



**HAL**  
open science

## Europe as a bridgehead in the worldwide invasion history of grapevine downy mildew, *Plasmopara viticola*

Michaël Fontaine, Frédéric Labbé, Yann Dussert, Laurent Delière, Sylvie Richart-Cervera, Tatiana E Giraud, François Delmotte

### ► To cite this version:

Michaël Fontaine, Frédéric Labbé, Yann Dussert, Laurent Delière, Sylvie Richart-Cervera, et al.. Europe as a bridgehead in the worldwide invasion history of grapevine downy mildew, *Plasmopara viticola*. *Current Biology - CB*, 2021, 31 (10), pp.2155-2166.e4. 10.1016/j.cub.2021.03.009 . hal-03184119

**HAL Id: hal-03184119**

**<https://hal.science/hal-03184119>**

Submitted on 29 Mar 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Europe as a bridgehead in the worldwide invasion history of grapevine downy mildew, *Plasmopara viticola*

Michael C. Fontaine<sup>1,2,3\*†</sup>, Frédéric Labbé<sup>1,2\*‡</sup>, Yann Dussert<sup>4§</sup>, Laurent Delière<sup>4</sup>, Sylvie Richart-Cervera<sup>4</sup>, Tatiana Giraud<sup>2</sup>, François Delmotte<sup>4†</sup>

1. Groningen Institute for Evolutionary Life Sciences (GELIFES), University of Groningen, PO Box 11103 CC, Groningen, The Netherlands
2. Ecologie Systématique et Evolution, UMR 8079, Université Paris-Saclay, CNRS, AgroParisTech, Orsay, F-91400, France
3. Laboratoire MIVEGEC (Université de Montpellier, CNRS 5290, IRD 229) et Centre de Recherche en Écologie et Évolution de la Santé (CREES), Institut de Recherche pour le Développement (IRD), F-34394, Montpellier, France
4. SAVE, INRAE, Bordeaux Sciences Agro, Université de Bordeaux, F-33140, Villenave d'Ornon, France

## ORCID numbers:

- MCF: <https://orcid.org/0000-0003-1156-4154>
- FL: <https://orcid.org/0000-0002-4064-2361>
- YD: <https://orcid.org/0000-0003-1900-7715>
- TG: <https://orcid.org/0000-0002-2685-6478>
- FD: <https://orcid.org/0000-0002-7138-8239>

† Corresponding authors. Emails: [michael.fontaine@cnrs.fr](mailto:michael.fontaine@cnrs.fr) and [francois.delmotte@inrae.fr](mailto:francois.delmotte@inrae.fr)

‡ Present address: Department of Ecology and Evolution, University of Chicago, IL, USA.

§ Present address: School of Biological and Chemical Sciences, Queen Mary University of London, London, UK.

\* These authors contributed equally to the study.

**Short title:** Global invasion history of grapevine downy mildew

**Keywords:** fungi, invasive plant pathogen, microsatellite, oomycetes, population genetics, recent introduction, *Vitis vinifera*, approximate Bayesian computation, emerging crop disease

5 **Author contributions:** FD, TG and MCF designed the study; FD and LD collected the biological materials; SRC conducted the molecular work in the laboratory and collected the data under the supervision of FD; MCF and FL analyzed the data with help from YD and FD; MCF, FL, TG and FD interpreted the results; MCF, TG, FL and FD wrote the manuscript, with feedback and final approval from all the co-authors.

10

## Abstract

Europe is the historical cradle of viticulture, but grapevines have been increasingly threatened by pathogens of American origin. The invasive oomycete *Plasmopara viticola* causes downy mildew, one of the most devastating grapevine diseases worldwide. Despite major economic consequences, its invasion history remains poorly understood. Comprehensive population genetic analyses of ~2000 samples from the most important wine-producing countries revealed very low genetic diversity in invasive downy mildew populations worldwide. All the populations originated from one of five native North American lineages, the one parasitizing wild summer grape. After an initial introduction into Europe, invasive European populations served as a secondary source of introduction into vineyards worldwide, including China, South Africa and, twice independently, Australia. Invasion of Argentina probably represents a tertiary introduction from Australia. Our findings provide a striking example of a global pathogen invasion resulting from secondary dispersal of a successful invasive population. It will help designing quarantine regulations and efficient breeding for resistance against grapevine downy mildew.

## Introduction

Global changes (e.g., climate warming and international exchanges) are favoring increases in the numbers of emerging diseases caused by invasive pathogens on crops worldwide, incurring substantial economic, social and environmental costs (1-3). Infamous recent examples include the emergence of new races of the stem rust fungus in Eastern Africa (4) and of the fungus causing wheat blast disease in Bangladesh (5), both threatening wheat production and becoming invasive. Most emerging diseases result from biological invasions bringing the native host-parasite association back together after crop introduction into new areas, or host shifts following pathogen introductions (6, 7). An understanding of the evolutionary processes responsible for emerging crop diseases is important for preventing further devastating biological invasions and for controlling introduced populations. This requires elucidation of the invasion mechanisms, pathways and demographic processes occurring during pathogen invasions (e.g., bottlenecks and hybridization). Important questions include whether pathogen invasions result from host shifts, whether limited genetic variation has been introduced, whether multiple introductions and admixture are required for successful invasions (7), and whether the invaded areas are colonized directly from native populations or whether an initial successful invasive population serves as the source for secondary introductions.

Grapevine downy mildew, caused by the oomycete *Plasmopara viticola*, is one of the most damaging diseases of grapevine, and is found in all grape-growing regions of the world (8). *Plasmopara viticola* is native to eastern North America (9), and was introduced into Europe in the 1870s (10), probably with American *Vitis* species used as rootstocks resistant to phylloxera, an insect pest (*Daktulosphaira vitifoliae*) (9, 11). After its first description in France (1878 in

Coutras), the disease rapidly reached Southern and Central Europe and was soon reported in nearly all wine-producing countries worldwide (12, 13).

Advances in *P. viticola* -omics (14-19) and population genetic studies (20-35) have improved our knowledge of the grapevine downy mildew pathosystem. In its native range, five cryptic species (also called *formae speciales*) have recently been identified in the *P. viticola* species complex, with genetic differentiation and contrasting host ranges on various *Vitis* and related species (33, 34). The five *P. viticola formae speciales* (f. sp.) are found on wild *Vitis* species across North America (33, 34), and it remains unknown which lineages were responsible for grapevine downy mildew invasions in vineyards across the world. In most temperate regions, *P. viticola* populations present widespread footprints of recombination, indicating the occurrence of frequent sexual reproduction (21, 24, 26, 27, 29, 31, 36). European invasive populations display little genetic diversity and have a weak but significant population structure at the continental scale (22, 25). Despite some notable advances, however, these studies have been restricted to a small number of countries and were performed with different markers, making it difficult to develop a comprehensive understanding of the pathways of *P. viticola* invasion worldwide.

In this study, we used phylogenetic and population genetic approaches, together with scenario testing by approximate Bayesian computation (ABC), to infer the routes of *P. viticola* invasion worldwide. We analyzed almost 2,000 *P. viticola* samples, collected from wild and cultivated grapes in Northeast America, and from the main grape-growing regions in which grapevine downy mildew occurs. Using nuclear and mitochondrial gene sequences, we found that all invasive grapevine downy mildew populations worldwide belonged to a small clade of the species *P.*

*viticola* f. sp. *aestivalis*, which parasitizes the *V. aestivalis* summer grape in Northeast America (Supplementary Photo S1). Using microsatellite markers on *P. viticola* f. sp. *aestivalis*, we found that there was little admixture between populations and that invasive populations had a low genetic diversity. The European populations had the highest level of diversity of all invasive populations and harbored all but one of the haplotypes present in other invaded areas. Using ABC scenario testing, we confirmed that *P. viticola* f. sp. *aestivalis* was first introduced into Western Europe, whence it spread to Central and Eastern Europe. The successful invasive populations in Europe then served as the source for secondary introductions into other grape-growing regions of the world, such as Northeast China, South Africa, and Australia. A third bridgehead effect occurred later, with the spread of the disease from Southeastern Australia to Argentina. All grapevine downy mildew invasions therefore stem from an initial single introduction event in Europe, followed by secondary and tertiary introductions via bridgehead effects. These introductions were probably mediated by the transfer of grapevine material by humans, as settlers imported European cultivars for the establishment of “New World” vineyards during the 19th century (37). Our findings of a strong bottleneck following the introduction into Europe and the common origin of all introduced populations worldwide provide essential knowledge for guiding breeding for resistance to grapevine downy mildew. The identification of these historical pathways also improves our understanding of biological pest and pathogen invasions.

