Inference of the worldwide invasion routes of the pinewood nematode, *Bursaphelenchus xylophilus*, using ABC analysis on microsatellite data

Sophie Mallez[1], Chantal Castagnone[1], Eric Lombaert[1], Philippe Castagnone-Sereno[1], Thomas Guillemaud[1]

[1] INRA, Université Côte d'Azur, CNRS, ISA, 06903 Sophia Antipolis, France

Corresponding Author:

Sophie Mallez
smallez.pro@gmail.com

ABSTRACT

Population genetics have been greatly beneficial to improve knowledge about biological invasions. Quantitative genetic methods, such as Approximate Bayesian Computation (ABC), have brought this improvement up a notch and are now essential tools to decipher the invasion routes of any invasive species. In this paper, we performed ABC analyses to shed light on the pinewood nematode (PWN) worldwide invasion routes and identify the source of the European populations. Originating from North America, this microscopic worm has been invading Asia since 1905 and Europe since 1999, causing tremendous damage on pine forests. Using microsatellite data, we demonstrated the existence of multiple independent events of introduction in Japan (one from the USA and one with an unknown origin) and China (one from the USA and one from Japan). We also found an American origin of the Portuguese samples. Although, we observed some discrepancies between traditional genetic methods and the ABC method, which are worth investigating and are discussed here, the ABC approach definitely helped clarify the worldwide history of the PWN invasion.

Keywords: population genetics, genetic diversity, differentiation, spread, Approximate Bayesian Computation, PWN.

INTRODUCTION

Biological invasions are recognized as one of the main threats to biodiversity (Walker & Steffen, 1997) and are fully integrated in global environmental changes induced by humans (Ricciardi, 2007; Sala et al., 2000; Vitousek, Dantonio, Loope, & Westbrooks, 1996; Wilcove, Rothstein, Dubow, Phillips, & Losos, 1998). Owing to their mainly irreversible aspect (Mooney & Cleland, 2001; but see Simberloff, 2009, for examples of successful eradications) and as their number is growing, studying them constitutes huge challenges at ecological, economical, societal and scientific scales (Mack et al., 2000).

Understanding biological invasions necessitates the understanding of the history of the invasive process, including the routes of invasion. The routes of invasion are defined as the pathways followed by organisms (individuals, seeds...) between their source populations and the invasive populations they formed in a new area. Their identification is a crucial step, whose accuracy is improved by the use of genetic data and analyses (Estoup & Guillemaud, 2010).

Lots of studies have shown the usefulness of genetic data and analyses to decipher the invasion routes (Boucher et al., 2013; Ciosi et al., 2008; Facon et al., 2003; Fontaine, Gladieux, Hood, & Giraud, 2013; Kelager, Pedersen, & Bruun, 2013; Papura et al., 2012; Perdereau et al., 2013; Rollins, Woolnough, Wilton, Sinclair, & Sherwin, 2009; Wan, Liu, & Zhang, 2012). Recently, quantitative methods were proposed, among which the Approximate Bayesian Computation (ABC) method (Beaumont, Zhang, & Balding, 2002). ABC relies on models and presents several advantages compared to traditional methods, including (i) the statistical comparison of different invasion scenarios, (ii) the integration of historical and biological knowledge and (iii) the evaluation of the confidence in the scenario choice. In brief, various models supposed to explain the data are compared based on the similarity

between summary statistics simulated from these models and from prior distributions of historical and genetic parameters on one hand and summary statistics computed from the actual samples on the other hand. This tool has already provided a great step forward in several invasion studies (Ascunce et al., 2011; Barres et al., 2012; Boissin et al., 2012; Boubou et al., 2012; Guillemaud et al., 2015; Lombaert et al., 2010a; Miller et al., 2005; Pascual et al., 2007; Rius, Turon, Ordonez, & Pascual, 2012).

In this study, we were interested in the invasion of the pinewood nematode (PWN), *Bursaphelenchus xylophilus* (Steiner & Buhrer, 1934; Nickle, 1970; Nematoda: Aphelenchoididae). The PWN is a microscopic worm, responsible of the pine wilt disease (Mamiya, 1972, 1976, 1983), which annually kills millions of pine trees worldwide (Mamiya, 1988; Soliman et al., 2012; Suzuki, 2002; Vicente, Espada, Vieira, & Mota, 2011). The PWN is native to North America, namely Canada and the USA (Dropkin et al., 1981; Kiritani & Morimoto, 2004). The first invasive outbreak of the PWN was observed in Asia, in Japan in 1905. The species was then observed in China in 1982, in Taiwan before 1985 and South Korea in 1988 (Futai, 2013; Mamiya, 1988; Moon, Cheon, & Lee, 2007). In 1999, the PWN was observed in Europe, first continental Portugal, and then Madeira Island and Spain (Abelleira, Picoaga, Mansilla, & Aguin, 2011; Fonseca et al., 2012; Mota et al., 1999). To date, the origin of PWN European populations is still unclear. Some authors have proposed an Asian origin for European invasive populations of PWN (Figueiredo et al., 2013; Valadas, Barbosa, Espada, Oliveira, & Mota, 2012). However, traditional genetic methods based on genetic distances and clustering analyses recently failed to distinguish between an American and an Asian origin for European outbreaks (Mallez et al., 2015). Analyses based on F_{ST} values (Weir & Cockerham, 1984) and mean individual assignment likelihoods (Paetkau, Slade, Burden, & Estoup, 2004) suggested an American origin for all Portuguese samples, while analyses based on Cavalli-Sforza and Edwards' distances (Cavalli-Sforza & Edwards,

1967), Neighbor-Joining tree (Saitou & Nei, 1987) and Bayesian clustering (Pritchard, Stephens, & Donnelly, 2000) suggested a Japanese origin for these samples. The aim of the present study was to use ABC to clarify the invasion history of the PWN by performing quantitative statistical comparison of various competing invasion scenarios, through the calculation of their posterior probabilities (Miller et al., 2005; Pascual et al., 2007).

MATERIAL & METHODS

Sampling and genotyping

In this study, 48 site samples, representing a total of 1080 individuals were analyzed: 28 site samples from the USA (554 individuals) representing the native area, and 7 site samples from Japan (210 individuals), 9 from Portugal/Madeira (169 individuals) and 4 from China (147 individuals) representing the invaded areas. The features of the samples are listed in Table 1. All the individuals were extracted from wood samples collected directly in the field, using a sieve or the Baermann method (Viglierchio & Schmitt, 1983). The historical knowledge about the invasion of the PWN and its spread within each invaded area (Abelleira et al., 2011; Fonseca et al., 2012; Futai, 2013; Mamiya, 1988; Moon et al., 2007; Mota et al., 1999) allowed us to associate a date of first observation to the different invasive outbreaks (Table 1).

The DNA extraction of each single individual was carried out by thermal shock (Castagnone, Abad, & Castagnone-Sereno, 2005) and 16 microsatellite loci were then amplified as described by Mallez et al. (2015; 2013). The locus M26 was excluded due to a

lack of amplification for one Chinese site sample. The analyses presented here were thus conducted with 15 microsatellite loci.

Standard genetic analyses

For each site sample, we computed the mean number of alleles (Mean *Na*) and the observed (*Ho*) and expected (*He*) heterozygosities per site sample with Genetix version 4.05 (Belkhir, Borsa, Chikhi, Raufaste, & Bonhomme, 1996-2004). We tested each site sample for a deviation to Hardy-Weinberg equilibrium (HWE) with Genepop version 4.1.3 (Rousset, 2008) and we quantified the inferred deviations to HWE by computing the F_{IS} estimate of Weir and Cockerham (1984) with Fstat version 2.9.3.2 (Goudet, 2002). Linkage disequilibria (LD) between all pairs of loci were also tested with Genepop (Rousset, 2008). The significance level of multiple tests of HWE and LD were adjusted by applying the false discovery rate (Benjamini & Hochberg, 1995) and sequential Bonferroni (Sokal & Rohlf, 1995) corrections, respectively.

Evolutionary relationships between the different site samples

We used traditional genetic methods to (i) determine the level of genetic structure existing in each area under study, (ii) select the site samples (especially in the native area) to include in the ABC analyses (see below) and (iii) guide the choice of the invasive scenarios to compare in these latter analyses.

First, we evaluated the level of genetic differentiation between all site sample pairs with Genepop (Rousset, 2008). Non independent multiple testing implied to adjust the significance level with the sequential Bonferroni method (Sokal & Rohlf, 1995). F_{ST} (Weir &

Cockerham, 1984), corrected for null alleles (observed for some of the site samples and loci, see Mallez et al., 2015) and D_{EST} of Jost (2008) between pairs of site samples were computed with FreeNA (Chapuis & Estoup, 2007) and SMOGD (Crawford, 2010), respectively.

Second, we performed Bayesian clustering analyses within each geographical area with Structure version 2.3 (Pritchard, Stephens, & Donnelly, 2000). These analyses were carried out only within the USA and China given that the other areas (i. e., Japan and Portugal) had been analyzed previously (Mallez et al., 2015). The number of clusters, *K*, varied from 1 to the total number of site samples per area, i.e., 18 for the USA and 4 for China. Twenty independent runs per *K* were carried out with 10^6 MCMC iterations each, following a burn-in period of $2x10^5$ iterations. For each Structure analysis, the Clumpak server (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015), which combines the Clumpp (Jakobsson & Rosenberg, 2007) and Distruct softwares (Rosenberg, 2004), was used to postprocess the Structure outputs, i.e., (i) to determine the highest level of genetic structure using the ΔK of Evanno (2005) and (ii) to identify the most frequent clustering pattern for each value of *K* (across the 20 independent runs) and to display the corresponding bar plots.

Third, the Structure software was also run with all site samples from all areas as previously described with *K* varying from 1 to 10 (which appeared to be high enough to capture the genetic pattern existing between the native and the invaded areas, see Results). For this Structure analysis, the Clumpak server (Kopelman et al., 2015) was also used to postprocess the Structure outputs as presented above. Examination of Structure results allowed us to propose putative historical relationships existing between the various site samples.

