# An easy and robust method for isolation and validation of single-nucleotide polymorphic markers from a first Erysiphe alphitoides draft genome

Cyril Dutech, N. Feau, Isabelle Lesur, François Ehrenmann, Thomas Letellier, B. Li, Charlotte Mouden, Marie Laure Desprez-Loustau, Erwan Guichoux, A. Gross

*This is a post-peer-review, pre-copyedit version of an article published in Mycological Progress The final authenticated version is available online at: http://dx.doi.org/10.1007/s11557-020-01580-w.*

# Isolation and validation of single nucleotide polymorphic markers from a first *Erysiphe alphitoides* draft genome.

Dutech C.[1#], Feau N.[2], Lesur I.[1, 3], Ehrenmann F.[1], Letellier T.[4], Jeziorski C.[5], Li B.[1], Mouden C.[1,], Guichoux[1] E., Desprez-Loustau M.-L.[1], Gross A[1,*].

[1] *UMR BIOGECO, INRAE, Univ. Bordeaux, F-33610 Cestas, France*

[2] *Department of Forest and Conservation Sciences, Forest Sciences Centre, University of British Columbia, Vancouver, BC, Canada*

[3] *HelixVenture, Mérignac, F33700, FRANCE*

[4] *Université Paris-Saclay, INRAE, URGI, 78026, Versailles, France*

[5] *INRAE, US 1426, GeT-PlaGe, Genotoul, Castanet-Tolosan, France (doi: 10.15454/1.5572370921303193E12)*

[*] *Present address : Eidg. Forschungstanstalt WSL, Zürcherstrasse 111, 8903, Birmensdorf, Switzerland*

**Isolating genetic markers is often costly and time-consuming for non model fungal species. However, these markers are of a primary importance to identify the origin of invasive species and to infer their reproductive mode and dispersal ability. We slightly modified a recent molecular method to quickly isolate and validate single nucleotide polymorphism (SNP) markers, from a first *Erysiphe alphitoides* draft genome, one of the main causal agent of oak powdery mildew in Europe. Although the draft assembly was strongly fragmented (555,289 contigs), we successfully isolated 1,700 SNPs from 75 single copy genes conserved in most fungal genomes. Ninety percent of them allowed to clearly distinguish the two main *Erysiphe* species reported on European oaks: *E. alphitoides* and *E. quercicola*. Thirty-six SNPs, located in distinct genes, were then validated using a strategy of MassArray genotyping on 95 *E. alphitoides* isolates sampled in Europe. This genotyping showed that only monospore isolates had the expected haploid signature, whereas direct genotyping from field leaves showed signature of mixed infection. Considering haploid isolates, these markers led to the first results of population genetic diversity, and suggested that *E. quercicola* may have a more asexual reproduction than its sister species, *E. alphitoides*.**

**Keywords:** Biological invasion, genome assembly, oak powdery mildew, fungal forest pathogen, single-copy genes, development of genetic markers.

[#] *Corresponding author : Cyril Dutech, cyril.dutech@inra.fr, 33+ 5 35 38 52 97*

## Introduction

Invasive fungal species represent a great threat for ecosystems, especially for forests (Gladieux et al. 2015, Desprez-Loustau et al. 2016). In the recent decades, a dramatic increase of emerging diseases, mostly caused by introductions of exotic pathogens, has been reported in forests worldwide, causing severe damage (Santini et al. 2013, Gross et al. 2014, Wingfield et al. 2017). One of the first crucial steps for risk management is the accurate taxonomic identification of the fungal species associated with the disease, since many emerging pathogens were unknown before their invasion (e.g. Gross et al. 2014). Taxonomic identification has been greatly simplified by the development of molecular markers, and especially the use of the nuclear ribosomal internal transcribed spacer (ITS), defined as the fungal taxonomic barcode (Schoch et al., 2012). However, additional genetic markers are sometimes required to accurately resolve more difficult species complexes (e.g. Feau et al. 2011, Queloz et al., 2011, Tsykun et al. 2013). A second step is the identification of the geographic origin of the emergent pathogen, assessment of its genetic diversity and reproduction regime (sexual/asexual) allowing to retrace the history of introduction, and to determine its adaptive potential in introduced areas (Gladieux et al. 2015). Identifying the geographic origin can be useful to find sources of natural regulation in native areas (for example genetic resistance of host species or occurrence of natural enemies), in view of their putative deployment in the introduced areas (e.g. Rouxel et al. 2013).

Traditionally, microsatellite loci, also called single sequence repeats (SSR), have been widely used for phylogeographic and population genetic studies (Selkoe and Toonen 2006). However, these markers may sometimes be difficult to isolate, especially in fungi (Dutech et al. 2007). In addition, routine genotyping with SSR may be costly because it needs inter-calibration of alleles between studies, taking into account stutter bands and the presence of null alleles (Ellis et al. 2011, Chapuis and Estoup 2007). Due to an unclear mutational model and potential homoplasy between some alleles (Estoup et al. 2002), it might also be difficult, in some cases, to infer a robust phylogeographic signal (but see Hardy et al. 2003). The recent advent of next

generation sequencing (NGS) has made possible the identification of thousands of single nucleotide polymorphisms (SNP) in non-model species, helping in deciphering their genetic diversity, reproductive regime and evolutionary history (e.g. Andrew et al. 2016, Elshire et al. 2011). Given the availability of these new tools and their ever decreasing costs, pathologists might be tempted to design population genetic studies of organisms without any previous knowledge of the reproduction biology and dissemination strategy of the organism. However, regardless of marker choice, an adequate sampling strategy and experimental design remain crucial in order to be able to answer basic population genetic questions (e.g. Arnaud-Haond et al. 2007, Meirmans 2015). Testing population genetic hypotheses may require in some cases taking into account spatial genetic structures to avoid coming to wrong conclusions (e.g. Meirmans 2012). The first step requires performing a pilot study by sampling several individuals; this can be done either with a random sampling or by using a hierarchical approach based on the spatial distribution of populations and their respective density (Arnaud-Haond et al. 2007, Barres et al. 2012). However, when using NGS technology this first study could represent a substantial waste of time and money, especially for clonal populations with many identical genotypes. By contrast, several studies have shown that genotyping a few SNPs, especially in fungal populations, may be sufficient for this first objective (e.g. Dutech et al. 2017, Tsykun et al. 2017). In addition, these SNPs may be quickly isolated using an easy, cheap, and robust strategy targeting conserved sets of single copy genes in a chosen taxonomic group (e.g. Feau et al. 2011, Tonnabel et al. 2014, Dutech et al. 2016). A significant advantage of this strategy is that it can be easily applied for a large panel of fungal species (Feau et al. 2018), and does not require bio-informatic expertise to deal with false positive SNPs generated by NGS, and associated with bias in computational analysis (e.g. Puritz et al. 2014, Ribeiro et al. 2015, Verdu et al. 2016).

Herein, we focused on the oak powdery mildew (OPM). This foliar disease affecting a large range of oak species in the northern hemisphere, is caused by various species in several genera, especially *Erysiphe* (Takamatsu et al. 2007). In Europe, the disease suddenly appeared in the early 20$^{th}$ century, and was thought to be caused by a previously unknown *Erysiphe* species described as *E. alphitoides* (Griffon and Maublanc 1912). Recent studies have determined that OPM in Europe is actually caused by a complex of three *Erysiphe* species: *E. alphitoides*, the most abundant and widely spread, *E. quercicola* and *E. hypophylla* (Mougou et al. 2008, Desprez-Loustau et al. 2018). These three species belong to a clade of at least five sister species all affecting oaks and present in Asia, suggesting an Asian origin of the species causing OPM in Europe (Desprez-Loustau et al. 2011, Takamatsu et al. 2007, 2015). Today, the disease is widespread in Europe but its impact has strongly decreased since its emergence, although it can still cause high mortalities on young seedlings in forests. Therefore, it may act as an important selective agent for oak populations at this early stage (Marçais and Desprez-Loustau 2014). Numerous questions remain to be investigated about this species complex such as the ecological niche differentiation of the species (Feau et al. 2012, Desprez-Loustau et al. 2018), their reproductive mode and winter survival (Feau et al. 2012, Hamelin et al. 2016), dispersal ability, or their history of introduction.

Only a few genetic markers are available for *E. alphitoides* and related species (Feau et al. 2011). They have been useful at the inter-specific level to describe the host range (Desprez-Loustau et al. 2017) and spatial distribution at different scales of the cryptic species (Desprez-Loustau et al. 2018). The objective of the present study was to develop genetic resources for population genetic analyses of OPM species by developing a pipeline allowing the detection of SNPs in a conserved set of single copy genes in fungal species, based on the principle described in Feau et al. (2011, 2018) with additional steps for species diversity analysis. The method presented here only requires basic bioinformatic analyses and public genetic resources. It can be applied to any fungal species by any research group even without strong expertise in genomics and bioinformatics. The preliminary population genetic analyses performed in this study confirmed that a first characterization of the population genetic structure and diversity of a non-model organism can be quickly estimated with the isolated SNPs, and provided useful guidelines of sampling for future genomic studies of the OPM European populations.

