped into small pieces. The cultures are chopped in the fermentation flask with a large pair of scissors or large forceps.
7. For liquid-state fermentation, 100 mL ethyl acetate is added into the cultures. For solid-state fermentation, the volume of ethyl acetate is about 100–150 mL; make sure that the ethyl acetate submerges the cultures.
8. For liquid-state fermentation, the mixture of ethyl acetate and cultures is transferred into a separatory funnel and allowed to stand until the organic phase and aqueous phase completely separates.
9. For liquid-state fermentation, pour out the organic phase after the lower phase flows out from the bottom of the separatory funnel. For solid-state fermentation since the mycelia and the rice are in solid state, the organic phase can be poured out directly.
10. The water bath temperature for the rotary evaporator should be under 40 °C to avoid decomposition of metabolites.
11. For extracting lower-yield compounds from liquid-state fermentation, the total volume is 3–5 L.

12. For solid-state fermentation, the rice medium consists of 100 g rice and 120 mL of double-distilled H_2O; the rice medium solidifies after sterilization. For extracting the lower-yield compounds, the total weight of rice is 3–5 kg.
13. For solid-state fermentation the volume of ethyl acetate is about 200–300 mL, make sure that the ethyl acetate submerges the cultures.
14. The fraction eluted with 30% EtOAc is one of the fractions. The separation of other fractions can also be done in this way. The flow rate is not fixed; we usually use the maximum velocity to wash the column and collect the fractions. The flow rate of sample injection is usually one drop per 4–5 s.
15. The LC/MS and NMR are used for further analyses to identify the structure of the metabolites.
16. As the second leading cause of hospital-acquired infections, Enterococcus species cause a wide variety of infections including urinary tract infections, bloodstream infections, and infective endocarditis. *Enterococcus faecium* is a vancomycin- and ampicillin-resistant bacterium. You can also use other bacteria or fungi. Select different pathogenic bacteria or fungi based on the purpose of your study.
17. According to the solubility of the metabolite, the initial concentration and the serial dilutions of the metabolite can be downregulated.
18. The exact incubation time of bacteria and fungi is dependent on the color change of the negative control. The optimum time is when the pink color disappears.

References

1. Hoffmeister D, Keller NP (2007) Natural products of filamentous fungi: enzymes, genes, and their regulation. Nat Prod Rep 24:393–416
2. Beck J, Ripka S, Siegner A, Schiltz E, Schweizer E (1990) The multifunctional 6-methylsalicylic acid synthase gene of *Penicillium patulum*. Its gene structure relative to that of other polyketide synthases. Eur J Biochem 192:487–498
3. Liu J, Farmer JD Jr, Lane WS, Friedman J, Weissman I, Schreiber SL (1991) Calcineurin is a common target of cyclophilin-cyclosporin a and FKBP-FK506 complexes. Cell 66:807–815
4. Martín JF, Ullán RV, García-Estrada C (2010) Regulation and compartmentalization of β-lactam biosynthesis. Microb Biotechnol 3:285–299
5. Shweta S, Zuehlke S, Ramesha BT, Priti V, Mohana KP, Ravikanth G et al (2010) Endophytic fungal strains of *Fusarium solani*, from *Apodytes dimidiata* E. Mey. ex Arn (Icacinaceae) produce camptothecin, 10-hydroxycamptothecin and 9-methoxycamptothecin. Phytochemistry 71:117–122
6. Panaccione DG, Coyle CM (2005) Abundant respirable ergot alkaloids from the common airborne fungus *Aspergillus fumigatus*. Appl Environ Microbiol 71:3106–3111
7. Daum M, Herrmann S, Wilkinson B, Bechthold A (2009) Genes and enzymes involved in bacterial isoprenoid biosynthesis. Curr Opin Chem Biol 13:180–188
8. Strobel G, Daisy B, Castillo U, Harper J (2004) Natural products from endophytic microorganisms. J Nat Prod 67:257–268

9. Strobel G, Daisy B (2003) Bioprospecting for microbial endophytes and their natural products. Microbiol Mol Biol Rev 67:491–502
10. Stierle A, Strobel G, Stierle D (1993) Taxol and taxane production by *Taxomyces andreanae*, an endophytic fungus of Pacific yew. Science 260:214–216
11. Metz AM, Haddad A, Worapong J, Long DM, Ford EJ, Hess WM, Strobel GA (2000) Induction of the sexual stage of *Pestalotiopsis microspora*, a taxol-producing fungus. Microbiology 146:2079–2089
12. Wilkinson HH, Siegel MR, Blankenship JD, Mallory AC, Bush LP, Schardl CL (2000) Contribution of fungal loline alkaloids to protection from aphids in a grass-endophyte mutualism. Mol Plant-Microbe Interact 13:1027–1033
13. Keller NP, Turner G, Bennett JW (2005) Fungal secondary metabolism—from biochemistry to genomics. Nat Rev Microbiol 3:937–947
14. Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S et al (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. Nature 438:1105–1115
15. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ et al (2005) The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature 434:980–986
16. Liu L (2011) Bioactive metabolites from the plant endophyte *Pestalotiopsis fici*. Mycology 2:37–45
17. Wang X, Zhang X, Liu L, Xiang M, Wang W, Sun X et al (2015) Genomic and transcriptomic analysis of the endophytic fungus *Pestalotiopsis fici* reveals its lifestyle and high potential for synthesis of natural products. BMC Genomics 16:28
18. Liu L, Li Y, Li L, Cao Y, Guo L, Liu G, Che Y (2013) Spiroketals of *Pestalotiopsis fici* provide evidence for a biosynthetic hypothesis involving diversified Diels-Alder reaction cascades. J Org Chem 78:2992–3000
19. Liu L, Liu S, Jiang L, Chen X, Guo L, Che Y (2008) Chloropupukeananin, the first chlorinated pupukeanane derivative, and its precursors from *Pestalotiopsis fici*. Org Lett 10:1397–1400



Chapter 4

Protocols for Investigating the Leaf Mycobiome Using High-Throughput DNA Sequencing

Shawn P. Brown, Devin R. Leopold, and Posy E. Busby

Abstract

High-throughput sequencing of taxon-specific loci, or DNA metabarcoding, has become an invaluable tool for investigating the composition of plant-associated fungal communities and for elucidating plant–fungal interactions. While sequencing fungal communities has become routine, there remain numerous potential sources of systematic error that can introduce biases and compromise metabarcoding data. This chapter presents a protocol for DNA metabarcoding of the leaf mycobiome based on current best practices to minimize errors through careful laboratory practices and validation.

Key words Fungal leaf endophytes, Next-generation sequencing, Metabarcoding, Mycobiome, Amplicons, Phyllosphere, Phytobiome, Sample bias, Primer bias, Mock community

1 Introduction

Fungal ecology has taken a monumental step forward with the advent of high-throughput DNA sequencing, which allows the simultaneous sequencing of millions of PCR-amplified gene fragments (amplicons). Using this technology to sequence amplicons from many samples in parallel while targeting taxonomically informative loci, a process also known as DNA metabarcoding, it is now possible to directly query unculturable and novel fungal communities to gain new insights into plant–fungal interactions. Dramatic reductions in cost and widespread availability of this technology have facilitated the democratization of DNA sequencing-based ecology [1] and many researchers now view high-throughput sequencing as an essential tool in their methodological toolbox. However, despite the seemingly straightforward process of obtaining sequences from environmental samples [2–4], there remain many opportunities to introduce errors and bias that can affect the balance of data across samples and among fungal taxa. Every step of the process, beginning with sample collection, has the potential

Wenbo Ma and Thomas Wolpert (eds.), *Plant Pathogenic Fungi and Oomycetes: Methods and Protocols*, Methods in Molecular Biology, vol. 1848, https://doi.org/10.1007/978-1-4939-8724-5_4,
© Springer Science+Business Media, LLC, part of Springer Nature 2018

to introduce biases that can then compound at each step. Although some biases can be identified and controlled for, for relatively unknown systems and taxa it can be difficult to identify problems simply because there is often a dearth of information about the communities under investigation.

In this chapter, we provide a protocol for sample preparation and amplicon library generation for investigating the leaf mycobiome, or the communities of fungal symbionts that live on and in plant leaves. The leaf mycobiome has been described as "hyperdiverse" [5] and often includes many novel taxa [6] whose ecologies and functions are unknown, making them challenging to study. Our protocol begins with sample collection and follows a flexible two-step PCR process for amplicon library generation [7], in which an initial PCR reaction amplifies template DNA with gene primers and a second PCR reaction adds sequencing adapters and sample-specific molecular identifiers. Notes are included to provide additional details, alternative approaches, and to emphasize techniques for minimizing error and user-generated biases.

2 Materials

2.1 Sample Collection and Processing

1. Individually labeled zip-top bags for leaf collection.
2. Cooler containing ice.
3. 70% EtOH (ethanol).
4. 1 L beakers.
5. Triton X-100.
6. Sterile water.
7. Hole punch or cork borer.

2.2 Genomic DNA Extraction

1. Plant or soil DNA extraction kits and required equipment and supplies (*see* kit protocols for these requirements).
2. Bead mill homogenizer (*see* **Note 1**).
3. DNA quantification equipment (e.g., Qubit or NanoDrop).

2.3 Amplification of Target Genes: PCR, and Sequencing

1. Stage 1 and 2 PCR primers specific to the targeted locus (*see* **Note 2**) and sequencing platform (i.e., Illumina MiSeq, *see* **Note 3**).
2. PCR reagents (*see* **Note 4**).
3. PCR product cleanup kits (96-well plate based cleanup and normalization kits such as SequalPrep; *see* **Note 5**).
4. Sequencing reagent kits—most sequencing facilities will provide this encompassed in the user cost.

2.4 Additional Supplies and Equipment

Sterile 96-well PCR plates, reagent reservoirs, a variety of sterile DNase/RNase-free low-retention filter micropipette tips, sterile microcentrifuge tubes, thermal cyclers with 96-well capability, gel electrophoresis equipment and supplies, reagent reservoirs, pipettes (multichannel pipettes preferred), sterile disposable gloves, and any additional equipment/supplies required for kit-based extraction and PCR cleanup kits (*see* kit protocols).

3 Methods

3.1 Sample Collection

1. Sample a minimum of three randomly (or haphazardly) selected leaves at a fixed height, at equal distances (e.g., 120°) around the plant to control for within-plant variation [8–10]. In some cases, depending on the experimental question, leaves with particular phenotypic traits (e.g., herbivore damage, necrotic tissue) may be avoided or targeted (*see* **Note 6**). Sampling leaves of the same age to prevent the introduction of a confounding phenological effect [11] is recommended. Monitor budburst and tag target leaves for later collection, or standardize sampling by the Leaf Plastochron Index (LPI) [12, 13], a measure of leaf age based on morphological development.
2. Place leaves into clean, individually labeled zip-top plastic bags and store immediately on ice or at 4 °C. Process leaves within 24 h to prevent fungi with saprobic capabilities from proliferating (*see* **Note 7**).

3.2 Sample Processing

1. In the laboratory, manually agitate leaves for 1 min in sterile ddH_2O containing Triton X-100 (1 mL Triton X-100 in 1 L ddH_2O), a surfactant that will remove spores and hyphal fragments on the leaf surface [14]. Next, rinse leaves by manually agitating for 30 s in sterile ddH_2O, in each of three consecutive washes (*see* **Note 8**).
2. Sub-sample a standardized mass of leaf material across sample units. The mass needed for DNA extraction varies with the extraction kit, but generally ranges from 0.25 g to 0.5 g (*see* extraction kit protocols). Excise an appropriate number of leaf discs (equal to desired mass) from the sampled leaves with a handheld hole-punch or a cork borer on a rubber mat (clean tool between samples by rinsing in 70% EtOH). If leaf tissue will not be used for DNA extraction immediately, store frozen at −20 °C for ~6 months or at −80 °C for ~3 years.

3.3 DNA Extraction

1. Genomic DNA extraction should follow kit protocols and include manual disruption on a ball mill homogenizer (*see* **Note 1**).

2. Examine DNA quantity and quality for all samples using fluorometic (e.g., Qubit) or spectrophotometric (e.g., NanoDrop) methods. For samples that were poorly extracted (consisting of low DNA concentrations, coextraction of PCR inhibitory compounds, or protein coextraction), reextract DNA or clean using a DNA clean up kit or DNA precipitation protocols (e.g., precipitation using isopropanol for low concentrations).
3. Store DNA in labeled 96-well 0.2 mL PCR plates or racks of 8-strip PCR tubes. For every 96 samples, a minimum of one well/tube should be reserved for sterile molecular grade water to be used as a negative control [15] with no fewer than four negative controls per sequencing reaction. Additionally, a "mock community" containing a known quantity of DNA from a mixture of known fungal isolates should be included as a positive control (*see* **Note 9**). Template DNA plates can be stored in the freezer (−20 °C) for up to 2 years or archived frozen (−80 °C) for long-term storage.

3.4 Amplicon Library Generation

1. Optimize PCR parameters: Use a random subset of at least eight samples to test parameters for stage 1 and 2 PCR (below), checking for robust amplification and nonspecific amplification by gel electrophoresis. To limit bias and incorporation of PCR errors, the number of cycles should be minimized. In addition, temperature ramp rate should be limited to 1 °C/s and the PCR enzyme manufacturer's recommended extension time should be increased by 15 s to limit formation of chimeras during PCR [16]. Other parameters that can be varied include annealing temperature, template concentration, and reaction volume (*see* **Note 10**). Other parameters, such as denaturation and extension temperature are determined by the polymerase and buffers used (*see* manufacturer's documentation).
2. First-stage PCR (~25–30 cycles): If practical, given the numbers of samples being processed, check for amplification in all stage 1 PCR reactions by gel electrophoresis. Otherwise, test at least one sample per column for each 96-well plate plus positive and negative control wells.
3. Optional: Cleanup after first-stage PCR (removes PCR buffer, remaining nucleotides, primer dimers, and polymerase). In some cases, primary PCR product includes an abundance of unused primers, which can increase the generation of primer dimers and amplification of short nonfungal amplicons during secondary PCR. To prevent these dimers from proliferating in the stage 2 PCR, primary PCR products should be cleaned prior to use as secondary PCR template (*see* **Note 5**).
4. Second-stage PCR (~5 cycles, *see* **Note 3**): Use gel electrophoresis to check for amplification of one sample per column for

each 96-well plate plus positive and negative control wells. Note that the second-stage PCR product will contain primary PCR product, so these primary amplicons may be visible via gel electrophoresis even if the second-stage PCR fails. When visualizing primary and secondary PCR products simultaneously, the secondary PCR product should be noticeably larger (i.e., the expected size increase should equal the length of overhanging sequencing adapters and molecular identifiers from the secondary PCR).

5. Cleanup, normalization, and pooling: Use a plate-based cleanup and normalization kit (e.g., SequalPrep) to bind a fixed quantity of PCR product within each well and to remove excess DNA, primer dimer, and remaining PCR reagents. After elution, combine equal volumes of the normalized PCR product from each well into a microcentrifuge tube to create the final amplicon library.
6. Library quality control: Before sequencing, ensure that the library contains amplicons of the expected length and is free of residual primer dimers or other short DNA fragments (<100 bp). Precise measurement of the amplicon size distribution can be achieved with a BioAnalyzer or TapeStation. If short fragments are detected (indicating a problem with earlier cleanup steps), an additional magnetic-bead-based size-selection will be needed. The molarity of the library is then determined using qPCR or a BioAnalyzer and adjusted to the concentration required by the sequencing facility. In most cases, final quantification, along with verification of amplicon size distribution, will be performed by the sequencing facility. However, at a minimum, the library concentration should be fluorometrically quantified (e.g., Qubit) and adjusted to the concentration required by the sequencing facility, usually ~10 nM for Illumina MiSeq.

3.5 Sequence Processing and Bioinformatics

A complete discussion of the computational methods used to convert high-throughput sequencing data into fungal community data for downstream ecological analyses is beyond the scope of this chapter. Fortunately, there are many computational tools and bioinformatics pipelines designed explicitly for this purpose (e.g., DADA2 [17], mothur [18], PipeCraft [19], PIPITS [20], QIIME [21], and UPARSE [22]), each of which is accompanied by tutorials and documentation. Irrespective of the choice of computational tools, any high-throughput sequencing pipeline will involve steps to filter low quality sequences and methodological artifacts (e.g., chimeras) and ultimately to assign sequences into operational taxonomic units (OTUs) based on sequence similarity. While robust patterns of community composition should be

recovered regardless of OTU clustering methods chosen, it is important to recognize that the assignment of sequences to OTUs will vary among clustering algorithms (e.g., agglomerative vs. greedy clustering [23]). In addition, because many tools were initially designed for bacterial 16S amplicon sequencing, default parameters in a sequence processing pipeline should be modified for fungal data (e.g., ITS amplicons). Most importantly, no default pipeline should be used without substantial data exploration and parameter optimization.

4 Notes

1. Disrupting leaf tissue with a bead mill homogenizer (i.e., bead beating) is critical for ensuring lysis of fungal cells within leaves, yet is not always included in plant and soil DNA extraction kit protocols. Bead beating can be incorporated into the first step of most DNA extraction kit protocols by adding the sample, initial lysis buffer, and 3–5 small (~2 mm) beads (stainless steel or zirconium oxide) into a reinforced screw cap tube. Samples should be homogenized at high speed in the bead mill for at least 1 min. Alternatively, lyophilized leaf tissue can be ground using a bead mill homogenizer prior to the addition of lysis buffer, though additional homogenization in lysis buffer will further improve yields. When working with new plant species, extra samples should be collected to optimize bead beating. Improved homogenization can be achieved through the use of heavier beads (e.g., stainless steel) and by increasing the oscillation rate or run time of the ball mill. In addition, greater homogenization efficiency can often be achieved by minimizing the volume of the bead tube by using 0.5–1 mL skirted, screw-cap tubes or 96-well plates. If a bead mill homogenizer is unavailable, tissue can be homogenized prior to DNA extraction under liquid nitrogen, using a micropestle and a 1.5 mL microcentrifuge tube rated for cryogenic temperatures
2. The most common gene targets used in fungal metabarcoding studies include variable regions of the small and large subunits (SSU and LSU) and Internal Transcribed Spacer regions (ITS1 and ITS2) of the rRNA gene repeat. While there is no single best locus, the ideal gene marker is conserved across fungi, has nonbiased priming sites, and has high interspecies sequence variation but low intraspecies variation [4]. The ITS regions fit these criteria and have been designated as the official fungal barcode for systematics [24]. Prior to the widespread use of high-throughput sequencing, most fungal sequencing libraries (either cloned or directly sequenced via Sanger sequencing) were sequenced with the robust primer pair ITS1F [25] and

ITS4 [26], which will amplify the entire ITS1 and ITS2 regions, including the 5.8S ribosomal subunit. Because the ITS1F primer is highly selective for fungi (i.e., excluding most plants) and amplifies a wide range of fungi with efficiency (with the exception of Glomeromycota), many studies have focused on the ITS1 region by pairing ITS1F with a reverse primer in the 5.8S region [26, 27]. Related degenerate primers, or mixes of primers with varying bases in some positions, have now been introduced to increase the taxonomic coverage of ITS amplification [28, 29]. However, due to the presence of introns near the ITS1F priming site in some fungi and significant length heterogeneity in this region, ITS2 is increasingly being chosen as a target locus [30, 31], though priming sites for this region are often less selective. As a result, PNA clamps [32] or blocking oligos [33] may be needed to prevent host ITS2 amplification.

The LSU region of the nrDNA gene is more phylogenetically conserved than the ITS region and is a better target if analyses will include questions on evolutionary ecology [34, 35]. However, the number of primers designed that target LSU specifically for short read Illumina chemistry are fewer compared to ITS regions (*but see,* [36]). Consequently, most studies utilize larger targets designed for 454-pyrosequencing, and only use the forward Illumina sequence read [37]. Because this region is highly conserved, host amplification blockers may be required for successful fungal community analyses in the leaf mycobiome.

While gene target selection should be optimized for the specific research questions, several studies have shown that ecological patterns inferred from fungal metabarcode data are generally consistent across gene markers [34, 38, 39]. Nonetheless, selection of gene primers to amplify target loci has potential to introduce taxon-specific amplification bias and skew the relationship between species abundance and the number of sequences obtained [30, 40, 41]. Additionally, the long primer overhangs used for amplicon library generation can result in disproportionate amplification due to interactions between primers and DNA template adjacent to priming sites [42]. To minimize the chances of large, unexpected primer bias, it is best to select primers based on previously reported error rates and community representation from systems similar to the one being studied.

3. While numerous approaches to amplicon library generation are available, a dual-indexed, two-stage PCR approach is recommended as a highly flexible and cost-effective option. This is because the stage 2 primers, which include sample specific molecular identifiers (MIDs), can be purchased once and used

for generating amplicon libraries targeting various loci by using different gene primers in stage 1. In addition, dual-indexed library constructs include two MIDs, allowing each unique pair of forward and reverse stage 2 primers to be used as a sample identifier for multiplexing. In this approach, primers for the first stage of amplicon library generation will include the gene target primer at the 3′ end with the addition of a 5′ overhang consisting of ~25 bp of the platform specific sequencing adapter. In order to increase base diversity during sequencing, stage 1 primers should include a variable length (3–6 bp) degenerate spacer between the partial sequencing adapter and the gene primer (5′-partial sequencing adapter–[3–6 bp Ns]–gene primer-3′) [32]. Separate stage 1 primers must be ordered for each length variant (i.e., 3 Ns, 4 Ns, etc.) and mixed in equal concentrations prior to use. Stage 2 primers complement the partial sequencing adapter added in stage 1 and include a 5′ overhang consisting of the sample-specific MID and the Illumina sequencing linker, which binds the amplicon to the Illumina flowcell in the correct orientation during sequencing (e.g., 5′-linker–MID–partial sequencing adapter-3′). For an example of two-stage PCR amplicon library preparation in practice, see [43].

4. We recommend use of a high-fidelity polymerase to limit PCR errors, which can lead to inflated species richness estimates [44, 45] and increased rates of chimera formation [46]. In addition, high-fidelity polymerase can facilitate the amplification of taxa with primer mismatches [45], resulting in broader taxonomic coverage. However, be aware that the ability of high-fidelity polymerase to "edit" primers to match template sequences [45] can have unintended consequences, such as host plant DNA amplification.
5. Magnetic bead-based cleanup and plate-based cleanup kits with normalization are common methods. Bead-based kits (e.g., AmPure) require an additional DNA quantification step, manual pooling of samples, and additional equipment (magnetic plate).
6. Specific methods for selecting leaves will depend on the particular study question. For example, leaves with visible pathogen damage should be avoided if the goal is to query endophytes. Inclusion of diseased tissues can flood the sample with pathogen DNA, potentially occluding amplification and sequencing of DNA from endophytes.
7. If working in areas where samples cannot be kept in ice-filled coolers after collection, it may be prudent to field-wash leaves (following the same laboratory method of washing in dilute Triton-X 100 solution followed by rinsing in sterilized water)

and punch leaf discs directly into microcentrifuge tubes. Tubes can then be placed into a small cryoshipper that can be attached to a backpack and keeps samples frozen for up to 7 days. Alternatively, leaf discs can be placed directly into tubes containing a lysis buffer (e.g., CTAB—Cetrimonium bromide), which will halt fungal growth and can be used directly for sample homogenization during the first step of DNA extraction.

8. Specific methods for washing leaves will depend on the particular study question. If targeting foliar endophytes (i.e., nonpathogenic fungi that reside within the leaf), epiphytes (i.e., fungi that reside on the surface of the leaf) and spore and hyphal fragments on the surface of the leaf should be removed using a strong surfactant. Surfactant solution and rinse water should be changed between each sample, or as often as possible. This method can be modified for high-throughput processing to expedite processing [14]. Alternatively, perturbation of surface dwelling fungi may be a concern if the study will also examine epiphytes. Exclusive sampling of epiphytes can be accomplished by washing leaves and collecting the liquid, which can be filtered or pelleted via centrifugation. Finally, note that many published studies on endophyte communities include a surface sterilization step using bleach. This works well for culture-based investigations; however, bleach ***will not*** destroy DNA on the leaf surface or remove fungal particles embedded in microcrevasses on the leaf surface and therefore will not eliminate epiphytes and surface adhered fungi from the results. If surface sterilization is desired, because samples will be used for both culturing and sequencing fungal endophytes, a mixture of bleach and a surfactant should be used.
9. In addition to serving as a positive control in PCR reactions, mock community data are useful for downstream bioinformatics and clustering optimization [15, 47] as well as calculating library-wide error rates [2]. Mock communities are generally constructed from pure culture isolates of 10–20 fungal taxa, representing a wide phylogenetic range. The choice of species to include will depend on the expected composition of the fungal communities being studied and the specific research questions, though they do not necessarily need to be members of the sampled communities [48]. Genomic DNA for each isolate is first extracted and used for Sanger sequencing of the target gene region to create a database of expected sequences. DNA extracts from individual isolates are then quantified and pooled by either normalizing to a consistent concentration or by staggering concentrations over several orders of magnitude. Normalized gDNA concentrations can be used to determine the amount of taxon-specific bias

stemming from nrDNA copy number, GC content, or other sources, while a staggered concentration can be used to determine the sensitivity for detecting species with low relative abundance.

10. Annealing temperature is dictated predominantly by the melting temperature of the primers, which is determined by their specific nucleotide composition and the chemical composition of the PCR reaction buffer. Most manufacturers of PCR reagents provide specific guidelines or online tools to calculate an appropriate annealing temperature. When using these tools, only the portion of the primers that complement the priming site should be considered, excluding any overhanging sequence (e.g., Illumina adapters, MIDs, etc.). Also, with mixed or degenerate primers it is important to consider the range of possible melting temperatures for the primer variants. Whenever possible, the annealing temperature should be set at least a few degrees lower than recommendations to facilitate amplification of mixed templates and reduce amplification biases. When visualizing PCR product during PCR optimization, length variation of the target loci in different taxa will result in a somewhat diffuse band or a few bands clustered within the expected range. If unexpected bands or excessive smearing are seen, gradient PCR can be used to help select an annealing temperature that will limit nonspecific amplification.

    Template concentration should be standardized as much as possible across samples to ensure consistent sampling effort and PCR efficiency. When DNA extraction efficiency is highly variable, normalization can be achieved by quantifying the amount of DNA recovered in each extraction and diluting to a consistent concentration across samples (generally 5 ng/μL–25 ng/μL). However, optimizing DNA extraction (beat beating, etc.) to minimize variation in extraction efficiency and using a consistent amount of plant tissue is often sufficient. Samples should also be tested for the presence of PCR inhibitors by running replicate PCR reactions with various dilutions of the same samples. If more dilute samples result in more robust amplification, inhibitors are present and all samples should be diluted by a fixed amount (i.e., the minimum amount necessary to achieve consistent amplification). When the concentration of fungal DNA is expected to be very low and diluting templates is not desirable, increasing the PCR reaction volume without increasing the volume of template used can often facilitate robust amplification, at the expense of using more PCR reagents. Additional modifications, such as the inclusion of DMSO or betaine can facilitate amplification when dealing with difficult templates and specific recommendations are often provided by polymerase manufacturers.

Acknowledgments

The authors are grateful to Maggie Wagner, Edward Barge, Kyle Gervers, Thomas Wolpert, and Wenbo Ma for helpful feedback in the development of this protocol. This research was supported by the Department of Energy Feedstock Genomics Award 219086 (PEB) and the National Science Foundation Science Engineering and Education for Sustainability Award 1314095 (PEB).

References

1. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N et al (2012) Ultra-hight-throughput microbial analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6:1621–1624
2. Schloss PD, Gevers D, Westcott SL (2011) Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS One 6:e27310
3. Lee CK, Herbold CW, Polson SW, Wommack KE, Williamson SJ, McDonald IR et al (2012) Groundtruthing next-gen sequencing for microbial ecology-biases and errors in community structure estimates from PCR amplicon pyrosequencing. PLoS One 7:e44224
4. Lindahl BD, Nilsson RH, Tedersoo L, Abarenkov K, Carlsen T, Kjøller R et al (2013) Fungal community analysis by high-throughput sequencing of amplified markers—a user's guide. New Phytol 199:288–299
5. Arnold AE, Maynard Z, Gilbert GS, Coley PD, Kusar TA (2000) Are tropical endophytes hyperdiverse? Ecol Lett 3:267–274
6. Hardoim PR, van Overbeek LS, Berg G, Pirttilä AM, Compant S, Campisano A et al (2015) The hidden world within plants: ecological and evolutionary considerations for defining functioning of microbial endophytes. Microbiol Mol Biol Rev 79:293–320
7. Herbold CW, Pelikan C, Kuzyk O, Hausmann B, Angel R, Berry D et al (2015) A flexible and economical barcoding approach for highly multiplexed amplicon sequencing of diverse target genes. Front Microbiol 6:1–8
8. Harrison JG, Forister ML, Parchman TL, Kock GW (2016) Vertical stratification of the foliar fungal community in the world's tallest trees. Am J Bot 103:2087–2095
9. Bahnweg G, Heller W, Stich S, Knappe C, Betz G, Heerdt C et al (2005) Beech leaf colonization by the endophyte *Apiognomonia errabunda* dramatically depends on light exposure and climatic conditions. Plant Biol 7:659–669
10. Rai M, Agarkar G (2016) Plant-fungal interactions: what triggers the fungi to switch among lifestyles? Crit Rev Microbiol 42:428–438
11. López-González RC, Gómez-Cornelio S, De la Rosa-García SC, Garrido E, Oropeza-Mariano O et al (2017) The age of lima bean leaves influences the richness and diversity of the endophytic fungal community, but not the antagonistic effect of endophytes against *Colletotrichum lindemuthianum*. Fungal Ecol 26:1–10
12. Erickson RO, Michelini FJ (1957) The plastochron index. Am J Bot 44:297–305
13. Meicenheimer RD (2014) The plastochron index: still useful after nearly six decades. Am J Bot 101:1821–1835
14. Greenfield M, Pareja R, Ortiz V, Gómez-Jiménez MI, Vega FE, Parsa S (2015) A novel method to scale up fungal endophyte isolations. Biocontrol Sci Tech 25(10):1208–1212
15. Nguyen NH, Smith D, Peay K, Kennedy P (2015) Parsing ecological signal from noise in next generation amplicon sequencing. New Phytol 205:1289–1393
16. Stevens JL, Jackson RL, Olson JB (2013) Slowing PCR ramp speed reduces chimera formation from environmental samples. J Microbiol Methods 93:203–205
17. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: high-resolution sample inferences from Illumina amplicon data. Nat Methods 13:581–583
18. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB et al (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541
19. Anslan S, Bahram M, Hiiesalu I, Tedersoo L (2017) PipeCraft: flexible open-source toolkit for bioinformatics analysis of custom high-throughput amplicon sequencing data. Mol

Ecol Resour 17(6):e234–e240. https://doi.org/10.1111/1755-0998.12692

20. Gweon HS, Oliver A, Taylor J, Booth T, Gibbs M, Read DS et al (2015) PIPITS: an automated pipeline for analysis of fungal internal transcribed spacer sequences from the Illumina sequencing platform. Methods Ecol Evol 6:973–980
21. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK et al (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336
22. Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10:996–998
23. Schmidt TSB, Matias Rodrigues JF, von Mering C (2015) Limits to robustness and reproducibility in the demarcation of operational taxonomic units. Environ Microbiol 17:1689–1706
24. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA et al (2012) Nuclear ribosomal internal transcribed space (ITS) region as a universal DNA barcode marker for fungi. PNAS 109:6241–6246
25. Gardes M, Bruns TD (1993) ITS primers with enhanced specificity for basidiomycetes—application to the identification of mycorrhizae and rusts. Mol Ecol 2:113–118
26. White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky J, White TJ (eds) PCR protocols: a guide to methods and applications. Academic Press, London, pp 315–322
27. Smith DP, Peay KG (2014) Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. PLoS One 9(2):e90234
28. Tedersoo L, Bahram M, Polme S, Koljalg U, Yorou NS, Wijesundera R et al (2014) Global diversity and geography of soil fungi. Science 346:1256688–1256688
29. Toju H, Tanabe AS, Yamamoto S, Sato H (2012) High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. PLoS One 7(7):e40863
30. Ihrmark K, Bödeker ITM, Cruz-Martinez K, Friberg H, Kubartova A, Schenck J et al (2012) New primers to amplify the fungal ITS2 region—evaluation by 454-sequencing of artificial and natural communities. FEMS Microbiol Ecol 82:666–677
31. Taylor DL, Walters WA, Lennon NJ, Bochicchio J, Krohn A, Caporaso JG et al (2016) Accurate estimation of fungal diversity and abundance through improved lineage-specific primers optimized for Illumina amplicon sequencing. App Environ Microbiol 82:7217–7226
32. Lundberg DS, Yourstone S, Mieczkowski P, Jones CD, Dangl JL (2013) Practical innovations for high-throughput amplicon sequencing. Nat Methods 10:999–1002
33. Vestheim H, Jarman SN (2008) Blocking primers to enhance PCR amplification of rare sequences in mixed samples—a case study on prey DNA in Antarctic krill stomachs. Front Zool 5:12
34. Brown SP, Rigdon-Huss AR, Jumpponen A (2014) Analyses of ITS and LSU gene regions provide congruent results on fungal community responses. Fungal Ecol 9:65–68
35. Halwachs B, Madhusudhan N, Krause R, Nilsson RH, Moissl-Eichinger C, Högenauer C et al (2017) Critical issues in mycobiota analysis. Front Microbiol 8:180. https://doi.org/10.3389/fmicb.2017.00180
36. Mueller RC, Gallegos LV, Kuske CR (2016) A new fungal large subunit ribosomal RNA primer for high-throughput sequencing surveys. FEMS Microbiol Ecol 92:fiv153
37. Amend AS, Martiny AC, Allison SD, Berlemont R, Goulden ML, Lu Y et al (2016) Microbial response to simulated global change is phylogenetically conserved and linked with functional potential. ISME J 10:109–118
38. Berriti A, Desirò A, Visentin S, Zecca O, Bonfante P (2017) ITS fungal barcoding primers vs 18S AMF-specific primers reveal similar AMF-based diversity patterns in roots and soils of three mountain systems. Environ Microbiol Rep 9(5):658–667. https://doi.org/10.1111/1758-2229.12574
39. Blaalid R, Kumar S, Nilsson RH, Abarenkov K, Kirk PM, Kauserud H (2013) ITS1 versus ITS2 as DNA metabarcodes for fungi. Mol Ecol Resour 13:218–224
40. Bellemain E, Carlsen T, Brochmann C, Coissas E, Taberlet P, Kauserud H (2010) ITS as an environmental DNA barcode for fungi: an *in silico* approach reveals potential PCR biases. BMC Microbiol 10:189
41. Anderson IC, Cairney JWG (2004) Diversity and ecology of soil fungal communities: increased understanding through the application of molecular techniques. Environ Microbiol 6:769–779
42. Berry D, Mahfoudh KB, Wagner M, Loy A (2011) Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. App Environ Microbiol 77:7846–7849
43. Toju H, Tanabe AS, Ishii HA (2016) Ericaceous plant-fungus network in a harsh alpine-subalpine environment. Mol Ecol 25:3242–3257

44. Oliver AK, Brown SP, Callaham MA Jr, Jumpponen A (2015) Polymerase matters: non-proofreading enzymes inflate fungal community richness estimates by up to 15%. Fungal Ecol 15:86–89
45. Gohl DM, Vangay P, Garbe J, MacLean A, Hauge A, Becker A, Gould TJ, Clayton JB, Johnson TJ, Hunter R, Knights D, Beckman KB (2016) Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. Nat Biotechnol 34(9):942–952
46. Lahr DJG, Katz LA (2009) Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. BioTechniques 47:857–863
47. Aas AB, Davey ML, Kauserud H (2017) ITS all right mama: investigating the formation of chimeric sequences in the ITS2 region by DNA metabarcoding analysis of fungal mock communities in different complexities. Mol Ecol Resour 17:730–741
48. Brown SP, Ferrer A, Dalling JW, Heath KD (2016) Don't put all your eggs in one basket: a cost-effective and powerful method to optimize primer choice for rRNA environmental community analysis using the Fluidigm access array. Mol Ecol Resour 16:946–956



Chapter 5

Characterizing Small RNAs in Filamentous Fungi Using the Rice Blast Fungus, *Magnaporthe oryzae*, as an Example

Vidhyavathi Raman, Blake C. Meyers, Ralph A. Dean, and Nicole M. Donofrio

Abstract

The goal of this chapter is to provide a framework of sequential steps for small RNA (sRNA) analysis in filamentous fungi. Here, we present protocols for (1) comparative analysis of sRNAs in different conditions, (2) comparisons of sRNA libraries to RNAseq data and (3) identification and analysis of methylguanosine-capped and polyadenylated sRNAs (CPA-sRNAs). This species of small RNA is particularly interesting in *Magnaporthe oryzae,* as they map to transcription start and end sites of protein-coding genes. While we do not provide specific command lines for scripts, we provide a general framework for steps needed to carry out all three types of analyses, including relevant references, websites and free online tools. Screenshots are provided from our own customized interface using *M. oryzae* as an example, to assist the reader in visualizing many of the steps.

Key words sRNAs, Small interfering RNAs, CPA-sRNAs, RNAseq, Fungus

1 Introduction

Since the discovery of quelling in the model, filamentous Ascomycete fungus *Neurospora crassa* [1], there has been great interest in understanding the role of small RNA (sRNA) species in fungi. While not exhaustive, our recent survey of the literature identified close to 500 published research and review articles on this topic. Species in which significant work has been done range from plant pathogens such as the rice blast fungus *Magnaporthe oryzae* and the basidiomycete wheat rust fungus *Puccinia striiformis* to model, nonpathogens like *N. crassa*, to human pathogens such as *Cryptococcus gattii*. Interest in sRNAs is driven by many biological questions, including what sRNA species do fungi contain, what are their roles, do they contribute to fungal fitness, and do they have roles in pathogenicity? Higher throughput methods

Wenbo Ma and Thomas Wolpert (eds.), *Plant Pathogenic Fungi and Oomycetes: Methods and Protocols*, Methods in Molecular Biology, vol. 1848, https://doi.org/10.1007/978-1-4939-8724-5_5,
© Springer Science+Business Media, LLC, part of Springer Nature 2018

to isolate sRNA species, coupled with genetic analyses are starting to provide answers to these questions. For example, fungi have different pathways to generate different species of sRNAs, such as miRNA-like sRNAs (milRNAs) and small interfering RNAs (siRNAs) [2]. A more recent study determined that fungal sRNAs from the necrotrophic plant pathogen *Botrytis cinerea* hijack the plant's sRNA machinery to silence plant genes involved in immunity [3]. Not only did the latter study define a new class of "effectors" from pathogenic fungi, but it demonstrated cross-kingdom transfer of sRNAs between a fungus and a plant.

The advent of next-generation sequencing made accessible many formerly unanswered questions about sRNAs in fungi. While initial studies of sRNAs began with the identification of genetic mutations and studying phenotypic changes as in the case of quelling in *N. crassa* [1, 4], eventually sequencing methods were paired with isolation and characterization of the sRNA species themselves. Recent research has centered around whole genome sequencing of sRNAs in fungi using next generation platforms, as in the case of the *B. cinerea* study mentioned above. However, the massive amounts of data generated from these fast-evolving techniques can lead to a so-called "bioinformatic bottleneck," or simply put, confusion over how to analyze copious amounts of data at once [5]. There is a multitude of freely available programs for analyzing deep sequencing data such as Bowtie, which aligns short sequence reads to a reference genome [6] and RepeatMasker, which identifies repetitive elements [7].

Our labs study the role of sRNAs in the devastating rice blast pathogen *M. oryzae*. Specifically, we characterized the types of sRNA species present in the fungus including how they are generated, their enrichment in different infective stages, and their abundance during environmental stresses [8–10]. Furthermore, we have also examined their role with regard to gene expression including the identification and characterization of methylguanosine-capped and polyadenylated sRNAs (CPA-sRNAs), which map to either end of protein-coding genes [9, 11]. CPA-sRNAs primarily map to the 5′ transcriptional start site (TSS) and 3′ transcription termination sites (TTS) of protein-coding genes and, since their accumulation is directly correlated with the mRNA level of the gene, they may have an important regulatory function. In all cases, bioinformatics approaches were employed, and in some cases developed or modified, in order to answer the biological questions of interest. Here, we describe the more commonly used analytic approaches, such as comparisons of sRNA libraries, as well as analyzing how sRNAs are associated with gene expression and the identification of CPA-sRNAs. In several instances, PERL or SQL scripts were used, and where possible, we provide references to these scripts. The purpose of this chapter, however, is not to list every command line in a script, but rather to provide the sequence of events needed to compare fungal sRNA libraries, and to analyze sRNAs in the context of their roles in modulating gene expression.

2 Materials

2.1 Basic Materials Required Are as Follows

1. Computer.
2. Internet connection.
3. Server to store and process gigabytes-worth of data.
4. Fastq data files, or other large-file format data files generated from an Illumina sequencer (we used the Illumina HiSeq 2000) or a 454 sequencer (454 Life Sciences).

2.2 Generation of the Data Files

Data files were generated from different fungal strains, e.g., WT7015 is the wild type strain of *M. oryzae*, to which all mutants were compared, and *dcl2* (where the Dicer 2 gene is knocked out) is one such mutant. All strains were grown in a complete, synthetic growth media for *M. oryzae* (10 g sucrose, 6 g yeast extract, 6 g casamino acids, 1 mL trace elements, per 1 L), which is conducive to vegetative, mycelial growth. RNA extracted using the TRIzol reagent according to the manufacturer's protocols (Invitrogen, New York, USA), and small RNAs were obtained using the Illumina TruSeq Small RNA Preparation Kit (San Diego, CA, USA). Small RNA mutants were generated as per previous studies by Raman and colleagues [10, 11], and were compared to wild type for the purpose of better defining their roles in generation of small RNAs, and the impacts on mRNAs. For CPA-sRNAs, data files were generated from WT7015 grown in complete liquid media (0.2% yeast extract and 1% sucrose) by shaking at 200 rpm, 25 °C for 3 days. RNA was extracted as described above and full-length cDNA prepared [9]. There are a number of commercial kits available to construct full-length cDNA libraries such as Cloneminer II (Invitrogen) or SMART (ClonTech). Any of these kits are appropriate, the details of construction are provided by the manufacturer. In all cases, specific sequences (linkers or adapters) are incorporated into the 5′ and 3′ ends of cDNA in order to provide directional cloning. Finally, size-selected cDNA (60–200 nt) are purified from agarose gels [9]. We matched genome reads to the reference genome of *Magnaporthe oryzae* version 8 assembly, which was at the time housed at the Broad Institute, but now can be found here: https://www.ncbi.nlm.nih.gov/genome/?term=magnaporthe+oryzae

For RNA-seq analysis, it is important to be familiar (or to familiarize yourself), with the Tuxedo Suite (at least Bowtie, TopHat, and Cufflinks; [15]), which is a commonly used set of analysis tools.

3 Methods

3.1 Analyzing and Comparing sRNA Libraries

This section discusses the analysis and comparison of deeply sequenced sRNA libraries (Fig. 1). We utilize a method [12], in which the entire genome is segmented into clusters (windows)

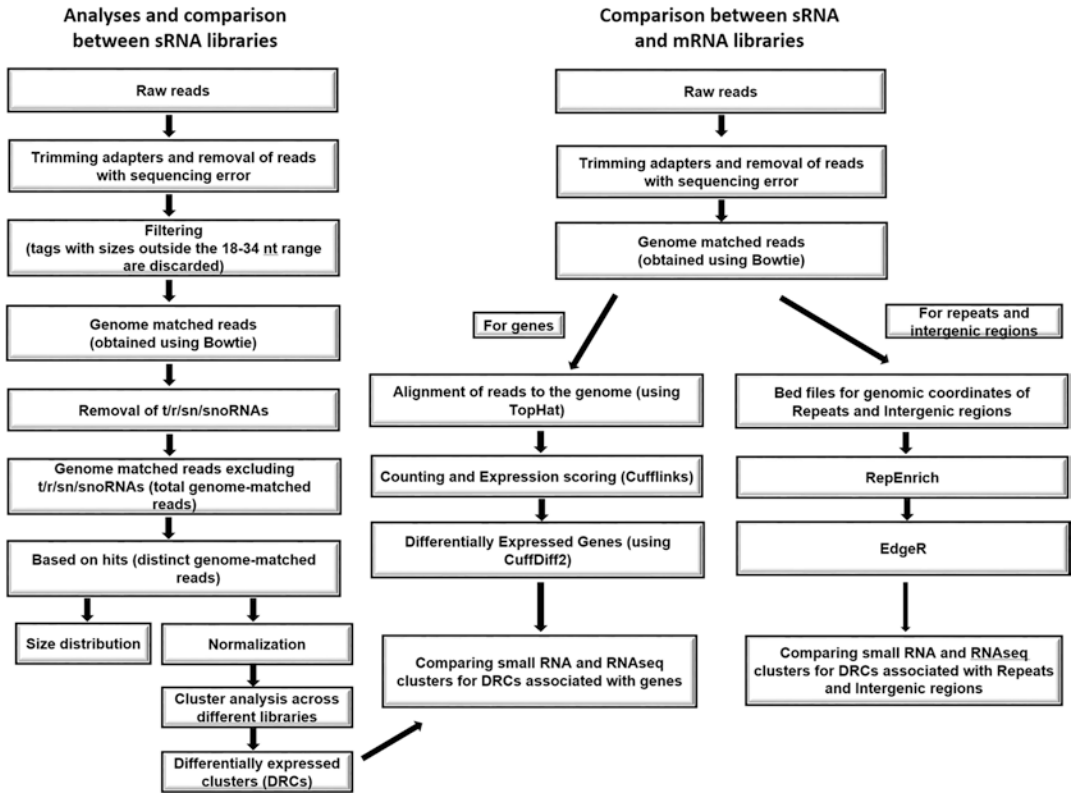


Fig. 1 A flow chart of sRNA-seq and RNA-seq analyses

using a proximity-based algorithm. These analyses lead to the identification of differentially regulated clusters (DRCs) of sRNAs, representing a direct measurement of the sRNAs reads in one sample versus another. In our case, a cutoff of >10-fold differences was used between two or more libraries. Because this protocol utilizes libraries generated with Illumina HiSeq sequencing, adaptors were removed before proceeding with the steps described below.

Note that **steps 1–6** are heavily computational and require various scripts that can be customized for personal use. These can be found in Nobuta et al. [13]. Listed below are the general steps provided in sequence.

1. Reads with any sequencing errors are filtered out using PERL scripts, as per Mathioni et al. [14].
2. Reads matched to tRNA, rRNA, snRNA, or snoRNA are removed using an annotation file (downloaded from the Broad Institute's website: ftp://ftp.broadinstitute.org/pub/annotation/fungi/magnaporthe/genomes/magnaporthe_oryzae_70-15_8/). tRNAscan is also utilized to identify tRNAs [15].

3. From the remaining collection, reads are matched to the reference genome using Bowtie, with parameters set to accept only perfect matches.
4. Distinct genome-matched reads are then calculated from total-genome matched reads.
5. The abundance of each read is normalized to transcripts per five million (TP5M), which accounts for differences in sequencing depth and is a standard practice in the Meyers' lab. Additional details may be found at our homepage: https://mpss.danforthcenter.org/web/php/pages/faq.php?SITE=mg_sRNA#5.
6. In order to compare different libraries, the next step is to parse the genome into clusters, or windows, of 500 bp each and calculate the abundance of sRNAs for each cluster, as explained in Lu et al. [12].
7. For the DRCs of interest, a comparison between sRNA libraries can be visualized by following the procedure below illustrated with the *Magnaporthe* sRNA database located at the following link as an example using a customized interface: https://mpss.danforthcenter.org/dbs/index.php?SITE=mg_sRNA.
   (a) Click the control panel for selecting the sRNA libraries to be visualized (Fig. 2a). The control panel can also be accessed from a different place that will be explained in **step 7c** Screenshots of the control panel with default settings and after selection are shown in Fig. 2b.
   (b) For DRCs of interest, enter the cluster ID (with the format of genome.chr.start.length, for example: MG_MIT8.1.1008501.500, here '1' is the chromosome number and 1,008,501 is the starting coordinate of the 500-bp window) (Fig. 2a). The results returned after selecting cluster ID is shown in Fig. 2c. By selecting the 'Tabular view' option shown with an arrow in Fig. 2c, the sum of abundance and counts of distinct sRNAs of selected libraries as well as individual size classes of sRNAs can be visualized as shown in Fig. 2d.
   (c) Clicking the control panel as marked in Fig. 2c is an alternate route to access the control panel. In the resulting window, "Abundance view" can also be selected from the drop-down menu of Fig. 2e. Available options for this step are "Sum of abundances," "Sum of abundances–hits-normalized-abundance," and "Individual abundances."
   (d) For the "Library abundance" page, select different display options for libraries ("All libraries (or the ones you selected) combined," "Individual libraries separately") that can be selected again by clicking the "control panel" (Fig. 2f).

Figure 2g shows the display page when the "Individual libraries" option is selected. For the analyzed cluster (MG_MIT8.1.1008501.500), the sRNAs present in the WT7015 library are depleted in the *dcl2* library, indicating that the sRNAs present in this cluster are DCL2 dependent.

3.2 Comparing sRNA- and RNA-seq Libraries

In this section, we present steps to compare sequencing results obtained from sRNA libraries with mRNA libraries in order to obtain a broad picture of genetic control (Fig. 1). For this section, we do not delineate the process of analyzing mRNA libraries, but rather assume that the analyzed data is already in-hand, and proceed from there with a comparison between sRNA and mRNA analysis.

1. Enter the gene ID in the marked area of Fig. 3a for sRNA abundance for a gene across different libraries. For multiple genes, a bulk query can be carried out. Select "Gene analysis viewer option" for this step as marked in the Fig. 3a.

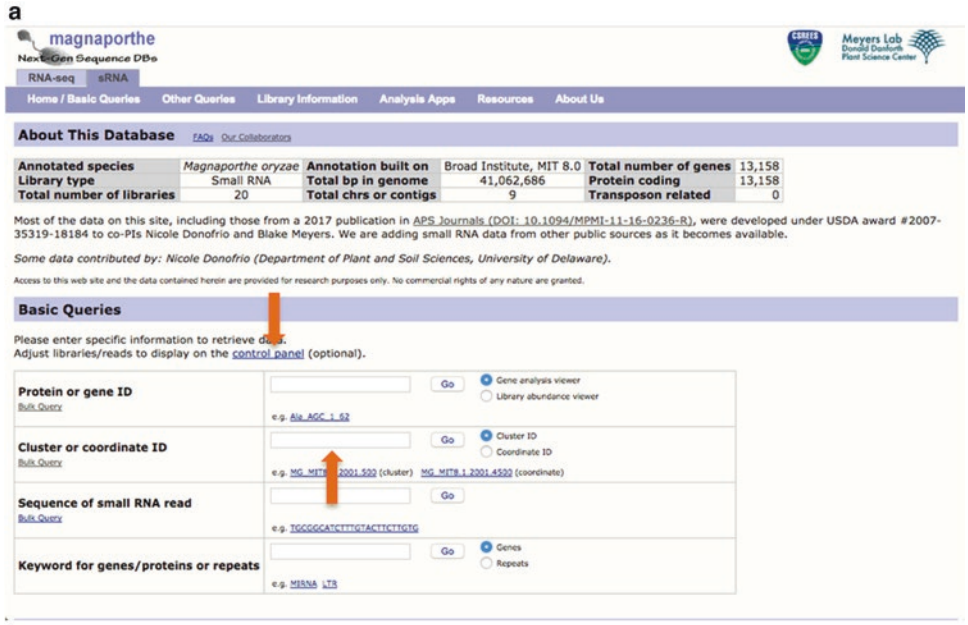


Fig. 2 User interfaces. (**a**) Initial user interface for selection of sRNA libraries. Down-facing orange arrow denotes the control panel link, and up-facing arrow denotes where the user may enter an individual DRC of interest. (**b**) User interface after clicking control panel. (**c**) Results returned after selecting a particular DRC of interest. Orange arrow closest to the left denotes the "Tabular View" button, which provides statistics on distinct sRNAs of previously selected libraries. (**d**) Visualization and tabular data for sum of abundance, and counts of distinct, selected sRNAs. (**e**) sRNA abundances may be viewed using the "Abundance view" drop-down menu (denoted by orange arrow). (**f**) Another view of the control panel to select additional libraries. (**g**) Visualization of having selected "individual libraries" from the preceding user interface (Fig. 2f). The graphics show sRNA abundances in one particular cluster in two different libraries, the wild type (WT7015) and the mutant (*dcl2*)

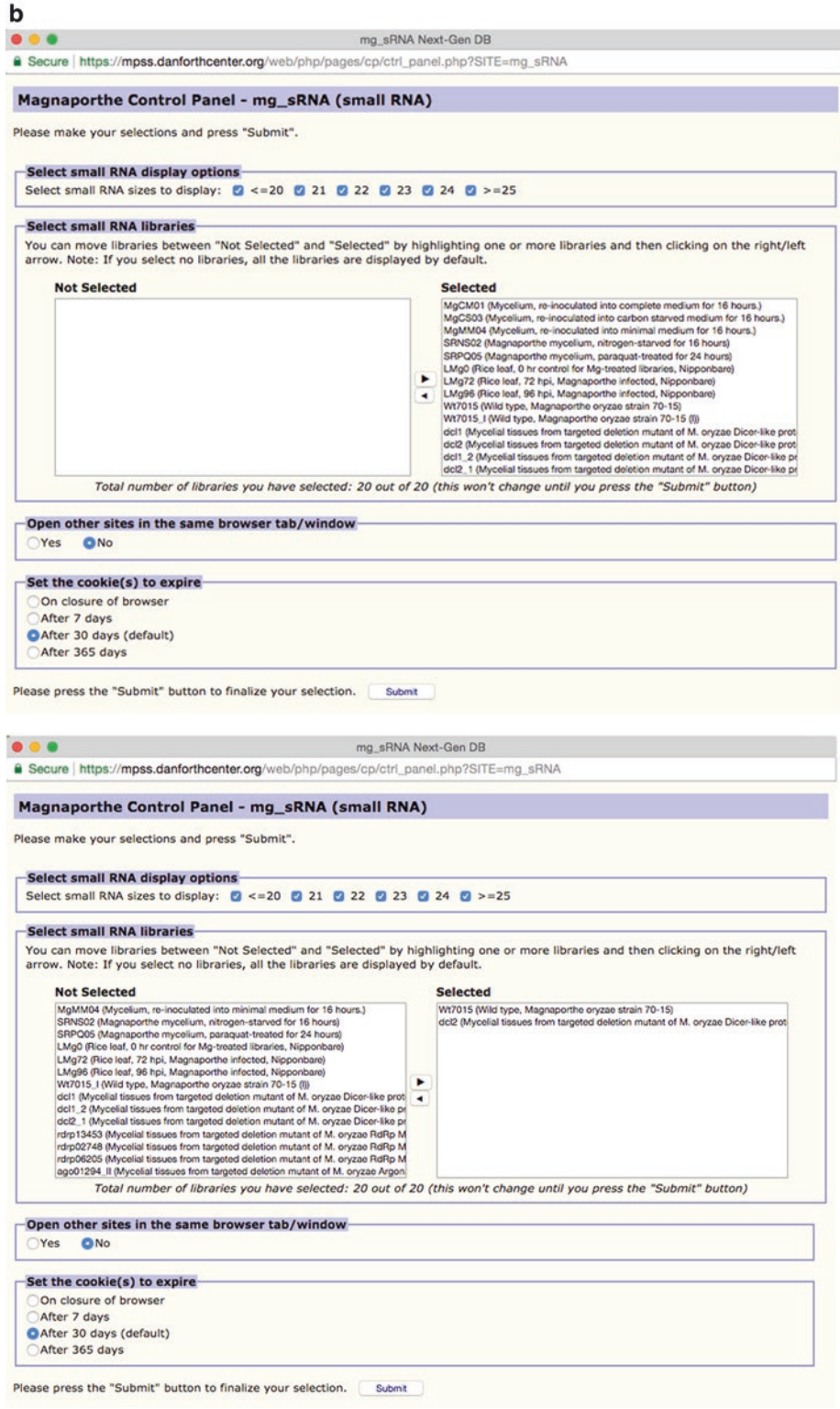


Fig. 2 (continued)

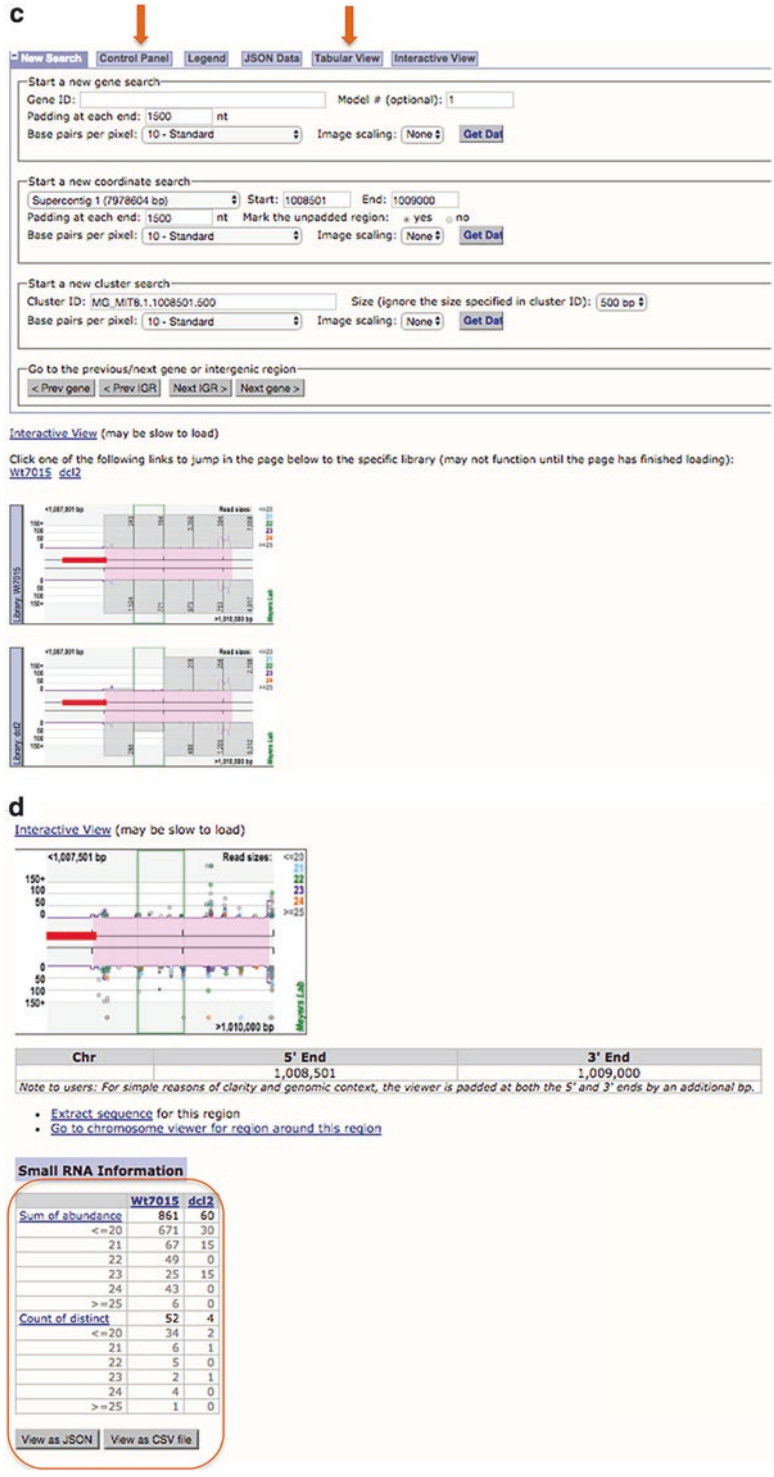


Fig. 2 (continued)

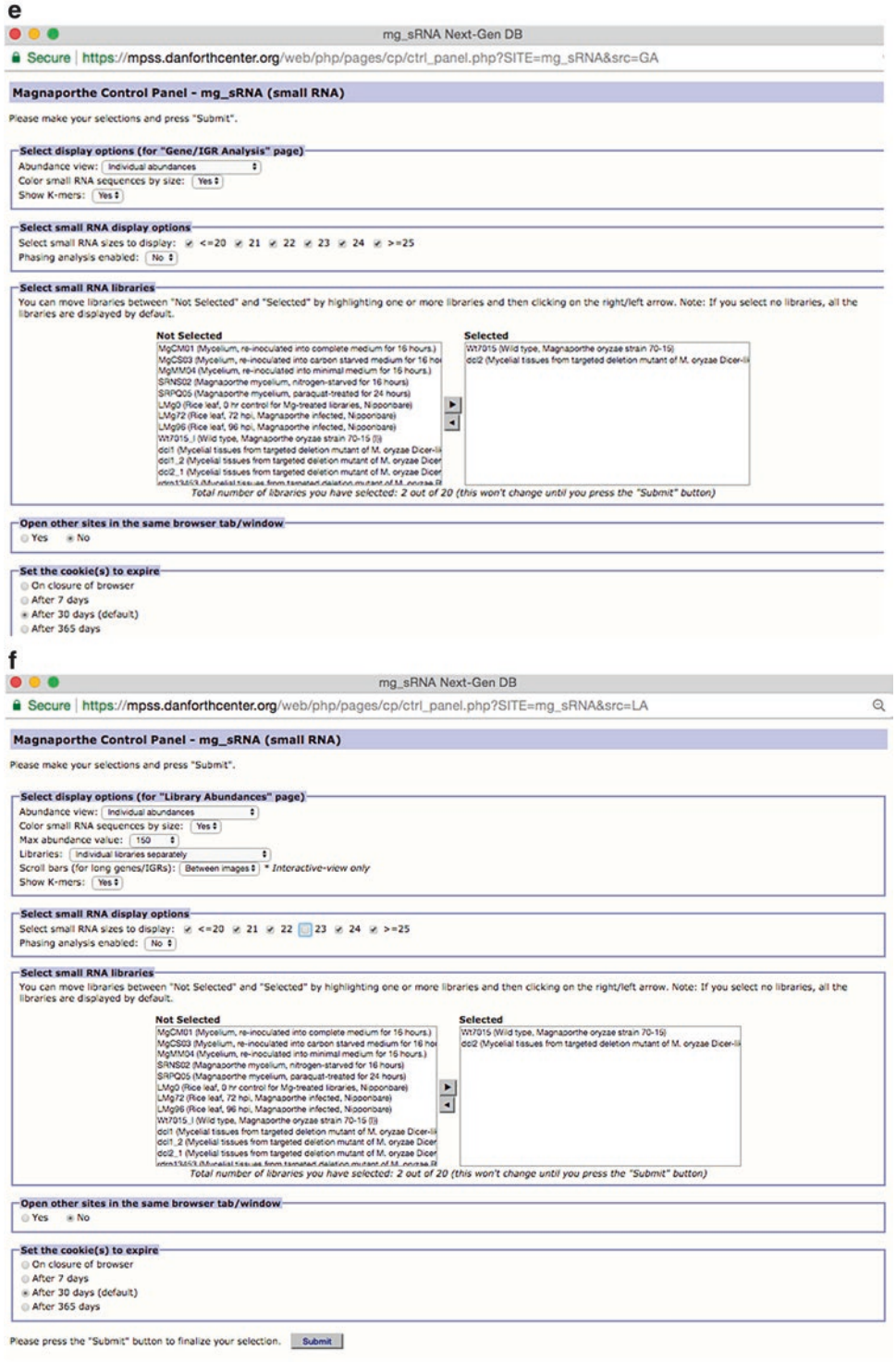


Fig. 2 (continued)

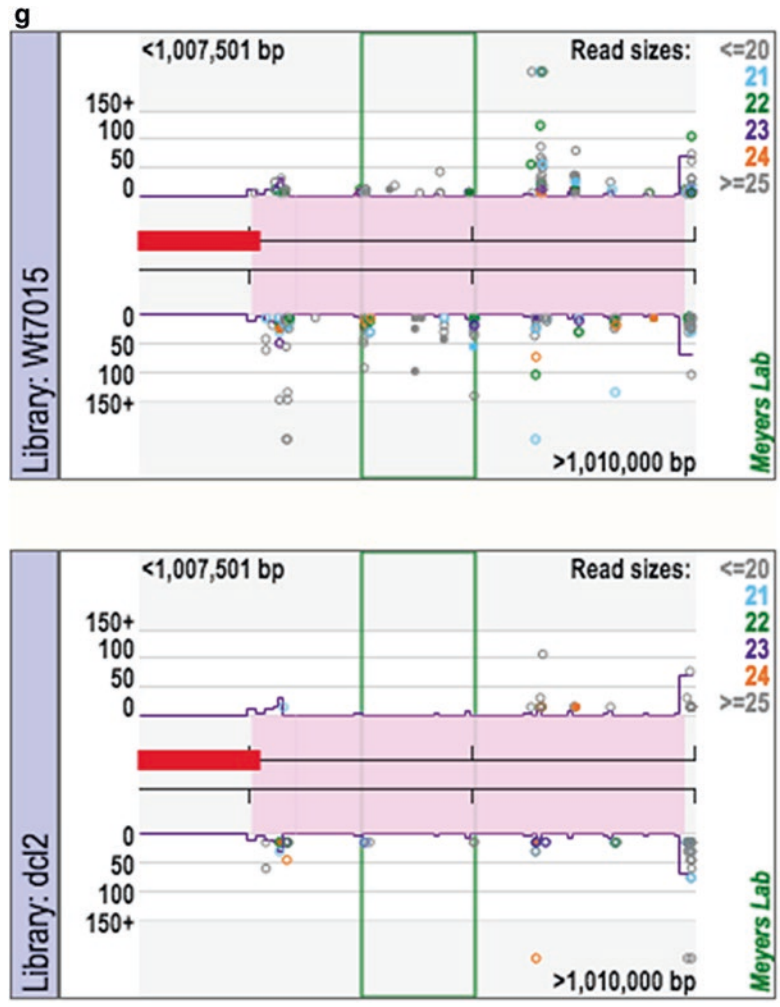


Fig. 2 (continued)

2. The window resulting from selections made as described in the previous step is shown in Fig. 3b. This page shows the sum of abundance and count of distinct sRNAs of selected libraries, as well as individual size classes of the sRNAs.
3. As described in Subheading 3.1, **step 7**, by clicking "Control panel" different selections on abundance view and display options for libraries can be made.
4. Once sRNA abundances are returned as shown in Fig. 3b, ratios or fold-changes of sRNAs from treatment vs. control libraries re calculated (e.g., sRNA abundance of MGG_02722 in *Δago01294* library divided by sRNA abundance in wild-type library)
5. Expression ratios or fold-changes of transcript abundance from RNA-seq data are analyzed by processing the data through the

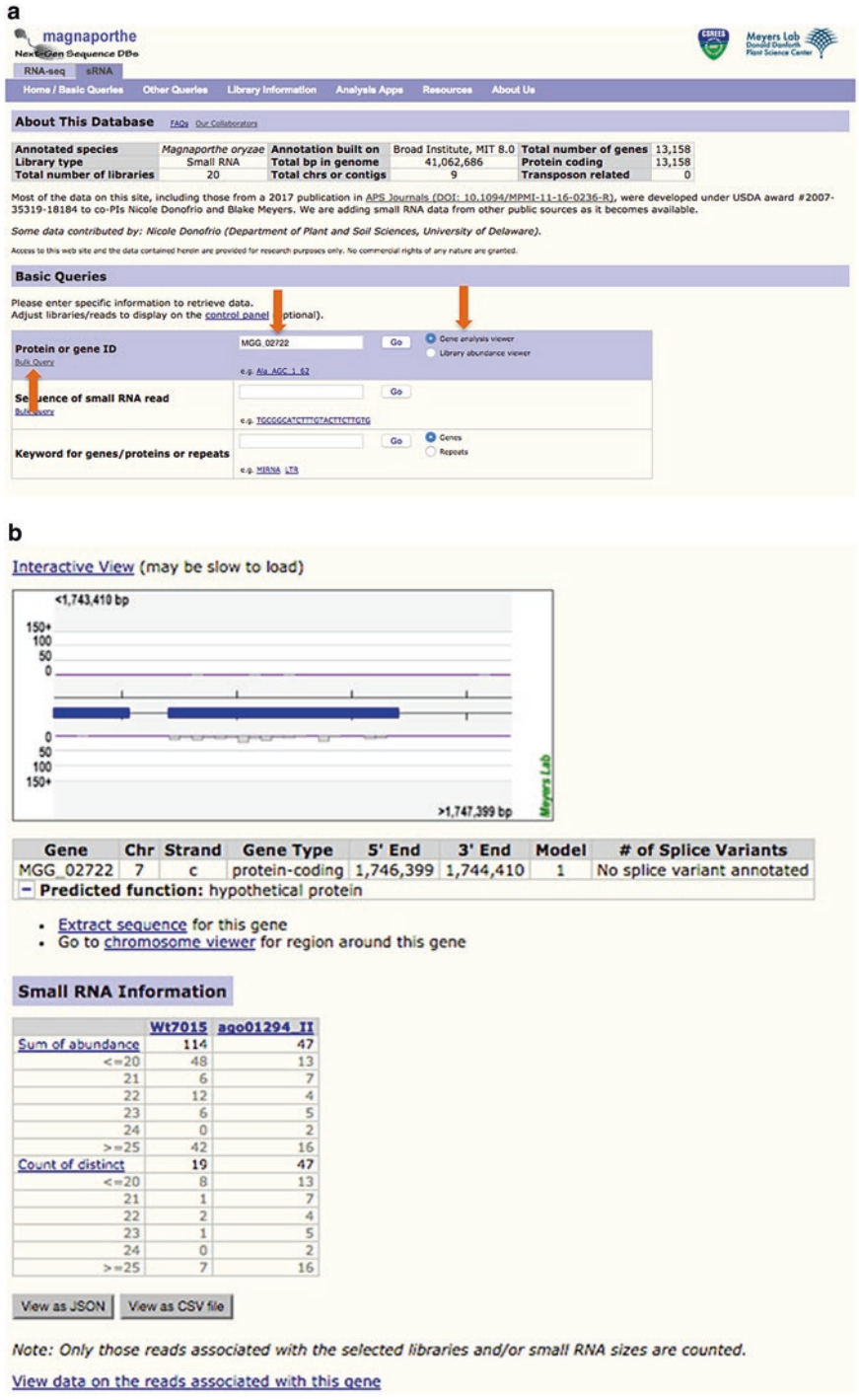


Fig. 3 (**a**) Down-facing orange arrows point to where gene ID number can be entered to observe sRNA abundance for a gene across different libraries. Up-facing arrow allows the user to enter gene IDs in bulk. (**b**) Visualization of sRNAs associated with a particular gene

Tuxedo suite [16], which is a collection of tools for analysis of small RNA data.

6. Comparison between sRNA-seq and RNA-seq data is visualized via scatter plotting $\log_2$ fold-changes of each gene as we have demonstrated in previous research by Raman and colleagues (refer to Fig. 7c in [11]).

3.3 Identification and Mapping of CPA-sRNAs

In this section, we provide guidance on how to identify CPA-sRNAs and evaluate their relationship to feature location and gene expression and function. The procedure requires the production of full-length cDNA from which small molecules are size selected and sequenced using Illumina or 454 sequencing technologies following manufacturers' protocols. In our work with mycelia of *M. oryzae*, ~100,000 raw reads from 454 sequencing were found to be sufficient for identifying and characterizing CPA-sRNAs.

1. 5′ and 3′ linker/adapter sequences are identified (exact String match), filtered out and sequence reads ≥16 bp are retained.
2. CPA-sRNAs are aligned to the *M. oryzae* genome and available ESTs using BLASTN with criteria of 80% coverage and 98% sequence identity. These criteria are used to minimize "false" location matches. Match locations (beginning and end nucleotide sites) to all known genomic features (contigs, gene, tRNAs, snRNAs, repeats, mitochondrial genome) are recorded and visualized on a genome browser, and one such example is gbrowse [17].
3. Because CPA-sRNAs may align to the genome more than one time, three metrics are applied to describe the alignments—alignment counts, read counts, and prorating. Alignment count means summation of all CPA-sRNA alignments to a particular genomic feature. Read count means each CPA-sRNA is counted only once for a given genomic feature, such as to a particular repetitive element. Prorating: CPA-sRNAs are apportioned across features associated with that alignment and weight values are derived. For example, if a particular CPA-sRNA maps to three genomic features, then that individual CPA-sRNA would have a weight score or 0.33 for each of the three features. Summation of so called prorated CPA-sRNA weight values for any feature thus yields a balanced picture of CPA-sRNA coverage for that feature (for additional examples, *see* [9]). Thus, read counts and prorating account for the ambiguity in determining genomic origin and deriving a reasonable estimate of number of CPA-RNAs mapping to a particular genomic feature.
4. To map the location of CPA-sRNAs to gene models or other genomic features such as tRNA, first, CPA-sRNAs are assigned to gene models supported by ESTs and other expression data (such as RNA-seq) when available. Second, CPAs

are mapped to gene models lacking expression data with the addition of 300 bp at both the 5′ and 3′ends, UTR lengths typically found in *Magnaporthe* protein-coding genes. One hundred and fifty base pairs 5′ and 3′ UTRs are used for tRNAs. The UTR positions and lengths are an estimated average based on predicted start and stop codons in gene models and on relatively few confirmed gene translation products and full-length cDNA clones. Summing the nucleotide positions matched by each CPA-sRNA provides frequency of base matches on both coding and noncoding DNA strand on each gene model. Plotting the frequency distribution of summed nucleotide match locations relative to the transcriptional start and stop provides visual evidence that CPA-sRNAs primarily match to 5′ TSS and 3′ TTS of protein coding genes. As an example, *see* Fig. 1c in [9].

5. To examine the relationship between CPA-sRNAs and gene expression, statistical analyses are conducted in readily available software packages such as JMP (SAS Institute, Inc., Cary, NC, 1989–2007). Correlation analyses are conducted for CPAs mapped in both the sense and antisense orientation against normalized values from gene expression data such as microarray or RNA-seq data. If the resulting correlations are weak, genes can be grouped into bins based on expression levels and the relationship between mean bin gene expression and the associated mean CPAs values compared.
6. Finally, correlation studies between mapped CPA-sRNAs and other criteria such as gene function can be examined. For example, Gene Ontology of genes are conducted and presented for Biological Process and Molecular Function. Data for GO annotation of *M. oryzae* is available at [18]. An example of the relationship between CPA-sRNAs from mycelia of *M. oryzae* and GO annotation can be seen in [[9], Fig. 4e and 4f].

References

1. Romano N, Macino G (1992) Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. Mol Microbiol 6(22):3343–3353
2. Lee HC, Li L, Gu W, Xue Z, Crosthwaite SK, Pertsemlidis A, Lewis ZA, Frietag M, Selker EU, Mello CC, Liu Y (2010) Diverse pathways generate microRNA-like RNAs and Dicer-independent small interfering RNAs in fungi. Mol Cell 38(6):803–814
3. Weiberg A, Wang M, Lin FM, Zhao H, Zhang Z, Kaloshian I, Jin H (2013) Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. Science 342(6154):118–123
4. Cogoni C, Macino G (1997) Isolation of quelling-defective (qde) mutants impaired in posttranscriptional transgene-induced gene silencing in *Neurospora crassa*. Proc Natl Acad Sci U S A 94(19):10233–10238
5. Scholz MB, Lo C-C, Chain PSG (2012) Next generation sequencing and bioinformatics bottlenecks: the current state of metagenomics data analysis. Curr Opin Biotechnol 23(1):9–15
6. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient align-

ment of short DNA sequences to the human genome. Genome Biol 10:R25

7. Smit AFA, Hubley R, Green P (1996-2000) RepeatMasker Open-3.0. http://www.repeatmasker.org
8. Nunes CC, Gowda M, Sailsbery J, Xue M, Chen F, Brown DE, Oh Y, Mitchell TK, Dean RA (2011) Diverse and tissue-enriched small RNAs in the plant pathogenic fungus, *Magnaporthe oryzae*. BMC Genomics 12:288
9. Gowda M, Nunes CC, Sailsbery J, Xue M, Chen F, Nelson CA, Brown DE, Oh Y, Meng S, Mitchell TK, Hagedorn CH, Dean RA (2010) Genome-wide characterization of methylguanosine-capped and polyadenylated small RNAs in the rice blast fungus *Magnaporthe oryzae*. Nucleic Acids Res 38:7558–7569
10. Raman V, Simon SA, Romag A, Demirci F, Mathioni SM, Zhai J, Meyers BC, Donofrio NM (2013) Physiological stressors and invasive plant infections alter the small RNA transcriptome of the rice blast fungus, *Magnaporthe oryzae*. BMC Genomics 14:326
11. Raman V, Simon SA, Demirci F, Nakano M, Meyers BC, Donofrio NM (2017) Small RNA functions are required for growth and development of *Magnaporthe oryzae*. MPMI 30(7):517–530
12. Lu C, Tej SS, Luo S, Haudenschild CD, Meyers BC, Green PJ (2005) Elucidation of the small RNA component of the transcriptome. Science 309:1567–1569
13. Nobuta K, McCormick K, Nakano M, Meyers BC (2009) Bioinformatics analysis of small RNAs in plantechnologies. In: Meyers BC, Green P (eds) Plant microRNAs. Methods in molecular biology (methods and protocols), vol. 592. Humana Press
14. Mathioni MM, Kakrana A, Meyers BC (2017) Characterization of plant small RNAs by next generation sequencing. Cur Protoc Plant Biol 2(1):39–63
15. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25(5):955–964
16. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7(3):562–578
17. Stein LC, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Staijch JE, Harris TW, Arva A, Lewis S (2002) The generic genome browser: a building block for a model organism system database. Genome Res 12(10):1599–1610
18. Meng S, Brown DE, Ebbole DJ, Torto-Alalibo T, Oh YY, Deng J, Mitchell TK, Dean RA (2009) Gene ontology annotation of the rice blast fungus, *Magnaporthe oryzae*. BMC Microbiol 9(Suppl 1):S8



Chapter 6

Plant Small RNAs Responsive to Fungal Pathogen Infection

Yun Jin and Hui-Shan Guo

Abstract

Accumulating evidence indicates that small noncoding RNAs (sRNAs) can be transferred across species for interkingdom communication. In addition to the artificial transgene-derived small interfering RNAs (siRNAs), endogenous microRNAs (miRNAs) can also influence interacting organisms to execute a regulatory function. For instance, we have recently found that, in response to infection with *Verticillium dahliae* (*V. dahliae*), cotton plants increase accumulation of miR166 and miR159, which can be exported to the fungal hyphae for specific silencing of virulence genes. These findings suggest a great potential for applying interkingdom mobile miRNAs for crop protection against fungal pathogens. The methods described here provide an approach to identify plant miRNAs and their potential targets in invading fungal pathogens, which will help in revealing the underlying mechanisms of these crosstalk phenomena.

Key words Plant miRNA, Fungal infection, Trans-kingdom regulation

1 Introduction

sRNAs, including siRNAs and miRNAs, are critical regulators of gene function. They are 20–30 nucleotides (nt) in length, and regulate gene expression by guiding mRNA cleavage, inhibiting translation or modifying chromatin [1, 2]. siRNA conventionally refers to endogenously produced sRNA duplexes or RNAs derived from exogenous double-stranded RNAs (dsRNAs) that are taken up by cells. Mature miRNA is single stranded and derived from endogenous stem-loop noncoding RNA. Both siRNA and miRNA are processed by the RNase Drosha or Dicer and then incorporated into the RNA-induced silencing complex (RISC) to negatively regulate target gene expression in a sequence-specific manner [3–5]. In plants, endogenous miRNAs are involved in diverse processes including organ development, signal transduction, and stress responses [6, 7]. It has also been noted that sRNA signals are transmitted between different kingdoms [8, 9].

Wenbo Ma and Thomas Wolpert (eds.), *Plant Pathogenic Fungi and Oomycetes: Methods and Protocols*, Methods in Molecular Biology, vol. 1848, https://doi.org/10.1007/978-1-4939-8724-5_6,
© Springer Science+Business Media, LLC, part of Springer Nature 2018

1.1 Double Stranded siRNA Transfer Across Species

Both animals and plants have been found to exchange double stranded siRNAs with closely interacting pathogenic, parasitic or symbiotic organisms. sRNAs could be trafficked from hosts to interacting organisms to silence their transcripts and suppress their growth. This has been extensively investigated over the last decade and is well known as host-induced gene silencing (HIGS). Transgenic plants and crops expressing dsRNAs that target essential growth and virulence genes of eukaryotic pathogens and pests are resistant/tolerant to disease [10–17]. Thus, the HIGS method provides us with a useful defense strategy for controlling invading pathogens. Moreover, scientists have also shown that this cross-kingdom RNA transfer is bidirectional, fungi are also sending endogenous sRNA to their plant hosts, and the covert operation could be aiding their invasion [18].

