

Genetic signature of a range expansion and leap-frog event after the recent invasion of Europe by the grapevine downy mildew pathogen *Plasmopara viticola*

MICHAEL C. FONTAINE,*† FRÉDÉRIC AUSTERLITZ,*† TATIANA GIRAUD,* FRÉDÉRIC LABBÉ,* DACIANA PAPURA,‡ SYLVIE RICHARD-CERVERA ‡ and FRANÇOIS DELMOTTE ‡

*Ecologie, Systématique et Evolution, UMR 8079 Université Paris Sud Laboratoire Ecologie, Systematique et Evolution, UMR8079, Orsay Cedex, F-91405, France, †Eco-Anthropologie et Ethnobiologie, UMR 7206 CNRS, MNHN, Univ Paris Diderot, Sorbonne Paris Cité, F-75231 Paris Cedex 5, France, ‡INRA, UMR1065 Santé et Agroécologie du Vignoble, ISVV, F-33883 Villenave d'Ornon Cedex, France

Abstract

Biologic invasions can have important ecological, economic and social consequences, particularly when they involve the introduction and spread of plant invasive pathogens, as they can threaten natural ecosystems and jeopardize the production of human food. Examples include the grapevine downy mildew, caused by the oomycete *Plasmopara viticola*, an invasive species native to North America, introduced into Europe in the 1870s. We investigated the introduction and spread of this invasive pathogen, by analysing its genetic structure and diversity in a large sample from European vineyards. Populations of *P. viticola* across Europe displayed little genetic diversity, consistent with the occurrence of a bottleneck at the time of introduction. Bayesian coalescent analyses revealed a clear population expansion signal in the genetic data. We detected a weak, but significant, continental-wide population structure, with two geographically and genetically distinct clusters in Western and Eastern European vineyards. Approximate Bayesian computation, analyses of clines of genetic diversity and of isolation-by-distance patterns provided evidence for a wave of colonization moving in an easterly direction across Europe. This is consistent with historical reports, first mentioning the introduction of the disease in Bordeaux vineyards (France) and subsequently documenting its rapid spread across Europe. This initial introduction in the west was probably followed by a 'leap-frog' event into Eastern Europe, leading to the formation of the two genetic clusters we detected. This study shows that recent population genetics methods within the Bayesian and coalescence frameworks are extremely powerful for increasing our understanding of pathogen population dynamics and invasion histories.

Keywords: fungi, invasive plant pathogen, microsatellite, oomycetes, population genetics, recent introduction, *Vitis vinifera*

Received 16 November 2012; revision received 11 February 2013; accepted 14 February 2013

Introduction

Biologic invasions can have important ecological, economic and social consequences (Pimentel *et al.* 2001), particularly when they involve the introduction and spread of pathogens (Anderson *et al.* 2004). Considerable

attention is focused on human and animal diseases, but outbreaks of disease due to invasive pathogens are also becoming increasingly frequent in plants (Desprez-Loustau *et al.* 2007). Plant diseases can threaten natural ecosystems and jeopardize the production of human food (Giraud *et al.* 2010). Many pathogens of agricultural systems are believed to be alien in the areas in which they cause damaging epidemics (Pimentel *et al.* 2001). With the same crops being cultivated worldwide and the

Correspondence: Michael C. Fontaine, Fax: +1-574-631-3996; E-mail: mikafontaine@gmail.com

globalization of trade in plant products, pathogenic species are increasingly being displaced out of their native ranges, providing recurrent opportunities for invasions by exotic fungal pathogens (Desprez-Loustau *et al.* 2007). Disease epidemics resulting from introductions of oomycete or fungal pathogens into previously unoccupied environments are well documented. Examples include potato late blight in Europe (Gomez-Alpizar *et al.* 2007), chestnut blight in America (Milgroom *et al.* 1996, Dutech *et al.* 2012), oilseed stem canker (Dilmaghani *et al.* 2012), grapevine powdery mildew in Europe (Brewer & Milgroom 2010), the Dutch elm disease (Brasier 1991) and the banana black leaf streak disease (Robert *et al.* 2012).

Such introduction events are often associated with founder effects, which have a major impact on pathogen population structure and diversity (Austerlitz *et al.* 1997; Le & Kremer 1998). However, the erosion of genetic diversity can be mitigated by multiple introductions from different source populations (Kolbe *et al.* 2004; Genton *et al.* 2005; Delmotte *et al.* 2008; Brewer & Milgroom 2010; Dutech *et al.* 2010). It is also important to study the complex genetic consequences of range expansions in introduced areas, to understand the dynamics of invasive populations (Excoffier *et al.* 2008). Range expansions generally lead to a loss of genetic diversity along the expansion gradient, due to the effects of recurrent bottlenecks (Austerlitz *et al.* 1997; Le & Kremer 1998), as have been demonstrated in humans (Prugnolle *et al.* 2005), *Arabidopsis thaliana* (Francois *et al.* 2008), trees (Demesure *et al.* 1996; Dumolin-Lapègue *et al.* 1997; Cornille *et al.* In press) and fungi (Vercken *et al.* 2010). However, the dynamics of spatially expanding populations typically involve complex interactions between migration and changing effective population size (Excoffier *et al.* 2008). The recent development of computationally intensive coalescent-based methods and the use of rapidly evolving microsatellite markers have resulted in a population genetics framework ideal for investigations of the demographic and genetic processes at work in invasive species subjected to rapid range expansion (Estoup & Guillemaud 2010). The signatures of such processes, when detected, provide us with clues to the distribution and demographic history of pathogens. An understanding of the changes in the genetic diversity and structure of invasive pathogen populations during their expansion is also important for the design of effective, sustainable disease management strategies.

In this study, we focused on a major invasive pathogen, the causal agent of grapevine downy mildew. Modern grapevine emerged from the domestication of the wild Eurasian species (*Vitis vinifera* L.) in Transcaucasia (Châtaignier 1995; Terral *et al.* 2010). The domesti-

cation of grapevines seems to have been linked to the discovery of wine, and grapevines have been cultivated in Europe since classical times (This *et al.* 2006). Since the middle of the 19th century, viticulture has been threatened by grapevine downy mildew, a disease caused by the oomycete *Plasmopara viticola* (Lafon & Clerjeau 1994). This pathogen, which is native to North America, was accidentally introduced into Europe in the 1870s (Viennot-Bourgin 1949). At the time, European agronomists grafted French vines onto American rootstocks to prevent another disease of grapevine caused by phylloxera, *Daktulosphaira vitifoliae*, another devastating pest of American origin. This importation of American phylloxera-resistant hybrid and wild *Vitis* species is thought to have caused the accidental introduction of grape downy mildew in Europe (Millardet 1881). The history of this introduction can be traced back from historical archives (Galet 1977). Grape downy mildew was first reported in France in 1878, in Bordeaux vineyards (Millardet 1881). It is thought that it then spread rapidly throughout European wine-producing regions: it was recorded in Northern Italy in 1879, in the Mosel region in Germany in 1880, in Portugal in 1881 and in Eastern Europe (Turkey, Greece) all the way to the Caucasian vineyards in 1887. However, historical observations may be incomplete or biased, and population genetics can help to trace introduction scenarios (Estoup & Guillemaud 2010).

Despite the economic importance of *P. viticola*, little was known about its genetic structure and diversity within and between European populations, until recently. In the last decade, several population genetic studies have been carried out with a set of four microsatellite markers, to assess the genetic diversity of *P. viticola* (Gobbin *et al.* 2003a,b, 2006, 2007; Rumbou & Gessler 2006). Most of these studies were conducted on a small spatial scale (i.e. within vineyards) and showed that *P. viticola* populations were sexual, with mating mostly random at the vineyard level, although footprints of clonal propagation were also found in some cases (Gobbin *et al.* 2007). Like most oomycetes (Kingdom Straminipila), *P. viticola* can in fact reproduce asexually through the production of sporangia, which release zoospores, and sexually through the production of thick-walled resistant oospores in the fall (Lafon & Clerjeau 1994). *P. viticola* being a heterothallic oomycete, sexual reproduction occurs by outcrossing between diploid individuals of the two mating types (A1 and A2) present on the same leaf (Wong *et al.* 2001; Billiard *et al.* 2012). Recently, cryptic-specialized species have been identified in North American populations of *P. viticola* on cultivated and wild hosts (Rouxel *et al.* 2013). In Europe, *Plasmopara muralis* has also been described on grape ivy (*Parthenocissus tricuspidata*) by Thines (2011).