## 20 **Results**

### **The cryptic species *Plasmopara viticola* f. sp. *aestivalis* is the origin of all invasive downy mildew populations worldwide**

We reconstructed the genetic relationships between an extensive set of *P. viticola* strains collected from wild or cultivated grape species in native areas and areas of introduction, by sequencing DNA

fragments from the mitochondrial cytochrome-*b* (*cytb*,  $n = 1,299$ ),  $\beta$ -tubulin (*tub*,  $n = 424$  samples) and ribosomal 28S (*r28S*,  $n = 536$ ) genes (Fig. 1, Supplementary Fig. S1 and Table S1). The  $\beta$ -tubulin tree provided the highest resolution (Fig. 1A, with 56 distinct haplotypes versus 32 haplotypes for *cytb* and 9 haplotypes for *r28S*, Supplementary Fig. S2). Nevertheless, all the trees presented highly differentiated lineages with contrasting host ranges (Fig. 1A, Supplementary Fig. S2), consistent with the previously demonstrated existence of five cryptic species (33, 34). Three *formae speciales* were found on cultivated grapes in North America, but only *P. viticola* f. sp. *aestivalis* was found in introduced populations worldwide (Fig. 1 and Supplementary Fig. S2). In North America, this *forma specialis* was found on *V. aestivalis* and on *V. labrusca sensu lato* (i.e. including *V. labrusca* and its main artificial hybrids), two *Vitis* species that have recently diverged (38).

### **A host shift from *V. aestivalis* at the origin of the invasive downy mildew populations and population subdivision in invasive populations**

We used highly polymorphic microsatellite markers to obtain a finer resolution of the population structure within *P. viticola* f. sp. *aestivalis*, to elucidate its worldwide population structure and invasion routes. We genotyped 1,974 diploid strains with eight microsatellite markers (Supplementary Table S2), which, together with the three sequenced fragments (Supplementary Table S1), revealed 1,383 distinct genotypes with this 11-marker dataset. We found that the native and introduced *P. viticola* f. sp. *aestivalis* populations were strongly differentiated, along the first axis of a discriminant analysis on principal components (DAPC) (39) (Supplementary Fig. S3). The second axis revealed a strong differentiation in North America between *P. viticola* f. sp. *aestivalis* populations collected on different *Vitis* species (Supplementary Fig. S3). Indeed, we identified specific genetic clusters on *V. aestivalis* (in yellow), *V. labrusca* (in black), and two



differentiated clusters (brown and red) on the cultivated grape *V. vinifera* (Supplementary Fig. S3). The cluster on cultivated grapes (in brown) was genetically close to the cluster on the wild species *V. labrusca* (in black), probably indicating a host shift from *V. labrusca* in the native range. We inferred that the invasive populations originated from a host shift from *V. aestivalis*, as the population from this wild summer grape (yellow cluster) appeared genetically closest, and even overlapping in the DAPC, with the invasive populations (Supplementary Fig. S3A). However, the eight microsatellite markers had a low power to resolve the genetic structure between invasive populations (Supplementary Fig. S3C).

10 We therefore genotyped a subset of strains ( $n = 181$ ) with 32 microsatellite markers, focusing on invasive populations and on the native *P. viticola* f. sp. *aestivalis* clusters that may have served as origin of the invasive populations (Supplementary Table S3, 174 unique genotypes). This 32-marker dataset confirmed the genetic patterns observed in the native range based on the 11-marker dataset, and, in particular, confirmed that the cluster on *V. aestivalis* (yellow) was the likely origin

15 of all invasive populations worldwide (Fig. 2). This cluster was, again, the closest to all the invasive populations on the DAPC (Fig. 2D). Furthermore, using the 32-marker dataset, we showed that the cluster on *V. aestivalis* was the only cluster having common genetic ancestry with the invasive populations in the STRUCTURE Bayesian clustering analysis (40, 41) (see the light blue, green or pink ancestry from  $K = 2$  to  $K = 6$  in genotypes collected from *V. aestivalis* on Fig.

20 2A). With the 32-marker dataset, we also showed that this was the closest cluster to all the invasive populations in the distance-based neighbor-joining tree (Fig. 2C). The 32-marker dataset further identified two genetic clusters on *V. aestivalis* in the native range (Fig. 2A): the yellow cluster identified above, genetically close to the invasive populations found on cultivated grapes worldwide (Fig. 2D), and the red genetic cluster, corresponding to strains collected on cultivated

grapes in North America (Fig. 2A). This suggests the occurrence of two distinct host shifts from the wild summer grape *V. aestivalis* to cultivated grapes, one in the native range, from the red genetic cluster, and the other, from the yellow cluster, giving rise to all invasive populations worldwide.

5

The 32-marker dataset also revealed genetic differentiation between invasive populations in the main wine-producing regions worldwide (Fig. 2). The two distinct genetic groups of *P. viticola* previously identified in Western and Eastern European vineyards (22), respectively, were also detected here in STRUCTURE clustering analyses (Fig. 2A) and with the DAPC (Fig. 2E).

10 Furthermore, the populations in the different invaded areas also displayed significant differentiation, as shown by the significant  $F_{ST}$  values (Supplementary Table S4 and S5), STRUCTURE analyses (Fig 2A) and the DAPC analyses focusing on invasive populations (Fig. 2E). We also found two differentiated populations in Australia (Fig. 2).

### 15 **Lower diversity in invasive *P. v. f. sp. aestivalis* populations**

Genetic diversity in invasive *P. v. f. sp. aestivalis* populations worldwide was much lower than that in native populations. Only a few closely related haplotypes of the three sequenced DNA fragments were present in the invaded areas, whereas considerable diversity was detected in the native *P. v. f. sp. aestivalis* populations (Fig. 1, Supplementary Fig. S2, Table S1). The resulting

20 low nucleotide ( $\pi$ ) and haplotype (H) diversities in the invasive populations (Fig. 3A, Supplementary Table S1) indicated the occurrence of severe bottlenecks during the invasion process. European populations displayed a smaller decrease in genetic diversity than the populations of other invaded areas (Figs. 1 and 3A, Supplementary Fig. S2, Table S1), suggesting that the bottlenecks occurring in Europe were milder. Furthermore, all but one of the invasive

haplotypes worldwide were present in Europe, for the three DNA fragments analyzed (Fig. 1 and Supplementary Fig. S2). Together, these findings suggest that Europe may have acted as a bridgehead for secondary invasions in the rest of the world.

5 Both microsatellite datasets confirmed the much lower levels of genetic diversity in invasive populations than in native populations, as shown by the compact clustering of invasive genotypes in the DAPC (Fig. 2D, Supplementary S3A), and by diversity indices (Fig. 3B, Supplementary Tables S2, S3, and S6 to S11). For example, private allelic richness was one order of magnitude lower in the invasive populations worldwide than in the native range. Microsatellite genetic  
10 diversity was also much lower in introduced than in native populations, with European populations having intermediate values (Fig. 3B, Supplementary Tables S2 and S3).

This pattern of diversity is, thus, consistent with the hypothesis that Europe acted as a bridgehead, with a first wave of invasion to Europe originating from the yellow *P. v. f. sp. aestivalis* cluster  
15 (Fig. 2 and Supplementary Fig. S3) occurring on the wild summer grape in the native Northern American range, and a second wave of invasions subsequently occurring from Europe to other vineyards throughout the world.

### **Worldwide invasion history of *P. viticola* reconstructed by ABC-RF scenario testing**

20 We formally compared the likelihoods of alternative invasion history scenarios involving the most likely population of origin in the native range (*i.e.*, the yellow genetic cluster on *V. aestivalis*), as identified based on previous analyses, and all the invasive introduced populations, using the 32-microsatellite marker dataset and an approximate Bayesian computation random forest (ABC-RF) statistical framework (42). In ABC-RF, genetic data are simulated under different demographic

scenarios, and summary statistics from the resulting simulated data are statistically compared with those obtained from the observed data (43-45). We used an iterative process to infer the various invasion events while keeping a tractable number of scenarios to be compared with the ABC-RF (46). We first identified the most likely demographic scenario(s) composed of bifurcation and admixture events considering the native populations and the invasive populations that corresponded to the most ancient introduction events (i.e., with dates of first disease records outside the native range between 1878 and 1889, Fig. 4B). We then considered the other invasive populations successively following increasing dates of first disease records and tested using ABC-RF what was their population of origin.

10

As a first step, we considered 18 scenarios of introduction from the most likely population-of-origin in the native range (the yellow cluster on *V. aestivalis*) and the first reported invasive populations from historical records (i.e., in Western and Eastern Europe, in ~1878 and ~1887, respectively, and in China, in ~1889; Supplementary Table S12 and Fig. S4). The most likely scenario identified (S11 in Supplementary Table S12 and Fig. S4) involved an initial introduction into Western European vineyards, followed by a spread to Eastern Europe (as previously shown (22)), and then an expansion to Eastern China from Eastern Europe. This scenario was the most strongly supported, with a posterior probability of  $0.42 \pm 0.01$  (Supplementary Table S12).

