



Microbe-ID: an open source toolbox for microbial genotyping and species identification

Javier F. Tabima¹, Sydney E. Everhart^{1,5}, Meredith M. Larsen²,
Alexandra J. Weisberg¹, Zhian N. Kamvar¹, Matthew A. Tancos³,
Christine D. Smart³, Jeff H. Chang^{1,4} and Niklaus J. Grünwald^{1,2,3,4}

¹ Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, United States

² Horticultural Crops Research Laboratory, USDA Agricultural Research Service, Corvallis, OR, United States

³ Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, Geneva, NY, United States

⁴ Molecular and Cellular Biology Graduate Program and Center for Genome Biology and Biocomputing, Oregon State University, Corvallis, OR, United States

⁵ Current affiliation: Department of Plant Pathology, University of Nebraska, Lincoln, NE, United States

ABSTRACT

Development of tools to identify species, genotypes, or novel strains of invasive organisms is critical for monitoring emergence and implementing rapid response measures. Molecular markers, although critical to identifying species or genotypes, require bioinformatic tools for analysis. However, user-friendly analytical tools for fast identification are not readily available. To address this need, we created a web-based set of applications called Microbe-ID that allow for customizing a toolbox for rapid species identification and strain genotyping using any genetic markers of choice. Two components of Microbe-ID, named Sequence-ID and Genotype-ID, implement species and genotype identification, respectively. Sequence-ID allows identification of species by using BLAST to query sequences for any locus of interest against a custom reference sequence database. Genotype-ID allows placement of an unknown multilocus marker in either a minimum spanning network or dendrogram with bootstrap support from a user-created reference database. Microbe-ID can be used for identification of any organism based on nucleotide sequences or any molecular marker type and several examples are provided. We created a public website for demonstration purposes called Microbe-ID (microbe-id.org) and provided a working implementation for the genus *Phytophthora* (phytophthora-id.org). In *Phytophthora*-ID, the Sequence-ID application allows identification based on ITS or *cox* spacer sequences. Genotype-ID groups individuals into clonal lineages based on simple sequence repeat (SSR) markers for the two invasive plant pathogen species *P. infestans* and *P. ramorum*. All code is open source and available on github and CRAN. Instructions for installation and use are provided at <https://github.com/grunwaldlab/Microbe-ID>.

Submitted 29 April 2016
Accepted 2 July 2016
Published 18 August 2016

Corresponding author
Niklaus J. Grünwald,
nik.grunwald@ars.usda.gov,
grunwaldn@science.oregonstate.edu

Academic editor
Blanca Landa

Additional Information and
Declarations can be found on
page 10

DOI 10.7717/peerj.2279

© Copyright
2016 Tabima et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Agricultural Science, Bioinformatics, Microbiology, Mycology, Taxonomy

Keywords Taxonomy, Identification, Molecular diagnostics, *Phytophthora*, Genotyping, Pathogen

BACKGROUND

Development of tools for identification of species, genotypes or strains is critical for monitoring emergence of invasive organisms such as *Phytophthora ramorum* causing

sudden oak death ([Grünwald, Goss & Press, 2008](#)), *Hymenoscyphus fraxineus* causing ash dieback ([Gross et al., 2014](#)), *Aphanomyces astaci* causing crayfish plague ([Holdich et al., 2009](#)), *Cryptococcus gattii* causing cryptococcosis and meningitis ([Byrnes III et al., 2010](#)), or Methicillin-resistant *Staphylococcus aureus* causing invasive MRSA disease ([Klevens et al., 2007](#)). Molecular markers provide a rapid means for identification, but require various bioinformatics tools for identification of species and/or novel genotypes. In eukaryotes, sequences from the rRNA internal transcribed spacer (ITS) region and various mitochondrial DNA regions are used to separate discrete species ([Coleman, 2003](#); [Coleman, 2007](#)). ITS and mtDNA markers are now the most widely used markers in plants ([Coleman, 2007](#); [Coleman, 2009](#)), fungi ([James et al., 2006](#)), corals ([Grajales, Aguilar & Sánchez, 2007](#)), and oomycetes ([Cooke et al., 2012](#); [Robideau et al., 2011](#)) and have been coined “DNA barcodes” because of their broad ability to distinguish species ([Schoch et al., 2012](#)). Classification of individuals using various molecular markers has recently increased. Multi-locus sequence types (MLST) are being widely used by researchers working with bacterial taxa to reveal the identity of samples by classification relative to known reference strains ([Maiden et al., 2013](#)). Other molecular markers or methods used to distinguish genotypes might include microsatellites (or simple sequence repeats) to identify strains and clonal lineages ([Cooke et al., 2012](#); [Ivors et al., 2006](#)), DNA sequences for specific genic regions ([Maiden et al., 2013](#)), single nucleotide polymorphism (SNP) genotyping using reduced representation approaches ([Grünwald, McDonald & Milgroom, in press](#)) such as RAD-seq ([Etter et al., 2010](#)) or genotyping by sequencing (GBS, [Elshire et al., 2011](#)), or genome wide SNP genotyping ([Huang et al., 2009](#)).

In addition to the molecular methods developed, different types of online databases have been implemented to identify species using these molecular methods within groups of organisms. Examples of these databases are FungiDB for fungi and fungal-like organisms ([Stajich et al., 2011](#)), EuPathDB for eukaryotic organisms ([Aurrecoechea et al., 2013](#)), and the *Phytophthora* database which allows entries by experts in the *Phytophthora* community from different labs or countries for different species of the genus ([Park et al., 2008](#)). We previously reported on our development of a database for *Phytophthora* species and genotype identification using web tools to identify species using common barcodes, enabling the conjunction of modern laboratory techniques with highly curated databases for species identification ([Grünwald et al., 2011](#)).