Table 1 Genetic variation at the eight microsatellite loci of *Plasmodium viticola* studied, within the European vineyards and in the western and eastern genetic clusters identified by TESS

	European vineyards ($n = 938$)						Western cluster ($n = 664$)					Eastern cluster ($n = 215$)				
	NA%	A_r	H_O	H_E	F_{IS}	F_{ST}	A_r	pA_r	H_O	H_E	F_{IS}	A_r	pA_r	H_O	H_E	F_{IS}
PV14	2.35%	4.90	0.68	0.67	-0.007	0.036***	3.00	0.00	0.66	0.67	0.023	4.85	1.85	0.65	0.67	0.010
ISA	0.75%	6.00	0.55	0.61	-0.054	0.085***	5.01	1.01	0.62	0.64	-0.025	4.00	0.00	0.51	0.50	0.030
PV17	1.71%	5.90	0.49	0.46	-0.063	0.043***	3.16	0.79	0.46	0.43	0.152	4.96	2.59	0.40	0.47	-0.072
PV39	0.64%	3.80	0.16	0.16	0.017	0.042***	2.91	0.91	0.15	0.16	-0.055	2.72	0.72	0.11	0.10	0.040
PV13	3.52%	4.80	0.29	0.33	0.054	0.068***	2.00	0.00	0.30	0.33	0.122	4.06	2.06	0.22	0.25	0.096
PV16	1.28%	4.00	0.45	0.41	-0.134	0.046***	2.00	0.00	0.45	0.42	-0.157	2.00	0.00	0.41	0.35	-0.070
PV31†	15.57%	4.00	0.07	0.13	0.345***	0.098***	—	—	—	—	—	—	—	—	—	—
PV7†	8.32%	4.40	0.11	0.31	0.561***	0.082***	—	—	—	—	—	—	—	—	—	—
All loci	4.26%	4.40	0.35	0.39	0.047***	0.060***	3.01	0.45	0.44	0.44	0.01	3.77	1.20	0.38	0.39	0.01

NA%, mean proportion of missing genotype over loci; A_r , Allelic richness; pA_r , private allelic richness; H_O and H_E , observed and unbiased expected heterozygosity.

Departure from HW expectations are measured using F -statistics computed within and between vineyards and tested using permutation tests (* $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; based on 105 permutations).

†These loci were excluded from the analyses as they were subject to potential null allele issues (see the text).

vineyard) values of A_r were calculated with ADZE (Szpiech *et al.* 2008), based on a standardized sample size of $n = 4$. We investigated geographic patterns of genetic diversity by computing the spatial interpolation of the local A_r and H_E values, using a thin plate spline method, with a smoothing parameter $\lambda = 0.005$, as implemented in the R package 'fields' (Furrer *et al.* 2011). Linkage disequilibrium between loci was assessed using permutation tests, with a nominal threshold fixed at 5% and Bonferroni correction for multiple comparisons. We used micro-checker software 2.2.3 (Oosterhout *et al.* 2004) to check for the occurrence of null alleles in the data.

Population genetic structure. We first used permutation tests (10^5 permutations) to determine, for each population, the departure from Hardy–Weinberg (HW) equilibrium; we calculated the F_{IS} statistic measuring this departure (Weir & Cockerham 1984) with FSTAT 2.9.1 (Goudet 1995, 2001). A pattern of isolation by distance (IBD) may emerge if dispersal is spatially restricted (Wright 1943). Under the hypothesis of IBD, genetic differentiation between demes (estimated as $F_{ST}/(1-F_{ST})$) is expected to increase with increasing geographic distance (Rousset 1997). We calculated the regression coefficient (b) between pairwise genetic and geographic distances (on both natural and logarithmic scales) between populations (vineyards in this case) and evaluated its significance with a permutation procedure (10^4 permutations), with SpaGeDi 1.3 (Hardy & Vekemans 2002).

We investigated the genetic structure in more detail with a Bayesian model-based clustering algorithm

implemented in TESS 2.3.1 (Durand *et al.* 2009). This model is a spatially explicit clustering model with admixture, in which individual genome proportions are estimated by incorporating spatial trends and spatial autocorrelation into the prior distribution of individual admixture proportions (q 's). These proportions are allowed to vary over space, and the variation is decomposed into effects at the regional and local scales. Trend surfaces account for clines in all geographic directions, and autocorrelated residuals account for isolation by distance. The parameters specifying the shape of the clines are also estimated from the data, together with the magnitude of spatial autocorrelation. The model – implemented in an Markov chain Monte Carlo (MCMC) algorithm – can therefore be used for the simultaneous detection of clines and clusters, through studies of the inferred variation of admixture proportions (Francois & Durand 2010).

We ran TESS under the conditional autoregressive (CAR) Gaussian model of admixture with a linear trend surface (Durand *et al.* 2009), with updating of the admixture parameter (α) and the interaction parameter (ρ), which were initially set to $\alpha = 1$ and $\rho = 0.6$. The algorithm was run with a burn-in period of 3×10^4 cycles and parameters were estimated with 3×10^5 additional cycles. We varied the maximal number of clusters (K_{max}) from 2 to 10, with 10 replicates for each value of K_{max} . The K value best accounting for the data can be determined by increasing K_{max} until no new subdivision is observed (Vercken *et al.* 2010), or with the deviance information criterion (DIC), a penalized measurement of the fit of the model to the data (Spiegelhalter *et al.* 2002). For the optimal K value, we performed

100 additional runs and investigated the various clustering solutions with CLUMPP version 1.1 (Jakobsson & Rosenberg 2007) ('greedy' algorithm, 100 random input sequences, G' statistic). CLUMPP accounts for label switching and identifies potentially different modes. We assessed the contribution of explicitly accounting for a spatial component in the clustering algorithm, by comparing the results of this spatial model with the results obtained for parameter setting in the absence of spatial information (i.e. $\rho = 0$ and constant spatial trend of admixture coefficient).

We also obtained an additional representation of the genetic structure using principal component analysis (PCA). This multivariate descriptive method is not dependent on any model assumption and can thus provide a useful validation of Bayesian clustering output (Patterson *et al.* 2006; McVean 2009, Francois & Durand 2010). We used the R package ADEGENET v1.3.4 (Jombart 2008) to carry out standard PCA and a modified version of this analysis accounting for spatial autocorrelation, hereafter referred to as spatial PCA or sPCA (Jombart *et al.* 2008). We used the global and local test procedures based on Monte Carlo permutations (10^4 permutations) to interpret the significance of the spatial principal components in the sPCA (Jombart *et al.* 2008). In the study by Jombart *et al.* (2008), 'global structure' relates to patterns of spatial genetic structure, such as patches, clines, isolation by distance and inter-mediate, whereas 'local structure' refers to strong differences between local neighbourhoods.

Demographic history. We first investigated whether the genetic data were consistent with the historical records-based scenario of a very recent introduction (about 140 years ago) and the subsequent expansion of *P. viticola* across Europe. We used the coalescent-based approach implemented in the MSVAR program (Beaumont 1999), which assumes that a population of size N_1 started to increase (or decrease) in size t_a generations ago, to reach its current size, N_0 . The change in population size was assumed to be exponential, and mutations were assumed to follow a stepwise-mutation model (SMM), with parameter $\theta = 4N_0\mu$, where μ is the mutation rate per locus. This method uses a MCMC approach to estimate the posterior probability distributions of (i) the magnitude of population size change $r = N_0/N_1$, (ii) the time (t_a) since the population started changing size scaled by N_0 , $t_f = t_a/N_0$ and (iii) the scaled population mutation rate $\theta = 4N_0\mu$. Wide uniform prior distributions were chosen, between -5 and 5 on a \log_{10} scale for $\log(r)$, $\log(\theta)$ and $\log(t_f)$.

We conducted the analysis on both the entire data set and within each identified genetic cluster. To keep manageable computing times, we performed the analyses

on restricted data sets generated by random resampling ($n = 150$). For each data set, several long runs of 10^{10} steps were carried out, with a thinning of 10^9 steps. The first 10% of the chain was discarded as the burn-in, the rest being considered to constitute a sample of the stationary distribution. For the analysis of MSVAR output, we used the 'locfit' (Loader 2010), 'coda' (Plummer *et al.* 2010), 'mcmc' and MCMCpack (Martin *et al.* 2010) packages in R (R Development Core Team Team 2011). Chain convergence was assessed using the Gelman–Rubin convergence statistics method (Gelman & Rubin 1992; Brooks & Gelman 1998). We plotted the Gelman–Rubin statistics against iteration number, to check that it was reliably converging to a stable value (Brooks & Gelman 1998). In addition, posterior densities from individual runs were examined visually, to check for overall consistency in shape. All runs that had reached convergence were combined for density estimations, conducted with the 'locfit' package (Loader 2010), for the estimation of modes and 90% highest probability density (HPD) limits. HPD limits are points of equal probability density enclosing regions in which the cumulative probability density equals 90%. As we were particularly interested in the detection of population size changes, we focused on the marginal posterior distribution of $\log(r) = \log(N_0/N_1)$.

Comparison of putative scenarios of P. viticola introduction in Europe by approximate Bayesian computation (ABC). We tested three plausible introduction scenarios of *P. viticola* introduction into Europe with the ABC statistical framework (Beaumont 2010; Bertorelle *et al.* 2010; Csilléry *et al.* 2010), on the basis of the observed genetic structure in Europe and historical information: 1) an initial introduction of *P. viticola* into the vineyards of Western Europe from an unsampled source population, with subsequent spread from western to eastern European vineyards (Fig. 2A); 2) an initial introduction into Eastern Europe, with subsequent spread to Western Europe (Fig. 2B); 3) independent introductions into Western and Eastern Europe from the same unknown source population (Fig. 2C). We excluded admixed genotypes – those with a TESS admixture coefficient (q) below 0.80.

