



Genome-Enhanced Detection and Identification (GEDI) of plant pathogens

Nicolas Feau¹, Stéphanie Beauseigle², Marie-Josée Bergeron³, Guillaume J. Bilodeau⁴, Inanc Birol⁵, Sandra Cervantes-Arango¹, Braham Dhillon⁶, Angela L. Dale^{1,7}, Padmini Herath¹, Steven J.M. Jones^{5,8,9}, Josyane Lamarche³, Dario I. Ojeda¹⁰, Monique L. Sakalidis¹¹, Greg Taylor⁵, Clement K.M. Tsui¹², Adnan Uzunovic⁷, Hesther Yueh¹, Philippe Tanguay³ and Richard C. Hamelin^{1,13}

¹ Department of Forest and Conservation Sciences, Forest Sciences Centre, University of British Columbia, Vancouver, BC, Canada

² Biopterre, Sainte-Anne-de-la-Pocatière, Quebec, Canada

³ Canadian Forest Service, Natural Resources Canada, Quebec city, Quebec, Canada

⁴ Ottawa Plant Laboratory, Canadian Food Inspection Agency, Ottawa, Ontario, Canada

⁵ BC Cancer agency, Genome Sciences Centre, Vancouver, BC, Canada

⁶ Department of Plant Pathology, University of Arkansas at Fayetteville, Fayetteville, AR, United States of America

⁷ FPIInnovations, Vancouver, BC, Canada

⁸ Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada

⁹ Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada

¹⁰ Department of Biology Unit of Ecology and Genetics, University of Oulu, Oulu, Finland

¹¹ Department of Plant, Soil & Microbial Sciences and Department of Forestry, Michigan State University, East Lansing, MI, United States of America

¹² Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

¹³ Foresterie et géomatique, Institut de Biologie Intégrative des Systèmes, Laval University, Quebec city, Quebec, Canada

ABSTRACT

Plant diseases caused by fungi and Oomycetes represent worldwide threats to crops and forest ecosystems. Effective prevention and appropriate management of emerging diseases rely on rapid detection and identification of the causal pathogens. The increase in genomic resources makes it possible to generate novel genome-enhanced DNA detection assays that can exploit whole genomes to discover candidate genes for pathogen detection. A pipeline was developed to identify genome regions that discriminate taxa or groups of taxa and can be converted into PCR assays. The modular pipeline is comprised of four components: (1) selection and genome sequencing of phylogenetically related taxa, (2) identification of clusters of orthologous genes, (3) elimination of false positives by filtering, and (4) assay design. This pipeline was applied to some of the most important plant pathogens across three broad taxonomic groups: Phytophthoras (Stramenopiles, Oomycota), Dothideomycetes (Fungi, Ascomycota) and Pucciniales (Fungi, Basidiomycota). Comparison of 73 fungal and Oomycete genomes led the discovery of 5,939 gene clusters that were unique to the targeted taxa and an additional 535 that were common at higher taxonomic levels. Approximately 28% of the 299 tested were converted into qPCR assays that met our set of specificity criteria. This work demonstrates that a genome-wide approach can efficiently identify multiple taxon-specific genome regions that can be converted into highly specific PCR assays. The possibility to easily obtain multiple alternative regions to design highly

Submitted 7 November 2017

Accepted 29 January 2018

Published 22 February 2018

Corresponding author

Richard C. Hamelin,
richard.hamelin@ubc.ca

Academic editor

Peter Solomon

Additional Information and
Declarations can be found on
page 17

DOI 10.7717/peerj.4392

Distributed under
Creative Commons Public
Domain Dedication

OPEN ACCESS

specific qPCR assays should be of great help in tackling challenging cases for which higher taxon-resolution is needed.

Subjects Genomics, Mycology, Plant Science

Keywords Plant pathogens, Genomics, Detection and identification, Diagnostics, Mycology, Plant pathogen

INTRODUCTION

Plant diseases caused by fungi and Oomycetes represent worldwide threats to crops and forest ecosystems. The inadequacy of disease diagnostics, reporting protocols and the lack of centralized recording mechanisms have been identified as causes for the increase in emerging diseases of plants and animals (*Fisher et al., 2012*). To prevent and manage those emerging diseases, rapid detection and identification of the causal pathogens are crucial. This is particularly important for tree diseases where large contiguous forest ecosystems can be threatened by invasive and emerging pathogens, resulting in ecosystem-wide irreversible damage when prevention and management fail (*Pautasso, Schlegel & Holdenrieder, 2015*). Plant pathogens can be transmitted by a variety of means, including natural dispersal in water, rain and wind. But because of their cryptic nature and their ability to remain dormant in plants or soil, introduction and spread of alien invasive species are often related to anthropogenic activities. The intensification of drivers such as international trade, combined with climate change contribute to the emergence and rapid increase of the threat that plant pathogens may cause to ecosystems (*Desprez-Loustau et al., 2016; Pautasso, Schlegel & Holdenrieder, 2015; Santini et al., 2013*).

DNA detection methods have provided powerful tools with broad applications for the detection and monitoring of invasive species. However, DNA-based assay design for plant pathogen detection relies mostly on the amplification and detection of a very small fraction of the target organism's genome. Assays are generally designed to amplify between one and three genes or genome regions that are conserved in all eukaryotes such as the internal transcribed spacer region (ITS rDNA), intergenic spacer region (IGS) and beta-tubulin (BT) (*Schena et al., 2013*). Single nucleotide polymorphisms (SNPs) in these conserved regions are sometimes abundant enough to design primers and/or probes that are specific to a targeted taxon and can discriminate closely related species (*Prévost-Bouré et al., 2011; Thonar, Erb & Jansa, 2012; Schena et al., 2013*). However, this method has limitations when applied to taxa that diverged recently since they often share highly homologous or even identical sequences in these conserved gene regions. This limits assay design and increases the risk of false positive results by making it difficult to find discriminant SNPs. This could be critical in cases where closely related microorganisms pose different risks and lead to different epidemiological or regulatory outcomes.

Genomics can play a valuable role in the discovery and characterization of emerging pathogens (*Firth & Lipkin, 2013*). The increase in genomic resources brought by next generation sequencing opens the way to mining entire genomes of pathogens and their close relatives to identify genes or genomic regions that are unique to a taxon or to a

group of taxa. This is promising and has been used to design assays for a wide range of microbe-host combinations. Applications include detecting genetic material shed by invasive fish in environmental (eDNA) water samples (Farrington *et al.*, 2015), human infecting bacteria (Ho *et al.*, 2011; Ho *et al.*, 2012; Hung *et al.*, 2012), zoonotic infectious bacteria (Hänsel *et al.*, 2015; Blaecher *et al.*, 2017) and plant pathogenic bacteria (Pritchard *et al.*, 2016).

Herein, we describe a pipeline to compare whole genomes of plant pathogens and phylogenetically related taxa to identify genome regions specific to targeted taxa or groups of related taxa and design highly accurate PCR assays. As a proof-of-concept, we have applied our assay development pipeline to a range of taxonomically diverse plant pathogens to demonstrate that it is robust, flexible and broadly applicable. Our method allowed for the rapid and efficient discovery of genome regions that provide highly specific targets for the development of assays to detect some of the most damaging plant pathogens across three broad taxonomic groups: Phytophthora (Heterokonts, Oomycota), Dothideomycetes (Fungi, Ascomycota) and Pucciniales (Fungi, Basidiomycota). This pipeline can be applied to a broad range of organisms, including pathogens of crops, forest trees and animals for which genomic resources are available.

MATERIALS AND METHODS

Pipeline development

Our approach required access to assembled and annotated genomes of the targeted organisms as well as closely related taxa. The underlying principle of our method is to compare the protein content within the genomes of phylogenetically related taxa to ensure the selection of targets that are discriminant towards the most closely related known species. Our bioinformatics pipeline is divided into four modules (Fig. 1). In the first module, genomic resources (i.e., sets of gene and protein models predicted from a whole genome sequence) are selected for the targeted species under investigation and for a group of related species. In the second module, clusters of homologous genes are generated from the protein sets of all available genomes to identify candidate protein clusters that are unique to the targeted species. The third module is a filtering step aimed at eliminating false positive candidates generated in the previous step. The fourth module consists of automated primer and probe design for the candidate genes retained in module 3. The four modules are detailed below and the pipeline is applied to plant pathogens.

Module 1: genomic resources

In order to maximize the accurate identification of candidate protein clusters that are unique to the targeted taxa and contain enough polymorphisms for the design of species-specific or group-specific qPCR assays, genomes of target and related non-target taxa are required (e.g., same genus and order). To achieve this, *de novo* genome assemblies and full protein sets are either produced and assembled for the targeted species under investigation and for a group of related species or recovered from public genome data repositories. Next-generation sequencing (NGS) techniques constitute a fast and cost-effective way of obtaining whole genome sequences, particularly in eukaryotic organisms with small

