

Inference of the worldwide invasion routes of the pinewood nematode Bursaphelenchus xylophilus using approximate Bayesian computation analysis

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

▶ **To cite this version:**

Sophie Mallez, Chantal Castagnone, Eric Lombaert, Philippe Castagnone-Sereno, Thomas Guillemaud. Inference of the worldwide invasion routes of the pinewood nematode Bursaphelenchus xylophilus using approximate Bayesian computation analysis. Peer Community Journal, 2021, 1, pp.e56. 10.24072/pcjournal.64 . hal-03515563

HAL Id: hal-03515563
https://hal.science/hal-03515563

Submitted on 6 Jan 2022

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

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



Distributed under a Creative Commons Attribution 4.0 International License

Peer Community Journal

Section: Zoology

RESEARCH ARTICLE

Published
2021-12-02

Cite as
Sophie Mallez, Chantal Castagnone, Eric Lombaert, Philippe Castagnone-Sereno and Thomas Guillemaud (2021) *Inference of the worldwide invasion routes of the pinewood nematode* Bursaphelenchus xylophilus *using approximate Bayesian computation analysis*, Peer Community Journal, 1: e56.

Correspondence
eric.lombaert@inrae.fr
philippe.castagnone@inrae.fr

Peer-review
Peer reviewed and recommended by PCI Zoology, https://doi.org/10.24072/pci.zool.100006

CC BY

This article is licensed under the Creative Commons Attribution 4.0 License.

Inference of the worldwide invasion routes of the pinewood nematode *Bursaphelenchus xylophilus* using approximate Bayesian computation analysis

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

Volume **1** (2021), article e56

https://doi.org/10.24072/pcjournal.64

Abstract

Population genetics have been greatly beneficial to improve knowledge about biological invasions. Model-based genetic inference methods, such as approximate Bayesian computation (ABC), have brought this improvement to a higher level and are now essential tools to decipher the invasion routes of any invasive species. In this paper, we performed ABC random forest analyses to shed light on the pinewood nematode (PWN) worldwide invasion routes and to identify the source of 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 introduction events in Japan (at least two involving individuals originating from the USA) and China (one involving individuals originating from the USA and one involving individuals originating from Japan). We also found that Portuguese samples had a Japanese origin. We observed some discrepancies between descriptive genetic methods and the ABC method, which are worth investigating and are discussed here. The ABC method helped clarify the worldwide history of the PWN invasion, even though the results still need to be considered with some caution because the features of the PWN and the genetic markers used probably push the ABC method to its very limits.

[1]Institut Sophia Agrobiotech, INRAE - Université Côte d'Azur - CNRS, Sophia Antipolis, France

Peer Community Journal is a member of the Centre Mersenne for Open Scientific Publishing
http://www.centre-mersenne.org/

e-ISSN 2804-3871

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 identifying the history of the invasive process, including 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).

Several studies of many invasive species 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 the simulation of genetic data following different demographic scenarios defined by the user, and presents several advantages compared to descriptive methods, including (i) the statistical comparison of different invasion scenarios, (ii) the integration of historical and biological knowledge and (iii) the evaluation of the ability to select the true scenario. 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 helped reconstructing invasion histories of multiple species (Ascunce et al., 2011; Barres et al., 2012; Boissin et al., 2012; Fraimout et al., 2017; Guillemaud et al., 2015; Lombaert et al., 2010a; Miller et al., 2005; Pascual et al., 2007; Rius, Turon, Ordonez, & Pascual, 2012; Sherpa et al., 2019).

The aim of the present study was to decipher the invasion history of the pinewood nematode (PWN), *Bursaphelenchus xylophilus* (Steiner & Buhrer, 1934; Nickle, 1970; Nematoda: Aphelenchoididae). The PWN is a microscopic worm, exclusively sexual (Futai, 2013), responsible for 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 Japan in 1905. The species was then observed in China in 1982, in Taiwan before 1985 and in South Korea in 1988 (Futai, 2013; Mamiya, 1988; Moon, Cheon, & Lee, 2007). In 1999, the PWN was observed in Europe, first in Portugal, and then Spain and Madeira Island in 2008 and 2009 respectively (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, descriptive 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) and Bayesian clustering (Pritchard, Stephens, & Donnelly, 2000) suggested a Japanese origin for these samples. In this study, we complemented the 34 population samples (770 individuals) used by Mallez et al. with 14 new ones (310 individuals), and used ABC to clarify the invasion history of the PWN (Miller et al., 2005; Pascual et al., 2007).

Methods

Sampling and genotyping

In this study, 48 site samples (among which 34 already used in Mallez et al., 2015), representing a total of 1080 individuals were analyzed: 28 site samples from the USA (554 individuals) representing the native area, with a strong focus on Nebraska state (14 site samples, 286 individuals), 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 samples are listed and described in Table 1, and their geographical locations are shown in Figure S1. No permission was required to collect samples of PWN from the infested areas and we obtained an official agreement from the French authorities (#2012060-0004) for the importation and manipulation of this quarantine organism at Institut Sophia Agrobiotech facilities. All the individuals were extracted from wood samples collected directly in the field, one tree per location, using a sieve or the Baermann method (Viglierchio & Schmitt, 1983), and each site sample consisted of seven to 41 individuals (without separation according to sex or developmental stage). All samplings were performed in 2012, except for China where they were carried out in 2013. Specimens were washed several times in distilled water and stored in DESS (Yoder et al., 2006) at 4°C before DNA was extracted. 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.