## Materials & Methods

***Fungal isolation, DNA extraction and genome sequencing.*** Fungal DNA for genome sequencing was obtained from conidia of the *E. alphitoides* monospore MS_42D strain, initially isolated from a leaf lesion of a young oak seedling (*Q. robur*) in south-western France (44.76 N; 0.71 W) in June 2013. To obtain sufficient quantities of spores, oak seedlings were inoculated with the MS_42D strain, and incubated in a growth chamber in plastic boxes. DNA was extracted using a CTAB method following Mougou et al. (2008), after grinding the spores in liquid nitrogen. Species identification was performed following ITS sequencing. Shotgun paired-end sequencing of genomic DNA was performed on the Illumina Hiseq2000 genome analyser, using the Truseq Genomic kit (Illumina Inc, USA) at the Genotoul facilities (Institut national de la recherche agronomique, Toulouse, France).

***Trimming, assembly and genome annotation.*** Quality of the Illumina paired-end reads was assessed using FastQC 0.10.0 (Andrews 2010). Reads were then trimmed and assembled into a draft genome using CLC Genomics Workbench 7.0.3 with the *de novo* assembler tool (CLC bio, Aarhus, Denmark) using default parameters. CD-HIT-est V4.6.0 (Fu et al. 2012) was used to remove redundancy in the contigs with the sequence identity threshold set to 95%. To remove possible oak sequence contamination, contigs were aligned against the oak transcriptome assembly OCV3 (Lesur et al. 2015). In addition, given the large number of contigs following the first assembly, we suspected microbiological contamination led to chimeric contigs (see Results). We thus decided to perform a two-steps supplementary filtering: first, we selected *E. alphitoides* contigs longer than 10Kb; then we kept only those with a sufficient protein homology with the 6,843 scaffolds of *Blumeria graminis* genome assembly V3 (available on https://mycocosm.jgi.doe.gov/Blugr1/Blugr1.home.html). This cereal powdery mildew genome assembly represents one of the most complete reference genomes among those sequenced for powdery mildew species (Spanu et al. 2010). The filtered set of *E. alphitoide*s contigs was aligned against the *Blumeria graminis* reference genome using BlastN (Altschul et al. 1990) with an E-value cut-off of 1e-5. Gene prediction was performed on this final set of contigs using Augustus V2.6.1 (Stanke et al. 2006), trained with the *Botrytis cinerea* genome (Amselem et al. 2011*)*. Genes identified by Augustus were finally validated by comparison of their nucleotidic sequences with the 6,470 *B. graminis* proteins using BlastX (Altschul et al. 1990) with an E-value cut-off of 1e-5.

***Identification of single copy genes and primer design.*** For SNP identification, we used candidate genes of the Funybase database (Marthey et al. 2008) which contains 246 reliable orthologous genes clusters present as single copy genes in 21 fungal genomes. Since *Sclerotinia sclerotiorum* is the closest species relative to

*Erysiphe* sp. in Funybase (http://genome.jouy.inra.fr/funybase/), we extracted all protein sequences of this species and searched them in the *E. alphitoides* contigs using TblastN (Altschul et al. 1990) with an E-value cut-off of 1e-20. All sequences with more than one hit in the *E. alphitoides* draft genome were rejected from subsequent analyses. We then checked whether these isolated nucleotidic sequences matched with the nucleotidic sequences of predicted genes obtained from Augustus, using BlastN and an E-value cut-off of 1e-20. Oligonucleotidic primers were then designed from selected candidate genes such that fragments of 300-450 bp are amplified, and that each future amplified sequence (hereafter called amplicon) contained at least one intron (assuming that introns have a higher rate of polymorphism than exons; Feau et al. 2011). Primer3 (Untergasser et al. 2012) was used for primer design using the default parameters. The Figure 1 describes the different steps of the workflow analysis to isolate SNPs from the *E. alphitoides* draft genome.

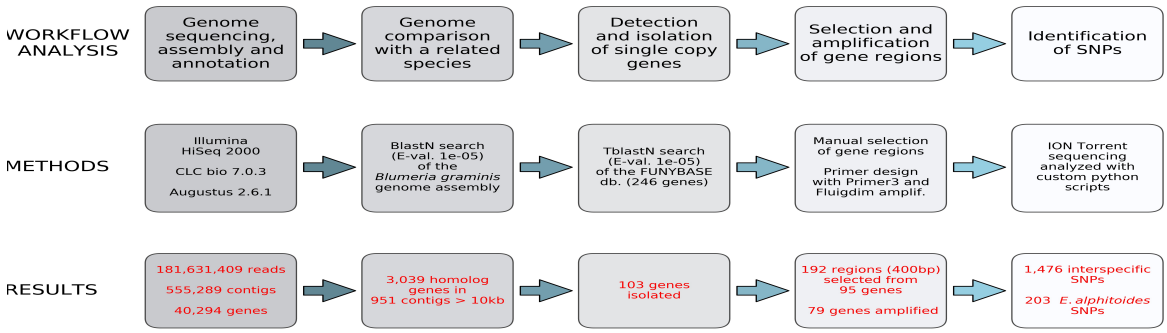


Figure 1. **Workflow analysis for isolation of SNPs using the draft *E. alphitoides* genome.**

***Gene sequencing and SNPs identification.*** A total of 47 monospores *E. alphitoides* and *E. quercicola* isolates were obtained from lesions on oak and mango leaves sampled in five European countries, by sub-culturing the monospore isolates on excised oak leaves in Petri dishes (Table 1). Each leaf inoculated with a single spore was desiccated and stored at INRAE-Pierroton. High molecular weight genomic DNA was extracted from these spores according to the protocol of Feehan et al. (2017). DNA was amplified using Targeted DNA Seq Library Reagent Kits (Fluidigm, San Francisco, USA) on an Access Array 48.48 IFC, following manufacturer's instructions. For each sample, two primer pairs were mixed together in the PCR reaction after the identification of the best combination among all the designed primer pairs of amplicons, using MultiPLX (Kaplinski et al., 2005). The quality and quantity of the final library were measured on a Bioanalyzer 2100 using High Sensitivity DNA kit (Agilent Technologies, Santa Clara, USA), and the final pool was diluted to 18 pM before sequencing on Ion Torrent PGM (Thermo Fisher Scientific, Waltham, USA). Ion Torrent Sequencing data were analyzed to identify SNPs using two custom Python (V2.7.6) scripts . Using the first script, Ion Torrent reads of each isolate were grouped into different amplicons, and then aligned to the corresponding reference sequences (*i.e.* the sequences extracted from the *E. alphitoides* draft genome) using MUSCLE (Edgar 2004). The second Python script was used to (i) build a consensus sequence for each amplicon, (ii) remove short indels which are often produced by the Ion Torrent Sequencing technology (Loman et al. 2012) and (iii) detect SNPs between the consensus sequences for each isolate and the reference amplicon. These scripts are available on https://doi.org/10.15454/UGMTBK.

***Validation of SNPs and population genetic analyses.*** Using the MassArray genotyping technology described in Chancerel et al. (2013), we designed two SNP arrays to validate a subset of the SNPs identified with the method described above. First, we selected SNPs following a manual check of the aligned consensus sequences for each amplicon, and removal of SNPs close to indels. We then selected the best combination of primers for amplification of the highest number of SNPs located in distinct genes, using the MassArray Designer V4.0.20.2. Ninety-five samples of *E. alphitoides* sampled in Europe were chosen for SNP validation (Table S1), comprising 23 isolates among the 47 monospore isolates previously used to identify SNPs, 48 samples from a previous pan-European study (Desprez-Loustau et al. 2018) (called hereafter “lesion samples”), and 24 herbarium samples of oak powdery mildew from distinct European herbaria (called hereafter “herbarium samples”). For lesion and herbarium samples, a 6-mm infected leaf disc was excised, and DNA extraction was performed according to Desprez-Loustau et al. (2018). After validation of SNPs by MassArray genotyping (*i.e.* positive amplification and polymorphism), we finally selected a single SNP per gene for population genetic analyses. Genetic diversity and population structure analyses were carried out using the R-CRAN packages PopR (Kamvar et al. 2015) and Adegenet (Jombart and Ahmed. 2011). Identical genetic analysis were also performed on the 13 *E. quercicola* sampled for this study (Table 1), using a subset of SNPs identified by Ion Torrent Sequencing technology, each one located in distinct genes.