20 The disease was later reported in South Africa in ~1907 and in eastern Australia in ~1917. For these two introduction events (steps 2 and 3 in Supplementary Table S12, Fig. S5 and S6), the ABC-RF again identified Eastern Europe as the most likely population of origin, with posterior probabilities of  $0.78 \pm 0.03$  and  $0.71 \pm 0.02$ , respectively. The disease was then reported in Argentina in ~1920, and the ABC-RF analysis suggested that this introduction was probably from

Eastern Australia, with a posterior probability of  $0.70 \pm 0.02$  (step 4 in Table S1, Supplementary Fig. S7). The most recent first report of the disease was in Western Australia in ~1917, for which an Eastern Australian origin has been suggested (28). However, the ABC-RF statistical framework, based on a larger set of microsatellite markers and more extensive sampling of *P. viticola* populations worldwide, again identified Eastern Europe as the most likely population of origin for the pathogen in Western Australia, with a posterior probability of 0.73 (step 5 in Supplementary Table S12 and Fig. S8).

The most likely global invasion history is summarized in Figure 4. The classification error rate (i.e., prior error rate) for the first step was relatively high (38%), reflecting the difficulty distinguishing between a first introduction into Eastern or Western Europe, as these populations are genetically very close. By contrast, all subsequent steps in the ABC-RF analysis were well resolved, with a low prior error rates, ranging from 7% to 12% (Supplementary Table S12). The final global scenario (Fig. 4) was strongly supported by the genetic data, as shown by the low prior error rate of 7% (Supplementary Table S12). Posterior model checking and goodness-of-fit assessment of the model-posterior distributions showed that the inferred global invasion scenario generated genetic summary statistics consistent with the observed data, providing high confidence in the inferred scenario. The 10,000 simulations from posterior parameter distributions under the best model produced summary statistics very close to those obtained from the observed dataset (Supplementary Fig. S9), with only 24 of 256 statistics falling in the tail of the predictive probability distribution of statistics calculated from the posterior simulations (i.e.,  $P < 0.05$  or  $P < 0.95$ ). Moreover, none of the *P-values* remained significant after correction for multiple comparisons (47). The inferred invasion scenario (Fig. 4) thus fitted the observed genetic data well.

## Discussion

Using extensive sampling and a powerful statistical framework, we have elucidated the worldwide invasion history of grapevine downy mildew. We show here that a specific genetic cluster of *P. v.* f. sp. *aestivalis*, which parasitizes the wild summer grape *V. aestivalis* in the native range, was the source of the invasion responsible for a devastating pandemic on cultivated grapevines in Europe at the end of the 19th century (9, 12). Severe bottlenecks occurred, with very few haplotypes from the native lineage introduced into Europe. In the inferred invasion scenario (Fig. 4), Europe then served as a bridgehead for a second wave of invasions worldwide, spreading the disease further afield, to Eastern China, South Africa and Eastern Australia. A third wave of expansion probably occurred from Eastern Australia to Argentina. The much more recent introduction in Western Australia (1998) also seems to have originated from Europe, raising questions about the efficacy of quarantine regulations. The contribution of European populations to grapevine downy mildew invasions worldwide reflects the key role of Europe in the trading of plant material during the development of “New World” viticulture. The phylloxera crisis intensified the importation of plant material from North America to Europe and from Europe to “New World” vineyards (48). A role for Europe as a hub for the invasion of “New World” vineyards by grapevine pathogens has also been suggested for phylloxera and for the soil-born nematode vector of the fanleaf degeneration virus (48, 49).

20

Controlling the introduction of pests and diseases is one of the greatest challenges in viticulture, with important economic and environmental consequences. The genetic relationships and diversity inferred here could be used to guide quarantine regulations for grape-growing countries, to prevent further invasions. Introducing new strains or lineages of *P. viticola* would significantly increase

the diversity of invasive populations and create opportunities for genetic admixture. Indeed, we found that, in North America, the cultivated grape (*V. vinifera*) was parasitized by a *P. viticola* lineage (f. sp. *vinifera*) other than the one identified here as invasive (f. sp. *aestivalis*). The *P. viticola* f. sp. *vinifera* lineage has not yet been found on other continents, but could also become  
5 invasive should it be introduced. An increase in genetic variation through migration and/or admixture would probably facilitate the adaptation of *P. viticola* to pesticides and resistant cultivars (50). New invasions by additional *P. viticola* strains or lineages would have a major effect on the wine industry, by destabilizing grapevine protection and canceling out the sustained efforts of breeders to obtain good-quality grapevine varieties resistant to downy mildew. The rich  
10 historical records available for grape diseases make these diseases excellent case studies for obtaining fundamental insight into the processes underlying biological invasions, showing that worldwide biological invasions can result from the secondary dispersal of a particularly successful invasive population, as for grape phylloxera (48, 51). Our findings show that even pathogens subject to very strong bottlenecks and without admixture as a means of generating diversity can  
15 achieve successful invasions worldwide, by retaining an ability to evolve rapidly, and thus to develop new virulence against plant resistance (52-54) and resistance to fungicides (21, 55, 56).

## Materials and methods

***Plasmopara viticola* sample collection and DNA extraction.** *Plasmopara viticola* isolates were  
20 collected as sporulating lesions from 163 sites in the native (North America) or invasive (Europe, China, South Africa, Australia, and Argentina; Fig. 1B and 1C, Supplementary Fig. S1; Table S1, S2 and S3) range. This sampling covers the main grape-growing regions in which the climate is favorable for the disease. In Northeast America, samples were collected from five wild *Vitis* species (*V. riparia*, *V. labrusca*, *V. aestivalis*, *V. vulpina* and *P. quinquefolia*), and cultivated

grapevines (*V. vinifera* and interspecific hybrids). In invaded areas, all samples were collected from diseased *V. vinifera* cultivars. Sampling was performed as previously described (22). Oil spots were freeze-dried overnight, and DNA was extracted with standard CTAB and phenol-chloroform methods (55, 57).

5

**DNA sequencing and analysis.** For the identification of cryptic *P. viticola* species, three DNA fragments, from the 28S ribosomal RNA (*r28S*),  $\beta$ -tubulin (*tub*) and mitochondrial cytochrome-*b* (*cytb*) genes, were amplified by polymerase chain reaction (PCR) and sequenced (Supplementary Table S1). DNA amplification, sequencing and assembly were performed as described by Rouxel *et al.*(34) for *r28S* and *tub*, and as described by Chen *et al.*(55) and Giresse *et al.*(58) for *cytb*. Sanger sequencing was performed at the Genoscope (French national sequencing center, Evry, France). For each DNA fragment, sequences were aligned with *Muscle* (59) implemented in Seaview v.4.6.2 (60). Final alignment lengths after cleaning were 499 base pairs (bp) for *tub*, 685 bp for *cytb* and 703 bp for *r28S*. Distinct haplotypes were identified with DnaSP v5.10.1 (61). For each sequence alignment, genetic relationships between haplotypes were inferred with a maximum likelihood (ML) method and a GTR substitution model implemented in PhyML v.3.0 (62). Node support was estimated by calculating 1,000 bootstraps. Each phylogeny was rooted with *Plasmopara hasltedii* sequences obtained from a recent genome assembly (63). Nucleotide ( $\pi$ , per site) and haplotype (H) genetic diversities were estimated with the DnaSP program.

20

**Microsatellite genotyping.** We constructed two different microsatellite datasets for analysis of the genetic structure and diversity of *P. viticola* f. sp. *aestivalis* populations across the main vineyards worldwide, identification of the most likely host-of-origin, and elucidation of the invasion routes followed. The first dataset included an extensive sampling of 1,974 strains



genotyped for eight microsatellite loci (ISA, Pv7, Pv13, Pv14, Pv16, Pv17, Pv31 and Pv39) (23, 57) (Supplementary Table S2). Microsatellite PCR amplification and genotyping were conducted as previously described (57). The repeatability of genotype scoring was checked by genotyping 5% of the samples twice. For the analyses, genotypes were built by combining the eight 5 microsatellite markers and the three DNA sequences. Genotypes with more than 60% missing data were discarded. A single genotype was retained when putative clone-mates (i.e., repeated multilocus genotypes or MLGs within a given vineyard) were identified. The final first genotype dataset encompassed 1,314 MLGs for the eight microsatellite markers and 1,383 MLGs for the 11 markers (combining the eight microsatellite markers with haplotypes for the three DNA fragments) 10 from 105 localities (Supplementary Table S1).