Descriptive 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 from 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 descriptive genetic methods to (i) determine the level of genetic structure existing in each area under study, (ii) reduce the data and select some site samples 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). To assess differentiation between site samples, pairwise F_{ST} (Weir & Cockerham, 1984) corrected for null alleles (observed for some of the site samples and loci, see Mallez et al., 2015) were computed with FreeNA (Chapuis & Estoup, 2007). As F_{ST} is dependent on the within-population variability (Charlesworth 1998; Jost 2008; Gerlach et al. 2010), we also calculated the pairwise D_{EST} of Jost (2008) with the SMOGD software (Crawford, 2010).

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 for samples within the USA and China given that the other areas (i. e., Japan and Portugal) had been analyzed previously (Mallez et al., 2015). We used the admixture model with correlated allele frequencies. 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.

Table 1: Characteristics and population genetics summary statistics of the 48 site 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	China - Guangdong	1988	1	-	-	-
GuangXi	32	China - GuangXi	?	1	-	-	-
Laoshan	39	China - Shandong - Laoshan	1982	1	-	-	-
ShanDong	35	China - 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. In the "Code" column, "§" indicates that the corresponding site sample was previously used by Mallez et al. (2015).

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 different 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 $L_{i \to s}$, Paetkau, Slade, Burden, & Estoup, 2004) of each invasive site sample *i* to each putative source sample *s* were computed with GeneClass2 version 2.0 (Piry et al., 2004).

ABC analyses

All ABC model-choice analyses were performed with DIYABC Random Forest v1.0 (Collin et al. 2021). This program implements in the context of ABC a supervised machine-learning algorithm called the Random Forest (Breiman, 2001). Briefly, this non-parametric classification method uses hundreds of bootstrapped decision trees (creating the so-called forest) to perform classification using a set of predictor variables, here the summary statistics.

To define the scenarios to compare, we took into account the historical knowledge (i.e., dates of first observation), 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 (i.e., the oldest invasive area) and then integrated Europe (i.e., the most recent invasive area) into the scenarios. Thus, we first focused on establishing the number of independent introduction events in Japan (step 1) and in China (step 2), separately. Then, we studied the relationships between Japan and China (steps 3 and 4). Finally, we integrated the Portuguese site sample into the inferred Asian invasion scenario to determine its origin among all possible sources (steps 5 and 6). See Results for details, and Figure S2 for a graphical representation of a few competing scenarios. All configuration files are available (Mallez et al. 2021) at https://doi.org/10.5281/zenodo.4681180.

Within ABC analysis, the number of simulated datasets for each scenario was not uniform, because we took into account the dates of first observations and their prior range. For example, a Chinese population could be the source of a Japanese population in one scenario, but the opposite was more likely *a priori*. For each analysis, we simulated a mean of 20,000 datasets per competing scenario. To reduce the datasets, we used all available summary statistics (SuSts): 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), the pairwise F_{ST} (Weir & Cockerham, 1984), the shared allele (Chakraborty & Jin, 1993) and $d\mu^2$ (Goldstein, Linares, Cavalli-sforza, & Feldman, 1995) distances between pairs of populations, the maximum likelihood estimate of admixture proportion for each trio of populations (Choisy et al. 2004) and the axes obtained from a linear discriminant analysis on summary statistics (Estoup et al. 2012). We grew a classification forest of 1,000 trees based on all simulated datasets. The random forest computation applied to the observed dataset provides a classification vote representing the number of times a model is selected among the 1000 trees. The scenario with the highest classification vote was selected as the most likely scenario. We then estimated its posterior probability by way of a second random forest procedure of 1,000 trees as described by Pudlo et al. (2016).

To evaluate the robustness of our inferences, we performed the ABC analyses with two different sets of prior distributions (hereafter referred to as "main" and "alternative" prior sets; Table 2). 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. Considering that the species reproduces from June to September and estimating an average summer temperature of 25°C where it is currently distributed, thirty generations per year was the most realistic, and this value was thus used for all other ABC analyses. However, we tested the effect of choosing other generation times, 15 and 45 generations per year, with the main prior set. We also evaluate the global performance of our ABC Random Forest scenario choice, we computed the prior error rate based on the available out-of-bag simulations (i.e. simulations that are not used in tree building at each bootstrap). The

prior error rate corresponds to the proportion of ABC analyses that do not give the highest probability to the true scenario, and thus informs us on our ability to select the true scenario. Finally, we used DIYABC 2.1.0 (Cornuet et al. 2014) to check 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 (model checking analysis, Cornuet et al. 2010). To do so, we simulated a new reference table consisting of 100,000 datasets of the selected scenario. Then, 1,000 new datasets were simulated by drawing the parameter values from the posterior distributions of the selected scenario. 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).

Table 2: Definition and prior distributions of parameters used in the scenarios of invasion considered in the ABC analyses.