| Species | Sampling site | Country | Host | Sampling Date | Geographical coordinates | | Names of isolates |
|---|---|---|---|---|---|---|---|
| *E.alphitoides* | Kersko | Czech Rep. | Q. robur | 12.06.2017 | 50.13366 °N | 14.95066 °E | AG288; AG292; AG324; AG325 |
| | Lussagnet, Landes | France | Q. robur and Q. petraea | 16.05.2017 | 43.763725 °N | -0.223289 °E | AG289; AG290; AG295 |
| | Algarrobo-Costa, La Majora | Spain | Mango, variety Kensington | 04.04.2017 | 36.756332 °N | -4.042761 °E | AG274; AG285; AG286 |
| | Alnarp | Sweden | Q. robur | 21.06.2017 | 55.6576 °N | 13.0824 °E | AG339; AG343; AG347; AG348; AG352 |
| | Ascona, Monte Veritas | Switzerland | Q. robur and Q. petraea | 09.05.2017 | 46.154716 °N | 8.753632 °E | AG283; AG296; AG299; AG300; AG302; AG304; AG306; AG307; AG310 AG314; AG315; AG316; AG317; AG318; AG319; AG320; AG321; AG322 AG323; AG338a |
| *E.quercicola* | Kersko | Czech Rep. | Q. robur | 12.06.2017 | 50.13366 °N | 14.95066 °E | AG335; AG334 |
| | Lussagnet, Landes | France | Q. robur and Q. petraea | 16.05.2017 | 43.763725 °N | -0.223289 °E | AG264; AG266; AG277; AG282 |
| | Algarrobo-Costa, La Majora | Spain | Mango, variety Kensington | 04.04.2017 | 36.756332 °N | -4.042761 °E | AG270; AG271; AG272; AG273; AG275; AG287 |
| | Ascona, Monte Veritas | Switzerland | Q. robur and Q. petraea | 09.05.2017 | 46.154716 °N | 8.753632 °E | AG313 |

Table 1. **Geolocation of the 34 *Erysiphe alphitoides* and 13 *E. quercicola* samples used to isolate and validate the single nucleotide polymorphism loci**

## Results

***Assembly and annotation of the Erysiphe alphitoides genome sequence.*** Sequencing of the *E. alphitoides* strain generated 2 x 184,631,409 100 bp paired-end reads. Average base qualities of the forward and reverse reads were 35.32 and 33.62, respectively, with quality ranging from 31 to 41 (encoding Sanger/Illumina 1.9). Following trimming, 181,631,409 paired-end reads were *de novo* assembled. A total of 592,884 contigs were obtained (N50= 1,646 bp, average length= 532 bp, total length = 315,529,701 bp; Table 2). Redundancy among contigs was removed using CD-HIT-est leading to a total of 555,289 contigs, with N50 = 1,735bp, contig length varying from 100 bp to 51,517 bp, a mean length of 532bp, and a total genome size higher than 300Mb (Table 2). A total of 3,088 contigs were removed due to sequence similarities with the oak genome or transcriptome. By similarity with *B. graminis* scaffolds (using a threshold of the Blast E-value = 10e-5), we identified 951 *E. alphitoides* contigs longer than 10Kb. These 951 contigs added up to a length of 15,296,844 bp, and a N50 equals to 16,172 bp. Using these contigs, gene annotation resulted in 3,045 gene models with a minimum and maximum length of 198 and 14,751bp respectively, and an average gene length of 1,624 bp (Table 2). Ninety-eight percent of these gene models (*i.e.* 3,039 out of 3,045) were validated by comparison with the 6,470 *B. graminis* proteins. Assembly and annotation data are available on https://arachne.pierroton.inra.fr/AlphiGeno/

| | Contigs >= 100bp | Contigs >= 500bp | Contigs >= 10Kb |
|---|---|---|---|
| Nb contigs | 555,289 | 103,796 | 951 |
| Total Assembly Size (Mb) | 308.4 | 220.5 | 15,3 |
| Mean Length (bp) | 532 | 2,124 | 16,085 |
| N50 (bp) | 1,735 | 3,449 | 16,172 |
| Nb Genes | - | - | 3,045 |
| Mean Size Genes | - | - | 1,624 |
| min size Gene | - | - | 198 |
| max size Gene | - | - | 14,751 |

Table 2. **Basic statistics of the *Erysiphe alphitoides* genome assembly**

Contigs >= 10Kb: contigs longer than 10kb, and homologs to Blumeria graminis contigs (Blast E-value less than 1e-5). Genes were detected with Augustus annotation software V2.6.1 using *B. graminis* as gene model.

***Identification of single copy genes in the draft E. alphitoides genome and selection of sequences for amplification.*** TBlastN searches of 246 *S. sclerotiorum* protein sequences against the 951 *E. alphitoides* contigs longer than 10Kb revealed a total of 147 (59.8%) *S. sclerotiorum* genes with an homolog in these contigs, among which 103 showed one single hit (Table S2). From this reduced list, we then selected 95 sequences homolog (BlastN E-value > 1e-50) to one of the automatically annotated genes of the *E. alphitoides* draft genome. Finally, using Primer3, we successfully designed 192 pairs of primers allowing the amplification of amplicons with a sequence size between 306 and 425 bp (Table S3).

***Gene sequencing and SNPs identification.*** Out of the 192 primer pairs tested on the 47 monospore *Erysiphe* isolates (*E. alphitoides* and *E. quercicola*), 165 pairs, targeting 79 distinct genes, produced amplicons (Table S3). We investigated the causes of failure of amplification by estimating phylogenetic relationships of the reference amplicons using PhyML 3.1 (Guindon et al. 2010), orthologs from the 21 complete genomes present in Funybase, and protein sequences from the *B. graminis* genome V3. Despite of our choice to only select the 10kb contigs homolog to Blumeria for limiting DNA amplification associated with putative DNA contamination from other microorganisms or host plants, we found that 11 reference amplicons isolated from the *E. alphitoide*s draft genome were related to *Aspergillus* sp. (Fig. S1). Our custom python-script detected 1794 SNPs in the 165 amplified sequences. These SNPs were located in 139 amplicons, corresponding to 79 distinct genes (Table S4). The median and maximum number of SNPs detected per amplicon was equal to 9 and 29, respectively. After filtering for putative spurious SNPs (*i.e.* not considering those located 30bp before the 3'-end and after 5'end of the alignment), we identified 203 intraspecific SNPs located in 53 genes for *E. alphitoides*, and 338 SNPs in 41 genes for *E. quercicola.* Among those, 22 SNPs were shared by the two species.

***Validation of SNPs for E. alphitoides and population genetic studies*.** For SNP validation in *E. alphitoides*, we selected a total of 58 SNPs within 39 distinct genes and grouped in two SNP arrays (Table 3). A set of 54 SNPs were polymorphic among the 23 monospore isolates (Table 3). The comparison between genotypes obtained by the Ion Torrent Sequencing and the MassArray genotyping Technology only revealed six discrepancies in the 23 control isolates, corresponding to an error rate of 0.0048. Most of these differences were concentrated on the isolate AG320. This isolate was probably not a monospore haploid isolate since several heterozygote SNPs were detected using the MassArray genotyping analysis.