The second dataset consisted of 181 samples genotyped for 34 microsatellite markers: ISA, Pv7, Pv14, Pv16, Pv17, Pv39, Pv65, Pv67, Pv74, Pv76, Pv83, Pv87, Pv88, Pv91, Pv93, Pv101, Pv103, Pv104, Pv124, Pv126, Pv127, Pv133, Pv134, Pv135, Pv137, Pv138, Pv139, Pv140, Pv141, Pv142, 15 Pv143, PvEST2, PvEST9 and PvEST10 (35, 57). The repeatability of genotype scoring was checked by genotyping 5% of the samples twice. The mean error rate was  $\leq 0.020 \pm 0.005$ . Genotypes with missing data for more than 16 markers ( $\geq 50\%$ ) were discarded. Significant linkage disequilibrium (LD) was detected between two pairs of microsatellite markers (Pv87 and Pv104; Pv124 and Pv133) with a permutation test (1,000 permutations) implemented in 20 GENEPOP v.4.2 (64). A single marker for each pair was therefore retained for data analyses (Pv104 and Pv124). After filtering for missing data and LD, the final dataset included 174 MLGs for 32 markers from 35 sites (Supplementary Table S3).

**Inference of population genetic structure and diversity.** We used the two datasets to investigate genetic structure: one with 11 markers (combining the eight microsatellite markers with haplotypes for the three DNA fragments), and the other with 32 microsatellite markers. We ran similar analyses on the two datasets. We used two complementary individual-based methods to estimate the population structure: the Bayesian model-based clustering method implemented in STRUCTURE v.2.3.4 (40, 41) and a discriminant analysis on PCA (DAPC) (39). We ran STRUCTURE with an admixture model with correlated allele frequencies, and uniform priors for the individual cluster of origin (40, 41). We performed simulations with a number of putative clusters ( $K$ ) ranging from 1 to 10. For each  $K$  value, we conducted 10 independent replicates and checked for convergence. Each analysis included a burn-in period of 50,000 Markov chain Monte Carlo (MCMC) iterations followed by 500,000 MCMC iterations for the estimation of model parameters. We determined the most relevant number of clusters ( $K$ ) using (i) the log likelihood of the data for each  $K$  value (40), (ii) the rate of change of the log likelihood of the data with increasing  $K$  (65), and (iii) the visual inspection of clusters newly generated with increases in  $K$  (66). We used STRUCTURE HARVESTER v.0.6.94 (67) to visualize the likelihood and its rate of change across  $K$  values and replicated runs. For the identification of potentially different clustering solutions, results were summarized and displayed with CLUMPAK v.1.1 (68). This analysis was performed on the two datasets combining native and introduced populations, and also on each area separately, to explore finer genetic structure.

20

DAPC(39) does not make any specific assumptions about mating system or mode of reproduction. It provides a visualization of the genetic structure complementary to that provided by STRUCTURE, by summarizing the variance in allele frequencies summarized using principal components (PCs) and partitioning it between populations relative to the within-group variance.

This analysis was performed on the two datasets, and on the native and introduced populations separately, to explore the finer genetic structure in the two areas. We used the ADEGENET v.2.0.1 R package (69) for conducting the DAPC. In accordance with the recommendations of the user guide, missing data were replaced by the mean value, the number of PCs used in the analysis was 5 identified by the alpha-score optimization procedure and was set to 20 and 10 for the 11- and 32-marker datasets, respectively. *A priori* groups were defined according to host plant or continent, and further partitioning was implemented in accordance with the findings of STRUCTURE or previous studies (e.g. split between Eastern and Western Europe (22)). We assessed how distinct or admixed the clusters of the DAPC were, using scatter plots along the top discriminant axes (DA) 10 and bar plots of individual membership probabilities for each group.

Genetic relationships between clusters were estimated with NJ population trees based on pairwise Cavalli-Sforza chord genetic distances (70) between populations computed with the program POPULATIONS v.1.2.32 (71). The node supports were estimated with 1,000 bootstraps over loci. 15 The trees were drawn with FigTree v.1.4.4 (72).

Allelic frequency differentiation between clusters was estimated with the Weir and Cockerham  $F_{ST}$  estimator (73) implemented in FSTAT v.2.9.4 (74). The significance of pairwise  $F_{ST}$  values was assessed with 100 random data permutations in FSTAT. We assessed the levels of 20 microsatellite genetic diversity within clusters, using allelic richness ( $A_r$ ) and private allelic richness ( $P_{Ar}$ ), both standardized by a rarefaction method (75), to account for differences in sample size. We also calculated observed and expected heterozygosities ( $H_o$  and  $H_e$ , respectively), and the  $F_{IS}$  fixation index (73). These statistics were calculated with ADZE v.1.0 (76), GENETIX v.4.05.2

(77), and FSTAT. Wilcoxon signed-rank tests were performed in the R statistical environment (78), to assess the significance of differences in genetic diversity between clusters.

### **ABC-RF-based inferences of global invasion history**

5 We reconstructed the worldwide invasion history of *P. viticola* f. sp. *aestivalis* using an approximate Bayesian computation (ABC) (43-45) random forest (RF) analysis (42, 79, 80). This ABC-RF analysis used the 32-microsatellite dataset and included eight populations: the seven genetic clusters identified outside the native range on the basis of the 32-microsatellite dataset and the wild population sampled in the native range on the summer grape *V. aestivalis*, which was the  
10 population genetically closest to the invasive clusters (Fig. 2). ABC-RF can estimate posterior probabilities of historical scenarios, based on historical data and massive coalescent simulations of genetic data. We used historical information (i.e., dates of first observation of invasive populations, Fig. 4) to define five sets of competing introduction scenarios, which were analyzed sequentially (Supplementary Table S12, Fig. S4 to S8). Step-by-step, each analysis made use of  
15 the results of the previous analyses, until the most recent invasive populations were considered. The scenarios for each analysis are detailed in Supplementary Table S12 and Figs. S4 to S8.

The scenario parameters (i.e., effective population size  $N$ , effective number of founders  $NF$ , admixture rate  $R_a$ , duration of the bottleneck  $DB$ , and time of population split and admixture  $T$ )  
20 were considered as random variables drawn from prior distributions (Supplementary Table S13). We assumed a generalized stepwise mutation process with possible single-nucleotide insertion, to model a realistic mutation process of microsatellite loci in the coalescent simulations (81). We used DIYABC v.2.1.0 (82) to simulate genetic data for ABC-RF analyses. Simulated and observed datasets were summarized using the whole set of summary statistics proposed by DIYABC for

microsatellite markers, describing the genetic variation for each population (e.g., mean number of alleles per locus, and mean genetic diversity), pair of populations (e.g., pairwise genetic diversity, mean  $F_{ST}$  across loci between two populations, and shared allele distance), or trio of populations (e.g., maximum likelihood admixture estimates) (see the full list and details of the summary statistics in Supplementary Table S14). Linear discriminant analysis (LDA) components were also used as additional summary statistics (83). The total number of summary statistics ranged from 70 to 256, depending on the analysis (Supplementary Table S12).

We used the random forest (RF) classification procedure to compare the likelihood of the competing scenarios at each step with the R package *abcrf* v1.8.1 (42). RF is a machine-learning algorithm that uses hundreds of bootstrapped decision trees to perform classification, using the summary statistics as a set of predictor variables. Some simulations are not used in tree building at each bootstrap (i.e., the out-of-bag simulations), and are used to compute the “prior error rate,” which provides a direct method for estimating the cross-validation error rate. We built a training set ranging from 10,000 to 50,000 simulated microsatellite datasets for each scenario, with the same number of loci and individuals as the observed dataset (Supplementary Table S12). We then grew a classification forest of 500 or 1,000 trees based on the simulated training datasets. The size of the training set and number of decision trees was increased until the results converged over ten independent replicated RF analyses. The RF computation applied to the observed dataset provides a classification vote (i.e., the number of times a model is selected from the decision trees). We selected the scenario with the highest classification vote as the most likely scenario, and we estimated its posterior probability (42). We assessed the global performance of our chosen ABC-RF scenario, by calculating the prior error rate based on the available out-of-bag simulations and we repeated the RF analysis 10 times to ensure that the results converged.

We then performed a posterior model checking analysis on the final scenario, including all eight populations, to determine whether this scenario matched the observed genetic data well. Briefly, if a model fits the observed data correctly, then data simulated under this model with parameters drawn from their posterior distribution should be close to the observed data (84). The lack of fit of the model to the data with respect to the posterior predictive distribution can be measured by determining the frequency at which the observed summary statistics are extreme with respect to the simulated summary statistics distribution (hence, defining a tail-area probability, or *P*-value, for each summary statistic). We simulated 100,000 datasets under the full final scenario (256 summary statistics), and obtained a “posterior sample” of 10,000 values of the posterior distributions of parameters through a rejection step based on Euclidean distances and a local regression post-treatment (43). We simulated 10,000 new datasets with parameter values drawn from this “posterior sample,” and each observed summary statistic was compared with the distribution of the 10,000 simulated test statistics; its *P*-value was computed, and corrected for multiple comparisons (47). The simulation steps, the computation of summary statistics, and the model checking analysis were performed in DIYABC v2.1.0. All scenario comparisons were carried out in R, with the *abcrf* v1.8.1 package (42)

### **Data availability**

Unique haplotypes from the  $\beta$ -tubulin, cytochrome *b*, and ribosomal 28S sequence data were deposited in NCBI-Genbank under the accession codes: [to be announced, TBA]. The two microsatellite datasets and the sequence alignments are available via the INRAE Grapevine Downy Mildew Genomics Dataverse (<https://data.inra.fr/dataverse/gdmg>; doi: [TBA]).