Parameter description	Parameter names	Gen/year	Prior distributions Main set / Alternative set
Stable effective size of the ancestral population	Na		LU[10 ; 10000] / LU[100 ; 5000]
Stable effective size of ghost populations	N_G ; NgJ1; NgJ2; NgC2		LU[10 ; 10000] / LU[100 ; 5000]
Stable effective size of the native populations	N1 ; N2 ; N3 ; N4		LU[10 ; 10000] / LU[100 ; 5000]
Stable effective size of the invasive populations	NJ1; NJ2; NJ3; NC1; NC2 ; NP		LU[10 ; 10000] / LU[100 ; 5000]
Effective size during the foundation of the invasive populations	NFJ1; NFJ2; NFJ3; NFC1; NFC2 ; NFP		LU[2 ; 10] / LU[1 ; 100]
Merging time into the ancestral population	ta		U[100 ; 10000]
Merging time of the sampled native populations	t_1 ; t_2 ; t_3 ; t_4		U[100 ; 10000]
Merging time of the source ghost native population of invasive population *i*	tgJ1 ; tgJ2 ; tgJ3 ; tC1 ; tC2 ; tP	30	U[x_i ; 10000]
		15	U[y_i ; 10000]
		45	U[z_i ; 10000]
Introduction time of invasive population *i*	tJ1 ; tJ2 ; tJ3 ; tC1 ; tC2 ; tP	30	U[x_i ; 3873]
		15	U[y_i ; 1905]
		45	U[z_i ; 5715]
Duration of the demographic bottleneck	DBJ1; DBJ2; DBJ3; DBC1 ; DBC2 ; DBP		U[0 ; 1000] / U[10 ; 500]
Rate of admixture	ra		U[0.1 ; 0.9]
Mean mutation rate	Mμ		U[10^{-6} ; 10^{-4}]
Mean parameter of the geometric distribution of the number of repeats	MP		U[0.1 ; 0.3]
Mean rate of one-nucleotide indels	Mμsni		LU[10^{-8} ; 10^{-4}]

Note: In the parameter names, J1, J2, J3, C1, C2 and P respectively stand for Japan1, Japan2, Japan3, China1, China2 and Portugal. The four native samples or NE2, NE5, NE9 and NE22. Effective sizes are given in number of diploid individuals. All times are expressed in backward numbers of generation. Time parameters x_i, y_i and z_i are respectively: 3294, 1620 and 4860 for Japan1; 1281, 630 and 1890 for Japan2; 1080, 520 and 1575 for Japan3; 945, 465 ans1395 for China1; 762, 375 and 1125 for China2; 427, 210 and 630 for Portugal. Depending on the scenarios considered, various conditions were applied to times so that coalescent times fit with each scenario's topology. The microsatellite loci were assumed to follow a generalized stepwise mutation model (Estoup et al. 2002). Each locus has a possible range of 40 contiguous allelic states and is characterized by individual μ_{loc} drawn from a gamma (mean = Mμ and shape = 2), P_{loc} drawn from a gamma (mean = MP and shape = 2) and μSNI_{loc} drawn from a gamma (mean = Mμsni and shape = 2) distribution. In the last column, U and LU respectively stand for "uniform" and "log-uniform" distributions.

Results

Descriptive 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* ranging 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). Raw data are publicly available (Mallez et al. 2021).

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 undifferentiated, 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 S1) and D_{EST} ranged from 0.0041 to 0.329 (Table S1). 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 S3). Different clustering patterns were identified with Clumpp, but they exhibited similar results, i.e., a strong genetic structure (Figure S3). In Japan, all site samples were also significantly differentiated, except the three undifferentiated ones mentioned above (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 S1). In China, all site samples were undifferentiated 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 S1). 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 S4). In Portugal/Madeira, all site samples were undifferentiated 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 S1). For samples from Japan and Portugal/Madeira, see also Mallez et al (2015; Figure S5) for additional information. In view of the presence of undifferentiated 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 1).

For the Bayesian clustering analysis, the Δ*K* inferred $K=2$ as the main structure (Figure S6A), in which all the invasive site samples pulled apart from the native American site samples (Figure 2). From $K=3$, different clustering patterns were observed (Figures 2 and S6B), 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 2). Finally, almost all the site samples from the invaded areas were assigned to NE2 with the highest mean individual assignment likelihood (Figure S7). Only Guangdong and Laoshan were assigned to different samples, NE5 and Kosa, respectively (Figure S7).