1.2 Plant miRNAs Responsive to Fungal Infection and Transferred from Host to Invader

Plant miRNAs play fundamental roles in abiotic and biotic stress responses [19]. To date, the expression profiles of many plant miRNAs have been investigated upon pathogen infection. For example, *Arabidopsis* miR393, induced by infection with the bacterial pathogen *Pseudomonas syringae*, was the first miRNA discovered to be involved in plant immunity [20]. In wheat (*Triticum aestivum L.*), miR393, miR444, miR827, miR2005, and miR2013 are upregulated during interactions with the powdery mildew fungus *Blumeria graminis* [21]. Also eight wheat miRNAs showed diverse transcript profiles when resistant and susceptible wheat cultivars were challenged with stem rust [22]. Both miR160 and miR164 were induced when rice (*Oryza sativa*) was inoculated with *Magnaporthe oryzae* [23]. Tae-miR1432 and a specific wheat miRNA named PC-3P-7484 are highly expressed in response to *Puccinia striiformis f. sp. tritici* (*Pst*) infection [24]. All these studies suggest that host endogenous miRNAs play an important role in plant defense.

Recently, we have found that upon infection with *V. dahliae*, cotton plants increase accumulation of miR166 and miR159 and both are exported to the fungal hyphae for specific silencing of virulence genes [25]. Other researches have revealed that animals can also deliver endogenous miRNAs into interacting organisms [26, 27]. Taken together, these findings suggest the transfer of single-stranded miRNA may be widespread between host and invaders. Although our lab's study stands alone as evidence that plants use RNAi to fight against fungi in the field, plant biologists are convinced that it is a widespread defense strategy. The discovery of this naturally occurring cross-kingdom RNAi phenomenon raises the question of how often it happens in nature. Investigations of more sRNA interkingdom regulation between plants and invading fungi will uncover this mystery.

In this chapter, we describe the cotton miRNA miR166 and its target gene, a Ca^{2+}-dependent cysteine protease (*Clp-1*) of *V. dahl-*

iae as an example to introduce the major experimental steps for the discovery and detection of plant miRNAs that regulate pathogen genes.

2 Materials

2.1 Fungal Strains, Culture Media, Fungal Infection of Plants, and Recovery from Infected Tissue

1. V592: A virulent defoliating *V. dahliae* strain, isolated from cotton in Xinjiang, China.
2. Cotton plants: 'Xinluzao No. 16'.
3. Liquid Czapek-Dox medium: 30 g/L sucrose, 3 g/L $NaNO_3$, 0.5 g/L $MgSO_4.7H_2O$, 0.5 g/L KCl, 0.1 g/L $FeSO_4.7H_2O$, and 1 g/L K_2HPO_4.
4. PDA (potato dextrose agar) medium.
5. 1/10 MS (Murashige and Skoog) liquid medium: 0.44 g/L MS including vitamins.
6. 70% ethanol.
7. 30% sodium hypochlorite.

2.2 RNA Gel Blotting

2.2.1 High Molecular Weight RNA Gel Blots for Detecting mRNA Levels

1. TRIzol reagent (Invitrogen) or other commercial RNA extraction kits.
2. 10× MOPS: 41.8 g/L MOPS, 6.56 g/L NaAc, 20 mM EDTA, adjust pH to 7.0 with NaOH.
3. 100 mL 1 M phosphate buffer, pH = 7.2: 68.4 mL 1 M Na_2HPO_4, 31.6 mL 1 M NaH_2PO_4.
4. 20 × SSC: 175.3 g/L NaCl, 88.2 g/L sodium citrate, adjust pH to 7.0 with HCl.
5. Methylene blue solution: 0.03% methylene-blue in 0.3 M pH = 5.2 NaAc.
6. Probe labeling reagents: [α-^{32}P]-CTP, Rediprime II DNA labeling system (Amersham).
7. Hybridization buffer: 43 mM pH = 7.2 phosphate buffer, 7% SDS, 1% BSA, 20 mM EDTA.

2.2.2 Low Molecular Weight RNA Gel Blots for Detecting miRNA Levels

1. 10× TBE buffer: 108 g/L Tris base, 55 g/L boric acid, and 20 mM EDTA.
2. 60 mL 17% polyacrylamide gel: 34 mL 30% polyacrylamide (acrylamide–bis = 29:1), 3 mL 10 × TBE buffer, 25.2 g urea, 2.5 mL ddH_2O.
3. 480 μL 10% APS (ammonium persulfate).
4. 20 μL TEMED (tetramethylethylenediamine).
5. 10 × MOPS–EDTA–sodium acetate buffer: 400 mM pH 7.0 MOPS, 100 mM sodium acetate, 10 mM pH 8.3 EDTA.

6. 6 × RNA loading buffer: 62.5% (v/v) deionized formamide, 1.14 M formaldehyde, 1.25× MOPS–EDTA–sodium acetate buffer, 200 μg/mL bromophenol blue, 200 μg/mL xylene cyanol FF.
7. Probe labeling reagents: [γ-^{32}P]-ATP, T4 PNK (polynucleotide kinase) (NEB).
8. Hybridization buffer: 50% Formamide,7% SDS, 50 mM pH 7.2 phosphate buffer, 0.3 M NaCl, 5× Denhardt's solution, 10 mg/mL salmon sperm ssDNA.

2.3 Quantitative Real-Time PCR

1. DNaseI.
2. SuperScript III reverse transcriptase (Promega).
3. EvaGreen 2* qPCR MasterMix (abm).
4. CFX96 Real-Time PCR system (Bio-Rad).

2.4 5′ RLM RACE Assay

1. FirstChoice RLM-RACE kit (Ambion).
2. pMD18-T vector.

2.5 Cloning of Fungal Target Genes

1. PCR primers for cloning genes of interest.
2. pGEM-T vector and binary plasmid (e.g., pRF-HU2).
3. Equipment and chemicals needed for PCR and agarose gel electrophoresis. DNA polymerase with high fidelity, such as Pfu enzyme, should be used for PCR.
4. LB medium (Luria–Bertani medium) and agar plates containing appropriate antibiotics.
5. Competent cells of a standard *E. coli* strain (e.g., DH5α).
6. Plasmid extraction kit or solutions (Tiangen).
7. DNA purification kit (Tiangen).
8. ClonExpress II One step cloning kit.
9. USER (Uracil-Specific Excision Reagent) II enzyme (NEB) for cloning a sequence containing uracil, due to its ability to generate a single nucleotide gap at the location of an uracil residue.
10. Competent cells of an *Agrobacterium tumefaciens* strain (e.g., EHA105).
11. Site-Directed Mutagenesis kit.

2.6 Agrobacterium tumefaciens-Mediated Transformation

1. LB/kanrif: LB medium with 50 mg/L kanamycin and 10 mg/L rifampicin.
2. M-N buffer: 30 g/L $MgSO_4 \cdot 7H_2O$, 15 g/L NaCl.
3. K buffer: 200 g/L K_2HPO_4, 145 g/L KH_2PO_4, pH 7.0.

4. IM (induction medium): 20 mL/L M-N buffer, 10 mL/L K buffer, 0.01 g/L $CaCl_2{\cdot}2H_2O$, 2 g/L glucose, 0.001 g/L $FeSO_4{\cdot}7H_2O$, 0.5 g/L NH_4NO_3, 5 mL/L glycerol. After high-pressure steam sterilization, add 2 mL/L 100 mM AS (acetosyringone) and 40 mL/L 1 M pH 5.3 MES.
5. CM (cocultivation medium): IM containing 1.5% (w/v) agar.
6. SM (selection medium): PDA containing 20 μg/mL 5FU (5-fluoro-2′-deoxyuridine), 75 mg/mL hygromycin B or 200 mg/mL G418 as a selection agent for transformants and 200 mg/mL cefotaxime and 200 mg/mL carbenicillin to kill the *Agrobacterium tumefaciens* cells.

3 Methods

3.1 Fungal Infection Using Unimpaired Root-Dip Inoculation Method [28]

1. The isolate V592 and its transformants are stored at −80 °C in the form of a microconidial suspension in 20% glycerol. Cultures are reactivated on PDA medium.
2. To prepare inocula, fungal cultures grow for 5 days in the liquid Czapek-Dox medium are passed through several layers of sterile cheesecloth (to remove mycelia), and the conidial concentration is adjusted to approximately 1×10^7 conidia per mL.
3. Twelve cotton seedlings per pot are planted in 1/10 MS liquid medium in an environmentally controlled growth room at 26 °C, 60–70% relative humidity, on a 16–8-h light–dark cycle, for about 2 weeks.
4. At the third true leaf stage, the seedlings are inoculated by immersing their roots in the conidial suspension for 50 min. The seedlings are then put back in the 1/10 strength MS liquid medium (*see* **Note 1**).
5. The disease ratings are classified on cotton leaves according to one of five levels of symptom severity: 0, no visible wilting or yellowing symptoms; 1, one or two cotyledons wilted or dropped off; 2 and 3, one or two true leaves wilted or dropped off, respectively; and 4, all leaves dropped off or the whole plant has died (*see* **Note 2**).

3.2 Fungal Recovery from Infected Cotton Tissues

1. Twenty days after inoculation with V592, the stem sections immediately under the cotyledons of the infected cotton plants are excised and cut into 1 cm slices (*see* **Note 3**, Fig. 1)
2. The slices are surface sterilized by soaking in 70% ethanol for 1 min and then 15 min in 30% sodium hypochlorite.
3. The samples are rinsed three times with sterile water, and cultured at 26 °C on PDA medium.

4. *V. dahliae* hyphae that grow from the cotton stems after 5 days are cut (away from the stems) and transferred to fresh PDA medium to reculture for another 20 days for morphology observation and RNA extraction (*see* **Note 4**, Fig. 1)

3.3 RNA Extraction and Illumina Hi-Seq Data Analysis

1. *V. dahliae* strains (colonies from PDA medium) are grown at 26 °C in liquid Czapek-Dox medium for 3 days with shaking in the dark at 200 rpm. The resulting mycelia are harvested, and RNAs are isolated using TRIzol according to the manufacturer's instructions.
2. *V. dahliae* recovered from V592-infected cotton (Vda^{Cotton}) are also grown in liquid Czapek-Dox medium and harvested as described above for RNA extraction.
3. sRNA library construction and sRNA sequencing are carried out by BGI (http://www.bgitechsolutions.com/).
4. All clean reads are mapped to the sequenced genome of Verticillium [29, 30] using SOAP (http://soap.genomics.org.cn/).
5. A Perl script is used to search for known miRNAs in cotton [31, 32]. The expression levels of miRNAs are normalized by the reads per million (rpm) value.
6. Fungal targets of miRNAs are predicted using psRNA target (http://plantgrn.noble.org/psRNATarget/) with default parameters.

3.4 RNA Gel Blotting, Quantitative Real-Time PCR, and 5′ RLM RACE Assay

Northern blotting and quantitative real time PCR (qPCR) are carried out to verify the expression of miR166 and its fungal target gene *Clp-1*. 5′ RLM RACE (RNA Ligase Mediated Rapid Amplification of cDNA Ends) is conducted to identify the cleavage sites.

3.4.1 High Molecular Weight RNA Gel Blots for Detecting mRNA Levels

1. 20 μg of total RNA extracted from Subheading 3.3 is separated on a 1.2% agarose gel containing 6% formaldehyde.
2. The RNA is transferred onto a Hybond-N+ membrane and UV cross-linked.
3. The membrane is stained with methylene blue solution for about 5 min and rinsed with distilled water until the bands are clearly visible (*see* **Note 5**).
4. Probe labeling: DNA fragments are labeled with [α-^{32}P] dCTP using the Rediprime II DNA labeling system following the manufacturer's instructions. Briefly, 25 ng DNA is denatured at 98 °C for 5 min and put on ice for 5 min before adding to the labeling mixture, then add 3 μL [α-^{32}P] dCTP, incubate at 37 °C for 1 h following 95 °C 5 min. Such treated DNA is denatured probe used for hybridization (*see* **Note 6**).

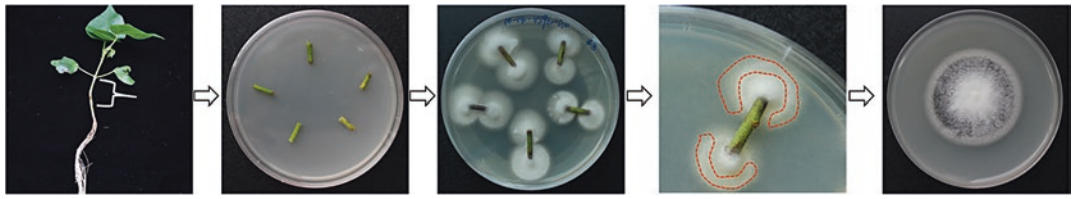


Fig. 1 The hyphae recovery process. The white brace marks the stems immediately under cotyledons. The red dotted line marks the recovery *V. dahliae* away from the stems

5. The blot is prehybridized with hybridization buffer at 65 °C for at least 1 h.
6. Add 20 μL of denatured probe to the hybridization buffer and hybridize overnight at 65 °C.
7. Wash membrane twice with 2 × SSC–2% SDS at 65 °C for 20 min.
8. Wash membrane with 0.2 × SSC–0.2% SDS at 65 °C for 20 min.
9. Expose the blots to X-ray imaging film.

3.4.2 Low Molecular Weight RNA Gel Blots for Detecting miRNA Levels

1. Gently mix the components of a 17% polyacrylamide gel (minus APS and TEMED) in a 100 mL flask, heat the mixture in microwave with high fire for 1 min. Shake gently until urea is dissolved completely.
2. Add 480 μL 10% APS (little by little, from bottom to top), and mix gently. Add 20 μL TEMED (little by little, from bottom to top), mix gently, transfer to a gel casting apparatus and then let it solidify for an hour. Assemble the gel apparatus, and add the running buffer (0.5 × TBE). Make sure there are no leaks.
3. Add 1 volume of deionized formamide to 40 μg of total RNA (dissolved in ddH_2O). Denature at 100 °C for 5–10 min, then put the sample on ice for 5–10 min.
4. Add RNA loading buffer to the small RNA solution and mix quickly. Load RNA samples to gel columns. Run in 0.5 × TBE at 80 V until the bromophenol blue band reaches the bottom of the gel (~16 h for a 20 × 20 cm gel).
5. Cut the appropriate region of the gel and appropriate size of the membrane (about 1 cm larger than the gel), then immerse the gel, membrane, and filter paper in 0.5 × TBE for about 30 s.
6. Set up a RNA transfer sandwich in the Trans-blot Semi-Dry Electrophoretic transfer cell as follows, from bottom to top: filter paper, membrane, gel, filter paper. Make sure to roll out any bubbles between the layers using a glass rod. Transfer membrane for about 30–45 min at X mA (X mA = area (cm^2 of membrane) × 3).

7. Remove the membrane and place it in a UV cross-linker and cross-link at an optimal setting (usually at 1200 × 100 μJ/cm^2 for 2 min).
8. Store the fixed membrane at 4 °C until use.
9. The probe is the reverse complement sequence of miR166. Probe labeling can be accomplished using T4 PNK. The labeling mixture contains 2 μL 10 × PNK buffer, 2 μL 10 μM Probe, 1 μL T4 PNK, 12 μL ddH_2O and 3 μL [γ-^{32}P]-ATP, then incubate at 37 °C for 1 h. For prehybridization, use 20 mL of hybridization buffer. Put the membrane in a hybridization tube and prehybridize at 40 °C for 2 h.
10. Add the labeled oligonucleotide as the probe to the hybridization buffer, hybridize at 40 °C overnight.
11. Wash the membrane with 2 × SSC containing 0.2% SDS at 50 °C for 15–20 min for two to three times.
12. Check radioactivity signals on the membrane with a Geiger counter.
13. Wrap the membrane with a sealing bag and expose to a Phosphoimager to detect hybridization signals.

3.4.3 Quantitative Real-Time PCR

1. Remove residual DNA from total RNA using DNase I, about 2 μL (10 units) DNaseI for 20–50 μg RNA, incubating at 37 °C for 30 min.
2. 2 μg of above treated RNA is then reverse-transcribed into cDNA using SuperScript III reverse transcriptase according to the manufacturer's instructions.
3. The RT-qPCR analysis is performed with a qPCR system (e.g., Biorad CFX96 Real-Time system using EvaGreen 2 × qPCR MasterMix).

3.4.4 5′ RLM RACE Assay

1. The 5' RACE assay is performed using the First Choice RLM-RACE kit. 2 μg total RNA is used for direct ligation to the 5' RACE RNA adaptor, and the subsequent steps are conducted according to the manufacturer's instructions.
2. PCR fragments obtained from 5' RACE can be cloned into the pMD18-T vector, and transformed into *E. coli* DH5α.
3. Individual clones are selected for DNA sequencing (*see* **Note 7**).

3.5 Cloning and Constructs of Target Genes

1. To generate the target gene *Clp-1* deletion construct, we use a homologous recombination method. The 1-kb upstream and downstream genomic sequences are amplified from V592 DNA using the Pfu enzyme. The primers used are as follows: *Clp-1* up-s: GGGTTTAAUGATGAATACTTCGCACCACG, *Clp-1* up-a: GGACTTAAUGTCAGTGGTGCTGCCATCAA,

Clp-1 dn-s: GGCATTAAUACGCAAACCCAGGGCAAAAC, *Clp-1* dn-a: GGTCTTAAUAACTCACGCGGCGGGATACT. Each primer contains the corresponding USER cloning sites at its 5′ end as indicated by the underlined sequence. The PCR products are ligated to pRF-HU2 with the USER enzyme as described previously [33, 34], producing pRF–Clp-1del. The construct is then used to transform *V. dahliae* and generate the *VdaΔclp-1* mutated strains (*see* **Notes 8** and **9**).

2. To obtain *Clp-1* and miRNA resistant *Clp-1 m*, *Clp-1* is amplified from V592 DNA with the primers *Clp-1* gene-s: ATGCACGGCTACAGCTCCTCCGA, *Clp-1* gene-a: CTAGCAGAGCTCCTTCTCAGACT. The PCR product is ligated to the pGEM-T vector to generate T-Clp-1. For anti-miR166 mutagenesis of *Clp-1 m*, oligonucleotide-directed mutagenesis is introduced into *Clp-1* with a pair of primers in opposite orientations, *Clp-1 m*-F: TCGTAAGGGTATCT GGGAAGGCGCGTGGAGTGATGGTTCTCGAG, *Clp-1 m*-R: CTCGAGAACCATCACTCCACGCGCCTTCC CAGATACCCTTACGA. Using a Site-Directed Mutagenesis Kit according to the manufacturer's instructions to generate T-Clp-1 m (*see* **Note 10**, Fig. 2).

3. For the fungi transformation construct of *Clp-1* with endogenous promoters, *Clp-1 m* is amplified from T-Clp-1 m with the following primers: Clp-1 mgene-s: CGGGATCCATGCAC GGCTACAGCTCCTCCGA Clp-1 mgene-a: CTGAATT CCTAGCAGAGCTCCTTCTCAGACT that contain the appropriate enzyme digestion sites (*Bam*HI/*Eco*RI). The *Bam*HI–*Eco*RI fragments of *Clp-1 m* is ligated into *Bam*HI/*Eco*RI-digested binary vector pNEOtef–GFP, derived from pSULPH–GFP [35] in which the chlorimuron-ethyl resistance gene has been replaced by a neo resistance cassette

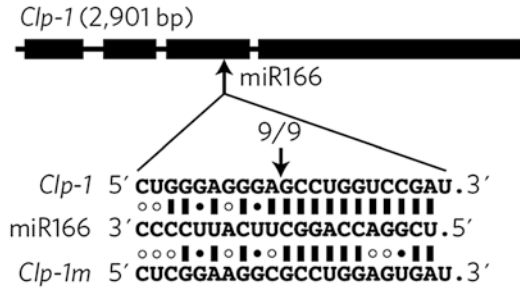


Fig. 2 Alignment of miR166 with fungal targets *Clp-1* and the miRNA-resistant mutated version *Clp-1 m*. The arrow indicates the cleavage site between two nucleotides of target mRNAs detected with the 5′RACE assay. The fraction above the arrow indicates the number of clones with an identical 5′ end detected in the total sequenced clones

from pKOV21 and TEF promoter from pPN94 plasmid [36] to produce pNEOtef–Clp-1 m. The promoters of *Clp-1* is amplified from V592 DNA with the following primers Clp-1pro-F: CACTAGTCAAGGCGGCGGCAGGCAACA Clp-1pro-R: CGGGATCCAGTGGCGGTCGCAGCTCGAG, the promoter PCR fragments is ligated into pNEOtef–Clp-1 m to replace the TEF promoter to produce pNEO–Clp-1 m. These constructs are then used to transfer *V. dahliae*, generating *VdaClp-1 m* mutant strains. pNEOtef–GFP is also transfer to *V. dahliae* to generate the vector control strain *VdaGFP* for infection assays.

3.6 Agrobacterium tumefaciens-Mediated Transformation of V. dahliae

This *A. tumefaciens*-mediated transformation method has been described previously [35, 37].

1. *A. tumefaciens* strain EHA105, containing an appropriate binary vector, such as pRF–Clp-1del and pNEO–Clp-1 m that constructed in Subheading 3.5 is grown at 28 °C for 2 days in liquid medium LB/kanrif.
2. The *A. tumefaciens* cells are suspended to (optical density) OD_{600} = 0.2 in IM and grown for an additional 6 h before mixing them with an equal volume of a conidial suspension of V592 (1×10^7 conidia per mL).
3. The mix (200 mL per plate) is plated on a 0.45-mm pore, 45-mm diameter nitrocellulose filter (Whatman) and placed on CM.
4. After coincubation at 26 °C for 36–48 h, the cultures are washed with 2 mL of sterile water per plate, and then transferred to PDA medium containing the appropriate antibiotics. The 2 mL aliquots are divided into several portions, and pipetted into the same sized PDA plates for about 200 μL/plates, then prompt to be absorbed by the PDA medium by using a spreader.
5. Individual transformants are transferred into PDA medium containing antibiotics and incubated until conidiogenesis.
6. The conidia of individual transformants are suspended in sterile water and plated on PDA medium. Spores from these monoconidial cultures are stored in 20% glycerol at −80 °C until further analysis.

3.7 Conclusion

After these experiments described above, we observed that in response to infection with *Verticillium dahliae*, cotton plants increase the production of miR166 and export it to the fungal hyphae for specific silencing fungal *Clp-1* gene. Furthermore, we found silencing *Clp-1* leads to reduced fungal virulence. Our findings identify a novel defense strategy of host plants by exporting specific miRNAs to induce cross-kingdom gene silencing in pathogenic fungi and confer disease resistance. Since efficient uptake of

engineered host RNAi triggers by distinct pathogenic fungi has been documented, we propose that the export of small silencing RNAs to downregulate virulence gene expression of pathogens represents a conserved host defense strategy against fungal and oomycete pathogens. Therefore, the methods we provide would be helpful in investigating this interkingdom regulation on other pathosystems.

4 Notes

1. The *V. dahliae* infection assay for other plant hosts are the same as cotton, only the immersing time varies. For instance, the roots of *Arabidopsis* seedlings are immersed in conidial suspension for only 2 min.
2. Disease progression in cotton should be recorded over time for at least 1 month. Disease severity is counted by the percent of leaves that show wilting symptoms at each time point. The infection assay should repeat at least three times.
3. The fungal recovery assay should be performed using 20 days inoculated cotton plants. In our experiences, fungal recovered from about 16–22 days inoculated cotton are appropriate for detecting the host delivered miRNAs. After 25 days inoculation, fungal recovered from cotton stem section can hardly detect these plant miRNAs by northern blotting. Another notice is cutting the stem sections immediately under cotyledons of cotton plants as the brace marks in Fig. 1.
4. The hyphae away from the stems are transferred to PDA medium to continue growth, as the area signed in dotted red line in Fig. 1. This is done to make sure to exclude plant tissues in the hyphae sample.
5. Roughly stain with fresh methylene blue solution for 5 min, if the solution is reused, prolong the time to 5–10 min. Generally, rinse with distilled water for five times, if the bands are still not clear, and the washed water is still dark blue, repeat more times.
6. The probe for probing mRNA is a gene specific PCR product about 300–500 bp in length. Generally, 25 ng DNA is sufficient for labeling and hybridization.
7. To avoid the inaccuracy of the direct sequencing of PCR products, we cloned the PCR products into pMD18-T vector and select the positive clones for sequencing by using the common primers like M13.
8. Traditional gene-disruption vectors require multiple cloning steps and elaborate characterization procedures, here we use a USER and dual-selection binary vector PRF-HU2 modified from pGKO-Gateway vector in our lab [34]. The USER sites

can realize single-step cloning, and the herpes simplex virus thymidine kinase negative-selection marker for high efficiency of gene deletion.

9. The USER cloning technique is an efficient one-step cloning strategy, which is convenient for construction of deletion lines by homologous recombination method. The underlined sequences are unique 9-bp extensions, including a deoxyuridine triphosphate, which are introduced into PCR primer pairs for homologous sequence amplification. Thus, each end of the two PCR-amplified homologous sequences carries a unique 9-bp extension for USER cloning. The vector backbone and HPT fragments, generated by restriction enzyme digestion, are incubated together with two PCR-amplified homologous sequences and USER enzyme mix to generate the gene-deletion construct.
10. *Clp-1 m* is the miRNA resistant version that mutated from *Clp-1* at the predicted binding sites. Constructing the miRNA resistant version is to verify the specific targeting of the fungal mRNAs by plant miRNA.

Acknowledgments

This work was supported by grants from the Natural Science Foundation of China (31730078 to H.S.G. and 31700131 to Y.J.).

References

1. Baulcombe D (2005) RNA silencing. Trends Biochem Sci 30:290–293
2. Sontheimer EJ, Carthew RW (2005) Silence from within: endogenous siRNAs and miRNAs. Cell 122:9–12
3. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297
4. He L, Hannon GJ (2004) MicroRNAs: small RNAs with a big role in gene regulation. Nat Rev Genet 5:522–531
5. Carthew RW, Sontheimer EJ (2009) Origins and mechanisms of miRNAs and siRNAs. Cell 136:642–655
6. Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP (2002) MicroRNAs in plants. Genes Dev 16:1616–1626
7. Jones-Rhoades MW, Bartel DP, Bartel B (2006) MicroRNAs and their regulatory roles in plants. Annu Rev Plant Biol 57:19–53
8. Knip M, Constantin ME, Thordal-Christensen H (2014) Trans-kingdom cross-talk: small RNAs on the move. PLoS Genet 10:e1004602
9. Wang M, Thomas N, Jin H (2017) Cross-kingdom RNA trafficking and environmental RNAi for powerful innovative pre- and post-harvest plant protection. Curr Opin Plant Biol 38:133–141
10. Nowara D, Gay A, Lacomme C, Shaw J, Ridout C, Douchkov D et al (2010) HIGS: host-induced gene silencing in the obligate biotrophic fungal pathogen Blumeria graminis. Plant Cell 22:3130–3141
11. Koch A, Kumar N, Weber L, Keller H, Imani J, Kogel KH (2013) Host-induced gene silencing of cytochrome P450 lanosterol C14alpha-demethylase-encoding genes confers strong resistance to Fusarium species. Proc Natl Acad Sci U S A 110:19324–19329
12. Ghag SB, Shekhawat UK, Ganapathi TR (2014) Host-induced post-transcriptional hair-

pin RNA-mediated gene silencing of vital fungal genes confers efficient resistance against Fusarium wilt in banana. Plant Biotechnol J 12:541–553

13. Nunes CC, Dean RA (2012) Host-induced gene silencing: a tool for understanding fungal host interaction and for developing novel disease control strategies. Mol Plant Pathol 13:519–529
14. Yin C, Jurgenson JE, Hulbert SH (2011) Development of a host-induced RNAi system in the wheat stripe rust fungus Puccinia striiformis f. Sp. tritici. Mol Plant-Microbe Interact 24:554–561
15. Zhang T, Jin Y, Zhao JH, Gao F, Zhou BJ, Fang YY et al (2016) Host-induced gene silencing of the target gene in fungal cells confers effective resistance to the cotton wilt disease pathogen Verticillium dahliae. Mol Plant 9:939–942
16. Westwood JH, Roney JK, Khatibi PA, Stromberg VK (2009) RNA translocation between parasitic plants and their hosts. Pest Manag Sci 65:533–539
17. Mao YB, Cai WJ, Wang JW, Hong GJ, Tao XY, Wang LJ et al (2007) Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. Nat Biotechnol 25:1307–1313
18. Weiberg A, Wang M, Lin FM, Zhao H, Zhang Z, Kaloshian I et al (2013) Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. Science 342:118–123
19. Sunkar R, Zhu JK (2004) Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. Plant Cell 16:2001–2019
20. Navarro L, Dunoyer P, Jay F, Arnold B, Dharmasiri N, Estelle M et al (2006) A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. Science 312:436–439
21. Xin M, Wang Y, Yao Y, Xie C, Peng H, Ni Z et al (2010) Diverse set of microRNAs are responsive to powdery mildew infection and heat stress in wheat (Triticum aestivum L.). BMC Plant Biol 10:123
22. Gupta OP, Permar V, Koundal V, Singh UD, Praveen S (2012) MicroRNA regulated defense responses in Triticum aestivum L. during Puccinia graminis f.sp tritici infection. Mol Biol Rep 39:817–824
23. Li Y, Lu YG, Shi Y, Wu L, Xu YJ, Huang F et al (2014) Multiple rice microRNAs are involved in immunity against the blast fungus Magnaporthe oryzae. Plant Physiol 164: 1077–1092
24. Feng H, Wang T, Feng CX, Zhang Q, Zhang XM, Huang LL et al (2016) Identification of microRNAs and their corresponding targets involved in the susceptibility interaction of wheat response to Puccinia striiformis f. Sp tritici. Physiol Plantarum 157:95–107
25. Zhang T, Zhao YL, Zhao JH, Wang S, Jin Y, Chen ZQ et al (2016) Cotton plants export microRNAs to inhibit virulence gene expression in a fungal pathogen. Nat Plants 2:16153
26. Liu SR, da Cunha AP, Rezende RM, Cialic R, Wei ZY, Bry L et al (2016) The host shapes the gut microbiota via fecal microRNA. Cell Host Microbe 19:32–43
27. LaMonte G, Philip N, Reardon J, Lacsina JR, Majoros W, Chapman L et al (2012) Translocation of sickle cell erythrocyte microRNAs into Plasmodium falciparum inhibits parasite translation and contributes to malaria resistance. Cell Host Microbe 12:187–199
28. Zhou BJ, Jia PS, Gao F, Guo HS (2012) Molecular characterization and functional analysis of a necrosis- and ethylene-inducing, protein-encoding gene family from Verticillium dahliae. Mol Plant-Microbe Interact 25:964–975
29. de Jonge R, Bolton MD, Kombrink A, van den Berg GCM, Yadeta KA, Thomma BPHJ (2013) Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. Genome Res 23:1271–1282
30. Klosterman SJ, Subbarao KV, Kang SC, Veronese P, Gold SE, Thomma BPHJ et al (2011) Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. PLoS Pathog 7:e1002137
31. Gong L, Kakrana A, Arikit S, Meyers BC, Wendel JF (2013) Composition and expression of conserved microRNA genes in diploid cotton (Gossypium) species. Genome Biol Evol 5:2449–2459
32. Liu NA, Tu LL, Tang WX, Gao WH, Lindsey K, Zhang XL (2014) Small RNA and degradome profiling reveals a role for miRNAs and their targets in the developing fibers of Gossypium barbadense. Plant J 80:331–344
33. Frandsen RJ, Frandsen M, Giese H (2012) Targeted gene replacement in fungal pathogens via agrobacterium tumefaciens- mediated transformation. Methods Mol Biol 835:17–45
34. Wang S, Xing H, Hua C, Guo HS, Zhang J (2016) An improved single-step cloning strategy simplifies the agrobacterium tumefaciens-mediated transformation (ATMT)-based gene-disruption method for Verticillium dahliae. Phytopathology 106:645–652

35. Gao F, Zhou BJ, Li GY, Jia PS, Li H, Zhao YL et al (2010) A glutamic acid-rich protein identified in Verticillium dahliae from an insertional mutagenesis affects microsclerotial formation and pathogenicity. PLoS One 5:e15319
36. Takemoto D, Tanaka A, Scott B (2006) A p67Phox-like regulator is recruited to control hyphal branching in a fungal-grass mutualistic symbiosis. Plant Cell 18:2807–2821
37. Mullins ED, Chen X, Romaine P, Raina R, Geiser DM, Kang S (2001) Agrobacterium-mediated transformation of Fusarium oxysporum: an efficient tool for insertional mutagenesis and gene transfer. Phytopathology 91:173–180



Chapter 7

Sequential Phosphopeptide Enrichment for Phosphoproteome Analysis of Filamentous Fungi: A Test Case Using *Magnaporthe oryzae*

Yeonyee Oh, William L. Franck, and Ralph A. Dean

Abstract

A number of challenges have to be overcome to identify a complete complement of phosphorylated proteins, the phosphoproteome, from cells and tissues. Phosphorylated proteins are typically of low abundance and moreover, the proportion of phosphorylated sites on a given protein is generally low. The challenge is further compounded when the tissue from which protein can be recovered is limited. Global phosphoproteomics primarily relies on efficient enrichment methods for phosphopeptides involving affinity binding coupled with analysis by fast high-resolution mass spectrometry (MS) and subsequent identification using various software packages. Here, we describe an effective protocol for phosphopeptide enrichment using an Iron-IMAC resin in combination with titanium dioxide (TiO_2) beads from trypsin digested protein samples of the filamentous fungus *Magnaporthe oryzae*. Representative protocols for LC-MS/MS analysis and phosphopeptide identification are also described.

Key words Mass spectrometry, Phosphopeptide enrichment, *Magnaporthe oryzae*, Iron-IMAC, TiO_2, Fungus

1 Introduction

Posttranslational modification (PTMs) of proteins such as phosphorylation adds a further layer of complexity to cellular regulation beyond transcription and translation. Although there are hundreds of possible PTMs of proteins, the addition of phosphate to serine, threonine or tyrosine is a primary and pervasive modification affecting protein activity, stability, location and interactions with other proteins [1]. Indeed, studies have suggested that half of an organism's proteome may be phosphorylated to some extent at any particular time [2, 3]. Given the importance and prevalence of protein phosphorylation in regulating proteins and cellular function, efforts in recent years have focused on protein phosphorylation dynamics at a global scale

Wenbo Ma and Thomas Wolpert (eds.), *Plant Pathogenic Fungi and Oomycetes: Methods and Protocols*, Methods in Molecular Biology, vol. 1848, https://doi.org/10.1007/978-1-4939-8724-5_7,
© Springer Science+Business Media, LLC, part of Springer Nature 2018

within cells and tissues [4–7]. Current techniques for exploration of the phosphoproteome are predominantly based on LC-MS/MS peptide analysis of trypsin digested protein samples following phosphopeptide enrichment [8, 9].

A number of challenges need to be overcome in order to achieve a holistic view of the phosphoproteome. Many proteins that are phosphorylated are regulatory proteins and of low abundance. Furthermore, the stoichiometry of phosphorylation of a given protein is low, thus modified peptides are buried in an excess of unmodified peptides [10]. Other issues are related to mass spectrometry. Loss of phosphate groups during phosphopeptide fragmentation can greatly affect the ability to identify phosphopeptides. Furthermore, assigning the exact location of the phosphorylation site in a peptide requires highly accurate MS/MS data. Many of these issues have been overcome through use of modern mass spectrometers such as the Q-Exactive HF Hybrid Quadrupole-Orbitrap mass spectrometer. Such instruments offer increased dynamic range and speed as well as the use of higher energy collision dissociation (HCD) which reduces neutral loss and enables the desired fragmentation for phosphopeptide identification [8, 11].

To overcome challenges associated with the relatively low representation of phosphoproteins in the proteome and obtain adequate coverage of modified sites, enrichment is needed. Protocols to enrich phosphoproteins are mainly based on use of antibodies and immunoprecipitation. Such strategies are unsuitable for large-scale studies and typically yield modest enrichment because only a small proportion of trypsin-derived peptides from a phosphorylated protein contain phosphate group(s) [12]. Strategies based on enrichment of phosphorylated peptides following trypsin digestion of proteins offer more flexibility and typically are superior, particularly when different phosphopeptide enrichment strategies are used in combination [13, 14].

Here, we describe a protocol developed for enrichment (and subsequent analysis) of phosphopeptides from the rice blast fungus *Magnaporthe oryzae*, the most destructive disease of rice worldwide. The protocol as shown in Fig. 1 employs sequential enrichment, first using immobilized metal affinity chromatography (IMAC) to capture phosphopeptides via interaction with chelated iron. Subsequently, the unbound fraction (or flow through) from the IMAC enrichment is subjected to metal oxide affinity chromatography (MOAC) using titanium dioxide (TiO_2) particles to capture additional phosphopeptides. The two phosphopeptide enriched samples are then pooled prior to LC-MS/MS.

The following references provide background on these enrichment approaches [9, 12]. IMAC has relatively poor selectivity for phosphopeptides; peptides with highly acidic residues such as aspartic and glutamic acids can bind nonspecifically. However,

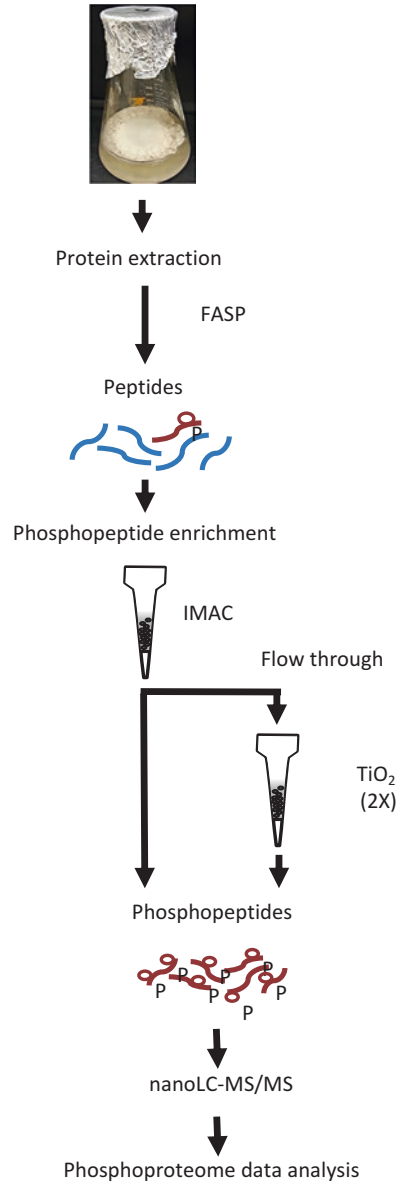


Fig. 1 Workflow of the sequential phosphopeptide enrichment and analysis protocol

binding of multiphosphorylated peptides is enhanced under acidic conditions. Conversely, metal oxides such as TiO_2 interact with negatively charged phosphate groups particularly with singly phosphorylated peptides under acidic conditions. Although considered to provide better selectivity than IMAC, acidic side chains of other amino acids can form ionic interactions with TiO_2. Binding specificity can be enhanced by modifying the acidic conditions (*see* [9]). Careful attention to pH and composition of loading and elution buffers during the sequential binding to first IMAC followed by

TiO_2 reduces nonspecific binding and affords excellent enrichment of both singly and multiphosphorylated peptides.

Using this approach to enrich phosphopeptides and using a nanoLC MS/MS high-resolution ion trap (Q-Exactive) mass spectrometer with data processed in MaxQuant, we identified and were able to examine changes in phosphorylation at 2924 phosphosites on 1914 phosphoproteins in conidia, mycelia and during conidial germination and appressorium formation of *M. oryzae* [15]. The protocol presented here is applicable to any fungal species or tissues and requires only 250 μg of protein input. This is of critical importance when protein samples from certain specialized tissues such as infection structures or sexual reproductive bodies are limited.

2 Materials

Various fresh or rapidly frozen tissues from *Magnaporthe oryzae* or other fungi can be used including but not limited to conidia, mycelium, developing appressoria, or sexual structures (*see* **Note 1**).

2.1 Protein Extraction

1. Cell lysis buffer: 1 × phosphate-buffered saline (PBS), 0.1% SDS, 2 M urea, and a PhosSTOP phosphatase inhibitor cocktail (Roche, Mannheim, Germany, 1 tablet per 10 mL buffer).
2. 0.5 mm zirconia/silica beads (BioSpec Products Inc., Bartlesville, OK).
3. 1.7 mL microcentrifuge tubes.
4. Bead mill homogenizer.
5. A refrigerated benchtop centrifuge (such as Eppendorf 5810R) equipped with a fixed angle rotor (such as F45-30-11, Eppendorf, Hauppauge, NY).
6. A bicinchoninic acid (BCA) assay kit.

2.2 Protein Digestion and Peptide Purification by Filter Aided Sample Preparation (FASP)

1. 50 mM dithiothreitol (DTT).
2. 8 M urea buffer: 8 M urea in 0.1 M Tris–HCl (pH 8.5) (*see* **Note 2**).
3. 0.05 M iodoacetamide in 8 M urea buffer (*see* **Note 2**).
4. 0.05 M ammonium bicarbonate in H_2O (ABC).
5. Stock trypsin solution (0.25 μg/μL) in ABC.
6. 100% acetic acid.
7. 1.7 mL microcentrifuge tubes.
8. Vivacon 500 30 kDa MW cutoff filter units (Vivaconproducts, Littleton, MA).

9. A refrigerated bench-top centrifuge (*see* Subheading 2.1, **item 5**).
10. Nanodrop 2000c (Thermo Scientific, Wilmington, DE).

2.3 Phosphopeptide Enrichment by Immobilized Metal Ion Affinity Chromatography (IMAC)

1. NTA Agarose (Qiagen, Valencia, CA).
2. 1% and 2% acetic acid.
3. 100 mM $FeCl_3$.
4. 10 micron filter paper.
5. 1.7 mL microcentrifuge tubes.
6. 10 mL Luer-Lock syringe.
7. 200 µL gel-loading tips.
8. A solution of 74% 100 mM NaCl, 25% acetonitrile and 1% acetic acid.
9. Molecular biology grade H_2O.
10. 5% NH_4OH.
11. Formic acid.
12. Benchtop tube rotator.

2.4 Phosphopeptide Enrichment by Titanium Dioxide (TiO_2) Chromatography

1. TiO_2 beads (10 µm Titansphere, GL Sciences, Torrance, CA).
2. A solution of 2% acetic acid and 200 mg/mL lactic acid.
3. 200 µL gel-loading tips.
4. 1.7 mL microcentrifuge tubes.
5. Pipette tip boxes.
6. Microcentrifuge rack.
7. A refrigerated benchtop centrifuge (*see* Subheading 2.1, **item 5**).
8. A solution of 74% 100 mM NaCl, 25% acetonitrile and 1% acetic acid.
9. Molecular biology grade H_2O.
10. 5% NH_4OH.
11. Formic acid.
12. Vacuum concentrator.
13. Mobile phase A (98% water, 2% acetonitrile, and 0.2% formic acid).

2.5 Analysis by Mass Spectrometry

1. A high-resolution, high mass accuracy mass spectrometer capable of performing MS/MS such as an Orbitrap based instrument (LTQ Orbitrap XL or Q Exactive HF, Thermo Fisher Scientific) with an online nanoHPLC system (such as Eksigent cHiPLC-nanoflex) using a 50–200 µm i.d. C18 column for peptide separation.

2. Mobile phase A: 98% water, 2% acetonitrile, 0.2% formic acid, and mobile phase B: 98% acetonitrile, 2% water, and 0.2% formic acid.
3. Software for processing of MS/MS output data such as Mascot Distiller (Matrix Science), Proteome Discoverer (Thermo Scientific) and MaxQuant ([16], www.maxquant.org) for identifying proteins from peptide sequence databases with the ability to search for peptide modifications.
4. Genome sequence of *M. oryzae* (version 8) downloaded from NCBI BioProject PRJNA13840.

3 Methods

3.1 Protein Extraction

1. Homogenize fungal tissue by bead beating for 2 min in 100 µL cell lysis buffer using ~150 mg of 0.5 mm zirconia/silica beads.
2. Heat the sample for 5 min at 95 °C and centrifuge at 13,000 × *g* at 4 °C for 5 min.
3. Transfer the supernatant to a new tube.
4. Measure the protein concentration using a bicinchoninic (BCA) acid kit following manufacturer's protocol.
5. Store protein sample at −80 °C before use.

3.2 Protein Digestion and Peptide Purification by Filter Aided Sample Preparation (FASP)

1. Up to 250 µg of *M. oryzae* protein sample is dried down to 27 µL and 3 µL of 50 mM Dithiothreitol (DTT) is added to reach a final DTT concentration of 5 mM in a 1.7 mL microcentrifuge tube.
2. Incubate the sample for 30 min at 56 °C to reduce the protein disulfide bonds.
3. Mix the sample with 100 µL of 8 M urea buffer and transfer to a Vivacon 500 30 kDa MW cutoff filter unit.
4. Centrifuge the filter unit at 21 °C at 14,000 × *g* for 15 min and discard the flow-through.
5. Add 200 µL of 8 M urea buffer and centrifuge the filter unit at 21 °C at 14,000 × *g* for 15 min. Discard the flow-through.
6. Add 100 µL of 0.05 M iodoacetamide to the filter unit and incubate for 30 min in the dark at room temperature for alkylation of the free thiols.
7. Centrifuge the filter unit at 21 °C at 14,000 × *g* for 15 min. Discard the flow-through.
8. Add 100 µL 8 M urea buffer to the filter unit and centrifuge at 14,000 × *g* for 10 min and discard the flow-through. Repeat this step two more times.

9. Add 100 µL 0.05 M ammonium bicarbonate (ABC) to the filter unit and centrifuge at 14,000 × *g* for 10 min and discard the flow-through. Repeat this step two more times.
10. Move the filter unit to new collection tube.
11. Add 30 µL ABC and 10 µL of stock trypsin solution (1 µg of enzyme per 100 µg of protein) to the filter unit and vortex.
12. Incubate the filter unit for 16 h at 37 °C.
13. Centrifuge the filter unit at 21 °C and 14,000 × *g* for 15 min. Do not discard the filtrate.
14. Add 40 µL ABC and centrifuge the filter unit at 21 °C and 14,000 × *g* for 15 min.
15. Acidify the filtrate with 1.7 µL of 100% acetic acid (final acetic acid concentration ~2%).
16. Measure peptide concentration using NanoDrop 2000c (Thermo Scientific, Wilmington, DE).

3.3 Phosphopeptide Enrichment by Immobilized Metal Ion Affinity Chromatography (IMAC) (See Note 3)

1. Add 400 µL of NTA Agarose resin to 1.7 mL microcentrifuge tube.
2. Centrifuge at 500 × *g* for 2 min and remove supernatant (*see* **Note 4**).
3. Mix 1 mL of 1% acetic acid with the resin, centrifuge at 500 × *g* for 2 min and remove supernatant.
4. To prepare the iron-IMAC resin, add 1 mL of 100 mM $FeCl_3$ in 1% acetic acid to the NTA Agarose resin and incubate for 4 h with rotation in the dark.
5. Store the iron-IMAC resin at 4 °C before use.
6. Load 100 µL of the iron-IMAC resin into a 200 µL gel-loading tip fitted with a 10 micron filter paper plug (*see* **Note 5**).
7. Wash the resin twice with 100 µL of 2% acetic acid (*see* **Note 6**).
8. Load total tryptic peptides onto the resin.
9. Collect the flow-through and reload it onto the column.
10. Set aside the flow-through for TiO_2 enrichment described below.
11. Wash the iron-IMAC resin twice with 100 µL 2% acetic acid.
12. Wash the resin twice with 100 µL of a solution containing 74% 100 mM NaCl, 25% acetonitrile and 1% acetic acid.
13. Wash the resin twice with 100 µL of molecular biology grade H_2O.
14. Elute phosphopeptides twice with 100 µL of 5% NH_4OH and combine eluents.
15. Acidify the eluate with 30 µL of formic acid and store at −80 °C.

3.4 Phosphopeptide Enrichment by Titanium Dioxide (TiO_2) Chromatography (see Note 8)

1. Suspend 1.5 mg of TiO_2 beads in a solution of 2% acetic acid and 200 mg/mL lactic acid in a 1.7 mL microcentrifuge tube.
2. Load the TiO_2 beads into a 200 µL gel-loading tip fitted with a 10 micron filter paper plug (*see* **Note 5**).
3. Wash the TiO_2 beads twice with 100 µL a solution of 2% acetic acid and 200 mg/mL lactic acid (*see* **Note 7**).
4. Apply the flow-through from the iron-IMAC enrichments to the TiO_2 beads.
5. Collect the flow-through and reapply to the TiO_2 beads.
6. Wash the TiO_2 beads twice with 100 µL of a solution of 2% acetic acid and 200 mg/mL lactic acid.
7. Wash the beads twice with 100 µL of a solution of 74% 100 mM NaCl, 25% acetonitrile and 1% acetic acid.
8. Wash the beads twice with 100 µL of molecular biology grade H_2O.
9. Elute phosphopeptides twice with 100 µL of 5% NH_4OH and combine eluents.
10. Acidify the eluate with 30 µL of formic acid (pH should be <4) (*see* **Note 9**).
11. Subject the eluate to a second round of TiO_2 enrichment using the same protocol and fresh TiO_2 beads (**steps 1–10**).
12. Pool the phosphopeptides from the iron-IMAC and TiO_2 enrichment steps and dry to near completion and store at −80 °C.
13. Resuspend the phosphopeptides in 25 µL mobile phase A prior to LC-MS/MS analysis.

3.5 Analysis by Mass Spectrometry (see Note 10)

1. A representative setup involves a 15 cm × 75 µm nanoLC analytical column containing 3 µm C18 resin coupled to a high-resolution ion trap (Orbitrap Q-Exactive) mass spectrometer as described in our previous work [15]. Phosphopeptides are typically eluted from the nanoLC with a linear gradient of 5% to 40% acetonitrile (ACN) over 3 h using mobile phase A and mobile phase B listed in materials. These conditions are optimal for use with a quadrupole Orbitrap mass spectrometer such as the Q-Exactive in data dependent mode using a top 12 method. Resolving power is set at $70{,}000_{FWHM}$ at 200 *m/z* for MS acquisition and $17{,}500_{FWHM}$ at 200 *m/z* for MS/MS acquisition. AGC settings for MS and MS/MS is 1E6 and 2E4, respectively (*see* [15]). The nanoLC setup may need modification for more complex samples and other mass spectrometers and may require longer columns with smaller chromatographic beads to improve phosphopeptide resolving power.

2. Phosphopeptide identification and analysis is performed using mass spectrometry software packages such as MaxQuant. MS/MS output data is first searched using the Andromeda search engine within MaxQuant against the *M. oryzae* genome sequence with reverse database functioned enabled set to protein, peptide and site false discovery rates of 1%. The Phospho (STY) Site output file of MaxQuant is used to obtain phosphorylation site information. High confidence sites are defined as having a location probability greater than or equal to 0.75 and a score difference greater than or equal to five. The Perseus module of MaxQuant is used for quantification and to examine differential changes in protein phosphorylation as described in [15]. Further statistical tests can be performed using the JMP software package (SAS).

4 Notes

1. High quality chemicals should be used.
2. Urea and iodoacetamide solutions must be freshly made.
3. If necessary, preliminary or trouble shooting experiments can be carried out with radiolabelled model peptides such as ^{32}P Kemptide added to the peptide sample to evaluate efficiency of binding to IMAC and TiO_2. Eluates can be counted by liquid scintillation, compared to starting material and efficiencies calculated. Typically, we found our sequential enrichment method yielded >95% recovery.
4. The agarose beads are fragile, so high-speed centrifugation should be avoided.
5. To generate enrichment columns prepared using gel loading tips, small disks are punched from 10-micron filter paper and used to plug the gel loading tips. Any brand is suitable providing the tip is narrow. Remove the top of gel loading tips with a razor blade so that it provides a tight fit to 1 mL Luer-Lock syringe barrel. Check flow rate with 30 μL acetonitrile using gentle pressure on the syringe barrel. There should be no or little backpressure. *See* [17] for more details.
6. For IMAC enrichment, all washes are performed by forcing the solution through the resin using a 10 mL Luer-Lock syringe fitted to the top of the gel-loader tip.
7. For TiO_2 enrichment, all washes can be performed as indicated for IMAC; however, the flow rate will be slow. If necessary, washes can be performed by centrifugation using a bench-top centrifuge (such as a Sigma 4K15C) fitted with swinging bucket rotors (such as Sigma EXG-1020) and deep well microtiter plate holders (such as Sigma 09366). In this instance,

a microcentrifuge rack is modified to fit inside a gel loading tip pipette box, which fits tightly in the microtiter plate holder. 1.7 mL microcentrifuge tubes are placed in the rack directly under each gel-loading tips. Samples are centrifuged at 1000 × *g* for a few minutes to collect eluent.

8. The method described here is suitable for processing between 1–12 samples simultaneously. For enrichment of more than 12 samples, collecting eluents from TiO_2 beads by centrifugation may be preferred, *see* **Note 7**.
9. Care should be taken in preparation and use of the tip loading columns to ensure no IMAC or TiO2 resin leaks from the columns during phosphopeptide elution as phosphopeptides may rebind to the resin. If necessary, the resin can be removed by centrifugation, prior to final acidification of the eluent.
10. Care must be exercised when performing quantitative phosphoproteome analyses involving different tissues or time course studies as protein levels may change. In such cases, global proteome studies should be conducted in conjunction with analyses of the phosphoproteome to enable accurate phosphosite normalization.

Acknowledgments

Support for this work was provided the National Science Foundation (MCB-0918611), the National Institute of Health Molecular Mycology and Pathogenesis Training program (5T32AI052080), and North Carolina State University to R.A.D.

References

1. Khoury GA, Baliban RC, Floudas CA (2011) Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. Sci Rep 1:90. https://doi.org/10.1038/srep00090
2. Yachie N, Saito R, Sugiyama N et al (2011) Integrative features of the yeast phosphoproteome and protein-protein interaction map. PLoS Comput Biol 7:e1001064. https://doi.org/10.1371/journal.pcbi.1001064
3. Wilhelm M, Schlegl J, Hahne H et al (2014) Mass-spectrometry-based draft of the human proteome. Nature 509:582–587. https://doi.org/10.1038/nature13319
4. Zhou H, Di Palma S, Preisinger C et al (2013) Toward a comprehensive characterization of a human Cancer cell Phosphoproteome. J Proteome Res 12:260–271. https://doi.org/10.1021/pr300630k
5. Rosenberg A, Soufi B, Ravikumar V et al (2015) Phosphoproteome dynamics mediate revival of bacterial spores. BMC Biol 13:76. https://doi.org/10.1186/s12915-015-0184-7
6. Humphrey SJ, Azimifar SB, Mann M (2015) High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics. Nat Biotechnol 33:990–995. https://doi.org/10.1038/nbt.3327
7. Lasonder E, Treeck M, Alam M, Tobin AB (2012) Insights into the Plasmodium falciparum schizont phospho-proteome. Microbes

Infect 14:811–819. https://doi.org/10.1016/j.micinf.2012.04.008

8. Leitner A (2016) Enrichment strategies in phosphoproteomics. In: Von Stecow L (ed) Methods in molecular biology. Springer, New York, pp 105–121
9. Dunn J, Reid G, Bruening ML (2010) Techniques for phosphopeptide enrichment prior to analysis by mass spectrometry. Mass Spectrom Rev 29:29–54. https://doi.org/10.1002/mas
10. Mann M, Ong S, Gr M et al (2002) Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. Trends Biotechnol 20:261–268. https://doi.org/10.1016/S0167-7799(02)01944-3
11. Olsen JV, Macek B, Lange O et al (2007) Higher-energy C-trap dissociation for peptide modification analysis. Nat Methods 4:709–712. https://doi.org/10.1038/nmeth1060
12. Fílla J, Honys D (2012) Enrichment techniques employed in phosphoproteomics. Amino Acids 43:1025–1047. https://doi.org/10.1007/s00726-011-1111-z
13. Engholm-Keller K, Birck P, Størling J et al (2012) TiSH — a robust and sensitive global phosphoproteomics strategy employing a combination of TiO2, SIMAC, and HILIC. J Proteome 75:5749–5761. https://doi.org/10.1016/j.jprot.2012.08.007
14. Thingholm TE, Jensen ON, Robinson PJ, Larsen MR (2008) SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. Mol Cell Proteomics 7:661–671. https://doi.org/10.1074/mcp.M700362-MCP200
15. Franck WL, Gokce E, Randall SM et al (2015) Phosphoproteome analysis links protein phosphorylation to cellular remodeling and metabolic adaptation during Magnaporthe oryzae appressorium development. J Proteome Res 14:2408–2424. https://doi.org/10.1021/pr501064q
16. Tyanova S, Temu T, Cox J (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc 11:2301–2319. https://doi.org/10.1038/nprot.2016.136
17. Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc 2:1896–1906. https://doi.org/10.1038/nprot.2007.261



Chapter 8

Assays for MAP Kinase Activation in *Magnaporthe oryzae* and Other Plant Pathogenic Fungi

Xue Zhang, Zhuyun Bian, and Jin-Rong Xu

Abstract

Mitogen-activated protein (MAP) kinases have the hallmark motif TXY and function in key signal transduction pathways in eukaryotic organisms. Most ascogenous plant pathogenic fungi have three MAPK pathways that regulate different developmental and infection processes. In the rice blast fungus *Magnaporthe oryzae*, the Pmk1 and Mps1 MAP kinases with the TEY motif are essential for appressorium formation, penetration, and invasive growth. Osm1 is the third MAP kinase that has the TGY motif and functions in osmoregulation. Although orthologs of Pmk1 and Mps1 are important for pathogenesis in all the plant pathogens studied, Osm1 orthologs have species-specific roles in stress responses and pathogenesis. Because of their functions in fungal development and pathogenesis, it is important to determine the expression and activation of MAP kinases under different growth conditions or infection stages. In this chapter, we describe methods for protein extraction and detection of the activation of the three MAP kinases in *M. oryzae* with the commercially available anti-TpEY or anti-TpGY phosphorylation-specific antibodies. Similar approaches can be used to monitor MAP kinase activation in other plant pathogenic fungi.

Key words Signal transduction, Fungal pathogens, Pathogenesis, Infection-related morphogenesis

1 Introduction

Like in other eukaryotic organisms, mitogen-activated protein (MAP) kinases play important roles in fungal pathogens in regulating various differentiation and infection processes in response to environmental stimuli or host signals [1, 2]. The hallmark of MAP kinases (MAPKs) is the TXY (Thr-any amino acid-Tyr) motif in the activation loop. Usually, an MAPK pathway consists of protein kinases in three tiers, in which an MAPK is activated by an MAPK kinase (MEK) that is activated by an upstream MEK kinase. The activated MAPK then regulates specific responses to external signals by phosphorylating various downstream targets [3, 4].

In the rice blast fungus *Magnaporthe oryzae*, a model organism that has been used to study fungal–plant interactions [5], three MAP kinase pathways have been characterized. The Pmk1 cascade

Wenbo Ma and Thomas Wolpert (eds.), *Plant Pathogenic Fungi and Oomycetes: Methods and Protocols*, Methods in Molecular Biology, vol. 1848, https://doi.org/10.1007/978-1-4939-8724-5_8,
© Springer Science+Business Media, LLC, part of Springer Nature 2018

is essential for appressorium formation, penetration, and invasive growth [6, 7]; the Mps1 pathway is important for cell wall integrity, conidiation, and plant infection [6, 8]; and the third MAPK, Osm1, is important for osmoregulation in hyphae but dispensable for appressorium turgor generation and plant infection [6, 9]. Unlike the Pmk1 and Mps1 MAPKs that have the TEY (Thr-Glu-Tyr) motif, Osm1 has the TGY (Thr-Gly-Tyr) motif, typical for MAPKs involved in stress responses in other organisms [9].

Studies in a number of plant pathogenic fungi have showed that the MAP kinase cascades are well conserved although their functions may differ among different pathogens [1, 10]. Like in yeast and other eukaryotes, the phosphorylated or activated form of MAP kinases are responsible for activating downstream targets in plant pathogenic fungi and regulating specific biological functions in response to signals recognized by cell surface receptors [11]. Therefore, it is often necessary to assay the expression and phosphorylation levels of MAP kinases in order to characterize their functions during fungal differentiation and infection processes. Commercial antibodies have been generated against synthetic phosphorylated peptides to aid detection of the activation of specific MAPKs in human and animal cells, including p44/42 or Erk1/2 (containing the TEY motif) and p38 (containing the TGY motif) [12, 13]. Because TEY and TGY motifs are conserved in MAPKs, the antibodies developed for human p44/42 and p38 can also be used to detect the phosphorylation of fungal MAPKs including *M. oryzae*, *Fusarium graminearum*, and other pathogens [14, 15].

This chapter describes the detection of Pmk1 and Mps1 activation in *M. oryzae* using the phosphorylation-specific anti-TpEY antibody (phospho-p44/42 MAPK), and Osm1 activation using the phosphorylation-specific anti-TpGY antibody (Phospho-p38 MAPK). The protein levels of Mps1/Pmk1 or Osm1 MAPKs are determined using an anti-MAPK ERK-2 antibody or an anti-Hog1 MAPK antibody respectively.

2 Materials

2.1 Protein Isolation from Fungal Hyphae

1. 5 × YEG liquid medium: 5 g yeast extract, 10 g glucose, 1 L double-distilled water (DDW). Autoclave for 20 min.
2. Miracloth.
3. Lysis buffer: 50 mM Tris–HCl (pH 7.4), 150 mM NaCl, 1 mM EDTA, 1% Triton X-100. Autoclave for 20 min.
4. Protease inhibitor cocktail (Sigma-Aldrich, St. Louis, MO). Store at −20 °C.
5. Phosphatase inhibitor cocktail 2 (Sigma-Aldrich). Store at 4 °C.

6. Phosphatase inhibitor cocktail 3 (Sigma-Aldrich). Store at 4 °C.
7. 0.5 mm acid-washed glass beads (Biospec Products, Bartlesville, OK). Autoclave for 30 min.
8. Mini-BeadBeater (Biospec Products).
9. *M. oryzae* wild-type strain Guy11 [16], *pmk1* mutant nn78 [7], *mps1* mutant M3H51 [8], and *osm1* mutant JH73 [9].

2.2 Polyacrylamide Gel Electrophoresis (SDS-PAGE) and Blotting

1. 30% acrylamide (acrylamide–bis-acrylamide, 37.5:1) stock solution (Bio-Rad Laboratories, Hercules, CA).
2. 10% (w/v) ammonium persulfate (APS) in DDW.
3. Resolving gel buffer: 1.5 M Tris–HCl (pH 8.8), 0.4% (w/v) SDS.
4. Resolving gel mixture: 4.2 mL 30% acrylamide stock, 2.5 mL resolving gel buffer, 3.3 mL DDW, 45 μL 10% APS, and 10 μL N,N,N,N′-tetramethyl-ethylenediamine (TEMED).
5. Stacking gel buffer: 0.5 M Tris–HCl (pH 6.8), 0.4% (w/v) SDS.
6. Stacking gel mixture: 0.75 mL 30% acrylamide stock, 1.88 mL stacking buffer, 4.75 mL DDW, 40 μL 10% APS, and 15 μL TEMED.
7. Bio-Rad Protein Assay Kit (Bio-Rad Laboratories).
8. Bovine serum albumin (BSA).
9. SDS-PAGE running buffer: 25 mM Tris–HCl (pH 8.3), 200 mM glycine, 0.1% (w/v) SDS.
10. 2× Loading buffer: 100 mM Tris–HCl (pH 6.8), 4% (w/v) SDS, 20% (v/v) glycerol, 200 mM β-mercaptoethanol, 0.2% (w/v) bromophenol blue.
11. Transfer buffer: 25 mM Tris–HCl (pH 8.5), 0.2 M glycine, 20% (v/v) methanol.
12. 10× Tris-buffered saline (10 × TBS): 0.5 M Tris–HCl (pH 7.4), 1.5 M NaCl.
13. TBST: 1× TBS with 0.05% Tween 20.
14. Broad range (10–250 kDa) prestained dual color protein marker.
15. Dual gel caster (10 × 10.5 cm) (GE Healthcare Life Sciences, Pittsburgh, PA) and SE 250 Mini-vertical Unit (GE Healthcare Life Sciences).
16. Vertical electrophoresis chamber (Labnet International, Edison, NJ).
17. Nitrocellulose membranes, 0.45 μm pore size (Bio-Rad Laboratories).

2.3 Detection of MAPK Expression and Phosphorylation

1. Blocking solution: 1× TBST with 5% nonfat dry milk. Store at 4 °C (*see* **Note 1**).
2. Antibody dilution buffer: 1× TBS with 5% nonfat dry milk.
3. Anti-rabbit IgG conjugated with horseradish peroxidase (HRP) (Sigma-Aldrich).
4. Phospho-p44/42 MAPK (anti-TpEY) antibody (Cell Signaling Technology, Danvers, MA).
5. Anti-MAPK ERK-2 antibody (Santa Cruz Biotechnology, Dallas, TX).
6. Phospho-p38 MAPK (anti-TpGY) antibody (Cell Signaling Technology).
7. Anti-Hog1 antibody (Santa Cruz Biotechnology).
8. Anti-Actin antibody (Sigma-Aldrich).
9. SuperSignal™ West Pico PLUS Chemiluminescent Substrate Kit (Thermo Fisher Scientific, Rockford, IL).
10. ECOMAX™ X-Ray Film Processor (PROTEC, Germany).

3 Methods

3.1 Isolation of Total Proteins from Fungal Hyphae

1. Harvest hyphae from 2-day-old cultures of wild-type strain Guy11, the *pmk1* deletion mutant (nn78) [7], the *mps1* deletion mutant (M3H51) [8], and the *osm1* deletion mutant (JH73) [9] that are grown in 5 × YEG [16] (*see* **Note 2**) by filtering through one layer of Miracloth. Rinse the hyphae with sterile distilled water at room temperature (*see* **Note 3**). Remove excessive water by pressing hyphal pads between layers of paper towels.
2. Transfer 300 mg of the semi-dry hyphae in a 2 mL screw cap microcentrifuge tube. Add 0.35 g of sterile, acid-washed glass beads, 1 mL protein lysis buffer (room temperature) containing 10 μL each of protease inhibitor cocktail, phosphatase inhibitor cocktail 2 and phosphatase inhibitor cocktail 3 (*see* **Note 4**). Fill the tube with additional lysis buffer (approximately 0.3 mL).
3. Grind hyphae with the Mini-BeadBeater at full speed for 40s then keep on ice for 2 min. Repeat the grinding five times (*see* **Note 5**).
4. Centrifuge the lysate at 25,000 × *g* for 30 min at 4 °C.
5. Transfer the supernatant to a prechilled microcentrifuge tube.
6. Mix 100 μL of the supernatant (i.e.. protein extract) with 100 μL 2× loading buffer and denature in boiling water for 5 min. Centrifuge at 25,000 × *g* for 1 min and keep the samples on ice (*see* **Note 6**).

7. Prepare a standard curve using BSA for assessing protein concentrations and determine the total protein concentrations with the Bio-Rad Protein Assay kit as described in [17]. Adjust the samples using 1× loading buffer so that all of them have the same protein concentration (*see* **Note 7**).

3.2 SDS-PAGE and Blotting

1. Prepare polyacrylamide gels (0.75 mm in thickness) with 6 cm resolving gel and 1 cm stacking gel (in height) as described in [18].
2. Load 10 μL boiled protein sample in each well (*see* **Note 8**). Run gel electrophoresis at 150 V for 1 h.
3. Prepare nitrocellulose membranes and Whatman filter papers with the same dimensions of the SDS-PAGE gels. Lay one piece of nitrocellulose membrane over the SDS-PAGE gel and remove air bubbles by rolling with a glass rod. Lay two pieces of Whatman filter paper over the nitrocellulose membrane and remove air bubbles by rolling with a glass rode.
4. Place the gel–nitrocellulose membrane package within the transfer cassette and transfer at 100 mA for 4 h or 50 mA overnight with stirring (*see* **Note 9**).
5. Transfer the nitrocellulose membrane to 20 mL blocking solution and incubate for 1 h at room temperature with gentle shaking.
6. Wash the membrane using 20 mL TBST for 5 min with gentle shaking.
7. Repeat **steps 1–9** using the same loading, gel electrophoresis, and blotting conditions to obtain three membranes for western blots: one for detecting MAPK protein levels; one for detecting MAPK phosphorylation levels and the third one for detecting actin as the loading control.

3.3 Detection of Pmk1 and Mps1 Protein and Phosphorylation Levels

1. To determine the phosphorylation status of Pmk1 and Mps1, add the anti-TpEY phosphorylation-specific antibody as the primary antibody at 1:1000 dilution in antibody dilution buffer and incubate overnight at 4 °C with gentle shaking (*see* **Note 10**).
2. Wash the membrane with 20 mL TBST for 5 min at room temperature. Repeat three times.
3. Add the secondary antibody (Anti-rabbit IgG conjugated with HRP) at 1:10,000 dilution in the antibody dilution buffer and incubate for 1 h at room temperature with gentle shaking (*see* **Note 11**).
4. Wash the membrane with 20 mL TBST for 5 min at room temperature. Repeat three times.
5. Prepare working solution of SuperSignal Chemiluminescent substrate by mixing an equal volume of the stable peroxide

solution and the luminol/enhancer solution provided in the kit. Incubate the membrane in the working solution at room temperature for 5 min and wrap the membrane with Saran wrap (keep the membrane wet).

6. Expose to X-ray film for 1 to 3 min (*see* **Note 12**). Develop the film with an X-Ray Film Processor. Signals from the 42 kDa and 46 kDa bands represent PmK1 and Mps1 proteins, respectively; intensity of the signals indicates their phosphorylation levels (Fig. 1).
7. Relative phosphorylation levels should be determined according to the protein levels. To assay the protein levels of Pmk1 and Mps1, use the anti-MAPK ERK-2 antibody as the primary antibody (*see* **Note 13**) at 1:500 dilution to a second blot and incubate the membrane overnight at 4 °C with gentle shaking. Repeat **steps 2–6** in this section. The intensities of the 42 kDa and 46 kDa bands represents the protein levels of Pmk1 and Mps1, respectively (Fig. 1).
8. In each sample, actin proteins are used as the loading control since its levels should remain constant. The actin proteins are detected using the anti-Actin antibody as the primary antibody at 1:1000 dilution to the third blot and incubate overnight at 4 °C with gentle shaking. Repeat **steps 2–6** to detect the actin band, which should be approximately 45 kDa in size (Fig. 1).

3.4 Detection of Osm1 Expression and Phosphorylation with an Anti-TpGY Antibody

1. To determine the phosphorylation status of Osm1, add the anti-TpGY (Phospho-p38 MAPK) phosphorylation-specific antibody at 1:1000 dilution in antibody dilution buffer and incubate the membrane overnight at 4 °C with gentle shaking.
2. Wash the membrane with 20 mL TBST for 5 min at room temperature. Repeat three times.

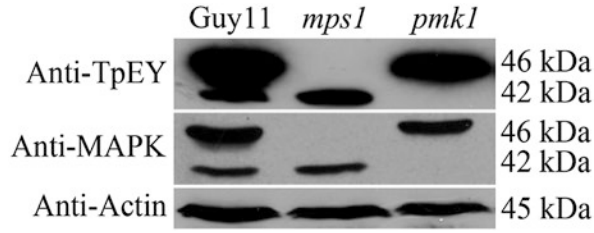


Fig. 1 Detecting the activation of Pmk1 and Mps1 MAP kinases in the rice blast pathogen *M. oryzae*. Western blots of total proteins extracted from *M. oryzae* strains including wild-type Guy11 [16], *pmk1* deletion mutant [7] and *mps1* deletion mutant [8] using anti-TpEY (top panel), anti-MAPK (middle panel), or anti-Actin antibody (lower panel). The molecular weights of Pmk1, Mps1, and actin are marked on the right

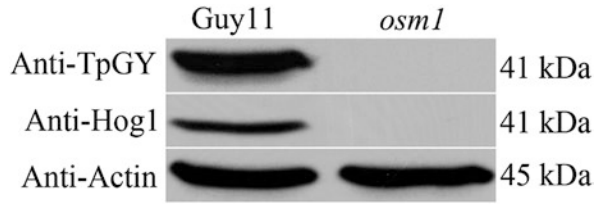


Fig. 2 Detecting the activation of Osm1 MAP kinase in *M. oryzae*. Western blots of total proteins extracted from *M. oryzae* strains including wild-type Guy11 and the *osm1* deletion mutant [9] using an anti-TpGY (top panel), anti-Hog1 (middle panel), or anti-Actin antibody (lower panel). The molecular weights of Osm1 and actin are marked on the right

3. Add secondary antibody (anti-rabbit IgG conjugated with HRP) at 1:10,000 dilution in diluent solution and incubate the membrane for 1 h at room temperature with gentle shaking.
4. Wash the membrane with 20 mL TBST for 5 min at room temperature. Repeat three times.
5. Incubate the membrane in the working solution of SuperSignal Chemiluminescent substrate for signal detection using X-ray film (*see* Subheading 3.3).
6. Develop the film using an X-Ray Film Processor. The 41 kDa band indicates the phosphorylated Osm1 (with the TGY motif) (Fig. 2).
7. To determine the protein levels of Osm1, add the anti-Hog1 antibody (*see* **Note 14**) at 1:500 dilution in antibody dilution buffer to a blot and incubate the membrane overnight at 4 °C with gentle shaking. Repeat **steps 2–6**. The intensity of signals at approximately 41 kDa represents Osm1protein levels (Fig. 2).
8. To detect actin proteins as a loading control, add the anti-Actin antibody at 1:1000 dilution in antibody dilution buffer to the third blot and incubate the membrane overnight at 4 °C with gentle shaking. Repeat **steps 2–6** to detect the actin band at around 45 kDa (Fig. 2).

4 Notes

1. 5% (w/v) bovine serum albumin (BSA) can be used as an alternative blocking reagent instead of milk.
2. It is important to use the same medium and culture conditions because they may affect the expression level and the activation of MAPKs in *M. oryzae*.
3. To avoid the activation of Mps1 and other MAPKs in response to temperature shifts, fungal hyphae should be collected and

washed at room temperature and used for protein extraction immediately.

4. Phosphatase inhibitors should be added to lysis buffer immediately before use. To reduce or avoid proteolysis and dephosphorylation, the samples should be kept on ice as soon as lysis buffer is added.
5. Continuous grinding will result in over-heating and should be avoided so that the samples will remain cold. If necessary, the interval on ice can be increased to 5 min between each grinding.
6. Centrifuge the samples to remove insoluble tissue debris before running the SDS-PAGE.
7. An alternative method to confirm equal loading of the samples is to stain the gels using 0.25% (w/v) Coomassie Brilliant Blue R-250 [19].
8. It is important to load an identical amount of proteins in each lane on all three gels for comparison.
9. Wet nitrocellulose membranes in the transfer buffer for a few minutes before making the sandwich package will facilitate the protein transfer.
10. If necessary, a higher concentration (500-fold dilution) of the primary antibody may be used to increase the signal intensity.
11. If the signals are weak, the concentration of secondary antibody can be increased to 2000-fold dilution. However, the background noise will also increase when using a higher concentration of the secondary antibody.
12. Adjust the exposure time based on signal strength to obtain optimal images. If necessary, the exposure time can be as long as 4 h.
13. The anti-P42/44 antibody (Cell signaling Technology) can also bind the Pmk1 and Mps1 proteins produced by *M. oryzae*.
14. The anti-P38 antibody (Cell signaling Technology) can also bind Osm1 produced by *M. oryzae*.

Acknowledgments

This work was supported by a grant from US Wheat and Barley Scab Initiative and a grant from the National Research Initiative of the United States Department of Agriculture National Institute for Food and Agriculture (award number 2013-68004-20378).

References

1. Zhao X, Mehrabi R, Xu JR (2007) Mitogen-activated protein kinase pathways and fungal pathogenesis. Eukaryot Cell 6(10):1701–1714
2. Hamel LP, Nicole MC, Duplessis S et al (2012) Mitogen-activated protein kinase signaling in plant-interacting fungi: distinct messages from conserved messengers. Plant Cell 24(4):1327–1351
3. Kondoh K, Nishida E (2007) Regulation of MAP kinases by MAP kinase phosphatases. Biochim Biophys Acta 1773(8):1227–1237
4. Pearson G, Robinson F, Gibson T et al (2001) Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions. Endocr Rev 22(2):153–183
5. Dean RA, Talbot NJ, Ebbole DJ et al (2005) The genome sequence of the rice blast fungus Magnaporthe grisea. Nature 434(7036):980–986
6. Zhao X, Kim Y, Park G et al (2005) A mitogen-activated protein kinase cascade regulating infection-related morphogenesis in Magnaporthe grisea. Plant Cell 17(4):1317–1329
7. Xu JR, Hamer JE (1996) MAP kinase and cAMP signaling regulate infection structure formation and pathogenic growth in the rice blast fungus Magnaporthe grisea. Genes Dev 10(21):2696–2706
8. Xu JR, Staiger CJ, Hamer JE (1998) Inactivation of the mitogen-activated protein kinase Mps1 from the rice blast fungus prevents penetration of host cells but allows activation of plant defense responses. Proc Natl Acad Sci U S A 95(21):12713–12718
9. Dixon KP, Xu JR, Smirnoff N et al (1999) Independent signaling pathways regulate cellular turgor during hyperosmotic stress and appressorium-mediated plant infection by Magnaporthe grisea. Plant Cell 11(10): 2045–2058
10. Li G, Zhou X, Xu JR (2012) Genetic control of infection-related development in Magnaporthe oryzae. Curr Opin Microbiol 15(6):678–684
11. Liu W, Zhou X, Li G et al (2011) Multiple plant surface signals are sensed by different mechanisms in the rice blast fungus for appressorium formation. PLoS Pathog 7(1):e1001261
12. Zhang X, Liu W, Li Y et al (2017) Expression of HopAI interferes with MAP kinase signalling in Magnaporthe oryzae. Environ Microbiol 19(10):4190–4204
13. Li G, Zhang X, Tian H et al (2017) MST50 is involved in multiple MAP kinase signaling pathways in Magnaporthe oryzae. Environ Microbiol 19(5):1959–1974
14. Zheng D, Zhang S, Zhou X et al (2012) The FgHOG1 pathway regulates hyphal growth, stress responses, and plant infection in Fusarium graminearum. PLoS One 7(11):e49495
15. Ramamoorthy V, Zhao X, Snyder AK et al (2007) Two mitogen-activated protein kinase signalling cascades mediate basal resistance to antifungal plant defensins in Fusarium graminearum. Cell Microbiol 9(6):1491–1506
16. Leung H, Borromeo ES, Bernardo MA et al (1988) Genetic analysis of virulence in the rice blast fungus Magnaporthe grisea. Phytopathology 78(9):1227–1233
17. Jones CG, Hare DJ, Compton SJ (1989) Measuring plant protein with the Bradford assay: 1. Evaluation and standard method. J Chem Ecol 15(3):979–992
18. Gallagher SR (2012) One-dimensional SDS gel electrophoresis of proteins. Curr Protoc Protein Sci Chapter 10:Unit 10.1.1-44
19. Gilda JE, Gomes AV (2015) Western blotting using in-gel protein labeling as a normalization control: stain-free technology. Proteomic Profiling: Methods and Protocols 1295: 381–391



Chapter 9

Visualizing the Movement of *Magnaporthe oryzae* Effector Proteins in Rice Cells During Infection

Kiersun Jones and Chang Hyun Khang

Abstract

During biotrophy, filamentous pathogens such as the rice blast fungus *Magnaporthe oryzae* deliver effector proteins into live host cells to facilitate colonization. We describe three complementary assays for visualizing *M. oryzae* effector translocation into the rice cytoplasm and cell-to-cell movement during infection. Our assays make use of live-cell confocal imaging of optically clear rice sheath cells infected with transgenic strains of *M. oryzae* that express the fluorescent protein-tagged effector known as PWL2. We highlight several important considerations for the analysis of effector translocation and movement dynamics during infection of host plants.