| Gene | Plex | SNP_ID | 2nd-PCRP | 1st-PCRP | AMP_LEN | UP_CONF | MP_CONF | Tm | PcGC | Nalon | Helon | NaSeq | HeSeq | Pganalysis |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FG1020 | W1 | FG1020_1P91 | ACGTTGGATGTGTCCTCTTGCCTATAAGCC | ACGTTGGATGATGGGTCATTCGGATCACTC | 100 | 100.0 | 62.4 | 46.1 | 44.4 | 21.7 | 0.03 | 13.7 | 0.02 | yes |
| FG1020 | W1 | FG1020_833P43 | ACGTTGGATGGTATGGGCGCCTGTTAATTC | ACGTTGGATGCGCCAATGATAATTTGCTCC | 89 | 98.1 | 62.4 | 45.1 | 31.6 | 2.2 | 0.4 | 28.4 | 0.38 | NO |
| FG1021 | W1 | FG1021_587P207 | ACGTTGGATGTCATCTGACTCTTCACTGCC | ACGTTGGATGCTGCAGCCGTAGGGAAATTG | 99 | 98.6 | 62.4 | 47.3 | 38.1 | 0 | 0.45 | 25.3 | 0.47 | yes |
| FG1021 | W1 | FG1021_587P349 | ACGTTGGATGTTTGCGAGCGCATACTACTG | ACGTTGGATGACGCGTGCATTGGTAACATC | 119 | 98.6 | 62.4 | 54.0 | 47.8 | 0 | 0.45 | 15.8 | 0.41 | NO |
| FG508 | W1 | FG508_1P275 | ACGTTGGATGGCGTCAAACCTGGGTAATTC | ACGTTGGATGGTCCCCAACTTCTTTTCCAG | 119 | 98.6 | 62.4 | 45.5 | 25.0 | 2.2 | 0.32 | 23.2 | 0.29 | yes |
| FG534 | W1 | FG534_1053P157 | ACGTTGGATGTCTCTTACTAGAGATGCCTG | ACGTTGGATGGGTTCGCTCAATATGCGATG | 104 | 98.5 | 62.4 | 45.8 | 50.0 | 19.6 | 0.47 | 30.5 | 0.49 | yes |
| FG543 | W1 | FG543_413P231 | ACGTTGGATGAGCTTGCAACAATGATGAGG | ACGTTGGATGTATGGCCAGACATCCAACAC | 112 | 98.0 | 62.4 | 46.6 | 41.2 | 28.3 | 0.42 | 23.2 | 0.48 | NO |
| FG552 | W1 | FG552_534P211 | ACGTTGGATGAGTTCTCAATGGCTCCCTTC | ACGTTGGATGCAGAAGAAGCTTTAGGTGTG | 121 | 96.8 | 62.4 | 47.1 | 42.1 | 2.2 | 0.42 | 48.4 | 0.49 | yes |
| FG559 | W1 | FG559_1524P364 | ACGTTGGATGGTGAGGTCTTACAGATTCCC | ACGTTGGATGTTCCTGGAGGTGGAATAGAG | 102 | 100.0 | 62.4 | 50.8 | 42.9 | 2.2 | 0.43 | 30.5 | 0.49 | yes |
| FG586 | W1 | FG586_903P110 | ACGTTGGATGACGACACATCATCACGACAC | ACGTTGGATGTCTTCATCTTGTAGGTGGCG | 92 | 99.7 | 62.4 | 47.3 | 47.1 | 26.1 | 0.03 | 30.5 | 0.49 | yes |
| FG591 | W1 | FG591_1P335 | ACGTTGGATGCTTCGTGCATTACACTGCTC | ACGTTGGATGCGATACGCATTGTATCGTGG | 99 | 100.0 | 62.4 | 50.4 | 50.0 | 15.2 | 0.32 | 31.6 | 0.41 | yes |
| FG662 | W1 | FG662_1P315 | ACGTTGGATGATTCAACTTGGCCAGGACAG | ACGTTGGATGACTCAATATAGAGTGGCTGG | 103 | 98.5 | 62.4 | 45.5 | 25.0 | 2.2 | 0.4 | 27.4 | 0.42 | yes |
| FG673 | W1 | FG673_378P209 | ACGTTGGATGCCCCAACTCTGTTTCAAGTG | ACGTTGGATGACGTTCGATCCATTATCGGG | 120 | 98.4 | 62.4 | 54.2 | 47.6 | 2.2 | 0.37 | 31.6 | 0.49 | NO |
| FG684 | W1 | FG684_1P190 | ACGTTGGATGTCTCAATTAGATTGCTGCCG | ACGTTGGATGTGATGACGTACCTTGTCAGC | 89 | 98.1 | 62.4 | 48.7 | 71.4 | 2.2 | 0.03 | 14.7 | 0.02 | NO |
| FG685 | W1 | FG685_1P175 | ACGTTGGATGCTTAGCAGCAGATACCGCAC | ACGTTGGATGGAGAAAAGGCACGAACCTTG | 100 | 98.6 | 62.4 | 45.1 | 38.9 | 30.4 | 0.45 | 22.1 | 0.47 | yes |
| FG702 | W1 | FG702_1P276 | ACGTTGGATGCGCCGTTAAGTATGTAACGC | ACGTTGGATGAGGCTCTCAGTCAGCATTAC | 99 | 100.0 | 62.4 | 49.8 | 42.1 | 2.2 | 0.21 | 25.3 | 0.36 | yes |
| FG722 | W1 | FG722_1P199 | ACGTTGGATGGACTGAGGAAGAGTTGCTTG | ACGTTGGATGACGTCAACATTGACCTTTCG | 122 | 96.7 | 62.4 | 45.4 | 20.8 | 6.5 | 0.46 | 29.5 | 0.49 | yes |
| FG722 | W1 | FG722_1P294 | ACGTTGGATGACCTAGGCCTAACATGATGC | ACGTTGGATGTAGAAGGGCTACGTTTAGAC | 121 | 96.8 | 62.4 | 46.4 | 35.0 | 6.5 | 0.46 | 32.6 | 0.49 | NO |
| FG752 | W1 | FG752_1P183 | ACGTTGGATGAAATCCGCTGTTCGGTTGAG | ACGTTGGATGGACTTGATGTTGCTGGTCG | 108 | 98.2 | 62.4 | 48.8 | 32.0 | 2.2 | 0.29 | 26.3 | 0.22 | yes |
| FG813 | W1 | FG813_1P214 | ACGTTGGATGACGTCTTACTCTAACGGGTC | ACGTTGGATGCATGAGAAGAGGTACGGATG | 98 | 100.0 | 62.4 | 46.0 | 44.4 | 21.7 | 0.21 | 23.2 | 0.34 | yes |
| FG870 | W1 | FG870_1P148 | ACGTTGGATGACCAATAGGCTCTTTAGTCG | ACGTTGGATGGACATCTGCATCTAGACCAC | 120 | 97.0 | 62.4 | 45.1 | 44.4 | 2.2 | 0.13 | 24.2 | 0.27 | yes |
| FG893 | W1 | FG893_1P271 | ACGTTGGATGAATGCTAGGTACGACGTCAC | ACGTTGGATGCTCTTTTTGTGCTTCTCTGC | 112 | 98.0 | 62.4 | 46.8 | 44.4 | 15.2 | 0.03 | 16.8 | **0** | NO |
| MS313 | W1 | MS313_1679P209 | ACGTTGGATGAATGTGGCCAGAAGGTAAGC | ACGTTGGATGTGCATTAACCAGCCAGGAG | 102 | 98.4 | 62.4 | 47.3 | 29.2 | 15.2 | 0.08 | 32.6 | 0.44 | NO |
| MS320 | W1 | MS320_752P195 | ACGTTGGATGAATTAGTGCCTCTGTAGATG | ACGTTGGATGAGGGCTGCTAAATGTTAGAG | 127 | 90.3 | 62.4 | 45.5 | 40.0 | 2.2 | 0.29 | 30.5 | 0.32 | yes |
| MS378 | W1 | MS378_1956P328 | ACGTTGGATGCCATGGGCTGTTAACCTTGC | ACGTTGGATGAATTCGTGGGCTGCAGATTG | 100 | 98.6 | 62.4 | 47.8 | 31.8 | 2.2 | 0.03 | 15.8 | 0.02 | yes |
| MS380 | W1 | MS380_1P252 | ACGTTGGATGTGAATACTGAGAATTGCGAG | ACGTTGGATGGCAAATGCCCCAGCAATTAT | 114 | 89.9 | 62.4 | 46.9 | 28.6 | 2.2 | 0.37 | 29.5 | 0.43 | yes |
| MS397 | W1 | MS397_500P187 | ACGTTGGATGTTCGACCCAATTCTGTCCTC | ACGTTGGATGGCATTCAATGGGTAGTCAAG | 108 | 98.3 | 62.4 | 45.4 | 56.3 | 2.2 | 0.08 | 28.4 | 0.43 | yes |
| MS437 | W1 | MS437_129P126 | ACGTTGGATGACCGGCACCTTCATTGTTTC | ACGTTGGATGCCCATGGTCTGGTGAATTTG | 105 | 99.9 | 62.4 | 45.2 | 57.1 | 2.2 | 0.03 | 16.8 | 0.02 | NO |
| MS437 | W1 | MS437_539P167 | ACGTTGGATGCAATTCGTCACTGCCAGTAG | ACGTTGGATGGTTTGAAGCCATAGATCGGG | 108 | 99.7 | 62.4 | 47.1 | 60.0 | 0 | 0.43 | 30.5 | 0.49 | NO |
| MS441 | W1 | MS441_500P231 | ACGTTGGATGCTAGGTGCTTTAGTGTTGTC | ACGTTGGATGTAGGGTCAGGTTTTGATGGC | 108 | 98.3 | 62.4 | 47.8 | 35.0 | 2.2 | 0.43 | 27.4 | 0.49 | yes |
| MS462 | W1 | MS462_3063P320 | ACGTTGGATGATCTGTGCCACAGCTGTACG | ACGTTGGATGTCCAACAATGTGATCCTGCC | 108 | 98.3 | 62.4 | 47.9 | 25.0 | 30.4 | 0.4 | 29.5 | 0.47 | yes |
| MS501 | W1 | MS501_1P62 | ACGTTGGATGTTTACGGATCCCATCAGGAC | ACGTTGGATGTTCGTATGCTTTCGCCTGAG | 121 | 98.3 | 62.4 | 46.2 | 30.4 | 2.2 | 0.42 | 23.2 | 0.49 | yes |
| MS501 | W1 | MS501_643P260 | ACGTTGGATGGCGGCATATTCGTTCACATA | ACGTTGGATGAGGTCAAGAAAATCCTGCTG | 129 | 91.5 | 62.4 | 49.5 | 34.8 | 2.2 | 0.42 | 27.4 | 0.49 | NO |