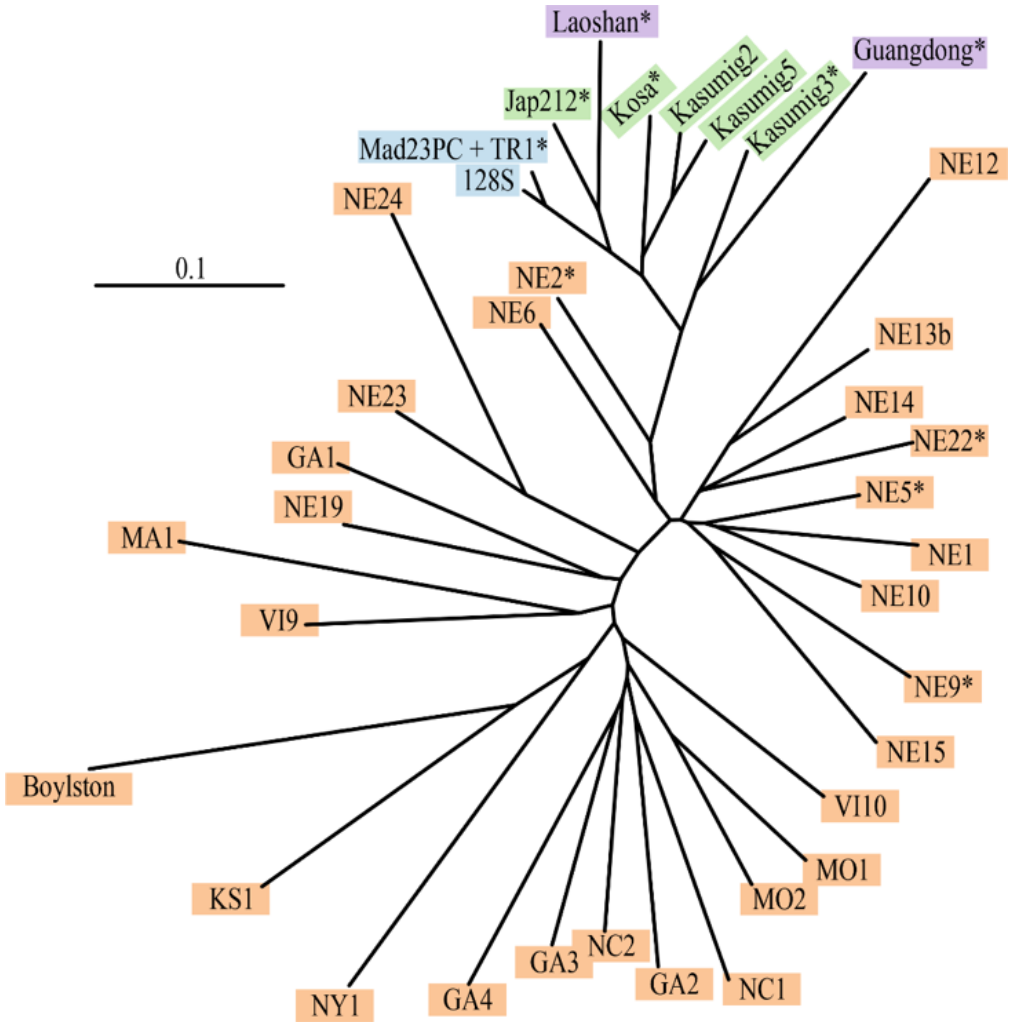


Figure 1: Neighbor-Joining tree built with all the pinewood nematode samples used in this study. Tree based on Cavalli-Sforza and Edwards distances (Cavalli-Sforza and Edwards 1967), corrected for null alleles (Chapuis and Estoup, 2007). The American (native) samples were highlighted in orange, the Japanese samples in green, the Chinese samples in purple and the Portuguese samples in blue. "*" indicates that the corresponding site sample was used in the ABC analyses.