Key words Biotrophic interfacial complex, Confocal imaging, Effector proteins, Fluorescent proteins, Host translocation assay, Live-cell imaging, *Magnaporthe oryzae*, *Oryza sativa*, Plant–pathogen interactions, Rice blast

1 Introduction

During biotrophic invasion of plant cells, filamentous pathogens such as fungi and oomycetes secrete effector proteins to suppress host immunity and facilitate colonization. Effectors are increasingly shown to be translocated into host cells across the host-derived membrane encasing effector-secreting cells such as haustoria or invasive hyphae and then move cell-to-cell in the host [1–3]. The mechanisms of effector translocation and cell-to-cell movement remain largely unknown. The ability to directly visualize effector movement and subcellular localizations can facilitate the investigation of such mechanisms and effector function.

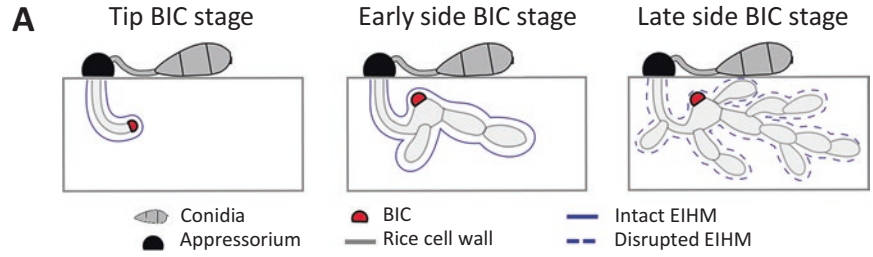
Magnaporthe oryzae is a hemibiotrophic filamentous fungus that causes the economically important blast disease on many crops such as rice and wheat. Using live-cell imaging of fluorescent protein (FP) reporters, subcellular localizations of several *M. oryzae* effectors with an N-terminal signal peptide have been

Wenbo Ma and Thomas Wolpert (eds.), *Plant Pathogenic Fungi and Oomycetes: Methods and Protocols*, Methods in Molecular Biology, vol. 1848, https://doi.org/10.1007/978-1-4939-8724-5_9,
© Springer Science+Business Media, LLC, part of Springer Nature 2018

visualized during fungal invasion of live rice cells [2, 4–8]. Apoplastic effectors such as BAS4 are localized in the *M. oryzae*-rice interfacial space enclosed by the tight-fitting plant-derived extra-invasive hyphal membrane (EIHM), whereas cytoplasmic effectors such as PWL2 enter the rice cytoplasm and move into adjacent cells. Cytoplasmic effectors preferentially accumulate at the biotrophic interfacial complex (BIC), a localized structure that has been hypothesized to be the site of effector translocation into host cells across the intact EIHM (Fig. 1a) [2, 4].

Here, we describe three complementary assays for visualizing *M. oryzae* effector movement during infection and discuss their advantages and potential limitations (Fig. 1b). All three assays make use of live-cell confocal imaging of optically clear rice sheath cells infected with transgenic strains of *M. oryzae* that express FP-tagged PWL2 proteins. In the first assay, mCherry-tagged PWL2 (PWL2:mCherry) is directly visualized in the host cytoplasm. In the second assay, the translocated PWL2:mCherry is more easily observed in the host cytoplasm after sucrose-induced plasmolysis. In the third assay, detection of translocated PWL2:mCherry is further improved by including a nuclear localization signal (NLS) to concentrate the proteins in the rice nucleus. Cell-to-cell movement of PWL2 can be detected in all three assays, particularly when NLS is included in the third assay.

There are a few points that are important to consider when studying translocation of FP-tagged effectors into host cells using live cell imaging. First, effector gene expression is coordinated with infection stages [1, 9], thus FP-tagged effectors should be expressed under the control of their native promoters. Second, to control for possible artifacts introduced by tagging, it is important to validate that FP-tagged effectors retain their biological function as shown for PWL2 and AVR-Pita1 [2]. For effectors whose functions are unknown, which is the case for most effectors, the localization of FP-tagged effectors must be interpreted with caution. Third, the EIHM loses integrity during infections, which results in leaking of EIHM matrix proteins into the host cytoplasm through the disrupted EIHM [2, 6, 10, 11]. Therefore, individual infection sites must be assessed for EIHM integrity when performing effector translocation studies. In our translocation assays, we introduced PWL2:mCherry or PWL2:mCherry:NLS constructs into an *M. oryzae* strain expressing the BAS4 signal peptide:EGFP fusion, which is secreted into the EIHM matrix, to monitor EIHM integrity. Exclusive localization of EGFP fluorescence around invasive hyphae indicates that the EIHM remains intact, whereas appearance of EGFP fluorescence in the host cytoplasm indicates that the EIHM has lost integrity [10].



B

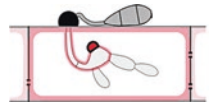
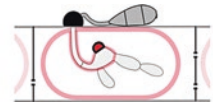
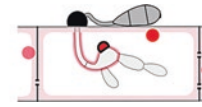
	Assay 1 	Assay 2 	Assay 3 
PWL2-associated mCherry localization	• Strongly at the BIC, weakly around the BIC-associated cell and absent from other invasive hyphae • In the cytoplasm of invaded and adjacent cells	• Strongly at the BIC, weakly around the BIC-associated cell and absent from other invasive hyphae • In the plasmolyzed cytoplasm of invaded and adjacent cells	• Strongly at the BIC, weakly around the BIC-associated cell and absent from other invasive hyphae • In nuclei of invaded and adjacent cells
Advantages	• Subcellular localization likely governed by the effector protein itself	• Easy to detect translocated effectors with host cytoplasm separated from cell walls • Plasmolyzed cells indicate live cells	• High sensitivity to detect effector translocation and cell-to-cell movement
Potential Limitations	• Not sensitive enough to detect effector translocation and cell-to-cell movement at low concentration • Possible confusion with cell wall autofluorescence	• Plasmolysis can cause structural alterations that may affect effector localization	• Endogenous effector localization not represented due to NLS-mediated nuclear targeting

Fig. 1 *M. oryzae* infection process and features of effector movement assays. (**a**) Schematic diagram showing hyphal growth, BIC development, and EIHM in the first-invaded rice cell. Tip BIC stage: A filamentous primary hypha with a tip BIC extends from the appressorium. The hypha is surrounded by an intact EIHM. Early side BIC stage: A few bulbous hyphal branches extend from the first bulbous invasive hyphal cell with a side BIC [2, 17]. Late side BIC stage: Invasive hyphae have branched to fill much of the space in the rice cell. The EIHM has disrupted, and the host cell may be dying or dead. (**b**) A table summarizing PWL2-associated mCherry localization as well as the advantages and potential limitations of each of the three effector movement assays

2 Materials

2.1 Cloning and Transformation

1. *Escherichia coli* competent cells and standard supplies for gene cloning.
2. *Agrobacterium tumefaciens* EHA105 competent cells.
3. *M. oryzae* strain CKF3816 (G418 resistance) expressing BAS4 signal peptide:EGFP fusion (secreted EGFP). CKF3816 is a transformant of *M. oryzae* wild-type strain O-137 [12].
4. Incubator(s) set at 24 °C and 37 °C.
5. Antibiotics (kanamycin for *E. coli* and *A. tumefaciens*; hygromycin for *M. oryzae*).
6. 24-well tissue culture plate containing V8 agar medium.

2.2 Plant and Fungal Materials and Inoculation

1. Rice (*Oryza sativa*) cultivar YT16: Grown in soil-filled pots in a growth chamber at 80% relative humidity with a 12/12 light/dark cycle at 28 °C during the day and 24 °C at night. Apply fertilizer (20-10-20 peatlite) once a week.
2. *M. oryzae* strain CKF3816: Grown on oatmeal agar (OMA [12]) in an incubator under constant light at 24 °C.
3. Adjustable pipettes and pipette tips (ranging 2–1000 μL).
4. Small scissors.
5. Miracloth.
6. Forceps.
7. Sterile distilled water.
8. 1.5 mL microcentrifuge tubes.
9. Hemacytometer.
10. Light microscope.
11. Humidity container (*see* **Note 1**).

2.3 Confocal Microscopy

1. Glass slides.
2. Coverslips 22 × 50 mm.
3. Forceps.
4. Razor blades (single edge).
5. Sterile distilled water.
6. 0.5 M sucrose.
7. Transfer pipettes.
8. Confocal microscope system. We use a Zeiss LSM 710 laser scanning confocal microscope and Zen software (Black edition) for image acquisition and analysis. However, the assay can be performed with most other confocal systems and epifluorescence microscopes (*see* **Note 2**).

3 Methods

3.1 Generate M. oryzae Transgenic Strains Expressing PWL2:mCherry or PWL2:mCherry:NLS

1. Generate PWL2:FP expression constructs by cloning the *PWL2* promoter and its entire 145-amino acid coding sequence fused to mCherry (*PWL2*: mCherry) or nuclear targeting mCherry (*PWL2*: mCherry:NLS) into the binary vector pBHt2 [13] (*see* **Note 3**).
2. Transform *M. oryzae* strain CKF3816 (G418 resistance) with each construct (hygromycin resistance) using *Agrobacterium*-mediated transformation [14]. CKF3816 expresses the BAS4 signal peptide:EGFP fusion (secreted EGFP) as a control for EIHM integrity as mentioned above.
3. Pick 10–12 transformants and grow them in a 24-well tissue culture plate with V8 agar.
4. Screen transformants after inoculating them onto rice sheaths (*see* Subheading 3.2 for inoculation and microscopy protocols)—check for morphology that is consistent with wildtype; typically growth on agar medium or in planta will not differ from wild-type. Confirm that all transformants for each construct show similar fluorescence patterns and then choose those with the brightest fluorescence to make filter disc stocks from monoconidia [12] and to use for further imaging.

3.2 Effector Localization Assay 1: Direct Visualization of Translocated PWL2:mCherry

3.2.1 Sheath Inoculation

1. Prepare rice sheaths (adapted from [15]): Choose 16–19 day old rice plants with their third leaf emerged about 4–10 cm. Excise the central sheath by cutting it just above where the tillers branch. Remove the outer and inner sheath layers carefully without bending the middle sheath layer. Cut the upper end of the middle sheath layer to produce a sheath segment ~8 cm in length. Make the cut at a 45° angle so that the midvein forms a point. Place the excised sheaths in a humidity container (*see* **Note 1**).
2. Prepare the inoculum: Harvest conidia from a one to 2 week old OMA culture of transgenic *M. oryzae* coexpressing PWL2:mCherry and secreted EGFP. Adjust the inoculum concentration to 5×10^4 spores/mL in water (*see* **Note 4**).
3. Inoculate excised sheaths (adapted from [15]): Homogenize the conidial suspension by vortexing and then pipette the suspension into the hollow space of an excised sheath from the opening at the end with the 45° cut. Completely fill each of the excised sheaths with inoculum from end-to-end. Each sheath typically requires about 200 μL of conidial suspension.
4. Incubate the inoculated sheaths in a humidity container in dark condition at 25 °C (*see* **Note 5**).

3.2.2 Live-Cell Confocal Imaging

1. Trim the inoculated rice sheath with a razor blade (*see* **Note 6**) at 26–28 h post inoculation (hpi) (*see* **Note 7**).
2. Immediately mount the trimmed sheath on a glass slide using sterile water. Ensure that the sheath is oriented correctly on the slide with the epidermal cells facing up (*see* **Note 8**). Use care when placing a coverslip over the trimmed sheath as not to introduce air bubbles or damage the sheath tissue (*see* **Note 9**).
3. Observe the trimmed sheath through the eyepiece using a 20× objective lens in bright-field for initial field selection (*see* **Note 10**). Dark pigmented appressoria can be easily recognized at this magnification. Be aware of the different types of epidermal cells (*see* **Note 11**).
4. Switch to a 40× objective lens and locate an infection site that is at an early growth stage with only the filamentous primary hypha (tip BIC stage) or up to a few bulbous IH branches (side BIC stage) (Fig. 1a).
5. After locating and focusing on the selected infection site in the bright-field, switch to image acquisition and use the appropriate laser lines for excitation along with the applicable emission spectra (e.g., 488 nm Argon laser line and 495–530 nm emission collection for EGFP and 543 nm Neon laser line and 560–615 nm emission collection for mCherry). Visualize the infection site using live scanning or an equivalent feature capable of acquiring fluorescence and bright-field channels at a fast scan speed to determine if it is an informative infection site (*see* **Note 12**).
6. After locating an informative infection site, set the pinhole size to 1 airy unit (AU). Adjust the pinhole to a higher AU if the fluorescence signal is too weak (*see* **Note 13**).
7. Set the zoom factor to a value of 1 for the 40× objective lens to capture the entire infected cell (*see* **Note 14**). To get optimal signal intensity and minimize background noise, adjust image acquisition settings (*see* **Note 15**). Choose a frame size (*see* **Note 16**).
8. Set the z-slice interval (*see* **Note 17**) and acquire a z-stack image of the informative infection site (*see* **Notes 12** and **18**).
9. Repeat **steps 3–8** to collect images of multiple informative infection sites in order to confirm the consistency of the protein localization patterns observed (*see* **Note 19**). When acquiring many images from the same trimmed sheath, check the slide about every 30 min to make sure the sheath does not dry out and add more water when necessary.
10. Use image analysis software such as Zen (Black edition) from Zeiss microscopy to confirm the EIHM integrity based on secreted EGFP distribution in the acquired images. Ensure

that fluorescence is associated with invasive hyphae and not located in the rice cytoplasm (*see* **Note 20**).

11. Use Zen to determine the PWL2:mCherry fluorescence localization in the acquired images. PWL2:mCherry fluorescence is strong at the BIC and relatively less intense in the cytoplasm of the infected cell (Fig. 2a) (*see* **Note 21**). Observing PWL2:mCherry fluorescence in the BIC and the infected cell's cytoplasm indicates effector secretion and translocation, respectively (*see* **Note 22**).

3.3 Effector Localization Assay 2: Visualizing Translocated PWL2:mCherry After Sucrose-Induced Plasmolysis

1. Prepare sheaths and inoculate them with the *M. oryzae* strain coexpressing PWL2:mCherry and secreted EGFP as described in Subheading 3.2. Follow all steps in Subheading 3.2.1.
2. Trim the inoculated sheaths by following **step 1** in Subheading 3.2.2.
3. Use 0.5 M sucrose as the slide mounting solution (*see* **Note 8**). Incubate the slide for 15 min to allow rice cell plasmolysis to occur (*see* **Note 23**).
4. Find suitable informative infection sites and optimize imaging settings by following **steps 3**–**7** in Subheading 3.2.2 (*see* **Note 24**).
5. Acquire z-stack images of informative infections by following **steps 8** and **9** in Subheading 3.2.2.
6. Follow **step 10** in Subheading 3.2.2 to confirm EIHM integrity by checking the localization of secreted EGFP.
7. To determine the PWL2:mCherry fluorescence localization in the acquired images, follow **step 11** in Subheading 3.2.2. The cytoplasm should be pulled away from the cell wall in plasmolyzed cells (Fig. 2b) (*see* **Note 25**).

3.4 Effector Localization Assay 3: Visualizing Host Nuclear-Targeted PWL2:mCherry:NLS

1. Prepare sheaths and inoculate them with *M. oryzae* coexpressing PWL2:mCherry:NLS and secreted EGFP by following all steps in Subheading 3.2.1.
2. Trim and mount the inoculated sheaths by following **steps 1** and **2** in Subheading 3.2.2.
3. Find infection sites and optimize imaging settings by following **steps 3**–**7** in Subheading 3.2.2 (*see* **Note 26**).
4. Acquire z-stack images of informative infections by following **steps 8** and **9** in Subheading 3.2.2.
5. To confirm that secreted EGFP fluorescence is associated with invasive hyphae and is not located in the rice cytoplasm in the acquired images, follow **step 10** in Subheading 3.2.2.
6. To determine the PWL2:mCherry:NLS fluorescence localization in the acquired images, follow **step 11** in Subheading

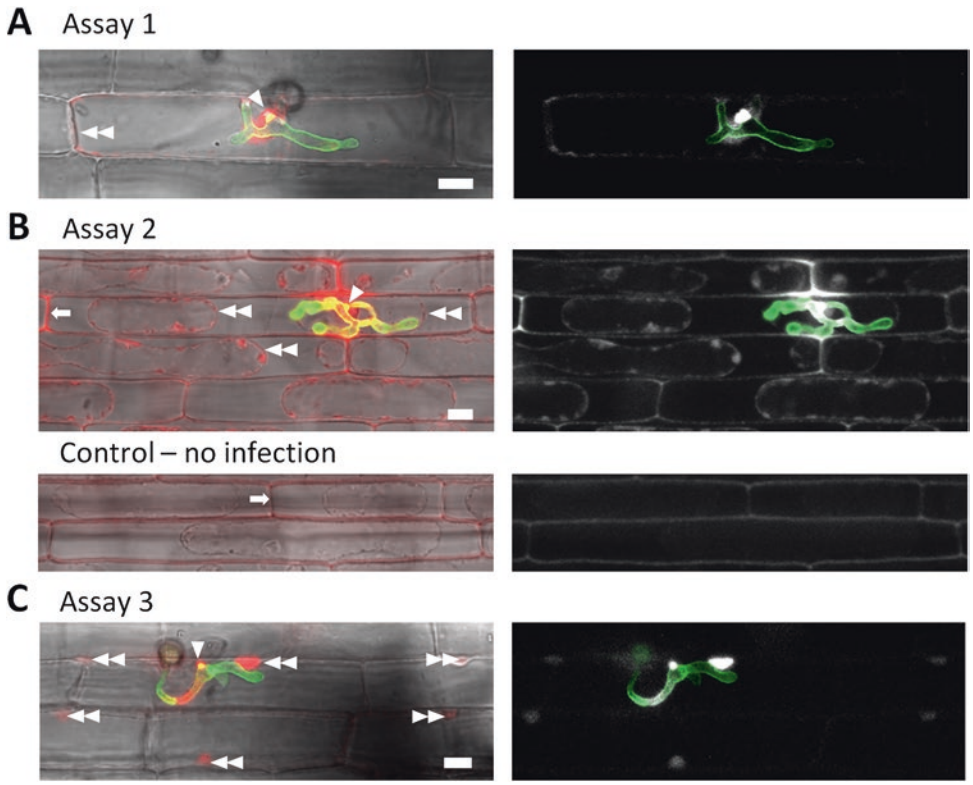


Fig. 2 Fluorescence-based assay to examine effector movement in plant cells. Three different assays for PWL2 movement during *M. oryzae* infection in rice. Shown are confocal images with merged fluorescence and bright-field (left) and merged fluorescence alone (right; red fluorescence pseudo-colored in white). Secreted EGFP (green) exclusively surrounds invasive hyphae in all images, indicating intact EIHM. Bars = 10 μm. (**a**) Direct visualization of PWL2:mCherry (red) preferentially accumulated at the BIC (arrowhead) and translocated into the rice cytoplasm (double arrowhead). (**b**) Plasmolysis-assisted detection of translocated PWL2:mCherry (red) accumulated at the BIC (arrowhead) and translocated into the rice cytoplasm of the infected cell and moved into surrounding cells (examples denoted by double arrowheads). The protoplast in this particular infected cell split into two upon plasmolysis. Note that some autofluorescence is associated with rice cell walls (example denoted by an arrow). In the control with no fungal infection shown below, plasmolyzed cells have the same cell wall autofluorescence (example denoted by an arrow) without cytoplasmic fluorescence. (**c**) PWL2:mCherry:NLS (red) accumulated at the BIC (arrowhead) as well as in the nuclei of the invaded cell and the surrounding cells (double arrowheads)

3.2.2. PWL2:mCherry:NLS fluorescence appears in BICs and also in rice nuclei for both the infected and adjacent uninfected rice cells (Fig. 2c).

4 Notes

1. Use any container that can maintain humidity and hold the excised sheaths upright. We found a convenient way to make a humidity container was to place a 10 cm filter paper at the

bottom of a 10 cm glass petri dish and add just enough water to wet the filter paper. To make a sheath holder, we cut the barrel of a plastic transfer pipette into ~3 cm sections and glued them together side-by-side.

2. Use a confocal microscope that is equipped with laser lines capable of 488 nm and 543 nm such as an Argon/2 laser (458, 488, 514 nm) and a HeNe1 laser (543/561 nm). Use the 488 nm laser for excitation of EGFP and collect emission between 495 and 530 nm (Ex/Em 489/508 nm) and a 543 nm laser for excitation of mCherry and collect emission between 560 and 615 nm (Ex/Em 587/610 nm).
3. We used a conventional cloning method to generate the constructs. Using PCR, we amplified the *PWL2* gene fragment, including the promoter (~1 kb) and the entire coding sequence (stop codon not included), from *M. oryzae* genomic DNA. The recognition sites for *Eco*RI or *Bam*HI were added to the 5′-end of the forward or reverse PCR primers, respectively. The *Bam*HI site serves to ligate *PWL2* and mCherry and also to encode a short linker peptide (Gly-Ser). Next, we cloned the *Eco*RI-*Bam*HI digested PCR product (1.4 kb *PWL2* gene fragment) together with mCherry:terminator (1 kb *Bam*HI-*Hind*III fragment isolated from pBV360; [2]) into *Eco*RI-*Hind*III sites of the binary vector pBHt2 to generate the *PWL2*:mCherry construct. Additionally, we cloned the same *PWL2* gene fragment together with mCherry:NLS:terminator (1.1 kb *Bam*HI-*Hind*III fragment isolated from pBV579; [2]) into *Eco*RI-*Hind*III sites of pBHt2 to generate the *PWL2*:mCherry:NLS construct.
4. To harvest conidia of *M. oryzae*, add 1 mL of sterile distilled water to the surface of the culture. Release conidia into the water by gently dislodging them from the outer edge of the mycelial colony with the tip of a sterile microcentrifuge tube. Filter the conidial suspension by pipetting it through miracloth into a new microcentrifuge tube. Vortex the spore suspension and load ~15 μL into a hemacytometer. Observe under a light microscope to determine the initial spore concentration. Conidial suspensions should be used within an hour after preparation.
5. Ensure that sheaths are positioned upright before leaving them for incubation so that conidia settle onto the epidermal cells along the midvein.
6. The ideal trimmed sheath should be (1) cut from the center of the inoculated sheath, (2) approximately 2 cm in length, (3) about 3–4 cell layers thick, (4) trimmed to a consistent thickness so that the coverslip lies flat over the top of it. To obtain ideal trimmed sheaths, cut each inoculated sheath in half and then excise an approximately 2 cm segment from the inner

portion of one of the pieces. A length of about 2 cm provides enough unperturbed infected cells between the cut ends of the sheath. Hold one end of the sheath gently with your hand and slice through the center of the hollow space from end-to-end with a single-edge razor blade. The remaining portion of the sheath should look U-shaped if viewed on end. Next, trim the sides of the sheath by slicing off the upright portions of the U-shape, leaving the mid vein. Then remove the lower cells of the midvein, leaving the inner epidermal layer and at least one intact mesophyll cell layer. Trimming the sheath too thin can damage the mesophyll layer directly beneath the epidermal layer, thus causing perturbation of the epidermal cells [16]. Trimming the sheath segment too thick reduces light transmission through the sample. Proper sheath trimming takes practice—we recommend practicing with noninoculated sheaths prior to performing the experiment.

7. We found that at 26–28 hpi infections of *M. oryzae* strain O-137 on the susceptible rice cultivar YT16 are typically at an early growth stage in the first-invaded rice cell (Fig. 1a Tip BIC stage and Early side BIC stage). That is, when there is just a filamentous primary hypha or up to a few bulbous IH branches. This is when EIHM is likely to be intact, and the translocation of PWL2:mCherry across the intact EIHM can be examined.
8. Maintain the orientation of the sheath (with the epidermal cells facing up) when mounting it on the slide. When observed on the microscope, the first layer of cells to come into focus should be the rectangular epidermal cells, in which *M. oryzae* initially colonizes. If large square cells (mesophyll cells) appear first, then the sheath was flipped over during mounting. To correct the orientation of the sheath, gently lift the coverslip and use forceps to reposition the sheath. Be sure to mount the trimmed sheath on a slide immediately after preparing it because it dries out quickly after trimmed.
9. Be aware that moving the coverslip can easily damage appressoria and cause the primary hypha to collapse when the infection is at the tip BIC stage or at the early side BIC stages before a septum is formed [17]. There is no need to seal the slide even if using an inverted microscope, provided the sheath was trimmed thin enough.
10. Initial field selection should be done using the bright-field rather than fluorescence to prevent unintentional photobleaching of the fluorescent proteins. We found that with a 20× objective lens (200× magnification), 20 or more rice cell rows are usually visible in the y-axis on the screen. The field of view is slightly larger when physically looking through the eyepiece. This magnification is good for finding infection sites but does

not reveal details of the invasive hyphae. With a 40× objective lens (400× magnification), less than 10 rows of rice cells are typically visible. However, the details of an infection site are much easier to view.

11. Typical trimmed rice sheaths contain more than 30 rows of cells with at least five alternating bands of grouped narrow cells and wide cells, though this can vary greatly based on the age of the plant and the location along the sheath, etc. These two different types of epidermal cells may show different degrees of effector cell-to-cell movement, for example, there is more movement of PWL2:FP in wider and shorter epidermal cells than in longer and thinner cells [2].
12. We found that the growth stage, the state of the EIHM, and the health of rice cells can vary between infection sites even at the same infection time point. These variations affect the interpretation of effector localization. Therefore, care must be taken to choose informative infection sites when assaying for effector translocation into host cells. Use the following criteria to choose informative infection sites: (1) Focus on infection sites at an early growth stage with just a filamentous primary hypha (tip BIC stage) or with up to a few bulbous IH branches (side BIC stage) (Fig. 1a) [10]. (2) Confirm that secreted EGFP fluorescence is localized exclusively around invasive hyphae. This indicates that the EIHM is intact. Do not use infection sites where EGFP fluorescence is dispersed in the host cytoplasm, which indicates that the EIHM has lost integrity [10]. (3) Choose rice cells that do not show excessive discoloration, granulation, or cytoplasmic aggregation. These are features of failed infections or resistance responses. (4) Avoid infection sites where the rice cell is damaged, for example at the cut end of the sheath. Also avoid rice cells with a high level of cell wall autofluorescence.
13. Lower pinhole settings allow for higher resolution in the z-axis and allow visualization of secreted EGFP fluorescence in an outline pattern around invasive hyphae, but this signal may be relatively weak. Higher pinhole settings allow for increased detection of fluorescent signals, although the outlining pattern around invasive hyphae will become indistinct, and secreted EGFP fluorescence will appear to be generally associated with invasive hyphae.
14. It is important to include at least the entire infected cell to determine whether secreted EGFP or PWL2:mCherry has entered the host cytoplasm because the cytoplasm is in close proximity to the cell wall. The 20× objective lens with a zoom factor of 2 or the 40× objective lens with a zoom factor of 1 can both be used to capture the entire infected cell together

with the surrounding uninfected cells, which is useful for determining effector cell-to-cell movement.

15. Optimize the signal-to-noise ratio to detect translocated PWL2:mCherry at low concentrations as well as instances of low intensity secreted EGFP fluorescence in the rice cytoplasm that occurs just after EIHM disruption. Be aware that with settings optimized for detection of low intensity signals, the BIC will likely be saturated because it has the brightest fluorescence intensity. To obtain the optimal settings, adjust the detector gain (maximum of 800), the digital offset, the digital gain, the scan speed, and the laser power. On confocal microscopes equipped with variable spectral filters, maximizing collection emission bandwidth for the fluorophore will also be helpful. Watch for potential autofluorescence associated with rice cell walls that could be mistaken for effector fluorescence.
16. We found that good results are obtained when starting with the 1024 × 1024 frame size. Reduce the scan field in the y-direction if possible because this reduces the image acquisition time. However, be sure to include some uninvaded cells above and below the infected cell to assay for cell-to-cell movement of effectors.
17. We found that a z-slice interval of 2 μm was sufficient for capturing all features of interest, including invasive hyphae, BICs, and rice nuclei. Make sure to set the z-stack parameters to capture the entire depth of the rice cell of about 10–15 μm, thus approximately 5–8 z-slices are needed.
18. Infection sites with a disrupted EIHM showing secreted EGFP spilled into the host cell can also be imaged if effector localization at the post-EIHM disruption stage and/or the state of the vacuole are of interest. The central vacuole has been shown to rupture after EIHM disruption during successful infections at the later stages of growth within the first-invaded cell [10]. This causes any translocated effectors or effectors spilled from the EIHM compartment to homogenize throughout the infected host cell.
19. The number of infection sites that need to be imaged is influenced by the level of variation in the fluorescence patterns. We suggest taking images from at least three different rice sheaths for a total of at least 10 infection sites. Observing comparable PWL2:mCherry fluorescence patterns across all infection sites constitutes strong evidence for effector translocation. The number of infection sites should be increased if major variation is observed across the infection sites. Some minor variations such as different fluorescence intensities or degrees of effector cell-to-cell movement may be observed and worth documenting.

20. To detect low intensity secreted EGFP fluorescence in the rice cytoplasm, lower the set white point in the green channel histogram. This results in an increase in the brightness and contrast, which allows low intensity signals to be more easily differentiated from background noise. Performing the white balance adjustment is important for identifying infections with a recently disrupted EIHM that have not accumulated high concentrations of spilled EGFP in the rice cytoplasm.
21. Some PWL2:mCherry fluorescence may be detected in the EIHM compartment along the primary hypha and the BIC-associated cell. However, PWL2:mCherry fluorescence is not found in the EIHM compartment surrounding other invasive hyphae. If testing a putative mCherry-tagged effector other than PWL2, observing exclusive colocalization of the effector with secreted EGFP in the EIHM compartment indicates that it is an apoplastic effector. However, the effector may still enter the cytoplasm together with secreted EGFP after the EIHM disrupts.
22. The degree of effector cell-to-cell movement can vary depending on the size of the protein and the size exclusion limit of the plasmodesmata [2].
23. We found that adding 0.5 M sucrose to a trimmed rice sheath was sufficient to induce convex plasmolysis, where the protoplast pulls away from the shorter traverse cell walls and forms convex ends [18]. For the purpose of this assay, convex plasmolysis is preferred over other forms of plasmolysis that can occur at higher concentrations of sucrose, such as concave, or sub-protoplasts [18]. It may be necessary to optimize plasmolysis if convex plasmolysis does not occur. This can be done by altering the concentration and/or incubation time to increase or decrease the intensity of plasmolysis. Step-wise plasmolysis, in which sucrose is added gradually at increasing concentrations, can lessen the stress of rapid plasmolysis and allow it to occur more gradually. This can be helpful in some instances when convex plasmolysis is difficult to produce with the addition of a single concentration of sucrose. Step-wise plasmolysis can be done by starting with 0.25 M sucrose and allowing 10 min for incubation before replacing with 0.5 M sucrose for another 10 min.
24. The optimal imaging settings should be the same for assays 1 and 2 because they use the same fungal strain.
25. Plasmolysis causes the plasma membrane and the enclosed protoplast of viable rice cells to pull away from the cell walls. The protoplast always shrinks around invasive hyphae rather than away from them [15], thus improving the distinction between

the cytoplasm and the apoplast as well as any autofluorescence associated with the cell wall.

26. The optimal imaging settings for the strain expressing PWL2:mCherry:NLS will be very similar to the optimal settings determined for the strain expressing PWL2:mCherry. Therefore, the settings for PWL2:mCherry can be used as a convenient starting point.

Acknowledgments

We acknowledge the assistance of the Biomedical Microscopy Core at the University of Georgia with imaging using a Zeiss LSM 710 confocal microscope. This work was supported by the Agriculture and Food Research Initiative competitive grants program, Award number 2014-67013-21717 from the USDA National Institute of Food and Agriculture.

References

1. Giraldo MC, Valent B (2013) Filamentous plant pathogen effectors in action. Nat Rev Micro 11:800–814
2. Khang CH, Berruyer R, Giraldo MC, Kankanala P, Park S-Y, Czymmek K, Kang S, Valent B (2010) Translocation of *Magnaporthe oryzae* effectors into rice cells and their subsequent cell-to-cell movement. Plant Cell 22:1388–1403
3. Presti LL, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S, Kahmann R (2015) Fungal effectors and plant susceptibility. Annu Rev Plant Biol 66:513–545
4. Giraldo MC, Dagdas YF, Gupta YK, Mentlak TA, Yi M, Martinez-Rocha AL, Saitoh H, Terauchi R, Talbot NJ, Valent B (2013) Two distinct secretion systems facilitate tissue invasion by the rice blast fungus *Magnaporthe oryzae*. Nat Commun 4:1996. https://doi.org/10.1038/ncomms2996
5. Mentlak TA, Kombrink A, Shinya T, Ryder LS, Otomo I, Saitoh H, Terauchi R, Nishizawa Y, Shibuya N, Thomma BP, Talbot NJ (2012) Effector-mediated suppression of chitin-triggered immunity by *Magnaporthe oryzae* is necessary for rice blast disease. Plant Cell 24:322–335
6. Mosquera G, Giraldo MC, Khang CH, Coughlan S, Valent B (2009) Interaction transcriptome analysis identifies *Magnaporthe oryzae* BAS1-4 as biotrophy-associated secreted proteins in rice blast disease. Plant Cell 21:1273–1290
7. Park C-H, Chen S, Shirsekar G, Zhou B, Khang CH, Songkumarn P, Afzal AJ, Ning Y, Wang R, Bellizzi M, Valent B, Wang G-L (2012) The *Magnaporthe oryzae* effector AvrPiz-t targets the RING E3 ubiquitin ligase APIP6 to suppress pathogen-associated molecular pattern–triggered immunity in rice. Plant Cell 24:4748–4762
8. Sharpee W, Oh Y, Yi M, Franck W, Eyre A, Okagaki LH, Valent B, Dean RA (2017) Identification and characterization of suppressors of plant cell death (SPD) effectors from *Magnaporthe oryzae*. Mol Plant Pathol 18:850–863
9. Kleemann J, Rincon-Rivera LJ, Takahara H, Neumann U, Ver Loren van Themaat E et al (2012) Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. PLoS Pathog 8:e1002643. https://doi.org/10.1371/journal.ppat.1002643
10. Jones K, Zhu J, Jenkinson CB, Kim DW, Khang CH (2017) Disruption of the interfacial membrane leads to *Magnaporthe oryzae* effector re-location and lifestyle switch during rice blast disease. bioRxiv. https://doi.org/10.1101/177147
11. Mochizuki S, Minami E, Nishizawa Y (2015) Live-cell imaging of rice cytological changes

reveals the importance of host vacuole maintenance for biotrophic invasion by blast fungus, *Magnaporthe oryzae*. MicrobiologyOpen 4:952–966

12. Valent B, Farrall L, Chumley FG (1991) *Magnaporthe grisea* genes for pathogenicity and virulence identified through a series of backcrosses. Genetics 127:87–101
13. Mullins ED, Chen X, Romaine P, Raina R, Geiser DM, Kang S (2001) Agrobacterium-mediated transformation of *Fusarium oxysporum*: an efficient tool for insertional mutagenesis and gene transfer. Phytopathology 91:173–180
14. Khang CH, Park S-Y, Lee Y-H, Kang S (2005) A dual selection based, targeted gene replacement tool for *Magnaporthe grisea* and *Fusarium oxysporum*. Fungal Genet Biol 42:483–492
15. Kankanala P, Czymmek K, Valent B (2007) Roles for rice membrane dynamics and plasmodesmata during biotrophic invasion by the blast fungus. Plant Cell 19:706–724
16. Jones K, Kim DW, Park JS, Khang CH (2016) Live-cell fluorescence imaging to investigate the dynamics of plant cell death during infection by the rice blast fungus *Magnaporthe oryzae*. BMC Plant Biol 16:1–8
17. Shipman EN, Jones K, Jenkinson CB, Kim DW, Zhu J, Khang CH (2017) Nuclear and structural dynamics during the establishment of a specialized effector-secreting cell by Magnaporthe oryzae in living rice cells. BMC Cell Biol 18:11. https://doi.org/10.1186/s12860-017-0126-z
18. Oparka KJ (1994) Plasmolysis: new insights into an old process. New Phytol 126:571–591



Chapter 10

Illuminating *Phytophthora* Biology with Fluorescent Protein Tags

Audrey M. V. Ah-Fong, Meenakshi Kagda, and Howard S. Judelson

Abstract

Phytophthora species cause diseases that threaten agricultural, ornamental, and forest plants worldwide. Explorations of the biology of these pathogens have been aided by the availability of genome sequences, but much work remains to decipher the roles of their proteins. Insight into protein function can be obtained by visualizing them within cells, which has been facilitated by recent improvements in fluorescent protein and microscope technologies. Here, we describe strategies to permit investigators to generate strains of *Phytophthora* that express fluorescently tagged proteins and study their localization during growth in artificial media and during plant infection.

Key words Fluorescent protein, Oomycete, Transformation, Cell biology

1 Introduction

The genus *Phytophthora* includes some of the world's most destructive pathogens of crops and forest trees [1]. The last decade has witnessed a revolution in *Phytophthora* genomics, with genome assemblies and predicted protein sequences available for several species [2]. A major challenge now is ascribing functions to those proteins. Determining the subcellular location of proteins provides clues to their roles and modes of regulation. The advent of genetically encoded fluorescent tags has made it possible to visualize proteins with minimal perturbation to the cell, in contrast to traditional immunolocalization methods which require not only highly specific antibodies but also cell fixation and permeabilization [3]. Efforts to engineer fluorescent proteins have resulted in variants that span the visible spectrum, allowing multiple proteins to be monitored simultaneously [4]. Combined with advances in confocal microscopy, these tags provide powerful tools to study protein interactions, interrogate biological processes, and monitor the abundance, localization, and trafficking of proteins at high spatial

Wenbo Ma and Thomas Wolpert (eds.), *Plant Pathogenic Fungi and Oomycetes: Methods and Protocols*, Methods in Molecular Biology, vol. 1848, https://doi.org/10.1007/978-1-4939-8724-5_10,
© Springer Science+Business Media, LLC, part of Springer Nature 2018

and temporal resolution. In *Phytophthora*, fluorescent tagging has helped visualize organelles [5], track proteins involved in development [6, 7], and study the targeting of proteins involved in plant interactions [8–11].

Although the expression of *Phytophthora* proteins with fluorescent tags (FFP, fluorescent fusion proteins) is now common in several laboratories, it is not without challenges. These include the formation of insoluble aggregates, mislocalization, and instability of transgene expression. Here, we provide protocols for labeling *Phytophthora* proteins that take these issues into consideration. Information is provided about designing constructs, generating transgenic strains, purifying heterokaryons, choosing the appropriate transformants, fixing tissues, and imaging FFPs. Along with continuing advances in fluorescent tags and new genome editing techniques, this will enhance our understanding of *Phytophthora* biology.

2 Materials

2.1 Making Fluorescent Fusion Protein Constructs

1. Transformation vectors containing selectable (drug resistance) markers.
2. Genes encoding fluorescent proteins. GFP, YFP, CFP, mCherry, tdTomato, and mRFP have already been incorporated into vectors adapted to *Phytophthora* [5, 8].

2.2 Electroporation of Zoospores

1. Glass spreader, sterile beaker, 15 μm nylon mesh, hemocytometer.
2. Disposables: sterile 50 and 15 mL polypropylene tubes, 4-mm electroporation cuvettes, 150 mm petri plates, disposable pipettes.
3. Equipment: refrigerated centrifuge with swinging bucket rotor, electroporator, light microscope for counting spores.
4. Regeneration media: For *P. infestans,* this is Rye A (rye-sucrose) broth [12] clarified by centrifugation plus 100 mM mannitol, 1 mM KCl, and 2.5 mM $CaCl_2$. Some species may prefer other media as the base [1].
5. Cold 5 M LiCl and cold modified Petri's solution: 0.25 mM $CaCl_2$, 1 mM $MgSO_4$, 1 mM KH_2PO_4, 0.8 mM KCl.
6. Selective media: 1.5% agar media with antibiotics, usually G418 or hygromycin. For *P. infestans,* this is Rye A media [12].

2.3 Identifying Transformants Expressing FFPs

1. Microscope slides and coverslips.
2. Water, paper towels.
3. Optical fluorescence or confocal microscope.

2.4 Single-Nucleus Purification

1. Glass spreader, 60 mm petri dish, 15 μm nylon mesh, hemocytometer.
2. Sterile water, beaker.
3. Selective media with 1.5% agar.

2.5 Fixing Tissues

1. Single and double-strength (2×) fixative; single-strength is 4% formaldehyde, 50 mM Pipes pH 6.8.
2. Washing solution: 50 mM Pipes pH 6.8.

2.6 Confocal Imaging

1. Glass slides.
2. Coverslips and/or glass-bottom culture dishes designed for confocal microcopy; these are typically described as being made from "precision glass". Brands of coverslips include Marienfeld, Warner and World Precision Instruments (WPI). Glass-bottom dishes are available from MatTek, Nunc, Warner, and WPI.
3. Clear nail polish.

3 Methods

3.1 Design of Fluorescent Fusion Protein (FFP) Construct

1. Select the appropriate vector system. Used most are plasmids bearing *nptII* or *hpt*, which confer resistance to G418 or hygromycin, respectively.
2. Select the desired fluorescent protein. We advise using the brightest proteins in order to increase the signal-to-noise ratio (*see* **Note 1**).
3. Choose between a constitutive, native, or inducible promoter. A strong promoter will maximize expression but may introduce artifacts (*see* **Note 2**).
4. Design a cloning strategy to place the tag at the amino or carboxyl terminus of the protein of interest, or within the protein (*see* **Note 3**). Information about the optimal strategy may be obtained by searching the literature for studies of related proteins. If the protein contains a functional or targeting domain, place the tag at a site that will not interfere with the function of those domains.
5. If the goal is to coexpress two proteins, express both from the same plasmid or introduce them on separate plasmids by cotransformation and/or using different selectable markers (*see* **Note 4**).

3.2 Phytophthora Transformation

Several methods have been developed for transforming *Phytophthora*. Those used most commonly involve treating protoplasts with DNA or electroporating zoospores [5, 13].

Agrobacterium-mediated transformation has also been described [14, 15]. These protocols need to be tailored to the species and sometimes isolate. Presented below is an electroporation method that works well in *P. infestans*.

1. Inoculate five 15-cm rye-sucrose plates by spreading ~1 × 10^4 sporangia on each with a sterile glass rod. Grow for 8–9 days in the dark at 18 °C (*see* **Note 5**).
2. Pour 20 mL of cold Petri's solution on each plate. Use a glass spreader to rub off the sporangia, decanting what should be a milky solution into a 150 mm plate, i.e., one plate per 20 mL of sporangia suspension.
3. Induce zoosporogenesis by incubation at 10 °C for about 2 h (*see* **Note 5**).
4. Harvest the zoospores by pouring the suspension through 15 μm nylon mesh into a 400-mL beaker. This and the following steps should be done on ice. A convenient holder for nylon mesh is sold by Biodesign of New York.
5. Divide the zoospores into two 50 mL tubes, add one-fiftieth the volume of 5 M LiCl to each, and mix by gentle inversion. Remove an aliquot and count the zoospore concentration using a hemocytometer.
6. Spin at 400 × *g* for 5 min at 4 °C in a swinging bucket rotor.
7. Decant most of the supernatant and resuspend the pellet by gently pipetting up and down. Add enough Petri's solution to set the concentration between 6 × 10^6 and 2.5 × 10^7 zoospores per mL.
8. Gently mix 800 μL of zoospores with 30 μg of DNA in a prechilled tube. If coexpressing two plasmids, use 15 μg of each.
9. Pipette the zoospores into a precooled 4 mm-gap cuvette. Cap the cuvette, wipe-dry with a tissue, and electroporate at 550 V, 50 μF, and 1575 Ω. The time constant is usually 1.8–2.6 ms.
10. Rapidly place the cuvette on ice, add 800 μL of regeneration media, and pipette into a 15 mL tube containing an additional 9 mL of regeneration media (*see* **Note 6**). Lay the tube on its side and incubate at 18 °C for 20 h. Longer incubation is not advised since the hyphae will mat together and be difficult to spread on selection plates.
11. After 20 h, count the concentration of germinated cysts. Multiply by the volume, and divide by the original number of zoospores to determine the regeneration rate. This is normally between 5 and 40%.
12. Concentrate the regenerated material by spinning the cultures at 1000 × *g* for 5 min at room temperature. Decant all but 1.6 mL of the liquid.

13. Gently resuspend the zoospores in the residual liquid by pipetting up and down, and spread 0.2 mL on each of eight rye-sucrose agar plates containing the appropriate antibiotics (*see* **Note 7**). Incubate at 18 °C. Colonies will appear after 8–12 days.

3.3 Identifying Transformants Expressing the FFP

1. Screen for expression using a fluorescence microscope (*see* **Note 8**). If the construct is expressed constitutively, it is usually sufficient to place a tuft of mycelia in a drop of water on a slide under a cover glass. Other life stages (zoospores, plant infection, etc.) will need to be examined if expression is not expected to occur in hyphae.
2. To avoid artifacts, verify that the localization is the same in multiple transformants and those with high and low levels of expression (*see* **Note 9**). Also check for the formation of protein aggregates (*see* **Note 10**).
3. Confirm that the FFP is of the expected size by immunoblotting (*see* **Note 10**).
4. For colocalization studies, it is best to choose transformants in which expression of the FFPs is well balanced.

3.4 Zoospore (Single-Nucleus) Purification

Expression within a culture is often not uniform, because some nuclei may not be transformed or may be epigenetically silenced. Single-nuclear purification is recommended in such cases. A procedure for *P. infestans* is described below. We have also purified transformants using a cell sorter with zoospore cysts.

1. Isolate zoospores as in Subheading 3.2. After passing the zoospores through 15 μm mesh, determine their concentration.
2. Spread about 25 zoospores on a 100 mm plate containing the appropriate selective media (*see* **Note 11**).
3. Incubate at 18 °C, select single colonies, and recheck for expression.

3.5 Fixing Phytophthora Tissues

Subcellular structures are often unstable during prolonged incubation on a slide, or under the heat of the microscope (*see* **Note 12**). This instability can be reduced by using the following protocols for fixing tissues, adapted from reference [16]. Fixation is less necessary for visualizing *Phytophthora* in plants, as it may increase autofluorescence of the host tissue.

1. To fix hyphae, place a tuft of mycelia in single-strength fixative solution for 1 h at room temperature. Remove the fixative with a pipette, and rinse three times in 50 mM Pipes buffer, 5 min each.
2. To fix sporangia, zoospores, or cysts, add an equal volume of double-strength fixative to the spore suspension. Mix by gently

inverting the tube, and incubate for 30 min at room temperature. Wash three times. For each wash, centrifuge at 1500 × *g* for 3 min, remove supernatant, resuspend the pellet in 3 mL of 50 mM Pipes buffer, and wait for 5 min.

3.6 Mounting Fixed and Live Specimens for Confocal Imaging

1. Place a drop of spore suspension, or a tuft of hyphae in a drop of liquid, on a microscope slide. Use a fine needle to tease apart the hyphae.
2. Add a #1.5 coverslip, taking care to avoid trapping air bubbles.
3. Blot the excess mounting liquid, and seal the coverslip on the slide by placing a minimal amount of clear nail polish around all edges. Although we have not experienced problems, there have been reports that the solvent in nail polish may reduce fluorescence [27].

3.7 Confocal Imaging of Phytophthora in Plants

The following methods have proved useful for imaging *P. infestans* in tomato and potato leaves, or in tubers. Normally, plant tissues are infected with zoospores or hyphae. Because infection protocols are often specialized to the pathogen, host, or host organ, their description is beyond the scope of this chapter.

1. Obtain infected tissues in which actively growing hyphae are abundant. Older hyphae are vacuolated and yield lower signals (*see* **Note 13**).
2. The mounting of plant tissue depends on the sample and microscope. On an upright microscope, a leaf (or portion) can be viewed by laying it on a slide and placing a water drop between the leaf and a water-dipping objective. If an inverted microscope is used, or only an oil immersion objective is available, place the leaf between a coverslip and slide, seal three sides of the coverslip with clear tape, fill the space between the slide and coverslip with water, and seal the fourth side with tape (*see* **Note 14**).
3. Flat, thicker samples such as potato tuber slices are more easily visualized by placing the tissue on a glass bottom culture dish, using an inverted microscope. One or two coverslips can help hold down the plant tissue.
4. Since plant tissues are thicker than hyphae, *z*-stack imaging can be useful.

3.8 General Guidelines for Confocal Imaging

The reader is advised to consult the manuals for their particular instrument as well as other resources [17] on confocal microscopy. Some additional tips from our experience with *Phytophthora* are as follows:

1. FFPs can be imaged in live or fixed samples. Unfixed tissues in water should be viewed immediately, while hyphae left in media can be viewed within a few hours. Fixed specimens should be viewed within 24 h.
2. The accurate detection of multiple FFPs can be challenging, and microscopes vary in their abilities to separate wavelengths. To avoid cross-talk, use the appropriate band-pass emission filters, test the settings using transformants expressing only one tag, and consider sequential scanning if possible.
3. Consider using additional methods to validate the FFP results, including placing tags at both the N and C-termini (*see* **Note 15**).

4 Notes

1. A list of fluorescent proteins and their spectral properties are available at Rodriguez et al. [4]. We have not observed superior expression of GFP or tdTomato optimized to *Phytophthora* codon usage compared to those optimized for plants or animals, but this might not be true for all tags. Some tags such as tdTomato and DsRed are popular due to their brightness; however, they should be used with caution as they form oligomers, which may cause artifacts in targeting or protein–protein interactions.
2. The protein should be expressed at a level that enables reliable imaging with minimal interference with cellular functions. However, some native promoters may not yield enough protein. Strong oomycete promoters include *ham34* or ribosomal protein promoters [5, 18, 19].
3. One widely used system allows the construction of N-terminal, C-terminal, or internal fusions with GFP, CFP, YPF, or mCherry in backbones containing *hpt* or *nptII* [5]. We have also expressed tdTomato and mScarlet in that system. The simplest cloning method is to use polymerase chain reaction to amplify the gene of interest with primers containing restriction sites compatible with the vector. Make sure that the protein is inserted in-frame, contains Kozak initiation sites, and no stop codon is between the gene of interest and tag.
4. Similar numbers of transformants are obtained when one plasmid is introduced compared to when two plasmids, each bearing a separate FPP, are cotransformed. Cotransformation is relatively efficient [20], and can be done with plasmids using the same or different selectable markers. Expressing two FFPs from one vector should minimize the unbalanced expression of the tags, which may complicate microscopic analysis. Examples

of transformants expressing GFP and tdTomato FFPs introduced by cotransformation are shown in Fig. 1a, b.

5. Incubation times for growing hyphae and releasing zoospores are species- and strain-specific. Most *P. infestans* should not be grown for more than 10 days, otherwise zoospore release may be slower and less synchronous. Most zoospores are usually released by 2 h, and longer incubations may reduce transformation efficiency.
6. It is often prudent to include antibacterials (e.g., 25 μg/mL penicillin G, or 50 μg/mL ampicillin plus 25 μg/mL vancomycin) and antifungals (e.g., 40 units/mL nystatin) in the media to avoid contamination.
7. For most *P. infestans* strains, we use 9 μg/mL G418 or 45 μg/mL hygromycin. The concentrations should be determined empirically for each strain.
8. A conventional fluorescence microscope is usually sufficient for the initial screening, although higher sensitivity might be achieved with a confocal. It is best to use young cultures because older ones tend to autofluoresce, and to visualize freshly mounted tissue. If using a digital (or confocal) microscope, avoid false positives by ensuring that the signal is present without having to turn the gain to high; viewing samples through an eyepiece may provide more reliable results.
9. Due to position and copy number effects, expression levels of the transgene will vary between transformants [21, 22]. Highly expressed FFPs may exhibit false localization. The transgene may also interfere with normal cell functions due to overexpression or silencing of the native gene. Therefore, the fitness of the strain (growth rate, sporulation rate, etc.) should be noted as this may influence interpretations of results.
10. Some FFPs may misfold or form aggregates; punctate signals may indicate the latter and should be interpreted cautiously. In addition, some FFPs are prone to proteolytic cleavage at the linker between the tag and protein of interest. While linkers containing small or polar amino acids (e.g., glycine or serine) may add flexibility and solubility to the FFP, other compositions may increase the stability of the FFP [23].
11. On average, plating 25 zoospores should result in several drug-resistant colonies. Because the efficiency of colony establishment from zoospores can vary depending on the species and strain, a greater number of zoospores may be required.
12. Live cell imaging is not always practical. Fixation will reduce stress responses including those resulting from desiccation and heating, is the only option for viewing zoospores, and is needed to accurately assess FFP localization in sporangia undergoing zoosporogenesis (Fig. 1c). Fixing plant tissue can be problem-

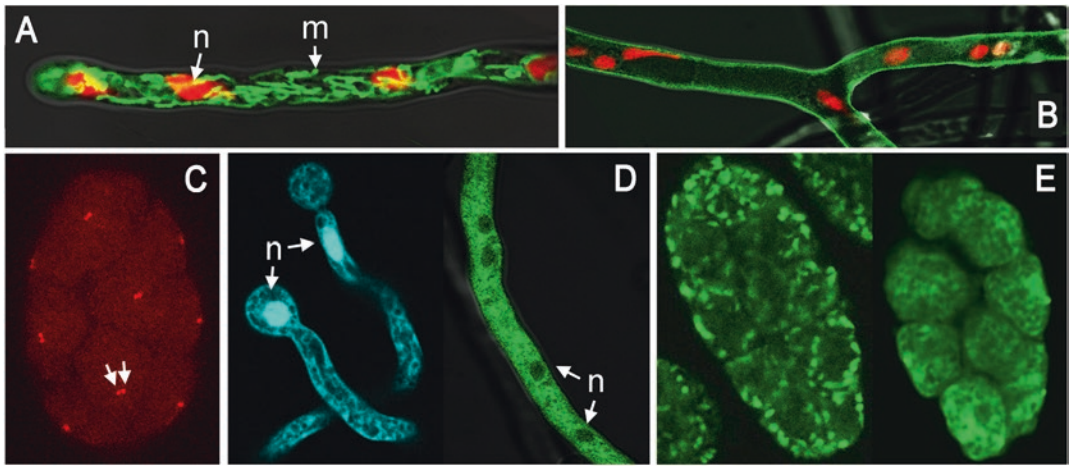


Fig. 1 Examples of FFPs in ***P. infestans*** transformants. (**a**) Colocalization in hyphae of GFP-tagged mitochondrial and tdTomato-tagged nuclear FFPs. Representative nuclei and mitochondria are denoted by n and m, respectively. (**b**) Cell wall and nuclear-localized FFPs in hyphae, expressing GFP and tdTomato, respectively. (**c**) tdTomato fused to basal body protein in sporangium cleaving into zoospores. Arrows point to a basal body doublet containing the FFP, and a weak signal is also in cytoplasm. With low signals, it is important to consider whether they are authentic or due to background noise. Here, examinations of multiple transformants and other FFPs that are targeted exclusively to basal bodies supported the veracity of the cytoplasmic signal. (**d**) The left image shows CFP in germinating cysts, with signals in cytoplasm (but not cytoplasmic vesicles) and nuclei. CFP, GFP, and YFP lack nuclear localization signals yet tend to accumulate in nuclei, so one must guard against drawing incorrect conclusions about protein targeting when expressing these small (27 kDa) markers alone or fused to small (<40 kDa) proteins. This is not an issue with larger FFPs, as illustrated in the right image where GFP was fused to a 68 kDa cytoplasmic protein; in this case, nuclei appear as nonfluorescing dark zones. Nuclear accumulation can also be reduced by using dimers or trimers of fluorescent protein such as GFP [5]. (**e**) Case where the location of the tag affects targeting. Both sporangia are undergoing zoosporogenesis and expressing the same ***P. infestans*** protein. The sporangium on the left is expressing the protein with N-terminal GFP, and resides in vesicles on the periphery of each developing zoospore. The right sporangium has a C-terminal GFP tag and is largely cytoplasmic. Subsequent experiments performed with other fluorescent tags indicated that the cytoplasmic location is more likely

atic as this usually damages its structure, making it harder to hold in place, and may increase autofluorescence.

13. Visualizing *Phytophthora* within plants can be challenging. It is helpful to scan at low magnification using eyepieces, and then switch to higher magnification and camera-assisted viewing. Autofluorescence can also be problematic. This derives mainly from chlorophyll (red), lignin (green), and compounds generated during wounding. It is thus preferable to minimize the number of cuts made in plant tissue, and use autofluorescence-blocking filters. Autofluorescence is greatest using blue or ultraviolet excitation wavelengths.
14. We recommend using inverted microscopes to visualize plant tissues. The objective on an upright instrument may damage the sample, and it is usually necessary to affix the sample in position using tape or a glue.

15. It is prudent to interpret with caution the results of experiments using FFPs. Confidence in subcellular localization data can be raised by testing both N and C-terminal tags, or different tags, since the site of fusion can affect targeting as shown in Fig. 1e. Members of the GFP family also show a propensity to translocate into nuclei on their own, unless they are part of a much larger protein (Fig. 1d). To help confirm the nature of an organelle, FFPs have been developed for *Phytophthora* basal bodies, cytoplasm, ER, Golgi, mitochondria, nuclei, peroxisomes, haustoria, and the cytoskeleton [5, 6, 8, 24]. Some organelles can also be labeled with fluorescent dyes, some of which are suited to imaging fixed or live cells. Dyes used for live-cell imaging are listed in Hickey et al. [25], but note that many of these have not been tested in *Phytophthora*. The use of FFPs to study dynamic cellular processes may also lead to erroneous conclusions since tags may affect a protein's half-life or its ability to interact with other proteins [26]. Therefore, researchers should consider supplementing studies using FFPs with other methods. For example, immunohistochemical analysis using an antibody against the native protein can confirm subcellular localization results from confocal microscopy.

Acknowledgments

This work has been supported by grants to H.S.J. from the National Institute of Food and Agriculture of the US Department of Agriculture, and the National Science Foundation.

References

1. Erwin DC, Ribeiro OK (1996) *Phytophthora* diseases worldwide. APS Press, St. Paul, MN
2. Panda A, Sen D, Ghosh A et al (2018) EumicrobeDBLite: a lightweight genomic resource and analytic platform for draft oomycete genomes. Molec Plant Pathol 19:227–237
3. Chalfie M, Tu Y, Euskirchen G et al (1994) Green fluorescent protein as a marker for gene-expression. Science 263:802–805
4. Rodriguez EA, Campbell RE, Lin JY et al (2017) The growing and glowing toolbox of fluorescent and photoactive proteins. Trends Biochem Sci 42:111–129
5. Ah-Fong AM, Judelson HS (2011) Vectors for fluorescent protein tagging in *Phytophthora*: tools for functional genomics and cell biology. Fungal Biol 115:882–890
6. Ah-Fong AM, Judelson HS (2011) New role for Cdc14 phosphatase: localization to basal bodies in the oomycete *Phytophthora* and its evolutionary coinheritance with eukaryotic flagella. PLoS One 6:e16725
7. Hua C, Meijer HJ, de Keijzer J et al (2013) GK4, a G-protein-coupled receptor with a phosphatidylinositol phosphate kinase domain in *Phytophthora infestans,* is involved in sporangia development and virulence. Molec Microbiol 88:352–370
8. Avrova AO, Boevink PC, Young V et al (2008) A novel *Phytophthora infestans* haustorium-specific membrane protein is required for infection of potato. Cell Microbiol 10:2271–2284
9. Wang S, Boevink PC, Welsh L et al (2017) Delivery of cytoplasmic and apoplast effectors from *Phytophthora infestans* haustoria by

distinct secretion pathways. New Phytol 216:205–215

10. Schornack S, van Damme M, Bozkurt TO et al (2010) Ancient class of translocated oomycete effectors targets the host nucleus. Proc Natl Acad Sci U S A 107:17421–17426
11. Dou D, Kale SD, Wang X et al (2008) RXLR-mediated entry of *Phytophthora sojae* effector Avr1b into soybean cells does not require pathogen-encoded machinery. Plant Cell 20:1930–1947
12. Caten CE, Jinks JL (1968) Spontaneous variability in isolates of *Phytophthora infestans*. I. Cultural variation. Can J Bot 46:329–348
13. Huitema E, Smoker M, Kamoun S (2011) A straightforward protocol for electrotransformation of *Phytophthora capsici* zoospores. Methods Molec Biol 712:129–135
14. Wu DL, Navet N, Liu YC et al (2016) Establishment of a simple and efficient *Agrobacterium*-mediated transformation system for *Phytophthora palmivora*. BMC Mcrobiol 16:204
15. Vijn I, Govers F (2003) *Agrobacterium tumefaciens* mediated transformation of the oomycete plant pathogen *Phytophthora infestans*. Molec Plant Pathol 4:456–467
16. Hardham A (2001) Investigations of oomycete cell biology. In: Talbot NJ (ed) Molecular and cell biology of filamentous fungi: a practical approach. Oxford University Press, Oxford, pp 127–155
17. Paddock SW (2014) Confocal microscopy methods and protocols, second edition. Methods in molecular biology, vol 1075. Humana Press, New York
18. Poidevin L, Andreeva K, Khatchatoorian C, Judelson HS (2015) Comparisons of ribosomal protein gene promoters indicate superiority of heterologous regulatory sequences for expressing transgenes in *Phytophthora infestans*. PLoS One 10:e0145612
19. Dou D, Kale SD, Wang X et al (2008) Conserved C-terminal motifs required for avirulence and suppression of cell death by *Phytophthora sojae* effector Avr1b. Plant Cell 20:1118–1133
20. Judelson HS (1993) Intermolecular ligation mediates efficient cotransformation in *Phytophthora infestans*. Mol Gen Genet 239:241–250
21. Judelson HS, Dudler R, Pieterse CMJ et al (1993) Expression and antisense inhibition of transgenes in *Phytophthora infestans* is modulated by choice of promoter and position effects. Gene 133:63–69
22. Dunn AR, Fry BA, Lee TY et al (2013) Transformation of *Phytophthora capsici* with genes for green and red fluorescent protein for use in visualizing plant-pathogen interactions. Austral Plant Path 42:583–593
23. Huang L, Pike D, Sleat DE et al (2014) Potential pitfalls and solutions for use of fluorescent fusion proteins to study the lysosome. PLoS One 9:e88893
24. Hua CL, Kots K, Ketelaar T et al (2015) Effect of Flumorph on F-actin dynamics in the potato late blight pathogen *Phytophthora infestans*. Phytopathology 105:419–423
25. Hickey PC, Swift SR, Roca MG et al (2005) Live-cell imaging of filamentous fungi using vital fluorescent dyes and confocal microscopy. Meth Microbiol 35:63–97
26. Yewdell JW, Lacsina JR, Rechsteiner MC et al (2011) Out with the old, in with the new? Comparing methods for measuring protein degradation. Cell Biol Int 35:457–462
27. Callis G. (2009) Nail polish is not suitable for sealing coverslips for fluorescence microscopy, but the bottle and brush are useful. Biotech Histochem 81:160



Chapter 11

Methods for Transient Gene Expression in Wild Relatives of Potato

Sidrat Abdullah and Dennis Halterman

Abstract

Genomics studies in potato and other plants have elucidated a large number of genes involved in a wide array of phenotypes. In particular, recent bioinformatic and genomic analyses of oomycetes and fungi have identified many effectors for which the corresponding host resistance-eliciting receptor remains to be found. Functional testing of host resistance gene candidates can be accomplished by generating whole plant transformants to either overexpress or silence these genes to obtain a visible phenotype. However, this is time consuming. Alternatively, *Agrobacterium tumefaciens* can be used to transiently express genes in plant tissue to observe phenotypic changes. Wild relatives of potato contain a large amount of genotypic diversity and are an excellent tool to identify important agronomic traits, including resistance to diseases. The methods presented here help to facilitate the screening of wild potato accessions using agroinfiltration.

Key words Agroinfiltration, Phenotype screening, Potato, Wild species relatives

1 Introduction

Transient gene expression using agroinfiltration is an effective tool for exploration of candidate gene function in plants [1]. Transient expression assays using agroinfiltration take advantage of the fact that *Agrobacterium tumefaciens* is able to introduce DNA into plant cells, which can be transcribed and translated without the need to produce transgenic plants. Methods for transient assays and agroinfiltration have been published previously [1–7] but most are focused on low-throughput testing of a single host genotype (e.g., *Nicotiana benthamiana*) rather than high-throughput screening of a large number of host genotypes. High-throughput assays for potato have been used, but rely on the potato virus X (PVX) virus to express the candidate gene beyond the initial site of inoculation [1, 2, 8], which could be problematic if the host genotype is PVX resistant. In several plant species, infiltration is amenable due to the even leaf surface, when the bacterial suspensions can easily penetrate through to the parenchyma tissue.

Wenbo Ma and Thomas Wolpert (eds.), *Plant Pathogenic Fungi and Oomycetes: Methods and Protocols*, Methods in Molecular Biology, vol. 1848, https://doi.org/10.1007/978-1-4939-8724-5_11,
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Over the past decade, development of "omics" technologies has revealed myriad oomycete and fungal proteins, termed effectors, that are essential for pathogen virulence [9, 10]. Identification and integration of the host resistance gene products that recognize pathogen effectors to elicit resistance responses is an essential component of breeding for improved crop varieties. Wild species relatives of cultivated potato contain a wealth of important agronomic traits, including disease resistance. Transient expression has been used previously to explore potato wild relatives to identify resistance to diseases such as *Phytophthora infestans* [2], but infiltration in many potato wild species is difficult compared to other plants. Here we present a method for agroinfiltration in wild potato accessions that takes into account factors such as age of the plant, leaflet structure, and environmental factors (temperature and humidity) to optimize efficient infiltration. Leaflets with variable morphologies were infiltrated at three different developmental ages at variable temperatures and humidity in the greenhouse. We have found that this approach is efficient in screening of wild potato accessions for the desired phenotype. Despite the ability to infiltrate most accessions, we have found that repetition is a critical factor. In our experience using wild potato relatives, there can be some variability in responses when using the same gene construct and host genotype across different experiments. It is therefore necessary to replicate agroinfiltrations extensively (10+ times occasionally) to properly determine a consensus of the correct phenotype.

2 Materials

Use deionized, autoclaved water to prepare all solutions, media, and reagents.

1. 1500 ppm gibberellic acid (GA3) for seed treatment
2. LB Media (for 1 L): 10 g Bacto tryptone, 5 g yeast extract, 10 g NaCl (plus 15 g Bacto agar for solid media)
3. 75 mg/mL kanamycin
4. *A. tumefaciens* strain GV3101 with appropriate binary vector constructs
5. MMA infiltration medium (for 1 L): Murashige and Skoog medium (*see* Table 1), 1.95 g MES, 20 g sucrose; adjusted pH to 5.6 with 1 M NaOH.
6. 1 M stock (196.2 mg/mL) 3′-5′ Dimethoxy-4′-hydroxy acetophenone (acetosyringone) (*see* **Note 1**).
7. 1 mL needleless syringe
8. Needle (to make wound, if necessary)
9. Shaking incubator: For growing bacterial cultures.

Table 1
Composition of Murashige and Skoog medium

Stock Name	Constituents	Conc. of stock solution g/L	Vol. of stock solution in final medium mL/L
A	NH_4NO_3	82.5	20
B	KNO_3	95.0	20
C	H_3BO_3	1.24	5
	KH_2PO_4	34.0	
	KI	0.166	
	$Na_2MoO_4.2H_2O$	0.05	
	$CoCl_2.6H_2O$	0.005	
D	$CaCl_2.2H_20$	88.0	5
E	$MgSO_4.7H_2O$	74.0	5
	$MnSO_4.1H_2O$	3.38	
	$ZnSO_4.7H_2O$	1.72	
	$CuSO_4.5H_2O$	0.005	
F	Na_2EDTA	7.45	5
	$FeSO_4 .7H_2O$	5.57	

10. Spectrophotometer: For OD_{600} measurements.
11. Centrifuge capable of up to 3,200 × *g*.
12. Plant materials: True potato seeds (TPS) of wild potato accessions. Wild accessions of TPS can be obtained from the NRSP-6 US Potato Genebank (https://www.ars-grin.gov/nr6/).

3 Methods

All the procedures for infiltration materials preparation are carried out at room temperature, unless otherwise noted.

3.1 Growing TPS (True Potato Seed)

1. Treat true potato seeds (30–50 based on germination percentage information) overnight in 500 μL of GA3 in a 1.5 mL tube.
2. The following day, sow seeds in the greenhouse in 4-in. pots containing soil (*see* Fig. 1a and **Note 2**).
3. When most seeds germinate (approximately 2 weeks), transfer individual seedlings to a flat containing 32 wells (*see* Fig. 1b). Start infiltrations 3 weeks after transplanting and infiltrate

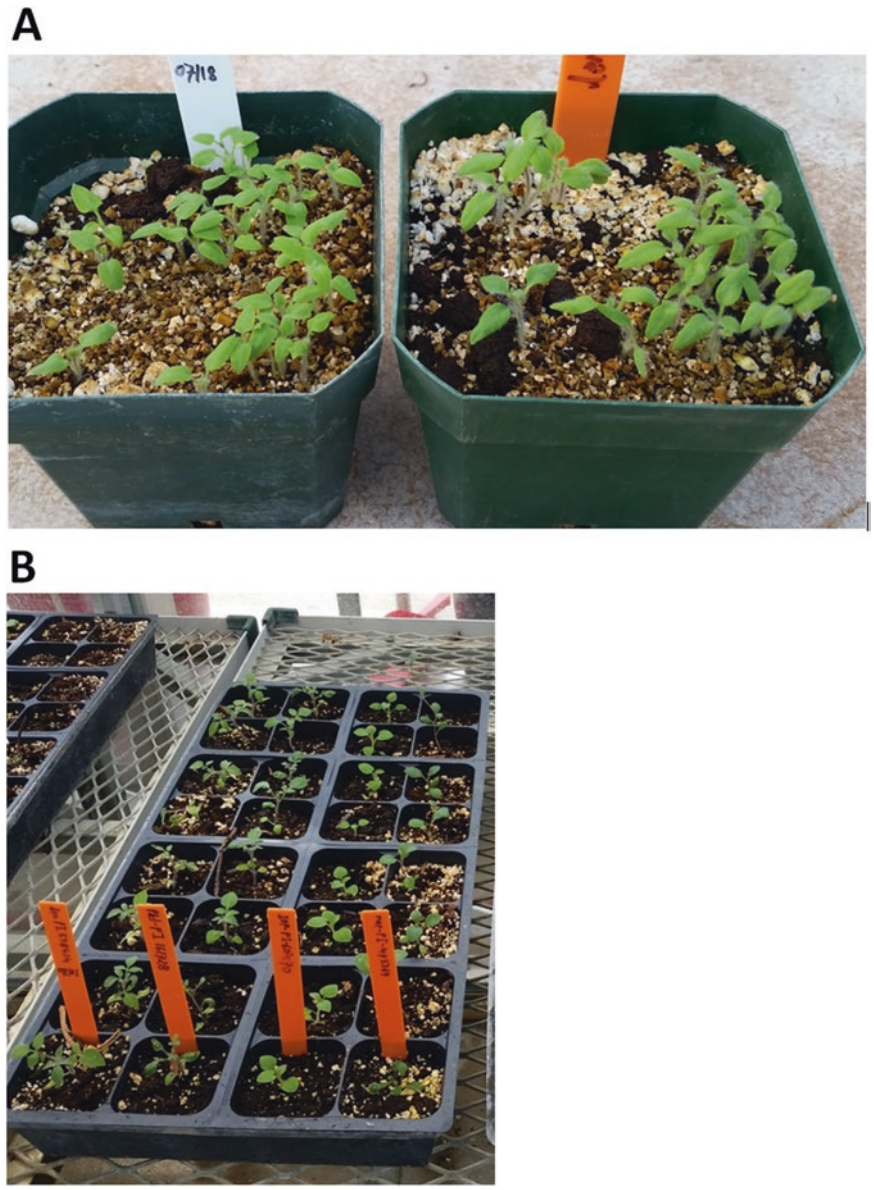


Fig. 1 Seedlings grown in pots and flats. (**a**) Seedlings in pots 12 days after sowing. (**b**) 2-week old seedlings transplanted to a 32-well flat

three separate times per individual plant at 1-week intervals (*see* **Note 3**).

3.2 Growing Agrobacterium Cultures

1. Inoculate LB solid agar media (supplemented with 50 μg/mL kanamycin, if necessary) with *A. tumefaciens* stock and incubate at 28 °C for 2 days.
2. Use a single colony to inoculate 20 mL liquid LB media (containing 50 μg/mL kanamycin, if necessary) and place in a shaking incubator at 28 °C at 200 rpm overnight

3. The next morning, harvest bacteria by centrifugation at 3,200 × *g* for 5 min.
4. Discard the supernatant, and resuspend the bacterial pellet in MMA (5–10 mL). Centrifuge again at 3,200 × *g* for 5 min and the pour off the supernatant.
5. Resuspend the cells with 5 mL MMA containing acetosyringone (*see* **Note 1**) and determine the cell concentration using the OD_{600}. Dilute the cells to an OD_{600} of 0.6 with MMA containing acetosyringone.
6. Incubate the culture at room temperature for 1–2 h before infiltration to induce expression of the Agrobacterium T-DNA. In our studies we start infiltrations after 1.5 h.

3.3 Choice of Plants/ Leaflets for Infiltration

Several factors should be considered for infiltration of wild potato accessions including:

1. Number of plants per accession: Accessions of potato exist as populations. Therefore, individual plants within an accession might be genetically different from one another. For this reason, all different gene constructs and controls should be tested on each plant, using 6–10 plants per accession. Individual plants should be marked and propagated for replications by cuttings, if the desired reaction is observed.
2. Growth conditions: Seedlings should be grown in a flat (as described in Subheading 3.1) in a greenhouse under 16 h light/8 h dark cycles at 17–24 °C and less than 60% relative humidity. (*see* **Note 4**). Seedlings should be treated with general-use fertilizer once a week and watered whenever necessary (*see* **Note 5**).
3. Plant physiological and developmental age: The ideal plant age for infiltration is typically 4 weeks after transplanting (*see* **Note 6**). Select the third/fourth leaf based on the plant's developmental age where the leaves have a smooth surface with no irregular shapes (*see* **Note 7**).
4. Morphology of the leaf: When selecting leaves also consider the morphological characteristics of individual plants including: trichomes, veins, structure of the mid rib, and interveinal distribution on the ventral side of the leaf. Some accessions have dense trichomes on the ventral side or develop prominent veins on the leaf surface that increase the difficulty of infiltrating bacterial suspensions (*see* Fig. 2 and **Note 8**). If these morphological characteristics are prominent, you may be able to find an older leaf on the lower part of the plant where interveinal space has increased.

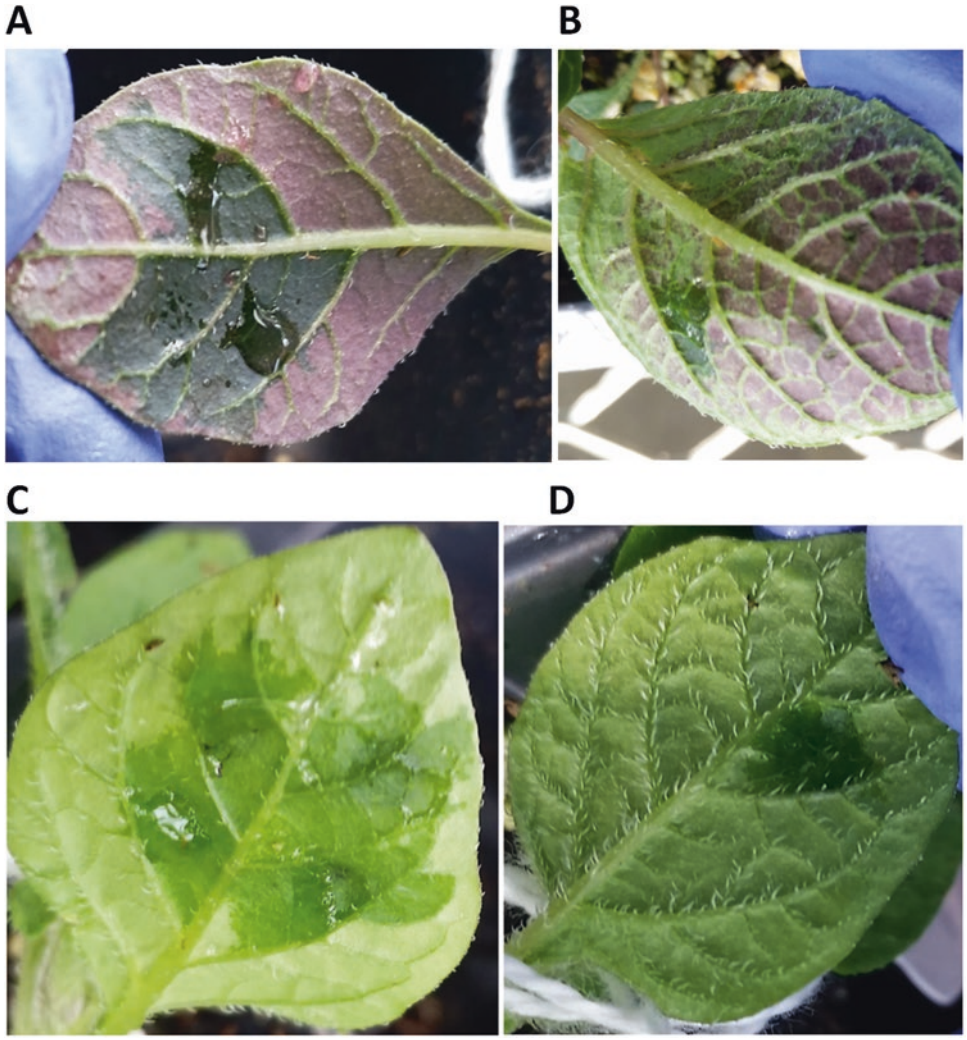


Fig. 2 Differences in leaf morphology affects infiltration efficiency. Individuals of the same species (**a**/**b** and **c**/**d**) can differ in interveinal space and vein prominence, leading to easy infiltration (**a**, **c**) or more difficult infiltration (**b**, **d**)

3.4 Syringe Infiltration

Eye protection should be used, as infiltrations can result in aerosols or spray. Always use a new syringe for each construct.

1. Infiltrate leaflets 3–5 weeks after transplanting. Briefly mix the agrobacterium suspension and draw 1 mL of the suspension into a 1 mL needleless syringe.
2. Select a plant with a fully expanded leaf for infiltration. For potato, the third/fourth leaf from the bottom of the plant typically works best for infiltration (as opposed to upper leaves (*see* ref. 6). In compound leaves, there is no difference with respect to the position of the leaflet on the leaf (terminal versus lateral leaflets).



Fig. 3 Water-soaked infiltration spots can be seen on the left side of the leaf. Circles surround spots that have been lightly wounded with a needle to facilitate infiltration

3. To infiltrate, turn the leaf (or leaflet) upside down and place the needleless syringe against the epidermis. Increase pressure using a thumb or finger on the other side of the leaflet and inject the suspension slowly into the interveinal space of the leaflet mesophyll. In some cases, infiltration efficiency can be increased by first wounding the leaf with a needle (*see* **Note 9**).
4. Successful infiltration is indicated by a darkened area in the leaf (water-soaking) (*see* Fig. 3).
5. Phenotypes should be visible 3–4 days after infiltration plants and continue up to 7–10 days (*see* **Note 10**).

4 Notes

1. To avoid multiple freeze–thaw cycles, aliquot 200 μM acetosyringone in small PCR tubes (20 μL/tube). Use 10 μL per 50 mL MMA. Add acetosyringone to MMA stocks immeadiately prior to carrying out experiments.
2. Wash GA treated seed with water by inserting a water bottle nozzle into the tube and spread the seeds over the top of the soil while expelling water. Spread a thin layer of vermiculite over the seeds and water gently so that the layer of vermiculite is dispersed evenly.
3. Best infiltration is generally observed 4 weeks after transplanting, but variability exists among accessions.
4. For ideal infiltration, humidity should be maintained below 60%. High humidity may increase the difficulty of infiltrations. Temperature should not exceed 24 °C (*see* refs. 4, 7).

5. Seedlings should not be watered the day of infiltration. Too much moisture/water enhances stomatal closure.
6. Seedling growth rate may vary between accessions and plants should be transferred and infiltrated accordingly.
7. Some leaves exhibit an irregular leaf surface with prominent veins and small interveinal spaces, which can hinder infiltration (*see* Fig. 2). In these cases, it might be necessary to allow plants to grow longer until older leaves with larger interveinal spaces develop.
8. Wild potato accessions have wide variation in leaf morphology. Some leaves are difficult to infiltrate due to small interveinal space. In these cases infiltration in different spots of the same leaf may be necessary.
9. Use a needle to wound the leaf before infiltrating with the syringe, if the leaf surface does not allow smooth infiltration (*see* Fig. 3). Infiltrations should be attempted without wounding, when possible, as wounding might result in nonspecific cell death (*see* ref. 1). If possible, wounding should not pierce through the other side of the leaf, as the bacterial suspension may go through the leaf, resulting in the use of more suspension than usual.
10. Some potato accessions take longer than 3 days to develop a reaction.