Finally, a neighbor-joining tree based on Cavalli-Sforza & Edwards' distances (Cavalli-Sforza & Edwards, 1967), corrected for null alleles with FreeNA (Chapuis & Estoup, 2007) was plotted and the mean individual assignment likelihoods (denoted Li→s, Paetkau,

Slade, Burden, & Estoup, 2004) of each invasive site sample to each putative native site sample were computed with GeneClass2 version 2.0 (Piry et al., 2004).

ABC analyses

All ABC analyses were performed with DIYABC version 2.1.0 (Cornuet et al., 2014). To define the scenarios to compare, we took into account the historical knowledge and results from this study and previous analyses (Mallez et al., 2015). Particular site samples representing genetic units inferred from Structure analyses were used for the ABC analyses. As proposed by Lombaert et al. (2014), we performed a step-by-step ABC analysis: we first tried to clarify the Asian invasion history and then integrated Europe into the scenarios. Thus, we first focused on establishing the number of introduction events in Japan and in China, independently (Figures S1 and S2 for a representation of the competing scenarios in each analysis). Then, we studied the relationships between Japan and China (Figures S3 and S4 for a representation of the competing scenarios in each analysis). Finally, we integrated the Portuguese site samples into the inferred Asian invasion scenario to determine their origin among the three possible sources: USA, Japan and China (Figure 1 for a representation of the competing scenarios in this analysis).

For each ABC analysis, we simulated one million datasets per competing scenario. The following summary statistics (SuSts) were used: the mean number of alleles, the mean expected heterozygosity and the mean allelic size variance per population and pairs of populations, the ratio of the number of alleles on the allelic size range (Garza & Williamson, 2001), the mean individual assignment likelihoods of population *i* to population *j* (Pascual et al., 2007) and the shared allele distances between populations (Chakraborty & Jin, 1993). Posterior probabilities were estimated using a polychotomous logistic regression on the 1% of

simulated datasets closest to the observed dataset (Cornuet, Ravigne, & Estoup, 2010; J. M. Cornuet et al., 2008). The LDA option, allowing to perform a discriminant analysis on the SuSts before the regression, was used (Cornuet et al., 2014; Estoup et al., 2012) to gain in computation time.

To evaluate the robustness of our inferences, we performed the ABC analyses with two different sets of prior distributions (referred to as prior set 1 and prior set 2 in the Results). The generation time of the PWN *in natura* is not well-known due to a lack of direct observations of the PWN inside the tree. Consequently, we tested the effect of choosing various generation times (15, 30 and 45 generations per year, Table S1). Thirty generations per year being the "reference", considering that the PWN reproduces from June to September and estimating an average summer temperature of 25°C where it is currently distributed. We also tested the effect of choosing different samples from the native area (NE5, NE9, NE22), the closest genetically to NE2, the sample selected based on traditional genetic analyses (see below). Then, we verified the non-overlapping of the 95 % confidence intervals of the posterior probabilities of scenarios within each ABC analysis. We calculated the posterior error rate by performing ABC analyses on 1000 simulated pseudo-observed datasets. These datasets were simulated by drawing with replacement the scenario identity and the parameter values among the 500 sets of scenarios and parameters values that generated simulated datasets closest to the observed dataset. The posterior error rate corresponds to the proportion of ABC analyses that wrongly identified scenarios and thus informs us on our ability to select the true scenario. Finally, we checked the adequacy between the observed dataset and the selected scenario, i.e., the capacity of this scenario to generate datasets similar to the observed one. Basically, 1000 new datasets were simulated by drawing the parameter values from the posterior distributions of the selected scenario. Then, each observed SuSt was confronted to the distribution formed by the SuSts obtained from the new simulated datasets, which gave a

rejection probability for each observed SuSt. A large number of probabilities being obtained, we adjusted the significance threshold with the false discovery rate method (Benjamini & Hochberg, 1995). For this purpose, two new SuSts were added to the ones used in ABC analyses: the F_{ST} (Weir & Cockerham, 1984) and the $d\mu^2$ distance (Goldstein, Linares, Cavallisforza, & Feldman, 1995) between pairs of populations.

RESULTS

Standard genetic analyses

The genetic features of the site samples are shown in Table 1. The American site samples displayed a genetic diversity low to moderate with mean Na from 1.27 to 3.87 and He from 0.08 to 0.44. More than half of the site samples were not at HWE and significant LD were observed for 13 tests out of the 1,084 realized, after correction. This may be due to the presence of null alleles, as already observed by Mallez et al. (2015).

All site samples from the invaded areas exhibited very low levels of genetic diversity owing to the presence of numerous monomorphic loci: 9 for Japan, 10 for China and 14 for Portugal/Madeira out of the 15 loci under study. In Japan, three site samples (Jap212, Jap308 and Jap120) appeared completely “identical”, with all the individuals sharing the same multi-locus genotype. In Portugal/Madeira, 162 individuals out of the 169 analyzed had the same multi-locus genotype and only one site sample, 128S, was different from the others. In China, all loci were monomorphic within site samples and three site samples out of the four studied presented the same multi-locus genotype for all their individuals.

Relationships between site samples within geographical areas

The analysis of the American site samples led to the general observation of a strong genetic structure among PWN populations in USA, in good agreement with data from a previous sampling in the same area (Mallez et al., 2015). Indeed, almost all the American site samples appeared significantly differentiated after correction for multiple tests (Fisher method of probability combination on loci, $p < 10^{-2}$). Only one test, between NC2 and GA3, was not significant (Fisher method, $p = 0.11$). F_{ST}, corrected for null alleles, reached high values and ranged from 0.036 to 0.76 (Table S2) and D_{EST} ranged from 0.0041 to 0.329 (Table S2). This strong genetic differentiation pattern was supported by the Structure analysis, which revealed biological relevant clustering of individuals within site samples for high values of K, although the ΔK inferred $K = 2$ as the main structure (Figure S5). Different clustering patterns were identified with Clumpp, but they exhibited similar results, i.e., a strong genetic structure (Figure S5). In Japan, all site samples were also significantly differentiated, except the three "identical" ones (Fisher method, $p < 10^{-5}$) with F_{ST} corrected for null alleles ranging from 0.627 to 0.995 and D_{EST} ranging from 0.0042 to 0.0938 (Table S2). In China, all site samples were "identical" except Guangdong that appeared significantly differentiated from the others (Fisher method, $p < 10^{-5}$) with a high mean corrected F_{ST} equal to 0.997 and a lower mean D_{EST} equal to 0.106 (Table S2). This result was confirmed by the Bayesian clustering analysis, which inferred a relevant clustering of individuals for $K = 2$, Guangdong being pulled apart from the others Chinese site samples (Figure S6). In Portugal/Madeira, all site samples were "identical" except the site sample 128S, that was significantly different from the others (Fisher method, $p < 10^{-5}$) with a mean corrected F_{ST} equal to 0.341 and a mean D_{EST} equal to 0.0005 (Table S2). In view of the presence of "identical" site samples within Japan, China and Portugal/Madeira, only one of these site samples was conserved in each invaded area for

simplification. The conserved site samples were the following: Jap212, Laoshan, TR1 and Mad23PC for Japan, China, Portugal and Madeira, respectively.

Relationships between site samples among geographical areas

All the site samples from different invaded areas appeared significantly differentiated between them and from the American site samples (Fisher method, $p < 10^{-5}$). NE2 was the American site sample with the lowest differentiation from site samples from invaded areas based on the D_{EST} (mean value = 0.034). Another American site sample, NE9, was selected based on the corrected F_{ST} (mean value = 0.469), NE2 being the second lowest one with this measure (mean value = 0.541). The neighbor-joining tree confirmed that the site sample NE2 is the closest site sample to all site samples from the invaded areas and that the Portuguese site samples are closer to a Japanese site sample than a Chinese or an American site samples (Figure 2). For the Bayesian clustering analysis, the ΔK inferred K=2 as the main structure (Figure S7A), in which all the invasive site samples pulled apart from the native American site samples (Figure 3). From K=3, different clustering patterns were observed (Figures 3 and S7B), in congruence with the sub-structure existing within each of the invaded and native areas. However, from K=8, the clustering pattern obtained within the invasive area appeared consistent between runs and identified four groups: one group with Mad23PC, 128S, TR1, Jap212 and Laoshan; another one with Kasumig2, Kasumig5 and Kosa; and Kasumig3 and Guangdong, that clustered separately from one another (Figure 3). Finally, almost all the site samples from the invaded areas were assigned to NE2 with the highest mean individual assignment likelihood (Figure S8). Only Guangdong and Laoshan were assigned to different samples, NE5 and Kosa, respectively (Figure S8).

Summarizing and considering the results presented above, the site sample NE2 was the closest site sample of the native area to all invasive site samples but Guangdong. Thus, it was used to represent the native area in the competing scenarios of the ABC analyses. In view of the incomplete sampling of the native area (no sample from Canada was obtained) and the strong genetic structuration observed in the USA, it is unlikely that NE2 is a sample of the true source population. An unsampled "ghost" population diverging from NE2 was thus included in all scenarios (see Figures 1, S1, S2, S3, S4 and Table S1). We also considered the following relevant genetic units to characterize the invaded areas in the following ABC analyses, such as:

- Japanese genetic unit 1, hereafter called Japan1, represented by Kosa, the closest sample to the introduction point in Japan and considered coming from the native area (Japan being the first known outbreak area of the PWN).
- Japanese genetic unit 2, hereafter called Japan2, represented by Kasumig3.
- Chinese genetic unit 1, hereafter called China1, represented by Laoshan.
- Chinese genetic unit 2, hereafter called China2, represented by Guangdong.
- Portuguese genetic unit, hereafter called Portugal, represented by TR1.