Table 3. **list of the 58 single nucleotide polymorphism loci genotyped with the MassArray genotyping technology in Erysiphe alphitoides.**

Table 3 continued

| Gene | Plex | SNP_ID | 2nd-PCRP | 1st-PCRP | AMP_LEN | UP_CONF | MP_CONF | Tm | PcGC | NaIon | HeIon | NaSeq | HeSeq | Pganalysis |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FG478 | W2 | FG478_7P107 | ACGTTGGATGTTGGAAGCTCCCCATACAAG | ACGTTGGATGGATGAGACGGACATGAAAGC | 97 | 100.0 | 80.9 | 46.8 | 38.1 | 4.3 | 0.08 | 30.5 | 0.49 | yes |
| FG478 | W2 | FG478_7P298 | ACGTTGGATGGATTTAGCGGCAGGCTTGAC | ACGTTGGATGTGATGAAGAGGAAGATCCCG | 122 | 96.7 | 80.9 | 45.4 | 31.6 | 4.3 | 0.08 | 30.5 | 0.49 | NO |
| FG487 | W2 | FG487_1000P96 | ACGTTGGATGTAGAGCAGCAGCAGTAACAG | ACGTTGGATGGAAGAACAGTGACACTTCGC | 103 | 100.0 | 80.9 | 47.7 | 52.9 | 2.2 | 0.43 | 29.5 | 0.49 | yes |
| FG543 | W2 | FG543_1P222 | ACGTTGGATGTTTTGAGGTCGCTGATCAGG | ACGTTGGATGATTGCTCACAAGCATGACCC | 118 | 98.7 | 80.9 | 48.3 | 38.1 | 2.2 | 0.42 | 27.4 | 0.48 | yes |
| FG673 | W2 | FG673_1P157 | ACGTTGGATGTTGAAGGATTTCCTGCGGAC | ACGTTGGATGTTGGTTGCATCCATCCCTTC | 101 | 100.0 | 80.9 | 48.4 | 47.4 | 2.2 | 0.4 | 33.7 | 0.49 | yes |
| FG673 | W2 | FG673_378P349 | ACGTTGGATGGGTACCTACCCTTATGTCAC | ACGTTGGATGCTTCGAGGATTCAATCCACC | 100 | 100.0 | 80.9 | 48.6 | 39.1 | 2.2 | 0.37 | 25.3 | 0.49 | NO |
| FG684 | W2 | FG684_1P103 | ACGTTGGATGAAGTCCTAATTATGAGATG | ACGTTGGATGGTCGTGGTATGTGAATGAGG | 102 | 79.5 | 80.9 | 47.0 | 24.0 | 2.2 | 0.21 | 20 | 0.22 | yes |
| FG685 | W2 | FG685_1P226 | ACGTTGGATGCGTGCCTTTTCTCACCTTTC | ACGTTGGATGAGAGCGGCAGATTAGCAAAG | 112 | 99.4 | 80.9 | 45.8 | 64.3 | 30.4 | 0.45 | 13.7 | 0.49 | NO |
| FG834 | W2 | FG834_642P234 | ACGTTGGATGTCACGAGTGTCTATCTTGGG | ACGTTGGATGAAGCTCCTGAAATTGCTGCC | 117 | 98.9 | 80.9 | 46.7 | 42.1 | 2.2 | 0.03 | 16.8 | 0.02 | yes |
| FG862 | W2 | FG862_1P45 | ACGTTGGATGATTTCCGATTCACCCTCTCC | ACGTTGGATGATCACCACGTTGAAATGCGG | 92 | 99.7 | 80.9 | 48.6 | 34.8 | 2.2 | 0.42 | 30.5 | 0.47 | yes |
| FG893 | W2 | FG893_1P237 | ACGTTGGATGAATGCTAGGTACGACGTCAC | ACGTTGGATGGTGCTTCTCTGCAATATTCG | 104 | 98.5 | 80.9 | 49.3 | 52.9 | 15.2 | 0.45 | 27.4 | 0.49 | yes |
| FG897 | W2 | FG897_1P184 | ACGTTGGATGGGCAGACGGAGAAGAAATTG | ACGTTGGATGCCGAGTAAAGAACAGAGAAG | 112 | 98.0 | 80.9 | 45.1 | 43.8 | 4.3 | 0.08 | 20 | 0.13 | yes |
| FG897 | W2 | FG897_1P343 | ACGTTGGATGGAAAGGTGCTGTTGGAGTTC | ACGTTGGATGAATTCAAGAGCTATCCTCTG | 98 | 94.3 | 80.9 | 46.8 | 31.8 | 4.3 | 0.08 | 21.1 | 0.21 | NO |
| MS307 | W2 | MS307_1P180 | ACGTTGGATGCTCGAGAGGCTGTAACAAAC | ACGTTGGATGGTGATGTCGGACTTCAATAC | 112 | 98.0 | 80.9 | 47.6 | 60.0 | 2.2 | 0.13 | 16.8 | 0.06 | yes |
| MS313 | W2 | MS313_50P250 | ACGTTGGATGGGATCAAACCCTTGACCAAC | ACGTTGGATGAAGCTCGACGAGTTGAGCTG | 99 | 98.6 | 80.9 | 51.4 | 47.6 | 23.9 | 0.37 | 10.5 | 0.49 | NO |
| MS313 | W2 | MS313_600P94 | ACGTTGGATGGCCTGTGGCATTAATTCCAG | ACGTTGGATGGGTGGGTCAAAAAAAGGCTC | 93 | 99.8 | 80.9 | 46.3 | 30.4 | 2.2 | 0.37 | 24.2 | 0.44 | yes |
| MS380 | W2 | MS380_312P34 | ACGTTGGATGGACCCATCTTGTGAGAGAAC | ACGTTGGATGTGCGCAGGCAATTTTTTACC | 99 | 98.6 | 80.9 | 45.3 | 30.4 | 13 | 0.03 | 13.7 | **0** | NO |
| MS397 | W2 | MS397_500P365 | ACGTTGGATGACTTGGTCTCCCAGGCATAG | ACGTTGGATGTCGCTGTTTATGGTCTAGAG | 100 | 97.1 | 80.9 | 52.0 | 71.4 | 2.2 | 0.13 | 41.1 | 0.44 | NO |
| MS408 | W2 | MS408_1P275 | ACGTTGGATGGCTGTGGATCTTGGTACTTG | ACGTTGGATGCATCCAAGATCCGAATGGAG | 108 | 99.7 | 80.9 | 47.7 | 44.4 | 28.3 | 0.03 | 15.8 | **0** | NO |
| MS424 | W2 | MS424_1P316 | ACGTTGGATGACAAGCTGGTGAGGAGTTAG | ACGTTGGATGCACATCATTATTCACTTCG | 128 | 85.2 | 80.9 | 47.6 | 29.2 | 2.2 | 0.4 | 17.9 | 0 | NO |
| MS437 | W2 | MS437_539P107 | ACGTTGGATGCCCGATCTATGGCTTCAAAC | ACGTTGGATGAGTTTTACAGTTGCCACCCC | 99 | 100.0 | 80.9 | 47.2 | 47.1 | 0 | 0.43 | 30.5 | 0.49 | yes |
| MS541 | W2 | MS541_1940P29 | ACGTTGGATGCCTCCACTAAAGTCACCAAC | ACGTTGGATGTTAGTGGAGCTGGTCGAGGT | 105 | 96.0 | 80.9 | 48.1 | 55.6 | 30.4 | 0.03 | 20 | 0.25 | NO |
| MS541 | W2 | MS541_1P152 | ACGTTGGATGGGCAAGCTGAAATCTTACGC | ACGTTGGATGGGAACCACAGTAGATTCTCC | 99 | 100.0 | 80.9 | 49.2 | 71.4 | 2.2 | 0.03 | 22.1 | 0.16 | yes |

**Gene**: Name of the targeted genes in Funybase (Mathey et al. 2008), **Plex**: first or second SNP MassArray, **SNP_ID**: name of the SNP, 2nd-PCRP: Secondary amplification primer ,**1st-PCRP**: Primary amplification primer, **AMP_LEN**: length of amplified sequence, **UP_CONF**: uniplex amplification score (quality of the amplicon design), **MP_CONF**: multiplex amplification score, **Tm**: Temperature of hybridization, **PcGC**: percent GC content of the extend primer, **NaIon**: percentage of missing data obtained from Ion-Torrent sequencing, **HeIon**: Gene diversity from Ion- Torrent sequencing data, **NaSeq**: percentage of missing data obtained from MassArray sequencing, **HeSeq**: Gene diversity from MassArray sequencing, **Pganalysis**: SNP used for population genetic analysis.