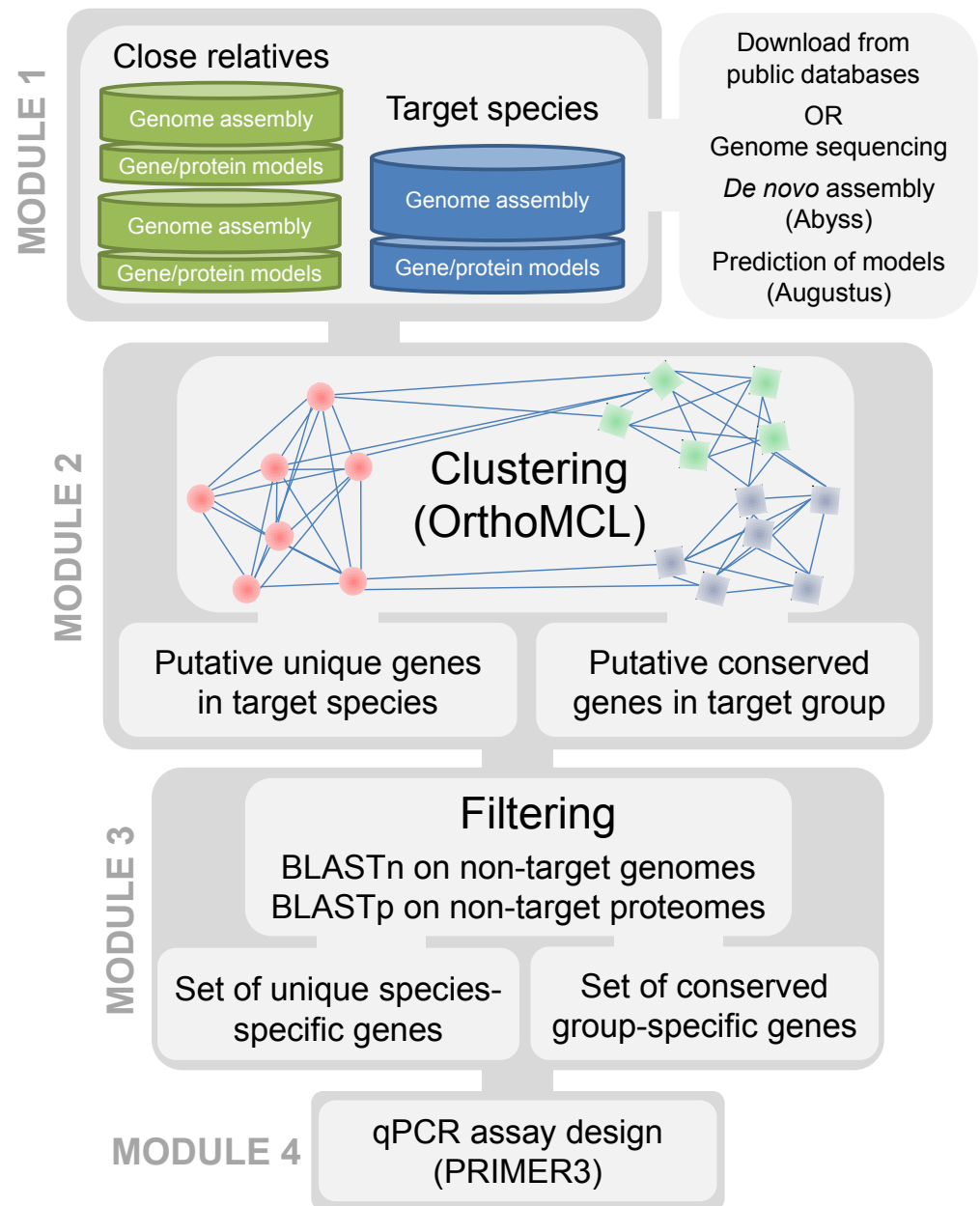


Figure 1 Pipeline for development of qPCR assays using whole genomes.

[Full-size](#) DOI: 10.7717/peerj.4392/fig-1

genomes such as pathogenic fungi (Faino & Thomma, 2014). *De novo* genome assemblies can quickly be achieved with assemblers such as ABySS (Simpson et al., 2012; Gallo et al., 2014; Abbas, Malluhi & Balakrishnan, 2014) (see: ‘Computational testing and experimental screening’); then, AUGUSTUS (Stanke et al., 2006) can provide a fast way to generate *ab initio* gene/protein model prediction (i.e., the “gene space”) from these *de novo* assemblies (Ma & Fedorova, 2010; Haas et al., 2011; Faino & Thomma, 2014). In addition, a collection of genomes of target and non-target taxa can be obtained through the publicly available

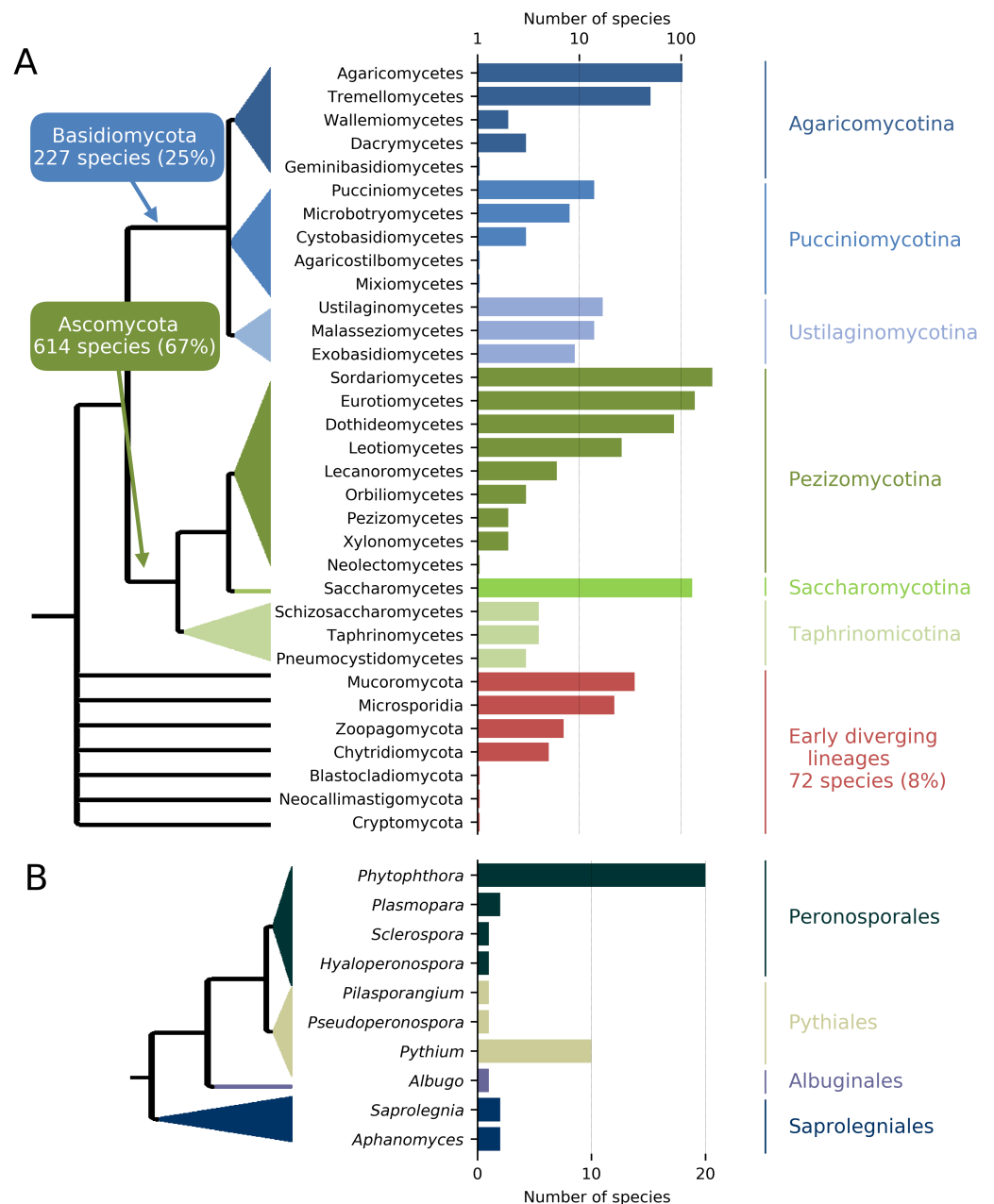


Figure 2 Number and phylogenetic coverage of fungal (A) and Oomycete (B) genomes available on the NCBI public database.

Full-size DOI: [10.7717/peerj.4392/fig-2](https://doi.org/10.7717/peerj.4392/fig-2)

fungal genome sequencing initiatives such as the Mycosom (*Grigoriev et al., 2014*) and the 1000 Fungal Genomes project (<http://1000.fungalgenomes.org/home/>). Currently, more than 2,100 *de novo* fungal genome assemblies are available in the NCBI database, representing about one thousand fungal species distributed along all the branches of the fungal tree of life (Fig. 2).

Module 2: discovering homologous gene clusters

OrthoMCL (*Li, Stoeckert & Roos, 2003*) is used to generate clusters of homologous genes (orthologs and paralogs) using the protein sets of all genomes previously selected. This allows for identification of candidate protein clusters that are unique to a taxon or that are common to members of a group and absent in the other taxa (“higher hierarchical levels”). Briefly, the OrthoMCL procedure comprises an all-against-all BLASTp step with a similarity cutoff defined by an e -value of $1e-20$ and minimum overlap of 50% followed by a Markov clustering (MCL) of groups of homologs. Cluster tightness is regulated by an inflation parameter arbitrarily selected between values of 1.1 (fewer clusters with more proteins in each) to 5.0 (smaller clusters of proteins with high similarity); by default, we used $I = 1.5$ to try to cluster together as many homologous proteins with relatively high similarity. Following MCL clustering the output file is parsed to identify candidate clusters. These can be taxa-specific clusters containing one or several proteins unique to a single taxon or higher level hierarchical clusters containing at least one protein homolog for each taxon in a group (e.g., multiples species within a clade); cDNA sequences are then retrieved from each candidate protein cluster, aligned using MUSCLE (*Edgar, 2004*) and stored in a fasta format until further processing.

Module 3: filtering false positives

This filtering step is crucial to account for putative unique genes that are present in other genomes but were mis-annotated and/or fragmented. Gene fragments may be caused by truncation of the protein-coding sequence in the genome due to the termination of a contig in the middle of an open reading frame. Such fragments may result in OrthoMCL clustering errors by clustering them in groups that are distinct from the group containing their true full-length homologs. Thus, to ensure that candidate clusters were truly unique, i.e., specific to the target genome, BLASTn and BLASTp searches were conducted using candidate proteins and cDNA sequences retrieved in Module 2 against all non-target genomes and protein sets, respectively ([Fig. 1](#)). The BLASTn search minimizes the likelihood of obtaining a candidate cluster that would result from an annotation error in the target genome or due to gene mispredictions in the non-target genomes. The BLASTp search aims to minimize the effect of truncated protein models ([Text S1](#)).