References

1. Thirumala DK, Edgar H, Sophien K (2007) In planta expression of oomycete and fungal genes. Methods in Molecular Biology Plant–Pathogen Interactions: Methods and Protocols 354:35–43
2. Du J, Rietman H, Vleeshouwers VGAA (2014) Agroinfiltration and PVX Agroinfection in potato and *Nicotiana benthamiana*. J Vis Exp 83
3. Tadeusz W, Anna T, Richard M (2005) Optimization of agrobacterium-mediated transient assays of gene expression in lettuce, tomato and Arabidopsis. Plant Biotechnol J 3:259–273
4. Qiang C, Huafang L, Jonathan H, Jake S, Kahlin L, Matthew D (2014) Agroinfiltration as an effective and scalable strategy of gene delivery for production of pharmaceutical proteins. Adv Tech Biol Med 1(1). https://doi.org/10.4172/atbm.1000103
5. Leuzinger K, Dent M, Hurtado J, Stahnke J, Lai H, Zhou X, Chen Q (2013) Efficient Agroinfiltration of plants for high-level transient expression of recombinant proteins. J Vis Exp:77
6. Vaghchhipawala ZMR, Muthappa SK, Kirankumar SM (2011) Agroinoculation and Agroinfiltration: simple tools for complex gene function analyses. Andy Pereira (ed.), plant reverse genetics: methods and protocols. Methods Mol Biol 678:65–76
7. Dillen W, De CJ, Kapila J, Zambre M, Van MM, Angenon G (1997) The effect of temperature on *Agrobacterium tumefaciens*-mediated gene transfer to plants. Plant J 12:1459–1463
8. Vleeshouwers VG, Rietman H, Krenek P, Champouret N, Young C, Oh SK, Wang M, Bouwmeester K, Vosman B, Visser RG, Jacobsen E, Govers F, Kamoun S, Van der Vossen EAG (2008) Effector genomics accelerates discovery and functional profiling of potato disease resistance and *Phytophthora infestans* avirulence genes. PLoS One 3:e2875
9. Gibriel HAY, Thomma BPHJ, Seidl MF (2016) The age of effectors: genome-based discovery and applications. Phytopathology 106:1206–1212
10. Varden FA, De la Concepcion JC, Maidment JHR, Banfield MJ (2017) Taking the stage: effectors in the spotlight. Curr Opin Plant Biol 38:25–33



Chapter 12

Host-Induced Gene Silencing (HIGS) for Elucidating *Puccinia* Gene Function in Wheat

Chuntao Yin and Scot H. Hulbert

Abstract

Biotrophic fungi (*Puccinia* spp.) cause devastating diseases of wheat and other cereal species globally. The function of large repertories of genes from *Puccinia* spp. still needs to be discovered to understand the infection process of these obligate parasites, eventually to protect plants from rust diseases. Functional analysis of targeted genes is challenging due to the inherent difficulties with culturing the fungus and transforming the host. RNA interference (RNAi) is a conserved gene regulation process in eukaryotes and known to be a powerful genetic tool in plant biotechnology. More recently, host-induced gene silencing (HIGS) has been developed to assess pathogen gene function in plants. HIGS is an RNAi-based process where double stranded RNA (dsRNA) homologous to a pathogen gene can be expressed in a plant to induce targeted silencing of the pathogen gene. Here we described a detailed HIGS protocol for functional analysis of rust genes from *Puccinia* species in cereals. As an example we describe an experiment silencing the tryptophan 2-monooxygenase gene (*Pgt-IaaM*) from *Puccinia graminis* f. sp. *tritici* (*Pgt*) that is involved in virulence to wheat.

Key words HIGS, *Puccinia*, Pathogenicity, *Barley mosaic stripe virus*, *Pgt-IaaM*

1 Introduction

RNA interference (RNAi), triggered by dsRNA, is a conserved cellular process in eukaryotic organisms which plays roles in gene regulation and defense against pathogen infection [1]. The dsRNAs that direct the process are either made in the organisms themselves to regulate expression of their own genes, or possibly by invading organisms like RNA viruses which replicate through dsRNA intermediates. To capitalize on these antiviral features of RNAi processes, gene fragments of host plants can be added to viral genomes to trigger silencing of the respective host gene after infection with the recombinant virus. This virus-induced gene silencing (VIGS) has emerged as a powerful and effective tool for the analysis of gene function in plants, especially those in which stable transformants are

Wenbo Ma and Thomas Wolpert (eds.), *Plant Pathogenic Fungi and Oomycetes: Methods and Protocols*, Methods in Molecular Biology, vol. 1848, https://doi.org/10.1007/978-1-4939-8724-5_12,
© Springer Science+Business Media, LLC, part of Springer Nature 2018

difficult or time consuming to produce [2, 3]. *Barley stripe mosaic virus* (BSMV), comprising the tripartite genome RNAα, RNAβ, and RNAγ, is a popular vector for VIGS in dicots and monocots. Furthermore, host-induced gene silencing (HIGS) has recently been shown to induce gene silencing in pathogenic fungi by *in planta* expression of dsRNA corresponding to a pathogen gene to direct silencing of that gene upon infection of the host plant. The technique can be used to test the essentiality of a gene or process in fungi or oomycetes for pathogenicity and ultimately to create resistant cultivars once the appropriate genes are identified. For example, expressing a dsRNA targeting transcripts of *Fusarium verticillioides* or *Blumeria* genes in cereals has been shown to enhance plant resistance to these fungal pathogens [4, 5]. HIGS has also been used to suppress fungal diseases caused by *Puccinia* spp. [6–9], other *Fusarium* species [10, 11], *Sclerotinia sclerotiorum* [12], or downy mildew caused by an oomycete pathogen [13]**.** These studies corroborate that HIGS is a feasible strategy to develop plants resistant to diseases and improve the agronomic value of crops.

A rapid HIGS assay can be very important for identifying a suitable targeted gene for engineering resistance. Genes involved in an essential process in a pathogen are often redundant and silencing one may have little effect. Fungal genes that are highly expressed in infection processes, including those essential for structural components like chitin and ergosterol, development, primary or secondary metabolism, and pathogenicity can be promising targets for HIGS [14]. In addition, the corresponding dsRNA and siRNA of targeted genes should avoid unintended targets in the host plant or potentially beneficial microbes. Selection of target gene sequences with little homology to the host or other fungal sequences in databases is a good starting point to achieve this. Alternatively, if the goal is to engineer resistance to multiple, closely related species, like the three *Puccinia* species that attack wheat, larger or more conserved gene fragments may be more appropriate [7].

The method described here has been used to silence the gene *Pgt-IaaM* of *Puccinia graminis*, which is highly expressed in haustorial cells and encodes a tryptophan 2-monooxygenase, causing overproduction of auxin in colonized plant cells. The *Pgt-IaaM* gene was silenced by HIGS to demonstrate that it was required for full virulence of *Puccinia* species in wheat [8].

2 Materials

2.1 Plant and Fresh Urediniospores of Pgt Preparation

1. Wheat (*Triticum aestivum*) cultivar McNair 701 seeds.
2. Potting mix and Miracle-Gro water soluble all-purpose plant food.
3. Urediniospores of *Pgt* CRL 75-36-700-3, race SCCL (Pgt7A).
4. Talcum powder.

5. Growth chamber.
6. Dew chamber.

2.2 Construction of BSMV-Derived Vectors

1. cDNA from *Pgt* infected-leaf.
2. *Pgt-IaaM* primers (Forward primer: 5′- ATAAGAATG CGGCCGCTAAACTATCAAGTCTTGGAGCATTCAC TCTGG-3′ and reverse primer: 5′-CCTTAATTAAGGGACAT TCATGGAAGTCCTCAACGC-3′).
3. Phusion high-fidelity PCR kit (New England Biolabs, MA).
4. Restriction endonuclease *Not*I and *Pac*I.
5. Three BSMV component vectors (pα, pβΔβa, and pγ) derived from the tripartite BSMV genome (RNAα, RNAβ, and RNAγ), respectively [15].
6. T4 DNA ligase enzyme and 10× reaction buffer.
7. One shot Top10 competent cells of *E. coli* (Thermo Fisher Scientific, MA) and LB agar plates with 50 mg/L carbenicillin.
8. S.O.C medium (0.5% Yeast Extract, 2% Tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose)
9. QIAquick gel extraction kit (Qiagen, CA).
10. Water bath.
11. Centrifuge, incubator, and shaker.

2.3 In Vitro Transcription of BSMV Viral RNAs

1. QIAprep spin miniprep kit (Qiagen, CA).
2. Restriction endonucleases *BssHII*, *MluI*, and *SpeI*.
3. The negative control plasmid DNA (pγ:MCS) carrying only a 121 bp fragment of the multiple cloning sites (MCS) from the plasmid pBluescript K/S (i.e., without *Puccinia* spp. Sequences [16].
4. mMESSAGE mMACHINE T7 transcription kit (Thermo Fisher Scientific, MA).

2.4 Inoculation of BSMV Viruses

1. 10× GP buffer: 18.77 g glycine, 26.13 g K_2HPO_4 dibasic, brought to 500 mL with ddH_2O, autoclave for 20 min.
2. FES Buffer: 100 mL GP buffer, 5 g sodium pyrophosphate decahydrate, 5 g bentonite, 5 g celite, brought to 500 mL with ddH_2O and autoclaved.

2.5 RNA Isolation from Infected Leaves

1. TRIzol Reagent (Thermo Fisher Scientific, Waltham, MA).
2. Chloroform, isopropanol and 70% ethanol.
3. 4 °C centrifuge.
4. NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, MA).

2.6 cDNA Synthesis

1. iScript™ reverse transcription supermix for RT-qPCR (Bio-Rad, CA).
2. Thermal cycler.

2.7 Reverse Transcriptase Quantitative PCR (RT-qPCR)

1. RT-qPCR primers for *Pgt-IaaM* (Forward primer: 5′-GGGCAACAAGAATGGGAAGA-3′ and reverse primer: 5′-CCACTAAGCGGCAGATGTAAG-3′).
2. RT-qPCR primers for *Pgt-actin* gene (Forward primer: 5′-TGTCGGGTGGAACGA CCATGTATT-3′ and reverse primer: 5′-AGCCAAGATAGAACCACCGATC CA-3′).
3. iTaq™ Universal SYBR® Green Supermix (Bio-Rad, CA).
4. Real-time PCR plates and plate sealing film.
5. Real-time PCR detection system.

3 Methods

3.1 Preparation of Plants and Fresh Urediniospores of Pgt

1. Plant the wheat cultivar McNair 701 (*see* **Note 1**) in two pots (10 cm in diameter, ~ten seeds/pot) containing potting mix. Grow the seedlings in a rust-free growth chamber at 24 °C during day (16 h) and 20 °C at night (8 h), 23–50% relative humidity. Water plants every 2 days and fertilize with a dilute nutrient solution weekly.
2. At the two-leaf stage (about 12 days after planting), uniformly inoculate the wheat seedlings with urediniospores of Pgt7A mixed with talcum powder at a ratio of 1:20 by dusting or rubbing the mix onto seedling leaves.
3. Place the inoculated plants in a dark dew chamber (*see* **Note 2**) for 24 h.
4. Move the plants back to the growth chamber for 12–14 days until fresh urediniospores of *Pgt* erupt from uredinia on the surface of the leaves (*see* **Note 3**).
5. Collect fresh urediniospores of *Pgt* and store at 4 °C for further uses.
6. Plant the wheat cultivar McNair 701 in 12 pots (as in **step 1** except with four seeds sown in each pot) and grow as in **step 1**.

3.2 Construction of BSMV-Derived Vectors

1. Amplify the targeted *Pgt-IaaM* fragment by PCR from *Pgt* infected-leaf cDNA using *Pgt-IaaM* primers. Perform PCR in a 25 μL reaction containing 5 μL 5× phusion HF buffer, 0.5 μL 10 mM dNTPs, 0.5 μL cDNA template, 1.25 μL 10 mM forward primer, 1.25 μL 10 mM reverse primer, 0.25 μL phusion DNA polymerase, and nuclease-free H_2O to bring the total volume to 25 μL. Incubate the reaction at 98 °C for 30 s, followed by 30 cycles of 98 °C for 10 s, 55 °C

for 30 s and 72 °C for 30 s, final extension at 72 °C for 5 min (*see* **Note 4**).

2. Load 25 μL of PCR amplicons on 1.5% w/v agarose gel along with a 100 bp ladder marker. Cut the desired fragment from the gel under UV light and purify PCR amplicons with QIAquick gel extraction kit according to the manufacturer's instructions.
3. Double digest PCR amplicons and the pγ vector with restriction endonucleases *NotI* and *PacI*. Perform double digestion in 30 μL reaction containing 20 μL of 20 ng/μL purified amplicons or 5 μg pγ DNA, 3 μL 10× NEBuffer 2, 10 units of *NotI*, 10 units of *PacI*, 0.3 μL 100× BSA, and nuclease-free H_2O to bring the total volume to 30 μL. Incubate the reaction at 37 °C for 2 h. and then load the DNA on 1.0% w/v agarose gel along with 1 kb and 100 bp ladder marker. Cut the desired DNA fragments from the gel under UV light and purify the DNA with QIAquick gel extraction kit according to the manufacturer's instructions.
4. Ligate vector and insert. Perform the ligation in 10 μL reaction containing 1 μL 10× T4 DNA ligase buffer, 30 ng vector DNA, 10 ng insert DNA, 1 μL T4 DNA ligase, and nuclease-free H_2O to bring total volume to 10 μL. Incubate the reaction at room temperature for 2–12 h. and store at 4 °C.
5. *E. coli* transformation of ligated reaction. Add 2 μL of the ligated reaction mix into 50 μL Top10 competent cells and mix completely. Incubate the reaction on ice for 30 min. Heat-shock at 42 °C for 1 min and return to ice for 2 min. Then add 0.25 mL S.O.C. medium to reaction and culture at 225 rpm in a 37 °C shaker for 1 h. Spread 100 μL of culture on an LB agar plate containing 50 mg/L carbenicillin.
6. Invert the plate and incubate at 37 °C overnight.
7. Select colonies, amplify cloned fragments by PCR using *Pgt-IaaM* primers and confirm the sequences.
8. Store glycerol stock of selected *E. coli* strains carrying the pγ:*Pgt-IaaM* plasmid at −80 °C.

3.3 In Vitro Transcription of Viral RNAs

1. Preparation of BSMV plasmid DNA. Streak LB plates containing 50 mg/L carbenicillin with each *E. coli* glycerol stock (pα, pβΔβa, pγ:MCS, and pγ:*Pgt-IaaM*). Incubate plates upside down at 37 °C overnight. Pick a single colony for each plasmid and inoculate a tube containing 10 mL of LB medium supplemented with 50 mg/L carbenicillin. Incubate at 37 °C overnight with shaking at 250 rpm. Extract plasmid DNA using the QIAprep spin miniprep kit according to the manufacturer's instructions.

2. Linearize plasmid DNA. Perform linearization in a 30 μL reaction containing 2.0 μg plasmid DNA, 3 μL 10× enzyme buffer, 10 units of enzyme, and nuclease-free H_2O to bring the total volume to 30 μL. Linearize pα with *MluI* at 37 °C, pβΔβa with *SpeI* at 37 °C, and pγ:MCS and pγ:*Pgt-IaaM* with *BssHII* at 50 °C. Incubate the digestion reactions for 2–3 h (*see* **Note 5**).
3. Purify the linearized plasmids. Add 70 μL of nuclease-free H_2O, 1 μL 0.5 M EDTA, 3 μL 3 M sodium acetate (pH = 5.2) and 200 μL cold 100% ethanol to the plasmid DNA digestion reactions. Mix well and keep at −20 °C for 30 min. Centrifuge at 21,000 × *g* for 15 min. Discard the supernatant. Add 300 μL of 70% ethanol and centrifuge at 18,000 × *g* for 5 min. Air-dry the pellet and resuspend the DNA in 15 μL of nuclease-free H_2O.
4. In vitro transcription of viral RNAs. Perform in vitro transcription in 30 μL reactions containing 15 μL NTP/CAP, 1.5 μg linearized plasmid DNA, 3 μL 10× reaction buffer, 3 μL enzyme mix, and nuclease-free H_2O to bring the total volume to 30 μL. Incubate the reaction at 37 °C for 2 h (*see* **Note 6**).
5. Quality check for the RNA products by running 1 μL of each reaction on 1% w/v agarose gel. There should be a bright band at approximately 1.5–2 kb (Fig. 1). Store the RNA products at −80 °C for plant inoculation (*see* **Note 7**).

3.4 Inoculation of Wheat Leaves by BSMV

1. 30 μL of the in vitro transcription products of pα, pβΔβa, and pγ:MCS or pγ:*Pgt-IaaM* are mixed in a 1:1:1 ratio (2.5 μL each for a total of 7.5 μL/plant). Place the transcript mix on ice, add 540 μL of FES buffer and mix them well on ice.
2. Inoculate the lower two leaves of 12-day-old wheat seedlings with the FES/transcript mix. Inoculate 12 plants with the FES/(pα, pβΔβa, and pγ:MCS mix) and 12 plants with the FES/(pα, pβΔβa, and pγ:*Pgt-IaaM* mix).
   (a) Wear gloves and place the tip of your forefinger and thumb together.
   (b) Add 50 μL of the FES/transcript mix between your forefinger and thumb.
   (c) Place forefinger and thumb on either side of the base of the leaf to be inoculated.
   (d) Holding the stem of plant with your other hand, using light pressure, rub the leaf with the inoculating hand from the base to the tip in a single motion.
   (e) Change gloves when inoculating plants using a different BSMV construct.
   (f) Lightly mist the plants with water and move the plants back to the growth chamber.

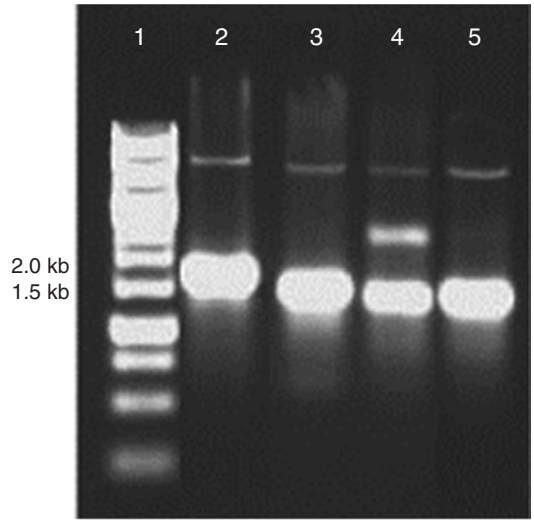


Fig. 1 An example 1% agarose gel loaded with in vitro transcription products from BSMV-derived vectors. Lane 1, 1 kb marker; lane 2, pα; lane 3, pβ; lane 4, pγ:MCS; lane 5, pγ:*Pgt-laaM*

(g) BSMV symptoms (yellow streaking) start to appear on the third and fourth leaves in 7–10 days and the inoculated leaves will appear wounded and die (Fig. 2).

3.5 Pgt Inoculation

Inoculate the third and fourth leaves of all BSMV-infected plants that show mild viral infection symptoms, as well as 12 additional healthy plants that have not been exposed to the virus, with freshly collected urediniospores of *Pgt* as described in Subheading 3.1, **steps 2–4** (*see* **Note 8**).

3.6 Tissue Collection for RNA Isolation and Gene Silencing Observation

1. *Pgt*-inoculated plants should display very slightly chlorotic spots at 4–5 days post inoculation (dpi) (Fig. 3A). Five days after inoculation with *Pgt*, harvest 100 mg of infected leaf tissue and store at −80 °C for RNA extraction (*see* **Note 9**).
2. The rest of the infected plants are kept in the growth chamber until 12–14 dpi. The infection types (IT) are assessed based on a 0–to–4 rating scale [17] (Fig. 3b) (*see* **Note 10**).

3.7 RNA Isolation from Infected Leaves

1. Grind tissue collected in Subheading 3.6, **step 1** in a tube with liquid nitrogen using a sterilized pestle. Use a new pestle for each sample.
2. Add 1 mL of TRIzol™ Reagent per 100 mg of tissue.
3. Vortex and incubate at room temperature for 5 min.
4. Add 0.2 mL of chloroform, securely cap the tube, mix and incubate at room temperature for 3 min.
5. Centrifuge the samples at 12,000 × *g* at 4 °C for 15 min.

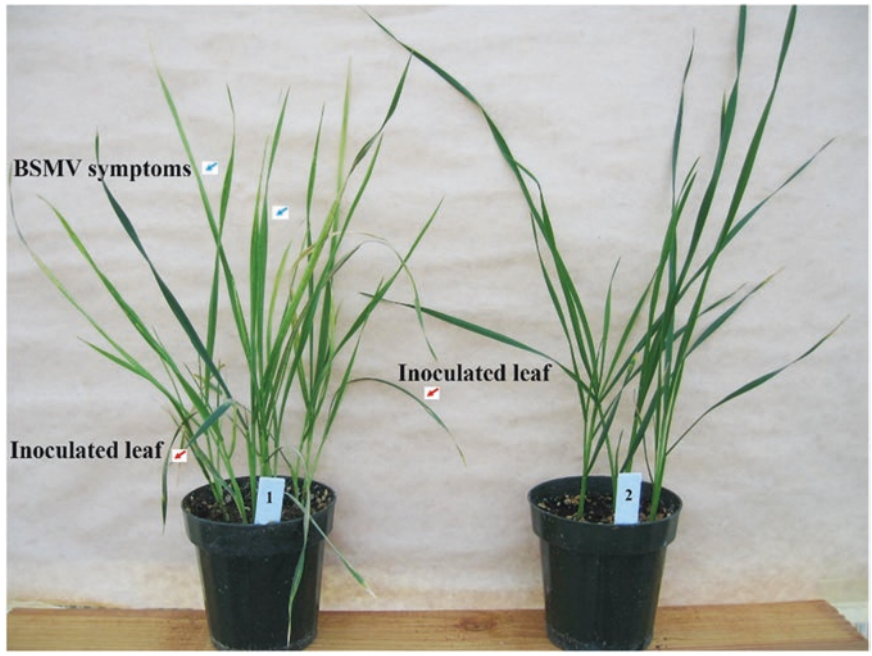


Fig. 2 BSMV symptoms appeared on wheat seedlings of cultivar McNair 701 at 10 days after inoculation with the in vitro transcribed viral RNAs. 1, BSMV-infected wheat; 2, Healthy wheat

6. Transfer the aqueous phase containing the RNA to a new 1.5 mL RNase-free tube.
7. Add 0.5 mL of isopropanol to the aqueous phase. Incubate at room temperature for 10 min.
8. Centrifuge at 12,000 × *g* at 4 °C for 10 min.
9. Discard supernatant and resuspend the pellet in 1 mL of 75% ethanol. Vortex the sample briefly, then centrifuge at 7,500 × *g* at 4 °C for 5 min.
10. Air-dry the RNA pellet for 10 min.
11. Resuspend the pellet in 50 μL of RNase-free water.
12. Measure RNA concentration using NanoDrop Spectrophotometers and store at −80 °C.

3.8 cDNA Synthesis

1. Clean up the above RNA samples via DNase I digestion. Perform DNA digestion in 50 μL reactions containing 10 μg RNA, 5 μL 10× DNase I buffer, 1 μL DNase I (10 units), and nuclease-free H_2O to bring the total volume to 50 μL. Incubate the reaction at 37 °C for 20 min. Stop reaction by adding 2 μL of 0.2 M EDTA (pH 8.0) and heating to 75 °C for 10 min. Then measure RNA concentration using the NanoDrop Spectrophotometer.

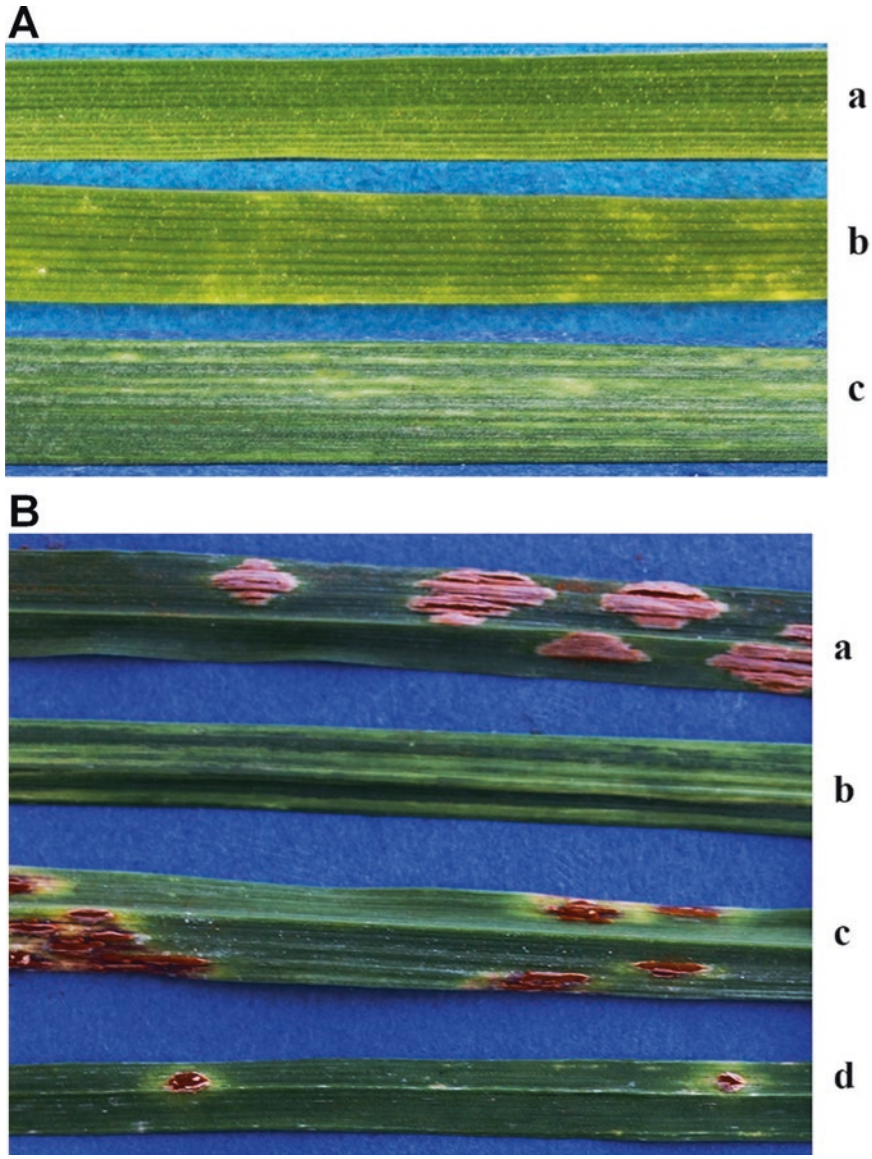


Fig. 3 Reduced virulence of *Pgt7*A on wheat cultivar McNair 701 after infection with BSMV:*Pgt-IaaM*. (A) Wheat leaves challenged with *Pgt7A* at 5 days post inoculation. (**a**) McNair 701 healthy plant; (**b**) McNair 701 challenged with *Pgt7*A; (**c**) McNair 701 inoculated with BSMV and then challenged with *Pgt7*A. (B) Wheat leaves challenged with *Pgt7A* at 12 days post inoculation. (**a**) McNair 701 challenged with *Pgt7*A; (**b**) McNair 701 inoculated with BSMV:MCS (negative control) only; (**c**) McNair 701 inoculated with BSMV:MCS and then challenged with *Pgt7*A; (**d**) McNair 701 inoculated with BSMV:*Pgt-IaaM* and then challenged with *Pgt7*A

2. Perform reverse transcription in 20 μL reactions containing 1 μg RNA template, 4 μL 5× iScript reaction mixes, 1 μL iScript reverse transcriptase, and nuclease-free H_2O to bring the total volume to 20 μL. Incubate the reaction mix at 25 °C for 5 min, followed by 30 min at 42 °C, and 5 min at 85 °C.

3.9 Reverse Transcriptase Quantitative PCR (RT-qPCR):

1. Perform RT-qPCR analysis in 96-well qPCR plate using *Pgt-Iaam* or *Pgt-actin* RT-qPCR primers and iTaq universal SYBR green supermix in 20 μL reactions containing 10 μL 2× Sybr Green supermix, 0.4 μL forward primer (10 mM), 0.4 μL reverse primer (10 mM), 5 μL 10× diluted cDNA, and nuclease-free H_2O to bring the total volume to 20 μL. Use *Pgt-actin* transcript as an internal control to normalize the amount of cDNA in each sample. Each treatment includes three replicates.
2. Seal the 96-well plate with micro-sealing film and spin the plate at 500 × *g* for 2 min to remove bubbles in the reaction mix.
3. Run qPCR at 95°C for 5 min; followed by 45 cycles of 95°C for 20 s, 58°C for 30 s, and 72°C for 30 s; followed by a melting curve program. The dissociation program was 95°C for 1 min and 55°C for 10 s followed by a slow ramp from 55 to 95°C.
4. Analyze fungal transcript abundance. Relative gene expression in *Puccinia* inoculated on seedlings expressing each putative silencing construct is determined by comparing with the transcript abundance in fungus inoculated onto control seedlings for each treatment. If no fungal gene silencing was observed on a given seedling preinfected with virus containing a silencing construct, the expected ratio of gene expression in fungus from that seedling to any of the control seedlings would be equal to 1.0. A t-test is conducted to determine if the ratios of fungal gene expression in the putative silencing seedlings compared with control seedlings are less than 1.0, which would indicate gene silencing occurred (*see* Fig. 4).

4 Notes

1. The compatible wheat cultivars are chosen for *Puccinia* species and isolates.
2. The temperature settings in the dew chamber can differ depending on the *Puccinia* species. 20–22 °C, 15–16 °C, and 13–15 °C are optimal air temperatures for urediniospore germination from stem rust, leaf rust, and stripe rust, respectively.
3. The surface of plant leaves should be uniformly covered with dew. The temperature settings in the growth chamber should be different depending on the *Puccinia* species. The optimal temperature of disease development is 25 °C (day and night) for stem rust and leaf rust, and 20 °C at daytime and 4 °C at night for stripe rust.

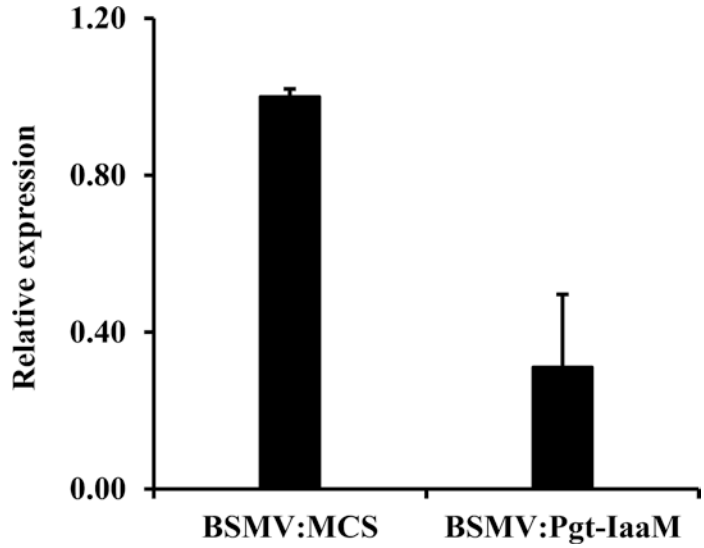


Fig. 4 *Pgt-IaaM* gene expression in *Puccinia* after inoculated on wheat seedlings infected by BSMV:*Pgt-IaaM* compared with the BSMV:MCS control. All data were normalized against the *actin* gene of *Pgt*

4. The targeted gene (e.g., *Pgt-IaaM*) fragment should be designed to avoid sequence similarity with plant sequences and other *Pgt* genes and can be developed by searching the appropriate sequence databases. The size of the cloned fragment should be in the range of 100–600 bp.
5. Check the completion of linearization via running 1 μL of the linearization reactions on 1% w/v agarose gel (linear α, β, and γ migrate at approximately 6–8 kb).
6. Add the 10× reaction buffer after adding the 2×NTP/CAP, linearized DNA template as well as water, and before the addition of the enzyme mix.
7. Do not freeze/thaw transcripts more than two times.
8. Perform *Pgt* inoculations at 10 days after viral infection which is sufficient for new seedling leaves to emerge and to allow initiation of viral symptoms. The virus-infected seedlings that display mild viral disease symptoms, typically on the third and fourth leaves, are chosen for *Pgt* inoculation.
9. The disease development time is different according to the *Puccinia* species. 12–14 days, 8–10 days, and 18–20 days are adequate for disease development for stem rust, leaf rust, and stripe rust, respectively. 100 mg of infected tissue is harvested before sporulation to confirm whether the target gene is silenced.
10. The infection types (IT) for stem rust are typically assessed based on a 0–to–4 rating scale. The infection types (IT) for stripe rust and leaf rust are often assessed on a 0-to-9 rating scale [18].

Acknowledgments

This work was funded by National Institute of Food and Agriculture to S.H. (Grant No. 2010-65108-20568).

References

1. Baulcombe DC (2004) RNA silencing in plants. Nature 431:356–363
2. Cakir C, Gillespie ME, Scofield SR (2010) Rapid determination of gene function by virus-induced gene silencing in wheat and barley. Crop Sci 50:S77–S84
3. Scofield SR, Nelson RS (2009) Resources for virus-induced gene silencing in the grasses. Plant Physiol 149:152–157
4. Nowara D et al (2010) HIGS: host-induced gene silencing in the obligate biotrophic fungal pathogen *Blumeria graminis*. Plant Cell 22(9):3130–3141
5. Tinoco MLP et al (2010) In vivo trans-specific gene silencing in fungal cells by in planta expression of a double-stranded RNA. BMC Biol 31:27
6. Yin C, Jurgenson JE, Hulbert SH (2011) Development of a host-induced RNAi system in the wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici*. Mol Plant-Microbe Interact 24:554–561
7. Yin C et al (2015) Identification of promising host-induced silencing targets among genes preferentially transcribed in haustoria of *Puccinia*. BMC Genomics 16(1):579
8. Yin C, Park JJ, Gang DR, Hulbert SH (2014) Characterization of a tryptophan 2-monooxygenase gene from *Puccinia graminis* f. sp. *tritici* involved in auxin biosynthesis and rust pathogenicity. Mol Plant-Microbe Interact 27:227–235
9. Panwar V, McCallum B, Bakkeren G (2013) Endogenous silencing of *Puccinia triticina* pathogenicity genes through in planta-expressed sequences leads to the suppression of rust diseases on wheat. The Plant J 73:521–532
10. Andrade CM et al (2015) Host-induced gene silencing in the necrotrophic fungal pathogen *Sclerotinia Sclerotiorum*. Plant Pathol 65(4):626–632
11. Koch A et al (2013) Host-induced gene silencing of cytochrome P450 lanosterol C14 ademethylase-encoding genes confers strong resistance to *Fusarium* species. Proc Natl Acad Sci U S A 110:19324–19329
12. Hu ZL et al (2015) Down-regulation of *Fusarium oxysporum* endogenous genes by Host-Delivered RNA interference enhances disease resistance. Front Chem 3:1
13. Govindarajulu M, Epstein L, Wroblewski T, Michelmore RW (2015) Host-induced gene silencing inhibits the biotrophic pathogen causing downy mildew of lettuce. Plant Biotechnol J 13:875–883
14. Koch A, Kogel KH (2014) New wind in the sails: improving the agronomic value of crop plants through RNAi-mediated gene silencing. Plant Biotech J 12:821–831
15. Scofield SR et al (2005) Development of a virus-induced gene-silencing system for hexaploid wheat and its use in functional analysis of the Lr21-mediated leaf rust resistance pathway. Plant Physiol 138:2165–2173
16. Brueggeman R et al (2008) The stem rust resistance gene Rpg5 encodes a protein with nucleotide-binding-site, leucine-rich, and protein kinase domains. Proc Natl Acad Sci U S A 105:14970–14975
17. Stakman EC, Stewart DM, Loegering WQ (1962) Identification of physiologic races of *Puccinia graminis* var. *tritici*. Agric Res Serv E 617. U. S. Dep. Agric, Washington, DC
18. Line RF, Qayoum A (1992) Virulence, aggressiveness, evolution, and distribution of races of Puccinia striiformis (the cause of stripe rust of wheat) in North America, 1968-87. In: Tech. Bull. No. 1788. United States Department of Agriculture, Agricultural Research Service, Washington, DC



Chapter 13

From Short Reads to Chromosome-Scale Genome Assemblies

Kyle Fletcher and Richard Michelmore

Abstract

A high-quality, annotated genome assembly is the foundation for many downstream studies. However, obtaining such an assembly is a complex, reiterative process that requires the assimilation of high-quality data and combines different approaches and data types. While some software packages incorporating multiple steps of genome assembly are commercially available, they may not be flexible enough to be routinely applied to all organisms, particularly to nonmodel species such as pathogenic oomycetes and fungi. If researchers understand and apply the most appropriate, currently available tools for each step, it is possible to customize parameters and optimize results for their organism of study. Based on our experience of de novo assembly and annotation of several oomycete species, this chapter provides a modular workflow from processing of raw reads, to initial assembly generation, through optimization, chromosome-scale scaffolding and annotation, outlining input and output data as well as examples and alternative software used for each step. The accompanying Notes provide background information for each step as well as alternative options. The final result of this workflow could be an annotated, high-quality, validated, chromosome-scale assembly or a draft assembly of sufficient quality to meet specific needs of a project.

Key words Genome assembly, Contig scaffolding, Genome annotation, Next-generation sequencing, Assembly quality, Hi-C, Optical maps

1 Introduction

The first fungal genome sequence was published in 1996 [1], followed by the publication of the first oomycete genome 10 years later [2]. Subsequently the technical difficulties of and time required for genome sequencing have decreased and computational resources and the number of tools able to utilize short- and long-read DNA sequencing technologies have increased. This has led to a rapid expansion in the number of fungal and oomycete genomes published. More than 3000 fungal and 100 oomycete assemblies are now available; however, not all of these assemblies are of high quality and most species have not benefited from all currently available tools.

Wenbo Ma and Thomas Wolpert (eds.), *Plant Pathogenic Fungi and Oomycetes: Methods and Protocols*, Methods in Molecular Biology, vol. 1848, https://doi.org/10.1007/978-1-4939-8724-5_13,
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Many of the early developments in sequencing and genome assembly of eukaryotes were pioneered on haploid fungi. The first chromosome-scale eukaryotic assembly was developed for *Saccharomyces cerevisiae* [1, 3]. Subsequent assemblies of plant pathogenic fungi exploited large-scale sequencing as well as optical and linkage mapping to generate chromosomal pseudomolecules [4–6]. The assembly of *Phytophthora sojae* is one the highest quality genomes of an oomycete; version 1 that was published in 2006 reported 1810 scaffolds [2], and the current version (3.0) has only 83 scaffolds, the fewest publically available for an oomycete. Another high quality oomycete assembly is for *Phytophthora capsici*, where 84% of the assembly has been genetically binned in to 18 linkage groups [7]. Despite dozens of additional oomycete species being assembled and published, they have rarely made it under 1000 scaffolds, likely due to difficulty in assembling.

Recent technological and computational advances have enabled the generation of high-quality, chromosome-scale assemblies and we are entering an era where it is possible to produce high-quality assemblies for newly and previously sequenced species. These advances include massively parallel, high-throughput sequencing [8], long read technologies [9, 10], and several scaffolding technologies as well as the ability to phase haplotypes [11–14]. There has also been a continual expansion of tools available for assembly using a variety of inputs and analysis of the resultant assemblies, examples of which are described below. Given sufficient budget and time, chromosome-scale assemblies are achievable; however, such complete genome assemblies may not always be required. Generation of an initial fragmented, draft assembly may be adequate; therefore not every module described in this chapter is necessary for all genome projects.

Genome assembly and annotation is not a one-time endeavor but rather a long, drawn out, reiterative process that continually assimilates new data and approaches. The use of high-quality data and assessments of quality and progress after each step are critically important. This chapter aims to assist researchers in going from raw reads through assembly to annotation of a genome. Some modules in this workflow are more important than others; multiple paths are possible from raw reads to an assembled, annotated genome. Example tools and syntax are provided at each step based on our experience with assembling genomes of downy mildew species ([15] and many in prep). This field continues to expand rapidly; this chapter should be used as a foundation for incorporating additional tools as they become available. The optimal combination of tools will vary from species to species and project to project. Multiple software programs and approaches should be tried, particularly at the genome assembly step. Alternative tools are provided in Subheading 4.

2 Prerequisites

Genome assembly and annotation typically requires some bioinformatics aptitude, access to high memory compute nodes (*see* Subheading 2.2), software, and input data.

2.1 Bioinformatic Aptitude and Best Practices

This workflow requires use of a terminal in a UNIX environment, including understanding of UNIX commands such as mkdir (make directory), cd (change directory) ls (list directory contents), and awareness of pipes (|) [16, 17]. Software often requires installation; manuals/webpages for the respective programs should be consulted for help and understanding of influential parameters.

Similar workflows can be completed through packages that provide graphical user interfaces (GUIs), e.g., CLCworkbench, Geneious. However, some parameters in these packages are not as customizable, or as flexible and are less evident than in command-line based workflows. Also, for completion of some tasks (e.g. de novo assembly) the purchase of a license is often required.

Best practices include maintaining organized workspaces. Each tool should be run in a separate directory so that results from each step can be easily identified. An example structure is shown in Table 1. For ease of reproducibility workflow managers, such as Snakemake [18] or Jupyter [19], could be used.

2.2 Access to High-Performance Computing

Computational requirements for assembly are dependent on the genome size and sequencing depth. An average desktop/laptop computer is not powerful enough to assemble genomes of oomycetes, whose genomes range from less than 40 megabases (Mb) (e.g., *Albugo laibachii*, *Pythium oligandrum*, and *Pythium periplocum*; [20–24]) to over 200 Mb (e.g., *Phytophthora infestans*; [24, 25]), or fungi with genome sizes ranging from less than 10 Mb (e.g., *Hansenual polymorpha*, *Pneumocystis carinii*; [26, 27]), up to over 3 gigabases (Gb) (e.g., *Jafnea semitosta*; [28]). For smaller genomes (<100 Mb) at moderate sequencing depth (~40×), high-performance computing clusters with 128 GB RAM nodes are usually adequate for assembly; however, larger genomes sequenced to higher depths may require up to 512 GB or 1 Tb of RAM. Parallelization of most tasks on one or more compute nodes will increase the speed of most steps in the workflow; 20 cores/threads are a good starting point. For annotation, less memory is required than for assembly; instead, having more cores or access to a cluster of nodes is beneficial. This allows independent annotation of chromosomes/scaffolds in parallel rather than sequentially.

Table 1
Example of directory structures for organization of preprocessing, assembly, and annotation

/…	Programs/	…			
	Data/	RawReads/			
	Downloads/	NCBI/	Databases/	nt/	nt.*
			ESTs/	…	
	Work/	Assembly/	00PreProcessing/	00FastQC/	
				01Deduplication/	
				02Merging/	
				…	
			01K-merAnalysis/	Heterozygosity/	
				Contamination/	
				K-merOptimization/	
				…	
			02Assembly/	ABySS/	
			03QualityControl/	00BlastFiltering/	
				01AssemblyStatistics/	
				02BUSCO/	
			04Annotation/	00Repeats/	00RepeatModeler/
					01RepeatMasker/
				01Annotation	00InitialMaker/
					01GeneTraining/
					02FinalMaker/

"/" is the root directory, usually a user will have to navigate through other directories to their writable area, indicated by "…"

2.3 Software

Example and alternative software required for each module are provided in Table 2. To run the syntax as written below, executables should be defined in the PATH variable in order to locate the executable. If executable is not in the PATH variable, then the full or relative PATH can be typed to run it. Dependencies required by programs also have to be defined in the PATH for successful execution. To understand more about PATH, refer to introductory UNIX guides such as [16].

Table 2

Tools used in this work-flow with links to the Section/Module and Notes for each process

Section/ Module	Note	Objective	Example tool(s)	Alternative tool(s)
3.1	*4.1*	*Quality control and preprocessing*		
3.1.1	4.1.1	Read quality assessment	FastQC [29]	FASTX toolkit [30]
3.1.2	4.1.2	PCR duplicate removal	SuperDeduper [31]	BBMap (dedupe.sh) [32], FASTX toolkit [30], FastUniq [33]
3.1.3	4.1.3	Read merging	FLASH2 [34, 35]	BBMap (bbmerge.sh) [32], PEAR [36], COPE [37]
3.1.4	4.1.4	Adapter trimming	BBMap (BBduk.sh) [32]	Scythe [38], Trimmomatic [39], AdapterRemoval [40], CutAdapt [41]
3.1.5	4.1.5	Quality trimming	BBMap (BBduk.sh) [32]	FASTX toolkit [30], Trimmomatic [39], Sickle [42]
3.1.6	4.1.6	Removal of contaminant reads	FastqFilter.sh [43], bwa [44], SAMtools [45], pairfq [46]	
	4.1.7	Additional considerations	Mate-Pair read preprocessing, NextClip [47] Normalization, BBMAP (bbnorm.sh) [32] Error correction, Quake [48] or QuorUM [49]	
3.2		*k-mer assessment*		
3.2.1	4.2.1	Estimation of heterozygosity and genome size	JELLYFISH [50], GenomeScope [51]	
3.2.2	4.2.2	Optimization of k-mer size for assembly	Kmergenie [52]	
		Assembly		
3.3	4.3	Assembly	ABySS [53]	DBG2OLC [54], MaSuRCA [55], Velvet [56], SOAPdenovo2 [57], ALLPaths-LG [58], Meraculous-2D [59], Platanus [60], SPAdes [61], DISCOvar [62], IDBA-UD [63]
3.4		*Assembly quality control*		

(continued)

Table 2
(continued)

Section/ Module	Note	Objective	Example tool(s)	Alternative tool(s)
3.4.1	4.4.1	BLAST filtering	AssemblyFilter.sh [64], BLAST [65]	
3.4.2	4.4.2	Assembly statistics	BBMap (stats.sh) [32]	QUAST [66]
3.4.3	4.4.3	Completeness statistics	BUSCO [67]	CEGMA [68]
3.4.4	4.4.4	Paired end coverage of assembly	bwa [44], SAMtools [45], BEDtools2 [69]	
3.4.5	4.4.5	Whole genome alignment	MUMmer [70], SyMap [71]	CoGe [72]
3.4.6	4.4.6	k-mer comparisons; reads vs. assembly	KAT [73]	
3.5	*4.5*	*Assembly optimization*		
3.5.1	4.5.1	Reassembly	*See* alternative tools for assembly	
3.5.2	4.5.2	(Re)scaffolding	SSPACE [74]	SGA [75], SOPRA [76], FragScaff [77], ARCs [78], Links [79], PBJelly [80], SSPACE-LongRead [81]
3.5.3	4.5.3	Gap-filling	GapFiller [82]	GapCloser [57], Sealer [83], PBJelly [80], variantCaller [84]
3.5.4	4.5.4	Merge alternative assemblies	QuickMerge [85]	
3.5.5	4.5.5	Removal of alternative haplotypes	Haplomerger2 [86]	Redundans [87]
3.5.6	4.5.6	Identify and break misjoins	REAPR [88]	
3.5.7	4.5.7	Scaffolding using Hi-C data	HiRise [12]/Proximo [13]	SALSA [89], Lachesis [90]
3.5.8	4.5.8	Scaffolding using optical maps	BioNano proprietary software [11]	
	4.5.9	Validation by genetic analysis		
3.6		*Repeat-masking and annotations*		
3.6.1	4.6.1	Repeat-masking	RepeatModeler [91], RepeatMasker [92]	

(continued)

Table 2 (continued)

Section/ Module	Note	Objective	Example tool(s)	Alternative tool(s)
3.6.2	4.6.2	Annotation	Maker [93], SNAP [94], RepeatMasker [92], BLAST [65], exonerate [95]	JAMg [96], Augustus [97], GeneMark [98], glimmerHMM [99]
		Additional considerations		
	4.7.1	Public release of data		
	4.7.2	Future perspectives		

Example software used in each module and alternatives are listed in columns 4 and 5 respectively

2.4 Input Data

The workflow will be dictated by:

1. Sequencing technology. Currently, most published genome assemblies have been generated from short reads (max. 250 bp) obtained from massively parallel sequencing provided by Illumina platforms. Most current assemblers have been designed to work with this type of data. Single molecule sequencing is being increasingly used, resulting in longer reads with higher error rates compared to Illumina-generated reads. Assemblers are being developed specifically for long-read technologies [100–102] and have recently been bench-marked [103], while others have been adapted to generate hybrid assemblies that incorporate both short and long reads [54, 55].
2. Types of library. Illumina platforms may provide short insert, paired-end fragments (typically up to 600 bp apart), mate-pair fragments (typically spaced up to 20 kb apart in the genome), or linked-read synthetic molecules [14]. Long read technologies require large genomic fragments (at least 40 kb and preferably greater). The type of library sequenced will produce reads with different information and types of errors.
3. Nucleotide template. Optimally, genome assemblies of eukaryotes require several types of reads from high-quality genomic DNA. Accurate gene annotation greatly benefits from a transcriptome assembly generated from RNAseq data. Knowledge of which genes in the genome are transcribed helps gene predictions and significantly improves determination of intron–exon boundaries.

In addition, to complete **steps 3.1.4** (Adapter removal), **3.1.6** (Contaminant removal), and **3.4.1** (Filtering by BLAST) of the workflow the following files are required:

1. adapters.fa. A file of Illumina adapter sequences which may be present in the reads in order to allow their removal (Module 3.1.4).
2. If the organism under study does not come from an axenic culture, a nucleotide database of potential contaminants and easily identified reference assemblies can be used for read filtering (Module 3.1.6). Filtering may reduce the impact of reads that originate from contaminating organisms and would otherwise complicate the assembly process and increase computational resources required.
3. A recent, local download of NCBI nt database and file of accession labels of closely related organisms present in this download. Postassembly filtering (Module 3.4.1) facilitates identification of similar sequences in closely related species as well as sequences from contaminating organisms. Even if the organism comes from an axenic culture, it is good practice to include this step to confirm that no additional organisms are present.

3 Workflow

The sections that follow are ones that the authors currently use for read preprocessing, genome assembly, and annotation (Fig. 1). Each section is split into multiple modules and quality control modules should be implemented throughout (Modules 3.1.1 [reads], 3.4.2–3.4.6 [assemblies]). When appropriate, alternative tools are provided in Table 2 and the Notes sections.

Simple syntax is provided for each module or a link to a Github repository containing shell scripts encasing multiple commands where the syntax is more complicated. Syntax is detailed for the benefit of the reader. Refer to the corresponding Notes for further discussion of what each module aims to accomplish (Table 2). Inexperienced researchers should refer to the "Notes" prior to using the Module syntax. The directory structure is not considered in this syntax; the output directory can usually be specified when running commands. Syntax is supplemented with "\ <carriage return>" for ease of reading. This indicates that the next line is still part of the previous command. When entering a command these features are not required if typed on a single line.

3.1 Quality Control and Preprocessing (See Note 4.1*)*

Reads should be assessed for quality and if necessary trimmed and filtered prior to use in assembly. Not all modules are necessary or guaranteed to have a positive impact on the final assembly. A routine first pass process is presented in Fig. 1. Additional modules may be included as necessary. Therefore, a diverse selection of preprocessing modules are provided with example syntax below.

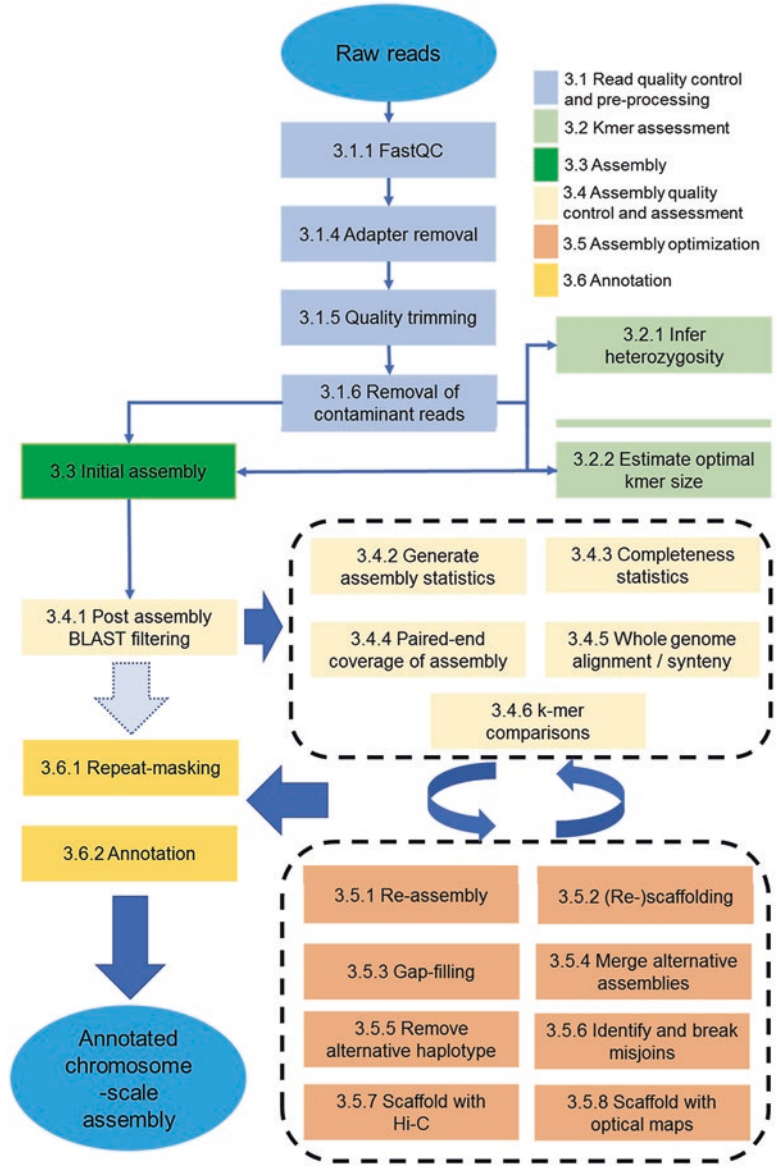


Fig. 1 Routine workflow for read preprocessing, assembly, quality control, and annotation

3.1.1 Read Quality Assessment (Strongly Recommended)

Objective:

To observe several quality statistics of read sets, both preprocessing and postprocessing (*see* **Note 4.1.1**).

Tools:

FastQC v0.11.2 [29]

Example syntax:

```
fastqc MySequence.fastq.gz
```

Output:

An html file for each read file input that may be opened in any web browser. Provides summary of read file, including average quality per base position, sequence GC content, base GC content, and duplication levels of the first 200,000 reads. Warnings are highlighted on the left summary bar (Fig. 2).

3.1.2 PCR Duplicate Removal (Optional; Recommended for Paired-End Reads Only)

Objective:

To remove PCR duplicate read-pairs (i.e., those that sequenced identical fragments) (*see* **Note 4.1.2**).

Tools:

SuperDeduper 2.0 [31]

Example syntax:

```
super_deduper -s 10 -l 25 -1 MyRawSeq.1.fastq -2
MyRawSeq.2.fastq -p MyDedupedSeq
```

Output:

Two files, corresponding to the two read sets. Both prefixed "MyDedupedSeq", suffixed "_nodup_PE1.fastq" and "_nodup_PE2.fastq".

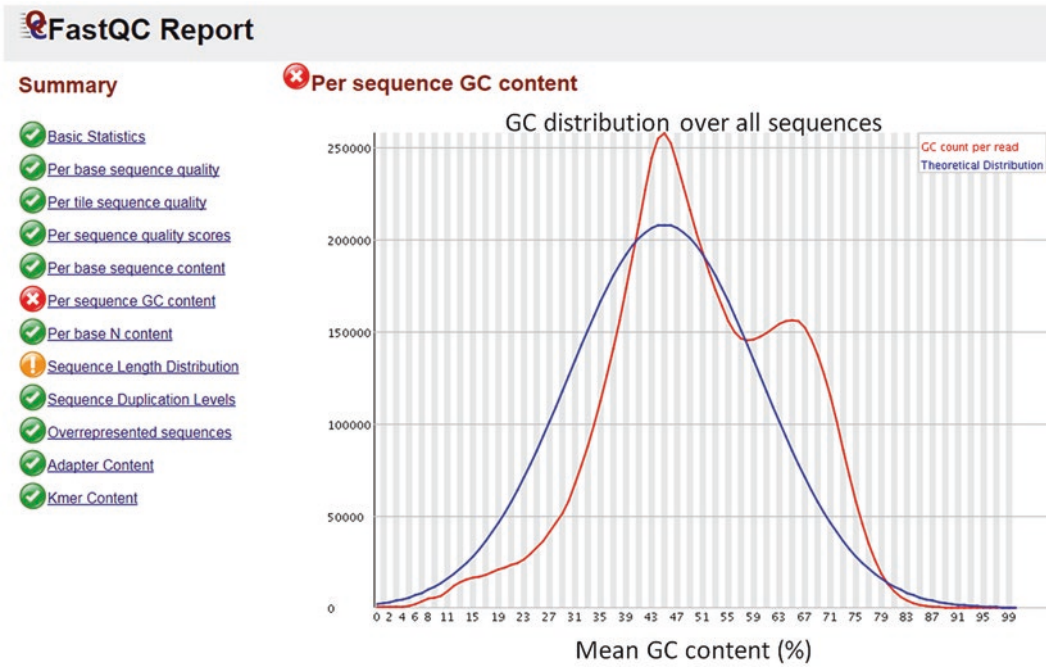


Fig. 2 Example output of FastQC report, GC content page [29]. The blue curve represents the expected theoretical distribution. The red curve represents the read set analyzed; the clear bimodal pattern indicates the presence of a second organism with a different GC content. Other available reports are listed at the left

3.1.3 Read Merging (Optional)

Objective:

To identify read pairs that overlap with one another and merge them into a single longer fragment (*see* **Note 4.1.3**).

Tools:

Flash2 commit c41a82e [34] (modified version: https://github.com/dstreett/FLASH2)

Example syntax:

```
flash2 -d FlashOut -o MyFlashedSeq -r 150 -f 350 -s 35
MyDedupedSeq_nodup_PE1.fastq \
MyDedupedSeq_nodup_PE2.fastq
# -r is the read length
# -f is the fragment size
# -s is the standard deviation length of fragment size
```

Output:

In a new directory, FlashOut will be three fastq files.

The reads that do not overlap:

MyFlashedSeq.notCombined_1.fastq

MyFlashedSeq.notCombined_2.fastq

The reads that did overlap and were merged:

MyFlashedSeq.extendedFrags.fastq

3.1.4 Adapter Trimming (Recommended)

Objective:

To remove Illumina sequencing adapters from reads (*see* **Note 4.1.4**).

Tools:

BBMap v36.86 (bbduk.sh) [32]

Example syntax:

```
bbduk.sh in1=MyFlashedSeq.notCombined_1.fastq \
in2=MyFlashedSeq.notCombined_2.fastq \
out1=MySequenceNoAdapters.1.fastq out2=MySequenceNoAdapters.2.f
astq \
ref=adapters.fa ktrim=r k=21 mink=11
bbduk.sh in1=MyFlashedSeq.notCombined_1.fastq \ in=MyFlashedSeq.
extendedFrags.fastq out=MySequenceNoAdapters.SE.fastq \
ref=adapters.fa ktrim=r k=21 mink=11
```

Output:

Adapter filtered, paired-end read files are output, named as specified on the command line. The above syntax produces the files:

MySequenceNoAdapters.1.fastq

MySequenceNoAdapters.2.fastq

MySequenceNoAdapters.SE.fastq

3.1.5 Quality Trimming (Recommended)

Objective:

To remove low quality regions of sequencing reads and reads without high-quality runs of sequence (*see* **Note 4.1.5**).

Tools:

BBMap v36.86 (bbduk.sh) [32]

Example syntax:

```
bbduk.sh in1=MySequenceNoAdapters.1.fastq in2=MySequenceNoAdapt
ers.2.fastq \
out1=MyTrimmedSequence.1.fastq out2=MyTrimmedSequence.2.fastq \
qtrim=rl trimq=20.
bbduk.sh in1=MySequenceNoAdapters.SE.fastq
out=MyTrimmedSequence.SE.fastq \
qtrim=rl trimq=20.
```

Output:

Trimmed read files are output, named as specified on the command line. The above syntax produces the files:

MyTrimmedSequence.1.fastq
MyTrimmedSequence.2.fastq
MyTrimmedSequence.SE.fastq

3.1.6 Removal of Contaminant Reads (Optional)

Objective:

To remove reads originating from contaminant organisms, preventing cross species chimeric contigs (*see* **Note 4.1.6**).

Tools:

FastqFilter.sh v0.1 [43]
bwa v0.7.12 + [44]
SAMtools v1.3 + [45]
pairfq v0.15 [46]

Other Inputs:

Contaminant database (*see* **Note 4.1.6** or Subheading 2.4)

String by which reference contaminants can easily be identified (*see* **Note 4.1.6** or Subheading 2.4)

Example syntax:

```
./FastqFilter.sh -d MyCustomDB.fa -1 MyTrimmedSequence.1.fastq \
-2 MyTrimmedSequence.2.fastq -M MyTrimmedSequence.SE.fastq \
-o MyReads -s TargetSeqID
```

Output:

Four read files are placed in the directory named MyReads-RFA:

Reads that could be paired after contaminant filtering:
MyReads.RFA.1.fq
MyReads.RFA.2.fq
Reads that could not be paired after contaminant filtering:
MyReads.RFA.NP.fq
Single end reads after contaminant filtering.
MyReads.RFA.SE.fq

The number of files output can be expanded, to produce the above files, but split on whether they mapped to a target organism or not, by adding the flag "-X".

3.2 K-mer Assessment (See Note 4.2*)*

Analyzing the k-mer content of the reads can reveal information about the heterozygosity and genome size as well as providing possible optimized settings for an assembler.

3.2.1 Estimation of Heterozygosity and Genome Size

Objective:

To produce a plot indicating the frequency of k-mers in the read set and calculate genome size (*see* **Note 4.2.1**).

Expectation:

A peak which is equal to 1× representing homozygous k-mers and a second peak equal to 0.5× representing heterozygous k-mers. Can be complicated by presence of multiple species and higher ploidies.

Tools:

JELLYFISH v2.2.3 [50]

GenomeScope (URL tool, requiring jellyfish histogram; [51])

Syntax:

```
jellyfish count -m 21 -s 1G -t 16 -C MyReads*.fq -o
ReadsForAssembly.jf21
jellyfish histo -f ReadsForAssembly.jf21 > ReadsForAssembly.
jf21.histo
# ReadsForAssembly.jf21.histo can then be plotted or provided
to GenomeScope
# (via URL). The kmer length should be set to how the -m flag in
JELLYFISH and max kmer # coverage altered for organisms with
high numbers of repeats
```

Output:

A JELLYFISH histogram file that can be ran through GenomeScope, producing outputs similar to Fig. 3.

3.2.2 Optimization of K-mer Size for Assembly

Objective:

Obtain an estimate of the best k-mer size to run assemblies at (*see* **Note 4.2.2**).

Tools:

kmergenie v1.7016 [52]

Example syntax:

```
#list all read files in a single file, called ReadNames below
ls MyReads*.fq > ReadNames
kmergenie ReadNames -k 149 -l 29 -t 12 -o AssemblyReads.kmerg-
enie
```

Output:

AssemblyReads.kmergenie_report.html—An html report summarizing the histograms generated at different sizes, with estimates of genome size and optimal k-mer size for assembly.

3.3 Assembly

Objective:

To produce a genome assembly from raw reads (*see* **Note 4.3**).

Tools:

ABySS v1.9.0 [53]

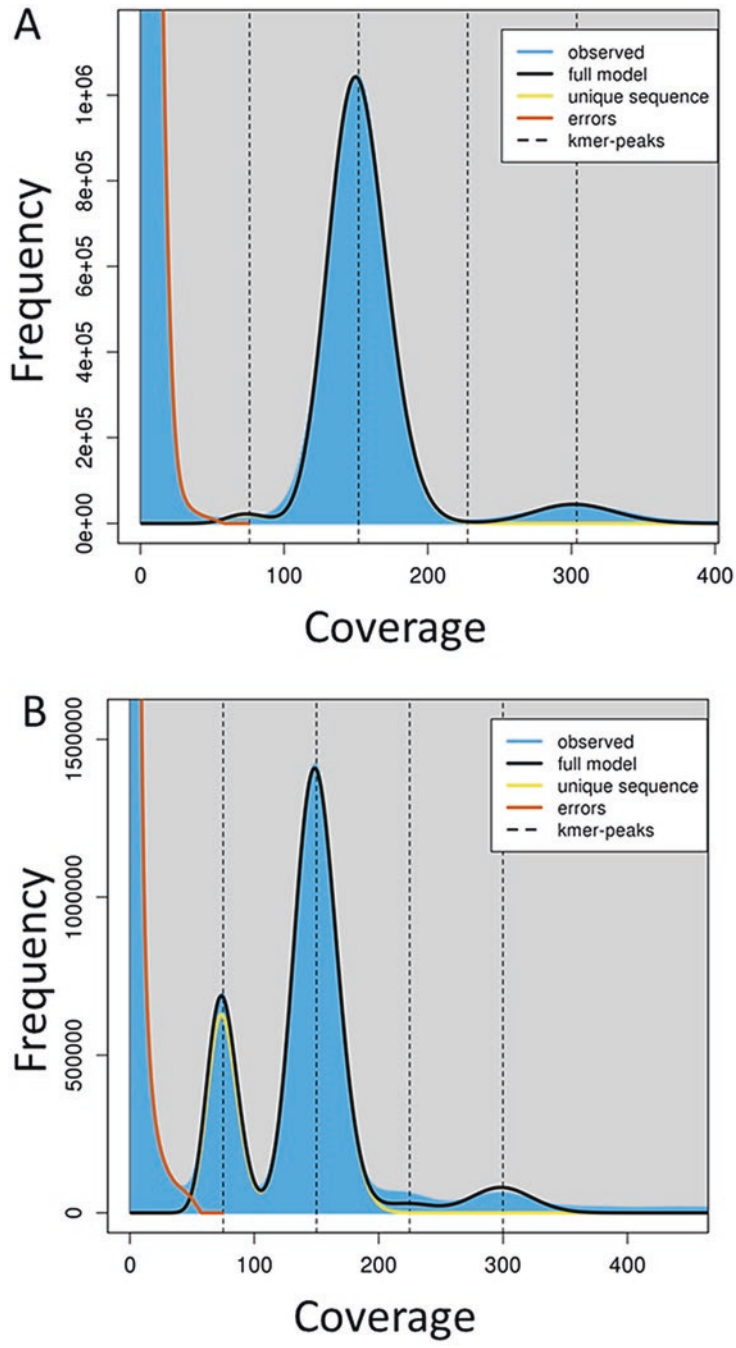


Fig. 3 Example outputs of GenomeScope [51] generated from Jellyfish 21-mer histograms [50] showing estimates of genome size and heterozygosity. (**a**) Reads of *Hyaloperonospora arabidopsidis* SRR3254744. The genome size of *H. arabidopsidis* was estimated to be 65 Mb with 0.033% heterozygosity. This low level of heterozygosity is indicated by the single peak present in the plot. (**b**) Reads of *Phytophthora infestans* ERR1990236. The genome size of *P. infestans* was estimated to be 105 Mb (smaller than that reported for isolate T30-4 [25]; *see* **Note 4.2.1**) with 0.695% heterozygosity. This high level of heterozygosity is indicated by two distinguishable peaks in the plot representing first the heterozygous then the homozygous k-mers. K-mers from repeated sequences are evident beyond the homozygous peaks in both species

Example syntax:

```
abyss-pe k=[k-mer-size] np=16 name=MySpecies lib="Paired" \
Paired='MyReads.RFA.1.fq MyReads.RFA.2.fq" \
se='MyReads.RFA.NP.fq MyReads.RFA.SE.fq'
# suggested k-mer size may be inferred by kmergenie in Module
3.2.2
# np is the number of processors/threads/cores to be used
# name is the output prefix
# lib defines paired end libraries to be processed, we call it
Paired
# Paired is then provided with the two paired end files, output
from Module 3.1.6
# se are all the single end reads, output from Module 3.1.6
#with mate pairs add a new library i.e. 'mp=Mates' Mates=…
```

Output:

An assembly will be produced in file MySpecies-scaffolds.fa

3.4 Assembly Quality Control (See Note 4.4)

After obtaining an assembly, additional quality control is required to further filter contaminants and determine the quality of the assembly.

3.4.1 BLAST Filtering

Objective:

Survey the output assembly file for contaminants and produce a filtered assembly (*see* **Note 4.4.1**).

Tools:

AssemblyFilter.sh v0.1 [64]

BLAST v2.6.0 [65]

SAMtools v1.3 [45]

Other requirements:

local download of blast nt database (available for download from NCBI; *see* **Note 4.4.1** or Subheading 2.4).

presorted list of accessions to blast for (could be obtained by querying NCBI Entrez; *see* **Note 4.4.1** or Subheading 2.4).

Example syntax:

```
./AssemblyFilter.sh -i MySpecies-scaffolds.fa -p MySpecies -d
NCBI.nt -t 8 \
-l AccessionsOfInterest.
```

Output:

Filtered assembly named MySpecies.filt.fasta

3.4.2 Assembly Statistics

Objective:

To obtain basic and commonly used statistics about the assembly produced (*see* **Note 4.4.2**).

Tools:

BBMap v36.86 [32]

Example syntax:

```
stats.sh MySpecies.filt.fasta
```

Output:

Single page report, printed to terminal screen including:

- Nucleotide and GC composition.
- Contig and scaffold length and number 50 and 90 (L50/N50/L90/N90).
- Total contig and scaffold count and size.
- Gap percentage.
- Maximum contig and scaffold length.
- Number of scaffolds larger than 50 Kb and percentage of the genome this includes.
- A table summarizing scaffold and contig length above different minimum scaffold length cutoffs.

3.4.3 Completeness Statistics

Objective:

To obtain completeness statistics of assembly, a useful comparative measure. BUSCO has over 40 libraries which may be queried, the most relevant will depend on the organism under study, please refer to the **Note 4.4.3** for specific recommendations for libraries.

Tools:

BUSCO v2.0 [67]

Example syntax:

```
BUSCO.py -i MySpecies.filt.fasta -o MySpecies.BUSCO -l BUSCO.library -m genome
```

Output:

In the directory "run_Reference.BUSCO" is a report outlining the results including percentage of genes identified as single copy, duplicated, fragmented, and missing. Additional supplementary information such as amino acid and nucleotide sequences of the genes is also provided.

3.4.4 Paired-End Coverage of Assembly

Objective:

Produce and plot a coverage histogram indicating the quality of the assembly (*see* **Note 4.4.4**).

Tools:

bwa v0.7.12 [44]

SAMtools v0.1.18 [45]

BEDtools2 v2.25.0 [69]

R v3.0.2 [104]

R package ggplot2 v2.2.1 [105]

Example syntax:

```
bwa index MySpecies.filt.fasta
bwa mem MySpecies.filt.fasta Reads1.fastq Reads2.fastq | \
samtools view -bT YourFilteredReference - | \
```

```
samtools sort - -o CovCalc.sorted
samtools index CovCalc.sorted.bam
bedtools genomecov -g MySpecies.filt.fasta -ibam CovCalc.sorted.
bam | \
grep 'genome' - > CovCalc.hist
#call R
R
library("ggplot2") # call install.packages("ggplot2") if not
already installed
CovTable <- read.table("CovCalc.hist")
names(CovTable)[c (2,5)] <- c ("coverage", " frequency")
jpeg("CovCalc.jpg")
ggplot(a, aes(x=coverage, y=frequency))+geom_line(colour="blue")
```

Output:

A line graph showing the read depth against the assembly.

3.4.5 Whole Genome Alignment

Objective 1:

To assess synteny between a high-quality reference assembly in order to test the quality of the assembly under study (*see* **Note 4.4.5**).

Tools 1:

SyMap v4.2 [71]

Methods 1 (SyMap):

Reference and test assemblies are imported to SyMap through the GUI interface [71]. Minimum cutoffs of scaffold sizes are declared (e.g., 50 Kb). The test assembly may initially be aligned against the reference. SyMap then uses MUMmer [70] or BLAT [106] to perform whole genome alignment and results are visualized upon completion.

Output 1:

A Circos-like plot indicating the extent of synteny between the two assemblies (Fig. 4a).

Objective 2:

To assess synteny on a finer scale, either between species, or between assembly parameters (*see* **Note 4.4.5**).

Tools 2:

MUMmer v3.23 [70]

Example syntax:

```
promer TestSpecies.fa MySpecies.filt.fasta -p PromerOut
# -p prefix
delta-filter -l 10000 -i 95 PromerOut.delta > PromerOut.df.delta
# -l length of alignment
# -i identity between alignments
mummerplot -l -p PromerOut -t png PromerOut.df.delta
# -l order along vertical
# -p prefix
# -t terminal (output format)
```

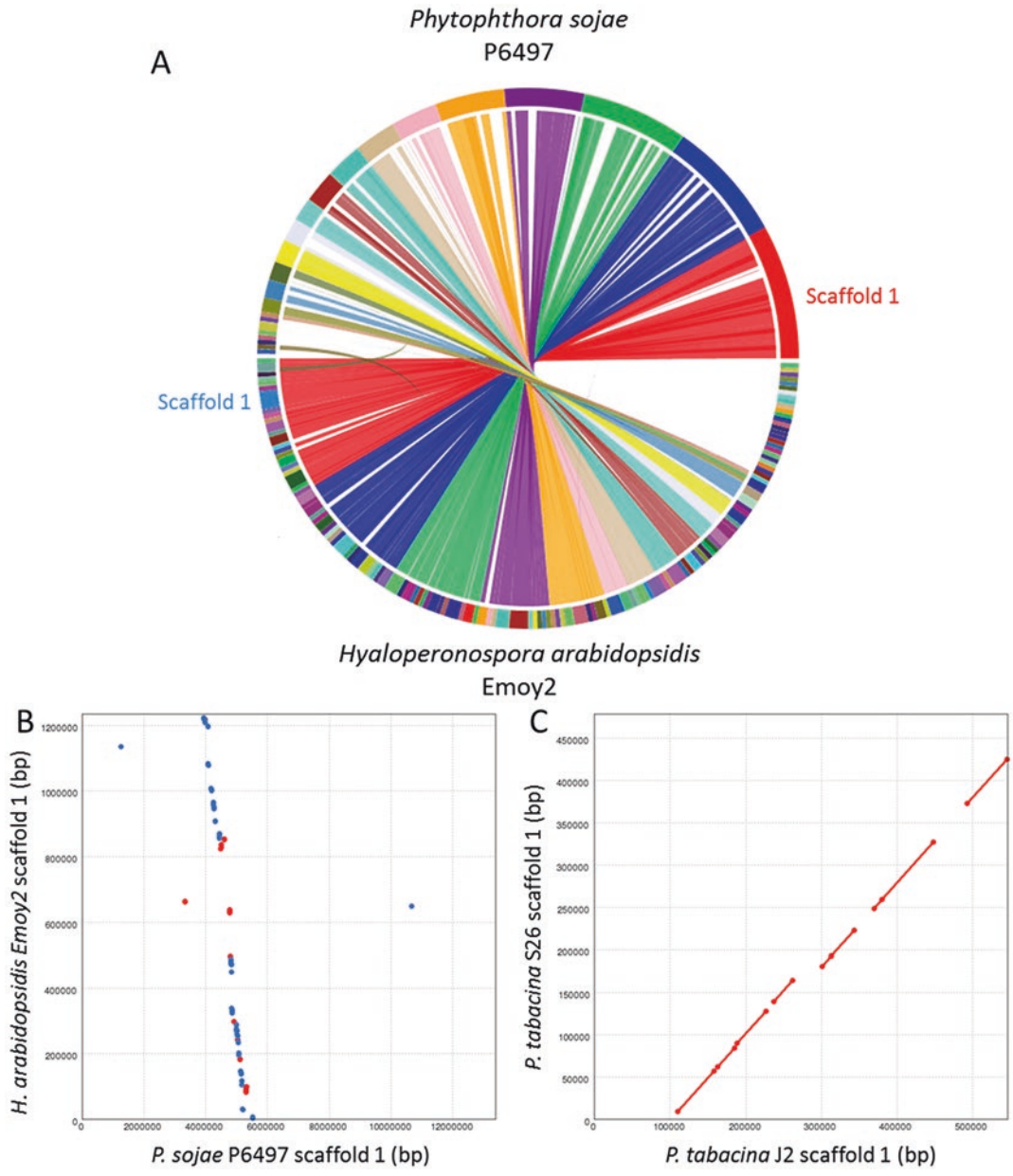


Fig. 4 Example outputs of two software packages for genome alignment that can be used to test for synteny as a measure of assembly quality. (**a**) SyMap output [71] comparing assemblies of *P. sojae* P6497 [2] and *H. arabidopsidis* Emoy2 [107]. The assemblies were aligned with PROmer through the SyMap GUI. The scaffolds of *P. sojae* were ordered by size and scaffolds of *H. arabidopsidis* were ordered based on anchors to the *P. sojae* assembly. This demonstrates that although the assembly of *H. arabidopsidis* is highly fragmented, the assembled scaffolds are of high quality because they are syntenic with large regions of the *P. sojae* genome. (**b**) The *P. sojae* supercontig 1 and *H. arabidopsidis* scaffold 1 highlighted in (**a**) independently aligned using PROmer within MUMmer [70]. Dots in this plot display regions with >70% identity at the translated nucleotide level. Blue dots indicate the alignment is reversed between the scaffolds, red that it is in the same direction in both. This demonstrates that scaffold 1 of *H. arabidopsidis* is largely collinear with part of the largest scaffold of *P. sojae*. (**c**) A NUCmer alignment [70] of the largest scaffolds of two isolates of *P. tabacina* J2 and S26 [15]. Red diagonals demonstrate alignments >97% at the nucleotide level between the scaffolds

Output:

PromerOut.png—A dotplot of the two assemblies showing regions of similarity (Fig. 4b, c).

PromerOut.gp—A gnuplot file which can be edited to change axis titles and font sizes.

3.4.6 K–mer Comparisons; Reads vs Assembly

Objective:

To compare the k-mer populations between raw reads and the assembly generated (*see* **Note 4.4.6**).

Tools:

KAT v2.3.4 [73]

Example syntax:

```
kat comp -t 12 -o ReadsxAssembly 'ReadsForAssembly?.fq'
MySpecies.filt.fasta
```

Output:

ReadsxAssembly.spectra-cn.png—An image depicting a k-mer histogram of the read set and the multiplicity of the k-mers represented in the queried assembly (Fig. 5).

3.5 Assembly Optimization (*See* Note 4.5)

Quality control of the initial assembly may find it not fit for the intended purpose. For example, the assembly may:

- Be highly fragmented (low N_{50}, high scaffold count or poor synteny with closely related, reference assemblies; Modules 3.4.2 and 3.4.5).
- Be highly duplicated (revealed by BUSCO, read mapping, and k-mer analysis; Modules 3.4.3, 3.4.4, and 3.4.6).
- Have suboptimal k-mer inclusion levels (indicated through k-mer analysis; Module 3.4.6).
- Have low completeness (low BUSCO score; Module 3.4.3).

This section provides strategies to address these problems and to utilize additional data, including progressing towards chromosome-scale scaffolds (Table 3). The use of this section should be tailored to the goals of the project, applying all modules below may not be needed, and could be costly.

Example syntax runs with the output file produced by assembly filtering (Module 3.4.1) as the input. This will not typically be the case; instead modules will be applied sequentially and feed into one another. Additionally, modules do not necessarily have to be run in the order outlined below and may be run multiple times depending on the goals of the project. For instance, it is often beneficial to run the scaffolding Module (3.5.2) before and after haplotype removal (Module 3.5.5). Therefore, the syntax may need to be modified to accommodate the output file from the preceding Module used. Quality should be assessed regularly (Modules 3.4.2–3.4.6) on intermediate assemblies generated by each module as well as on the final assembly, before moving onto Subheading 3.6.