Most of monospore isolates (95%) had less than three missing data, whereas 93% of the lesion and herbarium samples had more than five missing data (mean 11.8 and 33.1 per isolate for leaf and herbarium samplings respectively; Fig. 2). In addition, nearly all the lesions sampled from the field were diploid with at least one heterozygote SNP identified (Table S1) indicating that these lesions were likely a mix of (at least) two distinct haploid genotypes. It was therefore not possible to identify haplotypes, and to estimate allelic frequencies and genetic diversity in these two samplings.

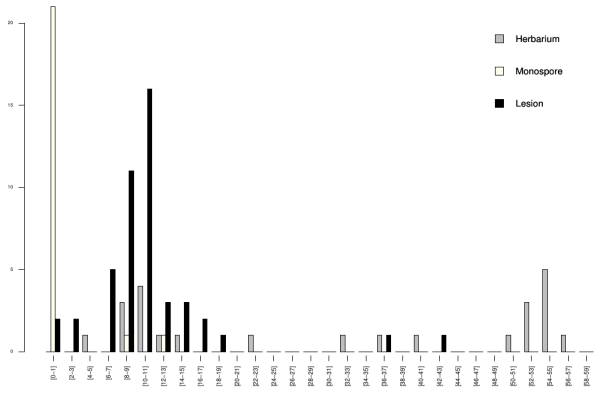


Figure 2. **Number of missing SNPs using the MassArray genotyping technology per type of sampling in *Erysiphe alphiltoides*.**

Out of the 58 SNPs genotyped by MassArray, we finally selected 36 SNPs from 36 distinct genes to remove the effect of physically linked SNPs within genes, for population genetic analysis (Table 3). Based on this subset of SNPs, we inferred no repeated genotype among the 34 *E. alphitoides* monospore isolates analyzed by the Ion Torrent sequencing. The average genetic diversity estimated in this sample was 0.37 (SD = 0.15), and the standardized index of association (*rD*), measuring the multilocus linkage disequilibrium (Agapow and Burt 2001) was 0.027, not significantly different from zero (*P-value* = 0.002). For *E. quercicola*, using a subset of 33 SNPs located in 33 genes detected in the 13 monospore isolates chosen for the SNPs identification, we found two isolates with the same genotype and two pairs of genotypes distinct by a single SNP. We estimated a genetic diversity of 0.24 (SD = 0.09) and *rD* equals to 0.11, significantly different from zero (*P-value* < 0.01).

## DISCUSSION

This study showed that reliable genetic markers can be quickly isolated from genomes hard to assemble such as those of *Erysiphales* (Spanu et al. 2010, Frantzeskakis et al. 2018). As obligate biotrophs non culturable on axenic media, these fungi are also more likely to be contaminated by an unknown source of micro-organisms since they are maintained on plant material. The method developed here allows to deal simply and efficiently with contamination by selecting sequences phylogenetically close to the target species. However, the E-value cut-off should be carefully tested in case of close proximity between target and putative contaminant species. As illustrated in our study, the choice of a relaxed E-value threshold for tBlastN (i.e. 10E-20) finally selected a few sequences unrelated to our target species. Most of the non-amplified sequences were associated to *Aspergillus* sp., likely present on the oak leaves from which *E. alphitoid*es was sampled (Unterseher et al. 2007). The phylogenetic analysis, such as presented in our study, should be performed before the final choice of candidate regions to better remove genes from contaminating taxa. Our method may also be applied to any other fungus since genomic resources can easily be generated today using high throughput sequencing. Moreover, gene databases are available for several phyla (Grigoriev et al. 2014).

Our approach may provide an alternative to the RAD-seq method since, on one hand, it does not require complex laboratory protocols (Harvey et al. 2016) and, on the other hand, it allows to easily control both the phylogenetic origin and duplications of the targeted sequences. The choice of focusing on a set of conserved

genes, generally present in single copy in most fungal genomes should prevent the risk of amplifying paralogs, reducing errors in estimation of allelic frequencies (Gayral et al. 2013). Controlling for gene duplicates with our method is dependent on the selected threshold for the Blast analysis (see for example Feau et al. 2011), and quality of the genome assembly. However, the results obtained from the MassArray genotyping showed no evidence of copy number variations for the chosen SNP (data not shown), and especially, we detected heterozygote SNPs for only one of the 23 monospore haploid isolates used as controls, consistent with absence of duplication of the chosen genes. Although it is likely that large parts of the *E. alphitoides* genome, especially regions of low complexity, have probably not been accurately assembled (Frantzeskakis et al. 2018), our results suggest that our targeted single-copy genes are frequently located in the easiest regions to assemble, greatly helping their isolation in draft genomes. In addition, we showed that a small quantity of DNA (which is usually what is obtained for obligate fungal biotrophs, like OPM) is not a limitation for the isolation of these targeted markers by using this method. This method might further be improved by increasing the depth of sequencing, by using long-read sequencing technologies which would allow better assembly of full genomes (Faino et al. 2015). Using additional gene databases to identify conserved genes such as OrthoMCL-DB (Li et al. 2003) might also allow to increase the number of targeted genes. Nevertheless, even a limited number of SNPs (i.e. around 40, such as obtained in this study) may allow to get a first insight on population genetic structure (e.g. Dutech et al. 2017, Tsykun et al. 2017). The study published by Tsykun et al. (2017) showed no major difference in population structure characterized with SSR or SNP markers.

DNA amplification using the Fluidigm methodology made it possible to obtain numerous SNPs for a preliminary population genetic study in *E. alphitoides*, and its sister species *E. quercicola*. A limitation of our study was due to the use of the Ion Torrent sequencing methodology which generates a large number of spurious indels linked to homopolymers (Loman et al. 2012). Our in-house script designed to removed these indels from the consensus sequences built for each individual, seemed to efficiently solve this problem. Actually, the validation of more than 90 % of the selected SNPs following MassArray genotyping confirmed that our method is robust. In addition, 72 % of the detected SNPs are fixed in each species, as expected in absence of recent gene flow between the two *Erysiphe* species (Feau et al. 2011), and support they are not false positive SNPs. By contrast, we also used another method implemented in DiscoSNP (Uricaru et al. 2015) which detected less than 300 SNPs, suggesting that numerous SNPs are missed relative to our method of SNP calling. Additional tests should be performed to define the best set of parameters of DiscoSNP on such sequence data with a large presence of indels generated by Ion Torrent sequencing, but these results suggest that this automatic method of SNP calling is unadapted to these sequence data.

Using 33 to 36 SNPs located in different genes, we could perform a preliminary population genetic study for the two main OPM species introduced in Europe. The genotyping of 36 SNPs with the MassArray genotyping in 34 *E. alphitoides* isolates from five regions in Europe, detected no repeated genotypes and absence of significant linkage disequilibrium between SNPs suggesting frequent sexual crossing and gene flow among populations. By contrast, in *E. quercicola*, several isolates had the same or a very close genotype and a significant deviation from zero was obtained for the standardized index of association. Although these results were obtained on a limited sampling, these findings are consistent with biological observations suggesting the quasi- absence of sexual reproduction in *E. quercicola*, which overwinters as mycelium and conidia in buds (Feau et al. 2012). The sister species *E. alphitoides* does sexual reproduction and differentiates chasmothecia, the sexual reproductive structures that also act as resistant structures (Feau et al. 2012). These first results should be further confirmed by a genetic analysis on a more thorough sampling at both local and regional geographical scales by using the SNPs isolated in this study. Our results on the lesions sampled directly on naturally infected oak leaves without monospore isolation also showed that field colonies of OPM often result from mixed infections of at least two distinct genotypes. A similar finding was reported for the Plantago powdery mildew (Tollenaere et al. 2012), whereas in a study performed on *E. necator*, the causal agent of powdery mildew on grapevines, only one genotype was generally detected in field lesions (Kisselstein et al. 2018). It would be interesting to further investigate whether these findings reveal differences in pathogen genetic structure between natural ecosystems and crops, with strong directional pressures in the latter maybe associated with lower diversity in powdery mildew populations. Our generic method could make it possible to study a wider range of powdery mildew species. From a practical point of view, the finding of mixed infections may hamper any population genetic study requiring the estimate of allele frequencies without prior monospore isolation and subculturing, which is a delicate and time-consuming step. This difficulty might be circumvented by an early sampling in the vegetative season when single colonies, putatively resulting from a single spore infection, are still clearly distinguishable.