## ACKNOWLEDGMENTS

We would like to thank all the people involved in *P. viticola* sampling: Anton Baudoin (Virginia Tech, USA), Odile Carisse (Agriculture and Agri-Food, Canada), David Gadoury (Cornell University, USA), Mizuho Nita (Virginia Tech, USA), Annemiek Schilder (University of California, Davis, USA), Andrew S. Taylor (DPI, Australia), Davide Gobbin (ETH, Switzerland), Marie-Laure Panon (Comité Champagne, France), Patrice Retaud (SRPV, France), Marie-Pascale Latorse (Bayer Crop-Science, France), Jorge Silva (Bayer CropScience, Portugal), Hanns Kassemeyer (Staatliches Weinbauinstitut Freiburg, Germany), Pal Kozma (University of Pécs, Hungary), Mauro Jermini (Agroscope, Switzerland), Sara Legler (UCSC, Italy), Atanas Atanassov (JGC, Bulgaria), Pere Mestre (INRAE, France), Jiang Lu (Shanghai JiaoTong University). This work was funded by ANR grant 07-BDIV-003 (*Emerfundis* project). YD received a grant from the French National Research Agency [GANDALF project, ANR-12-ADAP-0009] and MCF received a postdoctoral grant from the *Région Ile-de-France*. We thank the Center for Information Technology of the University of Groningen, and, in particular, Bob Dröge and Fokke Dijkstra, for their continuous support and for providing access to the *Peregrine* high-performance computing cluster.

## References

1. P. K. Anderson *et al.*, Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends Ecol Evol* **19**, 535-544 (2004).
- 5 2. M. L. Desprez-Loustau *et al.*, The fungal dimension of biological invasions. *Trends Ecol Evol* **22**, 472-480 (2007).
3. M. C. Fisher *et al.*, Emerging fungal threats to animal, plant and ecosystem health. *Nature* **484**, 186-194 (2012).
4. R. P. Singh *et al.*, The Emergence of Ug99 Races of the Stem Rust Fungus is a Threat to  
10 World Wheat Production. *Annual Review of Phytopathology* **49**, 465-481 (2011).
5. M. T. Islam *et al.*, Emergence of wheat blast in Bangladesh was caused by a South American lineage of *Magnaporthe oryzae*. *BMC Biology* **14**, (2016).
6. T. Giraud, P. Gladieux, S. Gavrilets, Linking the emergence of fungal plant diseases with ecological speciation. *Trends in Ecology & Evolution* **25**, 387-395 (2010).
- 15 7. P. Gladieux *et al.*, The population biology of fungal invasions. *Molecular Ecology* **24**, 1969-1986 (2015).
8. R. Lafon, M. Clerjeau, in *Compendium of Grape Diseases.*, R. C. Pearson, A. C. Goheen, Eds. (APS Press, St. Paul, 1994), pp. 11-13.
9. A. Millardet, *Notes sur les vignes américaines et opuscules divers sur le même sujet*.  
20 (Ferret et fils, Bordeaux, 1881).
10. G. Viennot-Bourgin, *Les champignons parasites des plantes cultivées*. (Masson, Paris, 1949).
11. C. Gessler, I. Pertot, M. Perazzolli, *Plasmopara viticola*: a review of knowledge on downy mildew of grapevine and effective disease management. *Phytopathol. Mediterr.* **50**, 3-44 (2011).
- 25 12. M. Cornu, *II. Le peronospora des vignes, observations sur le phylloxera et sur les maladies de la vigne*. Comptes rendus des séances de l'Académie des sciences, tomes XCI et XCIII (Institut de France, Académie des sciences, Paris, 1882).
13. P. Galet, *Les maladies et les parasites de la vigne, t. 1 : Les maladies dues à des végétaux : champignons, bactéries, viroses et phanérogames*. (Imprimerie du Paysan du Midi, Montpellier, 1977), pp. 871.
- 30 14. Y. Dussert *et al.*, Identification of the first oomycete mating-type locus sequence in the grapevine downy mildew pathogen, *Plasmopara viticola*. *Current Biology*, In press (2020).
- 35 15. M. Brilli *et al.*, A multi-omics study of the grapevine-downy mildew (*Plasmopara viticola*) pathosystem unveils a complex protein coding- and noncoding-based arms race during infection. *Sci Rep* **8**, 757 (2018).
16. Y. Dussert *et al.*, A High-Quality Grapevine Downy Mildew Genome Assembly Reveals Rapidly Evolving and Lineage-Specific Putative Host Adaptation Genes. *Genome Biology and Evolution* **11**, 954-969 (2019).
- 40 17. P. Mestre *et al.*, Comparative analysis of expressed CRN and RXLR effectors from two *Plasmopara* species causing grapevine and sunflower downy mildew. *Plant Pathology* **65**, 767-781 (2016).
18. Y. Liu *et al.*, In Planta Functional Analysis and Subcellular Localization of the Oomycete Pathogen *Plasmopara viticola* Candidate RXLR Effector Repertoire. *Frontiers in Plant Science* **9**, (2018).
- 45



19. L. Yin *et al.*, Genome sequence of *Plasmopara viticola* and insight into the pathogenic mechanism. *Sci Rep* **7**, 46553 (2017).
20. M. P. Camargo, C. F. Hong, L. Amorim, H. Scherm, Cryptic species and population genetic structure of *Plasmopara viticola* in São Paulo State, Brazil. *Plant Pathology* **68**, 719-726 (2019).
21. C. E. Delmas *et al.*, Soft selective sweeps in fungicide resistance evolution: recurrent mutations without fitness costs in grapevine downy mildew. *Mol Ecol* **26**, 1936-1951 (2017).
22. M. C. Fontaine *et al.*, Genetic signature of a range expansion and leap-frog event after the recent invasion of Europe by the grapevine downy mildew pathogen *Plasmopara viticola*. *Molecular Ecology* **22**, 2771-2786 (2013).
23. D. Gobbin, I. Pertot, C. Gessler, Identification of microsatellite markers for *Plasmopara viticola* and establishment of high throughput method for SSR analysis. *Eur J Plant Pathol* **109**, 153-164 (2003).
24. D. Gobbin, I. Pertot, C. Gessler, Genetic structure of a *Plasmopara viticola* population in an isolated Italian mountain vineyard. *Journal of Phytopathology* **151**, 636-646 (2003).
25. D. Gobbin, A. Rumbou, C. C. Linde, C. Gessler, Population genetic structure of *Plasmopara viticola* after 125 years of colonization in European vineyards. *Mol Plant Pathol* **7**, 519-531 (2006).
26. T. Koopman, C. C. Linde, P. H. Fourie, A. McLeod, Population genetic structure of *Plasmopara viticola* in the Western Cape Province of South Africa. *Mol Plant Pathol* **8**, 723-736 (2007).
27. X. Li *et al.*, Pathogenicity Variation and Population Genetic Structure of *Plasmopara viticola* in China. *Journal of Phytopathology* **164**, 863-873 (2016).
28. A. S. Taylor, B. J. Knaus, N. J. Grunwald, T. Burgess, Population Genetic Structure and Cryptic Species of *Plasmopara viticola* in Australia. *Phytopathology* **109**, 1975-1983 (2019).
29. W. Zhang *et al.*, Multiple gene genealogy reveals high genetic diversity and evidence for multiple origins of Chinese *Plasmopara viticola* population. *Sci Rep* **7**, 17304 (2017).
30. D. Gobbin *et al.*, Importance of secondary inoculum of *Plasmopara viticola* to epidemics of grapevine downy mildew. *Plant Pathology* **54**, 522-534 (2005).
31. C. L. Matasci, M. Jermini, D. Gobbin, C. Gessler, Microsatellite based population structure of *Plasmopara viticola* at single vine scale. *Eur J Plant Pathol* **127**, 501-508 (2010).
32. A. Rumbou, C. Gessler, Particular Structure of *Plasmopara viticola* Populations Evolved Under Greek Island Conditions. *Phytopathology* **96**, 501-509 (2006).
33. M. Rouxel *et al.*, Geographic Distribution of Cryptic Species of *Plasmopara viticola* Causing Downy Mildew on Wild and Cultivated Grape in Eastern North America. *Phytopathology* **104**, 692-701 (2014).
34. M. Rouxel *et al.*, Phylogenetic and experimental evidence for host-specialized cryptic species in a biotrophic oomycete. *The New Phytologist* **197**, 251-263 (2013).
35. M. Rouxel *et al.*, Microsatellite markers for characterization of native and introduced populations of *Plasmopara viticola*, the causal agent of grapevine downy mildew. *Applied and Environmental Microbiology* **78**, 6337-6340 (2012).
36. G. Maddalena, F. Delmotte, P. A. Bianco, G. De Lorenzis, S. L. Toffolatti, Genetic structure of Italian population of the grapevine downy mildew agent, *Plasmopara viticola*. *Annals of Applied Biology*, (2020).