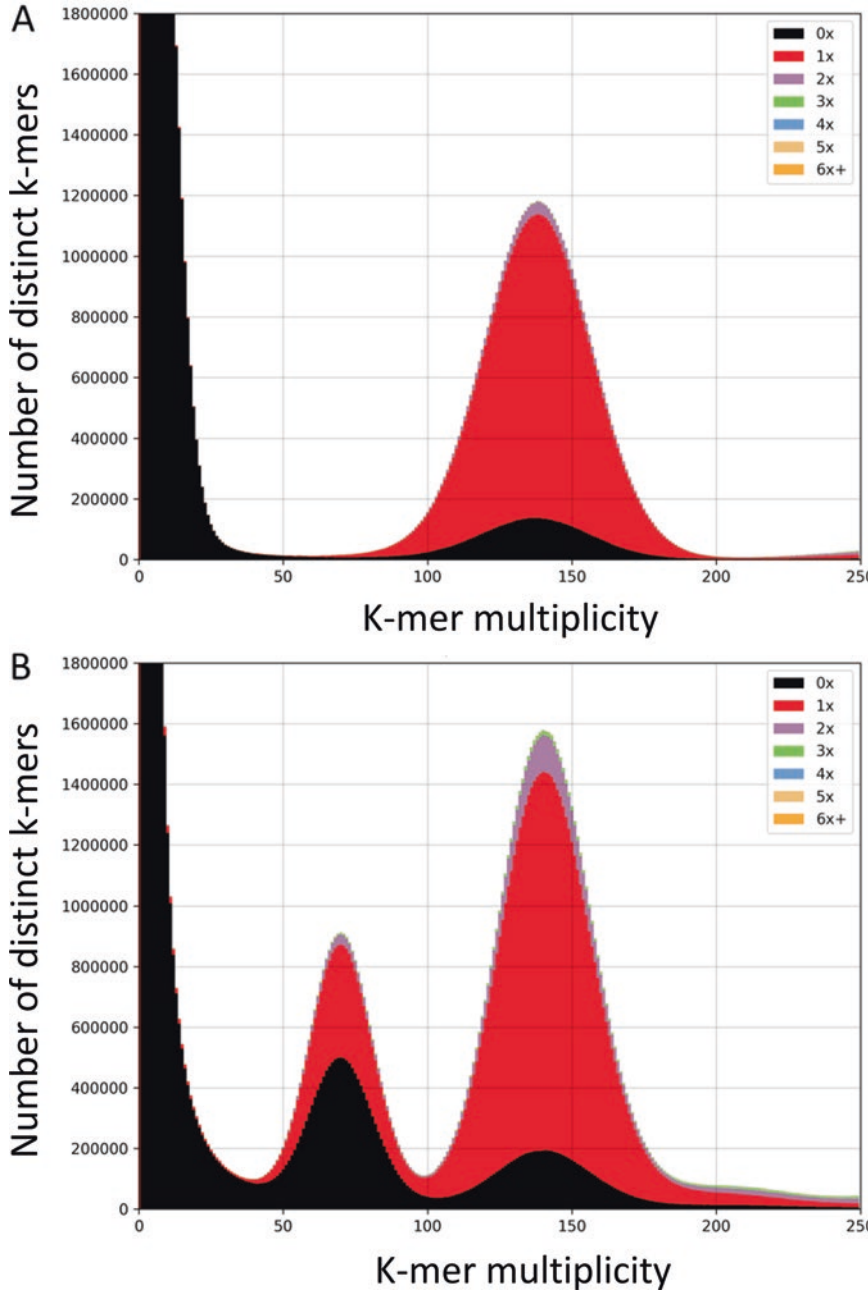


Fig. 5 Example KAT spectra-cn plots of hashes [73] generated to determine what proportion of the reads are represented in the assembly and consequently how complete it is. K-mers present in the reads but absent from the assembly are represented in black; single copy k-mers are red and overrepresented k-mers are purple (2×) or green (3×). (**a**) Reads of *H. arabidopsidis* SRR3254744 and the draft assembly of isolate Emoy2 [107]. The majority of k-mers are present only once as expected for a homozygous species and few k-mers are missing indicating a high-quality assembly. (**b**) Reads of *P. infestans* ERR1990236 and draft assembly of isolate T30-4 [25]. The first peak shows the k-mers representing heterozygous sites in genome; half of the k-mers are missing as expected in a good consensus assembly of a heterozygous diploid species. The second peak shows k-mers representing homozygous regions of the genome; the majority of k-mers are represented once and only a few k-mers are duplicated or absent

Table 3

Additional technologies and strategies for incorporation into an assembly

Technology	Description	Output	Use
Mate-pairs	Terminal sequences from large circularized genomic fragments that are usually under 20 kb	Pairs of Illumina reads	De novo assembly (most assemblers) Scaffolding [74]
Fosmid ends	Ends of cloned 35–40 kb genomic fragments	Pairs of Sanger sequences Pairs of Illumina reads	Scaffolding [74]
Bacterial artificial chromosome ends	Ends of cloned ~ 100 kb + genomic fragments	Pairs of Sanger sequences	Scaffolding [74]
Linked molecule reads	Illumina with 10× Genomics Chromium preparation. Genomic fragments isolated and tagged with unique barcodes through emulsion droplet PCR, sequenced on an Illumina platform and algorithmically reconstituted	Paired-end Illumina reads; Chromium barcodes contained within sequence of each read and removed with specialized assembly software [14]	Independent de novo assembly [14] Assembly merging [85] Scaffolding [77, 78] Phasing [108]
Single molecule reads	High molecular weight (HMW) DNA isolated and sequenced using Pacific Biosciences or Oxford Nanopore platforms	File(s) of long length reads and additional supplemental data. Fastq files of reads obtained with sequence providers proprietary software [9, 10]	Independent de novo assembly [100–102] Hybrid de novo assembly [54, 55] Assembly merging [85] Scaffolding [79, 80] Gap Filling [80] Phasing [108]
Hi-C	In vivo chromatin capture, followed by restriction digest, ligation and sequencing of linked, spatially distinct fragments	Pairs of Illumina reads whose frequency reflects their chromosomal separation	High level scaffolding [89, 90, 109, 110]
Optical map	HMW DNA isolated, nicked, and visualized; physical map created describing distances between nick sites	Optical map to which assembly can be aligned and orientated to	High level scaffolding [111–113]
Genetic map	Generation and sequencing of a mapping population to determine linkage between SNP markers	Genetic map to which scaffolds can be aligned and orientated against	Independent assembly validation. Identification of chimeric scaffolds. Ordering and orientation of scaffolds into chromosomal pseudomolecules [114]

3.5.1 Reassembly

Objective:

To reassemble the same data independently using an alternative assembler. ABySS is used initially because it is rapid and has low memory requirements compared to other assemblers and accommodates larger k-mers, if configured to do so during installation. However, for some oomycetes, MaSuRCA provided more contiguous assemblies than ABySS, although it did not always collapse haplotypes of highly heterozygous species efficiently (Fletcher and Michelmore, unpublished data). MaSuRCA [55] was reported to be the best assembler of a panel, which included ABySS [53], Velvet [56], SOAPdenovo2 [57], and AllPaths-LG [58], when applied to *Sclerospora graminicola* [115]. Assemblers designed for highly heterozygous genomes include Meraculous-2D [59] and Platanus [60]. It may also be necessary to reassemble including additional data (e.g., mate pairs or long single-fragment reads) as it becomes available (*see* **Note 4.5.1**).

Outputs:

A new assembly, which should be quality controlled as Modules 3.4.1–3.4.6.

3.5.2 (Re)scaffolding

Objective:

To improve contiguity of the draft assembly by including long distance paired information (from for example mate-pair reads, fosmid, and BAC ends). This can be performed using data not yet included in the assembly or reusing already incorporated data (*see* **Note 4.5.2**).

Tools:

SSPACE v3.0 [74]

Example syntax:

```
#Format read library file, a tab delimited file containing:
#libname, aligner, Read1, Read2, FragmentSize, Standard devia-
tion, library format
#i.e.
nano MPLib.tsv
MP bowtie MatePair.1.fq MatePair2.fq 4000 0.1 RF
#Full paths to reads must be specified if not in the current
working directory
SSPACE_Standard_v3.0.pl -l MPLib.tsv -s MySpecies.filt.fasta \
-b ScaffoldedAssembly
#Use flag T to specify number of threads.
```

Output:

ScaffoldedAssembly/ScaffoldedAssembly.final.scaffolds.fasta—Assembly file with new joins.

3.5.3 Gap-Filling

Objective:

To use paired reads (paired-end recommended) in order to fill gaps generated during assembling or scaffolding with sequence (*see* **Note 4.5.3**)

Tools:
GapFiller v1.0 [82]
Example syntax:

```
#Generate library file (named PELib.tsv in the syntax) as above,
this time containing paired-end reads. The format should be
changed to FR.
GapFiller.pl -l PELib.tsv -s MySpecies.filt.fasta -b
GapFillerAssembly -r 0.5
#Again -T will control how many threads are utilized
```

Output:
GapFillerAssembly/GapFillerAssembly.gapfilled.final.fasta—GapFilled Assembly

3.5.4 Merge Alternative Assemblies

Objective:
When two or more assemblies have been produced, align and merge together to form a consensus assembly (*see* **Note 4.5.4**).
Tools:
Quickmerge commit e4ea490 [85]
Example syntax:

```
merge_wrapper.py MySpecies.filt.fasta MySecondAssembly.fasta
```

Output:
merged.fasta—a new assembly file containing the uni

3.5.5 Removal of Alternative Haplotypes

Objective:
To remove alternative haplotypes where divergent regions have had assembled independently (*see* **Note 4.5.5**).
Tools:
Haplomerger2 v3.4 [86]
runHM2.sh v0.1 [116]
WinMasker v1.0.0 [117]
Example syntax:

```
#SoftMask assembly, RepeatMasker (Module 3.6.1) with the flag
-xsmall may be used as
# an alternative.
windowmasker -checkdup true -mk_counts -mem 6500 -in MySpecies.
filt.fasta \
-out library.ustat
# This will produce a repeat library called library.ustat re-
quired for the next step.
windowmasker -in MySpecies.filt.fasta -out MyMaskedReference.fa
\
-outfmt fasta -dust true -ustat library.ustat
# This will produce a soft masked version of the assembly named
# MyMaskedReference.fa .
./runHM2.sh -f MyMaskedReference.fa -t [threads]
```

Output:
MyMaskedReference_mp_ref_rt.fa.gz—Assembly file output by Haplomerger2, after misjoin processing (mp), haplotype

collapsing, and tandem removal (rt). It is often effective to rescaffold after haplotype collapsing.

MyMaskedReference_mp_alt.fa.gz—Alternative assembly file output processed up to step hm.batchB5 of Haplomerger2. Tandem removal has not been completed on this file, but can be done by separately running steps hm.batchD1-3.

3.5.6 Identify and Break Misjoins

Objective:

To use high depth, long fragment data identify misjoins and split chimeric scaffolds at these points (*see* **Note 4.5.6**).

Tools:

REAPR v1.0.17 [88]

FASTX toolkit v0.0.13 [30]

Example syntax:

```
#Check input assembly for naming conflicts that will cause program
failure.
reapr facheck MySpecies.filt.fasta
#Mate pair reads may need to be reverse complemented. Only 'in-
nies' expected.
fastx_reverse_complement -i MatePair.1.fq -o MatePair.RC.1.fq
fastx_reverse_complement -i MatePair.2.fq -o MatePair.RC.2.fq
reapr smaltmap MySpecies.filt.fasta MatePair.RC.1.fq MatePair.
RC.2.fq \
MateMap.bam
reapr perfectmap MySpecies.filt.fasta PaiedEnd.1.fq
PairedEnd.2.fq \
[Insert Size] perfect
reapr pipeline MySpecies.filt.fasta MateMap.bam ReaprOut perfect
```

Output:

ReaprOut/04.break.broken_assembly.fa—Assembly file with identified misjoins broken.

3.5.7 Scaffolding with Hi-C Data

Objective:

To order and orientate scaffolds through long-distance contact frequencies in chromatin (*see* **Note 4.5.7**).

Method:

From experience, the optimal approach for application of Hi-C data to oomycetes and other organisms is currently through the use of commercial platforms [12, 13]. This is the approach reported in recent studies [118, 119]. Other scaffolders are emerging in the public domain [89, 120]; when tested on an oomycete, they produced suboptimal results (Fletcher and Michelmore, unpublished).

Output:

Assembly ordered and orientated with Hi-C contact frequency signals.

3.5.8 Scaffolding with Optical Maps

Objective:

To order and orientate scaffolds with optical maps, derived from restriction digest patterns of HMW DNA (*see* **Note 4.5.8**).

Method:

The current approach of recent publications is to use proprietary BioNano platforms, to align optical maps to scaffolds and produce hybrid scaffolds from alignments [121–123].

Output:

Assembly ordered and orientated with optical maps.

3.6 Repeat Masking and Annotation

Once a high-quality genome assembly has been obtained, the subsequent processes typically include repeat masking and annotation. These processes are quite complex therefore researchers should refer to recent literature and manuals (as indicated below) to produce optimal annotation.

We have provided some example syntax for how to run repeat-masking and annotation in this section. For simplicity, the input assembly for all modules is the output file of BLAST filtering (Module 3.4.1); however, this will not be the case because additional assembly optimization (Subheading 3.5) is often required prior to repeat-masking and annotation. Therefore, the syntax should be modified to accommodate the final assembly output of the preceding section (Subheading 3.5) after quality control.

3.6.1 Repeat-Masking

Objective:

To identify (model) repeats in the assembly and mask repeated sequences by replacing with string of Ns (*see* **Note 4.6.1**).

Tools:

RepeatModeler v1.0.8 [91]

RepeatMasker v4.0.6 [92]

Example syntax:

```
BuildDatabase -name isolate -engine ncbi MySpecies.filt.fasta
RepeatModeler -engine ncbi -database isolate
RepeatMasker -l [RepeatModelerDirectory]/consensi.fa.classified
-norna -nolow \
-dir output MySpecies.filt.fasta
```

Output:

RepeatModeler will generate a new directory containing the repeat library, used in the RepeatMasker example syntax

RepeatMasker will generate a masked assembly in the directory specified. The example syntax would generate an assembly called:

output/MySpecies.filt.fasta .masked

To softmask repeats (encode repeats with lower case nucleotides) include the flag "-xsmall" in the RepeatMasker syntax

3.6.2 Annotation

Objective 1:

To set up Maker for generation of an SNAPhmm (*see* **Note 4.6.2**).

Tools:

Maker v2.31.8 [93]

SNAP v2006-07-28 [94]

RepeatMasker v4.0.6 [92]
BLAST v2.6.0 [65]
Exonerate v2.2.0 [95]
Example syntax.

```
#Generate the three maker control files.
maker -ctl
#Use a text editor (i.e. nano) to edit the maker_opts file
nano maker_opts.ctl
#Edit genome line
genome= MySpecies.filt.fasta
#Edit organism type if necessary, default is eukaryotic
#Edit EST evidence. If available for your organism provide
here. You may also provide EST evidence of other organisms un-
der altest.
#If no EST data available for your organism, it is suggested to
provide ESTs of closely related organisms from NCBI as the est
evidence.
est=$PATH_TO_ESTs/est.fasta
altest+$PATH_TO_EST2/est2.fasta
#Provide protein evidence for your organisms or a closely re-
lated organism.
protein=$PATH_TO_PROTEINS/Proteins.fasta
#Provide repeat library generated by RepeatModeler (Module
3.6.1)
rmlib=$PATH_TO_REPEATMODELERRESULTS/consensi.fa.classified
#Switch est2genome and protein2genome on
est2genome=1
protein2genome=1
#Turn on prediction of single exon proteins
single_exon=1
single_length=240
#Save and exit nano (ctrl + O) to save. (ctrl + X) to exit.
```

Output 1:

Three Maker control files in the current working directory, one of which is now updated for your organism.

Objective 2:

To run Maker to obtain initial gene models (*see* **Note 4.6.2**).

Syntax:

```
maker 2> maker.err
```

Output 2:

GFF files in a new directory labeled GFF_files. We will refer to this as maker.gff.

Protein and transcript sequences of predicted gene models in the maker_runs directory

Objective 3 (*see* **Note 4.6.2**):

To obtain SNAP HMM.

Syntax:

```
#The gff3 file produced by maker should be compatible with snap.
#If multiple runs of maker are to be joined then the maker pro-
gram gff3_merge should be used prior to running the hmm predic-
tion.
#maker gff3 file must be converted in to zff
maker2zff maker.gff
#create the snap hmm
fathom -categorize 1000 maker.zff MySpecies.filt.fasta
fathom -export 1000 -plus uni.ann uni.dna
#best to make a new directory here, due to large output of next
step.
mkdir params
cd params
forge ../export.ann ../export.dna
cd ../
hmm-assembler.pl MyGenome params > MyGenome.hmm
Objective 4:
Re run maker with SNAP hmm
#Edit the maker control file providing the snap hmm.
snaphmm=PATH_TO_SNAPhmm/MyGenome.hmm
#Rerun maker – preferably after saving the initial results
elsewhere
maker 2> maker.err
```

4 Notes

These provide explanations for each of the modules above and references for further reading.

4.1 Quality Control and Preprocessing (Subheading 3.1)

Assemblers designed for single molecule sequencing [100–102] and paired-end linked reads of synthetic long molecules [14] are typically part of packages which require raw read inputs and output assemblies. Such assemblers may benefit from read filtering but do not require additional preprocessing. Postassembly modules may still be applied to generated assemblies, as required.

4.1.1 Read Quality Assessment (Module 3.1.1)

FastQC [29] takes fastq files as an input and outputs a diagnostic report in html format. A good early indicator of high levels of contamination in read files is a non-Poisson curve in the GC content (Fig. 2). FastQC can be run at any point on the fastq files (i.e. before and after preprocessing modules). Persistent error messages upon completion of preprocessing, relating to k-mer content should not necessarily cause trepidation. Although it may indicate an error, it may also be a biological reality of the data.

FastX toolkit [30] can be used as an alternative to FastQC to plot summary graphs. These include nucleotide distribution and

quality box-plots, similar to information presented by FastQC "Per base sequence content" and "Per base sequence quality" reports respectively but not k-mer distributions and several other reports provided by FastQC.

4.1.2 PCR Duplicate Removal (Module 3.1.2)

Removing redundant read pairs, which have resulted from sequencing of identical fragments (typically caused by PCR amplification during library preparation or by sequencing libraries to high depth) can reduce the input data for assembly and possibly reduce RAM requirements. Only short sequences (coordinates 10–35) at the ends of each read are compared during de-duplication to increase the rate and efficiency of the process and because the quality of reads decreases over their length. There is the possibility that some distinct sequences will be removed; however, this is not detrimental when there is adequate depth of high-quality, nonredundant reads. When using paired-end reads, both ends must match for a read pair to be removed. De-duplication will work best with paired-end reads, it is not recommended to de-duplicate single-end reads because they do not contain information from both ends of the sequenced fragment.

Alternative tools include BBMap (dedupe.sh) [32], FASTX toolkit [30], , FastUniq [33].

4.1.3 Read Merging (Module 3.1.3)

Merging overlapping paired-end reads generates longer, single end reads. This may assist in sequencing error correction, accuracy of sequence mapping, and reduce assembly errors.

Alternative tools include BBMap (bbmerge.sh) [32], PEAR [36], and COPE [37].

4.1.4 Adapter Trimming (Module 3.1.4)

Preparation of Illumina libraries includes the addition of adapters to allow hybridization to the flow cell surface prior to sequencing and to prime the sequencing reaction. These adapter sequences may be read through while sequencing, particularly of short fragments, resulting in their presence in raw fastq files. They should be identified and removed after the removal of PCR duplicates because adaptors provide information for Module 3.1.2.

Alternative tools include Scythe [38], Trimmomatic [39], AdapterRemoval v2 [40], and CutAdapt [41].

4.1.5 Quality Trimming (Module 3.1.5)

In a fastq file, each base of the read has a complimentary ASCII character on the 4th line indicating its quality based on the signal strength observed during sequencing. Programs have been designed to interpret this string and remove low quality bases from the sequence. If this trimming results in reads under a user-specified length, then they will be removed. Some tools allow this module to be run in parallel with adapter trimming.

Alternative tools include Sickle [42], Trimmomatic [39], and FASTX toolkit [30].

4.1.6 Contaminant Read Removal (Module 3.1.6)

While not always necessary, contaminant removal may be beneficial, particularly when assembling the genome of organisms that cannot be cultured axenically, necessitating the sequencing of mixed species. Removal of reads from contaminants may simplify the data during the assembly process. The trade-off is that some reads may be wrongly identified as contaminants and removed.

The approach used here involves first mapping all reads to a database containing assemblies of organisms that are closely related to the target species as well as potential contaminant organisms (e.g., host/prokaryotic genomes). Importantly, in order to subsequently identify the reads that map to these closely related organisms, a common string to distinguish between sequences from related species and contaminants should be introduced into the header of each scaffold. This string should be provided with the -s flag to FastqFilter.sh [43]. Reads that map to the closely related species and reads that do not map are then extracted, checked to be in correct pairs, and used for assembly. FastqFilter.sh has been provided to automate this process applied to paired-end Illumina reads; the basic commands required for completion of this module are provided in the README. This approach will not exclude all contaminants and therefore postassembly filtering is still required. Additionally, the approach will have false positives and it is possible that reads originating from the target organism may be classified as contaminants.

An alternative approach is to assemble all of the reads, identify contigs from contaminants (e.g., by BLAST) and then reassemble the reads which do not map to these contaminant contigs. Theoretically this should prevent the assembly of contaminant contigs in the second assembly, although this should still be verified.

4.1.7 Additional Considerations

Mate-Pair Reads

A different cleanup protocol should be applied to mate-pair reads. Mate-pair sequences are formatted differently from paired-end reads, facing out, instead of in, when mapped back to a genome assembly. These may be referred to as RF (reverse-forward) or "outie" pairs. Mate-pair libraries will also include some proportion of reads in tandem FF (forward–forward), RR (reverse–reverse), or "innie"—FR (forward–reverse) orientation. Depending on the way mate-pair reads are generated, different programs are applicable to identify and remove adapter sequences. Tools such as NextClip [47] have been specifically designed to process reads from libraries generated with the Illumina Nextera kit. These fragments contain an adapter sequence at the junction between the reads, which is used to characterize the read-pairs based on the presence and location of the adapter in the read. This verifies the reliability of the mate-pairs as "outies" indicating they are from the ends of large genomic fragments, rather than paired-end reads contaminating the library.

Normalization

K-mer based normalization of the reads may be performed to reduce the presence of reads containing highly abundant k-mers. The main benefit of this is that k-mers originating from repeats will be reduced, possibly reducing the complexity of the de Bruijn graph and reducing the RAM required for assembly. This does not guarantee an improvement in assembly and should therefore be applied with caution. It may be performed by BBMap (bbnorm.sh) [32].

Error Correction

Removing low coverage k-mers is another way to reduce the complexity of the de Bruijn graph generated during assembly. Low occurrence k-mers, present in the reads are assumed to be present due to sequencing errors; this is supported by assemblers performing better with error-corrected reads [48, 49]. Tools to perform this include Quake [48] and QuorUM [49]. Some assemblers perform this step as part of the de novo assembly process.

4.2 K-mer Assessment (Subheading 3.2)

4.2.1 Estimating Heterozygosity and Genome Size (Module 3.2.1)

Counting the occurrence of k-mers and plotting the histogram will produce a profile of peaks. The profile of a haploid or highly homozygous diploid individual is expected to be a single peak displaying all the "single copy" k-mers of the organism. A heterozygous diploid, balanced heterokaryon (nuclei in a 1:1 ratio) or polyploid species should display the "single-copy" portion of the genome in two peaks; the first peak, at position x, will represent the heterozygous k-mers, while the second peak at approximately 2× will represent the homozygous k-mers (Fig. 3). The size of the first peak indicates the heterozygous portion of the genome and may indicate whether a genome will be hard to assemble; highly heterozygous genomes tend to result in more fragmented assemblies, with an inflated total assembly size due to the independent assembly of divergent portions of the genome. Collapsing haplotypes (Module 3.5.5) may help reduce divergent regions into consensus haplotypes.

Highly repetitive k-mers are plotted beyond the "single-copy" peak(s). The genome size of the organism may therefore be estimated by summing the product of the frequency (*y* axis) and density (*x* axis) of these histograms. The size of the "single-copy" region on the organism can be calculated by restricting this estimate to only use k-mers included in the "single-copy" peak(s). Typically, a high count of low frequency k-mers indicates a high number of contaminants.

GenomeScope [51] independently visualizes the k-mer histogram produced and applies a model to this estimating the genome size, levels of duplication and heterozygosity (Fig. 3). It is informative to query how well the model fits the input data as this will reflect how accurate the estimates provided are. To obtain an accurate estimate of genome size for a highly repetitive genome, it may be necessary to increase the maximum k-mer coverage of GenomeScope (set to 1000× as default) or increase the size of the

histogram generated by Jellyfish [50]. By default, Jellyfish will count k-mers up to 10,000× multiplicity, highly repetitive k-mers may be present in a genome more frequently than this, which it may be necessary to account for. However, high multiplicity k-mers could have other sources (i.e. organelles or contaminants), which may artificially inflate the apparent genome, so caution, or read filtering, may be required prior to running and interpreting Jellyfish.

4.2.2 K-mer Size Optimization for Assembly (Module 3.2.2)

A key parameter to be optimized for most assemblers is the k-mer size. This is the length the assembler will reduce reads down to, in a sliding window manner, to build the de Bruin graph. A suggested optimal k-mer size for this parameter may be generated by software such as kmergenie [52]. This provides a good starting point, with further optimization through reiterative assembly necessary, in order to obtain the best assembly.

4.3 Assembly (Subheading 3.3)

As stated previously, the assembler used will be dictated by the input data and what has been previously reported as successful on similar organisms. There is no one correct approach and multiple assemblers should be tried. High levels of synteny can often be found when different algorithms are used.

A critical parameter to optimize is the k-mer size. All assemblers have upper limits; some require specific instructions at installation to work at higher k-mers than set by default [53]. Some will work only on odd-numbered k-mers [56], while others will generate a de Bruijn graph using multiple k-mer sizes to guide contig assembly and scaffolding [59–61]. Some will build a de Bruijn graph and then attempt to build an Overlap Layout Consensus graph from the results [54, 55].

The optimal approach would be to run different assemblers using different k-mer sizes to generate the best assembly; however, realistically this is impractical. Provided that the assembly is well supported by the data available, has good statistics, a high completeness score, and is highly syntenic with closely related organisms (Module 3.4.5) and/or validated genetically (*see* **Note 4.5.9**), it may be a useful as a draft genome. ABySS [53] has been a useful tool for generating multiple draft assemblies because it is rapid and has a low memory consumption; MaSuRCA [55] output larger scaffolds, with higher N_{50}s compared to other assemblers at the expense of much longer run times and higher memory consumption (Fletcher and Michelmore, unpublished data).

4.4 Assembly Quality Control (Subheading 3.4)

4.4.1 BLAST Filtering (Module 3.4.1)

Even after applying preassembly read filtering, contaminant organisms are commonly present in an assembly. Therefore, an assembly should be filtered against the NCBI nucleotide database (nt) early during quality control. Failure to do so may result in deceptively good results because contaminant organisms, such as bacteria, may be more readily assembled resulting in contaminants

being the largest scaffolds. Downloading a copy of nt (ftp://ftp.ncbi.nlm.nih.gov/blast/db/), so that BLAST [65] may be run locally and results queried from the command-line is the most efficient way to filter an assembly. If nt already contains similar genomes to the target organism, then the default BLAST settings may be adequate to positively identify assembled scaffolds. However, if the organism under study is phylogenetically distinct from previously sequenced organisms, then the final assembly will benefit from the increased sensitivity of smaller word sizes. This adjustment comes with a trade-off, requiring more cpu time to run the filtering. Alternatively, or in addition, BLASTX may be used to query the NCBI nonredundant protein database to infer potential conserved gene models of unaligned scaffolds. A shell wrapper to complete this module has been provided on GitHub [64], the base commands for completion are present in the README there.

4.4.2 Assembly Statistics (Module 3.4.2)

Commonly used statistics are a useful indicator when assessing assemblies intrinsically (the same data set, run through different preprocessing modules or assemblers) and extrinsically (compared to closely related organisms). The N_{50} of contigs and scaffolds is often reported; however, this metric should be used with caution because this, like several other assembly statistics, is sensitive to the cutoff threshold used for the smallest scaffolds [124, 125].

4.4.3 Completeness Statistics (Module 3.4.3)

The BUSCO score is a good measure for assembly quality. Due to the use of taxon-specific (i.e. protist or fungal) databases to query assemblies, BUSCO [67] is preferred over the now superseded CEGMA [68], which used a single, general eukaryotic database. The specificity of BUSCO over CEGMA means fewer false positives are detected from contaminant organisms. Oomycetes may be tested against the single generalized protist library. Fungi, on the other hand, may be tested on one or more of at least ten fungal libraries. Such libraries range from a generalized fungi library to more specific libraries for Eurotiomycetes, Saccharomycetales, or Sordariomyceta. Both oomycetes and fungi can be tested against a general Eukaryota library. The duplicated score is a good indicator of whether a portion of the genome has assembled separately (i.e., where both haplotypes have not been collapsed into a consensus). Combining BUSCO scores with read coverage (Module 3.4.4) and k-mer assessment (Module 3.4.6) provides a good indication of the quality of the consensus assembly.

Authorial experience has highlighted the importance of investigating absent and especially lost BUSCO models when assessing different technologies. Incorporation of PacBio reads into assemblies of some oomycetes resulted in generation of indels in several BUSCO genes causing frameshifts and the failure to predict these gene models; these indels were not supported by Illumina reads (Fletcher and Michelmore, unpublished data).

4.4.4 Paired-End Coverage of Assembly (Module 3.4.4)

Mapping reads back to an assembly and producing a plot of the coverage histogram is an informative approach for assessing the amount of duplication in the assembly. In a consensus assembly, a Poisson curve should be observed around the expected coverage, regardless of ploidy, or heterozygosity. Additional peaks may indicate that haplotypes have been assembled separately or that there is a significant genomic variant present in your assembly. Contaminants, if present are expected at low coverage, as observed in equivalent k-mer plots, though this is not guaranteed.

4.4.5 Whole Genome Alignment (Module 3.4.5)

While chromosomal rearrangements between species are to be expected, variable levels of synteny between gene blocks will persist across evolutionary time. Therefore, synteny as detected through whole genome alignment with a closely related organism can be used to assess and compare the quality of assemblies. This can be done before annotation using MUMmer or SyMap (Module 3.4.5). Repeating synteny analysis after annotation using Symap or CoGe and alignments of gene models may also be informative. SyMap generates plots through a GUI, with or without gene models (Fig. 4a) [71]. CoGe [72] produces similar inferences through a web interface. A high-quality reference assembly is required in order to provide good estimates of quality through top-down measures of synteny between species. MUMmer [70] allows finer detailed investigation of fewer scaffolds (Fig. 4b) and can be useful when comparing assemblies of closely related organisms or multiple assemblies of the same isolate (Fig. 4c).

4.4.6 K-mer Comparisons; Reads vs Assembly (Module 3.4.6)

The assumption that relative k-mer presence in raw reads should be reflected in k-mer presence in the assembly can be tested by plotting a KAT [73] spectra-cn plot. KAT provides information similar to mapping reads back to the assembly, but is quicker because it removes the alignment process and more specific because it is not affected by mapping or assembly errors. KAT spectra-cn plots (Fig. 5) overlay k-mer histograms (as previously plotted by tools such as GenomeScope; Module 3.2.1) with representation information from the queried assembly, indicating the proportions of k-mers correctly represented against those overrepresented, underrepresented, or absent. The manual for KAT contains examples of perfect consensus and diploid assemblies and how they should be portrayed in these plots; however, such plots are rarely obtained. Figure 5 depicts two published oomycete assemblies [25, 107]. In a consensus assembly, half the heterozygous k-mers are expected to be absent, as depicted in Fig. 5b. Additional absent k-mers are depicted in black in the homozygous peaks, of both example assemblies, which also contain a small proportion of multiplied k-mers, again found in both assemblies. While KAT cannot directly address these issues, it can provide an indication as to what is required. An absence of k-mers may indicate that more data needs

to be incorporated or that the assembly has been too harshly filtered, while an overrepresentation homozygous or heterozygous k-mers may suggest that haplotype resolution is needed.

4.5 Assembly Optimization (Subheading 3.5)

Depending on the results of the initial QC, further optimization or even a complete overhaul of an assembly may be necessary. Ideally, before applying these techniques the biology of the organism under study should be well understood (heterozygosity, genome size, ploidy, chromosome number). The following techniques will enable the production of chromosome-scale assemblies, particularly the use of Hi-C scaffolding or optical maps, either independently or together.

4.5.1 Reassembly (Module 3.5.1)

As stated previously, producing multiple assemblies from the same data set is a robust approach to empirically determine the best assembler and/or parameters for the input data. Initially, the same data should be run through multiple assemblers or parameter variations and quality statistics generated, to identify the best assembler and proceed to optimization from there. Additional assemblies can be beneficial for comparative analysis and benchmarking expectations, especially if no closely related organism has been sequenced. Assemblies generated using different algorithms or parameters are also useful as templates for merging to generate consensus assemblies.

4.5.2 (Re)Scaffolding (Module 3.5.2)

Scaffolding is a built-in function of most assemblers. Long-distance data, for example in the form of mate pair reads, may be provided at the initial assembly step. However, it is often beneficial or necessary to apply additional information through scaffolding or gap-filling (Module 3.5.8) postassembly. Stand-alone scaffolders such as SSPACE [74], SGA [75], and SOPRA [76] allow for additional incorporation of mate-pairs and other paired reads (e.g., from the ends of fosmid clones or BACs) into the assembly. These tools have been bench-marked and shown to be effective [126]. Depending on the depth of the data available, this process may be optimized by altering the mapping parameters or the number of links required to join scaffolds.

Other sequences, including linked reads generated from 10× Genomics Chromium libraries and PacBio or Oxford Nanopore single molecule sequences may also be used to scaffold, although specialized programs including Scaff10x, FragScaff [77], ARCs [78], Links [79], or PBJelly [80] are required to incorporate such data. SSPACE-LongRead [81] is an alternative software to scaffold using single molecule sequencing; although this tool should be used with caution; in the authors' experience, it appeared to introduce several errors and made spurious joins, disobeying the number of user-specified links required between fragments.

4.5.3 Gap-Filling (Module 3.5.3)

Initial assemblies that have been derived from Illumina reads usually contain gaps, even in the absence of mate-pair data as input to the initial assembly. Gap filling software, such as GapFiller [82], GapCloser [57], or Sealer [83], are able to place read data into gaps thus adding new sequences to the assembly. This may reduce the proportion of gaps in an assembly, through complete or partial gap-closing and may improve the gene content. Observing whether BUSCO [67] completeness scores (Module 3.4.3) increase post gap-filling is an indication of the effectiveness of gap-filling.

Gap filling with PacBio read data can be performed in parallel with scaffolding, using PBjelly [80] as mentioned above. PacBio variantCaller [84] algorithms, Arrow and Quiver, may be applicable; however, these are recommended as being applied to haploid/diploid/polyploid assemblies, not consensus assemblies because the algorithms are not phase aware and may therefore produce chimeric gene models. In addition, the authors experienced a reduction in the BUSCO score when variantCaller was applied to a consensus assembly of an oomycete due to the introduction of erroneous indels that generated frameshifts in BUSCO genes (*see* **Note 4.4.3**).

Gap filling with 10× genomics data is performed with the same software as Illumina data. The only additional process is that the first read should have 23 bases trimmed from the 5′ end, because this contains barcode sequence. The following syntax achieves the required output for gap-filling:

```
bbduk.sh forcetrimleft=23 in=10xread.1.fq out 10xread.trim.1.fq
```

At present, no software is available that makes use of 10× linked-read information for gap-filling.

4.5.4 Merge Alternative Assemblies (Module 3.5.4)

Quickmerge was originally developed to combine assemblies of the same data source that had been assembled independently [85], through whole genome alignment. This method is distinct from that employed by assemblers that construct an assembly from a de Bruijn graph, built with multiple k-mer sizes to produce a final assembly [59–61]. Given good assemblies as inputs, Quickmerge produces merged assemblies, which retain a high degree of synteny with closely related species. This approach is worthwhile if combining assemblies generated using multiple parameters with the same assembler or when merging assemblies generated with different technologies or assemblers.

4.5.5 Removal of Alternative Haplotypes (Module 3.5.5)

Assembling highly heterozygous or dikaryotic species, such as some oomycetes and basidiomycetes, will result in a mixed consensus diploid assembly. Typically, this is attributed to some sequences being too divergent resulting in unresolved bubbles in the de Bruin graph of the assembler. Haplomerger2 [86] is well-equipped to deal with such instances, provided that the input assembly is of

high quality; an N_{50} greater than 50 kb is recommended. For ease of use, a shell script called runHM2.sh [116] is provided to run Haplomerger2 on a given, input assembly. This script runs the basic Haplomerger2 workflow, without the supplementary scaffolding or gap-filling processes, which have been described in Modules 3.4.2 and 3.5.3, respectively. The additional dependencies, SSPACE [74] and GapCloser [57], are required to use these supplementary features and the Haplomerger2 manual should be consulted for further guidance.

Haplomerger2 outputs two FASTA files that together represent the diploid genome, although the steps after haplotype removal (tandem removal and optional scaffolding and gap-filling) are not performed on the alternative assembly output. Importantly, despite the name alluding to it, separate haplotypes are not merged by this process; two alternative haplotype FASTA files are output instead. Methods for evaluating haplotype reduction include determining the BUSCO duplication score (Module 3.4.3), read depth of highly divergent effectors and KAT [73] spectra-cn (Module 3.4.6)/read coverage (Module 3.4.4) plots before and after this module. Performing multiple cycles of Haplomerger2 with rescaffolding either as part of the pipeline (hm.batchC processes) or as a separate process, is often beneficial for obtaining a single consensus assembly. Redundans [87] is an alternative tool and may be used in combination with Haplomerger2. It does not have the same high input N_{50} requirements as Haplomerger2 and so may be applied to more fragmented assemblies. When these software tools were applied to assemblies of heterozygous oomycetes, Redundans was not as effective independently as Haplomerger2; neither program entirely removed all alternate haplotypes, possible due to some haplotypes being too divergent (unpublished data).

4.5.6 Identify and Break Misjoins (Module 3.5.6)

REAPR [88] is one of the few tools available for automated quantification of assembly accuracy, taking mate-pair data sequenced to a reasonable depth (~15× recommended) and identifying misjoins through paired-data. However, it has not been benchmarked how increased heterozygosity of an organism affects the perfect mapping required and how inclusion of PacBio reads, assembling low Illumina coverage, AT-rich regions are scored by this tool.

4.5.7 Scaffolding Using Hi-C Data

Originally developed to investigate chromosomal interactions in interphase nuclei [127], Hi-C has more recently been utilized to scaffold assemblies into chromosome-scale molecules [89, 90, 109, 119, 120], as well as metagenome deconvolution [128, 129] and 3D nuclear visualization [110, 130, 131]. There are publicly available scaffolders [89, 120] which utilize Hi-C data; however, limited experience with such data applied to oomycetes and other species suggested that commercial offerings such as HiRise [12]

and Proximo [13] are currently superior for obtaining the best output assembly.

BWA v0.5.7 [44]
SAMtools v0.18 [45]
BEDtools2 [69]
SALSA [89]

4.5.8 Scaffolding Using Optical Maps (Module 3.5.8)

Optical maps are another approach for long-range chromosome scale scaffolding. Large single genomic DNA molecules are fingerprinted and chromosome length maps are constructed from overlapping signatures [132, 133]. Initially, this was done with immobilized strands of DNA and restriction digests [133]. Two methods are currently, commercially availble; either DNA is nicked at specific sequences followed by incorporation of fluorescently labeled nucleotides or DNA is directly labelled. High-throughput microfluidics are then used to produce the fingerprints [11, 111]. This map can then be aligned against an in silico digest of a genome assembly, so as to order and orientate scaffolds into chromosome-scale molecules [132]. Optical maps have been utilized for genome assembly validation, contig incorporation, de novo assembly, and methylome investigation [111–113, 132, 134, 135]. Most contemporary studies [121–123] use proprietary hardware and software to incorporate optical maps into their assembly [11].

4.5.9 Validation by Genetic Analysis

Genetic analysis provides a rigorous, independent method to validate the quality of an assembly. This is a multistep process beyond the scope of this chapter [114] and the generation of experimental segregating populations may not be possible for some oomycete and fungal pathogens. If feasible, genotyping by whole genome sequencing can provide large numbers of markers and high-resolution analysis of the inheritance of much of the assembled genome. The lack of congruity between segregation and sequence data can identify chimeric scaffolds and the points at which they should be split. In the absence of chromosome-scale scaffolds, genetic data can be used to order and orient scaffolds into chromosomal pseudomolecules. The number of progeny individuals required to do this depends on the number of scaffolds in the assembly. If the size of the progeny is too small, there will be an insufficient number of meioses to orient and order scaffolds within genetic bins.

4.6 Repeat-Masking and Annotation

4.6.1 Repeat-Masking (Module 3.6.1)

Repeats may make up a large proportion of the genome of a species and may provide false signals when performing annotation [136] or comparative analysis. Therefore, repeats should be masked prior to annotation or a repeat library be provided to the annotation pipeline. A detailed review of several programs for de novo repeat identification is available [136].

4.6.2 Annotation (Module 3.6.2)

Annotation is a complex process that often requires optimization. Due to its importance, a basic template is provided for how annotations may be produced. Multiple ab initio gene predictors in addition to SNAP [94] are available including Augustus [97], GeneMark [98], and glimmerHMM [99]. These may be used in conjunction with Maker [93] or with other gene predictions pipelines such as JAMg [96]. Alternative protocols for running Maker have been defined [137] and should be consulted in addition to publications utilizing the pipeline [15, 138–140].

Oomycetes also contain diverse effectors with single exons, which ab initio gene predictors may fail to detect. Applying string searches for motifs and HMMs generated from classes of effector are commonly used strategies that are constantly being updated [15, 25, 141, 142]. Incorporation of data from contemporary literature will be critical to obtaining the best possible effector predictions for de novo assemblies.

The highest quality annotations will be produced by providing multiple lines of evidence. Annotations using Maker greatly benefit from having transcriptome data generated from the organism under study, although ESTs of other organisms can be supplied instead of, or supplementary to such data. The protein models provided to Maker should be checked to ensure that they are of high quality. A robust option is to download reference genome sets from NCBI because these are usually stringently filtered. Finally, ab initio gene predictors can be trained after multiple rounds of prediction, as evidenced by the quality of predictions generated by SNAP increasing due to bootstrapping the data [137].

4.7 Additional Considerations

4.7.1 Public Release of Data

The raw reads, assembly, and gene models should be submitted to one or more to public databases (e.g., NCBI), once their quality has been established. These will be subjected to standardized filtering, with spurious results being flagged, indicating if something needs to be reanalyzed. Common filters include searching for contaminating organisms or removing gene models that are missing start/stop codons or with unusually small introns. Submitting to NCBI can be time-consuming; GAG [143] can assist with filtering and formatting annotation data to NCBI standards. Gene models should be submitted in support gene numbers reported in a publication because this greatly helps comparative genomics.

4.7.2 Future Perspectives

In the future, sequencing data will provide information on additional aspects of genome structure that are not covered in this chapter. Each technology will provide different supplemental insights. For example, single-molecule sequencing can provide genome-wide profiles of DNA modifications [144–146], not just methylation of cysteine residues that is accessible by bisulfite sequencing. It is possible that oomycetes and fungi have multiple DNA modifications that have yet to be characterized. Hi-C can provide information on topologically associated domains within

the nucleus that will aid in understanding regulation of gene expression [147, 148].

High-quality genome assemblies are the foundation for many downstream studies. However, it has become increasingly difficult to publish genome papers in the absence of a biological revelation enabled by the genome sequence [149]. Despite advances in technology and computational approaches, generating high-quality, useful assemblies remains an arduous, time-consuming task that requires careful implementation as illustrated here in order to provide accurate assemblies. Nonetheless, we are at the point where such resources can be reliably generated for most oomycetes and fungi (although some, for example diploid biotrophic oomycetes, will be more challenging than others, such as haploid fungi).

Acknowledgments

We would like to thank Sebastian Reyes-Chin-Wo (UC Davis) and Lida Derevnina (now The Sainsbury Laboratory, Norwich, UK) for their contributions in the initial setup of the workflow and William Palmer and Kelsey Wood (both UC Davis) for their reviews during preparation. We thank the UC Davis Bioinformatics Core for their computational and software support. We also thank Diane Saunders and members of her lab Guru Radhakrishnan, Daniel C.E. Bunting, and Antoine Persoons for their helpful comments. The work was supported by The Novozymes Inc. Endowed Chair in Genomics to RWM.

Conflicts of interest: The statements regarding Hi-C are based on experience resulting from collaboration with Dovetail Genomics. The authors declare that there are no other potential conflicts of interest.

References

1. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG (1996) Life with 6000 genes. Science 274(5287):546–567. https://doi.org/10.1126/science.274.5287.546
2. Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RH, Aerts A, Arredondo FD, Baxter L, Bensasson D, Beynon JL (2006) Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313. https://doi.org/10.1126/science.1128796
3. Bussey H, Kaback DB, Zhong W, Vo DT, Clark MW, Fortin N, Hall J, Ouellette BF, Keng T, Barton AB et al (1995) The nucleotide sequence of chromosome I from Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 92(9):3809–3813
4. Klosterman SJ, Subbarao KV, Kang S, Veronese P, Gold SE, Thomma BPHJ, Chen Z, Henrissat B, Lee Y-H, Park J, Garcia-Pedrajas MD, Barbara DJ, Anchieta A, de Jonge R, Santhanam P, Maruthachalam K, Atallah Z, Amyotte SG, Paz Z, Inderbitzin P, Hayes RJ, Heiman DI, Young S, Zeng Q, Engels R, Galagan J, Cuomo CA, Dobinson KF, Ma L-J (2011) Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. PLoS Pathog 7(7):e1002137. https://doi.org/10.1371/journal.ppat.1002137
5. Cuomo CA, Güldener U, Xu J-R, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma

L-J, Baker SE, Rep M, Adam G, Antoniw J, Baldwin T, Calvo S, Chang Y-L, DeCaprio D, Gale LR, Gnerre S, Goswami RS, Hammond-Kosack K, Harris LJ, Hilburn K, Kennell JC, Kroken S, Magnuson JK, Mannhaupt G, Mauceli E, Mewes H-W, Mitterbauer R, Muehlbauer G, Münsterkötter M, Nelson D, Donnell K, Ouellet T, Qi W, Quesneville H, Roncero MIG, Seong K-Y, Tetko IV, Urban M, Waalwijk C, Ward TJ, Yao J, Birren BW, Kistler HC (2007) The Fusarium graminearum genome reveals a link between localized polymorphism and pathogen specialization. Science 317(5843):1400

6. Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B, Houterman PM, Kang S, Shim W-B, Woloshuk C, Xie X, Xu J-R, Antoniw J, Baker SE, Bluhm BH, Breakspear A, Brown DW, Butchko RAE, Chapman S, Coulson R, Coutinho PM, Danchin EGJ, Diener A, Gale LR, Gardiner DM, Goff S, Hammond-Kosack KE, Hilburn K, Hua-Van A, Jonkers W, Kazan K, Kodira CD, Koehrsen M, Kumar L, Lee Y-H, Li L, Manners JM, Miranda-Saavedra D, Mukherjee M, Park G, Park J, Park S-Y, Proctor RH, Regev A, Ruiz-Roldan MC, Sain D, Sakthikumar S, Sykes S, Schwartz DC, Turgeon BG, Wapinski I, Yoder O, Young S, Zeng Q, Zhou S, Galagan J, Cuomo CA, Kistler HC, Rep M (2010) Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. Nature 464(7287):367–373
7. Lamour KH, Mudge J, Gobena D, Hurtado-Gonzales OP, Schmutz J, Kuo A, Miller NA, Rice BJ, Raffaele S, Cano LM (2012) Genome sequencing and mapping reveal loss of heterozygosity as a mechanism for rapid adaptation in the vegetable pathogen Phytophthora capsici. Mol Plant Microbe Interact 25
8. Shen R, Fan JB, Campbell D, Chang W, Chen J, Doucet D, Yeakley J, Bibikova M, Wickham Garcia E, McBride C, Steemers F, Garcia F, Kermani BG, Gunderson K, Oliphant A (2005) High-throughput SNP genotyping on universal bead arrays. Mutat Res 573(1–2): 70–82. https://doi.org/10.1016/j.mrfmmm.2004.07.022
9. Pacific Biosciences (PacBio). http://www.pacb.com/. Accessed 29 Sept 2017
10. Oxford Nanopore Technologies. https://nanoporetech.com/. Accessed 29 Sept 2017
11. Bionano Genomics. https://bionanogenomics.com/. Accessed 29 Sept 2017
12. Dovetail Genomics. https://dovetailgenomics.com/. Accessed 29 Sept 2017
13. Phase Genomics. https://phasegenomics.com/. Accessed 29 Sept 2017
14. Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB (2017) Direct determination of diploid genome sequences. Genome Res 27(5):757–767. https://doi.org/10.1101/gr.214874.116
15. Derevnina L, Chin-Wo-Reyes S, Martin F, Wood K, Froenicke L, Spring O, Michelmore R (2015) Genome sequence and architecture of the tobacco downy mildew pathogen Peronospora tabacina. Mol Plant-Microbe Interact 28(11):1198–1215. https://doi.org/10.1094/MPMI-05-15-0112-R
16. Bradnam K, Korf I (2012) UNIX and Perl to the rescue!: a field guide for the life sciences (and other data-rich pursuits). Cambridge University Press, Cambridge
17. Software Carpentry Foundation; The Unix Shell. http://swcarpentry.github.io/shell-novice/. Accessed 13 Nov 2017
18. Köster J, Rahmann S (2012) Snakemake—a scalable bioinformatics workflow engine. Bioinformatics 28(19):2520–2522. https://doi.org/10.1093/bioinformatics/bts480
19. Jupyter. https://jupyter.org/. Accessed 7 Nov 2017
20. Kushwaha SK, Vetukuri RR, Grenville-Briggs LJ (2017) Draft genome sequence of the mycoparasitic oomycete Pythium periplocum strain CBS 532.74. Genome Announc 5(12):e00057-00017
21. Berger H, Yacoub A, Gerbore J, Grizard D, Rey P, Sessitsch A, Compant S (2016) Draft genome sequence of biocontrol agent Pythium oligandrum strain Po37, an oomycota. Genome Announc 4(2):e00215-00216
22. Kushwaha SK, Vetukuri RR, Grenville-Briggs LJ (2017) Draft genome sequence of the mycoparasitic oomycete pythium oligandrum strain CBS 530.74. Genome Announc 5(21). https://doi.org/10.1128/genomeA.00346-17
23. Kemen E, Gardiner A, Schultz-Larsen T, Kemen AC, Balmuth AL, Robert-Seilaniantz A, Bailey K, Holub E, Studholme DJ, MacLean D, Jones JDG (2011) Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of Arabidopsis thaliana. PLoS Biol 9(7):e1001094. https://doi.org/10.1371/journal.pbio.1001094
24. Pais M, Win J, Yoshida K, Etherington GJ, Cano LM, Raffaele S, Banfield MJ, Jones A, Kamoun S, Saunders DGO (2013) From pathogen genomes to host plant processes: the power of plant parasitic oomycetes. Genome Biol 14(6):211. https://doi.org/10.1186/gb-2013-14-6-211

25. Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T, Bozkurt TO, Ah-Fong AMV, Alvarado L, Anderson VL, Armstrong MR, Avrova A, Baxter L, Beynon J, Boevink PC, Bollmann SR, Bos JIB, Bulone V, Cai G, Cakir C, Carrington JC, Chawner M, Conti L, Costanzo S, Ewan R, Fahlgren N, Fischbach MA, Fugelstad J, Gilroy EM, Gnerre S, Green PJ, Grenville-Briggs LJ, Griffith J, Grunwald NJ, Horn K, Horner NR, Hu C-H, Huitema E, Jeong D-H, Jones AME, Jones JDG, Jones RW, Karlsson EK, Kunjeti SG, Lamour K, Liu Z, Ma L, MacLean D, Chibucos MC, McDonald H, McWalters J, Meijer HJG, Morgan W, Morris PF, Munro CA, O'Neill K, Ospina-Giraldo M, Pinzon A, Pritchard L, Ramsahoye B, Ren Q, Restrepo S, Roy S, Sadanandom A, Savidor A, Schornack S, Schwartz DC, Schumann UD, Schwessinger B, Seyer L, Sharpe T, Silvar C, Song J, Studholme DJ, Sykes S, Thines M, van de Vondervoort PJI, Phuntumart V, Wawra S, Weide R, Win J, Young C, Zhou S, Fry W, Meyers BC, van West P, Ristaino J, Govers F, Birch PRJ, Whisson SC, Judelson HS, Nusbaum C (2009) Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. Nature 461(7262):393–398
26. Ramezani-Rad M, Hollenberg CP, Lauber J, Wedler H, Griess E, Wagner C, Albermann K, Hani J, Piontek M, Dahlems U, Gellissen G (2003) The Hansenula polymorpha (strain CBS4732) genome sequencing and analysis. FEMS Yeast Res 4(2):207–215
27. Gregory TR, Nicol JA, Tamm H, Kullman B, Kullman K, Leitch IJ, Murray BG, Kapraun DF, Greilhuber J, Bennett MD (2007) Eukaryotic genome size databases. Nucleic Acids Res 35(Database issue):D332–D338. https://doi.org/10.1093/nar/gkl828
28. Egertová Z, Sochor M (2017) The largest fungal genome discovered in Jafnea semitosta. Plant Syst Evol 303(7):981–986. https://doi.org/10.1007/s00606-017-1424-9
29. Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc. Accessed 13 Sept 2017
30. Hannon Lab FASTX Toolkit. doi:citeulike-article-id:9103573
31. Petersen KR, Streett DA, Gerritsen AT, Hunter SS, Settles ML (2015) Super deduper, fast PCR duplicate detection in fastq files. In: Proceedings of the 6th ACM conference on bioinformatics, computational biology and health informatics. ACM, pp 491–492
32. Bushnell B (2016) BBMap short read aligner. University of California, Berkeley, CA URL: http://sourceforgenet/projects/bbmap
33. Xu H, Luo X, Qian J, Pang X, Song J, Qian G, Chen J, Chen S (2012) FastUniq: a fast de novo duplicates removal tool for paired short reads. PLoS One 7(12):e52249. https://doi.org/10.1371/journal.pone.0052249
34. Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27(21):2957–2963. https://doi.org/10.1093/bioinformatics/btr507
35. Streett DA (2015) FLASH2. https://github.com/dstreett/FLASH2. Accessed 29 Sept 2017
36. Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30(5):614–620. https://doi.org/10.1093/bioinformatics/btt593
37. Liu B, Yuan J, Yiu S-M, Li Z, Xie Y, Chen Y, Shi Y, Zhang H, Li Y, Lam T-W (2012) COPE: an accurate k-mer-based pair-end reads connection tool to facilitate genome assembly. Bioinformatics 28(22):2870–2874
38. Buffalo V (2014) Scythe [Software]. https://github.com/vsbuffalo/scythe. Accessed 29 Sept 2017
39. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170
40. Schubert M, Lindgreen S, Orlando L (2016) AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes 9:88. https://doi.org/10.1186/s13104-016-1900-2
41. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17(1):10–12. https://doi.org/10.14806/ej.17.1.200
42. Joshi NA, Fass, J.N. (2011) Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. https://github.com/najoshi/sickle. Accessed 13 Sept 2017
43. Fletcher K (2017) FastqFilter.sh [Software]. https://github.com/kfletcher88/FastqFilter. Accessed 29 Sept 2017
44. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:13033997
45. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352

46. Staton SE (2013) Pairfq [Software]. https://github.com/sestaton/Pairfq. Accessed 29 Sept 2017
47. Leggett RM, Clavijo BJ, Clissold L, Clark MD, Caccamo M (2014) NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries. Bioinformatics 30(4):566–568. https://doi.org/10.1093/bioinformatics/btt702
48. Kelley DR, Schatz MC, Salzberg SL (2010) Quake: quality-aware detection and correction of sequencing errors. Genome Biol 11(11):R116. https://doi.org/10.1186/gb-2010-11-11-r116
49. Marçais G, Yorke JA, Zimin A (2015) QuorUM: an error corrector for Illumina reads. PLoS One 10(6):e0130821
50. Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27(6):764–770. https://doi.org/10.1093/bioinformatics/btr011
51. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC (2017) GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33(14):2202–2204. https://doi.org/10.1093/bioinformatics/btx153
52. Chikhi R, Medvedev P (2013) Informed and automated k-mer size selection for genome assembly. Bioinformatics 30(1):31–37
53. Simpson J, Wong K, Jackman S, Schein J, Jones S, Birol I (2009) ABySS: a parallel assembler for short read sequence data. Genome Res 19. https://doi.org/10.1101/gr.089532.108
54. Ye C, Hill CM, Wu S, Ruan J, Ma ZS (2016) DBG2OLC: efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. Sci Rep 6:31900
55. Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA (2013) The MaSuRCA genome assembler. Bioinformatics 29(21):2669–2677. https://doi.org/10.1093/bioinformatics/btt476
56. Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18(5):821–829
57. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1(1):18. https://doi.org/10.1186/2047-217x-1-18
58. MacCallum I, Przybylski D, Gnerre S, Burton J, Shlyakhter I, Gnirke A, Malek J, McKernan K, Ranade S, Shea TP, Williams L, Young S, Nusbaum C, Jaffe DB (2009) ALLPATHS 2: small genomes assembled accurately and with high continuity from short paired reads. Genome Biol 10. https://doi.org/10.1186/gb-2009-10-10-r103
59. Goltsman E, Ho I, Rokhsar D (2017) Meraculous-2D: haplotype-sensitive assembly of highly heterozygous genomes. arXiv preprint arXiv:170309852
60. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, Maruyama H, Kohara Y, Fujiyama A, Hayashi T, Itoh T (2014) Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. https://doi.org/10.1101/gr.170720.113
61. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19. https://doi.org/10.1089/cmb.2012.0021
62. Weisenfeld NI, Yin S, Sharpe T, Lau B, Hegarty R, Holmes L, Sogoloff B, Tabbaa D, Williams L, Russ C, Nusbaum C, Lander ES, MacCallum I, Jaffe DB (2014) Comprehensive variation discovery in single human genomes. Nat Genet 46:1350. https://doi.org/10.1038/ng.3121
63. Peng Y, Leung HC, Yiu SM, Chin FY (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28. https://doi.org/10.1093/bioinformatics/bts174
64. Fletcher K (2017) AssemblyFilter.sh [Software]. https://github.com/kfletcher88/AssemblyFilter. Accessed 29 Sept 2017
65. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410. https://doi.org/10.1016/s0022-2836(05)80360-2
66. Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29(8):1072–1075. https://doi.org/10.1093/bioinformatics/btt086
67. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210–

3212. https://doi.org/10.1093/bioinformatics/btv351
68. Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23. https://doi.org/10.1093/bioinformatics/btm071
69. Quinlan AR (2014) BEDTools: the Swiss-army tool for genome feature analysis. Current Protoc Bioinformatics 47:11.12.11–11.12.34. https://doi.org/10.1002/0471250953.bi1112s47
70. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL (2004) Versatile and open software for comparing large genomes. Genome Biol 5. https://doi.org/10.1186/gb-2004-5-2-r12
71. Soderlund C, Bomhoff M, Nelson WM (2011) SyMAP v3.4: a turnkey synteny system with application to plant genomes. Nucleic Acids Res 39. https://doi.org/10.1093/nar/gkr123
72. Lyons E, Freeling M (2008) How to usefully compare homologous plant genes and chromosomes as DNA sequences. Plant J 53(4):661–673
73. Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, Clavijo BJ (2017) KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. Bioinformatics 33(4):574–576. https://doi.org/10.1093/bioinformatics/btw663
74. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W (2011) Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27. https://doi.org/10.1093/bioinformatics/btq683
75. Simpson JT, Durbin R (2012) Efficient de novo assembly of large genomes using compressed data structures. Genome Res 22(3):549–556. https://doi.org/10.1101/gr.126953.111
76. Dayarian A, Michael TP, Sengupta AM (2010) SOPRA: scaffolding algorithm for paired reads via statistical optimization. BMC Bioinformatics 11:345–345. https://doi.org/10.1186/1471-2105-11-345
77. Adey A, Kitzman JO, Burton JN, Daza R, Kumar A, Christiansen L, Ronaghi M, Amini S, Gunderson KL, Steemers FJ, Shendure J (2014) In vitro, long-range sequence information for de novo genome assembly via transposase contiguity. Genome Res. https://doi.org/10.1101/gr.178319.114
78. Yeo S, Coombe L, Chu J, Warren RL, Birol I (2017) ARCS: Assembly Roundup by Chromium Scaffolding. bioRxiv. https://doi.org/10.1101/100750
79. Warren RL, Yang C, Vandervalk BP, Behsaz B, Lagman A, Jones SJ, Birol I (2015) LINKS: scalable, alignment-free scaffolding of draft genomes with long reads. GigaScience 4:35. https://doi.org/10.1186/s13742-015-0076-3
80. English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC (2012) Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLoS One 7. https://doi.org/10.1371/journal.pone.0047768
81. Boetzer M, Pirovano W (2014) SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC Bioinformatics 15:211. https://doi.org/10.1186/1471-2105-15-211
82. Boetzer M, Pirovano W (2012) Toward almost closed genomes with GapFiller. Genome Biol 13(6):R56. https://doi.org/10.1186/gb-2012-13-6-r56
83. Paulino D, Warren RL, Vandervalk BP, Raymond A, Jackman SD, Birol I (2015) Sealer: a scalable gap-closing application for finishing draft genomes. BMC Bioinformatics 16(1):230. https://doi.org/10.1186/s12859-015-0663-4
84. Pacific Biosciences (2017) SMRT-Link. https://github.com/PacificBiosciences/SMRT-Link. Accessed 29 Sept 2017
85. Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ (2016) Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. Nucleic Acids Res 44(19):e147–e147. https://doi.org/10.1093/nar/gkw654
86. Huang S, Kang M, Xu A (2017) HaploMerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. Bioinformatics. https://doi.org/10.1093/bioinformatics/btx220
87. Pryszcz LP, Gabaldon T (2016) Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic Acids Res 44(12):e113. https://doi.org/10.1093/nar/gkw294
88. Hunt M, Kikuchi T, Sanders M, Newbold C, Berriman M, Otto TD (2013) REAPR: a universal tool for genome assembly evaluation. Genome Biol 14(5):R47. https://doi.org/10.1186/gb-2013-14-5-r47
89. Ghurye J, Pop M, Koren S, Bickhart D, Chin C-S (2017) Scaffolding of long read assemblies using long range contact information. BMC Genomics 18(1):527. https://doi.org/10.1186/s12864-017-3879-z
90. Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J (2013) Chromosome-scale scaffolding of de novo

genome assemblies based on chromatin interactions. Nat Biotechnol 31. https://doi.org/10.1038/nbt.2727
91. Smit A, Hubley R (2008–2015) RepeatModeler Open-1.0
92. Smit A, Hubley R, Green P (2013–2015) RepeatMasker open-4.0
93. Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18(1):188–196. https://doi.org/10.1101/gr.6743907
94. Korf I (2004) Gene finding in novel genomes. BMC Bioinformatics 5(1):59. https://doi.org/10.1186/1471-2105-5-59
95. Slater GSC, Birney E (2005) Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6:31. https://doi.org/10.1186/1471-2105-6-31
96. Papanicolaou A (2013) Just annotate my genome (JAMg) v. RC1. http://jamg.sourceforge.net/. Accessed 1 Oct 2017
97. Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res 33(Web Server issue):W465–W467. https://doi.org/10.1093/nar/gki458
98. Besemer J, Borodovsky M (2005) GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res 33(Web Server issue):W451–W454. https://doi.org/10.1093/nar/gki487
99. Majoros WH, Pertea M, Salzberg SL (2004) TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics 20(16):2878–2879. https://doi.org/10.1093/bioinformatics/bth315
100. Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, Cramer GR, Delledonne M, Luo C, Ecker JR, Cantu D, Rank DR, Schatz MC (2016) Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods 13(12):1050–1054. https://doi.org/10.1038/nmeth.4035
101. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. https://doi.org/10.1101/gr.215087.116
102. Xiao C-L, Chen Y, Xie S-Q, Chen K-N, Wang Y, Luo F, Xie Z (2016) MECAT: an ultra-fast mapping, error correction and de novo assembly tool for single-molecule sequencing reads. bioRxiv. https://doi.org/10.1101/089250
103. Jayakumar V, Sakakibara Y (2017) Comprehensive evaluation of non-hybrid genome assembly tools for third-generation PacBio long-read sequence data. Brief Bioinform:bbx147. https://doi.org/10.1093/bib/bbx147
104. R Development Core Team (2012) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
105. Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer, New York
106. Kent WJ (2002) BLAT—the BLAST-like alignment tool. Genome Res 12(4):656–664. https://doi.org/10.1101/gr.229202 Article published online before March 2002
107. Baxter L, Tripathy S, Ishaque N, Boot N, Cabral A, Kemen E, Thines M, Ah-Fong A, Anderson R, Badejoko W (2010) Signatures of adaptation to obligate biotrophy in the Hyaloperonospora arabidopsidis genome. Science 330. https://doi.org/10.1126/science.1195203
108. Edge P, Bafna V, Bansal V (2016) HapCUT2: robust and accurate haplotype assembly for diverse sequencing technologies. Genome Res. https://doi.org/10.1101/gr.213462.116
109. Putnam NH, O'onnell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, Haussler D, Rokhsar DS, Green RE (2016) Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res 26(3):342–350. https://doi.org/10.1101/gr.193474.115
110. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL (2016) Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst 3(1):95–98. https://doi.org/10.1016/j.cels.2016.07.002
111. Cao H, Hastie AR, Cao D, Lam ET, Sun Y, Huang H, Liu X, Lin L, Andrews W, Chan S, Huang S, Tong X, Requa M, Anantharaman T, Krogh A, Yang H, Cao H, Xu X (2014) Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology. GigaScience 3(1):34. https://doi.org/10.1186/2047-217x-3-34
112. Nagarajan N, Read TD, Pop M (2008) Scaffolding and validation of bacterial genome assemblies using optical restriction maps. Bioinformatics 24(10):1229–1235. https://doi.org/10.1093/bioinformatics/btn102
113. Neely RK, Deen J, Hofkens J (2011) Optical mapping of DNA: single-molecule-based methods for mapping genomes. Biopolymers 95(5):298–311. https://doi.org/10.1002/bip.21579

114. Fierst JL (2015) Using linkage maps to correct and scaffold de novo genome assemblies: methods, challenges, and computational tools. Front Genet 6(220). https://doi.org/10.3389/fgene.2015.00220
115. Nayaka SC, Shetty HS, Satyavathi CT, Yadav RS, Kishor PBK, Nagaraju M, Anoop TA, Kumar MM, Kuriakose B, Chakravartty N, Katta AVSKM, Lachagari VBR, Singh OV, Sahu PP, Puranik S, Kaushal P, Srivastava RK (2017) Draft genome sequence of Sclerospora graminicola, the pearl millet downy mildew pathogen. Biotechnol Rep 16(Suppl C):18–20. https://doi.org/10.1016/j.btre.2017.07.006
116. Fletcher K (2017) runHM2.sh [Software]. https://github.com/kfletcher88/HM2-RunLight. Accessed 14 Nov 2017
117. Morgulis A, Gertz EM, Schaffer AA, Agarwala R (2006) WindowMasker: window-based masker for sequenced genomes. Bioinformatics 22(2):134–141. https://doi.org/10.1093/bioinformatics/bti774
118. Reyes-Chin-Wo S, Wang Z, Yang X, Kozik A, Arikit S, Song C, Xia L, Froenicke L, Lavelle DO, Truco M-J, Xia R, Zhu S, Xu C, Xu H, Xu X, Cox K, Korf I, Meyers BC, Michelmore RW (2017) Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. Nat Commun 8:14953. https://doi.org/10.1038/ncomms14953
119. Peichel CL, Sullivan ST, Liachko I, White MA (2016) Improvement of the threespine stickleback (Gasterosteus aculeatus) genome using a Hi-C-based Proximity-Guided Assembly method. bioRxiv:068528
120. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, Aiden EL (2017) De novo assembly of the *Aedes aegypti*; genome using Hi-C yields chromosome-length scaffolds. Science 356(6333):92–95
121. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, Lee J, Lam ET, Liachko I, Sullivan ST, Burton JN, Huson HJ, Nystrom JC, Kelley CM, Hutchison JL, Zhou Y, Sun J, Crisa A, Ponce de Leon FA, Schwartz JC, Hammond JA, Waldbieser GC, Schroeder SG, Liu GE, Dunham MJ, Shendure J, Sonstegard TS, Phillippy AM, Van Tassell CP, Smith TPL (2017) Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. Nat Genet 49(4):643–650. https://doi.org/10.1038/ng.3802
122. Jiao W-B, Accinelli GG, Hartwig B, Kiefer C, Baker D, Severing E, Willing E-M, Piednoel M, Woetzel S, Madrid-Herrero E (2017) Improving and correcting the contiguity of long-read genome assemblies of three plant species using optical mapping and chromosome conformation capture data. Genome Res 27(5):778–786
123. Mohr DW, Naguib A, Weisenfeld N, Kumar V, Shah P, Church DM, Jaffe D, Scott AF (2017) Improved de novo genome assembly: linked-read sequencing combined with optical mapping produce a high quality mammalian genome at relatively low cost. bioRxiv:128348
124. Earl D, Bradnam K, St. John J, Darling A, Lin D, Fass J, Yu HOK, Buffalo V, Zerbino DR, Diekhans M, Nguyen N, Ariyaratne PN, Sung W-K, Ning Z, Haimel M, Simpson JT, Fonseca NA, Birol İ, Docking TR, Ho IY, Rokhsar DS, Chikhi R, Lavenier D, Chapuis G, Naquin D, Maillet N, Schatz MC, Kelley DR, Phillippy AM, Koren S, Yang S-P, Wu W, Chou W-C, Srivastava A, Shaw TI, Ruby JG, Skewes-Cox P, Betegon M, Dimon MT, Solovyev V, Seledtsov I, Kosarev P, Vorobyev D, Ramirez-Gonzalez R, Leggett R, MacLean D, Xia F, Luo R, Li Z, Xie Y, Liu B, Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Yin S, Sharpe T, Hall G, Kersey PJ, Durbin R, Jackman SD, Chapman JA, Huang X, DeRisi JL, Caccamo M, Li Y, Jaffe DB, Green RE, Haussler D, Korf I, Paten B (2011) Assemblathon 1: a competitive assessment of de novo short read assembly methods. Genome Res 21(12):2224–2241. https://doi.org/10.1101/gr.126599.111
125. Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I, Boisvert S, Chapman JA, Chapuis G, Chikhi R, Chitsaz H, Chou W-C, Corbeil J, Del Fabbro C, Docking TR, Durbin R, Earl D, Emrich S, Fedotov P, Fonseca NA, Ganapathy G, Gibbs RA, Gnerre S, Godzaridis É, Goldstein S, Haimel M, Hall G, Haussler D, Hiatt JB, Ho IY, Howard J, Hunt M, Jackman SD, Jaffe DB, Jarvis ED, Jiang H, Kazakov S, Kersey PJ, Kitzman JO, Knight JR, Koren S, Lam T-W, Lavenier D, Laviolette F, Li Y, Li Z, Liu B, Liu Y, Luo R, MacCallum I, MacManes MD, Maillet N, Melnikov S, Naquin D, Ning Z, Otto TD, Paten B, Paulo OS, Phillippy AM, Pina-Martins F, Place M, Przybylski D, Qin X, Qu C, Ribeiro FJ, Richards S, Rokhsar DS, Ruby JG, Scalabrin S, Schatz MC, Schwartz DC, Sergushichev A, Sharpe T, Shaw TI, Shendure J, Shi Y, Simpson JT, Song H, Tsarev F, Vezzi F, Vicedomini R, Vieira BM, Wang J, Worley KC, Yin S, Yiu S-M, Yuan J, Zhang G, Zhang H, Zhou S, Korf IF (2013) Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. GigaScience 2(1):10. https://doi.org/10.1186/2047-217X-2-10
126. Hunt M, Newbold C, Berriman M, Otto TD (2014) A comprehensive evaluation of assembly scaffolding tools. Genome Biol 15(3):R42.

https://doi.org/10.1186/gb-2014-15-3-r42
127. Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. Science 295(5558):1306
128. Beitel CW, Froenicke L, Lang JM, Korf IF, Michelmore RW, Eisen JA, Darling AE (2014) Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. PeerJ 2:e415. https://doi.org/10.7717/peerj.415
129. Burton JN, Liachko I, Dunham MJ, Shendure J (2014) Species-level deconvolution of metagenome assemblies with Hi-C–based contact probability maps. G3 (Bethesda) 4(7):1339–1346. https://doi.org/10.1534/g3.114.011825
130. Paulsen J, Sekelja M, Oldenburg AR, Barateau A, Briand N, Delbarre E, Shah A, Sørensen AL, Vigouroux C, Buendia B, Collas P (2017) Chrom3D: three-dimensional genome modeling from Hi-C and nuclear lamin-genome contacts. Genome Biol 18(1):21. https://doi.org/10.1186/s13059-016-1146-2
131. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159. https://doi.org/10.1016/j.cell.2014.11.021
132. Howe K, Wood JMD (2015) Using optical mapping data for the improvement of vertebrate genome assemblies. GigaScience 4(1):10. https://doi.org/10.1186/s13742-015-0052-y
133. Schwartz DC, Li X, Hernandez LI, Ramnarain SP, Huff EJ, Wang YK (1993) Ordered restriction maps of Saccharomyces cerevisiae chromosomes constructed by optical mapping. Science 262(5130):110–114
134. Ananiev GE, Goldstein S, Runnheim R, Forrest DK, Zhou S, Potamousis K, Churas CP, Bergendahl V, Thomson JA, Schwartz DC (2008) Optical mapping discerns genome wide DNA methylation profiles. BMC Mol Biol 9:68. https://doi.org/10.1186/1471-2199-9-68
135. Dong Y, Xie M, Jiang Y, Xiao N, Du X, Zhang W, Tosser-Klopp G, Wang J, Yang S, Liang J (2013) Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). Nat Biotechnol 31(2):135–141
136. Jiang N (2013) Overview of repeat annotation and de novo repeat identification. In: Peterson T (ed) Plant transposable elements: methods and protocols. Humana Press, Totowa, NJ, pp 275–287. https://doi.org/10.1007/978-1-62703-568-2_20
137. Campbell MS, Holt C, Moore B, Yandell M (2014) Genome annotation and curation using MAKER and MAKER-P. Curr Protoc Bioinformatics 48:4.11.11–14.11.39. https://doi.org/10.1002/0471250953.bi0411s48
138. Adhikari BN, Hamilton JP, Zerillo MM, Tisserat N, Lévesque CA, Buell CR (2013) Comparative genomics reveals insight into virulence strategies of plant pathogenic oomycetes. PLoS One 8(10):e75072
139. Levesque CA, Brouwer H, Cano L, Hamilton JP, Holt C, Huitema E, Raffaele S, Robideau GP, Thines M, Win J (2010) Genome sequence of the necrotrophic plant pathogen Pythium ultimum reveals original pathogenicity mechanisms and effector repertoire. Genome Biol 11. https://doi.org/10.1186/gb-2010-11-7-r73
140. Rujirawat T, Patumcharoenpol P, Lohnoo T, Yingyong W, Lerksuthirat T, Tangphatsornruang S, Suriyaphol P, Grenville-Briggs LJ, Garg G, Kittichotirat W, Krajaejun T (2015) Draft genome sequence of the pathogenic oomycete pythium insidiosum strain Pi-S, isolated from a patient with pythiosis. Genome Announc 3(3):e00574-00515. https://doi.org/10.1128/genomeA.00574-15
141. Mestre P, Carrere S, Gouzy J, Piron MC, Tourvieille de Labrouhe D, Vincourt P, Delmotte F, Godiard L (2016) Comparative analysis of expressed CRN and RXLR effectors from two Plasmopara species causing grapevine and sunflower downy mildew. Plant Pathol 65(5):767–781. https://doi.org/10.1111/ppa.12469
142. Sharma R, Xia X, Cano LM, Evangelisti E, Kemen E, Judelson H, Oome S, Sambles C, van den Hoogen DJ, Kitner M, Klein J, Meijer HJG, Spring O, Win J, Zipper R, Bode HB, Govers F, Kamoun S, Schornack S, Studholme DJ, Van den Ackerveken G, Thines M (2015) Genome analyses of the sunflower pathogen Plasmopara halstedii provide insights into effector evolution in downy mildews and Phytophthora. BMC Genomics 16(1):741. https://doi.org/10.1186/s12864-015-1904-7
143. Hall B, DeRego T, Geib S (2014) GAG: the genome annotation generator (version 1.0) [Software]. http://genomeannotation.github.io/GAG. Accessed 26 Oct 2017
144. Mondo SJ, Dannebaum RO, Kuo RC, Louie KB, Bewick AJ, LaButti K, Haridas S, Kuo A, Salamov A, Ahrendt SR, Lau R, Bowen BP, Lipzen A, Sullivan W, Andreopoulos BB, Clum A, Lindquist E, Daum C, Northen TR, Kunde-Ramamoorthy G, Schmitz RJ, Gryganskyi A, Culley D, Magnuson J, James TY, O'Malley MA, Stajich JE, Spatafora JW, Visel A, Grigoriev IV (2017) Widespread adenine N6-methylation of active genes in fungi.

Nat Genet 49(6):964–968. https://doi.org/10.1038/ng.3859
145. Flusberg BA, Webster D, Lee J, Travers K, Olivares E, Clark TA, Korlach J, Turner SW (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods 7(6):461–465. https://doi.org/10.1038/nmeth.1459
146. Rand AC, Jain M, Eizenga JM, Musselman-Brown A, Olsen HE, Akeson M, Paten B (2017) Mapping DNA methylation with high-throughput nanopore sequencing. Nat Methods 14(4):411–413. https://doi.org/10.1038/nmeth.4189
147. Fraser J, Ferrai C, Chiariello AM, Schueler M, Rito T, Laudanno G, Barbieri M, Moore BL, Kraemer DC, Aitken S (2015) Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. Mol Syst Biol 11(12):852
148. Olivares-Chauvet P, Mukamel Z, Lifshitz A, Schwartzman O, Elkayam NO, Lubling Y, Deikus G, Sebra RP, Tanay A (2016) Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. Nature 540(7632):296–300. https://doi.org/10.1038/nature20158
149. Smith DR (2017) Goodbye genome paper, hello genome report: the increasing popularity of 'genome announcements' and their impact on science. Brief Funct Genomics 16(3):156–162. https://doi.org/10.1093/bfgp/elw026



Chapter 14

BLASTmap: A Shiny-Based Application to Visualize BLAST Results as Interactive Heat Maps and a Tool to Design Gene-Specific Baits for Bespoke Target Enrichment Sequencing

Katie Baker, Gordon Stephen, Shona Strachan, Miles Armstrong, and Ingo Hein

Abstract

Numerous genes that determine the outcome of plant–pathogen interactions are currently being discovered and include, for example, immune receptors, susceptibility factors and pathogen effectors and their host targets. Target enrichment sequencing provides a means to preferentially resequence these genes of interest without the need to first generate a genotype-specific genome assembly. The Basic Local Alignment Search Tool (BLAST), in combination with the here developed BLASTmap, can be used to design probes that specifically target such gene(s), either by using the target species or the closest related genus as a reference. BLAST is a ubiquitous tool in biological sequence analysis and a multitude of programs are available for the visualization of BLAST alignments. However, there are currently no dedicated programs for visual comparison of large-scale BLAST output attributes such as bit score. The need to quickly and efficiently compare many thousands of BLAST results led to the development of BLASTmap, an interactive web application created using the Shiny R package, customized for clustering and viewing BLAST results as an interactive heat map. Here we show an example of how BLASTmap was successfully applied to analyze custom DNA/RNA probe sequences and to visually determine that four probes are sufficient for the specific yet inclusive enrichment of the potato *R2* disease resistance gene family.