Our objective here was to report the development of a toolbox for microbe identification (Microbe-ID) that can readily be customized for sequence based species identification (Sequence-ID) or molecular marker-based identification of genotypes (Genotype-ID) for any group of organisms. Our objectives were two-fold: (1) to implement Microbe-ID as a demonstration site that is customizable for any group of organisms and (2) to demonstrate a working implementation at *Phytophthora*-ID.org version 2.0 with significant updates from version 1.0 ([Grünwald et al., 2011](#)). Microbe-ID includes two modules, Sequence-ID and Microbe-ID. Sequence analysis is implemented based on use of a well characterized barcode region for the genus *Phytophthora*, but can be implemented to use any barcode sequence of interest. Genotype analysis can be implemented to use a variety of marker data types. To demonstrate the breadth of the developed tools and applicability to the

diversity of microorganisms, the following three examples were included: codominant microsatellite data (SSR/Microsatellite) for the oomycete *P. ramorum* (Grünwald *et al.*, 2009), concatenated Multi Locus Sequence Type (MLST) or individual locus sequences for the bacterium *Clavibacter michiganensis* subsp. *michiganensis* (Tancos, Lange & Smart, 2015, Fig. S3), and dominant Amplified Fragment Length Polymorphism (Binary (AFLP) data, Fig. S4) for the oomycete *Aphanomyces euteiches* (Grünwald & Hoheisel, 2006). Moreover, Genotype-ID can be expanded to include other marker systems including gene sequences for resistance to antibiotics or fungicides as well as presence/absence polymorphisms for effector genes or other adaptive loci. Two sequence databases were developed that help us demonstrate the utility of Microbe-ID. These databases are for the genus *Phytophthora* (*Phytophthora*-ID) with two sequence databases containing over 110 species that are mostly plant pathogens (Kroon *et al.*, 2012), and two genotyping databases for populations of the potato late blight pathogen, *P. infestans*, and the sudden oak death pathogen, *P. ramorum* (Genotype-ID) (Grünwald, Goss & Press, 2008; Kamoun *et al.*, 2015). All of these tools are readily customizable and open source (<http://www.github.com/grunwaldlab/microbe-ID>), provided as a demonstration site (<http://microbe-id.org/>), and a working implementation of Sequence-ID and Genotype-ID that we use in our own work for the genus *Phytophthora* (phytophthora-id.org). Finally, a companion paper describes application of Microbe-ID for the implementation of a new website, Gall-ID, with novel tools for identification of gall-forming bacteria (Davis II *et al.*, 2016).

The Microbe-ID toolbox

We developed a template website named Microbe-ID (<http://microbe-id.org/>) with two separate modules, Sequence-ID and Genotype-ID, for sequence-based species identification and genotyping, respectively. The website is written using bootstrap (<http://getbootstrap.com/>), a HTML, CSS, and JS framework for developing responsive, mobile first projects on the web. Specific instructions, code, and resources necessary for implementation of Microbe-ID are provided on the GitHub repository (<http://www.github.com/grunwaldlab/microbe-ID>). The server currently hosting Microbe-ID is running Centos Linux release 6.6, NCBI BLAST 2.2.28+ (Altschul *et al.*, 1990), MAFFT version 7.221 (Katoh *et al.*, 2002), and R version 3.1.2. Below we describe specific tools required for customization of each module.

Navigating and using Microbe-ID

Navigating the demonstration site of Microbe-ID (<http://Microbe-ID.org>) the user will encounter a menu bar with links to Sequence-ID, Genotype-ID, and an about page. The home page has a general description of the functionality and components implemented in Microbe-ID, as well as links to the github site. Since input from the user is entered into forms as a query, each form implemented in Microbe-ID is encoded to check that the format of the data supplied by the user is supported. If the format is incorrect, the page will prompt the user with an error message, making the use of Microbe-ID more user-friendly.

Sequence-ID

We created a module called Sequence-ID that uses BLAST analysis of common sequence loci for species identification (Fig. S1). Sequence-ID includes a PERL_CGI script that permits the communication between a user interface form (implemented in HTML5) and a BLAST database for the desired sequence data. The form recognizes the input in FASTA format and uses BLASTN to search for the most similar DNA sequences in the marker database. The PERL-CGI script receives the information of the HTML tabulated output from BLAST and displays a table of hits to the end user. Sequence-ID is customizable as the FASTA data file can readily be updated using the makeblastdb program of the BLAST suite. It can also similarly be implemented for BLASTP analysis of amino acid sequences.

In the Sequence-ID webpage, the user will find two main tabs: A “Blast” tab in which the web-app is contained, and a “Help” tab which contains a link to a “Site Help” prompt. This “Site Help” prompt shows an example of the FASTA sequence format recognized by Sequence-ID. The user can copy the FASTA sequence and paste it in the “Blast” tab text form to perform a BLAST search on the example query, or can download the database in FASTA format. After the BLAST is performed, the web page will provide a table including the BLAST search results for the query of interest using the BLAST alignment format.

Genotype-ID

Genotype-ID is a web application designed to be user-friendly, and to facilitate the interaction with R using the set of tools listed in Table 1 (Fig. 1). To develop the web application, we used the R package ‘shiny’ (RStudio Team, 2015). Shiny facilitates the interactions between user, server, and R (Fig. 1). The shiny web framework relies on reactive programming, which allows dynamic deployment of traditional R scripts in response to data input through a website console whereby results generated in R are subsequently pushed to the end user. Thus, ordinary R packages, which otherwise require familiarity with the language, can be deployed behind a user-friendly interface. Genotype-ID interacts principally with the R packages ‘poppr’ (Kamvar, Tabima & Grünwald, 2014), ‘adegenet’ (Jombart, 2008), ‘ape’ (Paradis, Claude & Strimmer, 2004), and ‘pegas’ (Paradis, 2010) (Table 1).

Genotype-ID has three different modules, which are shown as tabs on the website, each specific to different molecular markers: SSR/microsatellite data (SSR/Microsatellite data), multilocus genotype sequence types (MLST data), and AFLP, RFLP, SNP and other binary datasets (Binary (AFLP) data). The user supplies a query via the web framework that is read into R and queried against a curated dataset. Genetic distances (Table 2) are calculated between the query and genotypes in the reference database, with relationship of the query to the database presented either as a UPGMA or neighbor-joining dendrogram with bootstrap support or a minimum spanning network. These methods reconstruct relationships of the query relative to the genotypes found in the curated database. Each module has its own customization scheme to analyze each particular molecular marker type. The SSR data tool calculates Bruvo’s distance (Bruvo et al., 2004) to reconstruct the UPGMA or NJ dendrogram and minimum spanning networks. The MLST data tool permits the comparison of user-submitted gene sequences in FASTA format. The MLST data tool uses MAFFT (Katoh et al., 2002) to align each query sequence to the corresponding curated gene

Table 1 Open source computational tools required to install and deploy Microbe-ID on a server.