37. K. Anderson, V. Pinilla, *Wine Globalization: A New Comparative History*. (Cambridge University Press, 2018).
38. G. Zecca *et al.*, The timing and the mode of evolution of wild grapes (*Vitis*). *Mol Phylogenet Evol* **62**, 736-747 (2012).
- 5 39. T. Jombart, S. Devillard, F. Balloux, Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**, 94-15 (2010).
40. J. K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959 (2000).
- 10 41. D. Falush, M. Stephens, J. K. Pritchard, Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587 (2003).
42. P. Pudlo *et al.*, Reliable ABC model choice via random forests. *Bioinformatics* **32**, 859-866 (2016).
- 15 43. M. A. Beaumont, W. Zhang, D. J. Balding, Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025-2035 (2002).
44. M. A. Beaumont, Approximate Bayesian Computation in Evolution and Ecology. *Annual Review of Ecology, Evolution, and Systematics* **41**, 379-406 (2010).
45. K. Csilléry, M. G. B. Blum, O. E. Gaggiotti, O. Francois, Approximate Bayesian
- 20 Computation (ABC) in practice. *Trends in Ecology & Evolution* **25**, 410-418 (2010).
46. S. F. Ryan *et al.*, Global invasion history of the agricultural pest butterfly *Pieris rapae* revealed with genomics and citizen science. *Proc Natl Acad Sci U S A* **116**, 20015-20024 (2019).
47. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and
- 25 Powerful Approach to Multiple Testing. *J. R. Statis. Soc. B* **57**, 289-300 (1995).
48. C. Rispe *et al.*, Genome sequence of the grape phylloxera : insights into the genome evolution and invasion routes of an iconic pest. *BMC Biology*, in press (2020).
49. V. C. Nguyen *et al.*, Phylogeography of the soil-borne vector nematode *Xiphinema index* highly suggests Eastern origin and dissemination with domesticated grapevine. *Sci Rep* **9**,
- 30 7313 (2019).
50. M. Rius, J. A. Darling, How important is intraspecific genetic admixture to the success of colonising populations? *Trends in Ecology & Evolution* **29**, 233-242 (2014).
51. J. Tello, R. Mammerler, M. Cajic, A. Forneck, Major Outbreaks in the Nineteenth
- Century Shaped Grape Phylloxera Contemporary Genetic Structure in Europe. *Sci Rep* **9**,
- 35 17540 (2019).
52. C. E. Delmas *et al.*, Adaptation of a plant pathogen to partial host resistance: selection for greater aggressiveness in grapevine downy mildew. *Evol Appl* **9**, 709-725 (2016).
53. E. Peressotti *et al.*, Breakdown of resistance to grapevine downy mildew upon limited deployment of a resistant variety. *BMC Plant Biology* **10**, (2010).
- 40 54. F. Delmotte *et al.*, Rapid and multiregional adaptation to host partial resistance in a plant pathogenic oomycete: Evidence from European populations of *Plasmopara viticola*, the causal agent of grapevine downy mildew. *Infection, Genetics and Evolution* **27**, 500-508 (2014).
55. W.-J. Chen *et al.*, At least two origins of fungicide resistance in grapevine downy mildew
- 45 populations. *Applied and environmental microbiology* **73**, 5162-5172 (2007).
56. M. Blum, H. A. Gamper, M. Waldner, H. Sierotzki, U. Gisi, The cellulose synthase 3 (CesA3) gene of oomycetes: structure, phylogeny and influence on sensitivity to carboxylic acid amide (CAA) fungicides. *Fungal Biology* **116**, 529-542 (2012).

57. F. Delmotte *et al.*, Microsatellite DNA markers for *Plasmopara viticola*, the causal agent of downy mildew of grapes. *Mol Ecol Notes* **6**, 379-381 (2006).
58. X. Giresse, S. Ahmed, S. Richard-Cervera, F. Delmotte, Development of New Oomycete Taxon-specific Mitochondrial Cytochrome *b* Region Primers for Use in Phylogenetic and Phylogeographic Studies. *Journal of Phytopathology* **158**, 321-327 (2010).
59. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797 (2004).
60. M. Gouy, S. Guindon, O. Gascuel, SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* **27**, 221-224 (2010).
61. P. Librado, J. Rozas, DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452 (2009).
62. S. Guindon *et al.*, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307-321 (2010).
63. Y. Pecrix *et al.*, Sunflower resistance to multiple downy mildew pathotypes revealed by recognition of conserved effectors of the oomycete *Plasmopara halstedii*. *Plant J* **97**, 730-748 (2019).
64. F. Rousset, genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106 (2008).
65. G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14**, 2611-2620 (2005).
66. E. Vercken *et al.*, Glacial refugia in pathogens: European genetic structure of anther smut pathogens on *Silene latifolia* and *Silene dioica*. *PLoS Pathogens* **6**, e1001229 (2010).
67. D. A. Earl, B. M. vonHoldt, STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359-361 (2011).
68. N. M. Kopelman, J. Mayzel, M. Jakobsson, N. A. Rosenberg, I. Mayrose, Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour* **15**, 1179-1191 (2015).
69. T. Jombart, adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403-1405 (2008).
70. L. L. Cavalli-Sforza, A. W. Edwards, Phylogenetic analysis. Models and estimation procedures. *Am J Hum Genet* **19**, 233-257 (1967).
71. O. Langella. (1999), vol. 2020.
72. A. Rambaut, A. J. Drummond. (2018), vol. 2020.
73. B. S. Weir, C. C. Cockerham, Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358-1370 (1984).
74. J. Goudet. (2001).
75. P. L. Leberg, Estimating allelic richness: effects of sample size and bottlenecks. *Mol Ecol* **11**, 2445-2449 (2002).
76. Z. A. Szpiech, M. Jakobsson, N. A. Rosenberg, ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* **24**, 2498-2504 (2008).
77. K. Belkhir, P. Borsa, L. Chikhi, N. Raufaste, F. Bonhomme. (2004).
78. R Core Team. (2020).
79. A. Fraimout *et al.*, Deciphering the routes of invasion of *Drosophila suzukii* by means of ABC random forest. *Molecular Biology and Evolution* **34**, 980-996 (2017).

80. L. Raynal *et al.*, ABC random forests for Bayesian parameter inference. *Bioinformatics* **35**, 1720-1728 (2019).
81. A. Estoup, P. Jarne, J. M. Cornuet, Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol Ecol* **11**, 1591-1604 (2002).
- 5 82. J.-M. Cornuet *et al.*, DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* **30**, 1187-1189 (2014).
83. A. Estoup *et al.*, Estimation of demographic model probabilities with Approximate Bayesian Computation using linear discriminant analysis on summary statistics. *Molecular Ecology Resources* **12**, 846-855 (2012).
- 10 84. J.-M. Cornuet, V. Ravignie, A. Estoup, Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics* **11**, 401 (2010).

## Figure legends

**Figure 1. Phylogenetic relationships, sampling, and geographic distribution of *Plasmopara viticola* haplotypes of the  $\beta$ -tubulin (*tub*) gene.** **A:** The maximum likelihood phylogenetic tree is rooted with the *Plasmopara hasltedii tub* sequence. Nodes with a star (\*) are supported with a bootstrap value greater than 90%. The branches of the tree are color-coded according to the five *P. viticola formae speciales* (ff. spp.). Colored and empty boxes on the right side of the tree show the host plant on which the haplotype was found in the native area or its geographic location on the introduced areas. Geographic codes include Europe (EU), Australia (AU), Argentina (AR), South Africa (ZA) and China (CN). The number of isolates carrying a given haplotype is indicated by the number within the boxes. Photos on the right show the grapevine downy mildew pathogen, *P. viticola*, infecting young grape berries (top) and typical lesions on a leaf (bottom) of *Vitis vinifera* in Europe. **B:** Map showing the sample size (N) and geographic location of origin of the strains from the five *formae speciales* occurring in Northeast America. **C:** Distribution of the sampling sites of the *P. v. f. sp. aestivalis* strains across the main wine-producing regions worldwide.