Key words BLAST, Enrichment, Interactive map, Bait design, DNA sequence, Protein sequence

1 Introduction

An advantage of target enrichment sequencing over whole genome sequencing is that the former achieves sufficient read-depth of targeted genes to facilitate the accurate identification of sequence polymorphisms [1] and to uncover copy number variation [2]. As shown for nucleotide-binding and leucine-rich repeat containing disease resistance (*R*) genes, target enrichment sequencing (RenSeq) can aid the annotation of gene families [3], the mapping

Wenbo Ma and Thomas Wolpert (eds.), *Plant Pathogenic Fungi and Oomycetes: Methods and Protocols*, Methods in Molecular Biology, vol. 1848, https://doi.org/10.1007/978-1-4939-8724-5_14,
© Springer Science+Business Media, LLC, part of Springer Nature 2018

of resistance gene loci [3], the identification of candidate genes in combination with long-read sequencing [4, 5] and can function as a diagnostic tool [6]. Target enrichment is a sequence-homology-dependent means of reducing the genome complexity and to focus resequencing on genes of interest which are represented by user-defined probes. Typically, whole genome DNA is fragmented and hybridized with bespoke baits that represent the desired target sequences. The hybridization event can involve "on-array" capture whereby fragmented DNA is washed over baits bound to a solid surface, or "in-solution" capture where baits are mixed with the whole fragmented genome in the absence of a solid surface. In all cases, hybridization requires sufficient sequence homology (*see* **Note 1**) between the "baits" designed for the specific enrichment and the "targeted sequences" [7]. For capturing a large gene space (e.g., the entire NB-LRRs or effector complement of hosts or pathogens), baits that are biotinylated for hybridization-based enrichment can be commercially ordered. In contrast, for small-scale enrichments (e.g., to capture single genes or small gene families), bespoke probes can be designed in the laboratory with the help of this application. BLASTmap can be utilized to ensure that baits conform to the hybridization requirements whilst being unique for the target(s). They can then be synthesized via, for example, reverse transcriptase run-offs using T7 or Sp6 sites that are incorporated into most cloning vectors [8]. Here we show how the Basic Local Alignment Search Tool (BLAST) can be used to design gene specific baits for target enrichment sequencing in combination with BLASTmap.

BLAST has been invaluable for DNA/protein sequence analysis since its inception in 1990 [9]. Many programs have been developed for the interpretation and visualization of BLAST alignments [10]. However, visualization of alignments alone is not ideal when thousands of BLAST results must be examined and compared. There are no programs specifically designed for visual comparisons of BLAST attributes, such as bit score, and analyses often rely on parsing text-based or XML BLAST output. Heat maps provide an intuitive way to visualize biological data with patterns easily discerned by color and clustering [11]. No dedicated tools for heat map visualization of BLAST results have been published, yet they would represent an effective way to compare the results of a large BLAST analysis.

Many programs have been developed to visualize multivariate biological data. For example HeatmapGenerator [12] produces static images, but interactive navigation and filtering of heat maps is not possible. HeatMapViewer [13] can produce interactive heat maps, but as a BioJS component, it requires substantial time to set up before use. ClustVis [14] and MORPHEUS (https://software.broadinstitute.org/morpheus/) are both specialized for matrix visualization of genomic data and can generate interactive heat

maps. However, BLAST output requires manual reformatting before meaningful heat map visualization in MORPHEUS, ClustVis, HeatmapGenerator, and HeatMapViewer.

Here, BLASTmap is introduced as a Shiny app which can produce customizable interactive BLAST heat maps with a user-friendly layout. BLASTmap can be run from the web and heat maps are produced shortly after importing the standard BLAST output with no software setup required by the user.

2 Materials

2.1 BLASTmap

It is written in R [15], a popular programming language and hosted online at https://ics.hutton.ac.uk/blastmap/. The source code is freely available from https://github.com/kb-bioinf/BLASTmap and distributed under the MIT licence.

2.2 Input Required for BLASTmap

A standard tab- or comma-delimited BLAST output. If accessed online, there are no further requirements. BLASTmap can also be run from the R command line or RStudio [16], a popular R interactive development environment.

2.3 R Packages Required for BLASTmap

1. *shiny* (the R web framework).
2. *shinythemes* for customization of the user interface.
3. *d3heatmap* for creating the interactive heat map (*see* **Note 2**).
4. *htmlwidgets* for exporting the heat map as an offline HTML page.
5. *gplots* for creating a static heat map image which can be exported.

3 Methods

3.1 User Interface

1. There are five tabs in the BLASTmap user interface (UI) and they all take the general form of a left sidebar with options and a main panel for data display (*see* **Note 3**).
2. BLAST data is imported in the "Import" tab and can be prefiltered using the "Heat map options" tab (*see* **Note 4**) if required. The app was successfully tested on 29,358 BLAST query–hit pairs (corresponding to a matrix size of 45,630). Larger data sets (>50,000 matrix cells) require prefiltering, which is available on the "Import" tab (*see* **Note 5**).
3. In the "Interactive heat map" tab the imported data is visualized using the *d3heatmap* package. BLASTmap output is customizable by a number of parameters (Fig. 1a).

(a) Firstly, the BLAST values to visualize must be chosen via a drop-down menu. Best BLAST-only visualization can be specified using a check box, in which case only one hit per query appears in the heat map.

(b) 3.2 The data set can be filtered using sliders to select specific bit scores, alignment lengths, percentage identities (% ID), e-values, gaps, mismatches, query, and hit positions. Particular query and hit names can also be selected using drop-down boxes or by importing text files. Heat maps are rendered based on the filtering parameters chosen, so clustering can change depending on which query–hit pairs are retained.

4. The features of *d3heatmap* are usable by BLASTmap and so the heat map is interactive. The heat map can be zoomed into by dragging a box around the desired cells and the chosen BLAST attribute is displayed via tooltip on mouse over.
5. BLASTmap uses the color-blind friendly blue and red" Spectral" color palette from the *RColorBrewer* package. Further customization is available by disabling row and/or column dendrograms, increasing the height of the heat map and increasing or decreasing the label margins.
6. In the "Export" tab graphics and tables can be downloaded (Fig. 1b).
7. The interactive heat map can be exported as an offline HTML page which retains the interactive elements of *d3heatmap* (*see* **Note 6**).
8. The static heat map can be downloaded in JPEG, PNG, or TIFF format and the data table can be exported in CSV or TSV format.

(a) displays the interactive heat map panel with BLAST filtering and user-modifiable display options on the left sidebar. The heat map reveals the (dis)similarity of the query sequences on the y-axis compared to the subject sequences on the x-axis. The y-axis depicts all putative probe sequences which were designed based on the *R2* gene family of potato. Each individual probe corresponds to 400 bp of a specific *R2* family member whilst overlapping with adjacent probes by 50 bp. The target genes (x-axis) contain members of the *R2* gene family. Fragmented putative bait sequences (y-axis) are compared to full length target sequences (x-axis) and pairs which are dissimilar are shaded blue and increase to a red shading when sequences are highly similar.

(b) shows the export options within the static heat map display options on the left sidebar and an image preview in main panel. Probes selected from display A) and derived from separate regions of the target genes are compared against the *R2* gene family (x-axis). The results reveal that, in this

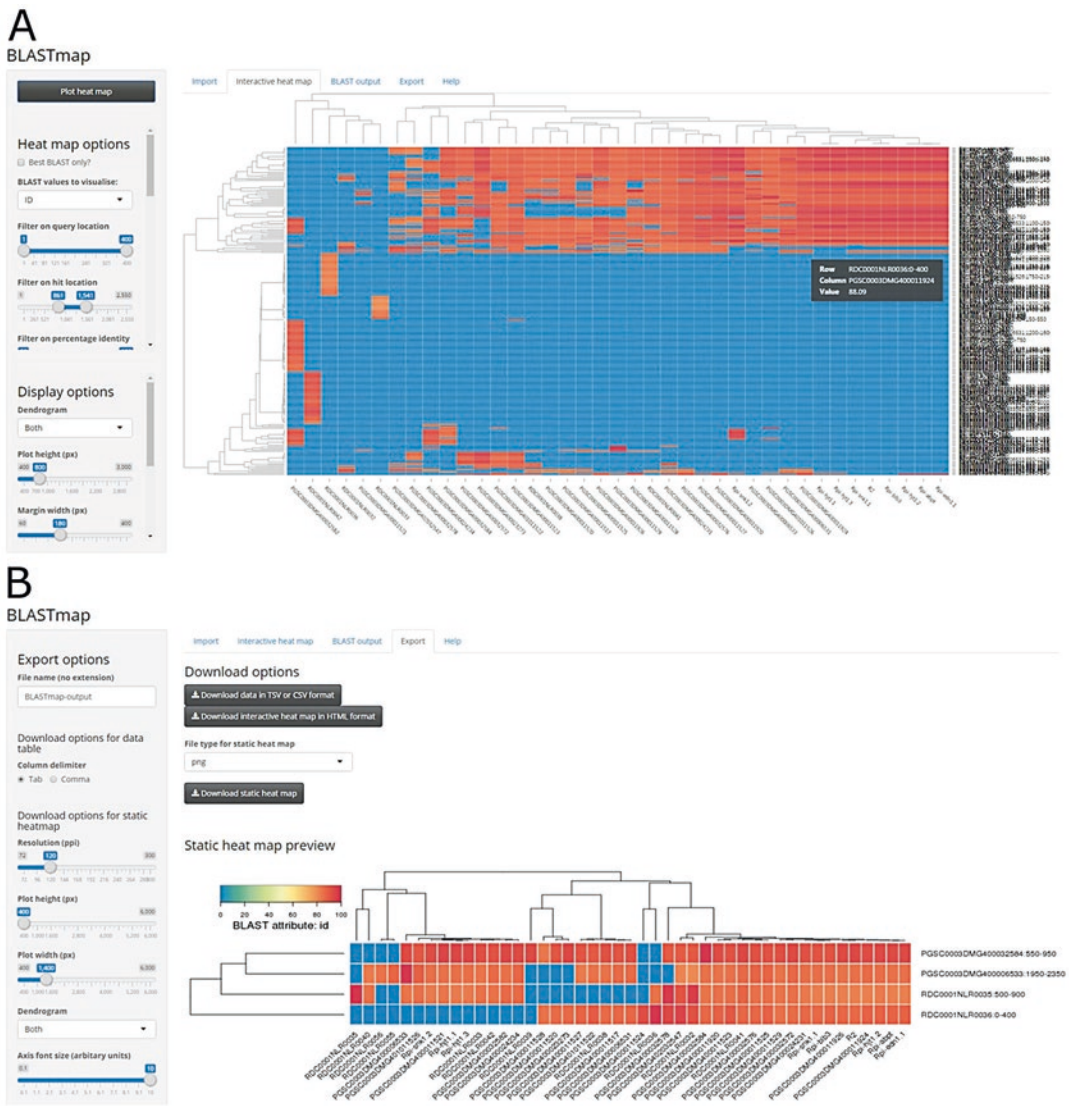


Fig. 1 Example of using Blastmap to identify the minimum number of probes required for targeted enrichment of the *R2* disease resistance gene family in potato

case, four probes are sufficient to hybridize and enrich for members of the *R2* gene family. Dendrograms show clustering of genes within the target and bait gene family (*see* **Note 7**).

3.2 Example Biological Application—Selection of DNA Probes for Targeted Enrichment Sequencing

1. DNA probes were designed against 36 previously described *Solanum* disease resistance genes with sequence homology to *R2* [17] by extracting the DNA sequence in a sliding windows of 400 bp with a 50 bp overlap.
2. These probe sequences were then aligned back to the same genes in an in silico hybridization using MEGABLAST via the

BLAST+ command line application [18] (v2.2.28) with default settings.

3. The tabular output was imported into BLASTmap for visualization (Fig. 1a).
4. Using the built- in heat map options, the query–hit pairs were filtered for a minimum % ID of 95% and an alignment length of 390–405 bp to account for gaps and insertions.
5. By inspecting the heat map output, four probe sequences were identified that are sufficient in length and % ID to be utilized for targeted enrichment sequencing (Fig. 1b) (*see* **Note 8**).

3.3 Conclusion

Visualization of BLAST results in heat map form is an underutilised way to display large volumes of BLAST data efficiently. BLASTmap was specifically designed for interactive exploration of BLAST data and utilizes the R package *d3heatmap* to produce heat maps which can be navigated and filtered dynamically. BLASTmap has advantages over existing heat map software for visualizing BLAST output, it requires no file formatting and performs clustering to easily discern patterns in the data. Here we have shown an example of how BLASTmap was successfully applied to design custom DNA/RNA probe sequences and to visually determine the minimum number of probes required for the specific enrichment of a large and diverse gene family. This approach can be applied to any gene family of interest across diverse taxa.

4 Notes

1. Typically, a minimum of 80% sequence identity between baits and targets is required.
2. Clustering is also possible using the default *d3heatmap* clustering method which is hierarchical clustering with Euclidean distance. For the data to be clustered and rendered correctly, missing query–hit pairs are given a value of zero. In cases where a query has multiple alignments to the same hit, the best query–hit pair, defined by highest bit score, is retained. The "heatmap.2" function from the *gplots* package is used to create a static heat map with the same parameters as the interactive heat map.
3. There is a "Help" tab with detailed instructions for using BLASTmap.
4. The input required is tab- or column-delimited BLAST output in the form: Query name, hit name, percentage identity, alignment length, mismatches, gaps, query start, query end, hit start, hit end, e-value, bitscore.

5. If the data set meets the matrix size limitation (50,000 cells), a heat map may be generated with the data. However, if the data exceeds this, the option to filter data sets is provided. The data may be filtered on query/hit names by selecting names from a drop-down box and/or importing files with a list of query/hit names. In addition to filtering by query/hit name, BLAST output attributes may be used to filter the dataset by using sliders to select thresholds.
6. In addition to the interactive heat map output, a static heat map created with the same parameters and an accompanying table of the BLAST data displayed in the heat map are available in other tabs of the UI.
7. The queries are on the y-axis and hits are on the x-axis. The heat map is interactive, so can be zoomed into by drawing a box around the cells. Selecting a hit or query highlights that row or column. Hovering with a cursor over a cell displays the query, hit name, as well as the value of the BLAST attributes.
8. There is an option to visualize only the best BLAST pairs; the default is to visualise all BLAST hits for each query. Each BLAST attribute can be filtered on, using the sliders to select ranges. Only BLAST pairs within these ranges will be visualized. Query and hit sequences can also be selected by name, either using drop-down box or uploading a files with a list of query/hit names (one name per line).

Acknowldgements

The authors would like to thank Iain Milne for advice on proxy configuration, Gaetan Thilliez for providing test BLAST data, and Peter Cock and Micha Bayer for feedback on the application.

References

1. Parla JS, Iossifov I, Grabill I, Spector MS, Kramer M, McCombie WR (2011) A comparative analysis of exome capture. Genome Biol 12:R97
2. Lighten J, van Oosterhout C, Paterson IG, Mcmullan M, Bentzen P (2014) Ultra-deep Illumina sequencing accurately identifies MHC class IIb alleles and provides evidence for copy number variation in the guppy (Poecilia reticulata). Mol Ecol Resour 14:753–767
3. Jupe F, Witek K, Verweij W, Sliwka J, Pritchard L, Etherington GJ et al (2013) Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. Plant J 76:530–544
4. Giolai M, Paajanen P, Verweij W, Percival-Alwyn L, Baker D, Witek K et al (2016) Targeted capture and sequencing of gene sized DNA molecules. BioTechniques 61:315–322
5. Witek K, Jupe F, Witek AI, Baker D, Clark MD, Jones JDG (2016) Accelerated cloning of a potato late blight–resistance gene using RenSeq and SMRT sequencing. Nat Biotechnol 34:656–660
6. Van Weymers PSM, Baker K, Chen X, Harrower B, Cooke DEL, Gilroy EM et al (2016) Utilizing "Omic" technologies to identify and

prioritize novel sources of resistance to the oomycete pathogen Phytophthora infestans in potato germplasm collections. Front Plant Sci 7:672

7. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A et al (2010) Target-enrichment strategies for next-generation sequencing. Nat Methods 7:111–118
8. Melton DA, Krieg PA, Rebagliati MR, Maniatis T, Zinn K, Green MR (1984) Efficient in vitro synthesis of biologically active RNA and RNA hybridisation probes from plasmids containing a bacteriophage SP6 promotor. Nucl Acids Res 12:7035–7056
9. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
10. Neumann RS, Kumar S, Shalchian-Tabrizi K (2014) BLAST output visualization in the new sequencing era. Brief Bioinform 15:484–503
11. Gehlenborg N, Wong B (2012) Points of view: into the third dimension. Nat Methods 9:851–851
12. Khomtchouk BB, Van Booven DJ, Wahlestedt C (2014) HeatmapGenerator: high performance RNAseq and microarray visualization software suite to examine differential gene expression levels using an R and C++ hybrid computational pipeline. Source Code Biol Med 9:30
13. Yachdav G, Hecht M, Pasmanik-Chor M, Yeheskel A, Rost B (2014) HeatMapViewer: interactive display of 2D data in biology. F1000Research 48:1–6
14. Metsalu T, Vilo J (2015) ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. Nucleic Acids Res 43:566–570
15. R Core Team (2015) R: A language and environment for statistical computing. Available from: http://www.r-project.org/
16. R Studio Team (2015) RStudio: Integrated Development for R. Available from: http://www.rstudio.com/
17. Destefanis M, Nagy I, Rigney B, Bryan GJ, McLean K, Hein I et al (2015) A disease resistance locus on potato and tomato chromosome 4 exhibits a conserved multipartite structure displaying different rates of evolution in different lineages. BMC Plant Biol 15:255
18. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K et al (2009) BLAST+: architecture and applications. BMC Bioinformatics 10:421



Chapter 15

A Computational Protocol to Analyze Metatranscriptomic Data Capturing Fungal–Host Interactions

Yong Zhang, Li Guo, and Li-Jun Ma

Abstract

Plant diseases cause significant losses to agricultural production and pose serious threats to food security worldwide. Understanding the mechanism of host–pathogen interaction is essential for the development of novel diagnostic methods and disease management strategies. RNA sequencing (or RNA-Seq) technology enables a global characterization and quantification of all transcripts of organisms from which RNA can be obtained, and it is particularly useful in identifying pathogen virulence factors involved in disease development and host immunity involved in the development of resistance. This chapter describes a computational protocol to manage, analyze and interpret RNA-Seq data. We have included two transcriptome analysis approaches, one reference-guided and the other de novo assembly-based, and discuss pros and cons for each method. We have also presented visualization methods to generate high quality figures as well as data mining strategies for identifying candidate genes/pathways involved in host immunity and pathogen virulence. In summary, this protocol captures the fungal-plant interactions at the transcriptional level and facilitates rapid gene discovery and expression analysis using next-generation sequencing data of mixed host and pathogen transcripts (i.e., metatranscriptomics). All bioinformatic tools used to build this protocol are publically available, and we strove to make them accessible to researchers with limited computational skills and applicable to metatranscriptomic data analysis in a wide range of plant–fungal interactions.

Key words RNA-Seq, Transcriptomics, Metatranscriptomics, Host–pathogen interactions, Differential gene expression

1 Introduction

The destructive effects of plant pathogens exert widespread consequences on human society. The Irish potato famine in the 1840s [1] and the Bengal famine of the 1940s [2] are two such examples. A variety of plant pathogens continue to afflict agricultural production today, as evidenced by the recent outbreak of Panama disease of banana spreading from South-East Asia into Mozambique and Jordan [3] and downy mildew of sweet basil in the USA [4]. To control plant

Wenbo Ma and Thomas Wolpert (eds.), *Plant Pathogenic Fungi and Oomycetes: Methods and Protocols*, Methods in Molecular Biology, vol. 1848, https://doi.org/10.1007/978-1-4939-8724-5_15,
© Springer Science+Business Media, LLC, part of Springer Nature 2018

diseases, it is essential to understand pathogenicity mechanisms and identify gene candidates associated with host–pathogen interactions. RNA sequencing (RNA-Seq) offers an unprecedented opportunity to completely characterize and quantify the active transcriptome involved in the development of disease [5]. Numerous powerful bioinformatic tools and programs have been developed to take advantage of this technology to successfully explore the functional elements of the genome and reveal molecular constituents involved in disease development [6, 7]. In this chapter, we present a protocol for RNA-Seq data analysis and discuss strategies in experimental design, library construction, choice of programs, and data interpretation for the purpose of achieving high-quality end results. Two data analysis pipelines, a reference-guided and a de novo assembly-based, for transcriptome analysis are included (Fig. 1). To ensure accessibility for most biological research laboratories, we have selected freely available bioinformatic packages.

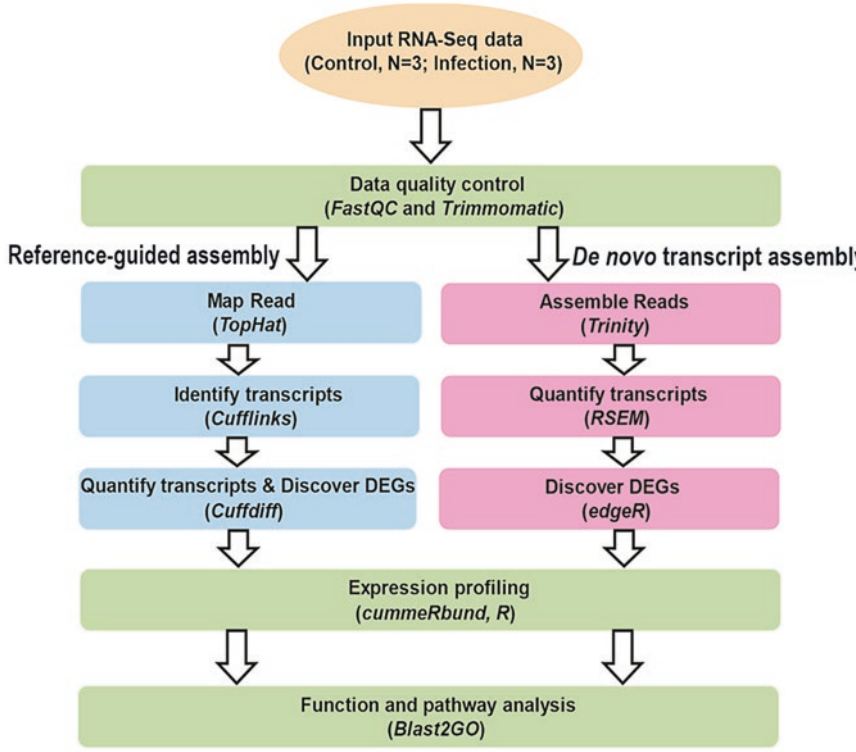


Fig. 1 Overview of RNA-Seq computational analysis workflow. Two RNA-Seq analysis pipelines, reference-guided and de novo assembly-based, are highlighted with light blue and pink, respectively. The steps highlighted with light green are commonly shared by the two pipelines. This chapter assumes two experimental conditions, control and infection, with three replicates each as an example to run the pipelines

2 Materials

2.1 Hardware and Software

1. *HPCC*: high-performance computer cluster, a remote server with the Linux operating system.
2. *PuTTY* (*http://www.putty.org*): a secure shell (SSH) client that supports several network protocols, including remote login, file copy, and transfer between local computer and remote HPC server environments. Other alternative SSH client software are listed in Table 1.
3. *FileZilla* (*http://filezilla-project.org/*): a cross-platform graphical FTP, SFTP, and FTPS file management tool for secure file transfer between local computer and remote servers.
4. *FastQC* (*https://www.bioinformatics.babraham.ac.uk/projects/fastqc/*): a quality control tool for high-throughput sequence data [8].
5. *Trimmomatic* (*http://www.usadellab.org/cms/?page=trimmomatic*): a read trimming tool for Illumina sequence data [9].
6. *Samtools* (*http://www.htslib.org/*): a suite of programs for Sequence Alignment/Map (SAM). The program can read/write/edit/index/view high-throughput sequence data in multiple file formats. It can also call/filter/summarize SNP and short indel sequence variants [10].
7. *Bowtie2* (*http://bowtie-bio.sourceforge.net/bowtie2/index.shtml*): an ultrahigh-throughput and memory-efficient tool for aligning sequencing reads to long reference sequences [11].
8. *TopHat* (*http://ccb.jhu.edu/software/tophat/index.shtml*): a fast, splice junction mapper for RNA-Seq reads. It aligns RNA-Seq reads to large (eukaryotic multicellular) genomes using short read aligner *Bowtie*, and then analyzes the mapping results to identify splice junctions between exons [12].

Table 1
Selected popular SSH clients

Clients	URL
PuTTY	http://www.putty.org
WinSCP	https://winscp.net/eng/docs/free_ssh_client_for_windows
SuperPuTTY	https://github.com/jimradford/superputty
KiTTY	http://www.9bis.net/kitty/
Bitvise	https://www.bitvise.com/ssh-client
OpenSSH	http://www.openssh.com/

9. *Cufflinks* (*https://github.com/cole-trapnell-lab/cufflinks*): a program accepts aligned RNA-Seq reads, assembles the alignments into a parsimonious set of transcripts, estimates their abundances, and tests for differential expression and regulation in RNA-Seq samples [13].
10. *Trinity* (*https://github.com/trinityrnaseq/trinityrnaseq/wiki*): a novel method for the efficient and robust de novo reconstruction of transcriptomes from RNA-Seq data [14].
11. *edgeR* (*https://bioconductor.org/packages/release/bioc/html/edgeR.html*): a program for differential expression analysis of digital gene expression data [15].
12. *RSEM* (*https://github.com/deweylab/RSEM*): a program for estimating the expression level of genes and isoforms [16].

2.2 Connecting a Local Computer to a Remote Server or HPCC

One of the foremost challenges in RNA-Seq analysis is the large quantity of data. Therefore, it is essential to establish a reliable connection from your particular environment to a remote server with such a capacity when conducting analyses of large data sets. Most institutions have an established HPCC or provide access to such a server. Of course, there are multiple third-party cloud computing platforms, including Amazon Elastic Compute Cloud, Google App Engine, Microsoft Azure, and the IBM SmartCloud. To establish the connection, users need to:

1. Understand the operating system of a local computer (Mac OS or Windows).
2. Gain access to a remote server (server IP address, connecting port number, user name and password) (*see* **Note 1**).
3. Access to the server using SSH:

   On Mac OS

   (a) Open up a terminal;

   (b) Type in the following command based on login information:

   ssh –p [port number] [account name]@[server IP address]

   (c) Press *Enter*;

   (d) Terminal will prompt you to enter *password*.

   On Windows

   (a) Open up *PuTTY* and the dialog box as shown in Fig. 2a will pop up;

   (b) Type the [*account name*]@[*server IP address*] and *port number* in the host name box and port box, respectively;

   (c) Select *SSH* under *Connection type* to access the server;

   (d) Once all required fields have been filled, click on *open*;

   (e) When the dialog box at the bottom in Fig. 2a pop up, enter *password* and press *enter*;

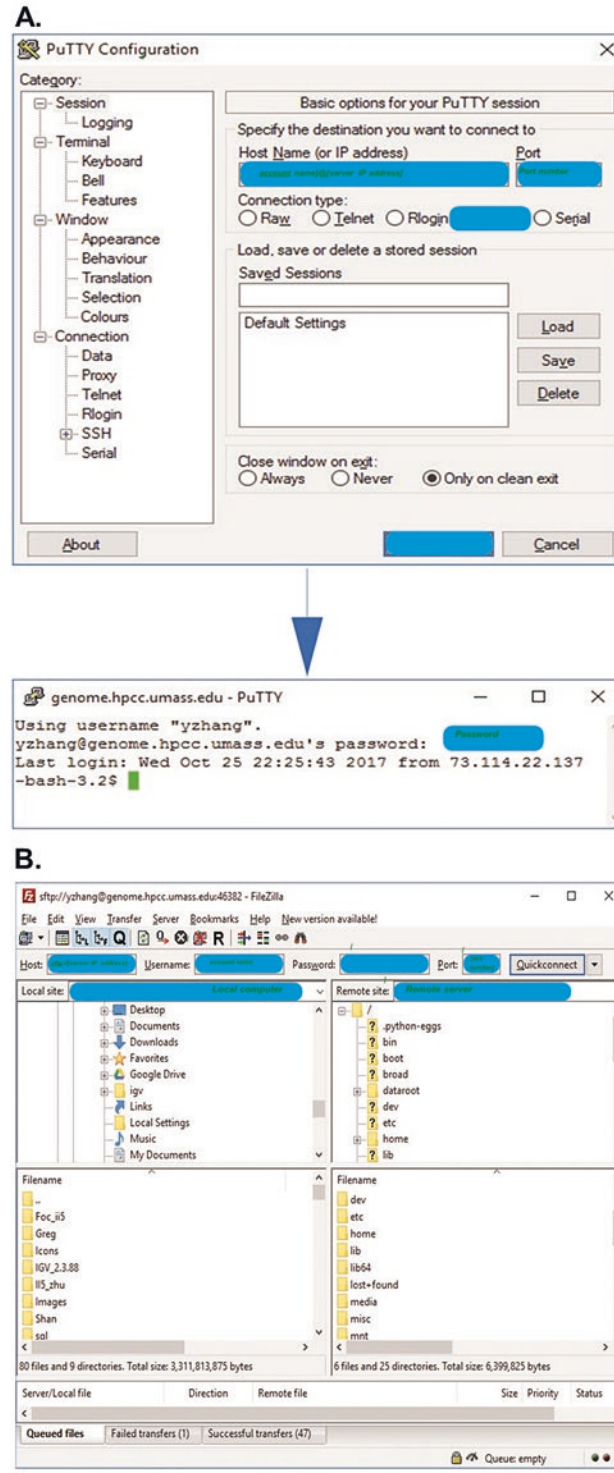


Fig. 2 Connecting to a remote server and setting up the file transferring process. (**a**) The ***PuTTY*** configuration dialog box. After filling the required information highlighted with blue, the login dialog box at the bottom will pop up to ask for the password. (**b**) The *Filezilla* dialog box. *Local site* and *Remote site* will appear after filling in the required information besides *Quickconnect*

(f) If it is your first time using *PuTTY*, press *yes* when a message appears asking you if you want to continue connecting.

2.3 Transferring Data Between a Local Computer and a Server or HPCC (Fig. 2b)

1. Open up *FileZilla*, on the top of *FileZilla* near the *Quickconnect* button, type the following in each box:

   Host: *sftp://*[*server IP address*]

   Username: [*account name*]

   Password: [*password*]

   Port: [*port number*]

2. Click the *Quickconnect* button;
3. On the left hand side of the dialog box, navigate to the directory where you wish to upload/download data in the pane right below *Local site*;
4. On the right hand side of dialog box, left click on the slash directly under the *Remote site* pane. Find the directory where you wish to move sequencing data to/from the remote server and double click;
5. *Uploading data from your computer*: Once the above steps have been completed, right click on the files in the left hand pane on your computer that you wish to transfer to the server, then left click on the *Upload* button; or *Downloading data from the server*: Once the above steps have been completed, right click on the files on the right hand pane from the server that you wish to transfer to your computer, then left click the *Download* button.
6. You should see your data that are stored on the server.

2.4 Practice Command Line with Command Line Interface (CLI)

Below are several most commonly used commands:

1. List contents of directory: *ls*
2. Change directory with a specific path: *cd /path/directory*
3. Copy file(s): *cp /path/file1 /path/file2*
4. Create a directory: *mkdir directory*
5. Delete file(s): *rm file*
   Additional commands are listed in Table 2.

2.5 Basic Software Installation

Many HPCC facilities have the most commonly used bioinformatic packages installed and tested. If that is the case, you may skip this section, but need to know where to access these tools and test the performance of each program with a testing dataset.

1. Upload downloaded software packages into remote server or HPCC via *FileZilla*.
2. Unzip each software package.

Table 2
List of basic Linux commands

Linux command	Description	Linux command example
cd	Change directory with a specified path	cd */path/directory1*
clear	Clear the screen	clear
cp	Copy file(s)	cp */path1/file1 /path2/file1*
diff	Compare the contents of files	diff *file1 file2*
exit	Log out of Linux	exit
grep	Find a string of text in a file	grep "word or phrase" *file1*
head	Display beginning of a file	head *file1*
less	View a file	less *file1*
ls	List contents of a directory	ls */path/directory1*
mv	Move file(s) or rename file(s)	mv */path1/file1 /path2/file2*
mkdir	Create a directory	mkdir *directory*
rm	Delete file(s)	rm *file1*
rmdir	Remove a directory	rmdir *directory*
tail	Display end of a file	tail *file1*
tar	Store, list or extract files in an archive	tar *file1*
vi	Edit file(s) with simple text editor	vi *file1*

tar -zxvf [xxx.tar.gz]

FastQC, *Trimmomatic*, and *Bowtie2* are ready to use after unzipping.

3. To install *Samtools*, change directory to the package directory:

cd [package directory]

Type the following commands:

./configure --prefix=/path/to/samtools/install

make

make install

"*./configure*": checks your build environment and requirements.

"*–prefix*": specifies where to install the software. If not specified, the software will be installed under */usr/local* directory. Other configure options such as "specify further optional external requirements" can be found with "*-help*".

4. Repeat the **step 3** for *TopHat*, *Cufflinks*, and *RSEM*.

For *Trinity* and *RSEM*, just two steps: *make* and *make install.*

5. Try a small sample data set provided within the package to test your installation.

2.6 Sample Preparation

Sample preparation is not the main focus of this chapter. However, there are a few key components we would like to bring to the readers' attention.

1. Experimental design: Define sample treatments based on biological questions and the knowledge of the host–pathogen interaction system under study. Our experiences with both a root pathogen system involving *Fusarium oxysporum* and a foliar pathogen system involving the obligate downy mildew parasite *Peronospora belbahrii* suggest that early infection time points during pathogens penetration capture the most critical biological processes important for the dissection of the host–pathogen interaction. You should have control samples that mirror your experimental conditions. Variables associated with plant growth, such as developmental stage, physiological condition, and circadian cycle of the plants, need to be considered as well. We also suggest a minimum of three biological replicates for each treatment (*see* Subheading 4). Adding synthetic external RNA as spike-in controls is also a valuable approach to evaluate sensitivity and accuracy of RNA-seq experiments.
2. Sequence depth and coverage: Sequence depth is determined by the goal of the RNA-Seq project. Because RNA-Seq is a global and unbiased sampling process, in principle we would agree that the larger the sample size (more sequencing reads), the better the data in capturing the dynamics of host–pathogen interaction. For instance, at a low sequencing coverage, only the most abundant transcripts will be detected. Of course, the increased sequencing coverage will increase the predictive accuracy of differentially expressed genes (DEGs). However, it is expensive to generate sequence reads, and it is equally computational expensive, if not more, to process sequencing data. Based on an extensive survey of the yeast genome, ~20× sequence coverage of a transcriptome will return about 80% all transcripts [17]. 20× coverage of the yeast genome means about two million Illumina paired-end sequence reads, and you can calculate your sequencing amount accordingly. Also, a saturation curve can be used to assess the relationship between transcriptome coverage at a given sequencing depth [18]. Moreover, the optimal sequencing depth of each project is determined by the complexity of the targeted transcriptome. For example, since single cell sequencing has more limited sample complexity than whole tissue sequencing, usually one million reads are enough for quantification analysis in single cell RNA-Seq rather than the at least five million reads for whole tissues [19, 20]. To investigate host–pathogen interactions, higher coverages should be considered for inoculated samples because: (1) these samples have high complexity as they contain both the pathogen and the host transcripts; and (2) these samples capture genes that play critical roles in the

host–pathogen interaction, but may not be highly abundant in the pool of RNA species.

3. Sample preparation and sequencing: most standard RNA extraction and cDNA synthesis protocols should work for preparing cDNA for sequencing. Due to the dynamic nature of transcription, avoiding delay between sample collection and RNA extraction is recommended. If it is impossible to process the samples in a timely manner, it is acceptable to immediately freeze the samples for future processing. We have successfully used TRIzol reagent (Thermo Fisher Scientific, Waltham, MA, USA) and the Spectrum™ Plant Total RNA Kit (Sigma Aldrich) for RNA extraction from leaf tissues. For soil-containing root samples, we would recommend the Zymo Research Soil/Fecal RNA MicroPrep™ or other similar kits to reduce RNA degradation caused by contaminants in the soil (*see* **Note 2**). To avoid potential contamination caused by genomic DNA, DNase I digestion is included before the reverse transcription step. Companies such as Illumina and New England Biolabs offer RNA-Seq library prep kits. It is also advisable to discuss with the sales representatives about your desirable coverage and the sequencing platform that will be used to generate your data. In general, single-end sequencing is enough for differential expression RNA-Seq analysis in host–pathogen interactions. However, paired-end reads are helpful for improving mapping accuracy and discovering transcript structure and alternative splicing.

3 Methods

The analysis of RNA-Seq data consists of several steps and could be done in different ways. In this section, we will discuss the major steps used for RNA-Seq analysis, including sequencing data quality control, transcript identification and quantification, DEGs detection, and approaches for an accurate interpretation of the results. Two different strategies of transcript identification are addressed. They are reference-guided identification and de novo transcript assembly (Fig. 1).

For this demonstration, we use *T1* and *T2* to represent the control group and infection group, respectively. *R1*, *R2*, and *R3* are three biological replicates. *XX_1.fq* and *XX_2.fq* refer to left (forward) and right read (reverse) data files, respectively. The command */path* defines the pathway to the directory that stores the software you want to use.

3.1 Sequencing Data Quality Control

Steps involved in sample prep and data generation, such as RNA fragmentation, cDNA synthesis, adapter ligation, PCR amplification, bar-coding, and sample loading, could all introduce biases and errors in the sequencing data and interfere with the downstream analysis. Therefore, data quality control is the critical first step.

1. Access sequence quality: facing a large quantity of sequence data, it is not feasible to inspect reads individually. However, it is important to evaluate the integrity of the data using some global parameters, such as GC content, *k-mer* distribution, and frequency of duplication. These values might vary from experiment to experiment, but they should be homogeneous for samples in the same experiment. *FastQC* is a popular tool providing all the necessary parameters to check sequencing data quality, for example:

```
/path/fastqc T1-R1_1.fq T1-R1_2.fq
/path/fastqc T1-R2_1.fq T1-R2_2.fq
/path/fastqc T1-R3_1.fq T1-R3_2.fq
/path/fastqc T2-R1_1.fq T2-R1_2.fq
/path/fastqc T2-R2_1.fq T2-R2_2.fq
/path/fastqc T2-R3_1.fq T2-R3_2.fq
```

Here: /path/fastqc specifies where the software *FastQC* is stored. The command will be run six times for three biological replicates (*R1*, *R2*, and *R3*) for two experimental treatments (*T1* and *T2*). After this process, the quality for both forward (*XX_1.fq*) and reverse reads (*XX_2.fq*) will be generated in one of the six *XXX_fastqc* folders, from *T1-R1_fastqc to T2-R3_fastqc*. Each folder contains a *summary.txt* file with a data quality summary. Also, you could open the *fastqc_report.html* file in a web browser to view the data summary. A check mark next to the checking parameter indicates good quality. The exclamation mark means acceptable. However, if you see a cross, it means there is a problem with the data and you need to pay attention to it.

2. Remove bad reads and trim poor-quality bases: most of the time, data problems can be solved by removing reads of low quality and trimming the poor-quality bases of a read. For Illumina sequencing technology, the base quality of a read decreases toward the 3′ end. Sequence trimmers, such as *Trimmomatic* [9], is frequently used to remove poor-quality bases and low-quality reads as well as sequencing adaptors at the beginning of reads. Of course, *FastQC* can be used to reevaluate the data quality (*see* **Note 3**).

   Here is an example:

```
java -jar /path/trimmomatic-xxx.jar PE -phred33 T1-R1_1.fq T1-
R1_2.fq Trimmed_T1-R1_1.fq Unpaired_ T1-R1_1.fq Trimmed_T1-
R1_2.fq Unpaired_T1-R1_2.fq LEADING:20 TRAILING:20 MINLEN:70
```

This command will run *Trimmomatic* over the forward and reverse reads for *T1-R1* sample, which will remove leading and trailing read bases with a quality below 20 and keep the reads with a trimmed length larger than 70. Other options you might want to use for your own data can be found in the software manual. The same command will need to be run repeatedly for *T1-R2, T1-R3, T2-R1, T2-R2, and T2-R3*. The command specifies input data as paired-end read (PE) and produces four output files: two files

-Trimmed_XXX_1.fq and *Trimmed_XXX_2.fq-* store the PE reads that have passed the trimming parameter process; two files *-Unpaired_XXX_1.fq* and *Unpaired_XXX_2.fq-* contain the unpaired reads that have passed the trimming parameter but the partner read.

3.2 Reference-Guided Transcript Identification

When a reference genome is available, the most commonly used strategy for RNA-Seq data analysis is to map the reads to the reference, and then to quantify the level of expression for each transcript based on the read coverage (as demonstrated in Fig. 1). However, the reference could be either the annotated transcriptome or the genomic sequence. A key difference between these two processes is whether gaps need to be considered during the mapping process. Specifically, if a genome is used as the reference, some RNA-Seq reads that cover more than one exon will need to be mapped across intron-exon boundaries. This is particularly challenging for short sequence reads. Even though there are multiple mappers available, *TopHat* [12]—a bioinformatic package designed to align RNA-Seq reads to large (eukaryotic multicellular) genomes and to discovering novel RNA splice sites and gene structure variation—remains the preferred choice for analysis from mammalian-size to bacterial genomes alike.

3.2.1 Mapping Reads with Annotated Transcriptomes

Mapping reads to the annotated transcriptome is a more straightforward approach, as the method is devoid of challenges in mapping reads across intron-exon boundaries. If an analysis is focused solely on the quantification of known transcripts, this should be the method of choice. Below are steps involved and commands used under a working directory that contains (1) genome sequence (*G_ref.fa*) and (2) genome annotation (*G.gtf*).

1. Prepare the reference:

```
/path/bowtie-build G_ref.fa G_ref_index
```

Through this command, *Bowtie* software will take the genome fasta as an input, and build a reference index file *G_ref_index* under the same working directory.

2. Build-up transcriptome index files:

```
/path/tophat -G G.gtf --transcriptome-index=transcriptome_data/
transcript G_ref_index
```

Through this process, *TopHat* creates the *transcriptome_data* folder in the current directory. The command extracts transcriptome sequences based on the input files of genome annotation *G.gtf* and the reference index file *G_ref_index*. All data are stored in the folder *transcriptome_data* named with the prefix *transcript*.

3. Map sequence reads:

```
/path/tophat -o T1-R1_T_output --no-novel-juncs --transcriptome-
index=transcriptome_data/transcript G_ref_index Trimmed_T1-
R1_1.fq Trimmed_T1-R1_2.fq
```

This example shows the mapping of one paired end read file after trimming for *T1-R1*. The command will generate an "*accepted_hits.bam*" file under *T1-R1_T_output* directory, which will be used along with the reference annotation file *G.gtf* to estimate transcript expression and identify DEGs in *cuffdiff*. The same command will be repeatedly run for *T1-R2, T1-R3, T2-R1, T2-R2,* and *T2-R3*. Through this process, a bam file is generated for each biological replicate. However, two experimental conditions, *T1, T2,* will be compared to identify DEGs.

4. Identify DEGs using cuffdiff:

```
/path/cuffdiff -o R_diff_output -u G.gtf -L T1,T2 \
T1-R1_T_output/accepted_hits.bam,T1-R2_T_output/accepted_hits.
bam, \
T1-R3_T_output/accepted_hits.bam T2-R1_T_output/accepted_hits.
bam, \
T2-R2_T_output/accepted_hits.bam,T2-R3_T_output/accepted_hits.
bam
```

The software package *cuffdiff* takes mapping results "*accepted_hits.bam*" from the previous step. –L option specifies two experimental datasets to be compared. Through this command, a new folder *R_diff_output* is created in the current working directory. Within the directory, the file *XX.exp.diff* captures DEGs and statistic measures used to support the call of DEGs.

Note: *TopHat* requires a specific working directory that contains input files as genome sequence (*G_ref.fa*), genome annotation (*G.gtf*) and reads used to map to the reference. If not, a path needs to be used to specify access to these files. Readers are also advised to check out the manual to find out other available options.

3.2.2 Mapping Reads with Genome Sequences

Using a genome as the reference is more challenging, but has its own merits: it helps identify novel and alternatively spliced transcripts and define intron/exon boundaries and improve genome annotation. To overcome the challenges associated with mapping reads that span across splice junctions, *TopHat* includes a two-step workflow: (1) map all the unspliced reads to exons; (2) unmapped reads are split and aligned independently to identify exon junctions. This tool also incorporates a genomic annotation file (gtf/gff file) to assist the identification of the splicing events and improve the mapping accuracy (*see* **Note 4**). Steps involved in this analysis are listed below:

1. Map the sequence reads to the reference genome using *tophat*.

Prepare the reference:

```
/path/bowtie-build G_ref.fa G_ref_index
```

Map sequence reads:

```
/path/tophat -G G.gtf -o T1-R1_T_output G_ref_index Trimmed_T1-
R1_1.fq Trimmed_T1-R1_2.fq
```

The key output from this process is the *accepted_hits.bam* file with the mapping information in *T1-R1_output* folder, which will be used as the input file to identify and estimate the expression of expressed transcripts in downstream analysis. The same command is needed to be run repeatedly for *T1-R2, T1-R3, T2-R1, T2-R2,* and *T2-R3.*

2. Assemble and identify transcripts using *Cufflinks*: there are additional challenges in accurately defining the full-length and inferring the transcription start and end sites of a gene based on multiple transcript fragments, or even alternatively spliced fragments after the initial mapping process. To add to the level of difficulty, a transcript with low expression usually does not have enough sequence coverage to create a reliable assembly. Our suggestions are (1) to use annotation to guide the assembly process; (2) to use pair-end reads to link fragments; (3) to increase sequence coverage if your research is particularly interested in low-abundant transcripts. *Cufflinks* is one of the programs designed to meet these challenges, by incorporating genome annotation, rigorously filtering out mapping false positives and artifacts, and then constructing the transcripts with the best evidence (*see* **Note 5**).
   Cufflinks assembles transcripts:

```
/path/cufflinks -o T1-R1_C_output T1-R1_T_output/accepted_hits.bam
```

After this step, a "*transcripts.gtf*" file will be generated under *T1-R1_C_output* folder.

The same command will be run repeatedly for *T1-R2, T1-R3, T2-R1, T2-R2,* and *T2-R3.* Then, the "*transcripts.gtf*" will be listed in "*assembly_GTF_list.txt*" and merged with genomic annotation file by *cuffmerge* to correct genomic annotation and calculate gene and transcript expression in each sample:

```
cuffmerge -g G.gtf -o R_M_output assembly_GTF_list.txt
```

The key output from this process is *merged.gtf* in *R_M_output* folder.

3. Estimate expression levels and identify differentially expressed genes using *cuffdiff:* The output from *Cufflinks* including the number of mapped reads to each assembly transcript can be used in *cuffdiff* to normalize and calculate the expression level for each gene. A normalization protocol called RPKM (reads per kilobase of exon model per million reads) and its subsequent derivatives FPKM (fragments per kilobase of exon model per million mapped reads) are commonly used (*see* **Note 6**). The protocol considers the length of a gene and the size of a library. A normalized expression of a gene enables a statistical comparison among different samples designed for the experiment (*see* **Note 7**). Similarly, *cuffdiff* also includes a differential gene expression analysis tool that enables comparative expres-

sion studies among samples from different experiments (*see* **Note 8**).

```
cuffdiff -o R_diff_output -b G_ref.fa -L T1,T2 -u R_M_output/
merged.gtf \
T1-R1_T_output/accepted_hits.bam,T1-R2_T_output/accepted_hits.
bam, \
T1-R3_T_output/accepted_hits.bam T2-R1_T_output/accepted_hits.
bam, \
T2-R2_T_output/accepted_hits.bam,T2-R3_T_output/accepted_hits.
bam \
```

The results of *cuffdiff* are stored in *R_diff_output* folder. Differential expression results (gene-/transcript-) with statistical analysis as well as correlated attributes are recorded in *XX.exp.diff* files.

3.3 De Novo Transcript Assembly

De novo assembly-based transcriptome analysis is particularly powerful when a reference genome is not available. It has been successfully used to study plant-pathogen interactions in many instances where a reference genome for neither the pathogen nor the plant host is available [6]. Several programs have been developed to perform this task, including *Trans-Abyss* [21], *SOAPdenovo-Trans* [22] and *Trinity* [14]. We will demonstrate the de novo assembly using *Trinity* (as demonstrated in Fig. 1), with the consideration that the *Trinity* pipeline takes advantage of *de Bruijn* graphs and generate graphs for each expressed gene, which guarantees the accuracy of transcript assembly and, meanwhile, reduces the computational running space and time. Using trimmed sequencing reads (as described in Subheading 3.1), the pipeline includes read concatenation, assembly, transcript quantification and differential expression analysis.

1. Read concatenation: The success of this approach depends on the initial identification of transcripts combining sequencing data from multiple samples. There are several advantages to pooling all data into one batch, and then subjecting it to the assembly process. First, combined reads provide a higher sequence coverage, and therefore increases the assembly power. This is particularly important for low expression genes. Second, this process allows for the generation of a set of unique transcript IDs, which can be used for comparing expression levels among different samples. However, there are drawbacks. For instance, larger data sets require much enhanced computing intensity. The solution is to request higher settings of RAM and CPU (*see* below). The process also pools errors together, which may result in more falsely assembled transcripts. Because most errors are created at random, most falsely assembled transcripts have low sequence coverage and can be removed with a coverage filter. For PE data, the batch is created separately for left (or forward) and right (or reverse) read data.

Here is an example for concatenating two datasets:

```
cat T1-R1_1.fq T1-R2_1.fq T1-R3_1.fq T2-R1_1.fq T2-R2_1.fq T2-
R3_1.fq > left.fq
cat T1-R1_2.fq T1-R2_2.fq T1-R3_2.fq T2-R1_2.fq T2-R2_2.fq T2-
R3_2.fq > right.fq
```

T1 and *T2*, are two experimental treatments, such as a control and an infected sample. Each condition has three biological replicates, represented as *R1*, *R2*, and *R3*.

XX_1.fq and *XX_2.fq* usually represents forward and reverse read files, respectively.

This process will produce two read files: *left.fq* and *right.fq* include *XX_1* and *XX_2* of paired-end reads from *XX_1.fq* and *XX_2.fq* files, respectively.

2. Trinity assembly: concatenated data are then assembled de novo using the command below. Make sure to specify a path to "*Trinity*," otherwise the error "*Trinity cannot be found*" will be raised.

```
/path/Trinity --max_memory 40G --CPU 20 --seqType fq --left
left.fq --right right.fq
```

Note: Higher settings of RAM (adjusting the "*--max_memory*") and CPU (adjusting the "*--CPU*") are needed if the data size is large (*see* **Note 9**).

The key output from this process is the *Trinity.fasta* file with all the assembled transcripts, which will be stored in *trinity_out_dir* directory as the default or one specified by users. This file will be used as the reference to estimate transcript expression in the next step.

3. Quantification of transcripts: Trimmed reads from each sample will be mapped to the assembled transcripts (*Trinity.fasta*) separately to estimate expression levels using *align_and_estimate_abundance.pl* within *Trinity* package.
For example:

```
/path/align_and_estimate_abundance.pl --transcripts trinity_
out_dir/Trinity.fasta --seqType fq --left T1-R1_1.fq --right
T1-R1_2.fq --est_method RSEM --aln_method bowtie2 --prep_refer-
ence --output_dir T1-R1_RSEM --thread_count 10
```

The output of this command will contain two files: *RSEM.genes.results* and *RSEM.isoforms.results*, both containing read coverage (count) and FPKM value for each gene and isoform in the output directory *T1-R1_RSEM*. In this process, make sure the perl script is able to find *bowtie2*, *RSEM*, and *Samtools*. The same command needs to be run repeatedly for *T1-R2, T1-R3, T2-R1, T2-R2,* and *T2-R3*.

4. Create a matrix table of read counts: The output from *Trinity RSEM.genes.results* (as shown in the following example) or *RSEM.isoforms.results* can be used for differential gene expression analysis via *edgeR* or *DEseq* in bioconductor. We use "prefix

infection" to indicate the comparison of differentially expressed genes in infected samples (*T2*) when compared to the control samples (*T1*).

The following command transforms the transcript expression levels into a read count matrix table:

```
/path/abundance_estimates_to_matrix.pl --est_method RSEM
--out_prefix infections T1_R1.genes.results T1_R2.genes.results
T1_R3.genes.results T2_R1.genes.results T2_R2.genes.results T2_
R3.genes.results
```

This step will result in three output files named after the prefix provided by the user:

```
Infections.counts.matrix
infections.TPM.not_cross_norm
infections.TMM.EXPR.matrix
```

The file "*infections.counts.matrix*", containing the count/coverage of raw reads mapped to each gene/isoform, is used for the downstream DEGs analysis by *edgeR* or *DESeq*. The other two files contain normalized read counts using TPM or TMM methods, and they are typically used for determining the cutoff for filtering out low expression genes/isoforms in downstream analyses.

5. Determining low expression genes/transcripts cutoff: Most often we see multiple transcript isoforms, varying in length and number of exons, assembled for one gene. As to which isoform should be used for downstream analysis, is a challenging question frequently asked and very much in debate. Simply picking the longest sequence as the major isoform is an option, but not necessarily the correct one. Among the multiple isoforms of a gene, usually only one or two are relatively high-expressed. Limited by sequencing abundance, low expressed transcripts usually are short and contain many assembly errors. Therefore coverage or FPKM/RPKM is considered a better filter to remove transcripts with a low level of abundance. The *contig_ExN50_statistic.pl* takes into account the RNA-Seq assembly quality (N50) and normalizes expression data statistically to indicate the cutoff for the low expression genes. For example:

```
/path/ contig_ExN50_statistic.pl infections.TMM.EXPR.matrix
Trinity.fasta | tee ExN50.stats
```

This step will quantify levels of transcript expression in relationship to the length of transcripts (Table 3). The relationship between *Ex* and *Ex-N50* could be reflected by a curve plot (*Ex* vs *Ex-N50*). The potential curve point will have a corresponding minimum expression in the table which will be the cutoff for the lowly expressed genes/transcripts.

Caution is needed as a hard cutoff could result in a false negative. *Trinity* package offers a program to filter the low expression genes/transcripts. The following is an example code:

```
/path/filter_low_expr_transcripts.pl --matrix infections.TMM.
EXPR.matrix --transcripts Trinity.fasta --min_expr_any 1.0
```

In this example, the filter, one that is most commonly used, requires a minimal expression (TPM or FPKM) of 1.0 across all samples.

6. Discovering differentially expressed transcripts: Here we demonstrate the identification of differentially expressed genes (DEGs) using the Bioconductor package *edgeR* [15]. For the DEG analysis using *edgeR* in this protocol, only the nonnormalized raw read counts should be used instead of normalized results.

First, an R session should be opened and the *edgeR* package needs to be installed using the following two lines:

```
> source("https://bioconductor.org/biocLite.R")
> biocLite("edgeR")
```

To use *edgeR*, we need to load the *edgeR* package:

```
 > library(edgeR)
```

Read count matrix (the first column has a header called "*genes*"):

```
 > count_data <- read.delim("./infections.counts.matrix",row.
names = "genes")
```

Note: make sure to use the matrix of raw read counts.

Provide sample information. The three columns immediately after the "*genes*" column are control samples (1), and the next three columns are treatment samples (2):

```
> group <- factor(c(1,1,1,2,2,2))
```

To calculate differentially expressed genes:

```
> Diff <- DGEList(counts=count_data,group=group)
> Diff <- calcNormFactors(Diff)
> design <- model.matrix(~group)
> Diff <- estimateDisp(y, design)
> fit <- glmQLFit(Diff, design)
> glf <- glmQLFTest(fit, coef=2)
```

To show how the first few differentially expressed genes:

```
> topTags(glf)
```

The output for *topTags(glf)* looks like:

ID	logFC	logCPM	F	*P-Value*	FDR
TRINITY_DN1020_c112_g1_i1	13.62494	13.242372	9339.25	2.29E−202	6.42E−197
TRINITY_DN1004_c131_g1_i1	12.12149	13.066616	7759.06	1.74E−180	2.44E−175
TRINITY_DN1504_c108_g2_i1	16.01237	10.078747	2405.222	2.15E−179	2.01E−174
TRINITY_DN1083_c45_g2_i1	15.2792	10.153407	2467.565	6.36E−179	4.46E−174
TRINITY_DN3020_c76_g1_i1	13.96	10.126923	2368.264	1.62E−173	9.10E−169
TRINITY_DN139_c1150_g1_i1	13.79549	9.848065	2085.537	8.88E−170	4.15E−165
TRINITY_DN457_c3820_g2_i1	12.54245	9.981079	2131.092	4.41E−165	1.77E−160
TRINITY_DN3490_c301_g1_i1	13.08786	9.666081	1885.495	2.27E−164	7.97E−160
TRINITY_D2538_c874_g2_i1	15.00401	9.068211	1482.866	3.50E−161	1.09E−156
TRINITY_DN7821_c412_g2_i1	14.27108	9.133886	1514.754	9.32E−161	2.61E−156

Table 3
Expression levels versus lengths of assembled transcripts

Ex[a]	Minimum expression[b]	*Ex-N50*[c]	Number of expressed genes/transcripts
E2	43,124.713	606	2
E3	13,240.900	964	3
…	…	…	…
E96	9.18	2400	14,261
E97	6.37	2600	16,646
E98	3.93	2800	23,471
E99	1.51	2400	66,696
E100	0	600	428, 285

[a] *Ex*: Ex is a division of all assembled transcript nucleotides based on their expression levels. *x* represents *x*% of all assembled transcript nucleotides. For instance, E100 includes all levels of expression for all 428,285 gene/transcripts assembled for this experiment. E2 represents top 2% of expression levels
[b] *Minimum expression*: the minimum level of expression of genes/transcripts in the group
[c] *Ex-N50*: at least half of genes/transcripts within the group are either equal or longer than the defined size

The first column, showing the gene and isoform ID, contains three parts: cluster, gene and isoform information. For example, TRINITY_DN1020_c112_g1_i1 indicates *Trinity* read cluster "TRINITY_DN1020", gene "g1" and isoform "i1." Differentially expressed genes are stored in a vector "DEGs", which contains first 50,000 rows, sorted by p. values and using a cutoff of 0.05:

```
> DEGs <- topTags(qlf,n = 50000,adjust.method = "BH",p.value =
0.05,sort.by = "PValue")
```

Finally, DEGs will be formatted in a tab-delimited file called "DEGs.txt" and exported to a local PC:

```
> write.table(DEGs, file="DEGs.txt",sep="\t", quote = FALSE,
row.names= TRUE)
```

Also, the latest version of *Trinity* provides the command line based on DEG analysis. You could check *Trinity* manual for detailed information.

3.4 Visualization for the Differential Gene Expression Analysis and Expression Profiling

Visualization is a very important component in interpreting large data sets. It brings life into quantitative data and facilitates intuitive decision-making process. There are many visualization tools. Here we present two:

1. Visualization of differentially expressed genes: *CummeRbund* is a visualization R package specifically designed with *Cufflinks* output to help researchers evaluate and visualize a large amount of data [23].

The following is an example:

Change to *R_diff_output* folder, open R session and load *cummeRbund*

```
> library ("cummeRbund")
```

Import Cuffdiff output data into R

```
> cuffdiff_data <- readCufflinks ('/path/ R_diff_output',
rebuild=T)
```

Extract FPKM for whole genome genes

```
> FPKM_matrix <- fpkmMatrix (genes(cuffdiff_data))
```

Save *FPKM_matrix* result in *fpkm_matrix.txt*

```
> write.table (FPKM_matrix, file=" fpkm_matrix.txt", sep="\t")
```

Identify significant DEGs and save results in *DEGs.txt*

```
> Diff_gene <- diffData (genes(cuffdiff_data))
> Sig_gene <- subset (Diff_gene, (significant == 'yes'))
> write.table (Sig_gene, file="DEGs.txt", sep="\t")
```

Visualize variations among biological replicates, including scatter plots and dendrograms between samples:

```
> csScatterMatrix (genes(cuffdiff_data), replicates=T)
> csDendro (genes(cuffdiff_data), replicates=T)
```

2. Heatmap clustering DEGs: Because of the complexity of transcriptomes, it is a challenge to display all the data information using one method. Users should adopt multiple visualization methods to overview expression patterns among biological samples and to explore genes of biological interests. A popular approach is to cluster genes and samples using Hierarchical clustering or K-means clustering algorithms and visualize the cluster(s) by heatmaps, which provide an effective way to navigate differential expression results and pinpoint interesting genes. One of the most popular tool is *heatmap.2*. Below is an example:

   Open R session and load *gplots* and *RColorBrewer*:

```
> install.packages ("gplots")
> install.packages ("RColorBrewer")
> library (gplots)
> library (RColorBrewer)
```

   Load and matrix DEGs expression results:

```
> DEG <- read.table ("DEGs.txt", header=T, row.names="gene_id")
> DEG_value <- DEG[,c("value_1","value_2")]#pick out values
(only values could be used in next steps)
> DEG_mat <- as.matrix (DEG_value)
```

   Generate heatmap:

```
> quantile(DEG_mat) #look at expression distribution
> z.DEG_mat <- t (scale(t(DEG_mat), center=TRUE, scale=TRUE))
#scale data, z-score transformation, z = (x - mean)/sd
> cols <- colorRampPalette (brewer.pal(10, "RdBu"))(256) #set
up plot color
> pdf(file='heatmap.pdf',height=12,width=10)
```

```
> heatmap.2 (z.DEG_mat, dendrogram="both", scale="none",
trace="none", col=rev (cols)) # using eucledian distance #
"Colv=F" do not order column
> dev.off()
```

The heatmap will be stored in *heatmap.pdf*.

3.5 Gene Ontology (GO) Enrichment

Functional characterization of DEGs involved in the host–pathogen interaction is the last step in a standard RNA-Seq analysis. Gene Ontology (GO) enrichment calculates the statistically significant enriched molecular functions, which provide insights into the mechanisms underlying the plant disease (*see* **Note 10**). The most popular and convenient tool for researchers without any bioinformatics experience is Blast2GO.

First, annotate whole genome genes (de novo assembled transcripts)

1. Open *Blast2GO*.
2. At the top left, click *File*.
3. Under *Load*, click *Load sequences*.
4. Click *Load Fasta File (.fasta)*.
5. Click *blast* and set up database (nr) and cutoff (1e-15).
6. After Blast, click *mapping*.
7. Then, click *annot*.

Perform GO enrichment analysis:

1. Click *Analysis* at the top panel.
2. Click *Enrichment Analysis (Fisher's Exact Test)* to upload *Test-Set Files*.
3. Click *Run*.

After these analyses, you will see results with your own data similar to those shown in Fig. 3. Functional annotation will be assigned to the sequence if a homologous sequence is found in the database. After the enrichment analysis process, a file named by user will be stored in the *blast2go* output folder with enriched GO terms.

4 Discussions

We need to consider several critical components:

1. Experimental design: the most critical component for the experimental design when using RNA-Seq to study host–pathogen interactions is to capture the dynamics of the interaction. The critical time points vary in each pathosystem. Our experiences point to early infection stages, covering host–pathogen recognition and getting in contact, to pathogen penetrating plant cells, till establishing infection in the host.

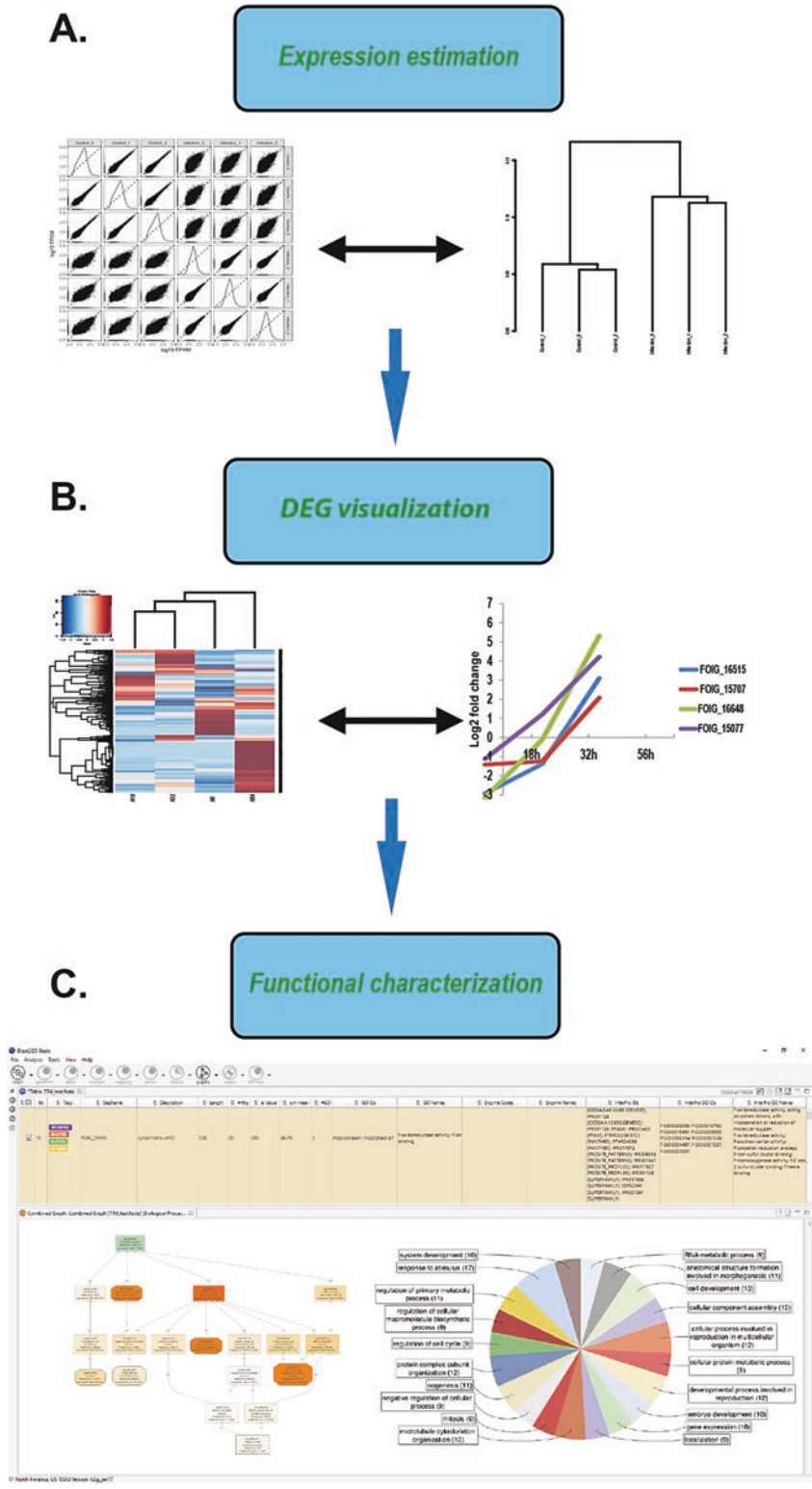


Fig. 3 Examples of DEG visualization and functional annotation. (**a**) The scatterplots (left) and dendrogram (right) Fig. 3 (continued) portray an estimation of the expression variation among biological replicates. (**b**) A visualization of DEG expression in a heatmap (left) or a clustering (right) format. (**c**) An example of applying *Blast2GO* annotations of DEGs. The GO network lists the GO terms and their potential relationships. A pie chart is used to summarize annotation statistics

For fungal systems we have studied, these events happen within 1, 2, and 4 days post inoculation respectively.

2. Create equal and sufficient sequence coverage for each experiment: this could be a challenge, especially with infected samples, as it is hard to know the proportion of RNAs that are from the pathogen versus the host. Our solution is to increase sequence coverage for infected samples. If the overall presentation of pathogen is low in the sample, FPKM normalization has the potential to amplify small changes. One solution is to look at the original mapping count, in addition to the FPKM values, for interesting candidate genes. Two factors are crucial for determining the number of replicates and sequencing depth in a study: variability (technical variability and biological variability) and statistical methods. Compared with technical variability, biological variability is the main source of variability in RNA-Seq experiments. Considering the experimental costs, it is not necessary to create technical replicates. A proper statistical analysis could eliminate noise and return a good false discovery rate. For example, the genes that are highly expressed among a set of genes under test may influence the statistical analysis results even after multiple corrections. Removing lowly expressed genes might be able to reduce false positives and improve the accuracy of differential expression analyses [24]. Otherwise, increasing the sequencing depth and the number of replicates may improve robustness of the analysis and minimize sample biases [25]. Thus, a proper statistical method with reasonable replicates in an experimental design is essential.
3. Linux is a user-friendly operating system. It provides Command Line (CLI) and graphical user interface (GUI) to execute commands. CLI is one of the many strengths of Linux and most of the programs used for RNA-Seq analysis are developed based on the CLI. Even though Linux has many different kinds of distributions, such as Ubuntu, Red Hat, and SUSE, commands used for data analysis are basically the same. So we strongly encourage users to overcome the initial learning curve and get comfortable with the Linux environment, learn the basic Linux commands and run the programs via the CLI. Remember that all the Linux commands are case sensitive.
4. RNA-Seq sequencing: Although third generation sequencing provides longer reads to support the identification of full-length transcripts without the assembly step, the currently high sequencing error rates and relatively low throughput limits its use in RNA-Seq. We encourage our readers to use Illumina sequencing to generate the data, as it provides more accurate reads which are essential to a more confident quantification of transcript expression.

5. Visualization of RNA-Seq mapping data is a good way to manually check the mapping results. Meanwhile, it provides a way to visually observe splice junctions and alternative splicing events. The process is similar to that of genome sequencing, which can be done at the level of unnormalized total read counts or normalized mapping coverage. Some programs have been developed to visualize RNA-Seq data, such as Integrative Genomics Viewer (IGV), UCSC browser and Genome Maps. These programs provide multiple ways to display transcript read abundance and exon and splicing junctions, and supplement each other in multiple ways. It is always good for users to visualize changes in gene read coverage that are consistent with the RNA-Seq mapping results to evaluate the robustness of the experiment.

5 Notes

1. HPCC facilities, which have been established in many universities and institutes, are indispensable for the analysis of large quantity RNA-Seq data. We would suggest checking with your university/institute to sign up for an account. If no resource is available, you could try cloud computing, like Amazon Web Services (AWS) (https://aws.amazon.com/). The elastic compute (EC2) service in AWS provides a cloud computational resource that allows you to configure and rent computers to fit your computing needs.
2. Because over 90% of the total RNA in a cell is ribosomal RNA (rRNA), a poly-A isolation method is recommended for RNA extraction in standard host–fungi (and oomycetes) experiments. However, this method will not include information about other RNA types such as microRNAs and long noncoding RNA. An alternative method is rRNA depletion, which is able to directly remove rRNAs in the extracts while maintaining other RNA categories. In general, rRNA depletion method results in more noise and requires more starting material (>200 ng of total RNA).
3. An alternative way to check the quality of the trimmed reads is to map the reads to the reference sequences to determine the overall sequencing accuracy and remove the contaminating sequences. Tools with quality control in mapping include *Picard*, *RSeQC* [26], and *Qualimap* [27].
4. Besides alternative splicing, some other factors should be considered during the mapping process, such as the percentage of mismatches in mapped reads, the length and type of reads and the library insert size. In the mapping process, one read could either be assigned to only one position or could be

mapped to multiple positions. When the reference is the transcriptome, multimapped reads will be frequent, because reads fall onto exons that are shared by different transcript isoforms of the same genomic sequence. However, if the reference is the genomic sequence, a read that has multimapping in different isoforms of a gene would be uniquely mapped to the genome. The genomic multimapping is primarily due to shared domains of homologous genes or repetitive sequences. Thus, we suggest using the longest isoform of each gene to estimate transcript expression when using transcriptomes as the reference. It is not wise to simply discard those reads during the analysis process. Currently, there is no simple way to solve this problem, which makes it a challenge in the analysis of the mapping results.

5. Accurately constructing transcripts with the short Illumina reads is one of the most difficult challenges in RNA-Seq analysis. To correctly assemble transcripts, the mapped RNA-Seq data can be analyzed with structural gene annotation tools such as *PASA*, *Augustus* to identify new expressed sequences and improve the genome annotation. Long reads such as those generated by PacBio that are long enough to cover the whole transcript could be another way to improve assembled transcripts.
6. Normalization could be used to remove other factors that could potentially affect the expression estimation. For instance, *DESeq* [28], Poisson-seq [29], and *UpperQuartile* [30] are designed to remove biases caused by highly expressed transcripts that could skew effective read counts. *Cufflinks* estimates positional biases caused by coverage along the transcript during its normalization process to remove inherited biases in the data. *EDAseq* [31] takes into account factors such as sample conditions, average fragment size, and GC content of the genes. For the programs of nonparametric models such as *NOISeq* [24], specific normalization procedures are applied after a wide variety of diagnoses of possible biases in RNA-Seq data. The appropriate experimental design is the best way to minimize biases. However, if you are worried that your samples may suffer from sequencing biases, batch effects, poor sample conditions, or GC biases in genes, you could explore additional normalization solutions.
7. *Cuffdiff* takes the advantage of PE reads as well as GTF information and uses expectation-maximization methods to decrease the biases caused by multimapped reads. This program is also able to find significant differences in gene length between samples and correctly estimate the gene length from the data. Besides this, other estimating expression

programs including *RSEM* [16], *Sailfish* [32] and *eXpress* [33] have been developed to estimate gene expression for within-sample and between-sample comparisons. While the choice of programs can have an impact on the gene expression estimation, each algorithm takes into account biases such as the nonuniform reads as well as sequencing depth. We strongly suggest readers conduct a comprehensive comparison of different programs when they are doing the estimation.

8. Several independent comparative studies have demonstrated that the choice of the programs (even the version of the software) can have a strong effect on the performance of the differential gene expression analysis. None of the programs is perfectly suitable for all datasets. The program *limma* has been tested under many circumstances with low computing memory and fast speed [34, 35]. However, some programs take more memory and a longer time to run differential expression analysis because of the algorithms implemented for the normalization procedure. Also, the algorithms determine the program sensitivity and specificity for different datasets. For sample comparison with multiple replicates, *edgeR* [15] takes the negative binomial model in the normalization step to generate the reference distribution using raw read counts as the input. However, this program is too liberal in controlling FDR and it may potentially introduce biases into the analysis. As a contrast, *DESeq* is more conservative [28], although it shares the same negative binomial model as *edgeR*. In the nonparametric program *NOISeq*, the null distribution is used to estimate the RNA-Seq data. When comparing data with no or few replicates, using a negative binomial model or null distribution model could be noisy. The alternative statistical models are Poisson distribution (*DESeq*) [28] and empirical distribution (*NOISeq*) [24]. Although the *p-value* could be inferred from those alternative statistical models, RNA-Seq samples with replicates are strongly recommended. Usually a minimal of three replicates are favored in RNA-Seq experiments to reduce false positive.
9. Depending on the amount of input data, it can take days or even weeks to finish this step. The *Trinity* program will generate many intermediate files during the whole process, which can take up disk space quickly. Therefore, it is important to have sufficient disk space in the working directory.
10. For the transcripts lacking protein-coding potential, the *Rfam* database that contains most of the well-characterized RNA families can be used for annotation [36].

References

1. Kinealy C (2001) The Great Irish Famine: impact, ideology and rebellion. Palgrave Macmillan, New York
2. Schumann GL, D'arcy CJ (2006) Essential plant pathology. American Phytopathological Society (APS Press), St. Paul, MN
3. Ploetz RC (2015) Fusarium wilt of banana. Phytopathology 105:1512–1521
4. Wyenandt CA, Simon JE, Pyne RM et al (2015) Basil downy mildew (Peronospora belbahrii): Discoveries and challenges relative to its control. Phytopathology 105:885–894
5. Ozsolak F, Milos PM (2011) RNA sequencing: advances, challenges and opportunities. Nat Rev Genet 12:87–98
6. Guo L, Allen KS, Deiulio G et al (2016) A de novo-assembly based data analysis pipeline for plant obligate parasite metatranscriptomic studies. Front Plant Sci 7:925
7. Zuluaga AP, Vega-Arreguin JC, Fei Z et al (2016) Transcriptional dynamics of Phytophthora infestans during sequential stages of hemibiotrophic infection of tomato. Mol Plant Pathol 17:29–41
8. Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc
9. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120
10. Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079
11. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359
12. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111
13. Trapnell C, Roberts A, Goff L et al (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562
14. Haas BJ, Papanicolaou A, Yassour M et al (2013) De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. Nat Protoc 8:1494–1512
15. Robinson MD, Mccarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140
16. Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323
17. Nagalakshmi U, Wang Z, Waern K et al (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320:1344–1349
18. Tarazona S, Garcia-Alcalde F, Dopazo J et al (2011) Differential expression in RNA-seq: a matter of depth. Genome Res 21:2213–2223
19. Pollen AA, Nowakowski TJ, Shuga J et al (2014) Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. Nat Biotechnol 32:1053–1058
20. Sims D, Sudbery I, Ilott NE et al (2014) Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet 15:121–132
21. Robertson G, Schein J, Chiu R et al (2010) De novo assembly and analysis of RNA-seq data. Nat Methods 7:909–912
22. Xie Y, Wu G, Tang J et al (2014) SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 30:1660–1666
23. Goff LA, Trapnell C, Kelley D (2012) CummeRbund: visualization and exploration of Cufflinks high-throughput sequencing data. R package version 2
24. Risso D, Schwartz K, Sherlock G et al (2011) GC-content normalization for RNA-Seq data. BMC Bioinformatics 12:480
25. Liu Y, Zhou J, White KP (2014) RNA-seq differential expression studies: more sequence or more replication? Bioinformatics 30:301–304
26. Wang L, Wang S, Li W (2012) RSeQC: quality control of RNA-seq experiments. Bioinformatics 28:2184–2185
27. Garcia-Alcalde F, Okonechnikov K, Carbonell J et al (2012) Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics 28:2678–2679
28. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106
29. Li J, Witten DM, Johnstone IM et al (2012) Normalization, testing, and false discovery rate estimation for RNA-sequencing data. Biostatistics 13:523–538
30. Bullard JH, Purdom E, Hansen KD et al (2010) Evaluation of statistical methods for

normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics 11:94

31. Steijger T, Abril JF, Engstrom PG et al (2013) Assessment of transcript reconstruction methods for RNA-seq. Nat Methods 10:1177–1184
32. Patro R, Mount SM, Kingsford C (2014) Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. Nat Biotechnol 32:462–464
33. Roberts A, Pachter L (2013) Streaming fragment assignment for real-time analysis of sequencing experiments. Nat Methods 10:71–73
34. Seyednasrollah F, Laiho A, Elo LL (2015) Comparison of software packages for detecting differential expression in RNA-seq studies. Brief Bioinform 16:59–70
35. Soneson C, Delorenzi M (2013) A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics 14:91
36. Nawrocki EP, Burge SW, Bateman A et al (2015) Rfam 12.0: updates to the RNA families database. Nucleic Acids Res 43:D130–D137



Chapter 16

Application of the Cre/*lox* System to Construct Auxotrophic Markers for Quantitative Genetic Analyses in *Fusarium graminearum*

Lanelle R. Connolly, Allyson A. Erlendson, Corinne M. Fargo, Kendra K. Jackson, Morgan M. G. Pelker, Jacob W. Mazzola, Mark S. Geisler, and Michael Freitag

Abstract

The bacteriophage P1 Cre/*lox* system has been utilized in diverse fungi for marker recycling and exchange, generation of targeted chromosome translocations, and targeted deletion of interstitial chromosome segments. Here we show the application of this tool in the wheat and maize pathogen, *Fusarium graminearum*. We explored three different ways to introduce Cre into strains with floxed genes, namely transformation with an episomal or integrative plasmid (pLC28), fusion of protoplasts of strains carrying floxed genes with strains expressing Cre by forcing heterokaryons, and crosses between strains with floxed genes and strains expressing Cre to isolate progeny in which the target genes had been deleted during the cross. We used this system for the construction of strains bearing auxotrophic markers that were generated by gene replacement with positively selectable markers followed by Cre-mediated marker excision. In addition, updated protocols for transformation and crosses for *F. graminearum* are provided. In combination, strains and tools developed here add to the arsenal of methods that can be used to carry out molecular genetics with *F. graminearum*.

Key words Cre/*lox*, DNA-mediated transformation, *Fusarium graminearum*, Genetic tools

1 Introduction

Since its first application in yeast 30 years ago [1], the bacteriophage P1 Cre/*lox* system has been utilized in plants, fungi, and animals to carry out a multitude of genetic manipulations, including genome rearrangements, conditional mutagenesis, and antibiotic marker recycling [2–10]. While there are well over a dozen recombination systems known to function in eukaryotes, the tyrosine recombinase-based *Saccharomyces cerevisiae* FLP/*FRT* [11–13] and the P1 Cre/*lox* system became the most popular because of the requirement for a single enzyme and relatively short

Wenbo Ma and Thomas Wolpert (eds.), *Plant Pathogenic Fungi and Oomycetes: Methods and Protocols*, Methods in Molecular Biology, vol. 1848, https://doi.org/10.1007/978-1-4939-8724-5_16,
© Springer Science+Business Media, LLC, part of Springer Nature 2018

recognition sequences that are usually absent from even complex genomes. Cre and FLP are the recombinases, whereas *lox* and *FRT* are the recombination sites [14]. The genetic crossover between recombination sites regenerates these sites, and in both systems the recombination reaction is fully reversible.