For each of the three scenarios considered, we simulated 10^6 data sets with DIY-ABC V1.0.4.46b software (Cornuet *et al.* 2008, 2010). The parameters defining each model (i.e. population size, timing of demographic events, mutation rates) were considered as random variables drawn from prior distributions (Table S2, Supporting information). DIY-ABC draws a value for each parameter from the prior distribution and performs coalescent-based simulations to generate simulated samples with the same number of gene copies and loci per

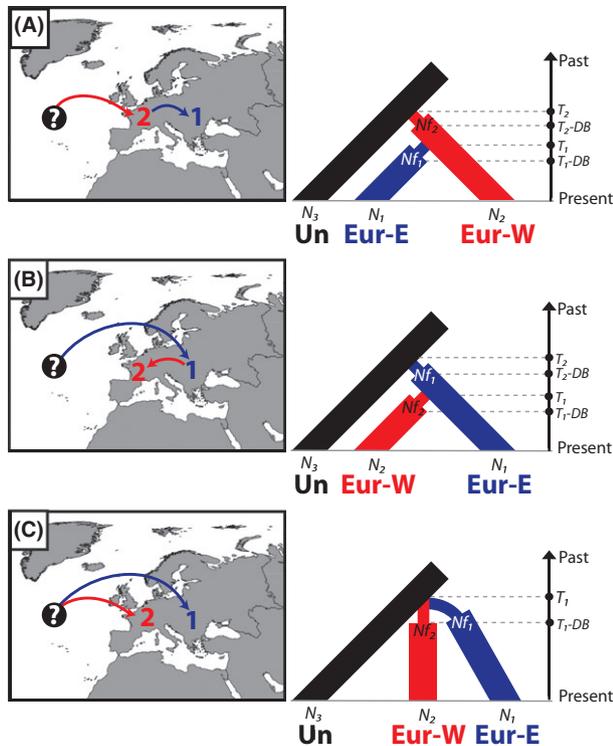


Fig. 2 Scenarios of *Plasmopara viticola* introduction in Europe tested by the approximate Bayesian computation (ABC) approach. A map depicting the successive introduction events is shown on the left and a schematic representation of the scenarios, including the model parameters, is shown on the right. The arrow on the left indicates the timescale, going backward. The first scenario (A) assumes that *P. viticola* was first introduced into Western Europe (Eur-W) from an unsampled source population (Un), from which it was subsequently introduced into Eastern Europe (Eur-E). The second scenario (B) assumes the reverse case, and the third scenario (C) assumes that the two groups were introduced into Europe independently, but from the same source population.

population as observed. It then calculates, for each simulated data set, a set of summary statistics, which are also calculated for the observed data. A Euclidean distance δ was calculated between the statistics obtained for each normalized simulated data set and those for the observed data set (Beaumont *et al.* 2002).

Mutation model

The mutation model for microsatellite loci was assumed to be a generalized stepwise-mutation (GSM) model (Estoup *et al.* 2002) with a mean mutation rate (μ) and mean parameter of the geometric distribution for the length, in repeat numbers, of mutation events (P) drawn from uniform prior distributions (μ : $[10^{-5}-10^{-3}]$ and P : $[0.1-0.3]$, see Table S2, Supporting information). Each locus had a possible range of 40 contiguous allelic

states; individual mutation rates (μ_{loc}) and geometric distribution parameters (P_{loc}) were drawn from gamma distributions with respective means μ and p , and shape parameter 2 (Verdu *et al.* 2009). We also considered mutations inserting or deleting a single nucleotide in the microsatellite sequence. We used default values for other mutation model settings (see details in Table S2, Supporting information).

Summary statistics

Summary statistics for genetic diversity were calculated with DIY-ABC, on the simulated and real data sets. Within populations, we calculated the mean number of alleles per locus (NAL), the mean expected heterozygosity (H_E) and the mean allele size variance (V) over all loci for each population. Between populations, we calculated F_{ST} (Weir & Cockerham 1984) and $(\delta\mu)^2$ distance (Goldstein *et al.* 1995) over all loci.

Model choice procedure and performance analyses

The posterior probability of each competing scenario was estimated using a polychotomous logistic regression (Cornuet *et al.* 2008, 2010) on the 1% of simulated data sets closest to the observed data set. The selected scenario was that with the highest significant probability value with a nonoverlapping 95% confidence interval. We evaluated the ability of our ABC approach to discriminate between scenarios, by analysing simulated data sets with the same number of loci and individuals as our real data set. As described by Cornuet *et al.* (2010), we estimated the Type-I error probability as the proportion of instances in which the selected scenario did not give the highest posterior probability among the competing scenarios, for 500 simulated data sets generated under the best-supported model. We also estimated the Type-II probability, by simulating 500 data sets for each alternative scenario and calculating the mean proportion of instances in which the best-supported model was incorrectly selected as the most likely model.

Parameter estimation and goodness-of-fit

We estimated the posterior distributions of demographic parameters under the best demographic model, by carrying out local linear regression on the 1% closest of 10^6 simulated data sets, after the application of a *logit* transformation to parameter values (Beaumont *et al.* 2002; Cornuet *et al.* 2008). Finally, as described by Gelman *et al.* (1995), we evaluated whether, under the best model-posterior distribution combination, we could reproduce the observed data with the model checking procedure available in DIY-ABC v.1.0.4.46b (Cornuet

et al. 2010). Model checking was carried out by simulating 10 000 pseudo-observed data sets under each studied model-posterior distribution combination, with sets of parameter values drawn with replacement from the 10 000 sets of the posterior sample. This generated a posterior cumulative distribution function for each simulated summary statistics, from which we were able to estimate the *P*-value of the deviation of the observed value of each statistics from its simulated distribution under the best demographic model. We also included, as summary statistics, the within-population *M* statistic of Garza & Williamson (2001), the two-sample mean number of alleles, gene diversity, allele size variance across loci and the shared allele distance between samples (Chakraborty & Jin 1993).

Regression of heterozygosity against geographic distance. In invading species, genetic diversity is expected to decrease with distance from the origin of introduction in newly invaded areas (Austerlitz *et al.* 1997; Le & Kremer 1998). A negative correlation is thus expected between the genetic diversity of a population and its distance from the origin of introduction (Prugnolle *et al.* 2005; Excoffier *et al.* 2008). We divided Europe into a 300 × 180 pixel two-dimensional lattice and considered each pixel in turn as a potential source for the geographic expansion of *P. viticola*. Like Francois *et al.* (2008), we calculated, for each pixel, Pearson's coefficient of the correlation, across all populations, between the genetic diversity of a given population and its geographic distance from the focal pixel. The pixel with the lowest negative correlation coefficient is thus indicative of the origin of introduction. Only sampled sites with a sample size greater than five were used in this regression analysis.

Results

Genetic polymorphism

We identified 515 distinct multilocus genotypes among the 1146 genotyped samples (*N*), giving a *G/N* value of 0.45. Twenty-two of 68 vineyards contained no duplicated genotypes and seven had more than ten duplicated genotypes (up to 16) (Table S1, Supporting information). The inclusion of clonal genotypes in the analysis can distort estimates of heterozygosity and *F*-statistics (Balloux *et al.* 2003). We therefore carried out all population genetic analyses after removing repeated genotypes at each given site. The final clone-corrected data set included 938 multilocus genotypes (Fig. 1 and Table S1, Supporting information).

Allelic richness ranged from 4.0 (PV16 and PV31) to 6.0 (ISA) for the various markers, and Nei's genetic diversity (H_E) was between 0.13 (PV31) and 0.67 (PV14,

Table 1). No obvious spatial trend in either parameter was observed across European vineyards, although some spots of higher diversity were observed (Fig. 3). The allele frequency spectrum for all markers other than PV14 and ISA was characterized by a dominant allele with a frequency above 0.7 and the presence of two or three alleles at low frequencies (Fig. S1, Supporting information).

No significant linkage disequilibrium between markers was detected after Bonferroni correction. Three loci (PV31, PV7 and PV13) displayed strong heterozygote deficiency across the whole sampling area compared to what would be expected under Hardy–Weinberg equilibrium (Table 1, $P < 0.001$). Two of these loci (PV31 and PV7) displayed significant, high levels of heterozygote deficiency within demes (with F_{IS} values of 0.35 and 0.56, respectively) and had much higher proportions of missing data than the other markers (15.6 and 8.3% vs. <3.5% for the other loci) (Table 1). The Micro-checker analysis (Oosterhout *et al.* 2004) identified these two markers as potentially affected by the occurrence of null alleles; they were therefore excluded from subsequent analyses.