**Figure 2. Worldwide population genetic structure of *Plasmopara viticola* inferred from the 32-microsatellite marker dataset.** **A:** Estimated individual ancestry according to the Bayesian clustering approach of STRUCTURE, for two to eight clusters (*K*). Each individual is represented by a thin vertical line, partitioned into *K* colored segments representing the estimated genome ancestry fractions for each cluster. Individuals from different continents (labeled at the top of the figure) are separated by black continuous lines; strains from different host species (labeled at the bottom of the figure) are separated by black dotted lines; within these categories, individuals are

sorted by increasing longitude. The figure shown for each  $K$  corresponds to the highest probability run from 10 replicates, with the best-fitting  $K$  value indicated by a star. **B**: STRUCTURE admixture proportions for samples averaged across populations for  $K = 8$ . Pie chart size is proportional to sample size. **C**: Neighbor-joining (NJ) tree based on the Cavalli-Sforza chord distance. Population branching with high bootstrap support (i.e., >75%), indicated by large widths. Wild and cultivated grapes are represented by different symbols. North America and other continents are separated into different gray circles. **D** and **E**: Scatterplot of the first two discriminant functions (DFs) from the discriminant analysis on principal components (DAPC), showing *P. viticola* individual genotypes for the full dataset (D) and for the subset excluding populations from the native North American range (E). The histogram insets of each DAPC scatter plot show the proportion of variance explained by each DF, according to their respective eigenvalues. The bar plots below each scatter plot represent the probability of the strain belonging to each group, based on all DFs of the DAPC. The populations studied are as follows: North American *P. viticola* strains collected on wild *V. labrusca*, *V. aestivalis* group 1 (yellow) and 2 (red); North American strains collected on cultivated *V. vinifera*. Strains from the rest of the world collected on cultivated *V. vinifera* including, strains from Western and Eastern Europe (WEUR and EEUR), Western and Eastern Australia (WAUS and EAUS), China (CN), South Africa (ZA), and Argentina (AR).

**Figure 3. Lower diversity in invasive populations of *Plasmopara viticola forma speciale aestivalis*.** **A**) Bar plot showing the nucleotide diversity ( $\pi$ , per site) of the  $\beta$ -tubulin (*tub*) gene sequence. **B**) Box plot showing the distribution of expected heterozygosity ( $H_{exp}$ ) for the 32 microsatellite markers.

**Figure 4. Worldwide invasion history of *Plasmopara viticola forma speciale aestivalis* inferred by the approximate Bayesian computation random forest (ABC-RF) approach. (A) The best population divergence scenario inferred by ABC-RF (Supplementary Table S12). (B) Geographic representation of the invasion scenario with the highest likelihood based on the ABC-RF analysis;**

5 areas are shown in color, based on their population assignment, as identified in Figure 2. Dates represent the first report of the disease in each area of introduction. Arrows indicate the most likely invasion pathways. Vineyards in other regions of the world not included in this study are indicated in black.