The Cre recombinase is endogenous to a temperate bacteriophage, P1, where it enables the double-stranded linear phage DNA to exist as an episome. Cre recombinase functions similarly to the Xer system of proteobacteria [15] by preventing chromosome dimers to form, ensuring faithful segregation of the phage DNA. Cre recombinase catalyzes recombination between two *loxP* sites, 34-bp recognition sequences that include an 8-bp asymmetric spacer, which is variable except for the middle two nucleotides and flanking 13-bp palindromic recognition motifs [16]. Cre is a monomer in solution, but cooperatively dimerizes at *loxP* sites to catalyze double strand breaks and recombination through the formation of a Holliday junction intermediate [17, 18]. The orientation of the *loxP* sites determines the final genomic rearrangement. If *loxP* sites are on the same DNA strand or chromosome, but are oriented in opposite directions the result of recombination will be an inversion. Two *loxP* sites oriented in the same direction on a chromosome will result in the excision of the DNA between sites. "Floxed" genes are thus genes flanked by *loxP* sites. Translocations occur when *loxP* sites are spread among chromosomes.

Some of the most notable advances in the use of the Cre/*lox* system have been accomplished with mammalian systems. Only one year after its introduction into budding yeast, Cre recombinase was shown to be active in mammalian cells [2]. One of the most impressive examples for Cre/*lox* technology is the "Brainbow," which enables differential labeling of individual neurons in mice in four or more colors, creating fantastically detailed maps of neuronal networks, and allowing the analysis of glial cells and neurons as a function of time [19]. One major problem facing the use of Cre/*lox* in multicellular organisms is induction of cell- or tissue-specific expression of the recombinase. To prevent the ubiquitous activity of Cre recombinase, split-Cre complementation was generated in vivo in mice. Inactive split-Cre fragments were generated to regain activity only when both were expressed in the same cell type [20]. In plants, use of this system is widespread for marker removal from genetically engineered individuals [3]. Different orientations of luciferase were integrated into the tobacco genome using differentially oriented *loxP* sites [21], and the use of Cre recombinase to fuse two tobacco chromosomes revealed that site-specific recombination can generate large scale genome rearrangements [22].

Development of efficient Cre/*lox* systems in fungi has enabled genome engineering. In the fission yeast, *Schizosaccharomyces pombe*, Cre/*lox* has been used for gene tagging and gene replacement by cassette exchange [23]. In *Candida* [24], *Yarrowia*

lipolytica [25], *Aspergillus* [9, 10], *Cryptococcus neoformans* [26, 27], various endophytes [28], *Neurospora crassa* [29], and *Trichoderma* [30], Cre/*lox* has been used to remove heterologous genes or swap alleles.

In our efforts to develop *Fusarium graminearum* into a more facile genetic system for chromatin studies we have made use of the Cre/*lox* system for several years. Here we summarize the various approaches taken to excise marker genes from the *Fusarium* genome after transformation with a series of marker genes that are flanked by *loxP* sites. We also present an improved transformation procedure for *F. graminearum*.

2 Materials and Methods

2.1 Strains

Transformation of *Escherichia coli* NEB5αF' was carried out according to standard procedures [31]. Transformants were selected on LB agar supplemented with 100 μg/mL ampicillin. Two wild-type *Fusarium graminearum* lab strains, PH-1 and Z-3639 [32–35], were used for transformations, protoplast fusions, and genetic crosses. Strains were maintained as hyphal fragments on agar pieces or asexual spores in 25% glycerol at −80 °C.

2.2 Molecular Biology

Methods for plasmid isolation, restriction enzyme digests, ligation, and cloning were standard [31]. Methods for fungal genomic DNA isolation and Southern analyses have been described previously [36].

2.3 Plasmids

Plasmids constructed for this study are based on pFOLT4R4 (kindly provided by Dr. H. Corby Kistler), which carries a partial duplication of pUC sequences [37, 38]. Based on deletion studies with pFOLT4R4, these sequences act as autonomously replicating sequences in *F. graminearum*.

2.3.1 pLC14

pFOLT4R4 was digested with *Pvu*II to isolate a 3.64 kb fragment (partial inverted pUC repeat with 20 telomeric DNA repeats) and ligated into pBluescript SKII+ that had been linearized with *Sma*I to obtain pLC14 (6.6 kb).

2.3.2 pLC15

The 1.4 kb fragment bearing the hygromycin phosphotransferase (*hph*) gene was released from pCT74 [39] by digestion with *Sal*I and ligated into the *Sal*I site of pLC14 to obtain pLC15 (7.6 kb).

2.3.3 pLC28

A 2.2 kb fragment containing the *N. crassa ccg-1* promoter and the *cre* recombinase gene was isolated from pCCG::Cre [29] by digestion with *Eco*R1 and *Pvu*II and inserted into pLC15 that had been linearized with *Eco*R1 and *Eco*RV to obtain pLC28 (9.8 kb). All constructs were validated by restriction digests and sequencing.

2.4 Transformations

Fusarium transformations were carried out using plasmids with autonomously replicating sequences, circular or linear plasmids, or overlapping PCR fragments ("split marker" transformation) [40, 41]. Primer sequences are listed in Table 1, and strains used or constructed are listed in Table 2. We provide an updated step-for-step transformation protocol (*see* **Protocol 1**). Transformation efficiencies for eight separate experiments obtained by following our protocol are shown in Table 3.

2.5 Auxotrophic Strains

Auxotrophs were constructed by a split marker fusion PCR method described previously [40, 41]. Selection was on regeneration medium (RM) supplemented with 100 μg/mL (1×) G418, and primary transformants were transferred to YPD agar (0.3% yeast extract, 1% Bacto peptone, 2% dextrose, 1.5% agar) with 1× G418. Conidiation was induced by shaking small agar plugs of mycelia in carboxymethylcellulose (CMC) medium supplemented with Ca-pantothenate (Pan; 10 mg/L), serine (Ser; 500 mg/L), thiamine (Thi; 10 mg/L), tryptophan (Trp; 40 mg/L), nicotinamide (Nic; 8 mg/L), aromatic amino acids (Aro; 40 mg/L Phe, 40 mg/L Trp, 250 μg/L PABA, 40 mg/L Tyr), methionine (Met; 500 mg/L), or histidine (His; 500 mg/L) at final concentrations suggested for *N. crassa* [42], as well as 1× G418 at room temperature (~22 °C) for 1 week. Conidia were spread on YPD agar with required supplements and 1× G418, allowed to germinate overnight at room temperature, and individual germlings picked into wells of 48-well plates filled with YPD agar containing required supplements and 1× G418. Germlings were considered single-spored homokaryotic strains for subsequent screening and DNA analysis. Spot assays were carried out on Vogel's minimal medium (VMM, [42]) supplemented with 2% (w/v) sucrose, and with or without amino acid or vitamin supplements to identify auxotrophic transformants (Table 2). Southern analyses confirmed targeted integration of the *neo* cassette into the marker loci; strains with additional unlinked fragments were not used for further studies (data not shown).

2.6 Crosses

An updated procedure for recovering recombinants from sexual crosses is provided (*see* **Protocol 2**). Crosses were carried out on carrot agar [43].

3 Results

There were multiple reasons to attempt expression of the Cre/*lox* system in *F. graminearum*, including the need for efficient "marker recycling," genome engineering by targeted translocations or inversions, and inducible excision of target genes or repetitive regions. Here we focus on efficient marker recycling. Even though

Table 1
Primers used in this study

Fragment	OMF	Name	Sequence (5′–3′)
loxP marker	OMF 1148	loxPF	GCGGATCCGACACATCCATGTTGATATGCG
	OMF 48	locPR	GATAAGCTTGATATCGAATTCTTACTTGTCTTGATTAACAGCAAGCTGCA
SM *neo*	OMF 3557	neo SM F	TTGAACAAGATGGATTGCACG
	OMF 3558	neo SM R	AGGCGATGCGCTGCGAATCGG
SM *hph*	OMF 1816	hph SM F	AAAAAGCCTGAACTCACCGCGACG
	OMF 1817	hph SM F	TCGCCTCGCTCCAGTCAATGACC
Fg his3	OMF 1696	Fghis35F	CCTCCACCGGCATTTCCTCCGCGC
	OMF 1712	Fghis35R	GATAAGCTTGATATCGAATTCTTACTTGTAGGCAGGTAAATTGAATTGAC
	OMF 1713	Fghis33F	TGCTATACGAAGTTATGGATCCGAGCTCGGAAATTTAGAATACAAAAGGC
	OMF 1697	Fghis33R	CACTACTTCCCTCCTGTAAGCACG
Fg pan2	OMF 1694	Fgpan25F	CCTCCACCGGCATTTCCTCCGCGC
	OMF 1710	Fgpan25R	GATAAGCTTGATATCGAATTCTTACTTGTGAGGGATAGGAGTGGGTCTTTGTG
	OMF 1711	Fgpan23F	TGCTATACGAAGTTATGGATCCGAGCTCGACCCTTAGAGTTGCTTGTGGA
	OMF 1695	Fgpan23R	GGTATTAGAAAAGTTGGCTGCTGG
Fg aro1	OMF 3716	Fg_aro1_5neoF_Bam	GCGGATCCGACACATCCATGTTGATATGCG
	OMF 3717	Fg_aro1_5neoR	GATAAGCTTGATATCGAATTCTTACTTGTCTTGATTAACAGCAAGCTGCA
	OMF 3718	Fg_aro1_3neoF	TGCTATACGAAGTTATGGATCCGAGCTCGGGATTTATGCTTGTCGATGAC
	OMF 3719	Fg_aro1_3neoR_Bam	GCGGATCCGGTCCATCTGCGACTGACATGCC
Fg met8	OMF 3724	Fg_met8_5neoF_Xba	GCCTCTAGAAGTCCGGAGCCAATGACTCCG
	OMF 3745	Fg_met8_5neoR	GATAAGCTTGATATCGAATTCTTACTTGTCTTTGCGGTTGTTGCAACTGC

(continued)

Table 1
(continued)

Fragment	OMF	Name	Sequence (5′–3′)
	OMF 3746	Fg_met8_3neoF	TGCTATACGAAGTTATGGATCCGAGCTCGATGTCAACCAGCACTATGTGT
	OMF 3747	Fg_met8_3neoR_Xba	GCCTCTAGAACAGGACGCTCATCAGGCGAGC
Fg ser3	OMF 3748	Fg_ser3_5neoF_Xho	CGGCTCGAGGAATTTACAGTATTATAGAGCG
	OMF 3749	Fg_ser3_5neoR	GATAAGCTTGATATCGAATTCTTACTTGTTGTATCGTGCGCGCGGGGTG
	OMF 3750	Fg_ser3_3neoF	TGCTATACGAAGTTATGGATCCGAGCTCGGTTGTTAGTGTTGGATGAGG
	OMF 3751	Fg_ser3_3neoR_Xho	CGGCTCGAGCCCCGAGATTCTGGTGCGTC
Fg thr2	OMF 3752	Fg_thr2_5neoF_Xba	GCCTCTAGACGGTTGTAGTCTCGGTTGGGC
	OMF 3753	Fg_thr2_5neoR	GATAAGCTTGATATCGAATTCTTACTTGTTTTGTCTGCTTGTTGATGCTTG
	OMF 3754	Fg_thr2_3neoF	TGCTATACGAAGTTATGGATCCGAGCTCGTAATTGTGAGATGAAAAGGTG
	OMF 3755	Fg_thr2_3neoR_Xba	GCCTCTAGAGGCTCAGCACTTGTAGCTGTC
Fg nic2	OMF 3756	Fg_nic2_5neoF_Xho	CGGCTCGAGGATAGGGCAAGAGGATACAGC
	OMF 3757	Fg_nic2_5neoR	GATAAGCTTGATATCGAATTCTTACTTGTACTGAAGTTGCCCTTGATTTG
	OMF 3758	Fg_nic2_3neoF	TGCTATACGAAGTTATGGATCCGAGCTCGGGCGACTCCTCGACGGGAATC
	OMF 3759	Fg_nic2_3neoR_Xho	CGGCTCGAGGATGTTTCTTATGGACCGGGC
Fg pdx1	OMF 3764	Fg_pdx1_5neoF_Bam	GCGGGATCCATGGTAAGACTGATCATGTTG
	OMF 3765	Fg_pdx1_5neoR	GATAAGCTTGATATCGAATTCTTACTTGTTGTGAATTGGATATAGAAGAT
	OMF 3766	Fg_pdx1_3neoF	TGCTATACGAAGTTATGGATCCGAGCTCGGGAGAAGGTTTATTCAACATG
	OMF 3767	Fg_pdx1_3neoR_Bam	GCGGGATCCGCCATATGCGACTGGTCCAG
Fg thi2	OMF 3760	Fg_thi2_5neoF_Bam	GCGGGATCCTTCGTCGCCAGGGCGATGTCC
	OMF 3761	Fg_thi2_5neoR	GATAAGCTTGATATCGAATTCTTACTTGTGATGTTGTGAGGGACTTTAG
	OMF 3762	Fg_thi2_3neoF	TGCTATACGAAGTTATGGATCCGAGCTCGGGGTTGAGATAACAGATGTAT
	OMF 3763	Fg_thi2_3neoR_Bam	GCGGGATCCTGTACGCATCCAACAGCTGAC

Table 2
Strains used and generated in this study

Strain	Genotype	Validation
FMF1	wild type; PH-1	Gift from Frances Trail, sequenced
FMF70	wild type; Z-3639	Gift from Corby Kistler, sequenced
FMF382	*F. asiaticum* (*Fa*) SCK04	Gift from Sung-Hwan Yun
FMF433	*Δmat1-1-1::hph*$^+$	Gift from Sung-Hwan Yun
FMF434	*Δmat1-1-2::hph*$^+$	Gift from Sung-Hwan Yun
FMF24	*hH4-2-sgfp*$^+$*-hph*$^+$	This study
FMF78	*hpo-tomato*$^+$*-neo*$^+$	This study
FMF159	*Δpan2::neo*$^+$	This study
FMF160	*Δhis3::neo*$^+$	This study
FMF177–FMF179	*pan2*$^+$ *neo*$^-$	This study
FMF180	*Δpan2::loxP; cre*$^+$*-hph*$^+$	This study
FMF189–FMF193	*his3*$^+$ *neo*$^-$	This study
FMF197	*Δpan2::loxP*	This study
FMF212, FMF213	*Δhis3::loxP*	This study
FMF220–FMF223	*Δhis3::loxP*	This study
FMF248	*Δkmt6::neo*$^+$	This study
FMF277	*Δsuz12::hph*$^+$	This study
FMF278	*Δeed::hph*$^+$	This study
FMF361	*Δkmt6::loxP*	This study
FMF373	*zea2::pan2*$^+$*; pan2::loxP*	This study
FMF388	*Δtrp3::neo*	This study
FMF403	*son1-rfp*$^+$*-neo*$^+$	This study
FMF474	*Δpdx1::neo*	This study
FMF475	*Δser3::neo*	This study
FMF476	*Δthi2::neo*	This study
FMF479	*Δaro1::neo*	This study
FMF519	*Δnic2::neo, Δmat1-1-1::hph*	This study
FMF520	*Δmet8::neo, Δmat1-1-1::hph*	This study
FMF597	*Δnic2::neo, Δpdx1::neo*	This study
FMF601	*Δnic2::loxP, Δpdx1::loxP*	This study

(continued)

Table 2
(continued)

Strain	Genotype	Validation
FMF632	*Δthi2::neo, Δmat1-1-1::hph*	This study
FMF633	*Δser3::neo, Δmat1-1-1::hph*	This study
FMF634	*Δtrp3::neo, Fa trp3*$^{+}$	This study
FMF635	*neo*$^{-}$*, Fa ser3*$^{+}$	This study
FMF636	*neo*$^{-}$*, Fa thi2*$^{+}$	This study
FMF638	*Δnic2::neo, Δser3::neo*	This study
FMF639	*Δnic2::neo, Δtrp3::neo*	This study
FMF883	*Δsuz12::loxP*	This study
FMF915	*hH4-2-sgfp*$^{+}$*::loxP*	This study
FMF921	*son1-rfp*$^{+}$*::loxP*	This study
FMF932	*mat1-1-2::hph*$^{+}$*; cre*$^{+}$*-hph*$^{+}$	This study
FMF961–FMF962	*hpo-rfp*$^{+}$*::loxP; cre*$^{+}$*-hph*$^{+}$	This study
FMF963 – FMF965	*hpo-rfp*$^{+}$*::loxP*	This study
FMF966	*Δeed::loxP*	This study

Strains will be submitted to the Fungal Genetics Stock Center and are also available upon request.

much molecular biology work has been done with *F. graminearum*, there is still a lack of useful positively selectable genetic markers when compared to model organisms such as budding or fission yeast, and *Neurospora* or *Aspergillus*. Most labs use resistance to hygromycin B (Hyg^{+}), encoded by the *hph* gene [44], resistance to G418 or geneticin (Neo^{+}), encoded by the *neo* gene [45, 46], or resistance to nourseothricin (Nat^{+}), encoded by the *nat* gene [47], limiting molecular genetic studies to three markers for DNA-mediated transformation. Thus, marker recycling is an attractive proposition for molecular genetics in this fungus, especially when complicated genotypes need to be examined.

3.1 Generating pan2 and his3 Auxotrophic Markers by Gene Deletion

Auxotrophic strains are expected to accelerate quantitative genetic analyses in *F. graminearum*, and thus we generated strains in which biosynthetic genes have been removed. We deleted the coding DNA sequences of *his3* (FGSG_00865), *pan2* (FGSG_09312), *aro1* (FGSG_04287), *met8* (FGSG_10825), *nic2* (FGSG_07058), *pdx1* (FGSG_05035), *ser3* (FGSG_08529), *thi2* (FGSG_07264), and *thr2* (FGSG_05713), and replaced them with the *neo*$^{+}$ or *hph*$^{+}$

Table 3
Transformation efficiency obtained with Protocol 1

	Test	1	2	3	4
Linear	cfu	4	1	5	8
	DNA (μg)	4.6	5.5	4.4	8.3
	Efficiency (cfu/μg)	0.9	0.2	1.1	1.0
Split marker	cfu	80	49	59	50
	DNA (μg)	1.2	2.1	2.7	1.8
	Efficiency (cfu/μg)	67.2	23.3	22.2	27.8

Transformation efficiency of gene replacement in an *eed*$^+$ strain with wild-type NHEJ system using a mutant allele of *eed-sgfp*$^+$*-hph*$^+$, supplied as linearized plasmid or as overlapping split marker fragment. The number of protoplasts used for each transformation was standardized to 2×10^7. Colony forming units (cfu) were counted after a week of incubation at room temperature following overlay. Transformation efficiency is expressed as colonies per μg of DNA used.

genes (Table 2). All gene names are in accordance with those in use for *Neurospora crassa* [42]. The new *F. graminearum* markers are localized on different regions of the four chromosomes (Fig. 1a), so that they are mostly unlinked and thus can be easily used as markers in crosses. Cassettes encoding the *neo* or *hph* selectable markers were derived from pLC13 [40] and pGFP*::hph::loxP* [29], respectively, and deletion cassettes were generated by fusion PCR (Fig. 1b).

To test the utility of the new markers, all relevant strains were complemented by reintroducing the wild-type alleles (Table 2). For example, we integrated *pan2* and *his3* fragments generated by using the 5′- and 3′-most primers from the split-marker scheme (Fig. 1b) at the endogenous locus of FMF159 and FMF160, respectively, and selected for growth on minimal medium. Growth phenotypes were validated by Southern blot and sequencing (data not shown), resulting in strains FMF177–179 or FMF189–193. The wild type *pan2* allele was also integrated into a *his3 pan2* double mutant, in this case targeted to disrupt the gene encoding *zea2*, required for the production of zearalenone (FMF373). Except for *nic2*, all wild-type alleles for the selectable markers obtained from *Fusarium asiaticum* (*Fa*; strain SCK04 was a gift from Sung-Hwan Yun) were able to complement the deletion alleles, mostly showing ectopic insertions of the complementing *Fa* alleles; *Fa* alleles were used because their sequences are sufficiently different to avoid

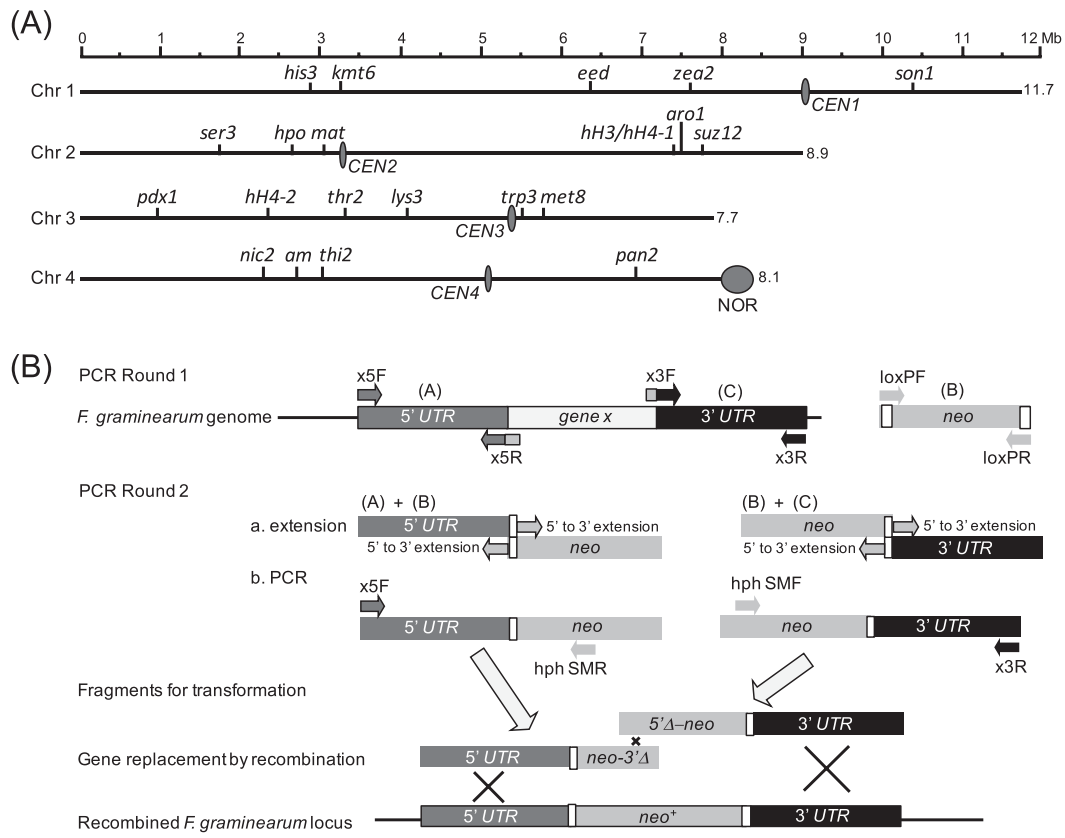


Fig. 1 Partial *F. graminearum* genetic map and gene replacement scheme based on fusion PCR and split marker selection. (**a**) Localization of genes mentioned or used in this study are shown on the four *F. graminearum* chromosomes. *CEN1–CEN4* mark the centromeric regions, and NOR marks the nucleolus organizing region, i.e., the single rDNA cluster on Chr. 4R. (**b**) The gene replacement scheme relies on PCRs (PCR Round 1) that amplify the 5′ and 3′ flanks (*5′ UTR* and *3′ UTR*, respectively) of the gene of interest (*gene x*) and the dominant selectable marker (here *neo*, although it could be *hph*, *nat*, *or* any of the auxotrophic markers developed as part of this study). The x5R and the x3F primers have 30-nt overlap with the loxPF and loxPR primers, respectively, which allows annealing and extension during the first step of PCR Round 2. After ten extension and melting cycles, the x5F + *hph* SMR or *hph* SMF + x3R primers are added to generate two separate fragments that contain the 5′ flank or 3′ flank with partially overlapping *neo*$^{-}$ marker fragments ("split marker"). These fragments are mixed and transformed into *F. graminearum* protoplasts as described in Protocol 1, relying on homologous recombination (marked by "X") in three regions to integrate the complete *neo*$^{+}$ fragment at the endogenous locus of *gene x*

efficient homologous recombination in future experiments. Strains FMF635 and FMF636 were Ser^{+} and Thi^{+} but G418-sensitive, and showed gene replacement of the *neo*$^{+}$ with the *Fa ser3* and *thi2* alleles (Table 2). In summary, our results suggest that the *his3*, *pan2*, *aro1*, *met8*, *pdx1*, *ser3*, *thi2*, and *thr2* genes can be used as new selectable markers in transformation experiments with *F. graminearum*.

3.2 Cre-Mediated Recombination via Transformation with an Episomal Cre Plasmid

To test whether the Cre/*lox* system can be applied in *F. graminearum*, we transformed the *pan2*::*neo*+ strain (FMF159) to Hyg resistance by using 2 μg of supercoiled pLC28, which contains an *hph* allele lacking *loxP* sites (Fig. 2). Transformants were plated on RM and overlaid the next day with VMM supplemented with 28% sucrose, 0.7% agar, Pan, and 100 μg/mL (1×) Hyg. After 1 week, 17 small and two normal-sized colonies grew through the second layer of medium. These colonies were picked to YPD agar supplemented with Pan and 1× Hyg. Conidiation was induced by transferring small agar plugs of mycelia in CMC medium supplemented with 0.5× Hyg and incubated with shaking at room temperature for 1 week. Spores were spread onto YPD agar supplemented with Pan and 1× Hyg, incubated overnight at room temperature, and germlings picked into wells of 48-well plates filled with YPD agar with Pan and 1× Hyg to obtain homokaryotic strains for subsequent screening and DNA analysis.

After 2 days of growth, single-spored transformants were spotted onto YPD agar supplemented with Pan and 1× G418. All twelve slow-growing and both normal-growing transformants had lost G418 resistance, suggesting removal by action of the Cre recombinase. To validate this, we amplified fragments spanning the deleted *neo* gene from one normal growing strain (FMF180), followed by Sanger sequencing, which confirmed the absence of the *neo* gene but presence of a single *loxP* site.

Purified transformants were also screened for loss of Hyg resistance, which was expected if the episomal pLC28 is lost. After a single passage on YPD medium (spanning a period of ~5 days), six slow-growing and both normal-growing strains were spotted onto YPD agar supplemented with either Pan or Pan with 1× Hyg. All six slow-growing strains were no longer Hyg-resistant, while both normal-growing strains retained Hyg resistance, suggesting stable integration into the *F. graminearum* genome. We had previously observed this with strains transformed with the original pFOLT4R4 and the pFOLT-derived pLC15 (Fig. 2b). Southern blotting analyses further verified the loss of the *hph* gene, and indeed all of pLC28, in the slow-growing transformants that originally contained the episome. In contrast, the *hph* gene was maintained as an ectopic integration in the genomes of the two normal-growing transformants.

In summary, these experiments showed that the original P1 Cre recombinase gene, driven by the *N. crassa ccg-1* promoter, can excise floxed genes. This was true when the Cre gene (*cre*$^+$-*hph*$^+$) was expressed from an episomal plasmid (pLC28) or when *cre*$^+$-*hph*$^+$ integrated ectopically in the genome. FMF180, a strain with an ectopic integration of *cre*$^+$-*hph*$^+$ and the *pan2::loxP* allele (i.e., *pan2* removed by Cre recombination) was chosen for additional experiments.

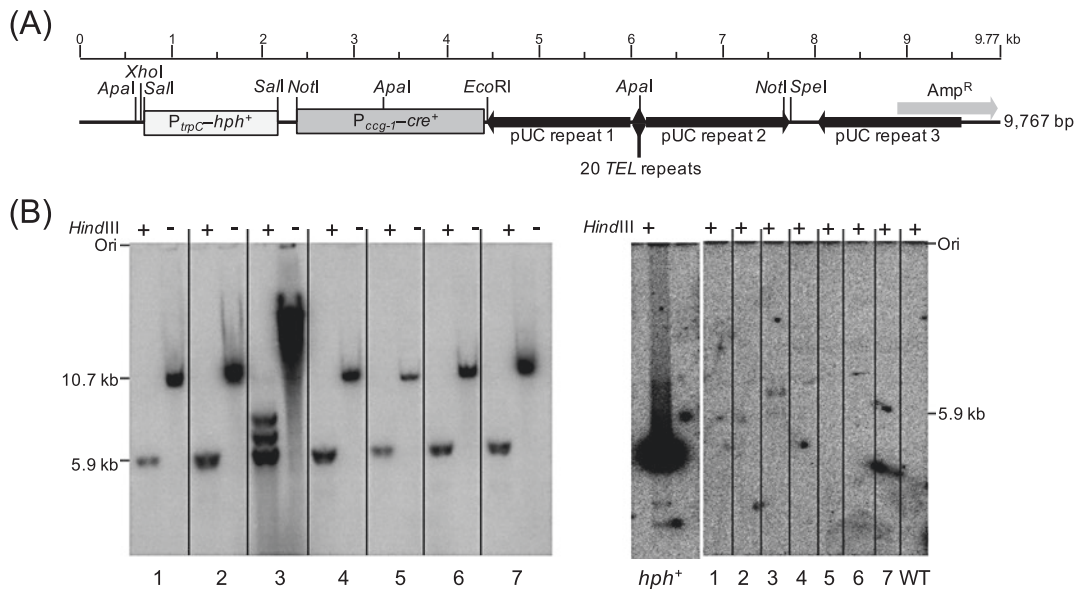


Fig. 2 Partial map of pLC28, an episomal Cre expression plasmid and analyses of the original episome, pFOLT4R4, in *F. graminearum*. (**a**) pLC28 is based on pFOLT4R4 [37], except that the wild-type P1 *cre* gene driven by the *N. crassa ccg-1* promoter [29] was inserted, and pUC repeats 1 and 2 were shortened (*see* Subheading 2). (**b**) Transformation with pFOLT4R4 resulted in strains in which the plasmid was maintained as an episome (# 1, 2, 4–7) or inserted into the genome (#3). Digestion of genomic DNA from strains grown on Hyg-containing medium resulted in the expected 5.9 kb fragment that contains the *hph* gene (probe), while a 10.7-kb fragment is expected for uncut extrachromosomally maintained plasmids; uncut genomic DNA of transformant #3 runs at ~60 kb, suggesting that the plasmid was inserted into the genome (left panel). After growth in the absence of selection, the episomal plasmids are quickly lost (transformants #1, 2, 4–7), while a ~8-kb band is visible for transformant #3, which has pFOLT4R4 inserted into the genome (right panel). Strain PH-1 without *hph* (WT) shows no hybridization, while an *hph*-containing plasmid shows strong hybridization (*hph*$^+$); lanes were run on the same gel but some lanes were removed for clarity. The molecular assay was reflected in growth (left panel) or no growth (right panel) on Hyg-containing medium for #1, 2, 4–7, suggesting the strains can be cured of pFOLT4R; this was also the case for pLC28, the *cre*$^+$-*hph*$^+$ plasmid

3.3 Cre-Mediated Recombination via Protoplast Fusion

Deleting floxed genes or chromosomal regions by transformation with a Cre-containing plasmid is not the most convenient or fastest way. We thus attempted to carry out heterokaryon fusions with *F. graminearum* auxotrophs (i.e., *his3* and *pan2* strains) on both minimal and YPD medium. This proved impossible as the observed rate of heterokaryon formation was very low; thus, we turned to protoplast fusion (Figs. 3 and 4).

To generate a *his3::loxP* strain, protoplasts were generated separately for FMF160 (*his3::neo*$^+$) and FMF180 (*pan2::loxP*, *cre*$^+$-*hph*$^+$) as previously described [40] and washed twice in STC (*see* **Protocol 1**). Protoplasts were resuspended in 1 mL of 30% PEG 8000 individually (2×10^6 as controls) and as a mixture of 2×10^6 protoplasts of each strain. Incubation was carried out at room temperature for 15 min. Cells were washed twice in 0.6 M NaCl, as described [48]. Protoplasts from FMF160 or FMF180 (2×10^6 of

each) were plated separately on VMM supplemented with 28% sucrose as controls. For protoplast fusions, aliquots of 4×10^5 protoplasts from the FMF160 and FMF180 mixture were spread on nine petri dishes containing VMM supplemented with 28% sucrose. Fused auxotrophs will form heterokaryons that are able to survive on minimal medium because of genetic complementation while nonfused protoplasts lacking histidine or pantothenic acid will die.

The FMF160–FMF180 mixture yielded colonies at a frequency of 4×10^3. In contrast, no colonies grew on the negative control plates (FMF160 or FMF180 incubated separately). After five days, heterokaryons were picked onto YPD agar and spot-tested on YPD supplemented with 1× G418, YPD supplemented with 1× Hyg, VMM supplemented with 2% sucrose and His, or VMM supplemented with Pan. No colonies survived exposure to G418, suggesting successful removal of the *neo* gene in all five colonies screened. A G418-sensitive, Hyg-resistant (*cre*$^+$-*hph*$^+$) colony able to grow without Pan (i.e., a FMF160/FMF180 heterokaryon) was subjected to sexual development on carrot agar to isolate the nuclear type where *his3* was replaced by a single *loxP* site, *his3::loxP*. Ascospores were germinated on VMM supplemented with His to eliminate the *cre*$^+$-*hph*$^+$ and *pan2::loxP* nuclei. Germlings were isolated the next day and spot-tested on all relevant media. A G418-sensitive strain, unable to grow on Hyg or VMM alone, but growing on VMM supplemented with His, was validated as *his3::loxP* (FMF220) by PCR assays (Fig. 5a).

Not all protoplast fusions occur rapidly, for example when the deleted gene results in extreme morphological phenotypes, as does deletion of the gene encoding lysine methyltransferase, *kmt6* [40]. To generate a *kmt6* deletion allele without *neo*, 5×10^5 protoplasts from each FMF180 (*pan2::loxP*, *cre*$^+$-*hph*$^+$) and FMF248 (*kmt6::neo*$^+$) were mixed, plated on VMM supplemented with 28% sucrose, and overlaid the next day with VMM supplemented with 28% sucrose and 1× Hyg. Twenty days later, one of two slow-growing colonies was picked and transferred to VMM supplemented with 2× Hyg. Mycelial fragments from a mature colony were spotted on VMM and incubated at room temperature for nine days. Agar plugs from this colony were grown in CMC medium without selection by shaking at room temperature for 5 days, and conidia were germinated on YPD for 2 days at room temperature. To speed up picking, a modified version of cellophane colony lifts [42] were used to transfer germlings from YPD plates to VMM. The cellophane disc was left on the VMM plate overnight and removed the next day leaving the colonial tissue in place; discs can be reused to test growth on different media. A pointed toothpick was used to transfer a plug from the center of each of 32 individual colonies from VMM back to YPD medium containing supplements and antibiotics. Spot tests were carried out on VMM, YPD supplemented with 2× G418, and YPD supplemented with

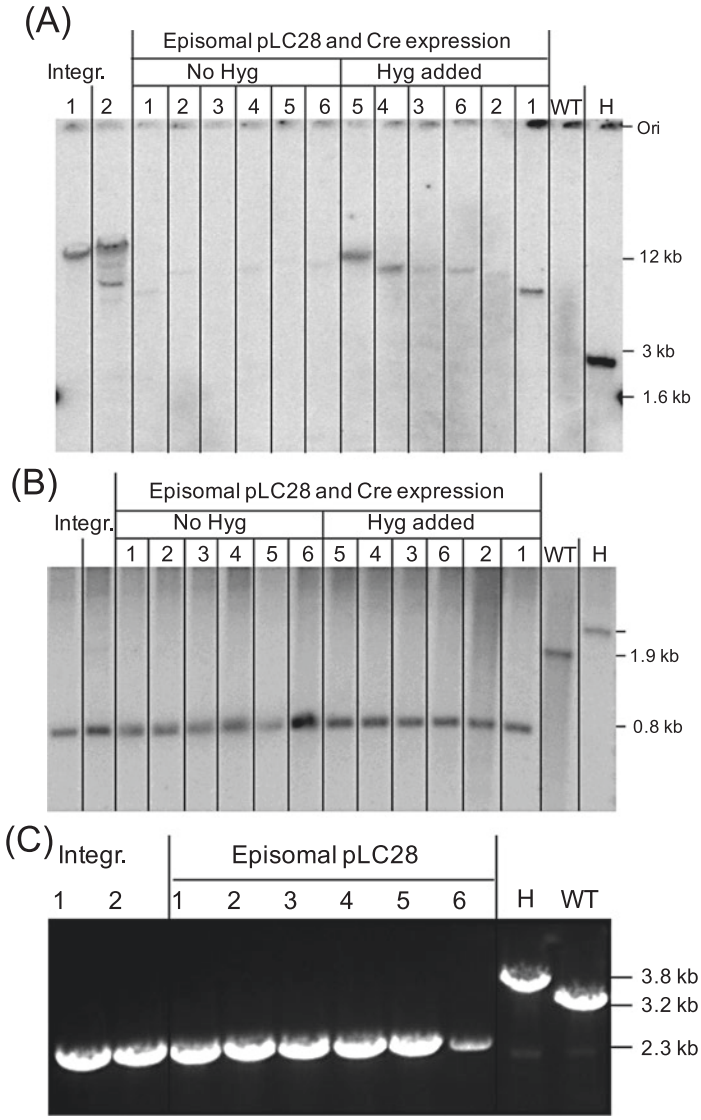


Fig. 3 Cre-mediated excision of the *pan2* gene in transformants with pLC28. (**a**) Strain FMF159 (Δ*pan2::neo*$^{+}$; host, H) was transformed with the Cre expression plasmid pLC28, resulting in two transformants in which pLC28 integrated (Integr.) and six transformants in which pLC28 was maintained extrachromosomally (#1–6). The probe for Southern analysis of genomic DNA digested with *Spe*I was the *neo*$^{+}$ gene, which is absent from PH-1 (WT). Note the lighter hybridization of the *neo* probe to DNA from strains that were no longer subjected to selection on Hyg-containing medium (~2 days of growth on "No Hyg" YPD medium; three passages to fresh medium resulted in loss of pLC28). Lane 1 shows strain FMF180 that was selected for further studies. (**b**) Hybridization of the same blot (after stripping with NaOH) with a *pan2* probe that includes 1 kb of 5′ and 3′ flanks shows the expected fragments, namely 1.9 kb for wild type (WT), 2.4 kb for the host (where *neo*$^{+}$ was integrated in place of *pan2*) and *pan2::loxP*, where the 1.7 kb *neo*$^{+}$ gene was deleted. (**c**) PCR validation of results shown in (**b**). Amplification with primers OMF1694 and OMF1695 results in a 3.2-kb band expected in wild-type PH-1 (WT). The *pan2* gene was deleted in both the integrated and extrachromosomally maintained pLC28-containing strains, resulting in a 2.3-kb band; amplification in FMF159 results in the 3.8-kb *pan2::neo*$^{+}$ gene fragment

2× Hyg to identify bright orange stunted colonies (reflecting the typical *kmt6* phenotype; [40]) growing on VMM, but not on medium containing G418 or Hyg. One strain (FMF361) was validated by PCR assays and Sanger sequencing, which confirmed the loss of the *neo* gene at the *kmt6* locus (*kmt6::loxP*).

3.4 Cre-Mediated Recombination During Crosses

Another convenient way to remove markers is by subjecting primary transformants to genetic crosses (Fig. 4). While generating single spore isolates by conidiation is usually sufficient to generate homokaryotic strains in *Fusarium*, there are instances when this is not possible, for example when a gene is essential (e.g., *F. fujikuroi kmt6*; [49]) or results in severe defects, which favors the formation

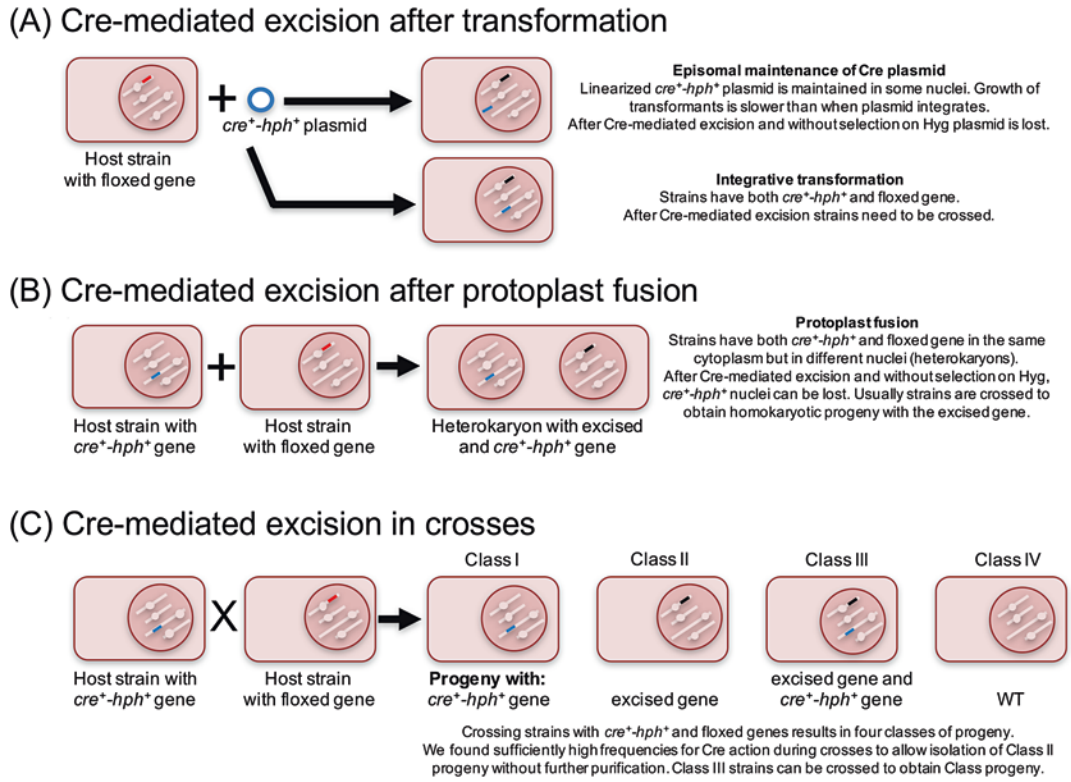


Fig. 4 Approaches used to show Cre-mediated marker excision in *F. graminearum*. (**a**) Transformation of strains bearing floxed genes with pLC28 will result in small colonies that carry pLC28 as an episome, or normal-growing strains in which pLC28 integrated into the genome of the host strain. (**b**) Protoplasts of FMF180 (*cre*$^{+}$*-hph*$^{+}$) and successor strains can be fused to protoplasts of strains with floxed genes, resulting in heterokaryons. Cre-mediated excision will occur at high frequency (>70%), and single spores of marker-less homokaryotic strains can be isolated on medium without Hyg. (**c**) Perhaps the simplest approach is to use Cre-mediated excision during crosses to isolate homokaryotic progeny. While we anticipated that we would need two successive crosses to first generate Class III progeny to excise the marker, and secondly to cross out the *cre*$^{+}$*-hph*$^{+}$ gene, we found that Cre was active during the cross and resulted in excision frequencies of ~15%, sufficient to find the desired progeny in most cases

of unusual heterokaryotic conidia. Crosses also allow combining desired traits by recombination. Even though *F. graminearum* is homothallic, recombination can be assayed easily by use of two markers in compatible strains, or by forcing crosses with parents where either one or both lack mating type idiomorphs [50–52].

We induce the sexual cycle by growth on carrot agar (*see* **Protocol 2**; [43]). For Cre-mediated recombination, FMF180 (*pan2::loxP*; *cre*$^{+}$*-hph*$^{+}$) was grown for 3–5 days at room temperature under ambient light conditions until the mycelium reached the edge of the Petri dish; for optimal fertility it is important not to stack plates that are incubated right side up. Mycelia were removed completely with a sterile toothpick and 800 μL of 2.5% Tween 80 with 1×10^5 conidia of the parent strain (bearing a floxed marker gene) was spread evenly on the surface of FMF180 with a sterile glass rod. Perithecia formed within a week, and once ascospores were shot to the plate lid (7–9 days after fertilization) plates were turned upside down. Ascospores were harvested and stored in sterile water at 4 °C for up to 3 months.

Because FMF180 bears both mating type idiomorphs it is self-fertile and strains in which genes are to be removed require easily selectable dominant or morphological markers; we found that selection against the *pan2* deletion allele of FMF180 increased the likelihood of finding recombinant progeny. Thus, germlings incubated in YPD without Pan but with supplements required for the desired genotype were picked to 48-well plates containing YPD with required supplements. After 2 days of growth, progeny were spot-tested on media containing antibiotics and required supplements. Progeny of desired phenotype were selected for DNA isolation to validate genotypes by both PCR assays and Sanger sequencing.

We found that Cre acted during the relatively short heterokaryotic or diploid phase during crosses in *F. graminearum*. Seven types of crosses, involving the single-spored but possibly still heterokaryotic FMF180 strain and different strains with floxed alleles resulted in one to five progeny each in which the *neo* or *hph* marker was excised by Cre (Table 2). The frequency of Cre-mediated excision varied between 1/169 to 1/16 total progeny isolated, or 1/12 to 3/5 preselected progeny that were tested based on their expected genotypes. While scoring of morphological defects (e.g., slow growth and orange pigmentation of *kmt6*) or scoring for fluorescence (e.g., HP1-RFP, SON1-RFP, hH4-GFP) was reasonably successful (1/12 to 1/4 strains showed marker excision), the highest frequencies were obtained by counterselecting on G418- or Hyg-containing medium (3/5 strains showed marker excision).

We generated backcrossed Cre^{+}-Hyg^{R} strains by crossing FMF180 to strains lacking different genes in the *mat1-1* idiomorph (FMF433 and FMF434, respectively), obtaining FMF932 (*mat1-1-2::hph*$^{+}$*; cre*$^{+}$*-hph*$^{+}$). Crossing homokaryotic FMF932 to FMF248

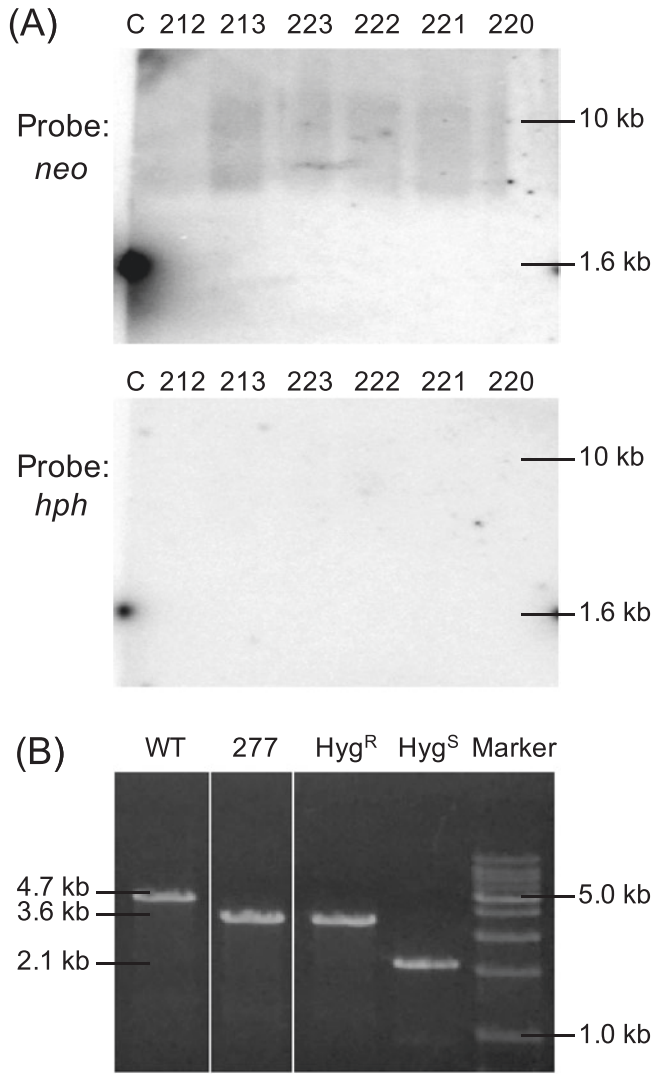


Fig. 5 Cre-mediated marker excision in protoplasts or during crosses of *F. graminearum.* (**a**) Protoplasts of FMF160 (Δ*his3::neo*$^+$) were forced into a heterokaryon with protoplasts of FMF180 (Δ*pan2::loxP*; *cre*$^+$*-hph*$^+$) by growing on RM medium without supplements. After growth for 3–5 days, Cre had excised the *neo*$^+$ gene from all strains tested (a 1.9-kb fragment would be expected; C is a control plasmid band of 1.7 kb). Strains were transferred to medium with His but lacking Pan to identify strains in which the heterokaryon had broken down to the Δ*his3::loxP* homokaryon. If *hph* were still present an 8.8-kb fragment would be expected; C is a control plasmid band of 1.7 kb). Strains FMF212–223 are described in Table 2. (**b**) Cre-mediated excision during a cross to FMF180 (Δ*pan2::loxP; cre*$^+$*-hph*$^+$) to FMF277 (Δ*suz12::hph*$^+$). SUZ12 is one subunit of the Polycomb Repressive Complex 2 [40]. Amplification with *suz12* primers OMF2775 and OMF2778 resulted in a 4.7-kb fragment in wild-type PH-1 (WT), while replacement of *suz12* with *hph*$^+$ resulted in the expected 3.6-kb band (FMF277; Δ*suz12::hph*$^+$). Hyg resistant (Hyg^R) or Hyg sensitive (Hyg^S) progeny were isolated, and several slow growing Hyg^S strains showed a 2.1-kb fragment expected after excision of *hph*$^+$ (FMF883; Δ*suz12::loxP*)

(*kmt6::neo*$^+$) resulted in predictable ratios of progeny, and an estimated Cre-mediated recombination frequency of ~16.3% for *kmt6::neo*$^+$ during the cross, while for a parallel cross with *dis2::neo*$^+$ this frequency was ~40% (Table 4). One example for molecular validation of removal of the *hph* gene from FMF277 is shown in Fig. 5b.

4 Discussion

4.1 Constructing Auxotrophic Markers for Genetic Studies with F. graminearum

Widely considered genetically "relatively intractable" [53], *F. graminearum* (aka *Gibberella zeae* in its sexual, teleomorph form; [54]) has been developed into a model organism to study basic biological questions, ranging from population and evolutionary biology to biosynthetic gene clusters, genetics of sexual development, and chromatin-mediated gene regulation [40, 54–58]. All studies have greatly benefited from the development of high-throughput DNA sequencing methods, which allow large-scale comparative approaches. For those who are interested in molecular mechanisms, however, availability of fundamental molecular genetic techniques is essential. While DNA-mediated transformation systems have been available for many years, the number of genetic markers and the development of quantitative genetics have lagged behind.

Outcrossing of *F. graminearum* occurs in nature but has been described as generally poor in most wild-type strains [34, 59]. Two strains from the USA have become the most widely used genetic backgrounds for molecular studies, namely Z-3639 [34] and PH-1, a strain selected for its synchronized and prolific fruiting [32]. By using dominant selectable markers (like *hph* and *neo*) or following segregation of *nit* (*nitrate nonutilizing*) mutations [34, 43] recombinants can be isolated by outcrossing of homothallic strains, but these procedures do not lend themselves to quantitative genetic analyses because most perithecia are derived from selfing events. Thus, several sets of matched strains with reciprocal deletions of the *mat1-1* and *mat1-2* locus (or individual genes in the two mating type idiomorphs) have been constructed [52, 60, 61]. Using these strains facilitates genetic analyses, in essence generating a heterothallic fungus, similar to *N. crassa*, the original genetic workhorse [42].

Earlier studies resulted in the development of nitrate nonutilizing (*nit*) mutants in four complementation groups (*nit1*, *nit3*, *nnu*, NitM; [34]) but the use of chlorate in the medium can result in additional unwanted spontaneous mutations. Traditional auxotrophs had also been selected by standard methods, resulting in isolation of mutants requiring adenine (*ade*), arginine (*arg*), histidine (*his*), a temperature-sensitive mutant (*uts*), and a mutant deficient in glutamate dehydrogenase (*gdh*) [53]. These mutants,

Table 4
Progeny derived from crosses between a Cre-expression strain and strains with an (a) floxed *kmt6::neo*$^+$ or (b) *dis2-RFP::neo*$^+$ allele

(a) FMF932 (*cre*$^+$*-hph*$^+$, Δ*mat1-1-2::hph*$^+$) × FMF248 (Δ*kmt6::neo*$^+$)						
Cre$^+$/Hyg^R	Mat$^-$/Hyg^R	G418^R	*kmt6*$^+$	Progeny (number)	Expected %	Observed %
+	+	−	+			
+	−	−	+			
−	+	−	+	25	25.0	27.2
−	−	+	−	25	25.0	27.2
+	+	+	−			
+	−	+	−			
−	+	+	−	16	25.0	17.4
−	−	−	+	18	25.0	19.6
+	+	−	−	6	0	6.5
−	−	−	−	2	0	2.2
				92	100	100.1
(b) FMF932 (*cre*$^+$*-hph*$^+$, Δ*mat1-1-2::hph*$^+$) × FMF942 (Δ*dis2::neo*$^+$)						
Cre$^+$/Hyg^R	Mat$^-$/Hyg^R	G418^R	*dis2*$^+$	Progeny (number)	Expected %	Observed %
+	+	−	+			
+	−	−	+			
−	+	−	+	21	25.0	25.6
−	−	+	−	10	25.0	12.2
+	+	+	−			
+	−	+	−			
−	+	+	−	21	25.0	25.6
−	−	−	+	11	25.0	13.4
+	+	−	−	15	0	18.3
−	−	−	−	4	0	4.9
				82	100	100

Both Δkmt6 and Δdis2 can be scored by growth phenotype (pigmentation and slow growth)

however, have not been widely used in the community, perhaps because the genes were not mapped, molecular genetics was still in the early stages, and many scientists, especially plant pathologists who were most likely to work with *Fusarium* species, were

reluctant to work with metabolic mutants that were likely affected in pathogenicity [62, 63]. Here we developed several well-characterized strains that are deficient in seven additional genes, generating auxotrophic markers on all four chromosomes that can be used for standard genetic analyses. We also showed that the wild-type genes from *F. graminearum* or its close relative, *F. asiaticum*, can be used to complement mutants in these seven genes by gene replacement at the endogenous locus or ectopic integration elsewhere in the genome. In combination with the Cre-mediated excision of traditional markers, such as *hph* and *neo*, this brings the number of useful markers in *F, graminearum* to at least a dozen, which will enable molecular geneticists to examine and construct strains with complicated genotypes.

4.2 Removing Floxed Markers from Laboratory Strains

In the original *S. cerevisiae leu2* assay, 100% *LEU2* excision was observed within 24 h after induction of a galactose-inducible *cre* gene [1]. Such high frequencies have not been observed in other fungi, although kinetics of marker excision vary or have not been studied as carefully as in yeast. In *A. nidulans* [10], in a different study in *S. cerevisiae* [4], and in *Yarrowia lipolytica* [25] efficiencies of 70–80% were obtained, the same as measured for a xylan-inducible *cre* gene in *Trichoderma reesei* that resulted in excision of the *hph* marker gene at the *mus53* locus [30]. In this study, *cre* was targeted to the *pyr4* locus, and was thus easily replaceable by retransformation with a functional *pyr4* gene, followed by selection on medium without uridine. In fission yeast, marker exchange was successful at 6–14% [23], and frequencies in *C. neoformans* were also low [27]. In *A. fumigatus* [9] and *N. crassa* [29] replacement or excision frequencies were around 20–25%. Experiments in *Neotyphodium coenophialum*, *Neotyphodium uncinatum*, and *Epichloë festucae* resulted in very low excision frequencies (0.5–2%), but these were obtained after unselected transient expression of the Cre recombinase, and the same efficiency was detected under these conditions for excision of the *A. nidulans yA* gene [28].

In our assays, all *F. graminearum* transformants with episomal pLC28 lost the floxed marker genes within 2–3 days of incubation, much better than what has been reported for Cre expression from episomal plasmids in *Cryptococcus* [27]. In some cases, however, the smaller colony size caused by maintenance of the episomal plasmid made use of the ectopically integrated *cre* gene a more attractive alternative, so we used FMF180 more frequently for protoplast fusions or crosses. Marker loss in forced heterokaryons was also efficient and successful in every case, although the kinetics was difficult to study because the generation of forced heterokaryons by protoplast fusion has not been optimized and was deemed too difficult as a routine procedure, even though it had been suggested as a valid alternative to crosses [53].

None of the previous studies investigated whether Cre would act during a cross. We first tested this by crossing Cre expression strains to strains with floxed genes. When we used heterokaryotic primary transformants as one or both parents the observed frequencies were low, from 0.6% to 6% of total progeny but increased to 8–60% when we were able to select or screen for recombinants. Crosses with a homokaryotic Cre expression strain that also lacked one mating type gene (and was thus useful for heterothallic crosses) resulted in excision frequencies of 8–20% of the total progeny. Analogous crosses with three additional floxed genes yielded similar frequencies (12–18%; data not shown). These results suggest that it is unnecessary to isolate progeny that have both the floxed gene and the *cre*$^{+}$-*hph*$^{+}$ gene to accomplish isolation of progeny without selectable marker. Cre-recombination during the cross appears frequent enough to yield the desired genotypes.

How can the observed frequency of Cre-mediated excision be improved further? One obvious possibility is that the P1 *cre* gene has poor codon bias when expressed in several fungi, for example *N. crassa* and *F. graminearum*. Generating a codon-optimized version was useful for application of the Cre/*lox* system in *Candida* [24]. We expect that the humanized version of P1 *cre*, called *iCre* [64], will result in increased excision frequencies in both *Neurospora* and *Fusarium*, and we are currently testing this in crosses.

4.3 Improved Transformation Protocol

Transformation protocols for *F. graminearum* can be traced back to three original protocols, those of Mary Case [65], Marian Beremand (cited in [66]), and Heinz Osiewacz [67]. In our lab, we benefited greatly from advice kindly provided by Heather Hallen and William Rittenour (while in the lab of Frances Trail), which resulted in the protocol provided here. More recently alternative protocols have been described as well [68, 69]. One protocol used electroporation of spheroplasts [46], which unfortunately does not avoid the expensive and time-consuming generation of protoplasts with cell wall lysis enzymes. No transformation of mycelium or spores by electroporation alone has been achieved yet. Major advantages of our updated method (*see* **Protocol 1**) are the use of much less plasmid or fusion PCR fragment DNA (~1–2 μg versus 5–30 μg DNA) and the selection on YPD or *Fusarium* minimal medium instead of V8 or PDA medium, which greatly reduces the amount of hygromycin or G418 required to isolate primary transformants (the poorer the medium, the lower the minimum concentration of antibiotic required).

Analysis by diagnostic PCR and Southern blots of the primary transformants or progeny obtained here did not support the findings of previous studies [70], which claimed unusual and unexpected tandem integrations of high amounts (20 μg) of transforming DNA. Instead, we overwhelmingly found expected single-copy gene replacements of target genes by marker genes when using

overlapping PCR fragments or linearized plasmids. Low amounts of circular plasmid DNA resulted in very few transformants (0–2 per μg of DNA). We found a 20- to 50-fold decrease in transformation efficiency when comparing transformations with linearized plasmid to split marker fragments (Table 3). As observed in a previous study [68], we found unusually high frequencies of homologous recombination, especially when using the split marker system or when using genotypes in which genes responsible for chromatin silencing had been deleted (data not shown). Integrative multicopy integration was relatively rare when 1-kb flanks were used. In combination, these results suggest that, unlike in *N. crassa* [71] or *Aspergillus* [72], inactivation of the NHEJ system by deletion or mutation of Ku70/Ku80 seems not necessary to achieve a relative increase of transformants with correct gene targeting events in *F. graminearum*.

In summary, we generated and tested new tools for *F. graminearum* quantitative genetic analyses (single- and double-mutant strains). We showed that even the original P1 Cre recombinase can be used to excise floxed genes from the genome by three different approaches, and we present updated versions of protocols for DNA-mediated transformation and sexual crosses for *F. graminearum*.

5 Notes

5.1 Protocol 1: Improved Transformation of F. graminearum

5.3.1 Generation of Protoplasts

1. Inoculate 50 mL of CMC medium in a 100-mL Erlenmeyer flask with a small fragment of fresh mycelium (or a few microliters of conidia from −80 °C storage) and shake at room temperature at 200 rpm for 3–7 days. A dense suspension of conidia should be visible; conidia will settle at the bottom of the flask when removed from shaking.

   CMC medium

1.0 g	NH_4NO_3
1.0 g	KH_2PO_4
0.5 g	$MgSO_4 \cdot 7H_2O$
1.0 g	Yeast extract
15 g	Carboxymethylcellulose[a] (CMC; Sigma C 5678)

   [a]Add CMC last and slowly to avoid clumping. Use only low viscosity CMC. Adjust to 1 L with H_2O and autoclave

2. Inoculate 75 mL of YPD medium in a 250-mL Erlenmeyer flask with ~5 × 10^6 conidia per flask, and shake overnight (room temperature, 200 rpm) to generate a grape-sized, ~5 g mycelium for protoplasting. For highest efficiency it takes

~24 h of growth, and slow-growing mutants may require 1×10^7 spores as inoculum to generate sufficient tissue. We often use more than one flask, as it is not advisable to grow the tissue for less than 12 or more than 24 h because the digestion with the cell wall lysing enzymes will be inefficient.

YPD broth

3 g	Yeast extract
10 g	Peptone
20 g	Glucose (dextrose)

Adjust to 1 L with H_2O and autoclave

3. Prepare protoplasting buffer. If fewer than five transformations will be done, make 10 mL (instead of 20 mL) of protoplasting buffer and use less mycelium to save expensive enzymes.

   In a 50-mL conical centrifuge tube add 1 mg chitinase (Sigma, catalog #C6137) and 100 mg lysing enzyme (Sigma, #L1412) to 10 mL of sterile 1.2 M KCl. Mix by rocking for 5 min. In a 15-mL conical centrifuge add 500 mg driselase (Sigma, #D8037) to 10 mL of sterile 1.2 M KCl. Mix by rocking for 5 min; centrifuge (4 °C, 5 min, 2,885 x *g* in a typical tabletop or Sorvall centrifuge with swinging bucket) to remove solids. Add the driselase solution to the chitinase solution for a final volume of 20 mL of protoplasting buffer. Sterile filtering is not necessary.

4. Digest ~5 g (wet weight) of mycelium in 20 mL of protoplasting buffer (28 °C, 90 rpm, 2.5 h). If too much tissue is used the digestion will be inefficient and the yield of protoplasts is reduced.

5. After 2.5 h, remove a small aliquot from the digest to a microscope slide and check under the microscope for digestion, and protoplast formation. Protoplasts will "bud" from the tip of hyphae. There should be ~1×10^8 round protoplasts with very little intact mycelium in the solution. Protoplasts will be destroyed if water or detergent is added to a small sample for microscopy; this is useful for verifying number of protoplasts.

6. Autoclave 30 μM Nitex membrane (http://www.dynamicaqua.com/nitex.html) that is held on 50-mL plastic or glass beakers by rubber bands and foil. Directly before use, lightly wash membrane with sterile water, and remove all water to avoid lysing protoplasts. Gently filter protoplasts through sterile Nitex membrane into the beaker, gently transfer to conical tube, and centrifuge at 4 °C and 1,923 x *g* for 5 min.

7. Remove the protoplasting buffer by gently pouring and resuspend protoplasts in 10 mL STC solution by gently pipetting up and down with a 10 mL pipette. Use an aliquot (12 μL) for

counting by hemacytometer. Expected are ~1 × 10^8 protoplasts in total. While counting, centrifuge protoplasts at 4 °C, 5 min, 1,923 x *g*. Gently pour off the supernatant.

STC solution

218.6 g	Sorbitol (1.2 M)
1.2 g	Tris–HCl (10 mM)
5.5 g	$CaCl_2$ (50 mM)

Adjust to 1 L with H_2O, pH 8.0, autoclave

8. Repeat STC wash once, discard supernatant and resuspend protoplasts in 1 mL of STC buffer by gently pipetting up and down with a wide-bore 10 mL pipette. Transfer to an Eppendorf (1.5–2 mL) tube that has been precooled on ice.
9. Resuspend protoplasts in STC so that the final concentration is 2.5 × 10^7 per mL (or 1 × 10^7 per 400 µL) of STC for each transformation. Return tubes to ice.

5.1.1 Transformation of Protoplasts

1. Prepare regeneration medium (RM) and fresh 30% PEG 8000 solution on the same day as the transformation.

Regeneration Medium

Solution A	
1.0 g	Yeast extract
1.0 g	Casein enzyme hydrolysate
7.5 g	Agar

Adjust to 500 mL with H_2O

Solution B	
273.5 g	Sucrose in ~300 mL of H_2O

Adjust to 500 mL with H_2O

Autoclave solutions A and B separately and combine after autoclaving.

30% PEG Solution

Dissolve 30% (w/v) polyethylene glycol (MW 8000) in STC buffer. Autoclave to sterilize.

2. For each transformation, mix gently by pipetting in one 15 mL conical centrifuge tube:
   400 µL (or 1 × 10^7) protoplasts resuspended in STC buffer.
   100 µL freshly prepared 30% PEG 8000 solution.
   20–40 µL DNA (1–10 µg linearized plasmid or split marker PCR fragments).
3. Incubate tubes at room temperature for 20 min. During incubation, remove regeneration medium (RM) from water bath to bench to allow to cool to about 42 °C (warm to the touch,

bottle can be held indefinitely in ungloved hand, which is not true for 50 °C).

4. After incubation, add 1 mL of 30% PEG solution, mix by gently inverting tube, and incubate at room temperature for 5 min. Label six empty petri plates for each transformation mix.
5. Add 2 mL of STC buffer and mix by gently inverting tube.
6. For each transformation, mix half of the protoplast mixture (~1.75 mL) with ~43 mL of warm (42 °C) RM by inverting gently in a 50 mL conical tube. Pour 15 mL per petri plate for the first three plates. Mix the remaining ~1.75 mL of protoplast mixture with ~43 mL of RM and pour the remaining three plates. Work quickly to keep RM from solidifying.

   Under these conditions, the plating density is about 0.5×10^6 cells per plate, which should not lead to overcrowding yet still yields up to 50 colonies per plate. For supercoiled plasmid transformations, or other high efficiency transformations, one can plate 10^4–10^5 cells per plate to prevent crowding.
7. Incubate at room temperature overnight.
8. The next day, overlay medium in every plate with selective medium. For selection against growth on Hygromycin B or G418, overlay each plate ~24 h after transformation with 15 mL of RM + 200 μg/mL antibiotic (for a final diffused concentration of about 100 μg/mL of total medium) by pouring the RM/antibiotic mixture from a 50 mL conical tube on top of the original medium in each plate. For selection for prototrophy, 15 mL of VMM supplemented with 28% sucrose and 0.7% agar is used as overlay for each transformation plate.
9. After 3-7 days, transfer a small mycelial fragment to YPD agar supplemented with 100 μg/mL antibiotic in 48-well plates. Primary transformants should be considered heterokaryons until purified by isolating single conidia or (better) by genetic crosses. Individual stocks should be made for each transformant and stored in 25% glycerol at −80 °C. Original plates can be wrapped with Parafilm and stored at 4 °C for up to 2 weeks.

5.2 Protocol 2: Genetic Crosses to Allow Cre-Mediated Recombination

Several protocols for crosses with *F. graminearum* have been described [34, 51]. Our procedures are based on those developed in the Trail lab [43]. We induce the sexual cycle by growth on carrot agar.

5.2.1 Carrot Agar

Wash, dice and boil (until soft) 400 g of fresh organic carrots in 1 L of water. Blend boiled carrot mix in a large Waring blender for about 60 s, add 15 g of agar and blend again. Pour into large Erlenmeyer flasks and autoclave for 40 min. Pour this viscous mix into petri plates (~20 mL per plate), let cool and store at 4 °C until needed. We found that several nonorganic batches of carrots did not support growth or sexual development of *Fusarium* strains.

1. Most strains, including FMF180 (*pan2::loxP*; *cre*$^{+}$-*hph*$^{+}$), are grown on carrot agar for 7 days at room temperature under ambient light conditions. For optimal fertility it is important not to stack plates. For Cre-mediated excision of marker genes we used FMF180, which carries both mating type idiomorphs and is therefore self-fertile. Selection against *pan2* thus increased the likelihood of finding recombinant progeny.
2. When the mycelium reaches the edge of the petri plate or after no more than 7 days, aerial mycelium is removed completely by scraping gently with a sterile toothpick and addition of 800 μL of 2.5% (v/v) Tween 80.
3. For fertilization, ~1×10^5 conidia of the donor ("male") parent, usually bearing the floxed allele, is spread evenly on the surface of the recipient ("female") strain with a sterile glass rod.
4. Plates are incubated right side up, in a single-layer under ambient light and at room temperature. Perithecia will form within a week. Crosses with PH-1 and derivatives usually will shoot ascospores to the plate lid (i.e., there is no need to inspect perithecia for cirrhi).
5. Once ascospores start collecting on the plate lid (~7–9 days after fertilization), plates are turned over. Ascospores should be harvested ~2 weeks after fertilization. Long-term storage on the plate lid for longer than 4 weeks after fertilization is not advised as germination will be inhibited.
6. Ascospores are harvested from the plate lid by adding 500 μL sterile water, scraping and pipetting up and down with a sterile pipette tip, and transferred to an Eppendorf tube.
7. Ascospores from most crosses are viable in water at 4 °C for up to 3 months. Freezing at −20 or −80 °C, even in 50% glycerol, kills ascospores.
8. Ascospores are spread on minimal medium (VMM + 2% agar) to select against strains that are *pan2*$^{-}$. The goal should be fewer than 5000 ascospores (usually 300–600) per plate to ensure sufficient separation between germlings to be able to pick single germinated spores.
9. Germlings are picked after ~24 h of incubation at room temperature to 48-well plates with YPD or VMM.
10. After 2–3 days of growth at room temperature, small hyphal fragments are transferred by pointed toothpick to medium with antibiotics. Progeny sensitive to the appropriate antibiotics, indicating marker excision by Cre during the cross, and demonstrating other desired phenotypes are selected to isolate DNA. A combination of diagnostic PCR, Sanger sequencing, and Southern analyses is carried out to validate correct genotypes.

References

1. Sauer B (1987) Functional expression of the cre-lox site-specific recombination system in the yeast Saccharomyces cerevisiae. Mol Cell Biol 7:2087–2096
2. Sauer B, Henderson N (1988) Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. Proc Natl Acad Sci U S A 85:5166–5170
3. Ow DW (2002) Recombinase-directed plant transformation for the post-genomic era. Plant Mol Biol 48:183–200
4. Guldener U, Heck S, Fielder T, Beinhauer J, Hegemann JH (1996) A new efficient gene disruption cassette for repeated use in budding yeast. Nucleic Acids Res 24:2519–2524
5. Yu Y, Bradley A (2001) Engineering chromosomal rearrangements in mice. Nat Rev Genet 2:780–790
6. van der Weyden L, Adams DJ, Bradley A (2002) Tools for targeted manipulation of the mouse genome. Physiol Genomics 11:133–164
7. Branda CS, Dymecki SM (2004) Talking about a revolution: The impact of site-specific recombinases on genetic analyses in mice. Dev Cell 6:7–28
8. Sorrell DA, Kolb AF (2005) Targeted modification of mammalian genomes. Biotechnol Adv 23:431–469
9. Krappmann S, Bayram O, Braus GH (2005) Deletion and allelic exchange of the Aspergillus fumigatus veA locus via a novel recyclable marker module. Eukaryot Cell 4:1298–1307
10. Forment JV, Ramon D, MacCabe AP (2006) Consecutive gene deletions in Aspergillus nidulans: application of the Cre/loxP system. Curr Genet 50:217–224
11. Golic KG, Lindquist S (1989) The FLP recombinase of yeast catalyzes site-specific recombination in the Drosophila genome. Cell 59:499–509
12. Chen Y, Rice PA (2003) New insight into site-specific recombination from Flp recombinase-DNA structures. Annu Rev Biophys Biomol Struct 32:135–159
13. Kopke K, Hoff B, Kuck U (2010) Application of the Saccharomyces cerevisiae FLP/FRT recombination system in filamentous fungi for marker recycling and construction of knockout strains devoid of heterologous genes. Appl Environ Microbiol 76:4664–4674
14. Turan S, Zehe C, Kuehle J, Qiao J, Bode J (2013) Recombinase-mediated cassette exchange (RMCE) – a rapidly-expanding toolbox for targeted genomic modifications. Gene 515:1–27
15. Carnoy C, Roten CA (2009) The dif/Xer recombination systems in proteobacteria. PLoS One 4:e6531
16. Missirlis PI, Smailus DE, Holt RA (2006) A high-throughput screen identifying sequence and promiscuity characteristics of the loxP spacer region in Cre-mediated recombination. BMC Genomics 7:73
17. Guo F, Gopaul DN, Van Duyne GD (1999) Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse. Proc Natl Acad Sci U S A 96:7143–7148
18. Van Duyne GD (2001) A structural view of cre-loxp site-specific recombination. Annu Rev Biophys Biomol Struct 30:87–104
19. Livet J, Weissman TA, Kang H, Draft RW, Lu J, Bennis RA, Sanes JR, Lichtman JW (2007) Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. Nature 450:56–62
20. Hirrlinger J, Scheller A, Hirrlinger PG, Kellert B, Tang W, Wehr MC, Goebbels S, Reichenbach A, Sprengel R, Rossner MJ, Kirchhoff F (2009) Split-cre complementation indicates coincident activity of different genes in vivo. PLoS One 4:e4286
21. Albert H, Dale EC, Lee E, Ow DW (1995) Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. Plant J 7:649–659
22. Qin M, Bayley C, Stockton T, Ow DW (1994) Cre recombinase-mediated site-specific recombination between plant chromosomes. Proc Natl Acad Sci U S A 91:1706–1710
23. Watson AT, Garcia V, Bone N, Carr AM, Armstrong J (2008) Gene tagging and gene replacement using recombinase-mediated cassette exchange in Schizosaccharomyces pombe. Gene 407:63–74
24. Dennison PM, Ramsdale M, Manson CL, Brown AJ (2005) Gene disruption in Candida albicans using a synthetic, codon-optimised Cre-loxP system. Fungal Genet Biol 42:737–748
25. Fickers P, Le Dall MT, Gaillardin C, Thonart P, Nicaud JM (2003) New disruption cassettes for rapid gene disruption and marker rescue in the yeast Yarrowia lipolytica. J Microbiol Methods 55:727–737
26. Baker LG, Lodge JK (2012) Multiple gene deletion in Cryptococcus neoformans using the Cre-lox system. Methods Mol Biol 845:85–98
27. Patel RD, Lodge JK, Baker LG (2010) Going green in Cryptococcus neoformans: the recycling of a selectable drug marker. Fungal Genet Biol 47:191–198

28. Florea S, Andreeva K, Machado C, Mirabito PM, Schardl CL (2009) Elimination of marker genes from transformed filamentous fungi by unselected transient transfection with a Cre-expressing plasmid. Fungal Genet Biol 46:721–730
29. Honda S, Selker EU (2009) Tools for fungal proteomics: multifunctional neurospora vectors for gene replacement, protein expression and protein purification. Genetics 182:11–23
30. Steiger MG, Vitikainen M, Uskonen P, Brunner K, Adam G, Pakula T, Penttila M, Saloheimo M, Mach RL, Mach-Aigner AR (2011) Transformation system for Hypocrea jecorina (Trichoderma reesei) that favors homologous integration and employs reusable bidirectionally selectable markers. Appl Environ Microbiol 77:114–121
31. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
32. Trail F, Common R (2000) Perithecial development by *Gibberella zeae*: a light microscopy study. Mycologia 92:130–138
33. Cuomo CA, Guldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma LJ, Baker SE, Rep M, Adam G, Antoniw J, Baldwin T, Calvo S, Chang YL, Decaprio D, Gale LR, Gnerre S, Goswami RS, Hammond-Kosack K, Harris LJ, Hilburn K, Kennell JC, Kroken S, Magnuson JK, Mannhaupt G, Mauceli E, Mewes HW, Mitterbauer R, Muehlbauer G, Munsterkotter M, Nelson D, O'Donnell K, Ouellet T, Qi W, Quesneville H, Roncero MI, Seong KY, Tetko IV, Urban M, Waalwijk C, Ward TJ, Yao J, Birren BW, Kistler HC (2007) The Fusarium graminearum genome reveals a link between localized polymorphism and pathogen specialization. Science 317:1400–1402
34. Bowden RL, Leslie JF (1999) Sexual recombination in Gibberella zeae. Phytopathology 89:182–188
35. Lee J, Jurgenson JE, Leslie JF, Bowden RL (2008) Alignment of genetic and physical maps of Gibberella zeae. Appl Environ Microbiol 74:2349–2359
36. Pomraning KR, Smith KM, Freitag M (2009) Genome-wide high throughput analysis of DNA methylation in eukaryotes. Methods 47:142–150
37. Kistler HC, Benny U (1992) Autonomously replicating plasmids and chromosome rearrangement during transformation of Nectria haematococca. Gene 117:81–89
38. Powell WA, Kistler HC (1990) In vivo rearrangement of foreign DNA by Fusarium oxysporum produces linear self-replicating plasmids. J Bacteriol 172:3163–3171
39. Lorang JM, Tuori RP, Martinez JP, Sawyer TL, Redman RS, Rollins JA, Wolpert TJ, Johnson KB, Rodriguez RJ, Dickman MB, Ciuffetti LM (2001) Green fluorescent protein is lighting up fungal biology. Appl Environ Microbiol 67:1987–1994
40. Connolly LR, Smith KM, Freitag M (2013) The Fusarium graminearum histone H3 K27 methyltransferase KMT6 regulates development and expression of secondary metabolite gene clusters. PLoS Genet 9:e1003916
41. Riquelme M, Bredeweg EL, Callejas-Negrete O, Roberson RW, Ludwig S, Beltran-Aguilar A, Seiler S, Novick P, Freitag M (2014) The Neurospora crassa exocyst complex tethers Spitzenkorper vesicles to the apical plasma membrane during polarized growth. Mol Biol Cell 25:1312–1326
42. Davis RH (2000) Neurospora: contributions of a model organism. Oxford University Press, Oxford
43. Cavinder B, Sikhakolli U, Fellows KM, Trail F (2012) Sexual development and ascospore discharge in Fusarium graminearum. J Vis Exp 61:3895
44. Punt PJ, Oliver RP, Dingemanse MA, Pouwels PH, van den Hondel CAMJJ (1987) Transformation of *Aspergillus* based on the hygromycin B resistance marker from *Escherichia coli*. Gene 56:117–124
45. Webster TD, Dickson RC (1983) Direct selection of Saccharomyces cerevisiae resistant to the antibiotic G418 following transformation with a DNA vector carrying the kanamycin-resistance gene of Tn903. Gene 26:243–252
46. Yoruk E, Albayrak G (2015) Geneticin (G418) resistance and electroporation-mediated transformation of Fusarium graminearum and F. culmorum. Biotechnol Biotechnol Equip 29:268–273
47. Hoff B, Kamerewerd J, Sigl C, Zadra I, Kuck U (2010) Homologous recombination in the antibiotic producer Penicillium chrysogenum: strain DeltaPcku70 shows up-regulation of genes from the HOG pathway. Appl Microbiol Biotechnol 85:1081–1094
48. Anne J, Peberdy JF (1976) Induced fusion of fungal protoplasts following treatment with polyethylene glycol. J Gen Microbiol 92:413–417
49. Studt L, Rosler SM, Burkhardt I, Arndt B, Freitag M, Humpf HU, Dickschat JS, Tudzynski B (2016) Knock-down of the methyltransferase Kmt6 relieves H3K27me3 and results in induction of cryptic and otherwise silent secondary metabolite gene clusters in Fusarium fujikuroi. Environ Microbiol 18:4037–4054

50. Leslie JF, Summerell BA (2006) The Fusarium laboratory manual, 1st edn. Blackwell Pub., Ames, IA
51. Klittich C, Leslie JF (1988) Nitrate reduction mutants of Fusarium moniliforme (Gibberella fujikuroi). Genetics 118:417–423
52. Lee J, Lee T, Lee YW, Yun SH, Turgeon BG (2003) Shifting fungal reproductive mode by manipulation of mating type genes: obligatory heterothallism of Gibberella zeae. Mol Microbiol 50:145–152
53. Leslie JF (1983) Some genetic techniques for *Gibberella zeae*. Phytopathology 73:1005–1008
54. O'Donnell K, Rooney AP, Proctor RH, Brown DW, McCormick SP, Ward TJ, Frandsen RJ, Lysoe E, Rehner SA, Aoki T, Robert VA, Crous PW, Groenewald JZ, Kang S, Geiser DM (2013) Phylogenetic analyses of RPB1 and RPB2 support a middle Cretaceous origin for a clade comprising all agriculturally and medically important fusaria. Fungal Genet Biol 52:20–31
55. Lofgren LA, LeBlanc NR, Certano AK, Nachtigall J, LaBine KM, Riddle J, Broz K, Dong Y, Bethan B, Kafer CW, Kistler HC (2017) Fusarium graminearum: pathogen or endophyte of North American grasses? New Phytol 217:1203–1212
56. Ma LJ, Geiser DM, Proctor RH, Rooney AP, O'Donnell K, Trail F, Gardiner DM, Manners JM, Kazan K (2013) Fusarium pathogenomics. Annu Rev Microbiol 67:399–416
57. Brown DW, Proctor RH (2016) Insights into natural products biosynthesis from analysis of 490 polyketide synthases from Fusarium. Fungal Genet Biol 89:37–51
58. Trail F, Wang Z, Stefanko K, Cubba C, Townsend JP (2017) The ancestral levels of transcription and the evolution of sexual phenotypes in filamentous fungi. PLoS Genet 13:e1006867
59. Chen Y, Zhou MG (2009) Sexual recombination of carbendazim resistance in Fusarium graminearum under field conditions. Pest Manag Sci 65:398–403
60. Kim HK, Cho EJ, Lee S, Lee YS, Yun SH (2012) Functional analyses of individual mating-type transcripts at MAT loci in Fusarium graminearum and Fusarium asiaticum. FEMS Microbiol Lett 337:89–96
61. Zheng Q, Hou R, Juanyu Z, Ma J, Wu Z, Wang G, Wang C, Xu JR (2013) The MAT locus genes play different roles in sexual reproduction and pathogenesis in Fusarium graminearum. PLoS One 8:e66980
62. Beremand MN, Van Middlesworth F, Taylor S, Plattner RD, Weisleder D (1988) Leucine auxotrophy specifically alters the pattern of trichothecene production in a T-2 toxin-producing strain of Fusarium sporotrichioides. Appl Environ Microbiol 54:2759–2766
63. Namiki F, Matsunaga M, Okuda M, Inoue I, Nishi K, Fujita Y, Tsuge T (2001) Mutation of an arginine biosynthesis gene causes reduced pathogenicity in Fusarium oxysporum f. sp. melonis. Mol Plant Microbe Interact 14:580–584
64. Shimshek DR, Kim J, Hubner MR, Spergel DJ, Buchholz F, Casanova E, Stewart AF, Seeburg PH, Sprengel R (2002) Codon-improved Cre recombinase (iCre) expression in the mouse. Genesis 32:19–26
65. Case ME, Schweizer M, Kushner SR, Giles NH (1979) Efficient transformation of Neurospora crassa utilizing hybrid plasmid DNA. Proc Natl Acad Sci U S A 76:5259–5263
66. Hohn TM, Desjardins AE (1992) Isolation and gene disruption of the Tox5 gene encoding trichodiene synthase in Gibberella pulicaris. Mol Plant Microbe Interact 5:249–256
67. Osiewacz HD, Skaletz A, Esser K (1991) Integrative transformation of the ascomycete Podospora anserina: identification of the mating-type locus on chromosome VII of electrophoretically separated chromosomes. Appl Microbiol Biotechnol 35:38–45
68. Maier FJ, Malz S, Losch AP, Lacour T, Schafer W (2005) Development of a highly efficient gene targeting system for Fusarium graminearum using the disruption of a polyketide synthase gene as a visible marker. FEMS Yeast Res 5:653–662
69. Moradi S, Sanjarian F, Safaie N, Mousavi A, Khaniki GRB (2013) A modified method for transformation of Fusarium graminearum. J Crop Prot 2:297–304
70. Watson RJ, Burchat S, Bosley J (2008) A model for integration of DNA into the genome during transformation of Fusarium graminearum. Fungal Genet Biol 45:1348–1363
71. Ninomiya Y, Suzuki K, Ishii C, Inoue H (2004) Highly efficient gene replacements in Neurospora strains deficient for nonhomologous end-joining. Proc Natl Acad Sci U S A 101:12248–12253
72. Nayak T, Szewczyk E, Oakley CE, Osmani A, Ukil L, Murray SL, Hynes MJ, Osmani SA, Oakley BR (2006) A versatile and efficient gene-targeting system for Aspergillus nidulans. Genetics 172:1557–1566



Chapter 17

Protocol of *Phytophthora capsici* Transformation Using the CRISPR-Cas9 System

Zhiwen Wang, Brett M. Tyler, and Xili Liu

Abstract

Phytophthora capsici is an important plant pathogen, which causes significant economic losses on multiple vegetable crops worldwide. It is an ideal model pathogen to study the role of important genes, plant–pathogen interactions, and fungicide resistance mechanisms etc. due to its wide range of hosts and genetic diversity. A more efficient gene editing tool is required to do these studies. Here, we describe a detailed experimental procedure using the CRISPR-Cas9 system to edit genes of interest in *P. capsici*, which has been proven to be an accurate and efficient gene editing method in *P. capsici*.