The e -value cutoff for BLASTp and BLASTn searches must be selected based on the level of identity between the genomes of target and non-target taxa, the estimated number of candidates, the quality of the genomes and the tolerance of potential false positives. Filtering simulations (see Results section ‘Computational testing’) helped to identify default parameters with “stringent” cutoff threshold values set at $1e-20$ for both BLASTn and BLASTp to retain a minimum of candidate clusters.

Module 4: assay design

The final module implements the assay design, using PRIMER3 (*Rozen & Skaletsky, 1999*) to generate unique or common primers and probes for PCR or qPCR assays. For each candidate gene, a primer pair and a probe are automatically designed from the cDNA alignment with PRIMER3 parameters as described in [Table 1](#).

Table 1 Oligonucleotide primer and probes default parameters used in qPCR automated assay design (Module 4).

| | Optimal | Range | Other |
|--|---------|-----------|-----------|
| Amplicon size | – | 80–150 bp | – |
| Primers | | | |
| Length | 20 nt | 18–24 nt | – |
| Melting temperature (T_m) | 60 °C | 60–63 °C | – |
| GC% | – | 50–60% | – |
| Maximum T_m difference between primers | 1.0 | – | – |
| Maximum 3' self-complementary | 1.0 | – | – |
| Maximum poly X | 3.0 | – | – |
| 3'-end | – | – | No G or C |
| Hybridization probe | | | |
| Length | – | 18–30 nt | – |
| Melting temperature (T_m) | – | 65–70 °C | – |
| GC% | – | 45–65% | – |
| Maximum poly X | 3.0 | – | – |
| 5'-end | – | – | No G or C |

Computational testing and experimental screening

Testing the cutoff filters

False positive filtering with BLASTn and BLASTp (Module 3) was first tested with a set of 11 *Phytophthora* genomes and subsequently using a second set of 16 Dothideomycetes genomes. For each dataset, Module 2 was run with default parameters. Then, filtering simulations with different BLASTn and BLASTp cutoff values were carried out as described in [Text S1](#) (“Filtering simulations with Module 3”).

Pipeline runs

We assessed the robustness of this pipeline by testing it in three divergent taxonomic groups. The genomes of 19 tree pathogens belonging to the genus *Phytophthora*, the class Dothideomycetes or the order Uredinales were sequenced, assembled and annotated ([Tables S1–S3](#)). Paired-end Illumina sequencing, read filtering and assembly procedures were described in [Lamarche et al. \(2015\)](#) and [Feau et al. \(2016\)](#). *Ab initio* gene and protein model predictions were obtained by using AUGUSTUS trained with (i) *de novo* transcriptome assemblies obtained using trans-ABYSS for the corresponding species of *Phytophthora* and Dothideomycetes ([Text S1](#) “RNA-seq libraries”) or (ii) gene and protein sets from closely related species for the Pucciniales. From each of these three genome collections, a subset was selected, which included the targeted species and other species in the same taxonomic group. This subset was then enriched with publicly available genomes from additional phylogenetically related species.

We tested the pipeline on the invasive species *Phytophthora ramorum* responsible for the sudden oak death and sudden larch death diseases, *P. lateralis* a pathogen of Port Orford Cedar (*Chamaecyparis lawsoniana*) and *P. kernoviae* a pathogen of european beech ([Hansen et al., 2000](#); [Brasier et al., 2005](#); [Brasier & Webber, 2010](#)). In order to perform

genome comparison to identify unique genome regions, we obtained the genomes of eight non-target taxa from phylogenetic clades 1, 2, 7, 8 and 10 (as defined in [Martin, Blair & Coffey, 2014](#)) (Table S3). In the Dothideomycetes, we obtained the genomes of 39 Dothideomycetes (mostly in the Capnodiales and one Pleosporales; Table S1) and searched for candidate genes in two poplar pathogens causing cankers (*Sphaerulina musiva*) and leaf spots (*S. populicola*; [Dhillon et al., 2015](#)) and in *Phaeocryptopus gaeumannii*, the causal agent of the Swiss needle cast of Douglas fir.

For the rust fungi, we obtained the genome sequences of 17 Pucciniales, including members of the genera *Melampsora*, *Cronartium*, *Endocronartium*, *Puccinia*, *Mixia*, *Sporobolomyces* and *Rhodotorula* (Table S2). We searched these genomes for candidate genes in three rusts attacking poplars and pines: the European (*Melampsora larici-populina*) and North American (*M. medusae* f. sp. *deltoidae*) poplar rust and the white pine blister rust (*Cronartium ribicola*).

***In silico* screening for intra-taxon variation**

We assessed the possibility that some of the candidate genes obtained in Module 3 may fail to generate the expected amplicon in all the individuals within the targeted taxa due to intra-taxon variability and presence/absence polymorphisms. BLASTn searches (*e*-value cutoff of $1e-20$) were conducted using the candidate genes obtained for *P. ramorum* after Module 3 filtering against 40 *P. ramorum de novo* genome assemblies (AL Dale, N Feau, SE Everhart, B Dhillon, B Wong, J Sheppard, GJ Bilodeau, A Brar, JF Tabima, D Shen, CM Brasier, BM Tyler, NJ Grunwald & RC Hamelin, 2013, unpublished data). Similarly, candidate genes targeting *S. musiva* were searched against 16 *de novo* genome assemblies obtained for this species (ML Sakalidis, N Feau, B Dhillon & RC Hamelin, 2013, unpublished data). We then conducted some simulations to identify the number of candidate genes that would need to be multiplexed to successfully target all individuals within the target species. This consisted in searching the *de novo* assemblies (BLASTn *e*-value cutoff of $1e-20$) with combination patterns of one to ten candidate genes. For each pattern, 500 random combinations were generated and tested.

Experimental screening

For each group of plant pathogens, we selected a subset of the candidate genes identified by the pipeline. Candidate genes from different scaffolds were selected to maximize the representation of genomic regions. Candidate genes were screened with biological material to eliminate genes that failed to generate the expected amplicon in the targeted species or generated amplicons in the most closely related species. Screening was conducted by performing standard PCR using the selected primers and DNA of the target and the most closely related species, followed by gel electrophoresis.

DNA was extracted from seventeen Dothideomycetes cultures representing 14 Capnodiales species, including five *Sphaerulina* species closely related to *S. musiva* and *S. populicola* ([Feau, Hamelin & Bernier, 2006](#)). For testing the candidate genes obtained for *P. ramorum* and *P. lateralis*, we used DNA extracted from 41 species distributed among nine phylogenetic clades (out of the ten clades known; [Martin, Blair & Coffey, 2014](#)), including 11 species from clade 8 which are closely related to the two targeted species. This collection

was reduced to 22 species for testing *P. kernoviae*-specific candidate genes, but included the other two species that are found in clade 10 (*P. boehmeriae* and *P. gallica*) with *P. kernoviae*. DNA from rust samples was obtained from 19 *Melampsora* and 12 *Cronartium* and *Endocronartium* taxa and 18 species from three additional Uredinales genera. For the Phytophthora and Dothideomycetes, 5 ng of DNA template was used in the PCR reactions (95 °C for 3 min, 29 cycles of 95 °C for 30 s, 62 °C for 30 s and 72 °C for 1 min, and a final extension step of 72 °C for 10 min) in a 25 µl volume reaction mix (10 µM primer concentration and 0.2 µl Platinum Taq DNA polymerase; Invitrogen, Life Science Technologies, Carlsbad, CA, USA). For the Pucciniales, primer concentration was at 1 µM and thermocycling conditions were slightly different : 94 °C for 3 min, 35 cycles of 92 °C for 30 s, 60 °C for 30 s and 72 °C for 1 min, and a final extension step of 72 °C for 10 min. PCR amplification was assessed running 5 µl of the PCR product in a 1.5–2% agarose gel stained with ethidium bromide.

RESULTS

Computational testing

Filtering simulations

Simulations with BLASTn and BLASTp filtering conducted on a set of 13 Phytophthora and 16 Dothideomycetes genomes suggested that the BLASTp *e*-value cutoff had a positive impact in removing false positive candidates when the genome sequences used were of poor quality (i.e., “fragmented genomes”; [Text S1](#) “Filtering simulations with Module 3”, [Figs. S1](#) and [S2](#)). When genomic distances between target and non-target taxa are short, rejection of candidate by the BLASTn filter can rapidly increase with diminution of the *e*-value cutoff ([Fig. S3B](#)). Use of stringent filters (i.e., high *e*-value cutoffs) rejected most of the potential false positive candidates, in particular when the genomic distances between the target and non-target taxa were short i.e., when taxa were closely related ([Text S1](#) “Filtering simulations with Module 3”, [Fig. S3](#)).

Pipeline run on Phytophthora

Nine *Phytophthora* genomes were selected in Module 1 ([Table 2](#); [Feau et al., 2016](#)). OrthoMCL search in Module 2 identified 52,280 clusters, including 1,624 (3.1%) and 5,578 (10.7%) putative unique clusters in *P. ramorum* and *P. lateralis*, respectively ([Table 2](#)). The filtering step in Module 3 decreased the number of candidates to 37 for *P. ramorum* and 180 for *P. lateralis*, suggesting that about 97% of the putative unique candidate clusters predicted in the OrthoMCL analysis were false positives according to our filtering criteria ([Table 2](#)). At a higher hierarchical level, the OrthoMCL search resulted in 110 clusters that were shared between *P. ramorum* and *P. lateralis* and not found in the six other species; nine (0.02%) of these clusters were retained after the Module 3 filtering with default values. For *P. kernoviae*, 830 putative unique clusters were identified out of 44,422 OrthoMCL clusters, which were reduced to 55 candidates after Module 3 filtering.