ABC analyses

In Japan, the results clearly indicated that a multiple introduction scenario is more likely than a single introduction scenario with posterior probabilities over 0.95 for prior sets 1 and 2 and non-overlapping confidence intervals, Japan2 being thus independent from Japan1 (Table 2). A similar result was observed in China. China1 was independent from China2 with posterior probabilities of 0.94 and 0.70 for prior sets 1 and 2 respectively and non-overlapping confidence intervals, suggesting two introduction events in China (Table 2). The ABC

analyses carried out between site samples from Japan and China revealed that: (i) Japan1 was probably the source of China1 and (ii) China2 is independent from Japan1 (high posterior probabilities, results robust to prior set changes, Table 2). The analysis between China2 and Japan2 gave inconsistent results, relative to the prior set used (Table 2). With the prior set 1, the scenario describing an admixture event was selected but the posterior probability was low, and a high posterior error rate was observed (Table 2). On the contrary, with the prior set 2, the scenario describing two independent events of introduction was selected with once again a low posterior probability and a high posterior error rate value (Table 2). Consequently, Japan2 was not included in the Asian scenario.

The analyses with all invaded areas studied here showed that an American origin for the Portuguese populations appeared more likely than an Asian origin, this scenario obtaining the highest posterior probability and this result being robust to prior set changes and generation time changes (Table 3). The use of the 3 samples (NE5, NE9, NE22) from the native area genetically closest to NE2, instead of NE2, gave mixed results and three analyses (out of six) pointed towards an American origin for the Portuguese populations (Table S3). Overall, there was a negative correlation between the posterior probability of scenario 1 and the genetic distance between the sample used and NE2 [correlation coefficient of -0.46 (prior set 1) and -0.64 (prior set 2) for $F_{ST,}$ and of -0.76 (prior set 1) and -0.79 (prior set 2) for D_{EST}]. We performed an independent ABC analysis with Japan2 and Portugal. In this latter analysis, we selected an American origin for Portugal based on the posterior probability (Table S4). We finally compared the final selected scenario (Scenario 1, Figure 1) to an “unique invasive bridgehead” scenario (Lombaert et al., 2010b), in which an unsampled “ghost” invasive population, originating from the native area, was the source for Japan1 (being then the source for China1), Japan2, China2 and Portugal. The scenario selected in the global analysis

(Scenario 1, Figure 1) obtained the highest posterior probability ($p > 0.85$) whatever the prior set used and thus appeared more likely than the "invasive bridgehead" scenario (Table S5).

The posterior error rates appeared quite low for all the analyses (except the Japan2 – China2 analysis, Tables 2 and 3), demonstrating that the competing scenarios were actually umiquedistinguishable. Finally, simulations from the selected scenarios and the posterior distributions of the parameters were able to produce data close to the observed ones: none of the observed SuSts was in the extreme 5% tails of the distribution of simulated SuSts after correction (Table S6).

DISCUSSION

The main objective of this study was to infer the invasion routes of the PWN using the ABC method. We demonstrated the existence of multiple independent introductions from the native area to Japan, China and Europe. This inference has not been possible using traditional population genetic methods performed in a previous study (Mallez et al., 2015). The ABC method helped clarify the worldwide history of the PWN invasion even though the post-hoc analyses tempered the conclusions. This highlights the importance of assessing the confidence in the results, especially when dealing with atypical genetic data. The discrepancies observed between traditional genetic methods and the ABC method are worth investigating and are discussed here.

Invasion routes of the PWN and the origin of Portuguese populations

Despite the low genetic diversity generally present in the PWN populations, we were able to reliably select scenarios (among the ones compared) with high posterior probabilities and good quality analyses in most cases. The most probable worldwide invasion history of the PWN identified in this study (Figure 4) includes multiple introductions in Asia and an American origin of the European samples.

The analyses carried out with the Asian site samples allowed us to reveal the existence of multiple independent events of introduction in China and Japan. Two events of introduction were inferred in both China (one from the USA and one from Japan) and Japan (one from the USA and one with an unknown origin). Especially, these results strengthen the hypothesis of two events of introduction in China, the first one near Nanjing in 1982 and the second one near Hong-Kong in 1988, already suggested in Robinet et al (2009; but see Cheng, Cheng, Xu, & Xie, 2008). Concerning Europe, the main ABC analyses of the PWN microsatellites data designated the USA as the most probable source of the Portuguese populations and excluded Japan and China as possible sources, contrary to what is often presented in the literature based on genetic distance trees (Figueiredo et al., 2013; Fonseca et al., 2012; Metge & Burgermeister, 2008; Valadas et al., 2012; Valadas, Laranjo, Mota, & Oliveira, 2013; Vieira, Burgermeister, Mota, Metge, & Silva, 2007). This result was quite robust to changes in the parameters' priors and to the posterior model checking analyses, but it depended on the samples considered in the native area, which may be considered a problem. Interestingly, our results displayed a pattern similar to that obtained by Guillemaud et al. (2010). These authors found a negative correlation between the genetic distance of the samples considered to the most probable source and the probability of an independent scenario when this scenario was true. In other words, when two invasive populations originate from two independent introductions from a native population, the more distant is the native sample from the true source, the more probable it is to erroneously point to a scenario with

two serial introductions. Therefore the "mixed" results obtained by changing the sample of the native area are expected if the true scenario includes an American origin of the European samples. Overall, the ABC method allowed us to clarify the history of the PWN invasion in Europe with good statistical confidence.

This study highlights the central role of the native area in the worldwide history of invasion of the PWN, adding up three as the minimum number of independent introductions from the USA. Only one detectable event of introduction seems to have occurred in Portugal since the first outbreak, i.e., 19 years ago. Considering the volume of global trade between Portugal and the worldwide infested areas, this result shows that quarantine and inspection efforts have been globally successful and efficient to date and confirms the importance of prevention and pro-active actions to limit new introductions (Simberloff & Rejmanek, 2011). At the European level, ecological and pedo-climatic conditions favorable to the PWN are encountered in many countries (Robinet, Van Opstal, Baker, & Roques, 2011). However, the invasion is currently limited to the Iberian Peninsula (plus Madeira Island). This reinforces the idea that the quarantine regulation measures have been efficient to date to protect Europe 1) from the entry of new invasive PWN populations from abroad, and 2) from a large dissemination of the PWN from the original Portuguese outbreak.

Different methods with different results

In the case of the PWN invasion, the results obtained from traditional genetic approaches pointed out different putative origins for the European populations of PWN (Mallez et al., 2015): American based on F_{ST} values (Weir & Cockerham, 1984) and mean individual assignment likelihoods (Paetkau et al., 2004), and Japanese based on Cavalli-Sforza and Edwards' distances (Cavalli-Sforza & Edwards, 1967), Neighbor-Joining tree

(Saitou & Nei, 1987) and Bayesian clustering (Pritchard et al., 2000). The ABC analyses, quantitative and probability-based, allowed us to improve greatly the history of the PWN invasion. Our study highlights discrepancies between the ABC methodology and some of the traditional approaches, namely the Bayesian clustering of individuals (Pritchard et al., 2000) and distance genetic trees (Saitou & Nei, 1987). To our knowledge, such discrepancies have not been observed yet, although a recent simulation study showed that classic clustering methods such as Structure analyses (Pritchard et al., 2000) may be erroneously interpreted in the context of invasion route inference (Lombaert, Guillemaud, & Deleury, 2018). The origin of the discrepancies we observed is not completely clear to us, but hypotheses may be proposed.

They may have arisen from the very low level of genetic diversity observed in the PWN populations, which might disturb the analyses of the traditional genetic methods and thus their outcomes. In an invasion context, a low genetic diversity is usually the translation of strong founder effects, which implies a strong genetic drift. Genetic drift, in turn, is a stochastic process whose consequences are random. Employing methods such as ABC that take these stochastic effects into account may thus be more reliable than other methods, such as Bayesian clustering or distance-based methods that do not (Guillemaud et al., 2010). In these conditions of low genetic diversity, we can expect traditional methods to lead to variable results when applied to various realizations of the same historical scenario. We may expect the low level of genetic diversity to produce a genetic structure pattern and/or relationship between site samples that we misinterpret (as in Lombaert et al., 2018). For instance, if the native area is weakly diversified so that it exhibits a few very frequent alleles, it is probable that two independent introductions from this native area (native → invasive 1 and native → invasive 2) lead to samples closer to each other than to their native area. When analyzed, such results would suggest serial introductions (native → invasive 1 → invasive 2, because

invasive1 and 2 are closer to each other than to native) instead of the true independent introductions scenario, because the interpretation is binary and not based on probability computation. Using a model-based stochastic method, such as ABC, a genetic structure with invading samples closer to each other than to the native area may lead to the choice of independent introduction scenario because it is probable to get such structure with this scenario when the diversity is low.

The observed discrepancies may thus just be the consequence of the absence of statistical support associated with traditional genetic methods permitting to quantitatively base the choice of evolutionary scenarios (criticism presented in A. Estoup & Guillemaud, 2010).

Carrying out simulations may help to disentangle the effect of the genetic diversity in the native area on the outcomes of clustering methods and ABC analyses, a work partly performed by Lombaert et al. (2018) who did not use ABC analyses. More precisely, analyzing dataset, simulated under chosen scenarios and known levels of genetic diversity, may allow to assess the behavior of these analyses in relation to the level of genetic diversity. These simulations would thus be useful to (i) determine the real impact of the genetic diversity on the outcomes and (ii) verify that the interpretations made from these methods in the biological invasion framework are appropriate.

ACKNOWLEDGEMENTS

We would like to thank Pedro Barbosa, Douglas LeDoux, Julia Thompson, Margarida Espada, Paulo Vieira, Jonathan D. Eisenback, Mark Harrell, Manuel Mota, Takuya Aikawa, Mitsuteru Akiba, Hajime Kosaka, Lilin Zhao and Jianghua Sun for technical support, precious help in sampling and fruitful discussions. This work was funded by the EU REPHRAME project (KBBE.2010.1.4-09).

REFERENCES

Abelleira, A., Picoaga, A., Mansilla, J. P., & Aguin, O. (2011). Detection of *Bursaphelenchus xylophilus*, causal agent of pine wilt disease on *Pinus pinaster* in Northwestern Spain. *Plant Disease, 95*(6), 776.