Key words *Phytophthora capsici*, CRISPR-Cas9, Gene edit, Protoplast transformation

1 Introduction

Phytophthora capsici, which causes root, crown, foliar, and fruit rot on a large number of important vegetables, is a filamentous oomycete first described in 1922 [1]. It has a broad host range, which includes pepper, tomato, eggplant, cucurbits, and lima beans [2]. Due to its economic importance, many investigators have chosen this pathogen to research functional genomics, fungicide resistance mechanisms, and plant–pathogen interactions [3–5]. As described below, a reliable gene editing system is desired. CRISPR-Cas9 is a potentially efficient gene editing system for *P. capsici*.

For the *P. capsici* genome project initiated in 2005, an inbreeding isolate, LT1534, was selected for sequencing [6]. Sequencing yielded a draft genome sequence covering 64 Mb of the predicted 65 Mb *P. capsici* genome. The high-quality genome database provides a very useful platform to enable biological studies of *P. capsici*. Precise gene prediction and structural information could benefit from accurate gene editing.

CRISPR-Cas9 (RNA-guided Cas9 nuclease from microbial clustered regularly interspaced short palindromic repeats) is a powerful tool to facilitate efficient genome engineering in eukaryotic cells by

Wenbo Ma and Thomas Wolpert (eds.), *Plant Pathogenic Fungi and Oomycetes: Methods and Protocols*, Methods in Molecular Biology, vol. 1848, https://doi.org/10.1007/978-1-4939-8724-5_17,
© Springer Science+Business Media, LLC, part of Springer Nature 2018

simply specifying a 20-nt targeting sequence within a guide RNA [7]. It has been regarded as the most efficient genome editing tool when compared to two other genome editing technologies which emerged in recent years, zinc-finger nucleases (ZFNs) [8–11] and transcription activator-like effector nucleases (TALENs) [11–18]. Cas9 is a nuclease guided by a small RNA through Watson–Crick base pairing with the target DNA [19–21]. The type II CRISPR-Cas9 system is derived from the adaptive immune system of *Streptococcus pyogenes*. It has DNA recognition and cleavage functions mediated by a single guide RNA (sgRNA) rather than a DNA recognition protein domain as in ZFNs and TALENs. Thus, it is much easier to design, and displays higher specificity and more efficiency than ZFNs and TALENs. Moreover, it is well suited for multiplexed gene editing and easy conducted in a variety of organisms [22].

Unlike the gene silencing triggered by hairpin, antisense, and sense RNA constructs, which has been the dominant system to study Phytophthora genetics in past years [23], the CRISPR-Cas9 system can completely knockout or precisely replace the targeted gene(s) rather than yielding only partial knockdowns. This provides higher confidence data. Consequently, the CRISPR-Cas9 system should be a powerful tool to facilitate the study of genes in *P. capsici*.

Here, we describe a detailed protocol for *P. capsici* transformation using CRISPR-Cas9 system. The protocol developed for *P. sojae* was followed [24], but slight modifications were made. Our results demonstrate that CRISPR-Cas9 mediated gene disruption and gene replacement is an efficient tool to explore not only specific gene function in *P. capsici* but also fungicide resistance mechanisms and sRNA function.

2 Materials

For preparing buffers and media, deionized water is used.

2.1 Plasmids

All plasmids were kindly provided by Brett M. Tyler. Some were slightly modified:

pYF2-PsNLS-hSpCas9: for Cas9 nuclease expression; pYF2.2-GFP-Ribo-gRNA: for sgRNA expression; and pBluscript II SK+::NPTII: to provide the templet for homologous recombination with the gene of interest.

2.2 P. capsici Growth Media

1. Pea broth (prepared for nutrient pea medium and regeneration medium): 125 g frozen peas (preferably unsalted) in 700 mL of water, Autoclave for 30 min, collect supernatant after straining through four layers of cheesecloth, bring up the volume to 1 L with ddH_2O.
2. Nutrient Pea Broth (NPB): 1.0 g K_2HPO_4 (final 0.1%, w/v), 3.0 g KNO_3 (final 0.3%, w/v), 0.5 g $MgSO_4$ (final 0.05%, w/v),

0.1 g $CaCl_2$ (final 0.01%, w/v), 2.0 g $CaCO_3$ (final 0.2%, w/v), 5.0 g D-sorbitol (final 0.5%, w/v), 5.0 g D-mannitol (final 0.5%, w/v), 5.0 g D-glucose (final 0.5%, w/v), 2.0 g yeast extract (final 0.2%, w/v), add Pea broth to 1 L, autoclave at 121 °C for 20 min. After the medium cools down, add the following components under sterile conditions: 2 mL vitamin stock (*see* protocol below), 2 mL trace elements (*see* protocol below).

3. Nutrient Pea Broth Agar (NPBA): 1.0 g K_2HPO_4 (final 0.1%, w/v), 3.0 g KNO_3 (final 0.3%, w/v), 0.5 g $MgSO_4$ (final 0.05%, w/v), 0.1 g $CaCl_2$ (final 0.01%, w/v), 2.0 g $CaCO_3$ (final 0.2%, w/v), 5.0 g D-sorbitol (final 0.5%, w/v), 5.0 g D-mannitol (final 0.5%, w/v), 5.0 g D-glucose (final 0.5%, w/v), 2.0 g yeast extract (final 0.2%, w/v), 15.0 g agar (final 1.5%, w/v), add Pea broth to 1 L, autoclave at 121 °C for 20 min. After the medium cools down, add the following components under sterile conditions: 2 mL vitamin stock (*see* protocol below), 2 mL trace elements (*see* protocol below).
4. Regeneration Medium (PM): 91.1 g D-mannitol, 1 g $CaCl_2$, 2 g $CaCO_3$, add Pea broth to 1 L. Autoclave at 121 °C for 20 min.
5. Regeneration Agar Medium (PAM): 91.1 g D-mannitol, 1 g $CaCl_2$, 2 g $CaCO_3$, 15.0 Bacto agar (final 1.5%, w/v), add Pea broth to 1 L. Autoclave at 121 °C for 20 min.
6. V8 medium: 340 mL commercial V8 vegetable juice (there should be about 250 mL supernatant left after centrifuged at 3,000 × *g* for 2 min), 3.5 g $CaCO_3$, 35 g Agar, add water to 2.5 L. Autoclave at 121 °C for 20 min.
7. Vitamin stock (prepared for NPB and NPBA): 10 μL of 0.02 g/mL Biotin (final 6.7×10^{-7} g/mL), 10 μL of 0.02 g/mL Folic Acid (final 6.7×10^{-7} g/mL), 0.012 g of L-inositol (final 4×10^{-5} g/mL), 0.06 g nicotinic acid (final 2×10^{-4} g/mL), 0.18 g pyridoxine–HCl (final 6×10^{-4} g/mL), 0.015 g Riboflavin (final 5×10^{-5} g/mL), 0.38 g thiamine–HCl (final 1.3×10^{-3} g/mL), add H_2O to 300 mL. Sterilize using a 0.45 μm filter, then store at 4 °C.
8. Trace elements (prepared for NPB and NPBA): 0.215 g $FeC_6H_5O_7 \cdot 3H_2O$ (final 5.4×10^{-4} g/mL), 0.15 g $ZnSO_4 \cdot 7H_2O$ (final 3.8×10^{-4} g/mL), 0.03 g $CuSO_4 \cdot 5H_2O$ (final 7.5×10^{-4} g/mL), 0.015 g $MgSO_4 \cdot H_2O$ (final 3.8×10^{-5} g/mL), 0.01 g H_3BO_3 (final 2.5×10^{-5} g/mL), 0.012 g $Na_2MoO_4 \cdot H_2O$ (final 3.0×10^{-5} g/mL), add H_2O to 400 mL. Sterilize using a 0.45 μm filter and then store at 4 °C.

2.3 P. capsici Transformation Buffers

1. 0.8 M mannitol: 145.76 g D-mannitol, add water to 1 L. Autoclave at 121 °C for 20 min. Store at 4–25 °C.
2. W5 solution: 0.093 g KCl, 4.6 g $CaCl_2 \cdot 2H_2O$, 2.25 g NaCl, 7.8 g glucose, add H_2O to 250 mL. Autoclave at 121 °C for 20 min. Store at 4–25 °C.

3. 0.5 M MES (pH 5.7): 21.325 g 4-morpholinoethanesulfonic acid (MES), dissolve in 100 mL H_2O, adjust pH to 5.7 using 1 M KOH, add H_2O to 200 mL, sterilize using a 0.45 μm filter. Store at 4 °C.
4. MMG solution: 18.22 g mannitol, 0.76 g $MgCl_2 \cdot 6H_2O$, 2 mL of 0.5 M MES (pH 5.7), add H_2O to 250 mL. Autoclave at 121 °C for 20 min. Store at 4–25 °C.
5. 0.5 M $CaCl_2$: 9.76 g $CaCl_2 \cdot 2H_2O$, add H_2O to 100 mL. Autoclave at 121 °C for 20 min. Store at 4–25 °C.
6. 0.5 M KCl: 7.5 g KCl, add H_2O to 200 mL. Autoclave at 121 °C for 20 min. Store at 4–25 °C.
7. Enzyme solution (20 mL needed for three flasks of *P. capsici* culture): 10 mL of 0.8 M mannitol, 8 mL H_2O, 0.8 mL of 0.5 M KCl, 0.8 mL of 0.5 M MES (pH 5.7), 0.4 mL of 0.5 M $CaCl_2$, 0.15 g lysing enzymes from *Trichoderma harzianum*, 0.06 g Cellulase from *Trichoderma viride*, total 20 mL. Prepare fresh each time.
8. PEG-calcium transformation solution (prepare 2 mL for each protoplast culture tube): 6 g PEG 4000, 3 mL H_2O, 3.75 mL of 0.8 M mannitol, 3 mL of 0.5 M $CaCl_2$. Mix well and sterilize using a 0.45 μm filter. Prepare fresh each time.

2.4 Antibiotics

1. 50 mg/mL G418: Dissolve 50 mg G418 in 1 mL H_2O, sterilize using a 0.45 μm filter and store at −20 °C.
2. 100 mg/mL Amp: Dissolve 100 mg Ampicillin in 1 mL H_2O, sterilize using a 0.45 μm filter and store at −20 °C.

3 Methods

3.1 sgRNA Design

We recommend using at least three different sgRNAs for each gene of interest.

For sgRNA target selection, identify the 20-bp sequence upstream of any 5'-NGG sequence. A web-tool (http://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design) created by Doench [25] is a powerful tool to design sgRNA. Choose the sgRNAs with the highest scores and desired positions. The optimal sgRNAs should avoid self-complementarity to prevent hybridization with the target DNAs. Analyze the secondary structure by using RNA STRUCTURE (http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html). The off-target effects should be checked though the FungiDB (www.fungidb.org) alignment search tool against the *P. capcisi* genome. Sequences that perfectly match the final 12 nt of the target sequence and NGG PAM (Protospacer Adjacent Motif) sequence should be discarded [22].

The scheme of constructing the sgRNA harboring plasmid is shown in Fig. 1 (all figures of plasmids from Fang and Tyler 2016 [24]), the annealed oligos generated by the designed sgRNA can

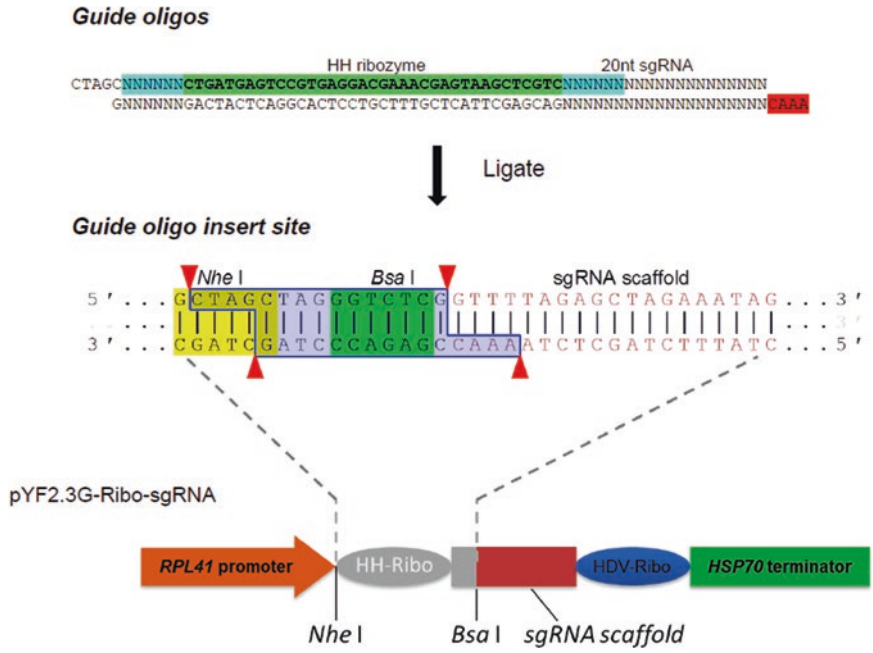


Fig. 1 Schematic for cloning of the guide sequence oligos into a plasmid containing sgRNA scaffold flanked by HDR-ribozyme

be ligated into the plasmid using *Nhe* I and *Bsa*I cleavage and ligation. The first six nucleotides of the Hammerhead (HH) ribozyme must be the reverse complement of the first six nucleotides of the sgRNA target sequence.

3.1.1 Oligo Annealing

Prepare system as below:

3 μL Sense oligo (100 μM in TE).

3 μL Antisense oligo (100 μM in TE).

3 μL 10 × T4 DNA ligase buffer.

21 μL water.

Incubate at 37 °C for 30 min, then add 4 μL 0.5 M NaCl, boil for 2 min, dilute 1 μL into 499 μL water after it has cooled down.

3.1.2 Ligation and Transformation

Prepare the system as below:

4 μL diluted annealed oligos.

50 ng pYF2.3G-Ribo-sgRNA (cut by *Nhe*I and *Bsa*I).

2 μL 10 × Ligase buffer.

1 μL T4 DNA Ligase (400,000 units/mL, NEB).

Add water to 20 μL.

Incubate at room temperature for 30 min, then transform 10 μL into 100 μL competent DH5α, plate on LB amended with 100 μg/mL Amp.

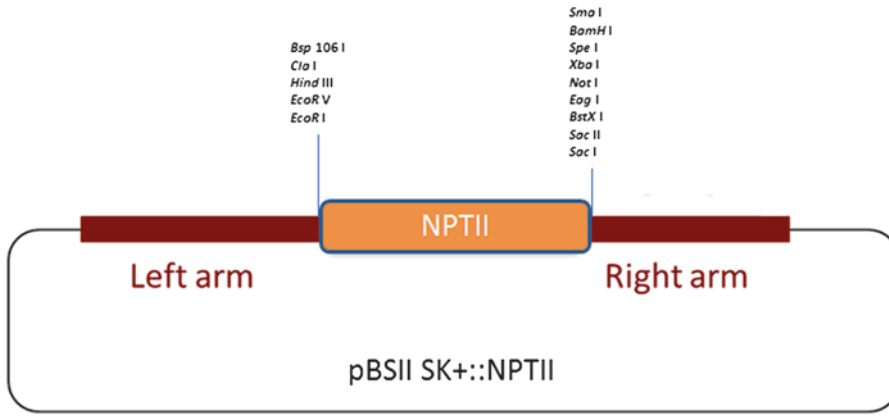


Fig. 2 A sketch map of HR vector pBSII SK+::NPTII

3.1.3 Colony PCR

Using M13F and the sgRNA forward primer for colony PCR.

3.2 Homologue Recombination (HR) Vector Construction

Homology directed repair mediated mutations are more accurate and efficient than random repair, thus the HR vector should be used when conducting gene editing in *P. capsici*. To accomplish this, a homologous donor must be provided. The circular plasmid pBS SK+::NPTII (Fig. 2) can be used as a carrier of the HR template in *P. capsici*. Two 1 kb homologous flanking arms located near the gene of interest (or sequence of interest) are enough to generate HR mutants. NPTII together with 1 kb of homologous 5′ and 3′ flanking arms outside the region of interest can be amplified and cloned into the plasmid as common approach.

3.3 CRISPR-Cas9 Plasmid

The hSpCas9 expression plasmid pYF2-PsNLS-hSpCas9 has been previously constructed (Fig. 3, [24]) and can be used for our method (*see* **Note 1**).

3.4 P. capsici Transformation

Polyethylene Glycol (PEG) mediated protoplast transformation was conducted.

3.4.1 P. capsici Growth

1. Culture *P. capsici* LT1534 on V8 solid plate for 3–4 days.
2. Place agar plugs (5 mm) from the periphery of actively growing cultures onto a NPBA plate with a cellophane membrane (*see* **Note 2**).
3. After 4 d, cut the mycelia grown on the cellophane membrane into small pieces (3 mm × 3 mm) using a sterile scalpel, inoculate three 250 mL flasks containing 50 mL NPB with about 20 pieces of *P. capsici* mycelia. Incubate at 25 °C for 1.5–2 days (*see* **Note 3**).

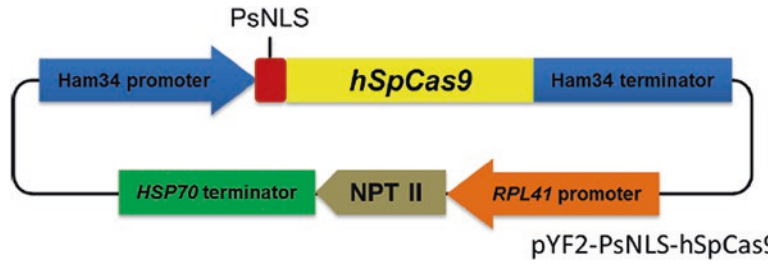


Fig. 3 A diagram of Cas9 expression plasmid

3.4.2 Protoplast Preparation

1. Harvest the 1.5-day-old *P. capsici* mycelial mats from the three flasks using two layers of cheesecloth, rinse twice with 0.8 M mannitol (*see* **Note 4**).
2. After the second wash, harvest all the mycelial mats into a 50 mL Falcon tube, wash once again on a shaker (60 rpm) with 35 mL 0.8 M mannitol for 10 min.
3. Prepare the enzyme solution (20 mL for mycelia from three flasks), sterilized using a 0.45-μm filter into a new 50 mL Falcon tube.
4. Harvest mycelial mat again using two layers of cheesecloth and transfer washed mycelia into the prepared enzyme solution, make sure you have dispersed the mycelia sufficiently by shaking the tube vigorously before digesting.
5. Digest for about 40 min at 25 °C on shaker (60 rpm).
6. Filter the digestion products through four layers of Miracloth to remove mycelial debris, collecting the flow-through into a new 50 mL Falcon tube.
7. Centrifuge the flow-through at 530 × *g* at 4 °C for 4 min.
8. Resuspend the pellet in 30 mL W5, and then spin at 530 × *g* for 3 min.
9. Pour off the supernatant and resuspend the protoplasts in 10 mL W5. Rest the protoplast by keeping them on ice for 30 min.
10. Spin at 530 × *g* for 3 min, remove the W5 completely. Resuspend the protoplast with 3 mL MMG; the concentration of protoplast should be around 10^6 per mililiter. Keep at room temperature for 10 min.

3.4.3 PEG–$CaCl_2$-Mediated Transformation

1. Mix all 3 plasmids well in 3 new Falcon 50 mL tubes (about 20 μg of each plasmid), respectively. The total volume should not exceed 100 μL (*see* **Note 5**).
2. Add 1 mL of protoplast contained MMG to each of the plasmid-containing tube, mix gently and keep on ice for 10 min.

3. Add 1.74 mL freshly prepared PEG-calcium transformation solution to each tube in three aliquots of 580 μL each. Add the PEG gently against the side of the tube wall into the protoplast solution, gently mixed and incubate on ice for 20 min.
4. Add 2 mL of cold PM to the tube, gently mixed and incubate on ice for 2 min.
5. Add 8 mL of cold PM to the tube, gently mixed and incubate on ice for 2 min.
6. Add 10 mL of cold PM to the tube, gently mixed and add Ampicillin to 100 μg/mL (final concentration).

3.4.4 Protoplast Regeneration and Mutant Screening

1. Incubate the tube at 25 °C for 14–18 h in the dark.
2. Collect the regenerated protoplasts by centrifugation at 700 × *g* for 5 min in 4 °C.
3. Pour off most of the supernatant and resuspend the regenerated protoplasts in the remaining 5 mL of PM medium.
4. Add warm PAM (about 50 °C) containing 30 μg/mL G418 and 100 μg/mL Amp to 45 mL, mix well.
5. Pour mix into three 10 cm petri dishes, after the medium solidifies, incubate plates at 25 °C for 2–3 days in the dark.
6. Cover emerged colonies with a layer (15–20 mL per petri dish) of V8 solid medium containing 50 μg/mL G418 and 100 μg/mL Amp, incubate at 25 °C for 3–4 days in the dark.
7. Transfer the visible transformants that have emerged from the surface of the V8 medium to a new V8 plate containing 50 μg/mL G418 and 100 μg/mL Amp, incubate at 25 °C for at least 3 days (according to the edited gene's function in the growth of *P.capsici*, for instance, if the edited gene is important for hyphal growth, you need wait more time to see the colonies grow out of the medium) in the dark.
8. Pick up single colonies and transfer to a new V8 plate containing 50 μg/mL G418, 100 μg/mL Amp and a cellophane membrane, collect hyphae for genotyping analysis after the colonies have grown out (*see* **Note 6**).

4 Notes

1. We have developed a new selection marker in pYF2.2-GFP-Ribo-gRNA besides NPTII in pYF2-PsNLS-hSpCas9 which is named ORP1, transformation of this gene (with a point mutation) into *P. capsici* will induce resistance to the fungicide Oxathiapiprolin. It is useful for carrying out double knock out and complementary gene editing with this double selection marker system. If you need any help, please contact us.

2. For subculturing *P. capsici* from mycelial plugs, cut agar slices from the outermost region of each petri dish, which contains newly grown mycelia.
3. To obtain enough potential mutants, culture 12 flasks of *P. capsici* each time to collect enough hypha for the next steps.
4. The hypha used for digesting should be very fresh (for LT1534, do not culture for more than 36 h, it may vary between different isolates that have different mycelial growth rates), the old hypha can cause incomplete digestion which can lead to insufficient protoplast.
5. The plasmids for transformation should also be very fresh, plasmids should be prepared ideally in the day before transformation.
6. You can pick single zoospores of potential mutants to obtain isolates with a pure genetic background.

References

1. Erwin DC, Ribeiro OK (eds) (1996) Phytophthora diseases worldwide. The American Phytopathological Society, St. Paul
2. Davidson CR, Carroll RB, Evans TA et al (2002) First report of *Phytophthora capsici* infecting lima bean (*Phaseolus lunatus*) in the mid-Atlantic region. Plant Dis 86:1049
3. Chidambareswaren M, Anu K, Gayathri GS et al (2016) Transcriptome-assisted label-free quantitative proteomics analysis reveals novel insights into Piper nigrum—*Phytophthora capsici* Phytopathosystem. Front Plant Sci 7:785
4. Miao JQ, Cai M, Dong X et al (2016) Resistance assessment for Oxathiapiprolin in *Phytophthora capsici* and the detection of a point mutation (G769W) in PcORP1 that confers resistance. Front Microbiol 7:615
5. Qiao YL, Shi JX, Zhai Y et al (2015) Phytophthora effector targets a novel component of small RNA pathway in plants to promote infection. Proc Natl Acad Sci U S A 112(18):5850–5855
6. Hurtado-Gonzales OP, Lamour KH (2009) Evidence for inbreeding and apomixis in close crosses of *Phytophthora capsici*. Plant Pathol 58:715–722
7. Ran FA Hsu PD, Wright J et al (2013) Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8(11):2281
8. Porteus MH, Baltimore D (2003) Chimeric nucleases stimulate gene targeting in human cells. Science 300:763
9. Miller JC, Holmes MC, Wang JB et al (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol 25:778–785
10. Sander JD, Dahlborg EJ, Goodwin MJ et al (2011) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat Methods 8:67–69
11. Wood AJ, Lo TW, Zeitler B et al (2011) Targeted genome editing across species using ZFNs and TALENs. Science 333:307
12. Christian M, Cermak T, Doyle EL et al (2010) Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186: 757–761
13. Zhang F, Cong L, Lodato S et al (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol 29:149–153
14. Hockemeyer D, Wang H, Kiani S et al (2011) Genetic engineering of human pluripotent cells using TALE nucleases. Nat Biotechnol 29:731–734
15. Reyon D, Tsai SQ, Khayter C et al (2012) FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol 30:460–465
16. Boch J, Scholze H, Schornack S et al (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326:1509–1512
17. Moscou MJ, Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. Science 326:1501
18. Sanjana NE, Cong L, Zhou Y et al (2012) A transcription activator-like effector toolbox for genome engineering. Nat Protoc 7:171–192
19. Garneau JE, Dupuis ME, Villion M et al (2010) The CRISPR/Cas bacterial immune

system cleaves bacteriophage and plasmid DNA. Nature 468:67–71

20. Jinek M, Chylinski K, Fonfara I et al (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821
21. Gasiunas G, Barrangou R, Horvath P et al (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A 109:E2579–E2586
22. Cong L, Ran FA, Cox D et al (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823
23. Ah-Fong AM, Bormann-Chung CA, Judelson HS (2008) Optimization of transgene-mediated silencing in *Phytophthora infestans* and its association with small-interfering RNAs. Fungal Genet Biol 45:1197–1205
24. Fang Y, Tyler BM (2016) Efficient disruption and replacement of an effector gene in the oomycete *Phytophthora sojae* using CRISPR/Cas9. Mol Plant Pathol 17:127–139. https://doi.org/10.1111/mpp.12318
25. Doench JG, Hartenian E, Graham DB et al (2014) Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat Biotechnol 32:1262–1267



Chapter 18

Generating Gene Silenced Mutants in *Phytophthora sojae*

Yao Zhao, Yufei Li, Min Qiu, Wenbo Ma, and Yuanchao Wang

Abstract

Phytophthora species are notorious pathogens of plants and cause enormous damage to agriculture. In order to understand *Phytophthora* pathogenesis, gene silencing and knockout methods are important for the investigation of gene functions. Although CRISPR/Cas9-based gene knockout procedures have been developed in *Phytophthora sojae* and *Phytophthora capsici*, it may not always be the best choice especially when knockout mutation leads to lethality. Therefore, gene silencing is a very useful tool for functional analysis of target genes in *Phytophthora*. This chapter introduces a gene silencing protocol for the soybean pathogen *P. sojae*. An expression cassette is incorporated into the *P. sojae* genome through PEG-mediated protoplast transformation, which leads to constitutive production of antisense RNA transcripts. These transcripts are able to target mRNAs through sequence complementarity and effectively reduce the expression of the target genes.

Key words Gene silencing, Antisense RNA, PEG-mediated transformation

1 Introduction

Phytophthora sojae is a destructive pathogen of soybean, causing a global economic loss of 1–2 billion dollars every year [1]. Different from most fungi, *Phytophthora* species have diploid chromosomes in the vegetative growth stage, making it difficult to conduct genetic manipulations. To understand the mechanism underlying pathogenesis, different approaches have been developed to modify gene expression in *P. sojae* [2–4]. Recently, a gene knockout system based on CRISPR/Cas9 has been established in *P. sojae* and another species *Phytophthora capsici*. This approach has been shown to efficiently generate gene knockout mutants [4]. However, the CRISPR system has not been successfully applied to other *Phytophthora* species such as the notorious *Phytophthora infestans*, which causes potato late blight. Furthermore, in some cases, gene silencing would be a better approach because it supports the investigation of the function of essential genes that are required for cell survival and for which a knockout mutation would lead to lethality [5].

Wenbo Ma and Thomas Wolpert (eds.), *Plant Pathogenic Fungi and Oomycetes: Methods and Protocols*, Methods in Molecular Biology, vol. 1848, https://doi.org/10.1007/978-1-4939-8724-5_18,
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Gene silencing is a very dependable tool for functional characterization of target genes in eukaryotic organisms. Suppression of gene expression can be achieved either by repression of transcription referred to as transcriptional gene silencing (TGS) or by mRNA degradation at the posttranscriptional level known as post-transcriptional gene silencing (PTGS) [6]. In studies of *Phytophthora* gene function, PTGS is commonly applied [7]. PTGS can work either by degradation of the transcripts or translational inhibition carried out by small RNAs.

In this method, we employ a plasmid called pTOR (GenBank: EU257520.1), which is derived from first-generation vectors for *Phytophthora* transformation [8]. This plasmid contains an *nptII* gene cassette that allows geneticin resistance as a selection marker. The *nptII* gene is controlled by the promoter and terminator from the *hsp70* gene of the oomycete pathogen *Bremia lactucae*. Another expression cassette contains the promoter and terminator of the *Ham34* gene, also from *B. lactucae* [9]. This cassette allows the constitutive expression of an antisense RNA fragment. Through PEG-mediated transformation, the vector can be transferred into *Phytophthora* cells. Production of antisense RNA molecules can lead to PTGS of a target gene(s) sharing sequence complementarity with the antisense RNA. During the gene silencing process, antisense RNA pairs with target mRNA and forms double-stranded RNA, will be futher cleaved into small fragments by DICER [10]. A single strand of the small RNA products is then loaded into the RNA-induced silencing complex (RISC), which silences target gene expression by transcript cleavage and/or translational repression [11]. Previous work shows that the expression of sense, antisense, or short hairpin RNA in *Phytophthora* all have a silencing effect on target genes. However, plasmid that directs expression of antisense RNA showed a higher efficiency of target gene silencing than sense and hairpin RNA [2, 12]. The mechanism(s) underlying the RNA-mediated gene silencing process in *Phytophthora* is not well understood. Nonetheless, silencing has been shown to work in several cases [2, 7, 12–16]. Here we describe a detailed protocol for gene silencing in *P. sojae* through PEG-mediated transformation.

2 Materials

2.1 Plasmid Preparation

1. Luria–Bertani medium (LB, 1 L): 10 g Bacto tryptone, 5 g yeast extract, 10 g NaCl. Adjust the volume to 1 L with ddH_2O. Sterilize by autoclaving.
2. Restriction enzymes.
3. T4 Ligase kit.
4. 1 L Erlenmeyer flasks.

5. Ampicillin (100 mg/mL).
6. 37 °C shaker.
7. NucleoBond® Xtra Maxi kits.

2.2 Propagation of P. sojae

1. Unclarified 10% V8 agar (1 L): 100 mL of V8 juice, 900 mL H_2O, 1 g $CaCO_3$, 0.05 g β-sitosterol (optional), 1.5% agar (w/v, for solid medium) (*see* **Note 1**).
2. Sterilize by autoclaving, Store up to 3 months at room temperature.
3. Pea broth (1 L): Autoclave 120 g frozen peas (preferably unsalted) in 700 mL of water at 121 °C for 20 min. Collect supernatant after straining through four layers of cheesecloth, squeeze the cheesecloth gently to collect the residual broth. Adjust the volume to 900 mL with ddH_2O (*see* **Note 2**).
4. Nutrient pea broth and agar medium (1 L): 1.0 g KH_2PO_4, 1.0 g K_2HPO_4, 3.0 g KNO_3, 0.5 g $MgSO_4$, 0.1 g $CaCl_2{\cdot}2H_2O$, 2.0 g $CaCO_3$, 5.0 g D-sorbitol, 5.0 g D-mannitol, 5.0 g D-glucose, 2.0 g yeast extract, add 900 mL pea broth (*see* recipe above), adjust to 1 L with ddH_2O. Mix for 30 min using a magnetic stirrer to dissolve ingredients (not all of the $CaCO_3$ will dissolve). Add 1% agar (optional for solid medium). Autoclave at 121 °C for 20 min. Add the following components under sterile conditions before use: 2 mL of vitamin stock (*see* recipe below), 2 mL trace elements (*see* recipe below). Store at 4 °C.
5. Vitamin stock (300 mL): 10 µL of 0.02 g/mL biotin, 10 µL of 0.02 g/mL folic acid, 0.012 g of L-inositol, 0.06 g nicotinic acid, 0.18 g pyridoxine–HCl, 0.015 g riboflavin, 0.38 g thiamine–HCl, H_2O to 300 mL. Filter-sterilize and then store (up to 1 year) at 4 °C.
6. Trace elements (400 mL): Dissolve 0.215 g $FeC_6H_5O_7{\cdot}3H_2O$, 0.15 g $ZnSO_4{\cdot}7H_2O$, 0.03 g $CuSO_4{\cdot}5H_2O$, 0.015 g $MgSO_4{\cdot}H_2O$, 0.01 g H_3BO_3, 0.012 g $Na_2MoO_4{\cdot}H_2O$, in ddH_2O to a final volume of 400 mL. Filter-sterilize and then store up to 1 year at 4 °C.
7. *P. sojae* strain P6497.
8. 250 mL Erlenmeyer flasks.
9. Cork borer (4 mm).
10. 25 °C incubator.

2.3 Protoplast Isolation

1. 0.8 M mannitol (autoclave sterilization, 121 °C for 20 min).
2. 0.5 M KCl (autoclave sterilization, 121 °C and 20 min).
3. 0.5 M MES (adjust pH to 5.7 using 1 M KOH; filter sterilization).

4. 0.5 M $CaCl_2$ (autoclave sterilization, 121 °C for 20 min).
5. Enzyme solution (20 mL): 10 mL of 0.8 M mannitol, 0.8 mL of 0.5 M KCl, 0.8 mL of 0.5 M MES, 0.4 mL of 0.5 M $CaCl_2$,0.15 g lysing enzymes from *Trichoderma harzianum* (Sigma, cat. no. L1412), 0.06 g CELLULYSIN cellulase, *Trichoderma viride* (Calbiochem, cat. no. 219466), 8 mL of ddH_2O (*see* **Note 3**).
6. W5 solution (250 mL): dissolve 0.093 g KCl, 4.6 g $CaCl_2{\cdot}2H_2O$, 2.25 g NaCl, and 7.8 g glucose in ddH_2O to a final volume of 250 mL. Store at 4 °C.
7. MMg solution (250 mL): dissolve 18.22 g mannitol, 0.76 g $MgCl_2{\cdot}6H_2O$, 2 mL of 0.5 M MES (pH = 5.7) in H_2O to a final volume of 250 mL. Store up to 6 months at 4 °C.
8. Miracloth (Millipore, Cat. 475855).
9. Long tweezers.
10. Cheesecloth.
11. 250 mL beaker.
12. 50 mL Falcon tube.
13. Parafilm.
14. Microscope.
15. Hemocytometer slide.
16. Tube tumbler rotating mixer.

2.4 PEG-Mediated Transformation and Regeneration

1. 20–25 mg plasmid DNA (*see* **Note 4**).
2. 40% PEG solution (w/v, 30 mL, enough for 15 samples): 12 g PEG 4000, 6 mL of ddH_2O, 7.5 mL of 0.8 M Mannitol, 6 mL of 0.5 M $CaCl_2$, Mix and filter sterilization. Store up to 2 weeks at 4 °C (*see* **Note 5**).
3. 0.45 μm cellulose acetate syringe filters.
4. Mannitol–pea broth (1 L): 91.1 g mannitol, 1.32 g $CaCl_2{\cdot}2H_2O$, 2 g $CaCO_3$, add 900 mL pea broth (*see* recipe in Subheading 2.2), add 1% bacto agar, adjust to 1 L with ddH_2O.
5. Ampicillin (50 mg/mL).
6. Parafilm.
7. 50 mL tube rack.

2.5 RNA Extraction, Reverse Transcription, and Real-Time PCR

1. TRIzol RNA Isolation Reagents.
2. Mini-Pestles for 1.5 mL tubes (autoclave-sterilized).
3. 1.5 mL tubes (RNase free).
4. Liquid nitrogen.
5. Chloroform.
6. Isopropyl alcohol.

7. 70% ethanol (v/v).
8. DEPC-treated ddH_2O.
9. SuperScript™ IV VILO™ Master Mix with ezDNase™ Enzyme.
10. Vortex mixer.
11. Refrigerated microcentrifuge.
12. SYBR Green supermix.
13. CFX Connect™ Real-Time PCR Detection System.
14. Beckman Allegra X-12 Centrifuge Benchtop w/SX4750 Swinging Bucket Rotor.

3 Methods

3.1 Constructing Plasmid Vectors for Gene Silencing.

1. The plasmid pTOR is maintained and isolated from *Escherichia coli* grown in Luria–Bertani (LB) media supplemented with ampicillin (final concentration 50 μg/mL).
2. Select an antisense sequence of about 200–800 bp in length from your target gene to make the silencing cassette based on the BLAST search results using the *P. sojae* transcriptome database (*see* **Note 6**). Multiple target genes can be silenced using one construct if they share a fragment with high sequence similarity.
3. Clone the PCR products into pTOR vector.
4. Extract the recombinant pTOR plasmid DNA using NucleoBond® Xtra Maxi kit (*see* **Note 7**).

3.2 Propagation of P. sojae

1. Grow *P. sojae* on 90 mm plate by placing an agar plug taken from the margin of a 3–5 days' colony grown on an unclarified V8 plate onto the surface of a nutrient pea agar plate. Seal these plates with parafilm.
2. Place the plates in a 25 °C incubator (constant darkness) for 3–5 days (Fig. 1a), before the hyphae reach the edge of the plate, inoculate four 250 mL Erlenmeyer flasks containing 50 mL nutrient pea broth with 5–6 mycelia discs (4 mm diameter) taken from the edge of the *P. sojae* colonies.
3. After 2–3 days grow at 25 °C in the dark without shaking (*see* **Note 8**), collect mycelium for protoplast isolation (Fig. 1b).

3.3 Protoplast Isolation

1. Harvest *P. sojae* mycelia grown in liquid medium (as described in Subheading 3.2) by pouring the nutrient pea broth and mycelium mix through four layers of sterilized cheesecloth. A 250 mL beaker is used for liquid waste collection.

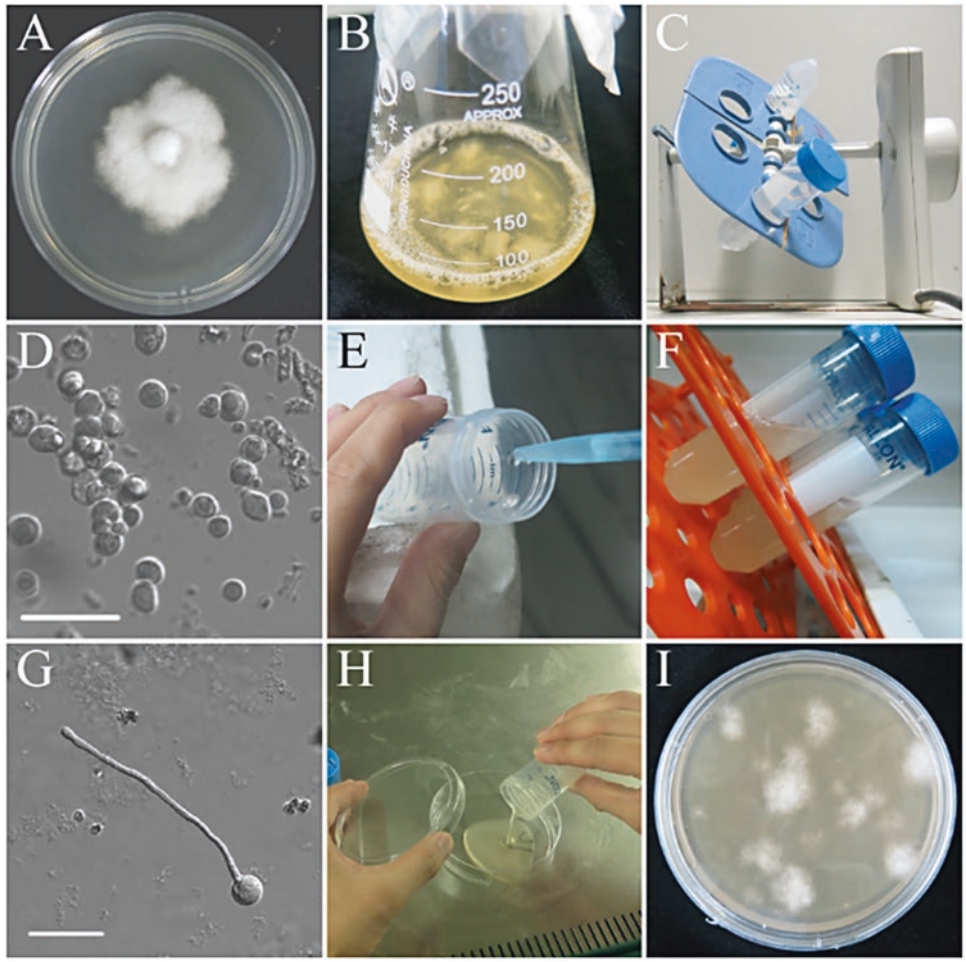


Fig. 1 PEG-mediated transformation procedure of *P. sojae*. (**a**) A colony of *P. sojae* (strain P6497) was grown on nutrient pea broth for 3–5 days. (**b**) An agar plug was used to inoculate nutrient pea broth in a flask. The image was taken after 2–3 days of inoculation. (**c**) *P. sojae* mycelium digestion using a rotating mixer. (**d**) Checking the efficiency of liberation of protoplasts under a microscope. (**e**) Add 40% PEG solution slowly on the wall of a Falcon tube, letting the droplet slide down to the bottom of the tube. (**f**, **g**) Protoplast regeneration and monitoring the progress of the protoplast regeneration under a microscope. (**h**) Pour plates with regenerated protoplasts. (**g**) Single colonies of transformants grown on selective medium. Size bars for the microscopy images are 50 μm

2. Transfer the mycelial mat to a 50 mL Falcon tube with a sterilized long tweezers, add 35 mL 0.8 M mannitol and shake several times to rinse the mycelia.
3. Collect the mycelial mat again using cheesecloth as in **step 1**, transfer it to a new 50 mL Falcon tube, add another 35 mL 0.8 M mannitol. Cover the tube and seal with Parafilm, and then place the tube on a shaker (50–60 rpm) for 10 min for plasmolysis (*see* **Note 9**).

4. During the break, prepare the enzyme solution based on the flask number of liquid culture, 5 mL for each flask of mycelia. Filter-sterilize the enzyme solution to a 50 mL Falcon tube.
5. Collect the mycelia again as in **step 1** and quickly transfer it into the enzyme solution and seal the tube (*see* **Note 10**).
6. Digest for 40 min to 60 min at 25 °C on a rotating mixer (50–60 rpm) (Fig. 1c). Start counting the time immediately after adding the mycelia into the enzyme solution (*see* **Note 11**).
7. Filter the mixture from **step 6** through a miracloth covered over a 50 mL beaker to remove mycelial debris. Protoplasts should be in the flow-through.
8. Transfer the flow-through to a 50 mL Falcon tube. Collect the protoplasts by centrifugation for 3 min at 750 × *g* at 4 °C in a swinging bucket centrifuge.
9. Discard the supernatant. Wash the pellet with 35 mL of ice-cold W5 solution, and centrifuge for 3 min at 750 × *g*, 4 °C (*see* **Note 12**).
10. Pour off the supernatant carefully and resuspend the protoplasts in 10 mL of ice-cold W5 solution. Count the protoplasts under a microscope using a hemacytometer. Rest the protoplasts by keeping them on ice for 30 min (*see* **Note 13**).
11. Centrifuge for 3 min at 750 × *g*, room temperature. Remove the supernatant and resuspend the protoplasts in ice-cold MMg solution to a minimum concentration of 2×10^6 protoplasts/mL (*see* **Note 14**).
12. Leave the protoplast suspension at room temperature for 10 min before transformation (*see* **Note 15**).

3.4 PEG-Mediated Transformation

1. Add 25–30 μg of plasmid DNA to the bottom of each 50 mL Falcon tube placed on ice.
2. Transfer 1 mL of protoplast suspension to each Falcon tube, mix thoroughly by gently tapping the tube, then incubate on ice for 5–10 min.
3. Slowly add 1.74 mL of 40% PEG solution into the mixture of protoplasts and plasmids (Fig. 1e). In the meanwhile, rotate the tube to let the PEG solution gently mix with the protoplast solution and the DNA.
4. Let the mixture sit on ice for 20 min.
5. Add 2 mL of ice-cold mannitol–pea broth into each tube and gently invert the tube to mix the solution. Place on ice for 2 min.
6. Add 8 mL ice-cold mannitol–pea broth to each tube and gently invert the tube to mix it. Place the tube on ice for 2 min.
7. Add 10 mL of ice-cold mannitol–pea broth to each tube and add ampicillin to a final concentration of 50 μg/mL.

3.5 Regeneration and Transformant Harvesting

1. Protoplasts are incubated at 22 °C in the dark for 12–18 h without shaking to allow regeneration of cell walls and germination of the protoplasts (Fig. 1f, g) (*see* **Note 16**).
2. Resuspend the regenerated hyphae in the falcon tube, pipetting up and down several times to dislodge the regenerated hyphae from the inner surface of the tube. Harvest hyphae by centrifuging for 3 min at 1000 × *g*, room temperature in a swinging bucket centrifuge.
3. Keep about 5 mL of supernatant to resuspend the pellet and discard the rest.
4. Disrupt the pellet by pipetting vigorously. Start with a tip cut 1 mL pipette tip (using a hot sterile scalpel) and then with an intact 1 mL pipette tip. It i important to get the regenerated hyphae off the wall of the tube.
5. Add 40 mL liquid (42 °C) mannitol–pea broth containing 1% agar and 40 μg/mL G418 to each Falcon tube, mix by inverting the tubes (*see* **Note 17**).
6. Pour 15 mL pea/mannitol broth containing germinated protoplasts into each of three 90 mm × 15 mm petri dishes (Fig. 1h).
7. Incubate at 25 °C in the dark. Mycelial colonies should be observable after 2 days incubation.
8. Overlay with ~15 mL of liquid (<50 °C) unclarified V8 medium containing 1% agar and 50 μg/mL G418 (*see* **Note 18**).
9. Incubate at RT in the dark and after 3–7 days, pick transformants and transfer to new unclarified V8 agar plates containing 50 μg/mL of G418 (Fig. 1i).

3.6 Collect Mycelial Mat of Transformants for RNA Extraction

1. After 3–7 days' growth at 25 °C in the dark, colonies of transformants will expand on the unclarified V8 agar plates containing 50 μg/mL of G418 (*see* **Note 19**).
2. To acquire the RNA of the transformants, two methods are commonly used to collect the mycelial mat.
3. (a) For transformants cultured on V8 agar plates:

   Place autoclaved sterilized porous cellophane (~16 cm^2) on unclarified V8 agar plates without G418 (*see* **Note 20**). Then inoculate a transformant onto the cellophane. Use a sterile toothpick to collect the mycelial mat when the hyphae reached the edge of the cellophane (time depends on the growth rate of the transformants, typically 5–7 days).

   (b) For transformants cultured in liquid V8 medium.

   Culture ~10 fresh mycelial plugs cut from the edge of a newly grown transformants and transfer the plugs into in Liquid V8

medium without G418 (*see* **Note 20**). After 3 days incubation at 25 °C in the dark, pour off the medium, transfer the mycelium onto a filter paper, fold the filter paper with the mycelium inside, rap the filter paper with tissue paper and press on it until no water can be squeezed out of the mycelial mat.

4. Transfer the mycelial mat to a 1.5 mL tube, freeze it in liquid nitrogen and grind it with a pestle into a powder for RNA extraction.

3.7 RNA Extraction from the Mycelial Mat

1. Add 1 mL of the TRIzol reagent to the tube containing mycelium powder and vortex for 30 s.
2. Incubate the homogenized sample for 5 min at room temperature to permit the complete dissociation of nucleoprotein complexes.
3. Centrifuge at 12,000 × *g* for 10 min at room temperature to remove cell debris. Transfer the supernatant to new tube.
4. Add 0.2 mL of chloroform per 1 mL of TRIzol Reagent. Vortex vigorously for 15 s, then centrifuge the samples for 15 min at 2–8 °C at no more than 12,000 × *g*.
5. The mixture separates into a lower red, phenol-chloroform phase, an interphase, and a colorless upper aqueous phase. RNA remains exclusively in the aqueous phase. Transfer about 600 μL (60% of the TRIzol reagent used) upper aqueous phase carefully without disturbing the interphase into fresh tube.
6. Add 0.5 mL of isopropyl alcohol to the tube containing upper aqueous phase. Incubate samples at 15–30 °C for 10 min and centrifuge for 10 min at 2–4 °C at no more than 12,000 × *g*.
7. Discard the supernatant and add 1 mL 70% ethanol to the tube. Gently inverted the tube several times to wash the pellet, centrifuge at 10,000 × *g* for 5 min at 4 °C.
8. Discard the supernatant and vacuum dry or air dry RNA for 5–10 min, add 40 μL DEPC water to dissolve the RNA before it is over-dried which will decrease the solubility. Use NanoDrop to measure the concentration and purity of the RNA.

3.8 Screening for Gene Silenced Mutants

1. Use SuperScript™ IV VILO™ Master Mix with ezDNase™ Enzyme (reverse transcription kit from Thermofisher) to prepare the cDNA for each RNA sample.
2. Use gene specific primer and Real-time PCR to test the silencing efficiency of the transformants. Usually less than 10% of the transformants can be confirmed as stable gene silenced mutants and individual mutants exhibit different silencing efficiency.

4 Notes

1. $CaCO_3$ and β-sitosterol cannot be completely dissolved in the media. Upon addition of all ingredients, as well as after sterilization, stir the media for 5–10 min. To make solid medium, add agar directly into the flasks before autoclaving.
2. 120 g peas are used for 1 L medium, but in this step, we only adjust the volume to 900 mL. Pea medium is used to make nutrient pea broth and mannitol–pea broth in the next step, where we bring the volume to 1 L. Pea medium needs to be used immediately after filtration because it can get contaminated easily.
3. Enzyme solutions should be made fresh. Weight the enzymes in a biosafety hood and filter-sterilize the solutions.
4. We routinely use the pTOR vector for gene silencing and overexpression in *P. sojae*. The pTOR cassette contains the NPTII gene for the selection of *Phytophthora* transformants on G418 (geneticin), a constitutive oomycete promoter (*ham34*) followed by a multiple cloning site to insert sense or antisense gene fragments and the *ham34* terminator.
5. PEG4000 can be dissolved in 65 °C water bath easily, and PEG4000 should be filter-sterilized. It is easier to filter PEG4000 immediately after removal from the water bath, as cold PEG4000 is very hard to filter.
6. One major bottleneck of gene silencing is associated with off target effects. Sequences of the antisense RNAs need to be designed for both effectiveness and specificity. The optimization for specificity (i.e., predicting and avoiding off-targets) should be based on the transcriptome information (*Phytophthora* transcriptional database V1.1). Ensure no other gene can be affected by the antisense RNA except for the target genes.
7. To prepare for maxi plasmid extraction, the *E. coli* strain is inoculated into a 1 L flask containing 200 mL of LB medium with 50 μg/mL of Ampicillin. Limiting the volume of LB in the flask will help increase the final OD of the cell culture. Typically, more than 0.75 mg of plasmid DNA can be obtained from 200 mL of cell culture.
8. To prevent the mycelium from making aggregates or attaching to the bottom of the flask, shake the flasks twice a day.
9. To treat the hyphae sufficiently with mannitol, it is important to disperse the mycelia uniformly by shaking the tube vigorously after placing the mycelia in the mannitol solution.

10. To enable effective digestion of the hyphae, mycelia should be uniformly mixed with the enzyme solution by shaking the tube vigorously.
11. Digestion efficiency can be monitored with a microscope by examining for the release of protoplasts into the solution. Digested *P. sojae* protoplasts are round and 10–30 μm in size (Fig. 1d).
12. To resuspend the protoplasts, use cut 5 mL pipette tips, add 10 mL of W5 first and pipet gently. Once the protoplasts are resuspended, add W5 to a final volume of 35 mL.
13. Prepare the 40% PEG solution during the waiting time.
14. The volume of MMg that should be used to resuspend the protoplast is based on the protoplast concentration determined in the previous step.
15. During the waiting time, place six to twelve 50 mL Falcon tubes on ice depending on the volume of protoplast and thaw the plasmid taken out from −20 °C.
16. Tilt the tubes to a 30–45° angle to the ground, which will help the regeneration of protoplast. In our experience, protoplasts regenerate better on the inner surface of the tube wall.
17. Do not over heat the liquid agar to over 42 °C. Higher temperatures can kill the regenerated hyphae.
18. Overlay with selection medium containing a higher G418 concentration will us to reduce the false positive rate for the mutant screening.
19. In some cases, when some specific genes are silenced, the growth rate will be impacted. Consequently, it may take a longer time to grow colonies on plates.
20. In some cases, tranformants that were originally resistant to G418 may lose their resistance after several generations. Silencing mutant which is G418 resistance independent will maintain the silencing efficiency better than the G418 resistance dependent mutant. To obtain stable silencing mutants, it is better to use medium without G418 for the growth of tranformants in this step.

Reference

1. Tyler BM (2007) Phytophthora sojae: root rot pathogen of soybean and model oomycete. Mol Plant Pathol 8(1):1–8. https://doi.org/10.1111/j.1364-3703.2006.00373.x
2. Pieter van West SK, Klooster JW v 't, Govers F (1999) Internuclear gene silencing in phytophthora infestans. Mol Cell 3:10
3. Judelson HS, Tyler BM, Michelmore RW (1991) Transformation of the oomycete pathogen, phytophthora infestance. Mol Plant Microbe In 4(6):6
4. Fang Y, Tyler BM (2016) Efficient disruption and replacement of an effector gene in the oomycete Phytophthora sojae using CRISPR/

Cas9. Mol Plant Pathol 17(1):127–139. https://doi.org/10.1111/mpp.12318

5. Tang C, Wang X, Duan X, Wang X, Huang L, Kang Z (2013) Functions of the lethal leaf-spot 1 gene in wheat cell death and disease tolerance to Puccinia striiformis. J Exp Bot 64(10):2955–2969. https://doi.org/10.1093/jxb/ert135
6. Sijen T, Vijn I, Rebocho A, van Blokland R, Roelofs D, Mol JN, Kooter JM (2001) Transcriptional and posttranscriptional gene silencing are mechanistically related. Current biology : CB 11(6):436–440
7. Qiao Y, Liu L, Xiong Q, Flores C, Wong J, Shi J, Wang X, Liu X, Xiang Q, Jiang S, Zhang F, Wang Y, Judelson HS, Chen X, Ma W (2013) Oomycete pathogens encode RNA silencing suppressors. Nat Genet 45:330. https://doi.org/10.1038/ng.2525
8. Torche S (2004) Isolation et etudes d'expression de genes potentiellement lis a la pathogenicite chez Phytophthora porri. Ph.D. thesis. University of Friborg, Switzerland
9. Judelson HS, Michelmore RW (1990) Highly abundant and stage-specific mRNAs in the obligate pathogen Bremia lactucae. Molecular plant-microbe interactions : MPMI 3(4):225–232
10. Giordano E, Rendina R, Peluso I, Furia M (2002) RNAi triggered by symmetrically transcribed transgenes in Drosophila melanogaster. Genetics 160(2):637–648
11. Wilson RC, Doudna JA (2013) Molecular mechanisms of RNA interference. Annu Rev Biophys 42:217–239. https://doi.org/10.1146/annurev-biophys-083012-130404
12. Judelson HS, Tani S (2007) Transgene-induced silencing of the zoosporogenesis-specific NIFC gene cluster of Phytophthora infestans involves chromatin alterations. Eukaryot Cell 6(7):1200–1209. https://doi.org/10.1128/EC.00311-06
13. Yang X, Zhao W, Hua C, Zheng X, Jing M, Li D, Govers F, Meijer HJ, Wang Y (2013) Chemotaxis and oospore formation in Phytophthora sojae are controlled by G-protein-coupled receptors with a phosphatidylinositol phosphate kinase domain. Mol Microbiol 88(2):382–394. https://doi.org/10.1111/mmi.12191
14. Blanco FA, Judelson HS (2005) A bZIP transcription factor from Phytophthora interacts with a protein kinase and is required for zoospore motility and plant infection. Mol Microbiol 56(3):638–648. https://doi.org/10.1111/j.1365-2958.2005.04575.x
15. Abrahamian M, Ah-Fong AMV, Davis C, Andreeva K, Judelson HS (2016) Gene expression and silencing studies in Phytophthora infestans reveal infection-specific nutrient transporters and a role for the nitrate reductase pathway in plant pathogenesis. PLoS Pathog 12(12):e1006097. https://doi.org/10.1371/journal.ppat.1006097
16. Zhang X, Zhai CH, Hua CL, Qiu M, Hao YJ, Nie PP, Ye WW, Wang YC (2016) PsHint1, associated with the G-protein alpha subunit PsGPA1, is required for the chemotaxis and pathogenicity of Phytophthora sojae. Mol Plant Pathol 17(2):272–285. https://doi.org/10.1111/mpp.12279

INDEX

A

ABySS 154, 155, 163, 172, 181
AdapterRemoval 155, 178
Affinity chromatography 82, 85, 87
Agrobacterium tumefaciens 70, 71, 76, 106, 131, 132, 134
Agroinfiltration 131, 132
Alamar blue 28, 35
Albugo laibachii 153
Alkaloid 25, 26
Amplicon 39, 40, 42–46, 143
Anastomosis 13
Antibiotic 1, 3, 5–10, 70, 76, 106, 120, 123, 235, 247, 250, 255, 259, 260, 268
Anti-sense RNA 276, 284
Arabidopsis 68, 77
Ascomycota 1–3, 9, 13
Ascospores 247, 250, 260
Aspergillus 237, 242, 256
Augustus 157, 188, 230
Autofluorescence 110, 114, 116, 123, 127
Auxotroph 238, 246, 247, 252
Auxotrophic markers 235–260

B

Bacteriophage P1 ..... 235, 236
Bacteroidetes 2
Bacto tryptone 132, 276
Barley stripe mosaic virus (BSMV) 140–145, 147, 149
Basic Local Alignment Search Tool (BLAST) 156, 157, 165, 175, 179, 181, 182, 199–205, 279
Basidiomycota 1, 13
BBMap 155, 156, 161, 165, 178, 180
BEDtools2 ..... 156, 166, 187
Bentonite 141
BioAnalyzer 43
Biological assay 33–35
Biotrophic fungi 103, 189
Biotrophic interfacial complex (BIC) 104, 105, 108–110, 112–115
Bisulfite sequencing 188
Blast2GO 226, 227
BLASTmap 199–205
Blumeria graminis 68
Botrytis cinerea 54
Bowtie 54, 55, 57, 217
Bremia lactucae 276
Buchner funnel 14, 16
BUSCO 156, 166, 169, 182, 185, 186
BWA 155, 156, 162, 187

C

Camellia sinensis 26
Carbenicillin 71, 141, 143
Carboxymethylcellulose (CMC) 238, 245, 247, 256
cDNA synthesis 20, 24, 142, 146, 215
Cellulose acetate filters 278
Cellulose chromatography 14, 23
Chitin 140
Chitinase 257
Chloroform 16, 18, 23, 141, 145, 278, 283
Chlorophyll 127
Chloropupukeananin 26
Chlorotic spots 145
CLCworkbench 153
ClustVis 200
Column chromatography (CC) 28, 32
Competent cells 70, 106, 141, 143
Conditional mutagenesis 235
Confocal 104, 106, 108–111, 114, 119–121, 124–126
Conidia 71, 76, 84, 107, 111, 238, 247, 250, 256, 259, 260
Conidiation 94, 238, 245, 249
Contig scaffolding 166, 181, 182
COPE 155, 178
Cre/lox system 235–260
Cre recombinase 236, 237, 245, 254, 256
CRISPR-Cas9 265–273
Cryphonectria parasitica 13
Cryptococcus gattii 53
Cryptococcus neoformans 237, 254
Cuffdiff 218, 219, 225, 230
Cufflinks 55, 213, 219, 224, 230
CummeRbund 224
CutAdapt 155, 178
Cyclohexamide 5, 7, 9
Cytoplasm 104, 109, 110, 113, 115, 116, 127, 128

Wenbo Ma and Thomas Wolpert (eds.), *Plant Pathogenic Fungi and Oomycetes: Methods and Protocols*, Methods in Molecular Biology, vol. 1848, https://doi.org/10.1007/978-1-4939-8724-5,

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Cytoskeleton....128
Czapek-Dox medium....69, 71, 72

D

DADA2....43
de Bruijn graph....180, 181, 185, 220
Deletion of interstitial chromosome segments....235
de novo assembly....153, 171, 180, 208, 220
DESeq....221, 222, 230, 231
Deuteromycota....13
Dicer....55, 67, 276
Dicots....140
Differentially expressed genes (DEGs)....214, 219, 222–225, 227
Differentially regulated clusters (DRCs)....56–58
Dikaryotic....185
3′-5′ Dimethoxy-4′-hydroxy acetophenone (acetosyringone)....132
DNase....15, 18, 22, 23, 41, 74, 146, 215
Double stranded RNA (dsRNA)....13–18, 20, 21, 23, 24, 67, 68, 139, 140, 276
Downy mildew....140, 152, 207, 214
Driselase....257
Drosha....67
DsRed....125

E

EDAseq....230
edgeR....221–223, 231
Effector....v, 54, 103–116, 132, 186, 188, 200, 266
EGFP....104, 106–111, 113–115
Electroporation....120, 122, 255
Endobacteria....1
Endofungal bacteria....1
Endohyphal bacteria (EHB)....1, 2, 5, 7, 9, 10
Endophyte....2, 3, 25, 26, 46, 47, 237
Enterococcus faecium....34, 37
Epichloe festucae....26, 254
Episome....236, 245, 246, 249
ER....128
Ergosterol....140
Erk1/2....94
Escherichia coli (E. coli)....14, 34, 70, 74, 106, 141, 143, 237, 279, 284
Eurotiomycetes....182
Exonerate....157, 176
eXpress....231
Expression profiling....224–226
Extra-invasive hyphal membrane (EIHM)....104, 105, 107–110, 112–115

F

FastQC....154, 155, 159, 160, 177, 209, 213, 216
FastqFilter.sh....162, 179
FastUniq....155, 178
FASTX....155, 174, 177, 178
Fermentation....26, 27, 29–33, 36, 37
FileZilla....209, 211, 212
Firmicutes....2
Fission yeast....236, 242, 254
Flash2 commit c41a82e....161
Fluorescent protein (FP)....103, 104, 107, 112, 113, 119–128
Fluorophore....10, 114
FPKM....219, 221–223, 225, 228
FragScaff....156, 184
FungiDB....268
Fusarium asiaticum....241, 243, 254
Fusarium oxysporum....33, 34, 214
Fusarium verticillioides....140

G

GapCloser....156, 185, 186
GapFiller....156, 173, 185
gbrowse....64
gDNA....47
Gene editing....265, 266, 270, 272
Gene knockdown....266
Gene knockout....275
Gene Ontology (GO)....65, 226, 227
Gene silencing....v, 68, 76, 139–149, 266, 275, 276, 279, 284
Geneious....153
GeneMark....157, 188
Gene replacement....236, 243, 244, 254, 255, 266
Geneticin (G418)....242, 284
Genome annotation....217–219, 230
Genome assembly....v, 152, 153, 158, 163, 175, 179, 187
Genome rearrangements....235, 236
GenomeScope....155, 163, 164, 180, 183
Gentamicin....4–8, 10, 34
Gibberella zeae....252
Gibberellic acid (GA3)....132, 133
GitHub....158, 182
glimmerHMM....157, 188
Glomeromycota....45
Golgi....128
Graphical user interfaces (GUIs)....153, 167, 168, 183, 228

H

Hansenual polymorpha....153
Haploid....152, 180, 185, 189
Haplomerger2....156, 173, 185, 186
Haplotypes....152, 156, 169, 172, 173, 180, 182–186
Haustoria....103, 128
Haustorial cells....140

HeatmapGenerator 200
HeatMapViewer 200
Hemacytometer 258
Hemibiotrophic 103
Hemocytometer 120–122, 278
Heterokaryons 120, 180, 246, 247, 249, 251, 254, 259
Hi-C 156, 171, 174, 184, 186, 188, 189
Higher energy collision dissociation (HCD) 82
High-performance liquid chromatography (HPLC) 26
High-throughput DNA sequencing 39–48
High-throughput screening 131
Host-induced gene silencing (HIGS) ..... v, 68, 140, 146, 147
Hygromycin 71, 106, 107, 120, 121, 126, 242, 255, 259
Hygromycin phosphotransferase (*hph*) 237, 239, 241–247, 249–255, 260
Hypovirus 13

I

Illumina 20, 24, 40, 43, 45, 46, 48, 55, 56, 64, 72, 157, 161, 171, 178, 179, 182, 185, 186, 209, 214–216, 228, 230
Immobilized metal affinity chromatography (IMAC) 82, 83, 85, 87–90
Immunolocalization 119
Immunoprecipitation 82
Infection-related morphogenesis 93
Integrative Genomics Viewer (IGV) 229
Internal Transcribed Spacer regions
ITS1 ..... 44, 45
ITS1F 44, 45
ITS2 ..... 44, 45
ITS4 45
Invasive hyphae 103, 109, 110, 113–115
Iron-IMAC resin 87
Isoform 210, 221, 222, 224, 230
Isopropanol 42, 141, 146

J

Jafnea semitosta 153
JAMg 157, 188
JELLYFISH 155, 163, 164, 181
Jupyter 153

K

Kanamycin 9, 70, 106, 132, 134
KAT 156, 169, 170, 183, 186
K-mer 155, 156, 162–164, 169, 170, 172, 177, 180–183, 185, 216
Kmergenie 155, 163, 181
Kozak initiation sites 125

L

Leaf Plastochron Index (LPI) 41
Lignin 127
limma 231
Liquid chromatography–mass spectrometry (LC/MS) 32, 37
Luria–Bertani medium 70, 276
Luteibacter 2, 3, 8–10
Lysogeny Broth (LB) 2, 3

M

M9 media 3–5, 7, 9
M9 Salts 2, 3
Magnaporthe oryzae 53–65, 68, 81–90, 93–100, 103–116
Magnetic beads 43, 46
Maker 157, 175, 176, 188
Mannitol 120, 267, 268, 271, 277, 278, 280–282, 284
Marker recycling and exchange 235
Mascot Distiller 86
MaSuRCA 155, 172, 181
Mate-pair fragments 157
MaxQuant 84, 86, 89
mCherry 104, 105, 107–116, 120, 125
MEGABLAST 203
Meioses 187
Meraculous-2D 155, 172
Mesophyll 112, 137
Metabarcoding 39, 44
Metal oxide affinity chromatography (MOAC) 82
Metatranscriptomics 207–231
Methylguanosine-capped and polyadenylated sRNAs (CPA-sRNAs) 54, 64–65
Methylome 187
MicroRNAs (miRNAs) 54, 67–70, 72–78
Mitochondria 127, 128
Mitogen-activated protein (MAP) kinases 93–100
Mock community 42, 47
Molecular identifiers (MIDs) 40, 43, 45, 46
Monocot 140
MORPHEUS 200
4-Morpholinoethanesulfonic acid (MES) 71, 132, 268, 277
Mps1 ..... 94, 97–100
mRNA 54, 55, 58, 67, 69, 72, 73, 75, 77, 78, 276
mScarlet 125
Mucoromycota 1
Mueller-Hinton
agar (MHA) 28, 33
broth (MHB) 28, 33, 35
MUMmer 156, 167, 168, 183
Murashige and Skoog medium 132, 133
Mycobiome 39–48
Mycoviruse 13–24

N

NAE buffer 14, 17
NanoDrop Spectrophotometers 146
Nanopore 171, 184
NB-LRRs 200
NCBI 86, 154, 158, 165, 181, 182, 188
Neotyphodium coenophialum 254
Neotyphodium uncinatum 254
Neurospora crassa 53, 237, 243
NextClip 155, 179
Next-generation sequencing 54
Nicotiana benthamiana 131
NOISeq 230, 231
Non-ribosomal peptide 25
Northern blotting 72, 77
Nuclear localization signal (NLS) 104, 107, 109–111, 116, 127
Nuclear magnetic resonance (NMR) 33, 37
Nucleic acid 14, 16–18, 22–24
Nutrient Pea Broth (NPB)/Nutrient Pea Broth Agar (NPBA) 266, 267, 270, 277, 279, 280, 284

O

Obligate parasites 214
Oomycetes ..... v, 77, 103, 125, 132, 140, 151–153, 172, 174, 182, 183, 185–189, 229, 265, 276, 284
Oospore 120–124, 126, 127, 272
Operational taxonomic units (OTUs) 43
Oryza sativa 68, 106
Osm1 ..... 94–96, 98–99
Oxathiapiprolin 272

P

p44/42 ..... 94, 96
Pacific Biosciences (PacBio) 182, 184–186, 230
Paired-end fragments 157
Palindromic recognition motifs 236
PASA 230
PATH 154
PCR 15, 19, 20, 22, 24, 39–43, 45–48, 70, 72–78, 111, 137, 141–143, 148, 155, 160, 171, 178, 215, 238, 243, 244, 247–250, 255, 258, 260, 270, 279
PEAR 155, 178
Perithecia 250, 252, 260
PERL 54, 56
Peronospora belbahrii 214
Peroxisomes 128
Pestalotiopsis 3, 8–10, 26
Petalotiopsis 3
Pezizomycotina 3
Phage 14, 236
Phenol 15, 16, 18, 23, 283
Phenotype screening 131
Phyllosphere 39
Phosphate-buffered saline (PBS) 84
Phosphopeptide 81–90
Phosphoproteome 81–90
Phosphoproteomics 81
Phytobiome 39
Phytophthora 119–128, 132, 153, 164, 265–273, 275–285
Phytophthora capsici 265–273, 275
Phytophthora infestans 132, 153, 164, 275
Phytophthora sojae 152, 275–285
Picard 229
PipeCraft 43
PIPITS 43
Platanus 155, 172
Platycladis orientalis 3
Pmk1 ..... 93–98, 100
Pneumocystis carinii 153
Polyethylene glycol (PEG) 15, 21, 258, 268, 270, 272, 276, 278, 280, 281, 285
Polyketide 25, 26, 31
Polyploid 180, 185
Post-transcriptional gene silencing (PTGS) 276
Posttranslational modification (PTMs) 81
Potato dextrose
  agar 14, 27, 69
  broth 14, 27
Potato virus X (PVX) virus 131
Primer bias 45
Primer ligation 14, 15, 20–22
Progeny 187, 249–253, 255, 260
Proteobacteria 1, 2, 9, 236
Proteome Discoverer 86
Protoplast 115, 237, 243, 244, 246, 247, 249, 251, 254, 256–259, 268, 270–273, 278–282, 285
Protoplasts 121
Pseudomonas syringae 68
Puccinia graminis f. sp. tritici (Pgt) 140
Puccinia striiformis 53, 68
PuTTY 209, 210, 212
Pythium oligandrum 153
Pythium periplocum 153

Q

QIIME 43
Qualimap 229
Quantitative real-time PCR (qPCR) 43, 70, 72–74, 142, 148
Quickmerge 173, 185

R

RACE 15, 21, 70, 72–75, 140
Reads per kilobase of exon model per million reads (RPKM) 219, 222

Real-time PCR....70, 142, 279, 283
REAPR....156, 174, 186
Reassociation....5, 8, 9
Recombinases / recombination....74, 78, 235–237, 244–247, 249, 250, 252, 254–256, 259–260, 266, 270
RepeatMasker....54, 154, 156, 173, 175
RepeatModeler....154, 156, 175
Restriction endonucleases....141, 143
Reverse transcriptase (RT)....15, 18, 20, 22, 70, 74, 142, 147, 148, 200, 282
Reverse transcriptase (RT-PCR)....14, 23
Ribosomal RNA (rRNA)....44, 56, 229
Rice blast....53–65, 82, 93, 98
Rice medium....27, 29, 31, 32, 36, 37
Rifampicin....3, 5, 7, 9, 70
RNA-induced silencing complex (RISC)....67, 276
RNA interference (RNAi)....68, 77, 139
RNase....14, 15, 18, 19, 23, 41, 67, 146, 278
RNA-Seq....157, 231
Rotary evaporator....28, 30, 32, 36
RSEM....213, 221, 231
RSeQC....229
runHM2.sh....173, 186
Rust diseases (stem rust, leaf rust, and stripe rust)....68, 148, 149

S

Saccharomyces cerevisiae....152, 235, 254
Saccharomycetales....182
Sailfish....231
SALSA....156
Sample bias....228
SAMtools....155, 156, 162, 165, 166, 187, 209, 213, 221
Schizosaccharomyces pombe....236
Sclerospora graminicola....172
Sclerotinia sclerotiorum....140
Scythe....155, 178
SDS-PAGE....95, 97, 100
Secondary metabolites....25–37
Sequencing....v, 24, 26, 151, 199, 265
Serological assay....13
Signal transduction....67
Small interfering RNAs (siRNAs)....54, 67, 68, 140
Small RNAs (sRNAs)....v, 53–65, 67–78, 266, 276
Snakemake....153
SNAP....157, 175, 176, 188
snoRNA....56
snRNA....56, 64
SOAPdenovo2....172
SOAPdenovo-Trans....220
Sordariomycetes....3
Spheroplasts....255
Spin column....4
Sporangia....122, 123, 126, 127
SQL....54
SSPACE....156, 172, 184, 186
ssRNA....13, 17
Streptococcus pyogenes....266
SuperDeduper....155, 160
SyMap....156, 167, 168, 183
Synteny....167–169, 181, 183, 185

T

Talcum powder....140, 142
TapeStation....43
Targeted chromosome translocations....235
Target enrichment sequencing (RenSeq)....199–205
Taxomyces andreanae....26
TBE buffer....15, 69
T4 DNA ligase....20, 141, 143, 269
tdTomato....120, 125–127
Terpenes....25
Titanium dioxide (TiO_2)....82, 85, 87–90
Tn7....10
TopHat....55, 209, 213, 217, 218
Trace elements....55, 267, 277
Trans-Abyss....220
Transcription activator-like effector nucleases (TALENs)....266
Transcriptional gene silencing (TGS)....276
Transcriptional start site (TSS)....54
Transcription termination sites (TTS)....65
Transcriptomics....207
Transient expression....v, 131, 132, 254
Trans-kingdom regulation....67
Transposon....10
Trichoderma
Trichoderma harzianum....268, 278
Trichoderma viride....268, 278
Trichomes....135
Trimmomatic....155, 178, 209, 213, 216
Trinity....210, 213, 221, 222, 224, 231
Triticum aestivum....68, 140
TRIzol....55, 69, 72, 141, 145, 215, 278, 283
tRNA....56, 64
Tryptophan 2-monooxygenase gene (*Pgt-IaaM*)....140–145, 148, 149
Tuxedo suite....55, 64

U

UNIX....153, 154
UPARSE....43
UpperQuartile....230
Uredinia....142
Urediniospores....140, 142, 145, 148
USER cloning....75, 78

V

Vacuum Liquid Chromatograph (VLC) 28, 32
Vectors 13, 15, 20, 22, 70, 74–78, 107, 111, 120, 121, 125, 132, 140–143, 145, 200, 224, 270, 276, 279, 284
Vein 112, 135, 136
Velvet 155, 172
Vermiculite 137
Verticillium dahliae 68, 69, 72, 73, 75–77
Virus-induced gene silencing (VIGS) 140
Vitamin stock 267, 277
V8 medium 267, 272, 282

W

Wheat (*Tricum aestivum*) 53, 68, 103, 139–149
Wild potato accessions 132, 133, 135, 138
WinMasker 173

Y

Yarrowia lipolytica 236–237, 254
Yeast extract 2, 27, 55, 94, 132, 141, 238, 256–258, 267, 276, 277

Z

Zinc-finger nucleases (ZFNs) 266
Zoosporogenesis 122, 126, 127