Population structure

A weak but significant genetic differentiation was observed among European vineyard sites ($F_{ST} = 0.054$,

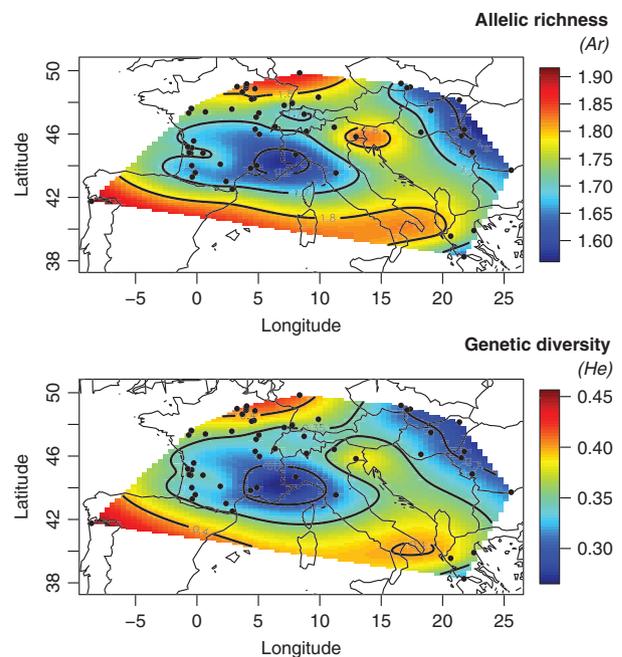


Fig. 3 Genetic diversity of *Plasmopara viticola* in Europe. The spatially interpolated values for allelic richness (A_r , based on a standardized sample size of $n = 4$) and genetic diversity (H_E) (Nei 1973) are provided in panels A and B, respectively.

95% CI: [0.039–0.073], $P < 0.001$, Table 1). We detected also a significant increase in genetic differentiation with geographic distance. The slope of the regression line (b) of $F_{ST}/(1-F_{ST})$ against linear geographic distance was 0.44×10^{-4} ($P = 0.004$) and that of the regression against log-distance was 0.025 ($P < 0.001$).

We investigated the genetic structure of *Plasmopara viticola* further, by performing model-based clustering analysis in TESS 2.3, accounting for spatial patterns in the genetic structure, such as clines, spatial autocorrelations and clusters (see Durand *et al.* 2009; Francois & Durand 2010). The TESS runs with the smallest DIC values were obtained for numbers of clusters (K_{max}) greater than 3 (Fig. S2, Supporting information). With numbers of clusters (K_{max}) from 2 to 10, we detected only two genetically distinct clusters (Fig. 4). We ran 100 additional runs at $K_{max} = 4$, to assess the robustness of the results, and found that two distinct clusters were identified in 75% of these runs. The remaining 25% identified a single cluster with a residual signal of a second cluster reminiscent of the dominant solution (Fig. 4A). The two genetic clusters were geographically separated: one was clearly predominant in Western European samples and the other in Eastern European samples (Fig. 4A,B). The two clusters met in Central Europe, forming a contact zone in which individual genotype admixture proportions progressively changed from one cluster to the other, along a longitudinal axis following a sigmoid pattern (Fig. 4A,B). No such genetic structure was observed when TESS was run under a model that did not account for space, probably due to the weak genetic structure. Indeed, although the genetic differentiation between the two clusters was highly significant ($P < 0.001$), the F_{ST} value was only 0.020.

Principal component analysis and spatial PCA (sPCA) (Jombart *et al.* 2008) supported the results obtained in TESS clustering analyses accounting for spatial proximity (Fig. S3, Supporting information). Individual genotypes from the two clusters identified by TESS were separated along the first components of the PCA and of the sPCA, the latter providing a slightly better discrimination. The overlap between the two groups further illustrated their weak genetic differentiation. We tested the significance of the spatial principal components in the *global* and *local* tests of Jombart *et al.* (2008). The first component was highly significant (global test, $P = 0.009$), providing further support for the existence of a significant genetic structure in European *P. viticola* populations.

The genetic diversities of the two clusters, estimated by calculating mean allelic richness (A_r) values, did not differ significantly (Wilcoxon signed-rank test $P = 0.15$). The mean A_r values per locus in the western and eastern clusters were 3.1 and 3.8, respectively

(Fig. S4, Supporting information). A significant IBD pattern was detected within the western cluster (b -linear = 1.1×10^{-4} , $P = 0.003$; b -log = 0.035, $P = 0.0003$, $n = 660$), but not within the eastern cluster (b -linear = 2.9×10^{-5} , $P = 0.121$; b -log = 0.012, $P = 0.082$, $n = 209$).

Demographic history

A signal of population expansion. We investigated whether the genetic data were consistent with a very recent introduction of *P. viticola* into Europe 140 years ago, followed by a large population expansion, as assumed on the basis of historical records. We used the coalescent approach implemented in MSVAR (Beaumont 1999). We conducted this analysis on a restricted data set constructed by resampling 150 genotypes randomly, first in the European sample as a whole, and then within each genetic cluster separately. We excluded microsatellite loci displaying possible null alleles (PV7 and PV31) or irregular mutational steps (ISA) (Fig. S1, Supporting information).

Most of the coalescent simulations on the full European data set (12 of 15) converged on the stationary distribution, as indicated by the Gelman–Rubin multivariate convergence statistic of 1.03 obtained, and provided consistent marginal posterior probability distributions (PPDs) for each parameter (Fig. 5). The PPDs of the log-ratio between ancestral and current population size ($\log_{10}(N_0/N_1)$) were well resolved and mostly concentrated in positive values (Fig. 5A), consistent with a population size expansion. The modal value from pooled PPDs of $\log_{10}(N_0/N_1)$ was 1.29, corresponding to a 20-fold population expansion. However, there was a high degree of uncertainty attached to this estimate, with the 90% highest probability density interval being [−3.64–4.16] corresponding to a change in effective population sizes (N_0/N_1) ranging from 0 to 14 454. The PPDs of the composite parameter $t_f = t_a/N_0$ were quite flat (Fig. 5B), suggesting that the genetic data contained little information about the timing of the onset of population expansion (t_a) or current effective population size. The mode at low values ($\log_{10}(t_f) = -3.06$, Fig. 5B) was consistent with a recent population expansion and/or a large current effective population size. Coalescent analyses on each genetic cluster separately revealed the same genetic signal of population expansion in the western cluster, but not in the eastern cluster (Fig. S4, Supporting information).

Scenarios of Plasmopara viticola introduction in Europe. We used the DIY-ABC program to compare three scenarios of *P. viticola* introduction in Europe (Fig. 2): the posterior probability of the first scenario (SC1, Fig. 2A),

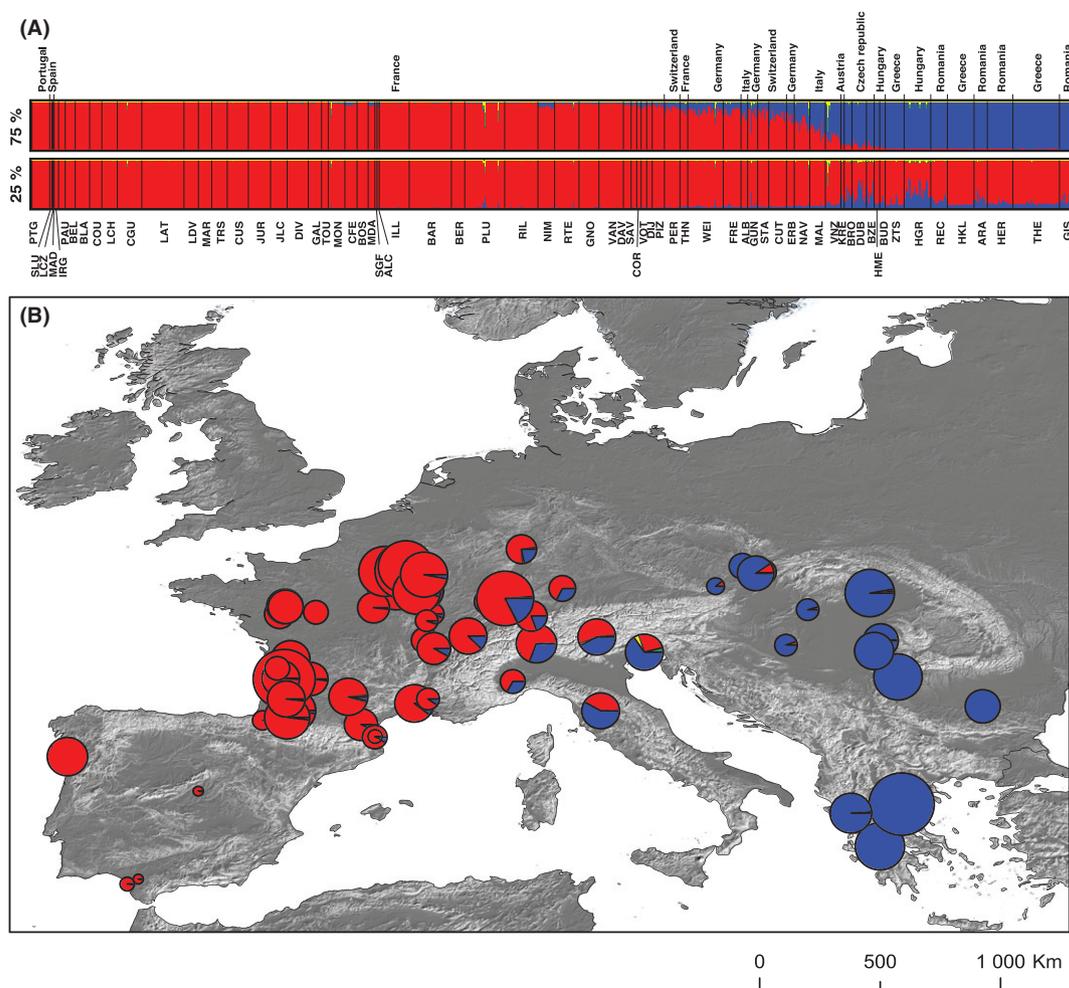


Fig. 4 Population genetic structure of *Plasmopara viticola* in Europe. (A) Individual membership coefficients assuming $K_{max} = 4$ putative populations, for the two clustering solutions obtained over 100 TESS runs: the dominant solution obtained in 75% of the runs (top) and the other solution observed in the remaining 25% of the runs (bottom). Each genotype is represented by a thin vertical line partitioned into K_{max} segments representing the individual genotype admixture proportions in each cluster. Sampling sites are separated by black lines and sorted by increasing longitude. The codes used are listed in Figure 1 and Table S1. (B) Admixture proportions of grapevine downy mildew samples averaged over sites. The relief is shown on a grey scale, with white indicating mountainous areas. Circle size is proportional to sample size.