Pipeline run on Dothideomycetes

Genomic resources for 13 Dothideomycete fungal species were selected in Module 1 of the pipeline ([Table 2](#)). Statistics for the *de novo* genome assemblies were within the range of

Table 2 Number of species and genus-specific protein models found after clustering with OrthoMCL and filtering with BLASTn.

| Targets | Module 1 | | Module 2 | | Module 3 |
|---|---|---------------------|---------------------|----------------------------|----------------------|
| | Non-target genomes | # of protein models | # OrthoMCL clusters | # OrthoMCL unique clusters | # of unique clusters |
| <i>Phytophthora</i> | | | | | |
| Species-specific | | | | | |
| <i>P. ramorum</i> | PINF ^a , PSOJ, PLAT, PCAP, PCIN, PHIB, PFOL | | 52,280 ^b | 1,624 (3.1%) | 37 |
| <i>P. lateralis</i> | PINF, PSOJ, PRAM, PCAP, PCIN, PHIB, PFOL | | 52,280 ^b | 5,578 (10.7%) | 180 |
| <i>P. kernoviae</i> | PINF, PSOJ, PRAM, PCAP, PCIN | | 44,422 | 830 (1.9%) | 55 |
| Group-specific | | | | | |
| <i>P. ramorum</i> + <i>P. lateralis</i> | PINF, PSOJ, PCAP, PCIN, PHIB, PFOL | | 52,280 | 110 (0.2%) | 9 |
| <i>Dothideomycetes</i> | | | | | |
| Species-specific | | | | | |
| <i>Sphaerulina musiva</i> | STON, SPOP, MPUN, PGAE, MGRAM, MLAR, MGIB, DPIN, DOTS, MDEAR, CLAF, MFIJ | 160,617 | 44,189 ^b | 885 (2.0%) | 131 |
| <i>S. populicola</i> | STON, SMUS, MPUN, PGAE, MGRAM, MLAR, MGIB, DPIN, DOTS, MDEAR, CLAF, MFIJ | 160,617 | 44,189 ^b | 765 (1.7%) | 134 |
| <i>Phaeocryptopus gaeumannii</i> | STON, SMUS, SPOP, MPUN, MGRAM, MLAR, MGIB, DPIN, DOTS, MDEAR, CLAF, MFIJ, CERZ, BAUCO, DIDZE | 193,449 | 51,248 ^b | 3,350 (6.5%) | 1,000 |
| Group-specific | | | | | |
| <i>S. musiva</i> + <i>S. populicola</i> | STON, MPUN, PGAE, MGRAM, MLAR, MGIB, DPIN, DOTS, MDEAR, CLAF, MFIJ | 160,617 | 44,189 | 392 (0.9%) | 68 |
| <i>S. musiva</i> + <i>S. populicola</i> + STON1 | STON, MPUN, PGAE, MGRAM, MLAR, MGIB, DPIN, DOTS, MDEAR, CLAF, MFIJ | 160,617 | 44,189 | 345 (0.8%) | 163 |
| <i>Rusts</i> | | | | | |
| Species-specific | | | | | |
| <i>Melampsora larici-populina</i> | All Pucciniales genomes in Table S3 , but <i>M. larici-populina</i> | | 13,355 ^b | 3,550 (26.6%) | 1,519 |
| <i>M. medusae</i> f. sp. <i>deltoidae</i> | All Pucciniales genomes in Table S3 , but <i>M. medusae</i> f. sp. <i>deltoidae</i> | | 20,713 ^b | 8,901 (43.9%) | 1,542 |

(continued on next page)

Table 2 (continued)

| Targets | Module 1 | | Module 2 | | Module 3 |
|----------------------------|---|---------------------|---------------------|----------------------------|----------------------|
| | Non-target genomes | # of protein models | # OrthoMCL clusters | # OrthoMCL unique clusters | # of unique clusters |
| <i>Cronartium ribicola</i> | All Pucciniales genomes in Table S3, but <i>C. ribicola</i> | | 9,633 ^b | 2,782 (28.8%) | 1,341 |
| Genus-specific | | | | | |
| <i>Melampsora</i> genus | All Pucciniales genomes in Table S3, but <i>Melampsora</i> spp. | | 1,870 ^c | 374 (20.0%) | 270 |
| <i>Cronartium</i> genus | All Pucciniales genomes in Table S3, but <i>Cronartium</i> spp. | | 1,027 ^c | 51 (5.0%) | 34 |

Notes.

^aSpecies name abbreviations: PINF, *Phytophthora infestans*; PSOJ, *P. sojae*; PCAP, *P. capsicii*; PCIN, *P. cinnamomi* var. *cinnamomi*; PHIB, *P. hibernalis*; PFOL, *P. foliorum*; PRAM, *P. ramorum*; PLAT, *P. lateralis*; STON, *Mycosphaerella* sp. STON; SMUS, *Sphaerulina musiva*; SPOP, *S. populicola*; PGAE, *Phaeocryptopus gaeumannii*; MPUN, *Ramularia endophylla*; MGRAM, *Zymoseptoria tritici*; MLAR, *M. laricina*; MGIB, *Pseudocercospora pini-densiflorae*; DPIN, *Dothistroma pini*; DOTS, *D. septosporum*; MDEAR, *Lecanostica acicula*; CLAF, *Cladosporium fulvum*; MFIJ, *M. fijiensis*; CERZ, *C. zeaе-maydis*; BAUCO, *Baudoinia compniacensis*; DIDZE, *Didymella zeaе-maydis*.

^bNumber of OrthoMCL clusters that include at least one gene from the targeted species.

^cNumber of OrthoMCL clusters that include at least one gene for each species of the genus.

those obtained from public databases with N50 ranging from 0.06 to 0.18Mb and BUSCO completeness over 99.3% (Table S1). From 160,617 protein models, OrthoMCL generated 16,103 protein clusters and 28,086 singletons, with 885 and 765 unique clusters for *S. musiva* and *S. populicola*, respectively. This represents about 8.0% of the protein content predicted for these two genomes (Dhillon et al., 2015). Module 3 filtering reduced this number to 131 and 134 candidates for *S. musiva* and *S. populicola*, respectively. At higher hierarchical levels, analysis of poplar pathogens *S. musiva*, *S. populicola* and *Mycosphaerella* sp. STON1 resulted in 163 candidate clusters (52.8% of the candidate clusters were eliminated) whereas for the two pathogens of the poplar section Aigeiros, *S. musiva* and *S. populicola*, 82.6% of the 392 candidate clusters were discarded. Adding three Dothideomycete genomes to our initial collection before searching candidates for the Swiss needle cast fungus *P. gaeumannii* increased the total number of clusters to 16,662 (17% increase) and singletons (34,586) found by OrthoMCL (Table 2). Among these, 3350 (6.5%) were unique *in silico* candidates for *P. gaeumannii* and 1,000 candidate clusters remained after Module 3 filtering.

Pipeline run on Pucciniales

In Module 1, a comprehensive proteome dataset from 17 rust species (including nine obtained in this project) was selected to identify candidates unique to the individual species within two rust fungi genera, *Melampsora* (*M. larici-populina* and *M. medusae*) and *Cronartium* (*C. ribicola*), and genes that are conserved at the genus level for these two groups. OrthoMCL clustering (Module 2) resulted in a similar proportion of candidate clusters in *M. larici-populina* (3,550 clusters; 26.6% of all clusters) and *C. ribicola* (2,782; 28.8%) (Table 2). The number of unique clusters obtained for *M. larici-populina* was only 9% lower than what was obtained in a pairwise comparison with the wheat rust *Puccinia graminis* f. sp. *tritici* (3,903 unique gene clusters) (Duplessis et al., 2011), suggesting that addition of new proteomes to the clustering analysis does not have a drastic impact on the number of species-specific clusters. A high proportion of unique clusters (about

twice the number found in the other two rust species) was obtained for *M. medusae* f. sp. *deltoidea*, likely resulting from the high number of protein models predicted in the genome of this pathogen (Table 2). After filtering in Module 3 the number of unique genes was reduced to 1,519, 1,542 and 1,341 for *M. larici-populina*, *M. medusae* f. sp. *deltoidea* and *C. ribicola*, respectively (Table 2). Filtering at the genus level resulted in only 270 candidate clusters shared between seven *Melampsora* species and 34 between the four *Cronartium/Endocronartium* species (Table 2 and Table S3). Module 4 rejected 73.3% to 76.4% (for *Melampsora* and *Cronartium*, respectively) of these candidates due to the presence of interspecific polymorphisms preventing the design of primers and probes with PRIMER3.

***In silico* screening for intra-taxa variation**

Twenty-one out of the 37 candidate genes (56.7%) obtained for *P. ramorum* were successfully retrieved in the 40 *de novo* assemblies tested for *P. ramorum* (Fig. 3A). The observed level of presence/absence of the candidate genes may be explained by polymorphism among the divergent *P. ramorum* lineages. Simulations carried out on the *de novo* assemblies with random combinations of one to nine of the 40 candidate genes indicated that a minimum of four candidate genes have to be multiplexed to successfully target the expected amplicon in 99.99% of the *P. ramorum* individuals (Fig. 3B). For *S. musiva*, 82% of the candidate genes were found in all 16 *de novo* genome assemblies tested. For this taxon, simulations indicated that the combination of three random candidate genes should be sufficient to successfully target all *S. musiva* individuals (Figs. 3C and 3D).

Experimental screening

For each targeted taxon or group studied, primer pairs for 5 to 65 candidate genes identified in the pipeline were evaluated for specificity. Overall, 76 (25.5%) candidate genes were validated out of the 297 tested. Within each group, the proportion of retained candidate genes after wet lab testing for specificity were in the same range (22.6, 28.0% and 33.3% for the Pucciniales, Dothideomycetes and Phytophthora, respectively) and not significantly different from the overall proportion of successful candidates (Chi-2 value = 1.76, P [3df] = 0.62) (Table 3) (Figs. S4 and S5). Furthermore, the number of non-target species considered in each group for PCR-testing (11–24 species for the Pucciniales, 11–14 species for the Dothideomycetes and 22–40 species for Phytophthora; Table 3) did not impact this success rate, as suggested by the lack of correlation between these two variables ($R^2 = -0.06$, $P = 0.55$). Similarly, the quality of the *de novo* genome assembly (genome N50-value; Tables S1–S3) used for the targeted taxa does not seem to impact this success ($R^2 = -0.06$, $P = 0.61$).