Ascunce, M. S., Yang, C. C., Oakey, J., Calcaterra, L., Wu, W. J., Shih, C. J., Shoemaker, D. (2011). Global Invasion History of the Fire Ant Solenopsis invicta. *Science, 331*(6020), 1066-1068. doi:10.1126/science.1198734

Barres, B., Carlier, J., Seguin, M., Fenouillet, C., Cilas, C., & Ravigne, V. (2012). Understanding the recent colonization history of a plant pathogenic fungus using population genetic tools and Approximate Bayesian Computation. *Heredity, 109*(5), 269-279. doi:10.1038/hdy.2012.37

Beaumont, M. A., Zhang, W. Y., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics, 162*(4), 2025-2035.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., & Bonhomme, F. (1996-2004). GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier Montpellier (France).

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological, 57*(1), 289-300.

Boissin, E., Hurley, B., Wingfield, M. J., Vasaitis, R., Stenlid, J., Davis, C., . . . Slippers, B. (2012). Retracing the routes of introduction of invasive species: the case of the Sirex noctilio woodwasp. *Molecular Ecology, 21*(23), 5728-5744. doi:10.1111/mec.12065

Boubou, A., Migeon, A., Roderick, G. K., Auger, P., Cornuet, J. M., Magalhaes, S., & Navajas, M. (2012). Test of Colonisation Scenarios Reveals Complex Invasion History of the Red Tomato Spider Mite Tetranychus evansi. *Plos One, 7*(4). doi:10.1371/journal.pone.0035601

Boucher, A. C., Mimee, B., Montarry, J., Bardou-Valette, S., Belair, G., Moffett, P., & Grenier, E. (2013). Genetic diversity of the golden potato cyst nematode Globodera rostochiensis and determination of the origin of populations in Quebec, Canada. *Molecular Phylogenetics and Evolution, 69*(1), 75-82. doi:10.1016/j.ympev.2013.05.020

Castagnone, C., Abad, P., & Castagnone-Sereno, P. (2005). Satellite DNA-based species-specific identification of single individuals of the pinewood nematode *Bursaphelenchus xylophilus* (Nematoda : Aphelenchoididae). *European Journal of Plant Pathology, 112*(2), 191-193. doi:10.1007/s10658-004-0580-2

Cavalli-Sforza, L. L., & Edwards, A. W. F. (1967). Phylogenetic analysis. Models and estimation procedures. *The American Journal of Humans Genetics, 19*, 233 - 257.

Chakraborty, R., & Jin, L. (1993). A unified approach to study hypervariable polymorphisms: statistical considerations of determining relatedness and population distances. *EXS, 67*, 153-175.

Chapuis, M.-P., & Estoup, A. (2007). Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution, 24*(3), 621-631. doi:10.1093/molbev/msl191

Cheng, X. Y., Cheng, F. X., Xu, R. M., & Xie, B. Y. (2008). Genetic variation in the invasive process of *Bursaphelenchus xylophilus* (Aphelenchida: Aphelenchoididae) and its possible spread routes in China. *Heredity, 100*(4), 356-365. doi:10.1038/sj.hdy.6801082

Ciosi, M., Miller, N. J., Kim, K. S., Giordano, R., Estoup, A., & Guillemaud, T. (2008). Invasion of Europe by the western corn rootworm, Diabrotica virgifera virgifera: multiple transatlantic introductions with various reductions of genetic diversity. *Molecular Ecology, 17*(16), 3614-3627.

Cornuet, J.-M., Santos, F., Beaumont, M. A., Robert, C. P., Marin, J.-M., Balding, D. J., . . . Estoup, A. (2008). Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics, 24*(23), 2713-2719. doi:10.1093/bioinformatics/btn514

Cornuet, J. M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., . . . Estoup, A. (2014). 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*(8), 1187-1189. doi:10.1093/bioinformatics/btt763

Cornuet, J. M., 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*. doi:40110.1186/1471-2105-11-401

Crawford, N. G. (2010). SMOGD: software for the measurement of genetic diversity. *Molecular Ecology Resources, 10*(3), 556-557. doi:10.1111/j.1755-0998.2009.02801.x

Dropkin, V. H., Foudin, A., Kondo, E., Linit, M. J., Smith, M., & Robbins, K. (1981). Pinewood nematode : a threat to US forests? *Plant Disease, 65*(12), 1022 - 1027.

Estoup, A., & Guillemaud, T. (2010). Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology, 19*(19), 4113-4130. doi:10.1111/j.1365-294X.2010.04773.x

Estoup, A., Lombaert, E., Marin, J.-M., Guillemaud, T., Pudlo, P., Robert, C. P., & Cornuet, J.-M. (2012). Estimation of demo-genetic model probabilities with Approximate Bayesian Computation using linear discriminant analysis on summary statistics. *Molecular Ecology Resources, 12*(5), 846-855. doi:10.1111/j.1755-0998.2012.03153.x

Evanno, G., Regnaut, S., & Goudet, J. (2005). - Detecting the number of clusters of individuals using the software structure: a simulation study. - *14*(- 8), - 2620.

Facon, B., Pointier, J. P., Glaubrecht, M., Poux, C., Jarne, P., & David, P. (2003). A molecular phylogeography approach to biological invasions of the New World by parthenogenetic Thiarid snails. *Molecular Ecology, 12*(11), 3027-3039. doi:10.1046/j.1365-294X.2003.01972.x

Figueiredo, J., Simões, M. J., Gomes, P., Barroso, C., Pinho, D., Conceição, L., . . . Egas, C. (2013). Assessment of the Geographic Origins of Pinewood Nematode Isolates via Single Nucleotide Polymorphism in Effector Genes. *Plos One, 8*(12), e83542. doi:10.1371/journal.pone.0083542

Fonseca, L., Cardoso, J. M. S., Lopes, A., Pestana, M., Abreu, F., Nunes, N., . . . Abrantes, I. (2012). The pinewood nematode, *Bursaphelenchus xylophilus*, in Madeira Island. *Helminthologia, 49*(2), 96 - 103.

Fontaine, M. C., Gladieux, P., Hood, M. E., & Giraud, T. (2013). History of the invasion of the anther smut pathogen on Silene latifolia in North America. *New Phytologist, 198*(3), 946-956. doi:10.1111/nph.12177

Futai, K. (2013). Pinewood nematode, Bursaphelenchus xylophilus. *Annual Review of Phytopathology, 51*, 61-83.

Garza, J. C., & Williamson, E. G. (2001). Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology, 10*(2), 305-318. doi:10.1046/j.1365-294X.2001.01190.x

Goldstein, D. B., Linares, A. R., Cavallisforza, L. L., & Feldman, M. W. (1995). An evaluation of genetic distances for use with microsatellite loci. *Genetics, 139*(1), 463-471.

Goudet, J. (2002). Fstat, a program to estimate and test gene diversities and fixation indices. Retrieved from http://www2.unil.ch/popgen/softwares/fstat.htm

Guillemaud, T., Beaumont, M. A., Ciosi, M., Cornuet, J.-M., & Estoup, A. (2010). Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity, 104*(1), 88 - 99. doi:doi:10.1038/hdy.2009.92

Guillemaud, T., Blin, A., Le Goff, I., Desneux, N., Reyes, M., Tabone, E., . . . Lombaert, E. (2015). The tomato borer, Tuta absoluta, invading the Mediterranean Basin, originates from a single introduction from Central Chile. *Scientific Reports, 5*, 8371. doi:10.1038/srep08371

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics, 23*(14), 1801 - 1806.

Jost, L. (2008). G(ST) and its relatives do not measure differentiation. *Molecular Ecology, 17*(18), 4015-4026. doi:10.1111/j.1365-294X.2008.03887.x

Kelager, A., Pedersen, J., & Bruun, H. (2013). Multiple introductions and no loss of genetic diversity: invasion history of Japanese Rose, Rosa rugosa, in Europe. *Biological Invasions, 15*(5), 1125-1141. doi:10.1007/s10530-012-0356-0

Kiritani, K., & Morimoto, N. (2004). Invasive insect and nematode pests from North America *Global Environmental Research, 8*(1), 75 - 88.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). -Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. - *15*(- 5), - 1191.

Lombaert, E., Guillemaud, T., Cornuet, J. M., Malausa, T., Facon, B., & Estoup, A. (2010a). Bridgehead Effect in the Worldwide Invasion of the Biocontrol Harlequin Ladybird. *Plos One, 5*(3). doi:e974310.1371/journal.pone.0009743

Lombaert, E., Guillemaud, T., Cornuet, J. M., Malausa, T., Facon, B., & Estoup, A. (2010b). Bridgehead Effect in the Worldwide Invasion of the Biocontrol Harlequin Ladybird. *Plos One, 5*(3), e9743. doi:10.1371/journal.pone.0009743

Lombaert, E., Guillemaud, T., & Deleury, E. (2018). Biases of STRUCTURE software when exploring introduction routes of invasive species. *Heredity, 120*(6), 485 - 499. doi:doi:10.1038/s41437-017-0042-1

Lombaert, E., Guillemaud, T., Lundgren, J., Koch, R., Facon, B., Grez, A., Estoup, A. (2014). Complementarity of statistical treatments to reconstruct worldwide routes of invasion: the case of the Asian ladybird *Harmonia axyridis*. *Molecular Ecology, 23*, 5979 - 5997.

Mack, R. N., Simberloff, D., Mark Lonsdale, W., Evans, H., Clout, M., & Bazzaz, F. A. (2000). Biotic invasions : Causes, Epidemiology, Global Consequences, and control. *Ecological Applications, 10*(3), 689-710.

Mallez, S., Castagnone, C., Espada, M., Vieira, P., Eisenback, J. D., Harrell, M., Guillemaud, T. (2015). Worldwide invasion routes of the pinewood nematode: What can we infer from population genetics analyses? *Biological Invasions, 17*(4), 1199 - 1213. doi:10.1007/s10530-014-0788-9

Mallez, S., Castagnone, C., Espada, M., Vieira, P., Eisenback, J. D., Mota, M., Castagnone-Sereno, P. (2013). First insights into the genetic diversity of the pinewood nematode in its native area using new polymorphic microsatellite loci. *Plos One, 8*(3), e59165. doi:10.1371/journal.pone.0059165

Mamiya, Y. (1972). Pine wood nematode *Bursaphelenchus lignicolus* Mamiya and Kiyohara, as a causal agent of pine wilting disease. *Review of Plant Protection Research, 5*, 46 - 60.