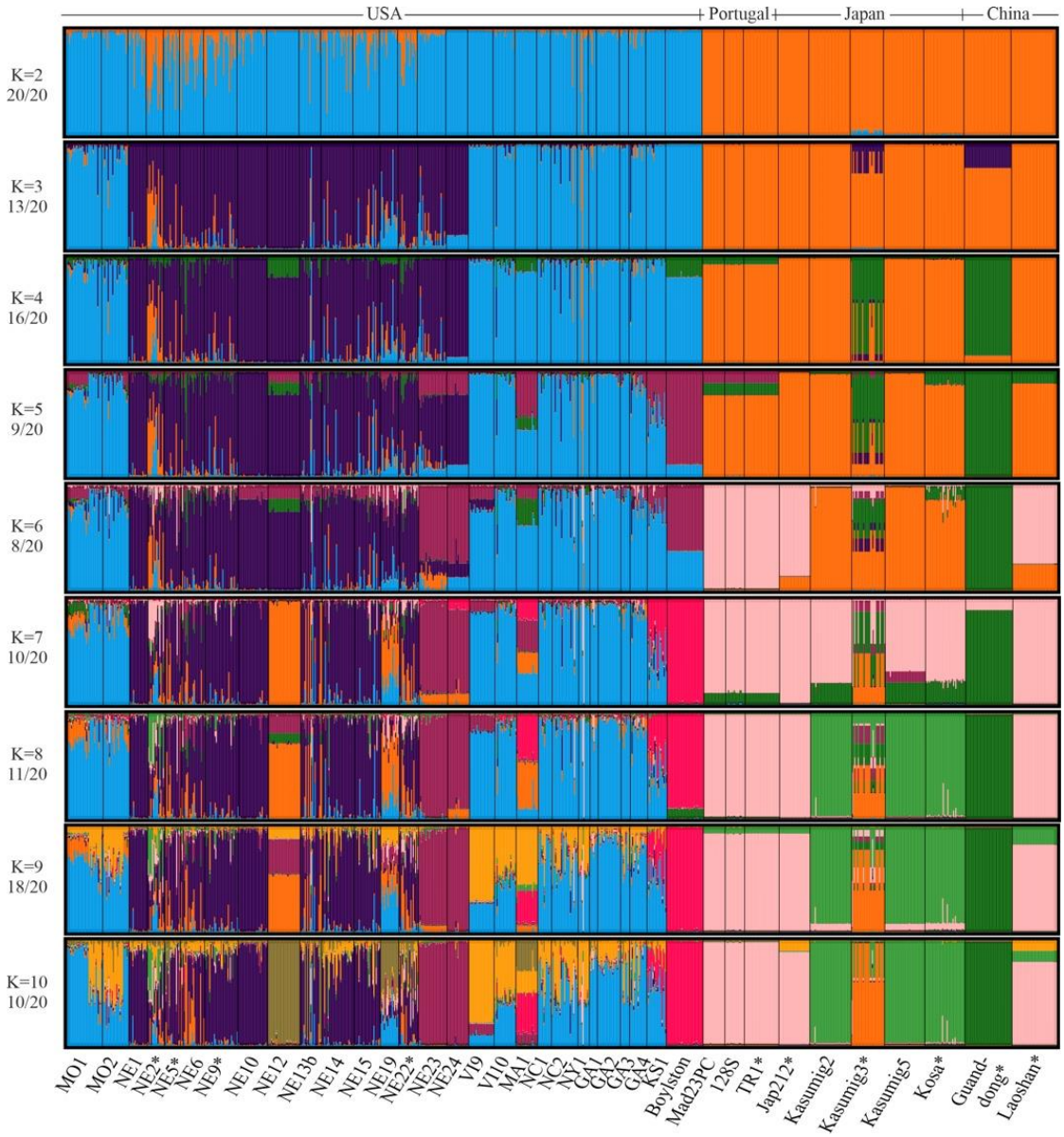


Figure 2: Genetic structure of the pinewood nematode samples used in this study. Bar plots of the coefficients of co-ancestry obtained in STRUCTURE analyses with several values of *K*. Each bar corresponds to one individual nematode and each cluster is represented with a particular color. The major clustering patterns are shown (see Figure S6B for the minor ones, when applicable). "*" indicates that the corresponding site sample was used in the ABC analyses.

Summarizing and considering the results presented above, the native site sample NE2 was the closest to all invasive site samples but Guangdong. Thus, it was used to represent the native area in the competing scenarios of the ABC analyses, together with three other genetically closed samples (NE5, NE9 and NE22). 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 one of our native sample is representative of the true source population. An unsampled "ghost" population diverging from the group of American samples was thus included in all scenarios (Figure S2). We also considered the following relevant genetic units to characterize the invaded areas in the following ABC analyses:

- 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;

- Japanese genetic unit 3, hereafter called Japan3, represented by Jap212;
- 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

The results of the first step indicated that at least two native introductions occurred in Japan, with Japan1 being independent from Japan2, and Japan1 being the source of Japan3, with posterior probability of 0.774 (Table 3, Table S2, Figure S2A). A similar result was observed in China, with the second step: China1 was independent from China2 with posterior probability of 0.963, suggesting two introduction events in China (Table 3, Table S2). The third step's ABC analysis, carried out between site samples from Japan and China, revealed that Japan1 was probably the source of China1 with posterior probability of 0.573 (Table 3, Table S2, Figure S2B). This result was then confirmed in the fourth step in which we tested, on the basis of the Structure patterns, the hypothesis that Japan2 may have derived from an admixture between China2 on the one hand, and either the native area or Japan1 on the other hand, with posterior probability of 0.567 for the scenario without admixture (Table 3, Table S2). In the fifth step, involving all invaded areas, we found that Japan1 was likely the source of the Portuguese population, with posterior probability of 0.682 (Table 3 , Table S2, Figure 3A). The same final scenario (Figures 3A and S2C) was selected in the sixth step, with posterior probability of 0.567, when it was compared to (i) a "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 itself the source for Japan3, China1 and Portugal), Japan2 and China2, and (ii) to a "unique native origin" scenario, in which Japan1, Japan2 and China2 all derived from the exact same native "ghost" population (Table 3, Table S2).

Table 3: Results of the ABC analyses carried out to infer the history of the pinewood nematode invasion. Prior error rates (PER) and posterior probabilities (post-prob) of the selected scenarios are given for the main prior set, assuming 30 generations per year. Full results for all scenarios, and all prior sets and time scales are shown in Table S2.

Step (nb SuSts)	Nb scenarios	PER	Selected scenario	Post-prob
1: Japan (301)	8	0.273	S4: Nat→Japan1→Japan3 ; Nat→Japan2	0.774
2: China (204)	3	0.171	S1: Native→China1 ; Native→China2	0.963
3: Japan/China (576)	24	0.338	S2: Nat→Japan1→Japan3 & China1 ; Nat→Japan2 ; Nat→China2	0.573
4: Japan/China admixture (576)	3	0.311	S1: Scenario 2 of step 3	0.567
5: Portugal (760)	8	0.208	S2: Nat→Japan1→Japan3 & China1 & Port ; Nat→Japan2 ; Nat→China2	0.682
6: Invasive bridgehead (760)	3	0.356	S1: Scenario 2 of step 5	0.567

Note: "Nat" stands for the native area, represented by a group of 4 samples (NE2, NE5, NE9 and NE22). "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. "Japan3" stands for the Japanese genetic unit 2, represented by the sample Jap212. "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. "Port" stands for the Portuguese genetic unit, represented by the sample TR1.