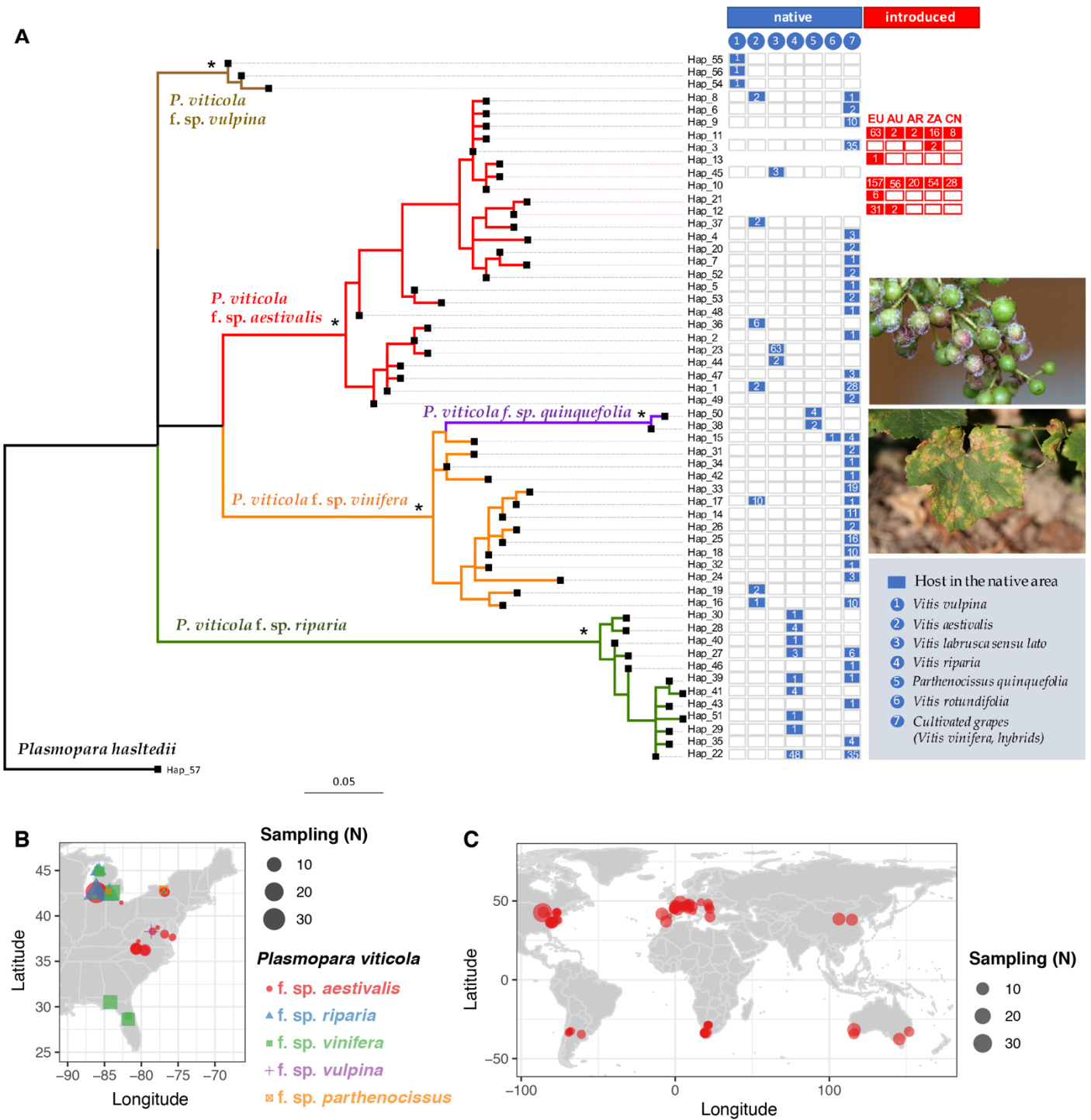


Figure 1



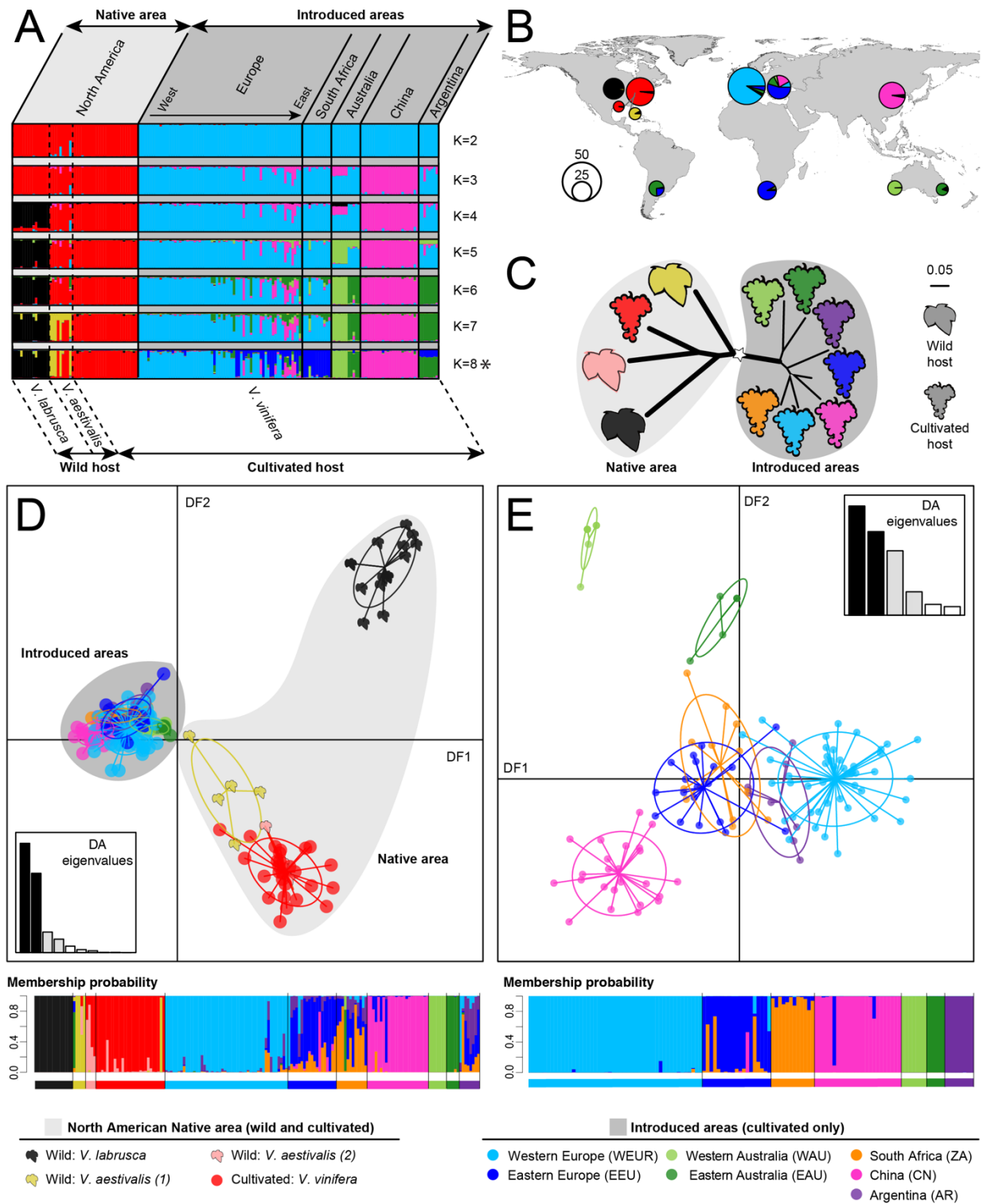


Figure 2

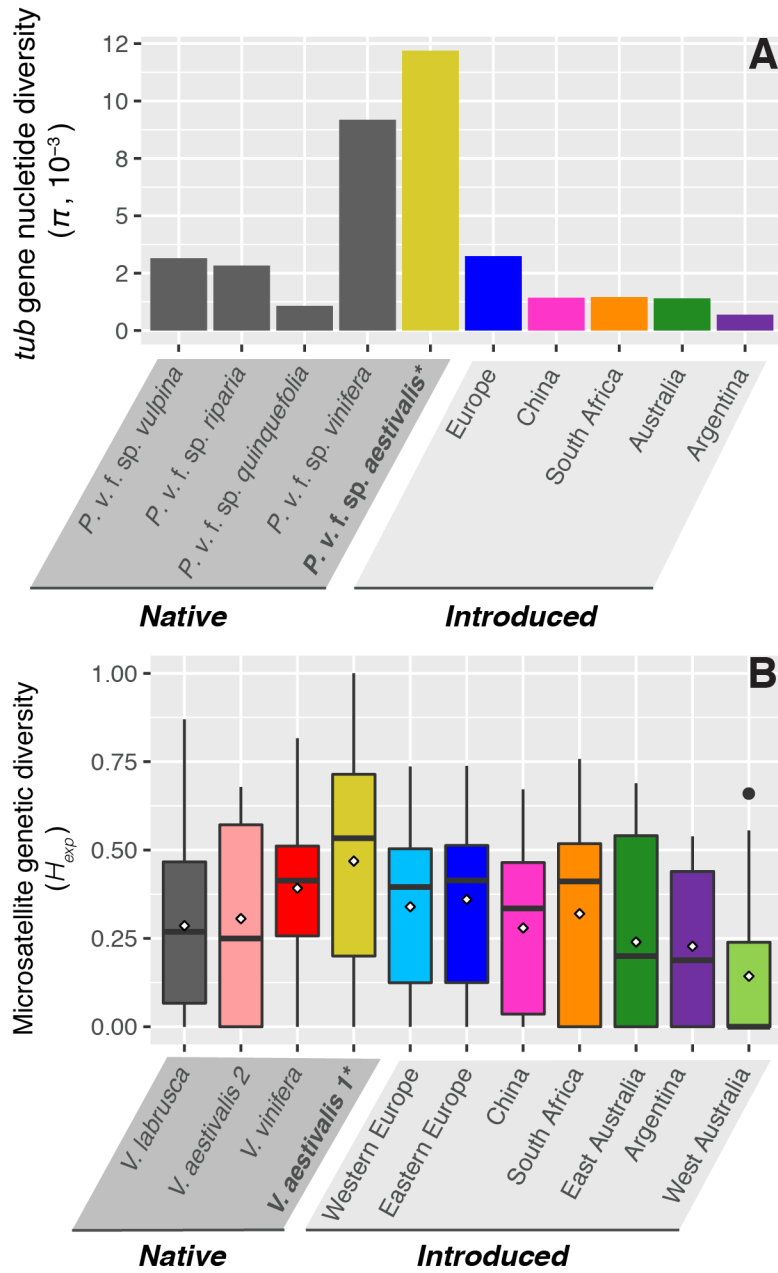


Figure 3

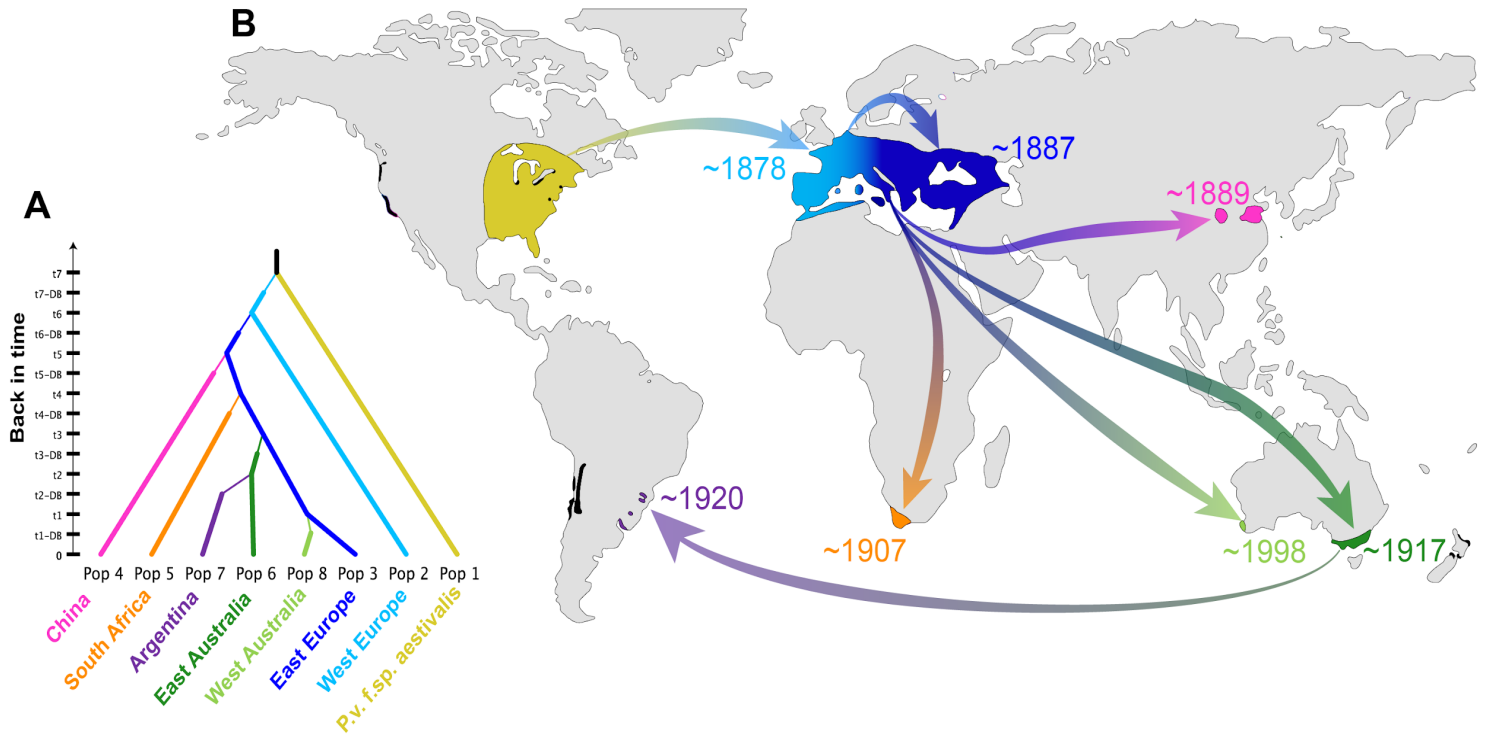


Figure 4.