Mamiya, Y. (1976). Pine wilt disease caused by the pine wood nematode, *Bursaphelenchus lignicolus*, in Japan. *Japan Agricultural Research Quarterly, 10*(4), 206 - 211.

Mamiya, Y. (1983). Pathology of the pine wilt disease caused by *Bursaphelenchus xylophilus*. *Annual Review of Phytopathology, 21*, 201-220. doi:10.1146/annurev.py.21.090183.001221

Mamiya, Y. (1988). History of pine wilt disease in Japan. *Journal of Nematology, 20*(2), 219 - 226.

Metge, K., & Burgermeister, W. (2008). Analysis of *Bursaphelenchus xylophilus* (Nematoda: Parasitaphelenchidae) provenances using ISSR and RAPD fingerprints. In M. Mota & P. Vieira (Eds.), *Pine wilt disease: a worldwide threat to forest ecosystems* (pp. 175-186).

Miller, N., Estoup, A., Toepfer, S., Bourguet, D., Lapchin, L., Derridj, S., Guillemaud, T. (2005). Multiple transatlantic introductions of the western corn rootworm. *Science, 310*(5750), 992-992. doi:10.1126/science.1115871

Moon, Y. S., Cheon, H. M., & Lee, S. (2007). Occurence of pine wilt disease, caused by *Bursaphelenchus xylophilus*, from *Pinus koraiensis* in Korea. *Journal of Nematology, 39*(1), 87 - 88.

Mooney, H. A., & Cleland, E. E. (2001). The evolutionary impact of invasive species. *Proceedings of the National Academy of Sciences of the United States of America, 98*(10), 5446-5451. doi:10.1073/pnas.091093398

Mota, M., Braasch, H., Bravo, M. A., Penas, A. C., Burgermeister, W., Metge, K., & Sousa, E. (1999). First report of *Bursaphelenchus xylophilus* in Portugal and in Europe. *Nematology, 1*(7 - 8), 727 - 734.

Paetkau, D., Slade, R., Burden, M., & Estoup, A. (2004). Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology, 13*(1), 55 - 65. doi:10.1046/j.1365-294X.2004.02008.x

Papura, D., Burban, C., van Helden, M., Giresse, X., Nusillard, B., Guillemaud, T., & Kerdelhué, C. (2012). Microsatellite and Mitochondrial Data Provide Evidence for a Single Major Introduction for the Neartic Leafhopper *Scaphoideus titanus* in Europe. *Plos One, 7*(5), e36882. doi:10.1371/journal.pone.0036882

Pascual, M., Chapuis, M. P., Mestres, F., Balanya, J., Huey, R. B., Gilchrist, G. W., Estoup, A. (2007). Introduction history of Drosophila subobscura in the New World: a microsatellite-based survey using ABC methods. *Molecular Ecology, 16*, 3069 - 3083.

Perdereau, E., Bagnères, A. G., Bankhead-Dronnet, S., Dupont, S., Zimmermann, M., Vargo, E. L., & Dedeine, F. (2013). Global genetic analysis reveals the putative native source of the invasive termite, *Reticulitermes flavipes*, in France. *Molecular Ecology, 22*(4), 1105-1119. doi:10.1111/mec.12140

Piry, S., Alapetite, A., Cornuet, J.-M., Paetkau, D., Baudouin, L., & Estoup, A. (2004). GENECLASS2: A Software for Genetic Assignment and First-Generation Migrant Detection. *Journal of Heredity, 95*(6), 536-539. doi:10.1093/jhered/esh074

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*(2), 945-959.

Ricciardi, A. (2007). Are modern biological invasions an unprecedented form of global change? *Conservation Biology, 21*(2), 329-336. doi:10.1111/j.1523-1739.2006.00615.x

Rius, M., Turon, X., Ordonez, V., & Pascual, M. (2012). Tracking Invasion Histories in the Sea: Facing Complex Scenarios Using Multilocus Data. *Plos One, 7*(4). doi:10.1371/journal.pone.0035815

Robinet, C., Roques, A., Pan, H. Y., Fang, G. F., Ye, J. R., Zhang, Y. Z., & Sun, J. H. (2009). Role of human-mediated dispersal in the spread of the pinewood nematode in China. *Plos One, 4*(2). doi:e464610.1371/journal.pone.0004646

Robinet, C., Van Opstal, N., Baker, R., & Roques, A. (2011). Applying a spread model to identify the entry points from which the pine wood nematode, the vector of pine wilt disease, would spread most rapidly across Europe. *Biological Invasions, 13*(12), 2981-2995. doi:10.1007/s10530-011-9983-0

Rollins, L. A., Woolnough, A. P., Wilton, A. N., Sinclair, R. O. N., & Sherwin, W. B. (2009). Invasive species can't cover their tracks: using microsatellites to assist management of starling (Sturnus vulgaris) populations in Western Australia. *Molecular Ecology, 18*(8), 1560-1573. doi:10.1111/j.1365-294X.2009.04132.x

Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes, 4*(1), 137-138. doi:10.1046/j.1471-8286.2003.00566.x

Rousset, F. (2008). GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources, 8*(1), 103-106. doi:10.1111/j.1471-8286.2007.01931.x

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution, 4*(4), 406-425.

Sala, O. E., Chapin, F. S., Armesto, J. J., Berlow, E., Bloomfield, J., Dirzo, R., . . . Wall, D. H. (2000). Biodiversity - Global biodiversity scenarios for the year 2100. *Science, 287*(5459), 1770-1774. doi:10.1126/science.287.5459.1770

Simberloff, D. (2009). We can eliminate invasions or live with them. Successful management projects. *Biological Invasions, 11*(1), 149-157. doi:10.1007/s10530-008-9317-z

Simberloff, D., & Rejmanek, M. (2011). *Encyclopedia of Biological Invasions*. In U. o. C. Press (Ed.).

Sokal, R. R., & Rohlf, F. J. (1995). *Biometry. The principles and practice of statistics in biological research.* (N. Y. W. H. Freeman and Compagny Ed. Third Edition ed.).

Soliman, T., Mourits, M. C. M., van der Werf, W., Hengeveld, G. M., Robinet, C., & Lansink, A. (2012). Framework for modelling economic impacts of invasive species, applied to pine wood nematode in Europe. *Plos One, 7*(9). doi:10.1371/journal.pone.0045505

Steiner, G., & Buhrer, E. M. (1934). *Aphelenchoides xylophilus*, N. SP. A nematode associated with blue-stain and other fungi in timber. *Journal of agricultural research, 48*(10), 949 - 951.

Suzuki, K. (2002). Pine wilt disease - a threat to pine forest in Europe. *Dendrobiology, 48*, 71 - 74.

Valadas, V., Barbosa, P., Espada, M., Oliveira, S., & Mota, M. (2012). The pinewood nematode, *Bursaphelenchus xylophilus,* in Portugal : possible introductions and spread routes of a serious biological invasion revealed by molecular methods. *Nematology, 14*(8), 899 - 911.

Valadas, V., Laranjo, M., Mota, M., & Oliveira, S. (2013). Molecular characterization of Portuguese populations of the pinewood nematode Bursaphelenchus xylophilus using cytochrome b and cellulase genes. *Journal of Helminthology, 87*(4), 457 - 466.

Vicente, C., Espada, M., Vieira, P., & Mota, M. (2011). Pine Wilt Disease : a threat to European forestry. *European Journal of Plant Pathology, 133*, 89 - 99.

Vieira, P., Burgermeister, W., Mota, M., Metge, K., & Silva, G. (2007). Lack of genetic variation of *Bursaphelenchus xylophilus* in Portugal revealed by RAPD-PCR analyses. *Journal of Nematology, 39*(2), 118-126.

Viglierchio, D. R., & Schmitt, R. V. (1983). On the methodology of nematode extraction from field samples : Baermann funnel modifications. *Journal of Nematology, 15*(3), 438 - 444.

Vitousek, P. M., Dantonio, C. M., Loope, L. L., & Westbrooks, R. (1996). Biological invasions as global environmental change. *American Scientist, 84*(5), 468-478.

Walker, B. H., & Steffen, W. (1997). An overview of the implications of global change for natural and managed terrestrial ecosystems. *Conservation Ecology [online]*. Retrieved from http://www.consecol.org/vol1/iss2/art2/

Wan, X. W., Liu, Y. H., & Zhang, B. (2012). Invasion History of the Oriental Fruit Fly, Bactrocera dorsalis, in the Pacific-Asia Region: Two Main Invasion Routes. *Plos One, 7*(5). doi:10.1371/journal.pone.0036176

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the analysis of population structure. *Evolution, 38*(6), 1358-1370.

Wilcove, D. S., Rothstein, D., Dubow, J., Phillips, A., & Losos, E. (1998). Quantifying threats to imperiled species in the United States. *BioScience, 48*(8), 607-615. doi:10.2307/1313420

Table 1: Characteristics and population genetics summary statistics of the 48 samples of the pinewood nematode used in this study.