assuming an introduction of *P. viticola* from an unsampled population into Western Europe, followed by a leap-frog event into Eastern Europe, was significantly higher than those of the other two scenarios (Table 2). The reciprocal scenario of an initial introduction into Eastern Europe followed by a leap-frog event into Western Europe (SC2, Fig. 2B) received slightly lower support, with no overlapping confidence interval with the posterior probabilities for SC1. The hypothesis of two independent introductions from the same unsampled source population (SC3, Fig. 2C) was the least supported scenario.

Despite the higher posterior probability of SC1 than of the other two scenarios, it was difficult to distinguish between these three possibilities based on evaluation of

the performance of the model choice procedure (Table 2). Indeed, although 31.6% of the 500 pseudo-observed data sets (PODs) simulated with SC1 were correctly identified as having been generated under this model, it was not possible to identify the model used for 24.6% of the PODs, and 43.8% were incorrectly identified as having been generated under SC2 (25.6%) or SC3 (18.2%) (*i.e.* Type-I or α -error rate). Of the 500 PODs simulated using each of the two alternative scenarios (SC2 or SC3), 22.6% and 22.8%, respectively, were wrongly identified as having been generated under SC1. This empirical estimation of the Type-II error rate of 45.7% yielded a power of 54.6% for our model choice procedure. However, the third scenario was less able to reproduce the observed data, with statistics deviating

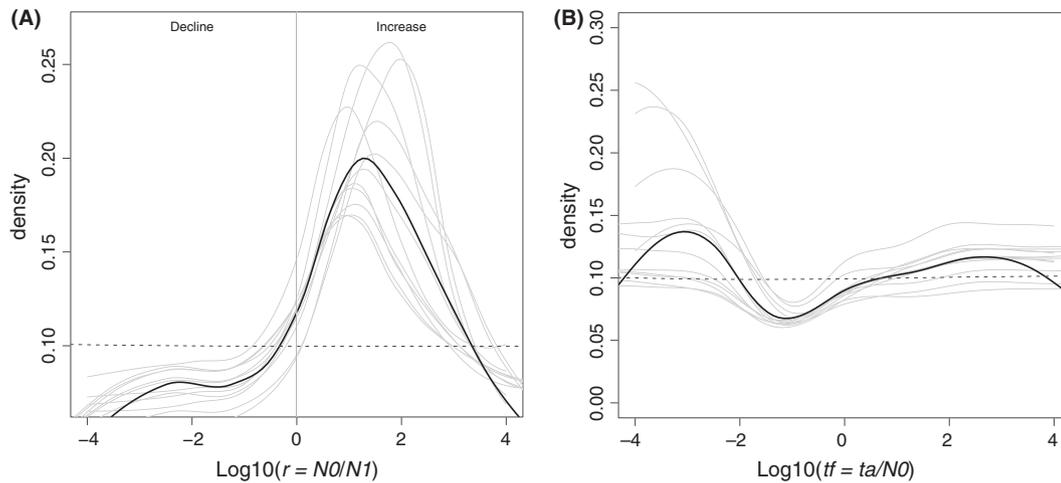


Fig. 5 Demographic history of *Plasmopara viticola* in Europe. Posterior probability density for the parameters of MSVAR analyses: (A) the \log_{10} ratio of current to ancestral population size (N_0 / N_1) and (B) the time (on a log-10 scale) at which population size began to change, scaled by N_0 ($t_f = t_a / N_0$). The grey curves show the individual runs and the black curves, the density estimation for all runs pooled together.

Table 2 Model choice procedure and ABC performance analysis by DIY-ABC

	SC1	SC2	SC3
Post Prob	46.61% (44.77–48.46)	34.77% (33.03–36.52)	12.62% (17.27–19.96)
Performance evaluation			
D(SC1)	31.6%	22.6% [†]	22.8% [†]
D(SC2)	25.6%*	33.0%	23.2%
D(SC3)	18.2%*	19.2%	33.4%
U	24.6%	25.2%	20.6%
Number of outlying statistics			
$P < 0.05$	2	2	2
$P < 0.01$	0	0	0
$P < 0.001$	0	0	1

D – proportion of case in which the model choice procedure was able to select a scenario as the most probable with non-overlapping confidence intervals of the posterior probabilities of each scenario. U – proportion of cases in which the model choice procedure was not able to select a scenario because of overlapping confidence interval in the posterior probabilities of each scenario.

*Type-I or α -error rate.

[†]Type-II or β -error rate and that $1 - \sum \beta_i$ provide the power of the model choice procedure.

considerably from the observed values (Table 2 and S3, Fig. S6, Supporting information).

Marginal posterior probability density and point estimates for each parameter of the best-supported model (SC1) are provided in supplementary material (Fig. S7, S8 and Table S4, Supporting information).

Regression of diversity on geographic distance. Genetic diversity is expected to decrease from the origin of introduction to newly invaded areas, so a negative correlation coefficient is expected between the genetic diversity of a population and its distance from the origin of introduction of the species (Prugnolle *et al.* 2005; Francois *et al.* 2008). Considering sampling sites with at

least five genotypes across the whole study range, the decrease in Nei's genetic diversity (H_E) with increasing geographic distance was maximal when North-West Europe was considered as the putative origin of introduction of *P. viticola* into Europe, with a Pearson's correlation coefficient of $r_p = -0.49$ (Fig. 6A). Within the western cluster, the decrease in genetic diversity with increasing geographic distance was maximal when northern France was considered as the putative origin of introduction, although the correlation coefficient was low ($r_p = -0.32$, Fig. 6B). This low correlation coefficient indicates that the genetic diversity of the western cluster includes little clinal variation. A similar analysis of the eastern cluster yielded a much stronger correlation

coefficient ($r_p = -0.80$, Fig. 6C) and identified the south-western part of this cluster (Italy) as the most probable zone of introduction.

Discussion

Based on a large sampling of the grapevine downy mildew pathogen (*Plasmopara viticola*) in European vineyards, this study revealed low genetic diversity and a weak genetic structure for this pathogen across Europe. The lack of significant deviation from Hardy–Weinberg equilibrium and of linkage disequilibrium within vineyards indicates that European *P. viticola* populations are panmictic at this scale. These results support the idea that sexual inocula (oospores) play a potentially

important role in the development of grapevine downy mildew epidemics in Europe (Gobbin *et al.* 2005, 2006). The rather low genetic diversity of *P. viticola* populations suggests the occurrence of a demographic bottleneck during its introduction into Europe. The rather low diversity and weak structure also strongly suggest that the isolates introduced into Europe in the 1870s came from a single source population of North American origin. Although multiple introductions could have been expected given the level of grapevines importation into Europe during the XIXth century, a single introduction is also consistent with historical records that reported the first observation of grapevine downy mildew localized in Bordeaux vineyards in 1878 (Millardet 1881; Galet 1977; Gessler *et al.* 2011). Previous studies