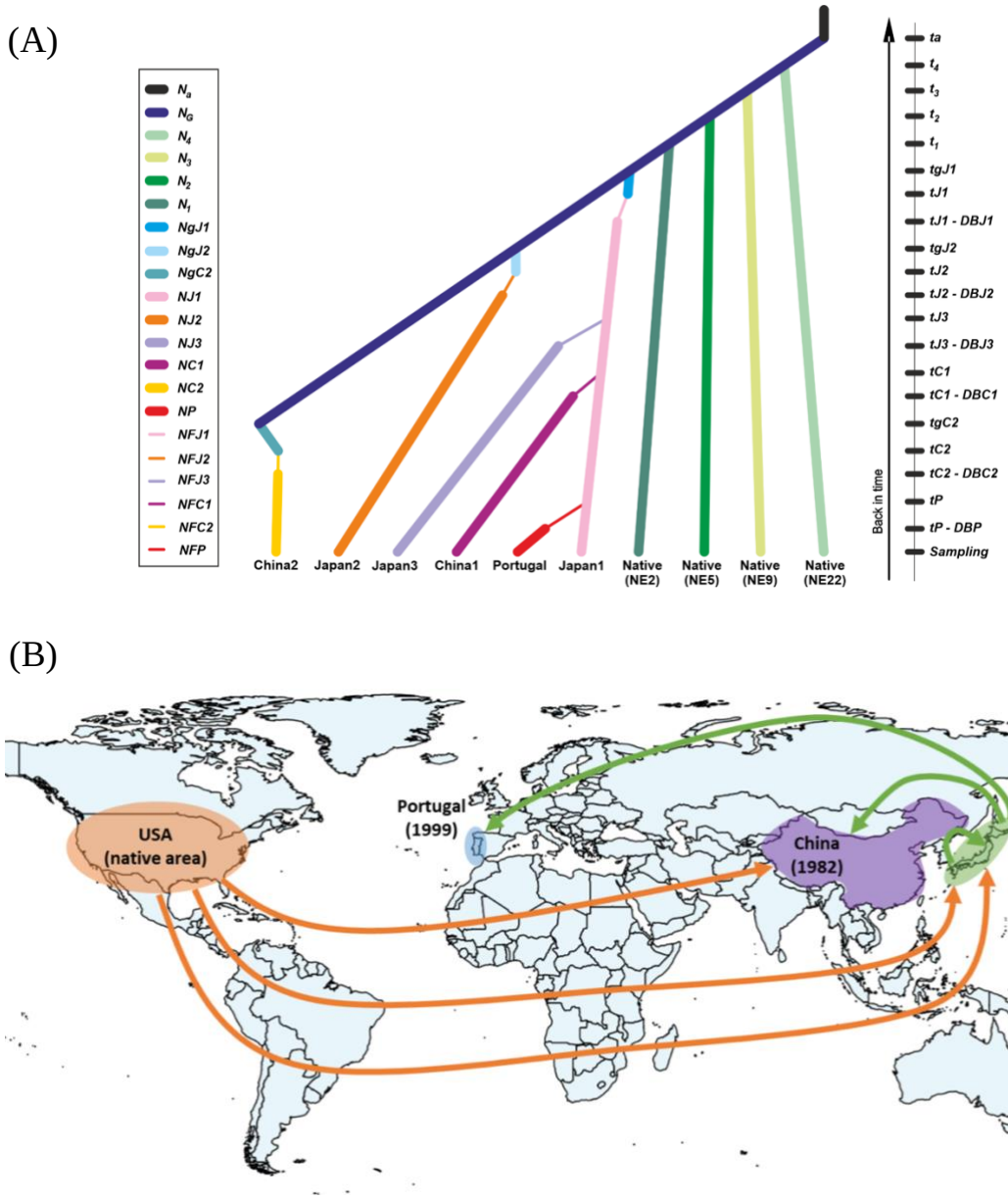


Figure 3: Worldwide invasion routes of the pinewood nematode as inferred from ABC-RF analyses. (A) Graphical illustration of the final selected scenario (step 6, scenario1); (B) Geographic representation of final selected scenario. The year of first infestation for every invaded area is indicated in parenthesis.

For each step, the use of an alternative prior set, a smaller time scale (15 generations/year) or a larger time scale (45 generations/year) essentially confirmed the main results. The exceptions were for steps 1 and 3 for which one analysis out of four provided the largest number of votes for a different scenario (Table S2). The prior error rates appeared moderate for all the analyses (Table 3, Table S2), reflecting a decent, albeit imperfect, ability to distinguish between competing scenarios. Finally, the model checking analysis showed that posterior simulations of the final selected scenario were able to produce data close to the observed ones: none of the 760 observed SuSts were in the extreme 5% tails of the distribution of simulated SuSts after correction by the false discovery rate method, whatever the prior set or time scale used (Table S3).