Code	Nb. Individuals	Location	First obs. (years)	Mean *Na*	*He*	*Ho*	F_{IS}
MO1	31	USA - Missouri - Columbia	-	3.33	0.35	0.29	0.19*
MO2	23	USA - Missouri - Columbia	-	3.53	0.38	0.28	0.27*
NE1	16	USA - Nebraska - Davey	-	2.07	0.21	0.19	0.09
NE2	15	USA - Nebraska - Davey	-	1.87	0.22	0.17	0.24
NE5	14	USA - Nebraska - Pawnee Lake	-	2.00	0.22	0.19	0.14
NE6	21	USA - Nebraska - Pawnee Lake	-	1.40	0.20	0.10	0.48*
NE9	29	USA - Nebraska - Pawnee Lake	-	2.73	0.33	0.25	0.26*
NE10	26	USA - Nebraska - Pawnee Lake	-	1.80	0.20	0.16	0.19*
NE12	28	USA - Nebraska - Conestoga Lake	-	1.53	0.16	0.11	0.34*
NE13b	19	USA - Nebraska - Pioneers Park	-	2.40	0.30	0.21	0.31*
NE14	28	USA - Nebraska - Pioneers Park	-	1.73	0.25	0.17	0.34*
NE15	23	USA - Nebraska - Pioneers Park	-	2.27	0.23	0.14	0.39*
NE19	16	USA - Nebraska - UNL East Campus	-	1.87	0.25	0.21	0.15
NE22	17	USA - Nebraska - Lincoln	-	2.07	0.30	0.22	0.28*
NE23	25	USA - Nebraska - Lincoln	-	1.27	0.08	0.07	0.09
NE24	19	USA - Nebraska - Lincoln	-	1.33	0.08	0.07	0.15
VI9	22	USA - Virginia - Midlothian	-	1.53	0.15	0.16	0.14
VI10	19	USA - Virginia - Virginia Beach	-	2.80	0.32	0.26	0.19
MA1	19	USA - Massachusetts - Worchester	-	1.53	0.17	0.15	0.00
NC1	12	USA - North Carolina	-	2.93	0.41	0.38	0.06
NC2	23	USA - North Carolina	-	3.27	0.38	0.30	0.22*
NY1	10	USA - New York	-	2.40	0.37	0.20	0.47*
GA1	7	USA - Georgia	-	2.00	0.27	0.23	0.16
GA2	20	USA - Georgia	-	2.67	0.36	0.31	0.13*
GA3	8	USA - Georgia	-	2.80	0.39	0.28	0.31
GA4	15	USA - Georgia	-	2.80	0.42	0.28	0.34*
KS1	17	USA - Kansas	-	3.87	0.44	0.36	0.18*
Boylston	32	USA - Mississippi - Boylston	-	2.20	0.34	0.24	0.29*
Jap120	23	Japan - Iwate - Shiwa	1978	1	-	-	-
Jap212	27	Japan - Iwate - Shiwa	1978	1	-	-	-
Jap308	25	Japan - Iwate - Shiwa	1978	1	-	-	-
Kasumig2	36	Japan - Ibaraki - Kasumigaura	1971	1	-	-	-
Kasumig3	29	Japan - Ibaraki - Kasumigaura	1971	1.20	0.07	0.07	-0.02
Kasumig5	35	Japan - Ibaraki - Kasumigaura	1971	1	-	-	-
Kosa	35	Japan - Kumamoto - Kosa	1905	1.27	0.06	0.06	0.03
GuangDong	41	Chine - Guangdong	1988	1	-	-	-
GuangXi	32	Chine - GuangXi	-	1	-	-	-
Laoshan	39	Chine - Shandong - Laoshan	1982	1	-	-	-
ShanDong	35	Chine - Shandong	1982	1	-	-	-

Mad23PC	12	Madeira Island - Porto da Cruz	2008	1	-	-	-
Mad24C	7	Madeira Island - Calheta	2008	1	-	-	-
128S	17	Portugal - Setubal - Grândola	1999	1.07	0.03	0.01	0.62*
TR1	30	Portugal - Setubal - Troia	1999	1	-	-	-
TR2	27	Portugal - Setubal - Troia	1999	1	-	-	-
AM2	21	Portugal - Setubal - Aguas de Moura	1999	1	-	-	-
Comporta	28	Portugal - Setubal - Comporta	1999	1	-	-	-
E182	13	Portugal - Coimbra - Penela	2008	1	-	-	-
E1069	14	Portugal - Viseu - Castro Daire	2008	1	-	-	-

Note: Mean *Na*, mean number of alleles per sample; *He*, expected heterozygosity and *Ho*, observed heterozygosity. F_{IS} was calculated as described by Weir & Cockerham (1984). "*" indicates that the result of the HWE test was significant at the 5% level after correction for multiple comparison. "-" indicates that *He*, *Ho* and F_{IS} were not computed for monomorphic samples on one hand and that the date of first observation was unknown or not applicable on the other hand.

Table 2: Results of the ABC analyses carried out to infer the history of the pinewood nematode invasion in Asia. Posterior probabilities and their 95% confidence intervals (CI) were obtained using a polychotomous logistic regression (J. M. Cornuet et al., 2010; J. M. Cornuet et al., 2008) and the LDA option (J. M. Cornuet et al., 2014; Arnaud Estoup et al., 2012). The selected scenario in each analysis is shown in bold.

Analysis and scenario	Posterior probability [95% CI]	PER
Analysis - Japan		
Prior set 1 - 30 generations/year		
Native → Japan1 ; Native → Japan2	**0.965 [0.962 ; 0.968]**	0.117
Native → Japan1 → Japan2	0.035 [0.032 ; 0.038]	
Prior set 2 - 30 generations/year		
Native → Japan1 ; Native → Japan2	**0.985 [0.983 ; 0.987]**	0.106
Native → Japan1 → Japan2	0.015 [0.013 ; 0.017]	
Analysis - China		
Prior set 1 - 30 generations/year		
Native → China1 ; Native → China2	**0.932 [0.928 ; 0.935]**	
Native → China1 → China2	0.018 [0.017 ; 0.020]	0.163
Native → China2 → China1	0.050 [0.048 ; 0.053]	
Prior set 2 - 30 generations/year		
Native → China1 ; Native → China2	**0.696 [0.686 ; 0.706]**	
Native → China1 → China2	0.088 [0.082 ; 0.093]	0.277
Native → China2 → China1	0.216 [0.207 ; 0.225]	
Analysis - China/Japan		
Prior set 1 - 30 generations/year		
Native → Japan1 ; Native → China1	0.256 [0.246 ; 0.265]	0.100
Native → Japan1 → China1	**0.744 [0.735 ; 0.754]**	
Native → Japan1 ; Native → China2	**0.827 [0.818 ; 0.836]**	0.092
Native → Japan1 → China2	0.173 [0.164 ; 0.182]	
Native → Japan2 ; Native → China2	0.342 [0.334 ; 0.350]	
Native → Japan2 → China2	0.170 [0.164 ; 0.176]	0.469
Native → China2 → Japan2	0.079 [0.074 ; 0.086]	
Native → China2 ; Native+China2 → Japan2	**0.409 [0.402 ; 0.416]**	
Prior set 2 - 30 generations/year		
Native → Japan1 ; Native → China1	0.112 [0.105 ; 0.119]	0.105
Native → Japan1 → China1	**0.888 [0.881 ; 0.895]**	
Native → Japan1 ; Native → China2	**0.794 [0.784 ; 0.805]**	0.089
Native → Japan1 → China2	0.206 [0.195 ; 0.216]	

Native → Japan2 ; Native → China2	**0.380 [0.372 ; 0.388]**	
Native → Japan2 → China2	0.222 [0.216 ; 0.229]	0.513
Native → China2 → Japan2	0.104 [0.099 ; 0.110]	
Native → China2 ; Native+China2 → Japan2	0.294 [0.287 ; 0.301]	

Note: PER stands for Posterior Error Rate. Native stands for the native area, represented by the sample NE2. Japan1 stands for the Japanese genetic unit 1, represented by the sample Kosa. Japan2 stands for the Japanese genetic unit 2, represented by the sample Kasumig3. China1 stands for the Chinese genetic unit 1, represented by the sample Laoshan. China 2 stands for the Chinese genetic unit 2, represented by the sample Guangdong. See Results for details.

Table 3: Results of the ABC analyses carried out to determine the origin of the pinewood nematode invasion in Portugal. Posterior probabilities and their 95% confidence intervals (CI) were obtained using a polychotomous logistic regression (J.-M. Cornuet, Ravigné, & Estoup, 2010; J.-M. Cornuet et al., 2008) and the LDA option (J. M. Cornuet et al., 2014; Arnaud Estoup et al., 2012). The selected scenario in each analysis is shown in bold.

Analysis and scenario	Posterior probability [95% CI]	PER
Prior set 1		
30 generations/year		
Native → Japan1 → China1 ; Native → China2 ; Native → Portugal	**0.529 [0.509 ; 0.550]**	0.231
Native → Japan1 → China1 ; Native → China2 → Portugal	0.013 [0.000 ; 0.040]	
Native → Japan1 → China1 and Portugal ; Native → China2	0.225 [0.203 ; 0.246]	
Native → Japan1 → China1 → Portugal ; Native → China2	0.233 [0.200 ; 0.266]	
15 generations/year		
Native → Japan1 → China1 ; Native → China2 ; Native → Portugal	**0.574 [0.546 ; 0.603]**	0.235
Native → Japan1 → China1 ; Native → China2 → Portugal	0.0210 [0.000 ; 0.067]	
Native → Japan1 → China1 and Portugal ; Native → China2	0.209 [0.174 ; 0.244]	
Native → Japan1 → China1 → Portugal ; Native → China2	0.196 [0.143 ; 0.250]	
45 generations/year		
Native → Japan1 → China1 ; Native → China2 ; Native → Portugal	**0.497 [0.479 ; 0.514]**	0.248
Native → Japan1 → China1 ; Native → China2 → Portugal	0.021 [0.000 ; 0.044]	
Native → Japan1 → China1 and Portugal ; Native → China2	0.258 [0.241 ; 0.275]	
Native → Japan1 → China1 → Portugal ; Native → China2	0.224 [0.197 ; 0.252]	
Prior set 2		
30 generations/year		
Native → Japan1 → China1 ; Native → China2 ; Native → Portugal	**0.744 [0.716 ; 0.772]**	0.158
Native → Japan1 → China1 ; Native → China2 → Portugal	0.039 [0.000 ; 0.118]	
Native → Japan1 → China1 and Portugal ; Native → China2	0.049 [0.000 ; 0.131]	
Native → Japan1 → China1 → Portugal ; Native → China2	0.168 [0.079 ; 0.257]	
15 generations/year		
Native → Japan1 → China1 ; Native → China2 ; Native → Portugal	**0.605 [0.568 ; 0.642]**	0.134
Native → Japan1 → China1 ; Native → China2 → Portugal	0.077 [0.023 ; 0.130]	
Native → Japan1 → China1 and Portugal ; Native → China2	0.123 [0.068 ; 0.178]	
Native → Japan1 → China1 → Portugal ; Native → China2	0.195 [0.123 ; 0.267]	

45 generations/year		
Native → Japan1 → China1 ; Native → China2 ; Native → Portugal	**0.598 [0.569 ; 0.627]**	
Native → Japan1 → China1 ; Native → China2 → Portugal	0.0372 [0.000 ; 0.078]	0.182
Native → Japan1 → China1 and Portugal ; Native → China2	0.082 [0.042 ; 0.124]	
Native → Japan1 → China1 → Portugal ; Native → China2	0.282 [0.234 ; 0.330]	

Note: PER stands for Posterior Error Rate. Native stands for the native area, represented by the sample NE2. Japan1 stands for the Japanese genetic unit 1, represented by the sample Kosa. Japan2 stands for the Japanese genetic unit 2, represented by the sample Kasumig3. China1 stands for the Chinese genetic unit 1, represented by the sample Laoshan. China 2 stands for the Chinese genetic unit 2, represented by the sample Guangdong. Portugal stands for the Portuguese genetic unit, represented by the sample TR1. See Results for details.