DISCUSSION

We provide a proof-of-concept for using whole genome sequence comparisons to design detection and diagnostic assays at different hierarchical levels in taxonomically diverse groups of crop and tree pathogens. The 73 whole genome sequences generated in this

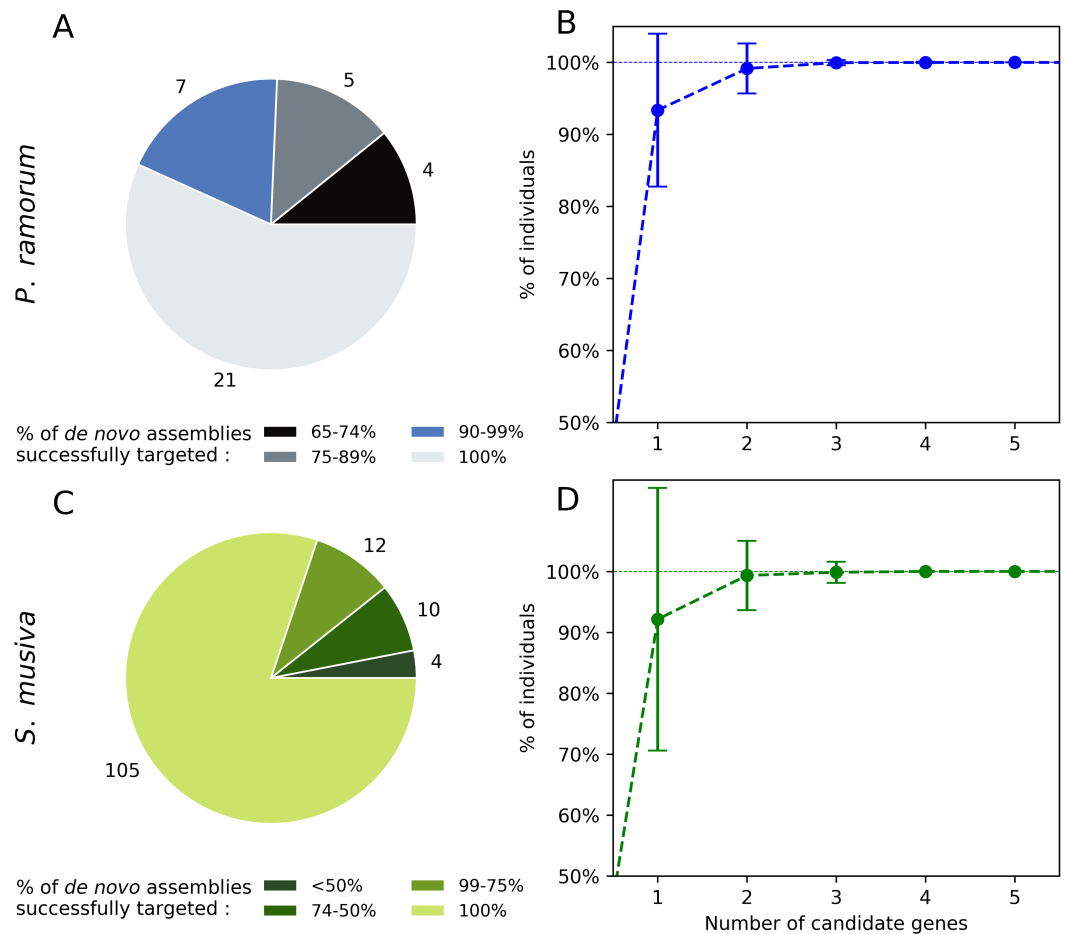


Figure 3 *In silico* screening for intra-taxon variation in *Phytophthora ramorum* (A and B) and *Sphaerulina musiva* (C and D). Number of candidate genes predicted for *P. ramorum* ($n = 37$) and *S. musiva* ($n = 134$) that targeted different proportions of the *de novo* genome assemblies of *P. ramorum* ($n = 40$) (A) and *S. musiva* ($n = 16$) (C). Minimum number of candidate gene required to successfully target all the *de novo* assemblies of *P. ramorum* (B) and *S. musiva* (C).

Full-size DOI: 10.7717/peerj.4392/fig-3

project or obtained from public databases were parsed to discover genes that are either unique to individual taxa or shared at different hierarchical levels within groups of taxa. Our approach, comprising *in silico* design and *in vitro* validation steps, generated 85 assays designed to specifically target species or groups of related species belonging to the Phytophthoras, Dothideomycetes and Pucciniales.

Having access to whole genome sequences provides an unprecedented opportunity to explore levels of variability among genomes, genes and taxa. With the expansion of the comparative genomics field and the development of tools such as Markov clustering (Li, Stoeckert & Roos, 2003; Szilagyi & Szilagyi, 2014), pairwise BLASTp and BLASTn (Lin et al., 2010; Yang et al., 2013; Xu et al., 2015), identification of the “non-core genome”, i.e., taxon-specific regions and genes, has become routine in computational genomics. Gene content-based differentiation between closely related taxa of filamentous microorganisms

Table 3 Experimental screening of the candidate clusters unique to species or group of taxa.

| Targeted taxa | # tested targeted taxa | # tested non-targeted taxa | # candidate genes tested | # success |
|---|--|--|--------------------------|------------|
| <i>Phytophthora</i> | | | | |
| <i>P. ramorum</i> | 11 <i>P. ramorum</i> | 40 <i>Phytophthora</i> spp. | 28 | 5 (17.9%) |
| <i>P. lateralis</i> | 4 <i>P. lateralis</i> | 40 <i>Phytophthora</i> spp. | 16 | 6 (37.5%) |
| <i>P. kernoviae</i> | 1 <i>P. kernoviae</i> | 22 <i>Phytophthora</i> spp. | 12 | 9 (75.0%) |
| <i>P. ramorum</i> + <i>P. lateralis</i> | 11 <i>P. ramorum</i> , 4 <i>P. lateralis</i> | 39 <i>Phytophthora</i> spp. | 19 | 5 (26.3%) |
| Dothideomycetes | | | | |
| <i>Sphaerulina musiva</i> | 2 <i>S. musiva</i> | 14 <i>Mycosphaerella</i> spp. | 51 | 14 (27.5%) |
| <i>S. populicola</i> | 2 <i>S. populicola</i> | 14 <i>Mycosphaerella</i> spp. | 65 | 16 (24.6%) |
| <i>Phaeocryptopus gaeumannii</i> | 10 <i>P. gaeumannii</i> | 14 <i>Mycosphaerella</i> spp. | 10 | 3 (30%) |
| <i>S. musiva</i> + <i>S. populicola</i> | 2. <i>S. musiva</i> , 2 <i>S. populicola</i> | 12 <i>Mycosphaerella</i> spp. | 39 | 13 (33.3%) |
| <i>S. musiva</i> + <i>S. populicola</i> + <i>Mycosphaerella</i> sp. STON1 | 2. <i>S. musiva</i> , 2 <i>S. populicola</i> , 1 <i>Mycosphaerella</i> sp. STON1 | 11 <i>Mycosphaerella</i> spp. | 6 | 2 (33.3%) |
| Rusts | | | | |
| <i>Melampsora larici-populina</i> | 13 <i>M. larici-populina</i> | 15 <i>Melampsora</i> spp., 1 <i>Coleosporium</i> sp., 1 <i>Pucciniastrum</i> sp., 1 <i>Cronartium</i> sp., 2 <i>Chrysomyxa</i> spp. | 10 | 2 (20%) |
| <i>M. medusae</i> f. sp. <i>deltoidae</i> | 10 <i>M. medusae</i> | 15 <i>Melampsora</i> spp., 1 <i>Coleosporium</i> sp., 1 <i>Pucciniastrum</i> sp., 1 <i>Cronartium</i> sp., 2 <i>Chrysomyxa</i> spp. | 10 | 2 (20%) |
| <i>Cronartium ribicola</i> | 10 <i>C. ribicola</i> | 10 <i>Cronartium</i> spp., 5 <i>Melampsora</i> spp., 3 <i>Coleosporium</i> spp., 3 <i>Pucciniastrum</i> spp., 2 <i>Chrysomyxa</i> spp. | 20 | 3 (15%) |
| <i>Melampsora</i> genus | 19 <i>Melampsora</i> spp. | 2 <i>Coleosporium</i> spp., 3 <i>Pucciniastrum</i> spp., 3 <i>Cronartium</i> spp., 3 <i>Chrysomyxa</i> spp. | 5 | 3 (60%) |
| <i>Cronartium</i> genus | 11 <i>Cronartium</i> spp. | 8 <i>Melampsora</i> spp., 4 <i>Coleosporium</i> spp., 5 <i>Pucciniastrum</i> spp., 7 <i>Chrysomyxa</i> spp. | 8 | 2 (25%) |

with different lifestyles and ecology was crucial to identify genomic factors underlying these traits and gain insights into the mechanism of genome evolution (Dhillon *et al.*, 2015; Goodwin *et al.*, 2011; Grandaubert *et al.*, 2014; Plissonneau, Stürchler & Croll, 2016). We used a similar conceptual approach to identify taxon-specific genome regions and validated the hypothesis that these regions could serve as targets for the development of molecular assays for taxa detection.

Most PCR assays for pathogen detection use genes that are conserved to allow for primer design but comprise polymorphisms which can be targeted for design of discriminant

oligonucleotide probes or primers. However, there is a trade-off in the selection of genome regions between low intra-specific heterogeneity (to allow design of universal primers) and interspecific divergence (to allow taxa discrimination). The internal transcribed spacer region (ITS) of the nuclear ribosomal repeat unit is widely used for PCR assays as well as for taxa discrimination and in DNA barcoding and metabarcoding studies (Schoch *et al.*, 2012). It has the advantages of providing a large database of sequences and occurring in multiple copies in genomes which enable the development of very sensitive detection assays. However, the ITS can be limited for discrimination of phylogenetically related species and specificity is often an issue (Gazis, Rehner & Chaverri, 2011). In particular, ITS resolution for taxa separation in species-rich Ascomycota (e.g., genera *Cladosporium*, *Penicillium* and *Fusarium*) has often been inferior to several of the protein-coding genes commonly used in fungal taxonomy (Seifert, 2009; Feau *et al.*, 2011; Schoch *et al.*, 2012; Vialle *et al.*, 2009). Similarly, some of the taxa within the groups that were used in this study are closely related and not readily distinguishable in assay design using the ITS region. Only eight SNPs [98.0% similarity] are found between the ITS sequences of *S. musiva* and *S. populicola* and six SNPs [99.03% similarity] are present between the ITS sequences of *P. ramorum* and *P. lateralis* (Feau *et al.*, 2005; Werres *et al.*, 2001). In addition, sub-specific lineages with different epidemiological and biological characteristics often have nearly identical ITS sequences, complicating or preventing the design of reliable and robust taxon-specific assays. This is the case of the four lineages of *P. ramorum* that differ in mating types and aggressiveness yet share identical ITS sequences (Ivors *et al.*, 2004; Eyre *et al.*, 2014).