Discussion

The main objective of this study was to infer the invasion routes of the PWN using the ABC method. We demonstrated the existence (i) of multiple independent introductions from the native area to Japan and China, and (ii) of a bridgehead population, located in Japan, and which was the secondary source of several populations in Asia and Europe. This inference has not been possible using descriptive population genetics methods performed in a previous study (Mallez et al., 2015). Conversely, the ABC method helped clarify the worldwide history of the PWN invasion, even though the results still need to be considered with some caution. The discrepancies observed between the various methods 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 good posterior probabilities in most cases and satisfactory quality analyses. The most probable worldwide invasion history of the PWN identified in this study (Figure 3B) includes multiple introductions in Asia and a Japanese origin of the European samples.

The analyses allowed us to reveal the existence of multiple independent events of introduction from the native area to Asia, the USA being the source of two outbreaks in Japan and one in China. Within Asia, the oldest Japanese population has partly spread across the country, but has also acted as a source for at least one population in China. 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 by modeling approaches in Robinet et al (2009; but see Cheng, Cheng, Xu, & Xie, 2008). Concerning Europe, the ABC analyses of the PWN microsatellites data designated the oldest Japanese population as the most probable source of the Portuguese populations and excluded the native area as a possible source. This result thus confirm 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 robust to changes in the parameters' priors and displayed good posterior model checking results.

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. The other striking result is the bridgehead status of the invasive population represented by our sample of Kosa in Japan, which has a significant role in Asia, but which is also the source of the European invasion. Noteworthy, only one detectable event of introduction seems to have occurred in Portugal. 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 to minimize the number of independent outbreaks 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 from the entry of new invasive PWN populations from abroad, and 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 descriptive 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) 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 descriptive approaches, such as the F_{ST} and the mean assignment likelihoods. It is interesting to notice that these two statistics are among those chosen to summarize the data in the ABC analyses. However, in addition to using other key statistics, the ABC method allows us to consider complex evolutionary scenarios, involving for example intense bottlenecks or unsampled populations.

Indeed, previous studies using simulated datasets have shown that F_{ST} and mean assignment likelihoods, used alone, can be misleading in some situations (Guillemaud et al., 2010; Lombaert et al., 2010; 2018).

On the contrary, the inference of multiple independent introductions from the native area in Asia (Japan1, Japan2 and China2) is not consistent with 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 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 likely have arisen from the very low level of genetic diversity observed in the PWN populations, which might disturb the analyses of the descriptive 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 that do not, such as Bayesian clustering or distance-based methods (Guillemaud et al., 2010; Estoup & Guillemaud, 2010). In these conditions of low genetic diversity, we can expect descriptive 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. 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 appear closer to each other than to native) instead of the true independent introduction scenario, because the interpretation is binary and not based on probability computation. Indeed, the use of simulated datasets enabled Lombaert et al. (2018) to show that multiple introductions from a native population with small effective size, and thus low diversity, could often lead to single introduction patterns with Structure. Conversely, 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.

Overall, the ABC method, explicitly based on models, seems appropriate for the PWN which is characterized by a markedly low intra-population diversity and a very strong inter-population genetic structure. However, one should remain cautious when interpreting our results. Indeed, the features of the PWN and of the genetic markers used probably push the ABC method to its very limits. Particularly noteworthy are the sometimes relatively high prior error rates, and the few – albeit rare – scenario selection shifts when using different time scales. Finally, a factor that is probably of particular importance is the incompleteness of our sampling scheme, especially in the native area. The genetic structure is so strong in the latter that the use of ghost populations in our analyses may not be sufficient to capture the full complexity of the true evolutionary history.

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 datasets, 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, (ii) assess the expected benefits of increasing the number of genetic markers and (iii) verify that the interpretations made from these methods in the biological invasion framework are appropriate.

Data accessibility

Complete dataset (microsatellite), STRUCTURE outputs and ABC configuration files were deposited at Zenodo: https://doi.org/10.5281/zenodo.4681180

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). A previous version of this article has been peer-reviewed and recommended by Peer Community In Zoology (https://doi.org/10.24072/pci.zool.100006)

Conflict of interest disclosure

The authors of this article declare that they have no financial conflict of interest with the content of this article. TG is co-founder of PCI and TG and EL are recommenders for PCI Evol Biol, PCI Ecol and PCI Zool.

Author Contributions

SM and CC performed microsatellite genotyping. SM performed the analyses, with input from EL. SM, EL, PCS and TG interpreted the results. SM wrote the manuscript, with guidance from PCS and TG, and input from EL. All authors have read and approved the final manuscript.