Tool	Description	Source	References
<i>Ape</i>	R package for phylogenetic and evolutionary analysis	http://ape-package.ird.fr/	<i>Paradis, Claude & Strimmer (2004)</i>
<i>BLAST</i>	Basic Local Alignment and Search Tool implemented as an algorithm for comparing DNA, RNA or protein query sequences against a reference database	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/	<i>Altschul et al. (1990)</i>
<i>Adegenet</i>	R package for multivariate analysis of genetic data	https://github.com/thibautjombart/adegenet/	<i>Jombart (2008)</i>
<i>Bootstrap</i>	A framework for developing responsive, mobile-first projects on the web	http://www.getbootstrap.com/	–
<i>Microbe-ID</i>	Set of web-apps for identification of species, genotypes, and strains of any organism	https://github.com/grunwaldlab/Microbe-ID	This paper.
<i>MAFFT</i>	Multiple sequence alignment algorithm to find homology between sequences using Fourier algorithms	http://mafft.cbrc.jp/alignment/software/	<i>Katoh et al. (2002)</i>
<i>Pegas</i>	R package for analysis of population genetic data	http://ape-package.ird.fr/pegas.html	<i>Paradis (2010)</i>
<i>Poppr</i>	R package for genetic analysis of populations with mixed reproduction	https://github.com/grunwaldlab/poppr	<i>Kamvar, Tabima & Grünwald (2014)</i> and <i>Kamvar, Brooks & Grünwald (2015)</i>
<i>Shiny</i>	Interactive web application framework for R	http://shiny.rstudio.com/	–

Notes.

–, reference not available; see website provided for information.

database. It then concatenates all separate alignments, calculates the genetic distance for the alignment, and reconstructs distance dendrograms and minimum spanning networks. Lastly, the binary tool uses molecular markers such as AFLPs or RFLPs to reconstruct relationships. This tool uses binary data (coded as 1 and 0) and different genetic distances (Table 2) to reconstruct the UPGMA or NJ dendrogram and the minimum spanning network.

The Microbe-ID implementation

Microbe-ID contains implementations of each of the modules for different custom databases: the SSR implementation uses nine diploid SSR loci for the oomycete species *P. ramorum* (Fig. S2); the MLST implementation uses eight multilocus sequence types for the bacterial species *Clavibacter michiganensis* subsp. *michiganensis*; and a binary implementation that uses 56 loci for AFLP data for the oomycete *Aphanomyces euteiches*. Each of the implementations of Genotype-ID contains collapsible instructions on how to format the query and a link to download a tabulated file with example queries formatted for use in Genotype-ID. The user can download the spreadsheet in order to edit, copy, and paste their custom queries into the “Data input” form. To make analyses that use a random seed repeatable, the user has the option to specify a seed number. In MLST-ID and Binary-ID, the user can select the genetic distance to be used in the analysis. After

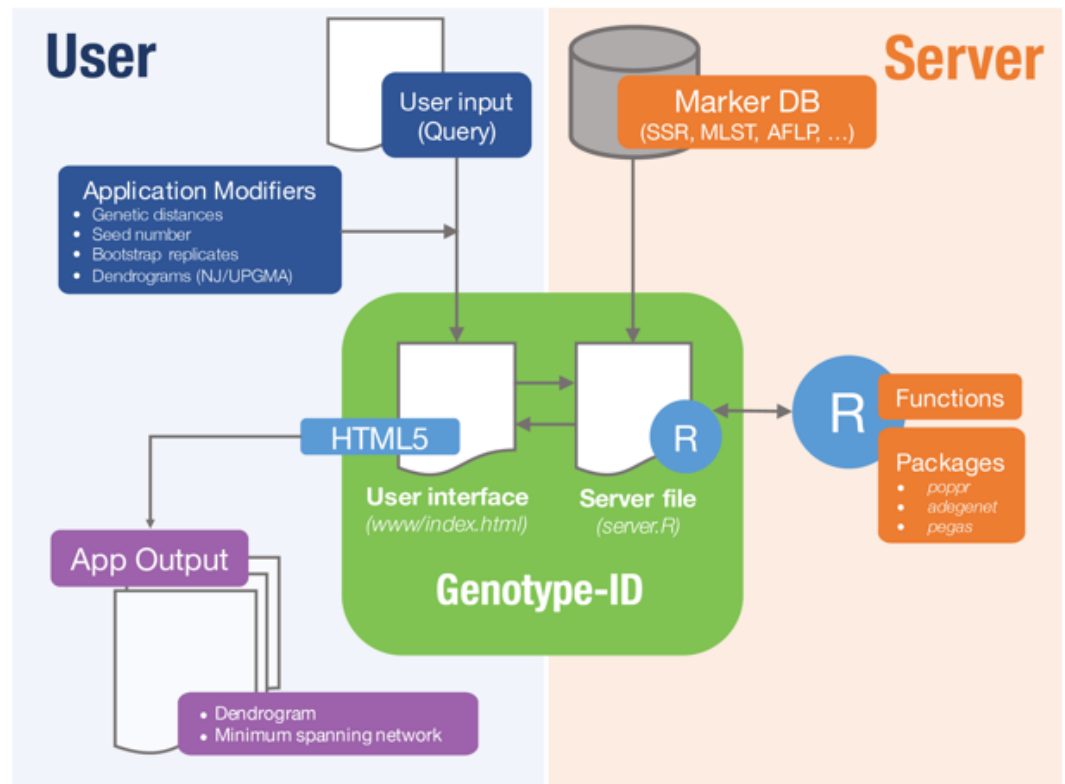


Figure 1 Diagram representing implementation of Genotype-ID, which is comprised of a user interface file (index.html) and a server file (server.R). Each file communicates with the R framework (via shiny) and user (via HTML5). On the user side (left side), user input is provided by copy/paste of a query and selects/specifies the desired application modifiers (seed number, genetic distance calculation). This information is subsequently received and processed by the server file, prompting the application to run in R. On the server side (right side) a database file (Marker DB), R packages, and functions are retrieved and executed. When the run is complete, the server file provides output to the user interface file and displayed on the app output.

the user inputs the query and runs the web application, the web page will proceed to the *Analysis* section, where the user can choose between two different visualizations, a distance tree with bootstrap support values or a minimum spanning network. If the distance tree is selected, the user can change the tree algorithm (either neighbor joining or UPGMA) and number of bootstrap replicates. The user can also download results as a PDF or in NEWICK format. For the minimum spanning network, the user can adjust the grey scale for edge distances and download results as a PDF file. Implementations for other genetic data and different visualizations can be added.