The method described herein provides benefits compared with assays conventionally designed on the basis of SNPs found in conserved genes among taxa. By focusing on taxon-specific genes for qPCR detection assays, we expected to lower type I errors, thereby maximizing specificity and reducing false positives. Despite its relative simplicity, our pipeline comprises several steps to limit both false positives and negatives. Filtering conducted in Module 3 greatly reduced the number of candidate genes by performing reciprocal blast searches across the genomes. This step was essential because of the variation in quality encountered in genome assembly and annotation. A second *in vitro* screening eliminated candidate genes that failed to amplify or discriminate close relatives. This two-step approach was efficient in quickly identifying potential unique and shared genes that are candidates for assay design and eliminating those that could yield false positives. It should be noted that the rate of conversion from *in silico* to *in vitro* validation steps (~28%) could be improved by fine-tuning the assay design and conducting additional *in silico* testing for intra-taxa variability using additional *de novo* assemblies in the targeted taxa. Moving or redesigning primers would make it possible to obtain amplicons for those candidate genes that failed the *in silico* validation step. However, since we obtained a large number of candidate genes for each taxonomic group targeted and our goal was to establish a proof-of-concept, we simply eliminated those that failed the *in vitro* screening step and we were still able to generate a large number of assays.

Another advantage of our approach is that primer design is less constrained since the genes targeted are unique, or at least they are not found in the genomes of phylogenetically related taxa. This could facilitate the design of multiplex qPCR assays that use internal

probes with different fluorochromes within a single reaction. Development of multiplex assays requires homogenization of annealing temperatures for the primers and probes targeting each amplicon. This design should be more efficient since the primers and probes in the unique targeted genes can be moved or their size can be customized without having the constraint of targeting discriminant SNPs. In addition, targeting unique genes reduces the likelihood of cross-interaction amongst the amplicons.

The conceptual and practical simplicity of qPCR, in combination with its speed and sensitivity, have made it the technology of choice in many diagnostic applications, including microbial quantification (Yu *et al.*, 2005; Narihiro & Sekiguchi, 2011; Thonar, Erb & Jansa, 2012) and pathogen detection (Orlofsky *et al.*, 2015). Multiple qPCR assays have become extremely powerful to detect fungi, bacteria and parasites (Kamau *et al.*, 2014; Gosiewski *et al.*, 2014; Bilodeau *et al.*, 2009) and we believe that this pipeline should enable development of qPCR assays in crop and tree pathogens. The pipeline described herein could be easily applied to any DNA-based detection methods (Yeo & Wong, 2002), such as loop-mediated isothermal amplification (LAMP), hybridization-based microarray (Huang *et al.*, 2006) or PCR-ELISA (Loeffler *et al.*, 1998). It would be particularly well-suited to target-enrichment methods by providing a large number of relevant genomic regions that can be used to enrich pathogen target genes in environmental samples. We expect that the impact of our approach and its efficiency in developing taxon-specific targets will improve quickly with the steady increase in the number of whole fungal and oomycete genome sequences (Fig. 2).

Given the wealth of genomic resources available and the increase in next-generation sequencing of microorganisms, this approach promises to be very useful to overcome some of the limitations of DNA detection assays based on conserved orthologous genes. Our work is a clear demonstration that a genome-wide approach can be useful to efficiently identify multiple, taxon-specific gene regions at different hierarchical levels that contain reliable priming site. We successfully discovered a large number of unique genes even between closely related species such as *P. ramorum* and *P. lateralis* (99.0% similarity in the ITS) and *S. musiva* and *S. populicola* (98.0% similarity). This approach is therefore likely to be broadly applicable to other fungi and Oomycetes. The possibility to easily obtain multiple alternative regions with equal or improved performance to the ITS region and other universal protein-coding genes commonly used for designing qPCR assays should be of great help in dealing with challenging cases for which higher taxonomic resolution is needed, such as for hybrids and races. Heteroploid organisms such as *Phytophthora xalni* (Aguayo *et al.*, 2016) represent more complicated cases in the development of qPCR diagnostic tools (e.g., Martin *et al.*, 2012). Homology to the putative parental species (Ioos *et al.*, 2005; Inderbitzin *et al.*, 2013) and deficiency of concerted evolution in homogenizing intra-individual copies (Lindner *et al.*, 2013) in the ITS region make this marker often unsuitable for assay design in such organisms. Multiple taxon-specific assays derived from whole genome sequences that target DNA inherited from the different lineages should help to obtain reliable identification. Such challenges should likely be resolved with the generation of good quality genomes based on long-read sequencing (Chin *et al.*, 2013).

CONCLUSIONS

We developed a pipeline that makes use of the increasing availability of whole genome sequences to identify unique taxon-discriminating genome regions as well as regions conserved across taxa that can be converted into PCR assays. We applied this approach to some of the most important plant pathogens and converted nearly one-third into qPCR assays. Our work demonstrates that a genome-wide approach can efficiently identify multiple taxon-specific genome regions that can be converted into highly specific DNA detection and identification assays.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was funded by Genome Canada, Genome British Columbia, the Canadian Forest Service (Genomics Research and Development Initiative, GRDI), Canadian Food Inspection Agency, FP Innovations and Boreal Genomics through a Large Scale Applied Research Project (LSARP 2112; Genome Canada) grant. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

Genome Canada.

Genome British Columbia.

Canadian Forest Service.

Genomics Research and Development Initiative, GRDI.

Canadian Food Inspection Agency.

Large Scale Applied Research Project (LSARP): 2112.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Nicolas Feau conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, approved the final draft and also wrote the paper.
- Stéphanie Beauseigle, Sandra Cervantes-Arango, Padmini Herath and Hesther Yueh performed the experiments, approved the final draft.
- Marie-Josée Bergeron, Monique L. Sakalidis and Clement K.M. Tsui performed the experiments, analyzed the data, authored or reviewed drafts of the paper, approved the final draft.
- Guillaume J. Bilodeau performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.

- Inanc Birol and Philippe Tanguay analyzed the data, contributed reagents/materials/-analysis tools, approved the final draft.
- Braham Dhillon and Angela L. Dale analyzed the data, authored or reviewed drafts of the paper, approved the final draft.
- Steven J.M. Jones analyzed the data, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Josyane Lamarche and Dario I. Ojeda performed the experiments, analyzed the data, approved the final draft.
- Greg Taylor performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, approved the final draft.
- Adnan Uzunovic contributed reagents/materials/analysis tools, approved the final draft.
- Richard C. Hamelin conceived and designed the experiments, approved the final draft and also wrote the paper.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences: Genbank Genome Assembly Accession numbers:

[GCA_002116355.1](#), [GCA_000504345.1](#), [GCA_000504365.1](#), [GCA_000504385.1](#), [GCA_002116395.1](#), [GCA_000504405.1](#), [GCA_002116385.1](#), [GCA_002157005.1](#), [GCA_000204055.1](#), [GCA_002157035.1](#), [GCA_002157085.1](#), [GCA_002157025.1](#), [GCA_002157015.1](#), [GCA_000464645.1](#), [GCA_000500795.1](#), [GCA_000464975.1](#), [GCA_000500245.1](#), [GCA_000149735.1](#), [GCA_000500205.2](#), [GCA_000468175.1](#), [GCA_000448265.2](#).

Data Availability

The following information was supplied regarding data availability:

The raw data used is already public and the pipeline only uses a series of already published software or scripts; these are described in the text and the figures.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.4392#supplemental-information>.

REFERENCES

- Abbas MM, Malluhi QM, Balakrishnan P. 2014. Assessment of *de novo* assemblers for draft genomes: a case study with fungal genomes. *BMC Genomics* 15:S10 DOI 10.1186/1471-2164-15-S9-S10.
- Aguayo J, Halkett F, Husson C, Nagy ZÁ, Szigethy A, Bakonyi J, Frey P, Marçais B. 2016. Genetic diversity and origins of the homoploid-type hybrid *Phytophthora ×alni*. *Applied and Environmental Microbiology* 82:7142–7153 DOI 10.1128/AEM.02221-16.
- Bilodeau G, Pelletier G, Pelletier F, Levesque CA, Hamelin RC. 2009. Multiplex real-time polymerase chain reaction (PCR) for detection of *Phytophthora ramorum*, the