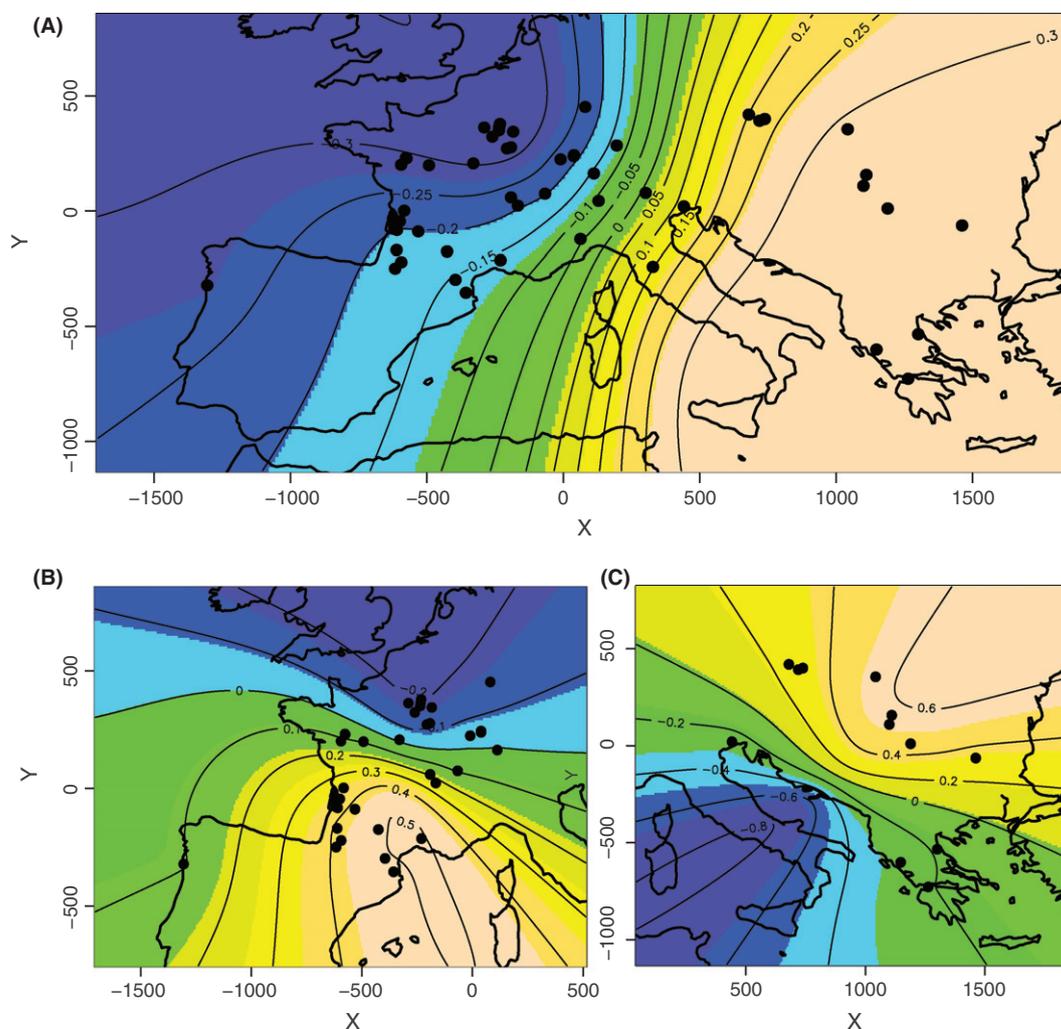


Fig. 6 Genetic diversity of *Plasmopara viticola* populations regressed on geographic distance across Europe (A) and per cluster (B and C). The value for each pixel corresponds to the correlation coefficient (r_p) between expected heterozygosity in each population and the geographic distance between this population and the pixel. We used 300×180 points on a two-dimensional lattice covering Europe. The black dots represent the sampling sites (where n was ≥ 5) used in the regression analysis. The per cluster analyses were performed on individual genotypes with proportions of ancestry for their respective clusters > 0.7 .

on other oomycete and fungal plant pathogens have also reported low overall diversity in the introduced ranges (Raboin *et al.* 2007; Goss *et al.* 2009; James *et al.* 2009; Mboup *et al.* 2009; Montarry *et al.* 2010; Ahmed *et al.* 2012; Fontaine *et al.* in press). In *P. viticola*, the mean genetic diversity ($H_E = 0.39$) was even lower than that measured in other invasive oomycete species, such as the human pathogen *Pythium insidiosum* ($H_E = 0.81$) (Supabandhu *et al.* 2007), the agent of potato late blight *Phytophthora infestans* ($H_E = 0.44$ in Montarry *et al.* (2010) and $H_E = 0.54$ in Brurberg *et al.* 2011) or the agent of sudden oak death *Phytophthora ramorum* ($H_E = 0.55$ in Vercauteren *et al.* 2011), in their areas of introduction.

The MSVAR coalescent-based analyses provided evidence for a demographic expansion of *P. viticola* in Europe. This is consistent with historical records reporting a rapid spread of the pathogen across this continent. Indeed, only 2 years after first the first reports in Bordeaux (1878), the disease had spread throughout all French vineyards and had reached Hungary and Italy (Galet 1977). This rapid expansion of grapevine downy mildew in Europe may have been fostered by the spread of infected plant material via human trade, in addition to the natural long-distance dispersal of sporangia. However, we found a weak but highly significant IBD pattern both at the European scale and within the western cluster, indicating that, despite the high level of gene flow in this species, dispersal was restricted at the continental scale.

The spatially explicit clustering analysis revealed the existence of two distinct, but weakly differentiated, genetic clusters, located in Western and Eastern European vineyards, respectively. The two genetic clusters in Europe were detected only when using a spatially explicit clustering algorithm (implemented in TESS) that has proved more powerful than nonspatial methods for revealing weak genetic structures (Durand *et al.* 2009; Guillot *et al.* 2009; Francois & Durand 2010). The PCA confirmed the differentiation between the two clusters, which largely overlapped. However, these clusters are unlikely to be an artefact resulting from the IBD pattern or from the sampling gap in central Europe (Schwartz & McKelvey 2009). As shown in Figure 1, the sampling gap in central Europe corresponds to the actual geographic distribution of grapevine in this region. The location of the suture zone between the two clusters, which lies farther west from the sample gap, indicates that this structure is not an artefact arising from the gap. Furthermore, the method implemented in TESS explicitly models clinal and spatially autocorrelated genetic variation (IBD pattern) to estimate admixture proportions for each genotype (Durand *et al.* 2009). The IBD patterns recovered by this method therefore cannot

give rise to artefactual clusters. The IBD patterns associated with sampling gaps can generate spurious genetic clusters in STRUCTURE analysis (Schwartz & McKelvey 2009), whereas TESS uses so-called hidden Markov random fields to model spatial dependence in allele frequencies. It detects geographic discontinuities in allele frequencies and uses them to find clusters, providing a spatially explicit alternative to STRUCTURE v2.x (Durand *et al.* 2009; Francois & Durand 2010).

There are several possible reasons for the existence of two weakly differentiated genetic clusters of *P. viticola* in Europe. First, after a single introduction and expansion across Europe, the populations from the two regions may have differentiated due to the presence of the Alps, a well-known biogeographic barrier to dispersal in Europe. However, the existence of admixed individuals in the suture zone of the two *P. viticola* clusters is not consistent with such a geographic barrier to gene flow, particularly as the hybrid zone is not located in this mountain range (Fig. 4). Differences in abiotic selective pressures are also unlikely to be strong enough to result in immigrant inviability or the evolution of extrinsic postzygotic barriers to gene flow, which would be necessary to generate genetic differentiation at neutral loci (Giraud *et al.* 2010). Second, two distinct introduction events may have occurred, followed by two waves of colonization converging on Central Europe. However, the low level of differentiation between the two clusters suggests that the isolates introduced originated from the same source population in North America. The two clusters may actually reflect a more complex process of colonization by this pathogen in Europe: an initial introduction event in Western Europe followed by a 'leap-frog' event allowing the colonization of Eastern Europe from Western Europe. The 'leap-frog scenario' was, indeed, the best-supported scenario in the ABC analysis. The power to distinguish between the two reciprocal leap-frog scenarios was low, but the ABC approach provided additional support for the hypothesis of an initial introduction event in Western Europe.

A simulation study has shown that allele surfing during such a population expansion may generate differentiated genetic clusters (Francois *et al.* 2010). However, the model assumed a homogenous colonization of a continuous habitat, whereas the habitat of *P. viticola* is highly fragmented, growing on grapes only. Furthermore, the findings of different IBD patterns in the western and eastern clusters and of a strong decline of genetic diversity with increasing distance in the eastern cluster indicate that the colonization process was not homogeneous in a single direction, instead involving a more complex scenario. It appears, therefore, unlikely that the patterns observed here could result from an allele surfing phenomenon.

The distinct patterns of genetic structure observed between the Western and Eastern clusters suggest that they have reached different stages towards migration–drift equilibrium (Rousset 1997, 2004, 2007; Hutchison & Templeton 1999; Leblois *et al.* 2004). Indeed, the significant IBD pattern, the absence of clinal variation in genetic diversity and the detectable signal of population expansion in coalescence-based analyses observed within the western cluster indicate that it is closer to this equilibrium than the Eastern cluster. This means that postcolonization dispersal processes and genetic drift have significantly impacted the current genetic diversity of the Western cluster, attenuating signatures of the early invasion process. The current effective population size in the western cluster is thus distinguishable from past effective population size at the time of the colonization, as shown by the MSVAR coalescence analysis even if the confidence interval remained quite large. In contrast, the lack of IBD pattern, the strong cline in genetic diversity and the lack of expansion signal in coalescence-based analyses suggest that the pattern of genetic diversity has kept a signature of the colonization wave, and that the past effective population size at the time of the introduction cannot be distinguished from the current one. The differences between the two clusters reinforce the view that two distinct waves of expansion have occurred, separately in Western and Eastern Europe.