Phytophthora-ID implementation

Tools provided in Microbe-ID were implemented in <http://Phytophthora-ID.org>, a functional website for identifying samples from the genus *Phytophthora*, a group of economically important plant pathogens in the stramenopile branch of the tree of life (Kamoun et al., 2015). Phytophthora-ID version 1.0 (Grünwald et al., 2011) had a BLAST script to identify samples using the ITS barcode. *Phytophthora-ID* version 2.0

Table 2 Genetic distances implemented in the Genotype-ID module of Microbe-ID. Each of the distances included in Microbe-ID are specific to a given molecular marker used in the web application.

Distance model	Module	R package	References
Felsenstein 81 (F81)	MLST-ID	ape	<i>Felsenstein (1981)</i>
Felsenstein 84 (F84)	MLST-ID	ape	<i>Felsenstein (1989)</i>
Indel	MLST-ID	ape	<i>Paradis, Claude & Strimmer (2004)</i>
Jukes-Cantor (JC69)	MLST-ID	ape	<i>Jukes & Cantor (1969)</i>
Kimura 80 (K80)	MLST-ID	ape	<i>Kimura (1980)</i>
Kimura 81 (K81)	MLST-ID	ape	<i>Kimura (1981)</i>
Raw	MLST-ID	ape	<i>Paradis, Claude & Strimmer (2004)</i>
Tamura and Nei 93 (TN93)	MLST-ID	ape	<i>Tamura & Nei (1993)</i>
Transitions (TS)	MLST-ID	ape	<i>Paradis, Claude & Strimmer (2004)</i>
Transversions (TV)	MLST-ID	ape	<i>Paradis, Claude & Strimmer (2004)</i>
Bruvo	SSR-ID	poppr	<i>Bruvo et al. (2004)</i>
Edwards	Binary-ID	poppr/adegenet	<i>Edwards (1971)</i>
Nei	Binary-ID	poppr/adegenet	<i>Nei (1972)</i>
Prevosti	Binary-ID	poppr/adegenet	<i>Prevosti, Ocaña & Alonso (1975)</i>
Reynolds	Binary-ID	poppr/adegenet	<i>Reynolds, Weir & Cockerham (1983)</i>
Rogers	Binary-ID	poppr/adegenet	<i>Rogers (1972)</i>

Notes.

MLST-ID, Multilocus sequence typing; SSR-ID, SSR/microsatellite loci; Binary-ID, AFLP/SNP loci.

was substantially revised and upgraded from the first iteration and now includes a faster BLAST search implemented in Sequence-ID and a new genotype identification system implemented in Genotype-ID.

The current version of *Phytophthora*-ID contains a Sequence-ID module customized for two molecular barcodes used for *Phytophthora* species identification (ITS and *cox* spacer). This particular version of the sequence identification tool was created in PERL-CGI, and permits the search of a FASTA sequence query against a curated database of *Phytophthora* species. The PERL-CGI for *Phytophthora*-ID version 2.0 was also redesigned to run directly on the server to make the web application more stable and faster. In contrast, in *Phytophthora*-ID 1.0, the web application was designed as a communication wrapper to an external cluster, thus rendering functionality dependent on the external server.

We implemented the Sequence-ID module to use sequences from two genetic regions: the nuclear internal transcribed spacer (ITS) and mitochondrial *cox* spacer region spanning the *cox1* and *cox2* loci. Databases were created using sequences from published *Phytophthora* species descriptions (*Blair et al., 2008; Martin & Tooley, 2003*) or those that belong to classical *Phytophthora* species described at least 25 years ago and are still recognized as valid (*Erwin & Ribeiro, 1996*).

For the ITS region, we gathered a total of 211 sequences representing 108 species. For the *cox* spacer region, we created a database of 150 sequences representing 106 species. Laboratory protocols for preparing samples for sequence analysis are available on the website and in a previous publication (*Grünwald et al., 2011*). Additionally, a file with the complete set of ITS sequences, *cox* spacer sequences, GenBank accession numbers, and

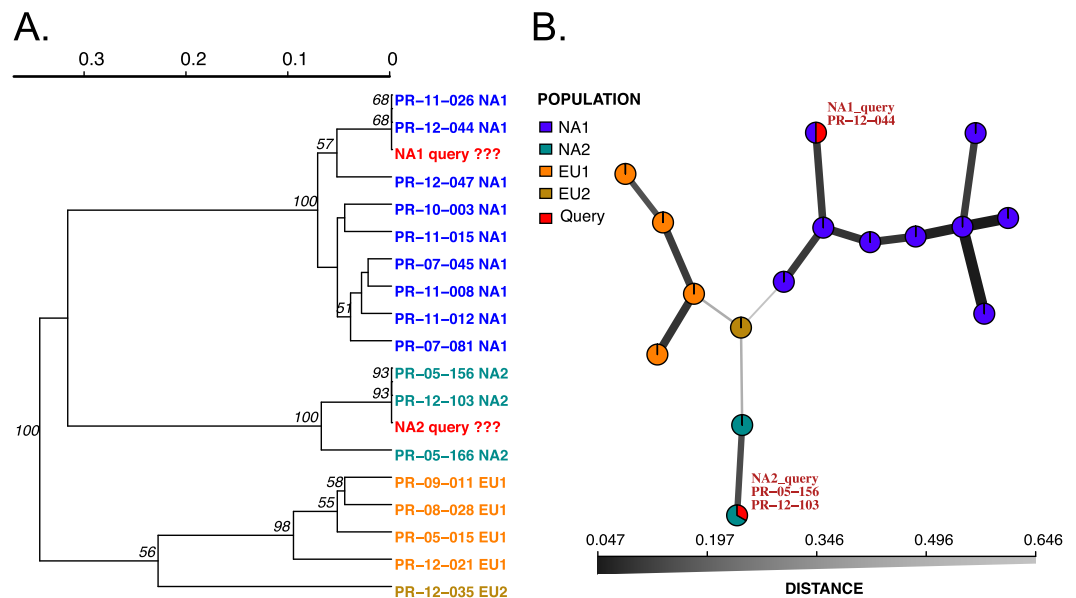


Figure 2 Results of SSR-ID for NA1 and NA2 queries of *P. ramorum* provided in the example data file. Each color represents a clonal lineage pre-assigned to each reference sample (NA1, NA2, EU1, EU2) with queries colored in red. (A) UPGMA tree with 1,000 bootstrap replicates and support values above branches. Queries are represented in red and all are correctly placed with reference samples of the presumptive clonal lineage while also representing the relationship between clonal lineages in the reference dataset. (B) Minimum spanning network reconstruction. Edge shade and width are inversely proportional to Bruvo's distance as shown in the horizontal scale bar. Queries are represented in red and placed in nodes with the most similar reference sample in the dataset, indicating the NA1 query is most similar to the PR-12-044 reference sample and the NA2 query is more closely related to the PR-05-156 and PR-12-103 samples, which also belong to the NA2 clonal lineage.