FIGURE LEGENDS

Figure 1: Competing scenarios in the ABC analyses performed to investigate the origin of the pinewood nematode invasion in Portugal. Native stands for the native area, represented by the sample NE2. Japan1 stands for the Japanese genetic unit 1, represented by the sample Kosa. Japan2 stands for the Japanese genetic unit 2, represented by the sample Kasumig3. China1 stands for the Chinese genetic unit 1, represented by the sample Laoshan. China2 stands for the Chinese genetic unit 2, represented by the sample Guangdong. Portugal stands for the Portuguese genetic unit, represented by the sample TR1. The different colors represent different effective population sizes. See Table S1 and Results for details.

Figure 2: Neighbor-Joining tree built with all the pinewood nematode samples used in this study. Tree based on Cavalli-Sforza and Edwards distances (Cavalli-Sforza and Edwards 1967), corrected for null alleles (Chapuis and Estoup, 2007). The American samples were highlighted in orange, the Japanese samples in green, the Chinese samples in purple and the Portuguese samples in blue.

Figure 3: Genetic structure of the pinewood nematode samples used in this study. Bar plots of the coefficients of co-ancestry obtained in STRUCTURE analyses with several values of K. Each bar corresponds to one individual nematode and each cluster is represented with a particular color. The major clustering patterns are shown (see Figure S7B for the minor ones, when applicable).

Figure 4: Worldwide invasion routes of the pinewood nematode as inferred from this study. The year of first infestation for every invaded area is indicated in parenthesis.

Figure 1

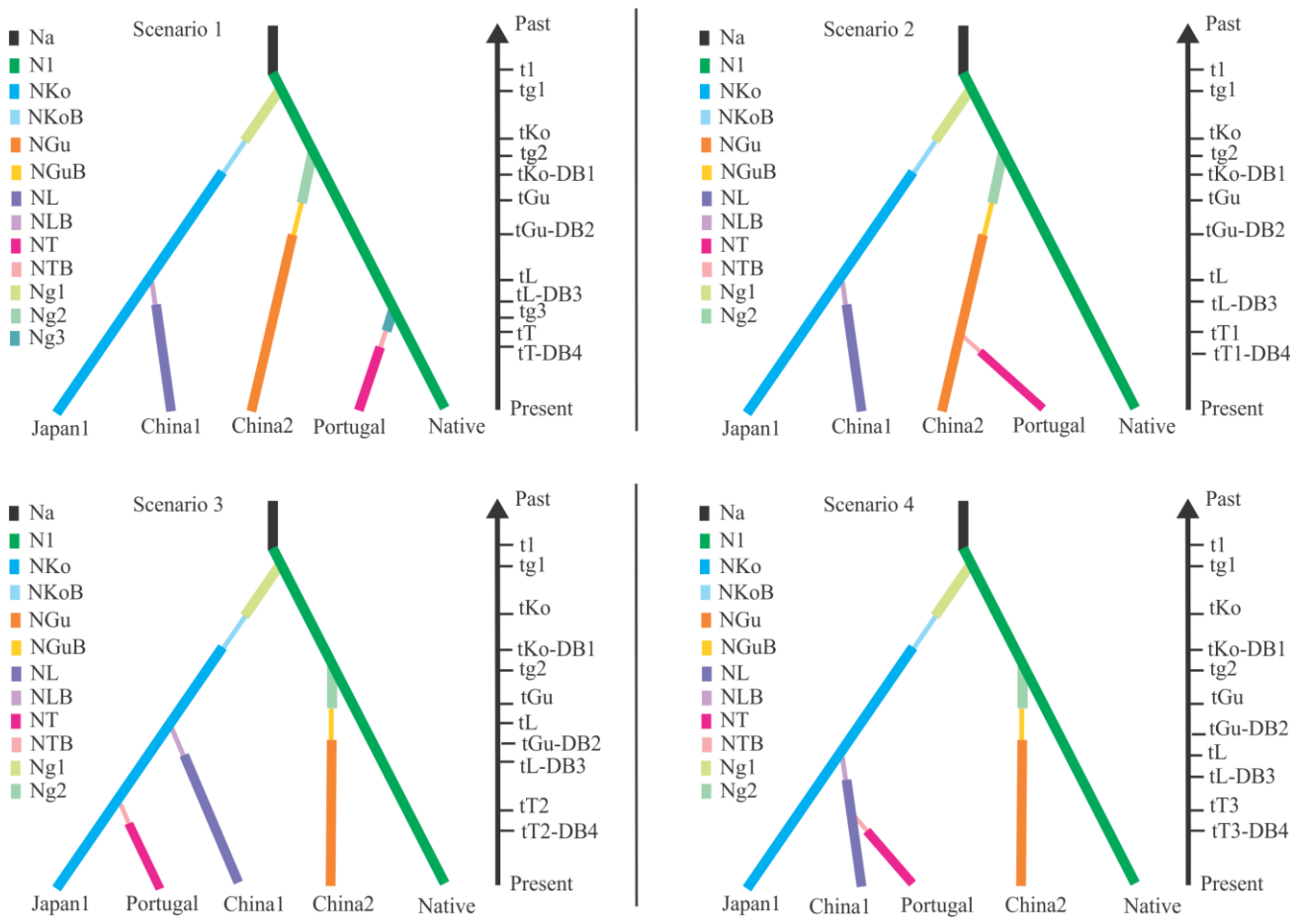

Scenario 1
Na
N1
NKo
NKoB
NGu
NGuB
NL
NLB
NT
NTB
Ng1
Ng2
Ng3
Japan1
China1
China2
Portugal
Native
Past
t1
tg1
tKo
tg2
tKo-DB1
tGu
tGu-DB2
tL
tL-DB3
tg3
tT
tT-DB4
Present
Scenario 2
Na
N1
NKo
NKoB
NGu
NGuB
NL
NLB
NT
NTB
Ng1
Ng2
Japan1
China1
China2
Portugal
Native
Past
t1
tg1
tKo
tg2
tKo-DB1
tGu
tGu-DB2
tL
tL-DB3
tT1
tT1-DB4
Present
Scenario 3
Na
N1
NKo
NKoB
NGu
NGuB
NL
NLB
NT
NTB
Ng1
Ng2
Japan1
Portugal
China1
China2
Native
Past
t1
tg1
tKo
tKo-DB1
tg2
tGu
tL
tGu-DB2
tL-DB3
tT2
tT2-DB4
Present
Scenario 4
Na
N1
NKo
NKoB
NGu
NGuB
NL
NLB
NT
NTB
Ng1
Ng2
Japan1
China1
Portugal
China2
Native
Past
t1
tg1
tKo
tKo-DB1
tg2
tGu
tGu-DB2
tL
tL-DB3
tT3
tT3-DB4
Present