Conclusion

Knowledge about the introduction pathways of invasive pathogens has long been largely based on historical records, which are often scarce and misleading. In this context, population genetics can greatly increase our understanding of pathogen dynamics, as shown here for the invasive pathogen responsible for grapevine downy mildew in Europe. However, we still need to obtain a full picture of invasion by this major grapevine pathogen worldwide and to assess the strength of the bottleneck in Europe by comparing the diversities of introduced and native populations. We also need to unravel the routes of introduction of *P. viticola* worldwide, to understand the history of its spread into New World vineyards (Argentina, Chile, South Africa, Australia, New Zealand).

Acknowledgements

We would like to thank all the people involved in *Plasmopara viticola* sampling across Europe: Davide Gobbin (ETH, Switzerland), Marc Raynal (Institut Français de la Vigne et du Vin, France), Marie-Laure Panon (Comité Champagne, France), Patrice Retaud (SRPV, France), Marie-Pascale Latorse (Bayer CropScience, France) and Jorge Silva (Bayer CropScience, Portugal).

We thank Mathilde Paré for assistance with data analyses during her Masters work. This work was funded by ANR grant 07-BDIV-003 (Emerfundis project) and MCF received a post-doctoral grant from the Région Ile-de-France.

References

- Ahmed S, de Labrouhe DT, Delmotte F (2012) Emerging virulence arising from hybridisation facilitated by multiple introductions of the sunflower downy mildew pathogen *Plasmopara halstedii*. *Fungal Genetics and Biology*, **49**, 847–855.
- Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P (2004) Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution*, **19**, 535–544.
- Austerlitz F, JungMuller B, Godelle B, Gouyon PH (1997) Evolution of coalescence times, genetic diversity and structure during colonization. *Theoretical Population Biology*, **51**, 148–164.
- Balloux F, Lehmann L, de Meeus T (2003) The population genetics of clonal and partially clonal diploids. *Genetics*, **164**, 1635–1644.
- Beaumont MA (1999) Detecting population expansion and decline using microsatellites. *Genetics*, **153**, 2013–2029.
- Beaumont MA (2010) Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 379–406.
- Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.
- Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology*, **19**, 2609–2625.
- Billiard S, Lopez-Villavicencio M, Hood ME, Giraud T (2012) Sex, outcrossing and mating types: unsolved questions in fungi and beyond. *Journal of Evolutionary Biology*, **25**, 1020–1038.
- Brasier CM (1991) *Ophiostoma novo-ulmi* sp. nov., causative agent of current Dutch elm disease pandemics. *Mycopathologia*, **115**, 151–161.
- Brewer MT, Milgroom MG (2010) Phylogeography and population structure of the grape powdery mildew fungus, *Erysiphe necator*, from diverse *Vitis* species. *BMC Evolutionary Biology*, **10**, 268.
- Brooks SP, Gelman A (1998) General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, **7**, 434–455.
- Brurberg MB, Elameen A, Le VH *et al.* (2011) Genetic analysis of *Phytophthora infestans* populations in the Nordic European countries reveals high genetic variability. *Fungal Biology*, **115**, 335–342.
- Chakraborty R, Jin L (1993) A unified approach to study hypervariable polymorphisms: statistical considerations of determining relatedness and population distances. *EXS*, **67**, 153–175.
- Châtagnier C (1995) La Transcaucasie au Néolithique et au Chalcolithique. *British Archaeological Series*, **624**, 1–240.
- Cornille A, Giraud T, Bellard C *et al.* (In press) Post-glacial recolonization of the European crabapple (*Malus sylvestris* Mill.), a wild relative of the domesticated apple. *Molecular Ecology*, DOI: 10.1111/mec.12231.

- Cornuet JM, Santos F, Beaumont MA *et al.* (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics*, **24**, 2713–2719.
- Cornuet JM, Ravigne V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, **11**, 401.
- Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, **25**, 410–418.
- Delmotte F, Chen WJ, Richard-Cervera S *et al.* (2006) Microsatellite DNA markers for *Plasmopara viticola*, the causal agent of downy mildew of grapes. *Molecular Ecology Notes*, **6**, 379–381.
- Delmotte F, Giresse X, Richard-Cervera S *et al.* (2008) Single nucleotide polymorphisms reveal multiple introductions into France of *Plasmopara halstedii*, the plant pathogen causing sunflower downy mildew. *Infection Genetics and Evolution*, **8**, 534–540.
- Demesure B, Comps B, Petit RJ (1996) Chloroplast DNA phylogeography in the common beech (*Fagus sylvatica* L.) in Europe. *Evolution*, **50**, 2515–2520.
- Desprez-Loustau ML, Robin C, Buee M *et al.* (2007) The fungal dimension of biological invasions. *Trends in Ecology & Evolution*, **22**, 472–480.
- Dilmaghani A, Gladieux P, Gout L *et al.* (2012) Migration patterns and changes in population biology associated with the worldwide spread of the oilseed rape pathogen *Leptosphaeria maculans*. *Molecular Ecology*, **21**, 2519–2533.
- Dumolin-Lapègue S, Demesure B, Fineschi S, Le Corre V, Petit RJ (1997) Phylogeographic structure of white oaks throughout the European continent. *Genetics*, **146**, 1475–1487.
- Durand E, Jay F, Gaggiotti OE, François O (2009) Spatial inference of admixture proportions and secondary contact zones. *Molecular Biology and Evolution*, **26**, 1963–1973.
- Dutech C, Fabreguettes O, Capdevielle X, Robin C (2010) Multiple introductions of divergent genetic lineages in an invasive fungal pathogen, *Cryphonectria parasitica*, in France. *Heredity*, **105**, 220–228.
- Dutech C, Barres B, Bridier J, Robin C, Milgroom MG, Ravigne V (2012) The chestnut blight fungus world tour: successive introduction events from diverse origins in an invasive plant fungal pathogen. *Molecular Ecology*, **21**, 3931–3946.
- Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology*, **19**, 4113–4130.
- Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, **11**, 1591–1604.
- Excoffier L, Foll M, Petit RJ (2008) Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 481–501.
- Fontaine MC, Gladieux P, Hood ME, Giraud T (In press) History of the invasion of the anther smut pathogen on *Silene latifolia* in North America. *New Phytologist*, (2013) doi: 10.1111/nph.12177
- Francois O, Durand E (2010) Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources*, **10**, 773–784.
- Francois O, Blum MG, Jakobsson M, Rosenberg NA (2008) Demographic history of European populations of *Arabidopsis thaliana*. *PLoS Genetics*, **4**, e1000075.
- Francois O, Currat M, Ray N, Han E, Excoffier L, Novembre J (2010) Principal component analysis under population genetic models of range expansion and admixture. *Molecular Biology and Evolution*, **27**, 1257–1268.
- Furrer R, Nychka D, Sain S (2011) Fields: Tools for spatial data. *R package version 6.6*.
- Galet P (1977) *Mildiou*. Lavoisier, Paris.
- Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, **10**, 305–318.
- Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–472.
- Gelman A, Carlin JB, Stern HS, Rubin DB (1995) *Bayesian Data Analysis*. Chapman & Hall, London.
- Genton BJ, Shykoff JA, Giraud T (2005) High genetic diversity in French invasive populations of common ragweed, *Ambrosia artemisiifolia*, as a result of multiple sources of introduction. *Molecular Ecology*, **14**, 4275–4285.
- Gessler C, Pertot I, Perazzolli M (2011) *Plasmopara viticola*: a review of knowledge on downy mildew of grapevine and effective disease management. *Phytopathologia Mediterranea*, **50**, 3–44.
- Giraud T, Gladieux P, Gavrillets S (2010) Linking the emergence of fungal plant diseases with ecological speciation. *Trends in Ecology and Evolution*, **25**, 387–395.
- Gobbin D, Pertot I, Gessler C (2003a) Genetic structure of a *Plasmopara viticola* population in an isolated Italian mountain vineyard. *Journal of Phytopathology*, **151**, 636–646.
- Gobbin D, Pertot I, Gessler C (2003b) Identification of microsatellite markers for *Plasmopara viticola* and establishment of a high throughput method for SSR analysis. *European Journal of Plant Pathology*, **109**, 153–164.
- Gobbin D, Jermini M, Loskill B, Pertot I, Raynal M, Gessler C (2005) Importance of secondary inoculum of *Plasmopara viticola* to epidemics of grapevine downy mildew. *Plant Pathology*, **54**, 522–534.
- Gobbin D, Rumbou A, Linde CC, Gessler C (2006) Population genetic structure of *Plasmopara viticola* after 125 years of colonization in European vineyards. *Molecular Plant Pathology*, **7**, 519–531.
- Gobbin D, Bleyer G, Keil S, Kassemeyer HH, Gessler C (2007) Evidence for sporangial dispersal leading to a single infection event and a sudden high incidence of grapevine downy mildew. *Plant Pathology*, **56**, 843–847.
- Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics*, **139**, 463–471.
- Gomez-Alpizar L, Carbone I, Ristaino JB (2007) An Andean origin of *Phytophthora infestans* inferred from mitochondrial and nuclear gene genealogies. *Proceedings of the National Academy of Sciences*, **104**, 3306–3311.
- Goss EM, Larsen M, Chastagner GA, Givens DR, Gruenwald NJ (2009) Population genetic analysis infers migration pathways of *Phytophthora ramorum* in US nurseries. *PLoS Pathogens*, **5**, e1000583.
- Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.

- Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3.2). Available at: <http://www2.unil.ch/popgen/softwares/fstat.htm>. (accessed 3 November 2012).
- Guillot G, Leblois R, Coulon A, Frantz AC (2009) Statistical methods in spatial genetics. *Molecular Ecology*, **18**, 4734–4756.
- Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual of population levels. *Molecular Ecology Notes*, **2**, 618–620.
- Hutchinson DW, Templeton AR (1999) Correlation of pairwise genetics and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, **53**, 1898–1914.
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.
- James TY, Litvintseva AP, Vilgalys R *et al.* (2009) Rapid global expansion of the fungal disease Chytridiomycosis into declining and healthy Amphibian populations. *PLoS Pathogen*, **5**, e1000458.
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.
- Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, **101**, 92–103.
- Kolbe JJ, Glor RE, Rodriguez SL, Lara AC, Larson A, Losos JB (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, **431**, 177–181.
- Lafon R, Clerjeau M (1994) *Downy Mildew*. APS Press, St. Paul.
- Le Corre V, Kremer A (1998) Cumulative effects of founding events during colonisation on genetic diversity and differentiation in an island and stepping-stone model. *Journal of Evolutionary Biology*, **11**, 495–512.
- Leblois R, Rousset F, Estoup A (2004) Influence of spatial and temporal heterogeneities on the estimation of demographic parameters in a continuous population using individual microsatellite data. *Genetics*, **166**, 1081–1092.
- Loader C (2010) Locfit: Local Regression, Likelihood and Density Estimation. *R package 1.5-6*.
- Martin AD, Quinn KM, Hee Park J (2010) MCMCpack: Markov chain Monte Carlo (MCMC) Package. *R package version 1.0-8*.
- Mboup M, Leconte M, Gautier A *et al.* (2009) Evidence of genetic recombination in wheat yellow rust populations of a Chinese overwintering area. *Fungal Genetics and Biology*, **46**, 299–307.
- Meirmans PG, Van Tienderen PH (2004) Genotype and genotype: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.
- Milgroom MG, Wang KR, Zhou Y, Lipari SE, Kaneko S (1996) Intercontinental population structure of the chestnut blight fungus, *Cryphonectria parasitica*. *Mycologia*, **88**, 179–190.
- Millardet A (1881) *Notes sur les Vignes Américaines et Opuscles Divers sur le Même Sujet*, Ferret edn. Feret, Bordeaux.
- Montarry J, Andrivon D, Glais I, Corbiere R, Mialdea G, Delmotte F (2010) Microsatellite markers reveal two admixed genetic groups and an ongoing displacement within the French population of the invasive plant pathogen *Phytophthora infestans*. *Molecular Ecology*, **19**, 1965–1977.
- Nei M (1973) Analysis of gene diversity in subdivided population. *Proceedings of the National Academy of Sciences of the United States of America*, **70**, 3321–3323.
- Oosterhout CV, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.
- Pimentel D, McNair S, Janecka J *et al.* (2001) Economic and environmental threats of alien plant, animal, and microbe invasions. *Agriculture Ecosystems & Environment*, **84**, 1–20.
- Plummer M, Best N, Cowles K, Vines K (2010) Coda: Output analysis and diagnostics for MCMC. *R package version 0.13-5*.
- Prugnolle F, Manica A, Balloux F (2005) Geography predicts neutral genetic diversity of human populations. *Current Biology*, **15**, R159–R160.
- R Development Core Team Team (2011) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN: 3-900051-07-0.
- Raboin LM, Selvi A, Oliveira KM *et al.* (2007) Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen *Ustilago scitaminea*. *Fungal Genetics and Biology*, **44**, 64–76.
- Robert S, Ravigne V, Zapater MF, Abadie C, Carlier J (2012) Contrasting introduction scenarios among continents in the worldwide invasion of the banana fungal pathogen *Mycosphaerella fijiensis*. *Molecular Ecology*, **21**, 1098–1114.
- Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.
- Rousset F (2004) *Genetic Structure and Selection in Subdivided Populations*. Monographs in population biology edn. Princeton University Press, Princeton and Oxford.
- Rousset F (2007) Inferences from spatial population genetics. In: *Handbook of Statistical Genetics* (eds Balding D, Bishop M, Cannings C), pp. 945–979. Wiley & Sons, Chichester, England.
- Rouxel M, Papura D, Nogueira M *et al.* (2012) Microsatellite markers for the characterization of native and introduced populations of *Plasmopara viticola*, the causal agent of grapevine downy mildew. *Applied and Environmental Microbiology*, **78**, 6337–6340.
- Rouxel M, Mestre P, Comont G, Lehman B, Schilder A, Delmotte F (2013) Phylogenetic and experimental evidence for host-specialized cryptic species in a biotrophic oomycete. *New Phytologist*, **197**, 251–263.
- Rumbou A, Gessler C (2006) Particular structure of *Plasmopara viticola* populations evolved under Greek island conditions. *Phytopathology*, **96**, 501–509.
- Schwartz M, McKelvey K (2009) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics*, **10**, 441–452.
- Spiegelhalter DJ, Best NG, Carlin BP, Avd L (2002) Bayesian measures of model complexity and fit. *Journal of The Royal Statistical Society Series B*, **64**, 583–639.
- Supabandhu J, Fisher MC, Vanittanakom N (2007) Polymorphic microsatellite markers for the human oomycete pathogen *Pythium insidiosum*. *Molecular Ecology Notes*, **7**, 1088–1090.
- Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, **24**, 2498–2504.
- Terral JF, Tabard E, Bouby L *et al.* (2010) Evolution and history of grapevine (*Vitis vinifera*) under domestication: new mor-

- phometric perspectives to understand seed domestication syndrome and reveal origins of ancient European cultivars. *Annals of Botany*, **105**, 443–455.
- Thines M (2011) Recent outbreaks of downy mildew on grape ivy (*Parthenocissus tricuspidata*, Vitaceae) in Germany are caused by a new species of *Plasmopara*. *Mycological Progress*, **10**, 415–422.
- This P, Lacombe T, Thomas MR (2006) Historical origins and genetic diversity of wine grapes. *Trends in Genetics*, **22**, 511–519.
- Vercauteren A, De Dobbelaere I, Gruenwald NJ *et al.* (2011) Clonal expansion of the Belgian *Phytophthora ramorum* populations based on new microsatellite markers. *Molecular Ecology*, **19**, 92–107.
- Vercken E, Fontaine MC, Gladieux P, Hood ME, Jonot O, Giraud T (2010) Glacial refugia in pathogens: European genetic structure of anther smut pathogens on *Silene latifolia* and *Silene dioica*. *PLoS Pathogen*, **6**, e1001229.
- Verdu P, Austerlitz F, Estoup A *et al.* (2009) Origins and genetic diversity of pygmy hunter-gatherers from Western Central Africa. *Current Biology*, **19**, 312–318.
- Viennot-Bourgin G (1949) *Les Champignons Parasites des Plantes Cultivées*. Masson, Paris.
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Wong FP, Burr HN, Wilcox WF (2001) Heterothallism in *Plasmopara viticola*. *Plant Pathology*, **50**, 427–432.
- Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.

M.C.F., F.D. and T.G. designed research; F.D., D.P. and S.R.C. conducted the sampling and the laboratory work; F.A. and F.L. contributed new reagents or analytical tools; M.C.F. analysed data; M.C.F., F.D., T.G. and F.A. wrote this work.

Data accessibility

Genotype data from this project are deposited at Dryad: doi:10.5061/dryad.52j07.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Sampling of *Plasmopara viticola* in vineyards.

Table S2 Priors for the ABC.

Table S3 Model checking procedure.

Table S4 Point estimates from the marginal posterior distributions for each parameter of the best-supported model (A).

Fig. S1 Allelic frequency spectrum over the European vineyards for each microsatellite locus.

Fig. S2 Estimated number of populations (K_{\max}) in the dataset assessed, from TESS clustering analyses based on the deviation index criterion (DIC).

Fig. S3 Principal component analysis (left panel) and spatial principal component analysis (right panel) conducted at individual genotype level.

Fig. S4 Allele frequency spectrum for each microsatellite locus, for the western and eastern clusters

Fig. S5 Demographic history of *Plasmopara viticola* in the western and eastern genetic clusters.

Fig. S6 Principal component analysis on the summary statistics in model checking for the 3 tested scenarios detailed in Fig. 2.

Fig. S7 Marginal posterior distributions for the demographic parameters of the best-supported demographic scenario estimated by ABC analysis.

Fig. S8 Marginal posterior distributions for the mean microsatellite mutation rate (μ_{micro}), the parameter P determining the shape of the gamma distribution of the mutation rate for individual loci, and the rate of single nucleotide insertions (SNI).