Phytophthora spp. can be downloaded from the website. Documentation on updates to the databases is provided on the website.

We implemented two modules for genotype identification of two *Phytophthora* species: *P. infestans* that causes potato late blight, and *P. ramorum* that causes sudden oak death. *P. infestans* has more than 18 reported clonal lineages (Hu et al., 2012), while *P. ramorum* has 4 reported clonal lineages (Grünwald et al., 2009; Van Poucke et al., 2012). To establish a database with a wide representation of clonal lineages in each species, we obtained and prepared DNA samples from *P. infestans* and *P. ramorum* that were collected from various regions of the world. A total of 48 *P. ramorum* isolates representing the 4 reported clonal lineages (NA1, NA2, EU1, EU2) were genotyped at nine SSR loci (Prospero et al., 2007; Vercauteren et al., 2010; Grünwald et al., 2009; Ivors et al., 2006). Similarly, 11 *P. infestans* clonal lineages that are dominant in the US (including US11, US12, US8, US20, US21, US23, US24, EU4, EU5, EU8, EU13) represented by a total population of 96 isolates were genotyped at 11 SSR loci, using protocols from Lees et al. (2006) and Li et al. (2013). We constructed SSR reference databases compatible with the poppr/adeget R packages that provide dendrograms and minimum spanning networks for queries. Bootstrap support can be calculated for dendrograms. To test the datasets, we used two queries per dataset, one sample of NA1 and NA2 clonal lineages each for *P. ramorum* and one sample of clonal

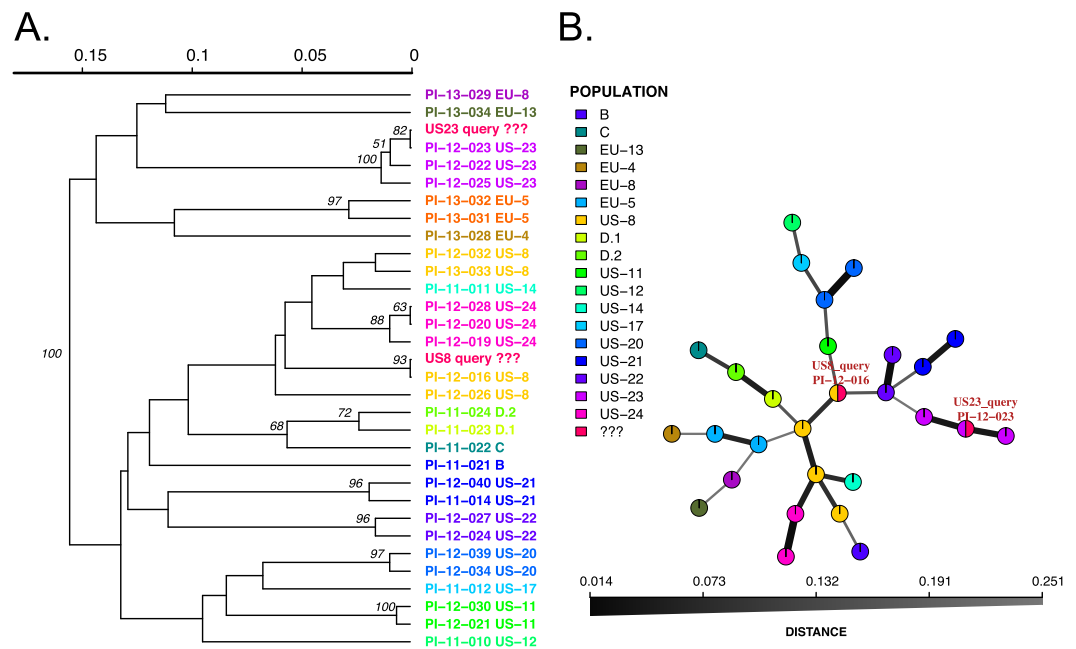


Figure 3 Results of SSR-ID queries for strains placed into the US8 and US23 clonal lineages of the potato late blight pathogen, *P. infestans*. Colors correspond to clonal lineages assigned to each reference sample (B, C, EU-13, EU-14, etc.) except for the queries which are colored in red. (A) UPGMA tree with 1,000 bootstrap replicates with support values above branches. Queries are represented in red and all are correctly placed with samples of the presumptive clonal lineage while also representing relationships between clonal lineages in the reference dataset. (B) Minimum spanning network reconstruction. Edge shade and width are proportional to Bruvo's distance shown in the horizontal scale bar. Queries are represented in red nodes and appear in legend as '???'. Queries placed in nodes with the most similar reference sample, indicating that the US8 query is most similar to the PI-12-016 reference sample (US-8 clonal lineage) and the US23 query is most closely related to the PI-12-023 sample, part of the US-23 lineage.

lineages US-23 and US-8 clonal lineages for *P. infestans*. All samples grouped with the corresponding reference lineages with high support values (Figs. 2 and 3), demonstrating functionality of the Microbe-ID tool to identify samples correctly.

CONCLUSIONS

We constructed a web framework that can use a wide array of molecular markers to rapidly determine identities of species and genotypes. This web framework is provided as open source code on GitHub (<https://github.com/grunwaldlab/microbe-ID>). We implemented a demonstration available at <http://Microbe-ID.org>. New implementations for any organisms require reference databases, scripts, and web pages including choices from the set of computational tools shown in Table 1. The Microbe-ID framework is currently implemented for two fully functional sites, *Phytophthora*-ID (<http://Phytophthora-ID.org/>) and Gall-ID (<http://Gall-ID.cgrb.oregonstate.edu/>; Davis II et al., 2016).