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. https://doi.org/10.1094/PDIS-12-10-0902

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. https://doi.org/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. https://doi.org/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. https://www.genetics.org/content/162/4/2025

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. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

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. https://doi.org/10.1111/mec.12065

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. https://doi.org/10.1016/j.ympev.2013.05.020

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

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. https://doi.org/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. https://doi.org/10.1007/978-3-0348-8583-6_14

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

Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, *15*(5), 538–543. https://doi.org/10.1093/oxfordjournals.molbev.a025953

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. https://doi.org/10.1038/sj.hdy.6801082

Choisy, M., Franck, P., & Cornuet, J. M. (2004). Estimating admixture proportions with microsatellites: comparison of methods based on simulated data. *Molecular Ecology*, *13*(4), 955–968. https://doi.org/10.1111/j.1365-294X.2004.02107.x

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. https://doi.org/10.1111/j.1365-294X.2008.03866.x

Collin, F.-D., Durif, G., Raynal, L., Lombaert, E., Gautier, M., Vitalis, R., … Estoup, A. (2021) Extending Approximate Bayesian Computation with Supervised Machine Learning to infer demographic history from genetic polymorphisms using DIYABC Random Forest. Manuscript submitted for publication.

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. https://doi.org/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. https://doi.org/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, 401*. https://doi.org/10.1186/1471-2105-11-401

Crawford, N. G. (2010). SMOGD: software for the measurement of genetic diversity. *Molecular Ecology Resources, 10*(3), 556-557. https://doi.org/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. https://doi.org/10.1094/PD-65-1022

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

Estoup, A., Jarne, P., & Cornuet, J. M. (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, 11(9), 1591–1604. https://doi.org/10.1046/j.1365-294X.2002.01576.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. https://doi.org/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), 2611-2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

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. https://doi.org/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. https://doi.org/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. https://doi.org/10.2478/s11687-012-0020-3

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. https://doi.org/10.1111/nph.12177

Fraimout, A., Debat, V., Fellous, S., Hufbauer, R. A., Foucaud, J., Pudlo, P., … Estoup, A. (2017). Deciphering the Routes of invasion of *Drosophila suzukii* by Means of ABC Random Forest. *Molecular Biology and Evolution*, *34*(4), 980–996. https://doi.org/10.1093/molbev/msx050

Futai, K. (2013). Pinewood nematode, *Bursaphelenchus xylophilus*. *Annual Review of Phytopathology, 51*, 61-83. https://doi.org/10.1146/annurev-phyto-081211-172910

Gerlach, G., Jueterbock, A., Kraemer, P., Deppermann, J., & Harmand, P. (2010). Calculations of population differentiation based on GST and D: Forget GST but not all of statistics. *Molecular Ecology*, *19*(18), 3845–3852. https://doi.org/10.1111/j.1365-294X.2010.04784.x

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

Goldstein, D. B., Linares, A. R., Cavalli-sforza, L. L., & Feldman, M. W. (1995). An evaluation of genetic distances for use with microsatellite loci. *Genetics, 139*(1), 463-471. https://doi.org/10.1093/genetics/139.1.463

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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1093/bioinformatics/btm233

Jost, L. (2008). G(ST) and its relatives do not measure differentiation. *Molecular Ecology, 17*(18), 4015-4026. https://doi.org/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. https://doi.org/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), 1179-1191. https://doi.org/10.1111/1755-0998.12387

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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1111/mec.12989

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. https://doi.org/10.1890/1051-0761(2000)010[0689:BICEGC]2.0.CO;2

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. https://doi.org/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. https://doi.org/10.1371/journal.pone.0059165

Mallez, S., Castagnone, C., Lombaert, E., Castagnone-Sereno, P., and Guillemaud, T. (2021). Inference of the worldwide invasion routes of the pinewood nematode: Datasets, STRUCTURE outputs, ABC configuration files, sup information [Data set]. Zenodo. http://doi.org/10.5281/zenodo.4681180

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. https://doi.org/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). https://doi.org/10.1007/978-1-4020-8455-3_15

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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1163/156854199508757

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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1111/j.1365-294X.2007.03336.x

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. https://doi.org/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. https://doi.org/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.

Pudlo, P., Marin, J. M., Estoup, A., Cornuet, J. M., Gautier, M., & Robert, C. P. (2016). Reliable ABC model choice via random forests. *Bioinformatics*, *32*(6), 859–866. https://doi.org/10.1093/bioinformatics/btv684

Ricciardi, A. (2007). Are modern biological invasions an unprecedented form of global change? *Conservation Biology, 21*(2), 329-336. https://doi.org/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). https://doi.org/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), e4646. https://doi.org/10.1371/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1126/science.287.5459.1770

Sherpa, S., Blum, M. G. B., Capblancq, T., Cumer, T., Rioux, D., & Després, L. (2019). Unravelling the invasion history of the Asian tiger mosquito in Europe. *Molecular Ecology*, *28*(9), 2360–2377. https://doi.org/10.1111/mec.15071

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

Simberloff, D., & Rejmanek, M. (2011). *Encyclopedia of Biological Invasions*. In U. o. C. Press (Ed.). https://doi.org/10.1525/9780520948433

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). https://doi.org/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. https://doi.org/10.1163/156854112X632673

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. https://doi.org/10.1017/S0022149X12000673

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. https://doi.org/10.1007/s10658-011-9924-x

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). https://doi.org/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. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

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. https://doi.org/10.2307/1313420

Yoder, M., Tandingan De Ley, I., King, I.W., Mundo-Ocampo, M., Mann, J., Blaxter, M., Poiras, L. & De Ley, P. (2006). DESS: A versatile solution for preserving morphology and extractable DNA of nematodes. *Nematology*, 8(3), 367-376. https://doi.org/10.1163/156854106778493448