## LITERATURE


Agapow PM, Burt A (2001). Indices of multilocus linkage disequilibrium. Mol Ecol Notes 1: 101–102.

Altschul S,Gish W, Miller W, et al. (1990) Basic Local alignment search tool. J. Mol. Biol. 215: 403–410. https://doi.org/10.1006/jmbi.1990.9999

Amselem J, Cuomo CA, van Kan JAL, et al. (2011) Genomic Analysis of the Necrotrophic Fungal Pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. PloS Genet. 7: e1002230 https://doi.org/10.1371/journal.pgen.1002230

Andrews S. (2010) FastQC: a quality control tool for high throughput sequence data. Available online http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Arnaud-Haond S, Duarte CM, Alberto F, Serrao EA (2007) Standardizing methods to address clonality in population studies. Mol. Ecol. 16:5115–5139. https://doi.org/10.1111/j.1365-294X.2007.03535.x

Barrès B, Dutech C, Andrieux A, Halkett F, Frey P (2012) Exploring the role of asexual multiplication in poplar rust epidemics: impact on diversity and genetic structure. Mol. Ecol. 21: 4996-5008. https://doi.org/10.1111/mec.12008

Chancerel E, Lamy J-B, Lesur I, et al. (2013) High-density linkage mapping in a pine tree reveals a genomic region associated with inbreeding depression and provides clues to the extent and distribution of meiotic recombination. BMC Biol. 11:50 https://doi.org/10.1186/1741-7007-11-50

Chapuis M-P, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. Mol. Biol. Evol. 24:621–631. https://doi.org/10.1093/molbev/msl191

Desprez-Loustau M-L, Feau N, Mougou-Hamdane A, Dutech C (2011) Interspecific and intraspecific diversity in oak powdery mildews in Europe: coevolution history and adaptation to their hosts. Mycoscience. 52:165–173. https://doi.org/10.1007/s10267-010-0100-5

Desprez-Loustau M-L, Aguayo J, Dutech C, et al. (2016) An

evolutionary ecology perspective to address forest pathology challenges of today and tomorrow. Ann. For. Sci. 73:45–67. https://doi.org/10.1007/s13595-015-0487-4

Desprez-Loustau M-L, Massot M, Feau N, et al. (2017) Further Support of Conspecificity of Oak and Mango Powdery Mildew and First Report of *Erysiphe quercicola* and *Erysiphe alphitoides* on Mango in Mainland Europe. Plant Dis. 101:1086–1093. https://doi.org/10.1094/PDIS-01-17-0116-RE

Desprez-Loustau M-L, Massot M, Toigo M, et al. (2018) From leaf to continent: The multi-scale distribution of an invasive cryptic pathogen complex on oak. Fung. Ecol. 36:39–50. https://doi.org/10.1016/j.funeco.2018.08.001

Dutech C, Enjalbert J, Fournier E, et al. (2007) Challenges of microsatellite isolation in fungi. Fung. Genet. Biol. 44:933–949. https://doi.org/10.1016/j.fgb.2007.05.003

Dutech C., Prospero S., Heinzelmann R., Fabreguettes O., Feau N. (2016) Rapid identification of polymorphic sequences for both phylogenetic and population genetic analyses in non-model species: The PHYLORPH method tested in Armillaria species. For. Pathol. 46:298-308. https://doi.org/10.1111/efp.12256

Dutech C, Labbe F, Capdevielle X, Lung-Escarmant B (2017) Genetic analysis reveals efficient sexual spore dispersal at a fine spatial scale in Armillaria ostoyae, the causal agent of root-rot disease in conifers. Fung. Biol. 121:550–560. https://doi.org/10.1016/j.funbio.2017.03.001

Edgar R (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. 32:1792–1797. https://doi.org/10.1093/nar/gkh340

Ellis JS, Gilbey J, Armstrong A, et al. (2011) Microsatellite standardization and evaluation of genotyping error in a large multi-partner research programme for conservation of Atlantic salmon (Salmo salar L.). Genetica 139:353–367. https://doi.org/10.1007/s10709-011-9554-4

Elshire RJ, Glaubitz JC, Sun Q, et al. (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLOS One 6.: e19379 https://doi.org/10.1371/journal.pone.0019379

Estoup A, Jarne P, Cornuet J (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Mol. Ecol. 11:1591–1604. https://doi.org/10.1046/j.1365-294X.2002.01576.x

Faino L, Seidl MF, Datema E, et al. (2015) Single-Molecule Real-Time Sequencing Combined with Optical Mapping Yields Completely Finished Fungal Genome. MBIO 6.: e00936-15 https://doi.org/10.1128/mBio.00936-15

Feau N, Beauseigle S, Bergeron M-J, et al. (2018) Genome-Enhanced Detection and Identification (GEDI) of plant pathogens. PEERJ 6: e4392 https://doi.org/10.7717/peerj.4392

Feau N, Decourcelle T, Husson C, et al. (2011) Finding Single Copy Genes Out of Sequenced Genomes for Multilocus Phylogenetics in Non-Model Fungi. PLOS One 6.: e0018803 https://doi.org/10.1371/journal.pone.0018803

Feau N, Lauron-Moreau A, Piou D, et al. (2012) Niche partitioning of the genetic lineages of the oak powdery mildew complex. Fung. Ecol. 5:154–162. https://doi.org/10.1016/j.funeco.2011.12.003

Feehan JM, Scheibel KE, Bourras S, et al. (2017) Purification of High Molecular Weight Genomic DNA from Powdery Mildew for Long-Read Sequencing. JOVE-J. Visual. Experim. https://doi.org/10.3791/55463

Frantzeskakis L, Kracher B, Kusch S, et al. (2018) Signatures of host specialization and a recent transposable element burst in the dynamic one-speed genome of the fungal barley powdery mildew pathogen. BMC Genomics 19:381 https://doi.org/10.1186/s12864-018-4750-6

Fu L, Niu B, Zhu Z, et al. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Gayral P, Melo-Ferreira J, Glemin S, et al. (2013) Reference-Free Population Genomics from Next-Generation Transcriptome Data and the Vertebrate-Invertebrate Gap. PLOS Genet. 9: e1003457 https://doi.org/10.1371/journal.pgen.1003457

Gladieux P, Feurtey A, Hood ME, et al. (2015) The population biology of fungal invasions. Mol. Ecol. 24:1969–1986. https://doi.org/10.1111/mec.13028

Griffon E, Maublanc A. (1912) Les *Microsphaera* des Chênes. Bull. Soc. Mycol. F. 28: 88–104.

Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar 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 Res. 42:699-704

Gross A, Holdenrieder O, Pautasso M, et al. (2014) *Hymenoscyphus pseudoalbidus*, the causal agent of European ash dieback. Mol. Plant Pathol. 15:5–21. https://doi.org/10.1111/mpp.12073

Guindon S, Dufayard J-F, Lefort V, et al. (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst. Biol. 59:307–321. https://doi.org/10.1093/sysbio/syq010

Hamelin FM, Bisson A, Desprez-Loustau M-L, et al. (2016) Temporal niche differentiation of parasites sharing the same plant host: oak powdery mildew as a case study. Ecosphere 7.: https://doi.org/10.1002/ecs2.1517

Hardy O, Charbonnel N, Freville H, Heuertz M (2003) Microsatellite allele sizes: A simple test to assess their significance on genetic differentiation. Genetics 163:1467–1482

Harvey MG, Smith BT, Glenn TC, et al. (2016) Sequence Capture versus Restriction Site Associated DNA Sequencing for Shallow Systematics. Syst. Biol. 65:910–924. https://doi.org/10.1093/sysbio/syw036

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27:3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Kamvar ZN, Brooks JC, Gruenwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front. Genet. 6:208 https://doi.org/10.3389/fgene.2015.00208

Kaplinski L, Andreson R, Puurand T, Remm M (2005) MultiPLX: automatic grouping and evaluation of PCR primers. Bioinformatics 21:1701–1702. https://doi.org/10.1093/bioinformatics/bti219

Kisselstein B, Cadle-Davidson LE, Gadoury DM (2018) Investigation of *Erysiphe necator* Population Structure using Amplicon Sequencing (AmpSeq) without Clonal Isolation. Phytopath. 108:96

Lesur I, Le Provost G, Bento P, et al. (2015) The oak gene expression atlas: insights into Fagaceae genome evolution and the discovery of genes regulated during bud dormancy release. BMC Genomics 16:112

https://doi.org/10.1186/s12864-015-1331-9

Li L, Stoeckert C, Roos D (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res. 13:2178–2189. https://doi.org/10.1101/gr.1224503

Loman NJ, Misra RV, Dallman TJ, et al. (2012) Performance comparison of benchtop high-throughput sequencing platforms. Nat. Biotech. 30:434. https://doi.org/10.1038/nbt.2198

Marcais B, Desprez-Loustau M-L (2014) European oak powdery mildew: impact on trees, effects of environmental factors, and potential effects of climate change. Ann. For. Sci. 71:633–642. https://doi.org/10.1007/s13595-012-0252-x

Marthey S, Aguileta G, Rodolphe F, et al. (2008) FUNYBASE: a FUNgal phYlogenomic dataBASE. BMC Bioinformatics 9.:e456 https://doi.org/10.1186/1471-2105-9-456

Meirmans PG (2012) The trouble with isolation by distance. Mol. Ecol., 21: 2839–2846.