Modularity of Microbe-ID permits implementation of a range of markers using curated databases of known species or genotypes to determine identity of specimens. Use of a custom BLAST database permits identification of any sample of a species of interest by using any sequence-based molecular marker. Note that the amount of time required for

BLAST execution is dependent on size of the databases. For the Sequence-ID ITS region implementation for *Phytophthora*, each run was completed and loaded to the web-app in less than one second, with results shown as a BLAST output table.

Most innovative in this web framework is the Genotype-ID module, which permits the researcher to use, in a simple, web-based interface, any type of molecular marker with a corresponding and curated reference dataset. Use of a shiny server for strain identification is completely novel and modularity permits use of virtually any molecular marker suitable for existing R packages. For genotyping, population genetic markers demonstrated include SSR, MLST, and AFLP/RFLP data, but can also be expanded to presence/absence of genes or alleles or any other genetic features of interest. Custom implementations must include a reference dataset and custom R scripts. Implementations can be customized for virtually any organism and any molecular marker that can be analyzed in R.

Development of tools for species identification has increased in recent years; but lack of modularity and complicated methodological procedures make the process of species identification tedious and challenging for researchers without sufficient skills in computational biology. We developed an easy-to-set-up web framework, which permits quick and flexible deployment of species identification and genotyping tools using sequence, microsatellite/SSR, AFLP/RFLP, or MLST data for any organism. Given use of the bootstrap html framework, webpages are light in terms of computer resources required and are thus usable on any device including smartphones and tablets, and function with any browser. These tools have been deployed and used successfully for the genus *Phytophthora* and gall-forming bacteria, providing examples of two curated websites ([Grünwald et al., 2011](#); [Davis II et al., 2016](#)).

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This research is supported in part by the US Department of Agriculture (USDA) Agricultural Research Service Grant 5358-22000-039-00D (NJG), USDA National Institute of Food and Agriculture (NIFA) Grant 2011-68004-30154 (NJG), the USDA ARS Floriculture Nursery Research Initiative (NJG), USDA NIFA Grant 2014-51181-22384 (JHC and NJG), and USDA NIFA 2012-67012-19844 (SEE). 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:

US Department of Agriculture (USDA) Agricultural Research Service: 5358-22000-039-00D.

USDA National Institute of Food and Agriculture (NIFA): 2011-68004-30154.

USDA ARS Floriculture Nursery Research Initiative.

USDA NIFA: 2014-51181-22384, 2012-67012-19844.

Competing Interests

Niklaus Grünwald and Jeff Chang are Academic Editors for PeerJ.

Author Contributions

- Javier F. Tabima conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Sydney E. Everhart conceived and designed the experiments, analyzed the data, wrote the paper, reviewed drafts of the paper.
- Meredith M. Larsen and Matthew A. Tancos performed the experiments, contributed reagents/materials/analysis tools, reviewed drafts of the paper.
- Alexandra J. Weisberg and Zhian N. Kamvar performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, reviewed drafts of the paper.
- Christine D. Smart and Jeff H. Chang wrote the paper, reviewed drafts of the paper.
- Niklaus J. Grünwald conceived and designed the experiments, wrote the paper, reviewed drafts of the paper.

Data Availability

The following information was supplied regarding data availability:

<https://github.com/grunwaldlab/Microbe-ID>.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.2279#supplemental-information>.

REFERENCES

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403–410 DOI 10.1016/S0022-2836(05)80360-2.
- Aurrecoechea C, Barreto A, Brestelli J, Brunk BP, Cade S, Doherty R, Fischer S, Gajria B, Gao X, Gingle A, Grant G. 2013. EuPathDB: the eukaryotic pathogen database. *Nucleic Acids Research* 41:D684–D691.
- Blair JE, Coffey MD, Park S-Y, Geiser DM, Kang S. 2008. A multi-locus phylogeny for *Phytophthora* utilizing markers derived from complete genome sequences. *Fungal Genetics and Biology* 45:266–277 DOI 10.1016/j.fgb.2007.10.010.
- Bruvo R, Michiels NK, D’Souza TG, Schulenburg H. 2004. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Molecular Ecology* 13:2101–2106 DOI 10.1111/j.1365-294X.2004.02209.x.
- Byrnes III EJ, Li W, Lewit Y, Ma H, Voelz K, Ren P, Carter DA, Chaturvedi V, Bildfell RJ, May RC, Heitman J. 2010. Emergence and pathogenicity of highly virulent *Cryptococcus gattii* genotypes in the northwest United States. *PLoS Pathogens* 6(4):e1000850 DOI 10.1371/journal.ppat.1000850.
- Coleman AW. 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends in Genetics* 19:370–355 DOI 10.1016/S0168-9525(03)00118-5.
- Coleman AW. 2007. Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Research* 35:3322–3329 DOI 10.1093/nar/gkm233.