- causal agent of sudden oak death. *Canadian Journal of Plant Pathology* **31**:195–210 DOI [10.1080/07060660909507593](https://doi.org/10.1080/07060660909507593).
- Blaecher C, Bauwens E, Tay A, Peters F, Dobbs S, Dobbs J, Charlett A, Ducatelle R, Haesebrouck F, Smet A. 2017.** A novel isolation protocol and probe-based RT-PCR for diagnosis of gastric infections with the zoonotic pathogen *Helicobacter suis*. *Helicobacter* **22**(3):e12369 DOI [10.1111/hel.12369](https://doi.org/10.1111/hel.12369).
- Brasier CM, Beales PA, Kirk SA, Denman S, Rose J. 2005.** *Phytophthora kernoviae* sp. nov., an invasive pathogen causing bleeding stem lesions on forest trees and foliar necrosis of ornamentals in the UK. *Mycological Research* **109**:853–859 DOI [10.1017/S0953756205003357](https://doi.org/10.1017/S0953756205003357).
- Brasier C, Webber J. 2010.** Plant pathology: sudden larch death. *Nature* **466**:824–825 DOI [10.1038/466824a](https://doi.org/10.1038/466824a).
- Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW. 2013.** Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods* **10**:563–569 DOI [10.1038/nmeth.2474](https://doi.org/10.1038/nmeth.2474).
- Desprez-Loustau ML, Aguayo J, Dutech C, Hayden KJ, Husson C, Jakushkin B, Marçais B, Piou D, Robin C, Vacher C. 2016.** An evolutionary ecology perspective to address forest pathology challenges of today and tomorrow. *Annals of Forest Science* **73**:45–67 DOI [10.1007/s13595-015-0487-4](https://doi.org/10.1007/s13595-015-0487-4).
- Dhillon B, Feau N, Aerts AL, Beauseigle S, Bernier L, Copeland A, Foster A, Gill N, Henrissat B, Herath P, LaButti KM, Levasseur A, Lindquist EA, Majoor E, Ohm RA, Pangilinan JL, Pribowo A, Saddler JN, Sakalidis ML, De Vries RP, Grigoriev IV, Goodwin SB, Tanguay P, Hamelin RC. 2015.** Horizontal gene transfer and gene dosage drives adaptation to wood colonization in a tree pathogen. *Proceedings of the National Academy of Sciences of the United States of America* **122**:3451–3456 DOI [10.1073/pnas.1424293112](https://doi.org/10.1073/pnas.1424293112).
- Duplessis S, Cuomo C, Lin Y-C, Aerts A, Tisserant E, Veneault-Fourrey C, Joly DL, Hacquard S, Amselem J, Cantarel BL, Chiu R, Chiu R, Coutinho PM, Feau N, Field M, Frey P, Gelhaye E, Goldberg J, Grabherr MG, Kodira CD, Kohler A, Kües U, Lindquist EA, Lucas SM, Mago R, Mauceli E, Morin E, Murat C, Pangilinan JL, Park R, Pearson M, Quesneville H, Rouhier N, Sakthikumar S, Salamov AA, Schmutz J, Selles B, Shapiro H, Tanguay P, Tuskan GA, Henrissat B, Van de Peer Y, Rouzé P, Ellis JG, Dodds PN, Schein JE, Zhong S, Hamelin RC, Grigoriev IV, Szabo LJ, Martin F. 2011.** Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences of the United States of America* **108**:9166–9171 DOI [10.1073/pnas.1019315108](https://doi.org/10.1073/pnas.1019315108).
- Edgar RC. 2004.** MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**:113.
- Eyre CA, Hayden KJ, Kozanitas M, Grunwald NJ, Garbelotto M. 2014.** Lineage, temperature, and host species have interacting effects on lesion development in *Phytophthora ramorum*. *Plant Disease* **98**:1717–1727 DOI [10.1094/PDIS-02-14-0151-RE](https://doi.org/10.1094/PDIS-02-14-0151-RE).

- Faino L, Thomma BPHJ. 2014. Get your high-quality low-cost genome sequence. *Trends in Plant Science* **19**:288–291 DOI [10.1016/j.tplants.2014.02.003](https://doi.org/10.1016/j.tplants.2014.02.003).
- Farrington HL, Edwards CE, Guan X, Carr MR, Baerwaldt K, Lance RF. 2015. Mitochondrial genome sequencing and development of genetic markers for the detection of DNA of invasive bighead and silver carp (*Hypophthalmichthys nobilis* and *H. molitrix*) in environmental water samples from the United States. *PLOS ONE* **10**:e0117803 DOI [10.1371/journal.pone.0117803](https://doi.org/10.1371/journal.pone.0117803).
- Feau N, Decourcelle T, Husson C, Desprez-Loustau M-L, Dutech C. 2011. Finding single copy genes out of sequenced genomes for multilocus phylogenetics in non-model fungi. *PLOS ONE* **6**(4):e18803.
- Feau N, Hamelin RC, Bernier L. 2006. Attributes and congruence of three molecular data sets: inferring phylogenies among *Septoria*-related species from woody perennial plants. *Molecular Phylogenetics and Evolution* **40**:808–829 DOI [10.1016/j.ympev.2006.03.029](https://doi.org/10.1016/j.ympev.2006.03.029).
- Feau N, Mottet MJ, Périnet P, Hamelin RC, Bernier L. 2010. Recent advances related to poplar leaf spot and canker caused by *Septoria musiva*. *Canadian Journal of Plant Pathology* **32**:122–134 DOI [10.1080/07060661003740009](https://doi.org/10.1080/07060661003740009).
- Feau N, Taylor G, Dale AL, Dhillon B, Bilodeau GJ, Birol I, Jones SJM, Hamelin RC. 2016. Genome sequences of six *Phytophthora* species threatening forest ecosystems. *Genomics Data* **10**:85–88 DOI [10.1016/j.gdata.2016.09.013](https://doi.org/10.1016/j.gdata.2016.09.013).
- Feau N, Weiland JE, Stanosz GR, Bernier L. 2005. Specific and sensitive PCR-based detection of *Septoria musiva*, *S. populicola* and *S. populi* the causes of leaf spot and stem canker on poplars. *Mycological Research* **109**:1015–1028 DOI [10.1017/S0953756205003242](https://doi.org/10.1017/S0953756205003242).
- Firth C, Lipkin WI. 2013. The genomics of emerging Pathogens. *Annual Review of Genomics and Human Genetics* **4**:281–300 DOI [10.1146/annurev-genom-091212-153446](https://doi.org/10.1146/annurev-genom-091212-153446).
- Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ. 2012. Emerging fungal threats to animal, plant and ecosystem health. *Nature* **484**:186–194 DOI [10.1038/nature10947](https://doi.org/10.1038/nature10947).
- Gallo JE, Muñoz JF, Misas E, McEwen JG, Clay OK. 2014. The complex task of choosing a *de novo* assembly: lessons from fungal genomes. *Computational Biology and Chemistry* **53**:97–107 DOI [10.1016/j.compbiolchem.2014.08.014](https://doi.org/10.1016/j.compbiolchem.2014.08.014).
- Gazis R, Rehner S, Chaverri P. 2011. Species delimitation in fungal endophyte diversity studies and its implications in ecological and biogeographic inferences. *Molecular Ecology* **20**:3001–3013 DOI [10.1111/j.1365-294X.2011.05110.x](https://doi.org/10.1111/j.1365-294X.2011.05110.x).
- Goodwin SB, Ben M'Barek S, Dhillon B, Wittenberg AHJ, Crane CF, Hane JK, Foster AJ, Van der Lee TAJ, Grimwood J, Aerts A, Antoniw J, Bailey A, Bluhm B, Bowler J, Bristow J, Van der Burgt A, Canto-Canché B, Churchill ACL, Conde-Ferrández L, Cools HJ, Coutinho PM, Csukai M, Dehal P, De Wit P, Donzelli B, Van de Geest HC, Van Ham RCH, Hammond-Kosack KE, Henrissat B, Kilian A, Kobayashi AK, Koopmann E, Kourmpetis Y, Kuzniar A, Lindquist E, Lombard V, Maliepaard C, Martins N, Mehrabi R, Nap JPH, Ponomarenko A, Rudd JJ, Salamov A, Schmutz J,

- Schouten HJ, Shapiro H, Stergiopoulos I, Torriani SFF, Tu H, De Vries RP, Waalwijk C, Ware SB, Wiebenga A, Zwiers L-H, Oliver RP, Grigoriev IV, Kema GHJ. 2011. Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Genetics* 7(6):e1002070 DOI 10.1371/journal.pgen.1002070.
- Gosiewski T, Jurkiewicz-Badacz D, Sroka A, Brzychczy-Włoch M, Bulanda M. 2014. A novel, nested, multiplex, real-time PCR for detection of bacteria and fungi in blood. *BMC Microbiology* 14:144 DOI 10.1186/1471-2180-14-144.
- Grandaubert J, Lowe RGT, Soyer JL, Schoch CL, De Wouw AP, Fudal I, Robbertse B, Lapalu N, Links MG, Ollivier B, Linglin J, Barbe V, Mangenot S, Cruaud C, Borhan H, Howlett BJ, Balesdent M-H, Rouxel T. 2014. Transposable element-assisted evolution and adaptation to host plant within the *Leptosphaeria maculans*-*Leptosphaeria biglobosa* species complex of fungal pathogens. *BMC Genomics* 15:891 DOI 10.1186/1471-2164-15-891.
- Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otilar R, Riley R, Salamov A, Zhao X, Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I. 2014. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research* 42(Database issue):D699–D704.
- Haas BJ, Zeng Q, Matthew D, Pearson MD, Cuomo CA, Wortman JR. 2011. Approaches to fungal genome annotation. *Mycology* 2:118–141.
- Hänsel C, Mertens K, Elschner MC, Melzer F. 2015. Novel real-time PCR detection assay for *Brucella suis*. *Veterinary Record Open* 2:e000084 DOI 10.1136/vetreco-2014-000084.
- Hansen EM, Goheen DJ, Jules ES, Ullian B. 2000. Managing Port-Orford-cedar and the introduced pathogen *Phytophthora lateralis*. *Plant Disease* 84:4–14 DOI 10.1094/PDIS.2000.84.1.4.
- Ho CC, Lau CC, Martelli P, Chan SY, Cindy WS, Wu AK, Yuen KY, Lau SK, Woo PC. 2011. Novel pan-genomic analysis approach in target selection for multiplex PCR identification and detection of *Burkholderia pseudomallei*, *Burkholderia thailandensis*, and *Burkholderia cepacia* complex species: a proof-of-concept study. *Journal of Clinical Microbiology* 49:814–821 DOI 10.1128/JCM.01702-10.
- Ho CC, Wu AK, Cindy WS, Yuen KY, Lau SK, Woo PC. 2012. Automated pangenomic analysis in target selection for PCR detection and identification of bacteria by use of ssGeneFinder webserver and its application to *Salmonella enterica* serovar typhi. *Journal of Clinical Microbiology* 50:1905–1911 DOI 10.1128/JCM.06843-11.
- Huang A, Li JW, Shen ZQ, Wang XW, Jin M. 2006. High-throughput identification of clinical pathogenic fungi by hybridization to an oligonucleotide microarray. *Journal of Clinical Microbiology* 44:3299–3305 DOI 10.1128/JCM.00417-06.
- Hung G-C, Nagamine K, Li B, Lo S-C. 2012. Identification of DNA signatures suitable for use in development of real-time PCR assays by whole-genome sequence approaches: use of *Streptococcus pyogenes* in a pilot study. *Journal of Clinical Microbiology* 50(8):2770–2773.