Figure 2

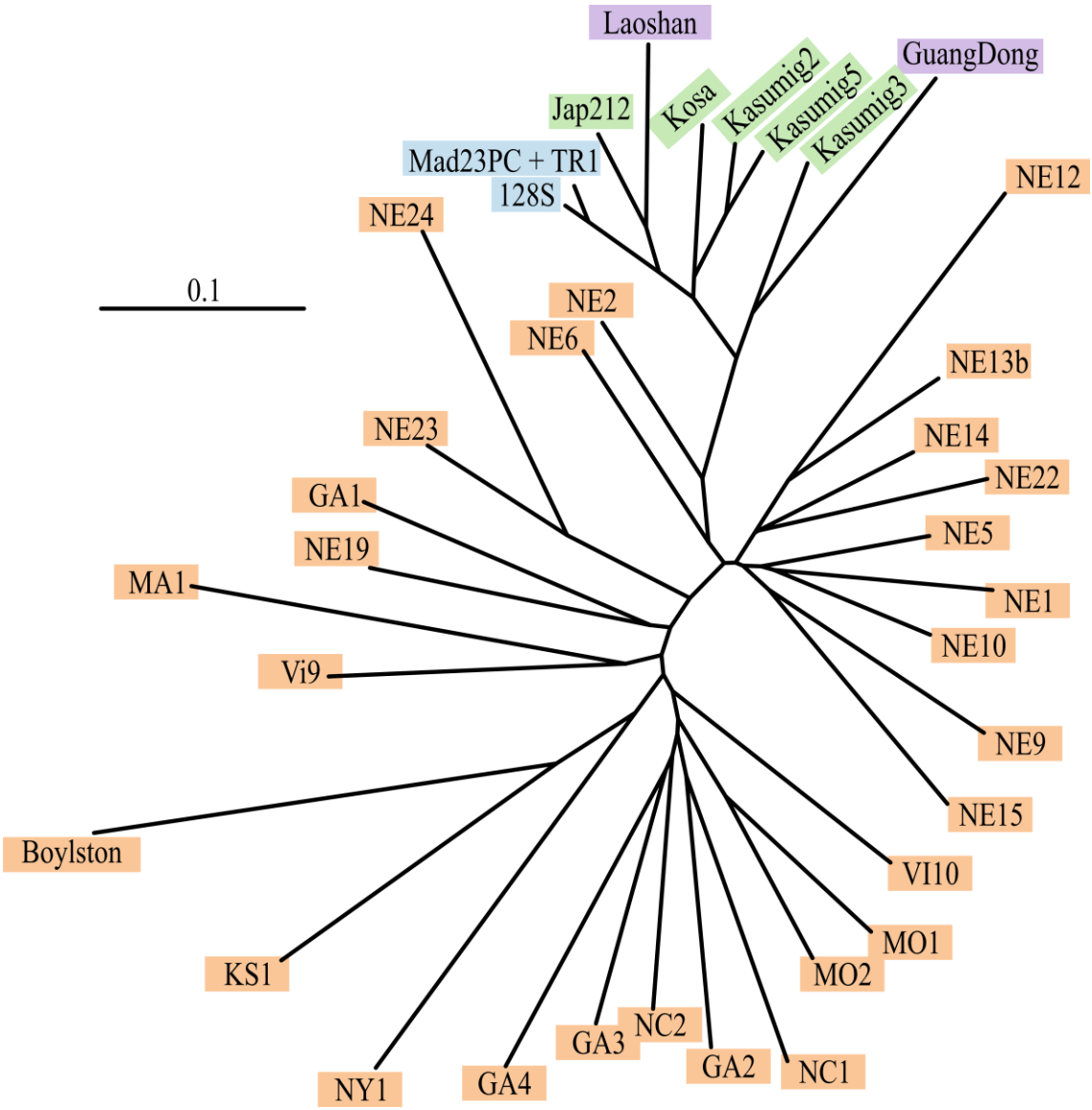

Laoshan
GuangDong
Jap212
Kosa
Kasumig2
Kasumig5
Kasumig3
Mad23PC + TR1
128S
NE12
NE24
0.1
NE2
NE6
NE13b
NE23
NE14
NE22
GA1
NE5
NE19
MA1
NE1
NE10
Vi9
NE9
NE15
Boylston
VI10
MO1
KS1
MO2
NC2
GA3
GA2
NC1
NY1
GA4

Figure 3

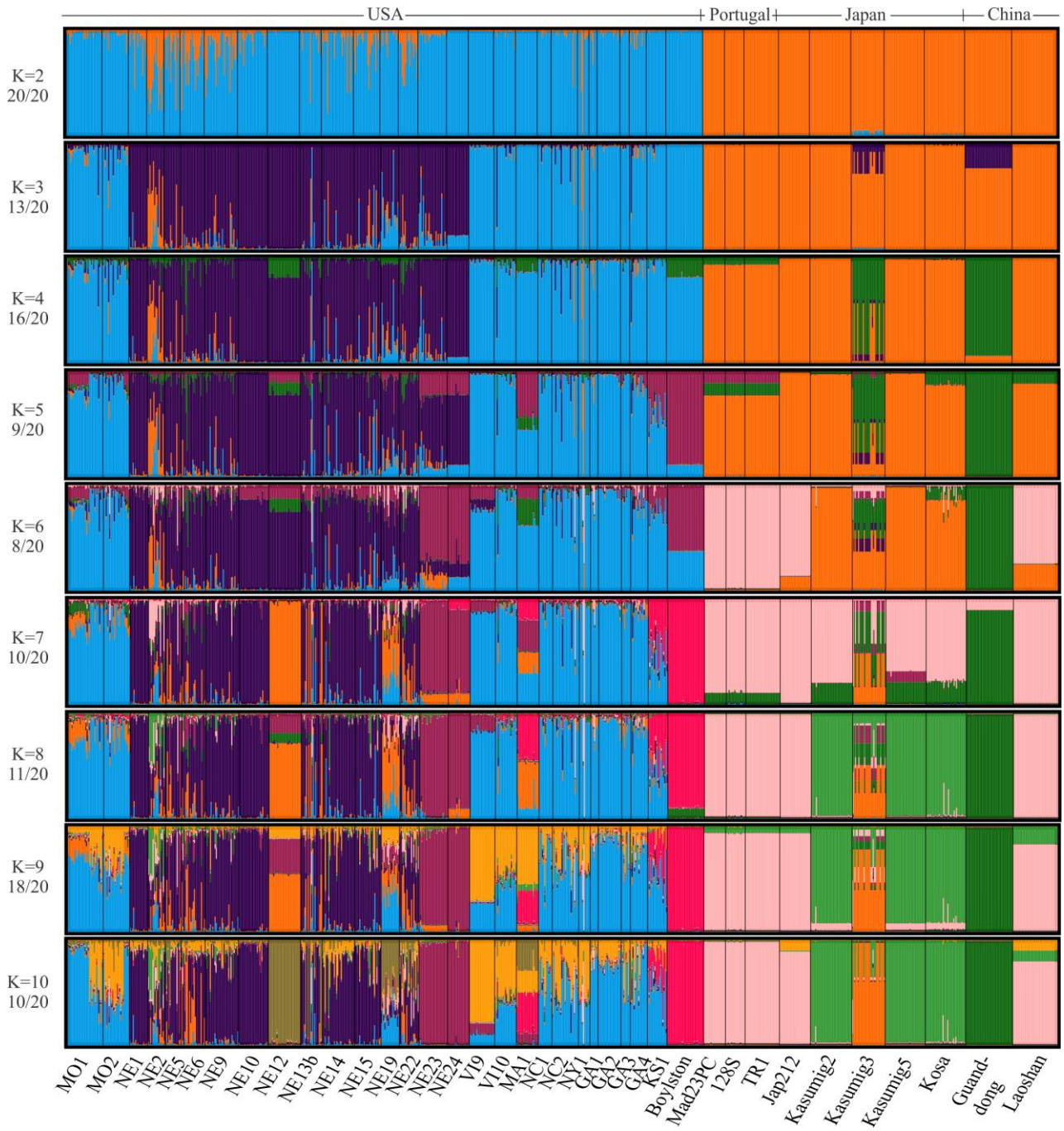
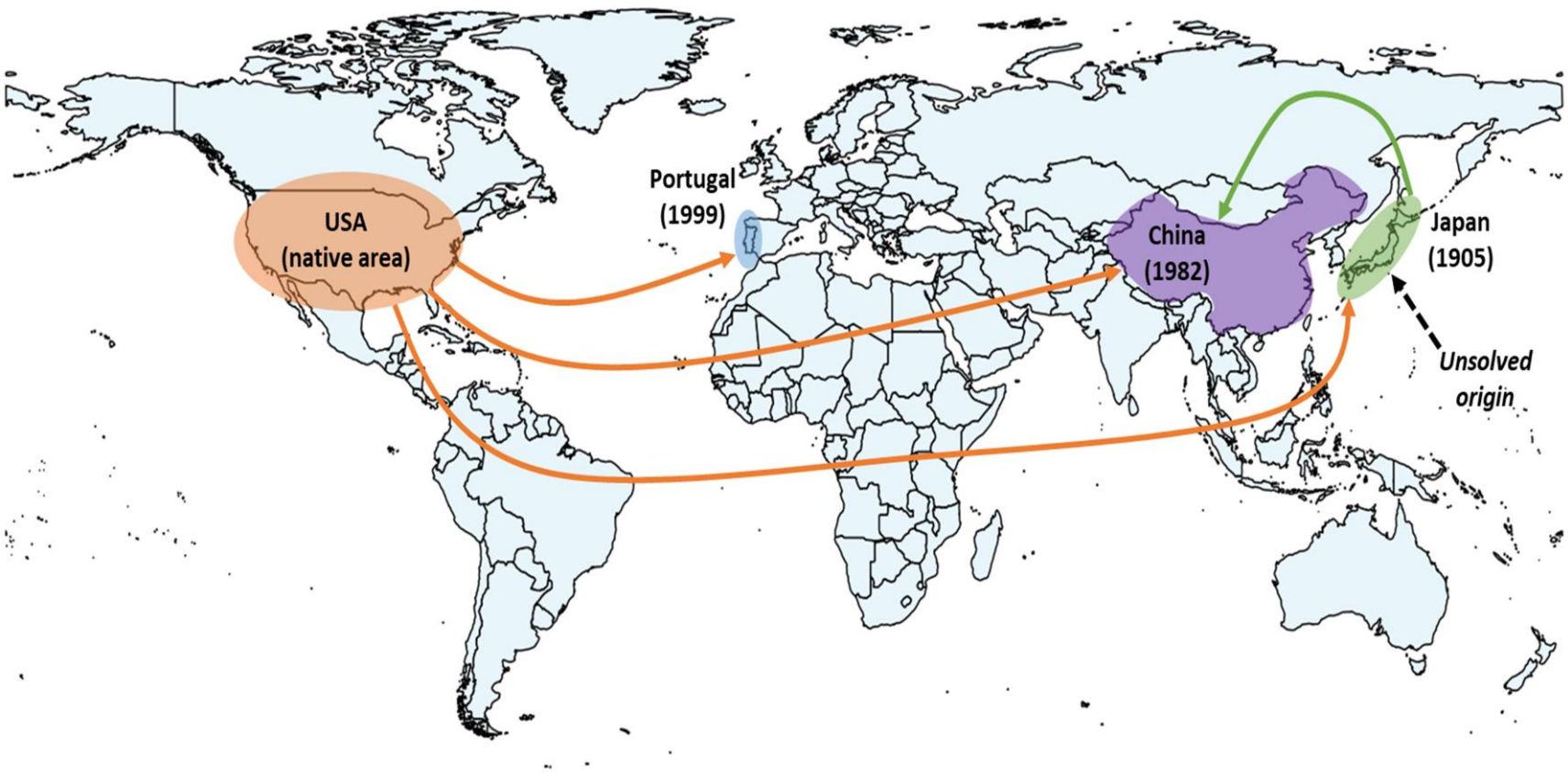

USA
Portugal
Japan
China
K=2
20/20
K=3
13/20
K=4
16/20
K=5
9/20
K=6
8/20
K=7
10/20
K=8
11/20
K=9
18/20
K=10
10/20
MO1
MO2
NE1
NE2
NE5
NE6
NE9
NE10
NE12
NE13b
NE14
NE15
NE19
NE22
NE23
NE24
VI9
VI10
MA1
NC1
NC2
NY1
GA1
GA2
GA3
GA4
KS1
Boylston
Mad23PC
128S
TR1
Jap212
Kasumig2
Kasumig3
Kasumig5
Kosa
Guand-
dong
Laoshan

Figure 4


USA
(native area)
Portugal
(1999)
China
(1982)
Japan
(1905)
Unsolved
origin