- Coleman AW. 2009.** Is there a molecular key to the level of “biological species” in eukaryotes? A DNA guide. *Molecular Phylogenetics and Evolution* **50**:197–203 DOI [10.1016/j.ympev.2008.10.008](https://doi.org/10.1016/j.ympev.2008.10.008).
- Cooke DE, Cano LM, Raffaele S, Bain RA, Cooke LR, Etherington GJ, Deahl KL, Farrer RA, Gilroy EM, Goss EM, Grünwald NJ. 2012.** Genome analyses of an aggressive and invasive lineage of the Irish potato famine pathogen. *PLoS Pathogens* **8**:e1002940 DOI [10.1371/journal.ppat.1002940](https://doi.org/10.1371/journal.ppat.1002940).
- Davis II EW, Weisberg AJ, Tabima JF, Grünwald NJ, Chang JH. 2016.** Gall-ID: tools for genotyping gall-causing phytopathogenic bacteria. *PeerJ* **4**:e2222 DOI [10.7717/peerj.2222](https://doi.org/10.7717/peerj.2222).
- Edwards AWF. 1971.** Distances between populations on the basis of gene frequencies. *Biometrics* **27**:873–881 DOI [10.2307/2528824](https://doi.org/10.2307/2528824).
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011.** A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**:e19379 DOI [10.1371/journal.pone.0019379](https://doi.org/10.1371/journal.pone.0019379).
- Erwin DC, Ribeiro OK. 1996.** *Phytophthora diseases worldwide*. St. Paul: American Phytopathological Society.
- Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA. 2010.** SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Methods in Molecular Biology* **772**:157–178 DOI [10.1007/978-1-61779-228-1_9](https://doi.org/10.1007/978-1-61779-228-1_9).
- Felsenstein J. 1981.** Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* **17**:368–376 DOI [10.1007/BF01734359](https://doi.org/10.1007/BF01734359).
- Felsenstein J. 1989.** PHYLIP- Phylogeny Inference Package. In: *Cladistics*. Version 3.2. Vol. 5. 164–166.
- Grajales A, Aguilar C, Sánchez JA. 2007.** Phylogenetic reconstruction using secondary structures of Internal Transcribed Spacer 2 (ITS2, rDNA): finding the molecular and morphological gap in Caribbean gorgonian corals. *BMC Evolutionary Biology* **7**:90 DOI [10.1186/1471-2148-7-90](https://doi.org/10.1186/1471-2148-7-90).
- Gross A, Holdenrieder O, Pautasso M, Queloz V, Sieber TN. 2014.** *Hymenoscyphus pseudoalbidus*, the causal agent of European ash dieback. *Molecular Plant Pathology* **15**:5–21 DOI [10.1111/mpp.12073](https://doi.org/10.1111/mpp.12073).
- Grünwald NJ, Goss EM, Ivors K, Garbelotto M, Martin FN, Prospero S, Hansen E, Bonants PJ, Hamelin RC, Chastagner G, Werres S. 2009.** Standardizing the nomenclature for clonal lineages of the sudden oak death pathogen, *Phytophthora ramorum*. *Phytopathology* **99**:792–795 DOI [10.1094/PHYTO-99-7-0792](https://doi.org/10.1094/PHYTO-99-7-0792).
- Grünwald NJ, Goss EM, Press CM. 2008.** *Phytophthora ramorum*: a pathogen with a remarkably wide host-range causing sudden oak death on oaks and ramorum blight on woody ornamentals. *Molecular Plant Pathology* **9**:729–740 DOI [10.1111/j.1364-3703.2008.00500.x](https://doi.org/10.1111/j.1364-3703.2008.00500.x).
- Grünwald NJ, Hoheisel G-A. 2006.** Hierarchical analysis of diversity, selfing and genetic differentiation in populations of the oomycete *Aphanomyces euteiches*. *Phytopathology* **96**:1134–1141 DOI [10.1094/PHYTO-96-1134](https://doi.org/10.1094/PHYTO-96-1134).

- Grünwald NJ, Martin FN, Larsen MM, Sullivan CM, Press CM, Coffey MD, Hansen EM, Parke JL. 2011. *Phytophthora*-ID.org: a sequence based *Phytophthora* identification tool. *Plant Disease* 95:337–342 DOI 10.1094/PDIS-08-10-0609.
- Grünwald NJ, McDonald BM, Milgroom MG. 2016. Population genomics of fungal and oomycete pathogens. *Annual Review of Phytopathology* In Press DOI 10.1146/annurev-phyto-080614-115913.
- Holdich DM, Reynolds JD, Souty-Grosset C, Sibley PJ. 2009. A review of the ever increasing threat to European crayfish from non-indigenous crayfish species. *Knowledge and Management of Aquatic Ecosystems* 11:394–395 DOI 10.1051/kmae/2009025.
- Hu CH, Perez FG, Donahoo R, McLeod A, Myers K, Ivors K, Secor G, Roberts P, Deahl KL, Fry WE, Ristaino JB. 2012. Recent genotypes of *Phytophthora infestans* in the eastern United States reveal clonal populations and reappearance of mefenoxam sensitivity. *Plant Disease* 96:1323–1330 DOI 10.1094/PDIS-03-11-0156-RE.
- Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T, Dong G. 2009. High-throughput genotyping by whole-genome resequencing. *Genome Research* 19:1068–1076 DOI 10.1101/gr.089516.108.
- Ivors K, Garbelotto M, Vries IDE, Ruyter-Spira C, Te Hekkert B, Rosenzweig N, Bonants P. 2006. Microsatellite markers identify three lineages of *Phytophthora ramorum* in US nurseries, yet single lineages in US forest and European nursery populations. *Molecular Ecology* 15:1493–1505 DOI 10.1111/j.1365-294X.2006.02864.x.
- James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Gueidan C, Fraker E, Miadlikowska J, Lumbsch HT. 2006. Reconstructing the early evolution of fungi using a six-gene phylogeny. *Nature* 443:818–822 DOI 10.1038/nature05110.
- Jombart T. 2008. Adegnet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405 DOI 10.1093/bioinformatics/btn129.
- Jukes TH, Cantor CR. 1969. Evolution of protein molecules. *Mammalian Protein Metabolism* 3:12–132.
- Kamoun S, Furzer O, Jones JDG, Judelson HS, Shad Ali G, Dalio RJD, Guha Roy S, Schena L, Zampounis A, Panabières F, Cahill D, Ruocco M, Figueiredo A, Chen X-R, Hulvey J, Stam R, Lamour K, Gijzen M, Tyler BN, Grünwald NJ, Tor M, Mukhtar SM, Tome D, Van den Ackerveken G, McDowell J, Daayf F, Fry WE, Lindqvist-Kreuzer H, Meijer HJG, Petre B, Ristaino J, Yoshida K, Birch P, Govers F. 2015. The top 10 oomycete pathogens in molecular plant pathology. *Molecular Plant Pathology* 16:413–434 DOI 10.1111/mpp.12190.
- Kamvar ZN, Brooks JC, Grünwald NJ. 2015. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics* 6:208 DOI 10.3389/fgene.2015.00208.
- Kamvar ZN, Tabima JF, Grünwald NJ. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2:e281 DOI 10.7717/peerj.281.
- Katoh K, Misawa K, Kuma KI, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30:3059–3066 DOI 10.1093/nar/gkf436.