Meirmans PG (2015) Seven common mistakes in population genetics and how to avoid them. Mol. Ecol. 24:3223–3231. https://doi.org/10.1111/mec.13243

Mougou A, Dutech C, Desprez-Loustau M-L (2008) New insights into the identity and origin of the causal agent of oak powdery mildew in Europe. For. Pathol. 38:275–287. https://doi.org/10.1111/j.1439-0329.2008.00544.x

Puritz JB, Matz MV, Toonen RJ, et al. (2014) Demystifying the RAD fad. Mol. Ecol. 23:5937–5942. https://doi.org/10.1111/mec.12965

Queloz V, Gruenig CR, Berndt R, et al. (2011) Cryptic speciation in *Hymenoscyphus albidus*. For. Pathol. 41:133–142. https://doi.org/10.1111/j.1439-0329.2010.00645.x

Ribeiro A, Golicz A, Hackett CA, et al. (2015) An investigation of causes of false positive single nucleotide polymorphisms using simulated reads from a small eukaryote genome. BMC Bioinformatics 16:382 https://doi.org/10.1186/s12859-015-0801-z

Rouxel M, Mestre P, Comont G, et al. (2013) Phylogenetic and experimental evidence for host-specialized cryptic species in a biotrophic oomycete. New Phytol. 197:251–263. https://doi.org/10.1111/nph.12016

Santini A, Ghelardini L, De Pace C, et al. (2013) Biogeographical patterns and determinants of invasion by forest pathogens in Europe. New Phytol. 197:238–250. https://doi.org/10.1111/j.1469-8137.2012.04364.x

Schoch CL, Seifert KA, Huhndorf S, et al. (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. PNAS 109:6241–6246. https://doi.org/10.1073/pnas.1117018109

Selkoe K, Toonen R (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol. Lett. 9:615–629. https://doi.org/10.1111/j.1461-0248.2006.00889.x

Spanu PD, Abbott JC, Amselem J, et al. (2010) Genome Expansion and Gene Loss in Powdery Mildew Fungi Reveal Tradeoffs in Extreme Parasitism. Science 330:1543–1546. https://doi.org/10.1126/science.1194573

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 https://doi.org/10.1186/1471-2105-7-62

Takamatsu S, Braun U, Limkaisang S, et al. (2007) Phylogeny and taxonomy of the oak powdery mildew *Erysiphe alphitoides* sensu lato. Mycol. Res. 111:809–826. https://doi.org/10.1016/j.mycres.2007.05.013

Takamatsu S, Ito (Arakawa) H, Shiroya Y, et al. (2015) First comprehensive phylogenetic analysis of the genus *Erysiphe* (*Erysiphales*, *Erysiphaceae*) I. The *Microsphaera* lineage. Mycologia 107:475–489. https://doi.org/10.3852/15-007

Tollenaere C, Susi H, Nokso-Koivisto J, et al. (2012) SNP Design from 454 Sequencing of Podosphaera plantaginis Transcriptome Reveals a Genetically Diverse Pathogen Metapopulation with High Levels of Mixed-Genotype Infection. PLOS One 7:e52492 https://doi.org/10.1371/journal.pone.0052492

Tonnabel J, Olivieri I, Mignot A, et al. (2014) Developing nuclear DNA phylogenetic markers in the angiosperm genus Leucadendron (Proteaceae): A next-generation sequencing transcriptomic approach. Mol. Phylogenet. Evol. 70:37–46. https://doi.org/10.1016/j.ympev.2013.07.027

Tsykun T, Rellstab C, Dutech C, et al. (2017) Comparative assessment of SSR and SNP markers for inferring the population genetic structure of the common fungus *Armillaria cepistipes*. Heredity 119:371–380. https://doi.org/10.1038/hdy.2017.48

Tsykun T, Rigling D, Prospero S (2013) A new multilocus approach for a reliable DNA-based identification of *Armillaria* species. Mycologia 105:1059–1076. https://doi.org/10.3852/12-209

Untergasser A, Cutcutache I, Koressaar T, et al. (2012) Primer3-new capabilities and interfaces. Nucleic Acids Res. 40:e115 https://doi.org/10.1093/nar/gks596

Unterseher M, Reiher A, Finstermeier K, et al. (2007) Species richness and distribution patterns of leaf-inhabiting endophytic fungi in a temperate forest canopy. Mycol. Progress. 6:201–212. https://doi.org/10.1007/s11557-007-0541-1

Uricaru R, Rizk G, Lacroix V, et al. (2015) Reference-free detection of isolated SNPs. Nucleic Acids Res. 43:e11 https://doi.org/10.1093/nar/gku1187

Verdu CF, Guichoux E, Quevauvillers S, et al. (2016) Dealing with paralogy in RADseq data: in silico detection and single nucleotide polymorphism validation in *Robinia pseudoacacia* L. Ecol. Evol. 6:7323–7333. https://doi.org/10.1002/ece3.2466

Wingfield MJ, Slippers B, Wingfield BD, Barnes I (2017) The unified framework for biological invasions: a forest fungal pathogen perspective. Biol. Invasions 19:3201–3214. https://doi.org/10.1007/s10530-017-1450-0

## ACKNOWLEDGMENT

We thank Olivier Fabreguettes, Gilles Saint-Jean and Martine Martin-Clotté for their help in the laboratory, especially for *Erysiphe alphitoides* sub-cultures, and DNA extraction, Nacer Mohellibi for helping in the genome browser, and Françoise Alfama-Depauw for blast banks, using the facilities of the URGI platform (https://urgi.versailles.inra.fr/). MassArray, Access and Ion Torrent experiments were performed at the Genome Transcriptome Facility of Bordeaux (Grants from Investissements d'Avenir, Convention attributive d'aide EquipEx Xyloforest ANR-10-EQPX-16-01), with the help of Christophe Boury. Illumina sequencing was performed in collaboration with the GeT core facility, Toulouse, France (http://get.genotoul.fr), and was supported by France Génomique National infrastructure, funded as part of "Investissement d'avenir" program managed by Agence

Nationale pour la Recherche (contract ANR-10-INBS-09). Andrin Gross benefited from the IdEx Bordeaux post-doctoral fellowship programme. This work was also supported by the ANR-12-ADAP-0009 (Gandalf project), and an innovative project INRA department EFPA. Part of the sampling was issued from a European BiodivERsA project (RESIPATH: Responses of European Forests and Society to Invasive Pathogens" – BIODIVERSA/0002/2012) with a national grant ANR-13-EBID-0005-01. We are also grateful to the genotoul bioinformatics platform Toulouse Midi-Pyrénées (Bioinfo Genotoul) for providing computing and storage resources. We finally thank colleagues from RESIPATH and the herbaria curators D. Triebel (Bot. Staatssammlungen Munich, Germany), C. Lange (Natural history museum of Denmark), A. D. Bond (Royal Botanical gardens fungarium, United Kingdom), R. Berndt (Fungarium Z + ZT, Switzerland),V. Queloz and V. Dubach (forest protection service herbarium, WSL, Switzerland), and Nathalie Séjalon-Delmas et Paul Seimandi (Paul Sabatier Toulouse university, France) for *Erysiphe* sp. samples used in the genotyping test.

**DATA AVAILABILITY**:

Raw sequencing data of the draft genome are available on the Bioproject PRJNA593204. The first assembly, contigs > 10 Kb homologs to *Blumeria graminis*, and predicted genes are available on arachne.pierroton.inra.fr/AlphiGeno. *Erysiphe alphitoides* Jbrowse is available on https://urgi.versailles.inra.fr/jbrowse/gmod_jbrowse/?data=myData/AlphiGeno. Raw Ion Torrent sequencing data, scripts for detecting SNP, reference sequences for each amplicon from the *E. alphitoides* draft genome, MassArray genotyping data, and suplementary tables of this paper are available on https://doi.org/10.15454/UGMTBK.