- Inderbitzin P, Davis RM, Bostock RM, Subbarao KV. 2013.** Identification and differentiation of *Verticillium* species and *V. longisporum* lineages by simplex and multiplex PCR assays. *PLOS ONE* **8**:e65990 DOI [10.1371/journal.pone.0065990](https://doi.org/10.1371/journal.pone.0065990).
- Ioos R, Husson C, Andrieux A, Frey P. 2005.** SCAR-based PCR primers to detect the hybrid pathogen *Phytophthora alni* and its subspecies causing alder disease in Europe. *European Journal of Plant Pathology* **112**:323–335 DOI [10.1007/s10658-005-6233-2](https://doi.org/10.1007/s10658-005-6233-2).
- Ivors KL, Hayden KJ, Bonants PJM, Rizzo DM, Garbelotto M. 2004.** AFLP and phylogenetic analyses of North American and European populations of *Phytophthora ramorum*. *Mycological Research* **108**:378–392 DOI [10.1017/S0953756204009827](https://doi.org/10.1017/S0953756204009827).
- Kamau E, Alemayehu S, Feghali KC, Juma DW, Blackstone GM, Marion WR, Obare P, Ogutu B, Ockenhouse CF. 2014.** Sample-ready multiplex qPCR assay for detection of malaria. *Malaria Journal* **13**:Article 158 DOI [10.1186/1475-2875-13-158](https://doi.org/10.1186/1475-2875-13-158).
- Lamarche J, Potvin A, Pelletier G, Stewart D, Feau N, Alayon DIO, Dale AL, Coelho A, Uzunovic A, Bilodeau GJ, Brière SC, Hamelin RC, Tanguay P. 2015.** Molecular detection of 10 of the most unwanted alien forest pathogens in Canada using real-time PCR. *PLOS ONE* **10**:e0134265 DOI [10.1371/journal.pone.0134265](https://doi.org/10.1371/journal.pone.0134265).
- Li L, Stoeckert CJ, Roos DS. 2003.** OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* **13**:2178–2189 DOI [10.1101/gr.1224503](https://doi.org/10.1101/gr.1224503).
- Lin H, Moghe G, Ouyang S, Iezzoni A, Shiu S-H, Gu X, Buell CR. 2010.** Comparative analyses reveal distinct sets of lineage-specific genes within *Arabidopsis thaliana*. *BMC Evolutionary Biology* **10**:41 DOI [10.1186/1471-2148-10-41](https://doi.org/10.1186/1471-2148-10-41).
- Lindner DL, Carlsen T, Henrik Nilsson R, Davey M, Schumacher T, Kauserud H. 2013.** Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi. *Ecology and Evolution* **3**:1751–1764 DOI [10.1002/ece3.586](https://doi.org/10.1002/ece3.586).
- Loeffler J, Hebart H, Sepe S, Schumacher U, Klingebiel T, Einsele H. 1998.** Detection of PCR-amplified fungal DNA by using a PCR-ELISA system. *Medical Mycology* **36**:275–279.
- Ma L-J, Fedorova ND. 2010.** A practical guide to fungal genome projects: strategy, technology, cost and completion. *Mycology* **1**:924 DOI [10.1080/21501201003680943](https://doi.org/10.1080/21501201003680943).
- Martin FN, Abad ZG, Balci Y, Ivors K. 2012.** Identification and detection of *Phytophthora*: reviewing our progress, identifying our needs. *Plant Disease* **96**:1080–1103 DOI [10.1094/PDIS-12-11-1036-FE](https://doi.org/10.1094/PDIS-12-11-1036-FE).
- Martin FN, Blair JE, Coffey MD. 2014.** A combined mitochondrial and nuclear multi-locus phylogeny of the genus *Phytophthora*. *Fungal Genetics and Biology* **66**:19–32 DOI [10.1016/j.fgb.2014.02.006](https://doi.org/10.1016/j.fgb.2014.02.006).
- Narihiro T, Sekiguchi Y. 2011.** Oligonucleotide primers, probes and molecular methods for the environmental monitoring of methanogenic archaea. *Microbiology and Biotechnology* **4**:585–604 DOI [10.1111/j.1751-7915.2010.00239.x](https://doi.org/10.1111/j.1751-7915.2010.00239.x).
- Orlofsky E, Benami M, Gross A, Dutt M, Gillor O. 2015.** Rapid MPN-Qpcr screening for pathogens in air, soil, water, and agricultural produce. *Water, Air, Soil Pollution* **226**:Article 303 DOI [10.1007/s11270-015-2560-x](https://doi.org/10.1007/s11270-015-2560-x).

- Pautasso M, Schlegel M, Holdenrieder O. 2015.** Forest health in a changing world. *Microbial Ecology* 69:826–842 DOI 10.1007/s00248-014-0545-8.
- Plissonneau C, Stürchler A, Croll D. 2016.** The evolution of orphan regions in genomes of a fungal pathogen of wheat. *mBio* 7(5):e01231-16 DOI 10.1128/mBio.01231-16.
- Prévost-Bouré NC, Christen R, Dequiedt S, Mougél C, Lelièvre M, Jolivet C, Shahbazkia HR, Guillou L, Arrouays D, Ranjard L. 2011.** Validation and application of a PCR primer set to quantify fungal communities in the soil environment by real-time quantitative PCR. *PLOS ONE* 6:e24166 DOI 10.1371/journal.pone.0024166.
- Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. 2016.** Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Analytical Methods* 8(1):12–24 DOI 10.1039/C5AY02550H.
- Rozen S, Skaletsky H. 1999.** Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics Methods and Protocols* 132:365–386.
- Santini A, Ghelardini L, De Pace C, Desprez-Loustau ML, Capretti P, Chandelier A, Cech T, Chira D, Diamandis S, Gaitniekis T, Hantula J, Holdenrieder O, Jankovsky L, Jung T, Jurc D, Kirisits T, Kunca A, Lygis V, Malecka M, Marcais B, Schmitz S, Schumacher J, Solheim H, Solla A, Szabó I, Tsopelas P, Vannini A, Vettrano AM, Webber J, Woodward S, Stenlid J. 2013.** Biogeographical patterns and determinants of invasion by forest pathogens in Europe. *New Phytologist* 197:238–250 DOI 10.1111/j.1469-8137.2012.04364.x.
- Schena L, Li Destri Nicosia MG, Sanzani SM, Faedda R, Ippolito A, Cacciola SO. 2013.** Development of quantitative PCR detection methods for phytopathogenic fungi and oomycetes. *Journal of Plant Pathology* 1:7–24 DOI 10.4454/JPP.V95I1.016.
- Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Fungal Barcoding Consortium, Fungal Barcoding Consortium Author List. 2012.** Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America* 109:6241–6246 DOI 10.1073/pnas.1117018109.
- Seifert KA. 2009.** Progress towards DNA barcoding of fungi. *Molecular Ecology Resources* 9:83–89 DOI 10.1111/j.1755-0998.2009.02635.x.
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. 2012.** ABySS: a parallel assembler for short read sequence data structures assemblies and assembly algorithms. *Genome Research* 19:1117–1123.
- Stanke M, Schoffmann O, Morgenstern B, Waack S. 2006.** Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7:62 DOI 10.1186/1471-2105-7-62.
- Szilagyi SM, Szilagyi L. 2014.** A fast hierarchical clustering algorithm for large-scale protein sequence data sets. *Computational Biology and Medicine* 48:94–101 DOI 10.1016/j.combiomed.2014.02.016.
- Thonar C, Erb A, Jansa J. 2012.** Real-time PCR to quantify composition of arbuscular mycorrhizal fungal communities—marker design, verification, calibration and field validation. *Molecular Ecology Resources* 12:219–232 DOI 10.1111/j.1755-0998.2011.03086.x.

- Vialle A, Feau N, Allaire M, Didukh M, Martin F, Moncalvo J-M, Hamelin RC. 2009.** Evaluation of mitochondrial genes as DNA barcode for Basidiomycota. *Molecular Ecology Resources* **9**:99–113 DOI [10.1111/j.1755-0998.2009.02637.x](https://doi.org/10.1111/j.1755-0998.2009.02637.x).
- Werres S, Marwitz R, Man In't veld WA, De Cock AWAM, Bonants PJM, De Weerd M, Themann K, Ilieva E, Baayen RP. 2001.** Phytophthora ramorum sp. nov., a new pathogen on Rhododendron and Viburnum. *Mycological Research* **105**:1155–1165 DOI [10.1016/S0953-7562\(08\)61986-3](https://doi.org/10.1016/S0953-7562(08)61986-3).
- Xu Y, Wu G, Hao B, Chen L, Deng X, Xu Q. 2015.** Identification, characterization and expression analysis of lineage-specific genes within sweet orange (*Citrus sinensis*). *BMC Genomics* **16**:995 DOI [10.1186/s12864-015-2211-z](https://doi.org/10.1186/s12864-015-2211-z).
- Yang L, Zou M, Fu B, He S. 2013.** Genome-wide identification, characterization, and expression analysis of lineage-specific genes within zebrafish. *BMC Genomics* **14**:65 DOI [10.1186/1471-2164-14-65](https://doi.org/10.1186/1471-2164-14-65).
- Yeo SF, Wong B. 2002.** Current status of nonculture methods for diagnosis of invasive fungal infections. *Clinical Microbiology Reviews* **15**:465–484 DOI [10.1128/CMR.15.3.465-484.2002](https://doi.org/10.1128/CMR.15.3.465-484.2002).
- Yu Y, Lee C, Kim J, Hwang S. 2005.** Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. *Biotechnology and Bioengineering* **89**:670–679 DOI [10.1002/bit.20347](https://doi.org/10.1002/bit.20347).