- Kimura M. 1980.** A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*. **16**:111–120 DOI [10.1007/BF01731581](https://doi.org/10.1007/BF01731581).
- Kimura M. 1981.** Estimation of evolutionary distances between homologous nucleotide sequences. *Proceedings of the National Academy of Sciences of the United States of America* **78**:454–458.
- Klevens RM, Morrison MA, Nadle J, Petit S, Gershman K, Ray S, Harrison LH, Lynfield R, Dumyati G, Townes JM, Craig AS. 2007.** Invasive methicillin-resistant *Staphylococcus aureus* infections in the United States. *JAMA* **298**:1763–1771 DOI [10.1001/jama.298.15.1763](https://doi.org/10.1001/jama.298.15.1763).
- Kroon LP, Brouwer H, De Cock AW, Govers F. 2012.** The genus *Phytophthora* anno 2012. *Phytopathology* **102**:348–364 DOI [10.1094/PHYTO-01-11-0025](https://doi.org/10.1094/PHYTO-01-11-0025).
- Lees AK, Wattier R, Shaw DS, Sullivan L, Williams NA, Cooke DEL. 2006.** Novel microsatellite markers for the analysis of *Phytophthora infestans* populations. *Plant Pathology* **55**:311–319 DOI [10.1111/j.1365-3059.2006.01359.x](https://doi.org/10.1111/j.1365-3059.2006.01359.x).
- Li Y, Cooke DEL, Jacobsen E, Van der Lee T. 2013.** Efficient multiplex simple sequence repeat genotyping of the oomycete plant pathogen *Phytophthora infestans*. *Journal of Microbiological Methods* **92**:316–322 DOI [10.1038/nrmicro3093](https://doi.org/10.1038/nrmicro3093).
- Maiden MCJ, Van Rensburg MJJ, Bray JE, Earle SG, Ford SA, Jolley KA, McCarthy ND. 2013.** MLST revisited: the gene-by-gene approach to bacterial genomics. *Nature Reviews Microbiology* **11**:728–736 DOI [10.1038/nrmicro3093](https://doi.org/10.1038/nrmicro3093).
- Martin FN, Tooley PW. 2003.** Phylogenetic relationships among *Phytophthora* species inferred from sequence analysis of mitochondrially encoded cytochrome oxidase I and II genes. *Mycologia* **95**:269–284 DOI [10.2307/3762038](https://doi.org/10.2307/3762038).
- Nei M. 1972.** Genetic distance between populations. *The American Naturalist* **106**:283–292 DOI [10.1086/282771](https://doi.org/10.1086/282771).
- Paradis E. 2010.** Pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* **26**:419–420 DOI [10.1093/bioinformatics/btp696](https://doi.org/10.1093/bioinformatics/btp696).
- Paradis E, Claude J, Strimmer K. 2004.** APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**:289–290 DOI [10.1093/bioinformatics/btg412](https://doi.org/10.1093/bioinformatics/btg412).
- Park J, Park B, Veeraghavan N, Jung K, Lee Y-H, Blair JE, Geiser DM, Isard S, Mansfield MA, Nikolaeva E, Park S-Y, Russo J, Kim SH, Greene M, Ivors KL, Balci Y, Peiman M, Erwin DC, Coffey MD, Rossman A, Farr D, Cline E, Grünwald NJ, Luster DG, Schrandt J, Martin F, Ribeiro OK, Makalowska I, Kang S. 2008.** *Phytophthora* database: a forensic database supporting the identification and monitoring of *Phytophthora*. *Plant Disease* **92**:966–972 DOI [10.1094/PDIS-92-6-0966](https://doi.org/10.1094/PDIS-92-6-0966).
- Prevosti A, Ocaña J, Alonso G. 1975.** Distances between populations of *Drosophila subobscura*, based on chromosome arrangement frequencies. *Theoretical and Applied Genetics* **45**:231–241 DOI [10.1007/BF00831894](https://doi.org/10.1007/BF00831894).
- Prospero S, Hansen EM, Grünwald NJ, Winton LM. 2007.** Population dynamics of the sudden oak death pathogen *Phytophthora ramorum* in Oregon from 2001 to 2004. *Molecular Ecology* **16**:2958–2973 DOI [10.1111/j.1365-294X.2007.03343.x](https://doi.org/10.1111/j.1365-294X.2007.03343.x).

- Reynolds J, Weir BS, Cockerham CC. 1983.** Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* **105**:767–779.
- Robideau GP, De Cock AW, Coffey MD, Voglmayr H, Brouwer H, Bala K, Chitty DW, Désaulniers N, Eggertson QA, Gachon CMM. 2011.** DNA barcoding of oomycetes with cytochrome c oxidase subunit I and internal transcribed spacer. *Molecular Ecology Resources* **11**:1002–1011 DOI [10.1111/j.1755-0998.2011.03041.x](https://doi.org/10.1111/j.1755-0998.2011.03041.x).
- Rogers JS. 1972.** Measures of genetic similarity and genetic distance. *Studies in Genetics* **7**:145–153.
- RStudio Team. 2015.** *R Studio: integrated development for R*. Boston: R Studio, Inc. Available at <http://www.rstudio.com/>.
- Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W. 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](https://doi.org/10.1073/pnas.1117018109).
- Stajich JE, Harris T, Brunk BP, Brestelli J, Fischer S, Harb OS, Kissinger JC, Li W, Nayak V, Pinney DF, Stoeckert CJ. 2011.** FungiDB: an integrated functional genomics database for fungi. *Nucleic Acids Research* **40**:D675–D681 DOI [10.1093/nar/gkr918](https://doi.org/10.1093/nar/gkr918).
- Tamura K, Nei M. 1993.** Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10**:512–526.
- Tancos MA, Lange HW, Smart CD. 2015.** Characterizing the genetic diversity of the *Clavibacter michiganensis* subsp. *michiganensis* population in New York. *Phytopathology* **105**:169–179 DOI [10.1094/PHYTO-06-14-0178-R](https://doi.org/10.1094/PHYTO-06-14-0178-R).
- Van Poucke K, Franceschini S, Webber JF, Vercauteren A, Turner JA, McCracken AR, Heungens K, Brasier CM. 2012.** Discovery of a fourth evolutionary lineage of *Phytophthora ramorum*: EU2. *Fungal Biology* **116**:1178–1191 DOI [10.1016/j.funbio.2012.09.003](https://doi.org/10.1016/j.funbio.2012.09.003).
- Vercauteren A, De Dobbelaere I, Grünwald NJ, Bonants P, Van Bockstaele E, Maes M, Heungens K. 2010.** Clonal expansion of the Belgian *Phytophthora ramorum* populations based on new microsatellite markers. *Molecular Ecology* **19**:92–107 DOI [10.1111/j.1365-294X.2009.04443.x](https://doi.org/10.1111/j.1365-294X.2009.04443.x).