1   **Inference of the worldwide invasion routes of the pinewood nematode *Bursaphelenchus***
2   ***xylophilus* using approximate Bayesian computation analysis**
3
4
5   Sophie Mallez[1], Chantal Castagnone[1], Eric Lombaert[1], Philippe Castagnone-Sereno[1], Thomas
6   Guillemaud[1]
7
8
9   [1] INRAE, Université Côte d'Azur, CNRS, ISA, Sophia Antipolis, France
10
11
12
13   **Corresponding Authors:**
14   Data analysis: eric.lombaert@inrae.fr
15   Pinewood nematode biology: philippe.castagnone@inrae.fr
16
17
18
19
20   **ABSTRACT**
21
22   Population genetics have been greatly beneficial to improve knowledge about biological invasions.
23   Model-based genetic inference methods, such as approximate Bayesian computation (ABC), have
24   brought this improvement to a higher level and are now essential tools to decipher the invasion
25   routes of any invasive species. In this paper, we performed ABC analyses to shed light on the
26   pinewood nematode (PWN) worldwide invasion routes and to identify the source of European
27   populations. Originating from North America, this microscopic worm has been invading Asia since
28   1905 and Europe since 1999, causing tremendous damage on pine forests. Using microsatellite data,
29   we demonstrated the existence of multiple introduction events in Japan (one involving individuals
30   originating from the USA and one involving individuals with an unknown origin) and China (one
31   involving individuals originating from the USA and one involving individuals originating from Japan).
32   We also found that Portuguese samples had an American origin. Although we observed some
33   discrepancies between descriptive genetic methods and the ABC method, which are worth
34   investigating and are discussed here, the ABC approach definitely helped clarify the worldwide
35   history of the PWN invasion.
36
37
38
39   **Keywords:** population genetics, genetic diversity, differentiation, spread, approximate Bayesian
40   computation, PWN.

41 **INTRODUCTION**
42

43 Biological invasions are recognized as one of the main threats to biodiversity (Walker & Steffen, 1997)
44 and are fully integrated in global environmental changes induced by humans (Ricciardi, 2007; Sala et
45 al., 2000; Vitousek, Dantonio, Loope, & Westbrooks, 1996; Wilcove, Rothstein, Dubow, Phillips, &
46 Losos, 1998). Owing to their mainly irreversible aspect (Mooney & Cleland, 2001; but see Simberloff,
47 2009, for examples of successful eradications) and as their number is growing, studying them
48 constitutes huge challenges at ecological, economical, societal and scientific scales (Mack et al.,
49 2000).

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

55   Several studies of many invasive species have shown the usefulness of genetic data and
56 analyses to decipher the invasion routes (Boucher et al., 2013; Ciosi et al., 2008; Facon et al., 2003;
57 Fontaine, Gladieux, Hood, & Giraud, 2013; Kelager, Pedersen, & Bruun, 2013; Papura et al., 2012;
58 Perdereau et al., 2013; Rollins, Woolnough, Wilton, Sinclair, & Sherwin, 2009; Wan, Liu, & Zhang,
59 2012). Recently, quantitative methods were proposed, among which the approximate Bayesian
60 computation (ABC) method (Beaumont, Zhang, & Balding, 2002). ABC relies on the simulation of
61 genetic data following different demographic scenarios defined by the user, and presents several
62 advantages compared to descriptive methods, including (i) the statistical comparison of different
63 invasion scenarios, (ii) the integration of historical and biological knowledge and (iii) the evaluation
64 of the confidence in the scenario choice. In brief, various models supposed to explain the data are
65 compared based on the similarity between summary statistics simulated from these models and from
66 prior distributions of historical and genetic parameters on one hand and summary statistics computed
67 from the actual samples on the other hand. This tool has helped reconstructing invasion histories of
68 multiple species (Ascunce et al., 2011; Barres et al., 2012; Boissin et al., 2012; Fraimout et al., 2017;
69 Guillemaud et al., 2015; Lombaert et al., 2010a; Miller et al., 2005; Pascual et al., 2007; Rius, Turon,
70 Ordonez, & Pascual, 2012; Sherpa et al., 2019).

71   In this study, we were interested in the invasion of the pinewood nematode (PWN),
72 *Bursaphelenchus xylophilus* (Steiner & Buhrer, 1934; Nickle, 1970; Nematoda: Aphelenchoididae).
73 The PWN is a microscopic worm, responsible for the pine wilt disease (Mamiya, 1972, 1976, 1983),
74 which annually kills millions of pine trees worldwide (Mamiya, 1988; Soliman et al., 2012; Suzuki,
75 2002; Vicente, Espada, Vieira, & Mota, 2011). The PWN is native to North America, namely Canada
76 and the USA (Dropkin et al., 1981; Kiritani & Morimoto, 2004). The first invasive outbreak of the PWN
77 was observed in Japan in 1905. The species was then observed in China in 1982, in Taiwan before
78 1985 and in South Korea in 1988 (Futai, 2013; Mamiya, 1988; Moon, Cheon, & Lee, 2007). In 1999,
79 the PWN was observed in Europe, first in Portugal, and then Madeira Island and Spain (Abelleira,
80 Picoaga, Mansilla, & Aguin, 2011; Fonseca et al., 2012; Mota et al., 1999). To date, the origin of PWN
81 European populations is still unclear. Some authors have proposed an Asian origin for European
82 invasive populations of PWN (Figueiredo et al., 2013; Valadas, Barbosa, Espada, Oliveira, & Mota,
83 2012). However, descriptive genetic methods based on genetic distances and clustering analyses
84 recently failed to distinguish between an American and an Asian origin for European outbreaks
85 (Mallez et al., 2015). Analyses based on F_{ST} values (Weir & Cockerham, 1984) and mean individual
86 assignment likelihoods (Paetkau, Slade, Burden, & Estoup, 2004) suggested an American origin for all
87 Portuguese samples, while analyses based on Cavalli-Sforza and Edwards' distances (Cavalli-Sforza &
88 Edwards, 1967) and Bayesian clustering (Pritchard, Stephens, & Donnelly, 2000) suggested a Japanese
89 origin for these samples. The aim of the present study was to use ABC to clarify the invasion history
90 of the PWN by performing quantitative Bayesian statistical comparison of various competing invasion

91 scenarios, through the calculation of their posterior probabilities (Miller et al., 2005; Pascual et al.,
92 2007).
93
94
95 **MATERIAL & METHODS**
96
97 *Sampling and genotyping*
98
99 In this study, 48 site samples, representing a total of 1080 individuals were analyzed: 28 site samples
100 from the USA (554 individuals) representing the native area, and 7 site samples from Japan (210
101 individuals), 9 from Portugal/Madeira (169 individuals) and 4 from China (147 individuals)
102 representing the invaded areas. The features of the samples are listed in Table 1. All the individuals
103 were extracted from wood samples collected directly in the field, using a sieve or the Baermann
104 method (Viglierchio & Schmitt, 1983). The historical knowledge about the invasion of the PWN and
105 its spread within each invaded area (Abelleira et al., 2011; Fonseca et al., 2012; Futai, 2013; Mamiya,
106 1988; Moon et al., 2007; Mota et al., 1999) allowed us to associate a date of first observation to the
107 different invasive outbreaks (Table 1).
108
109
110 **Table 1:** Characteristics and population genetics summary statistics of the 48 site samples of the
111 pinewood nematode used in this study.

Code	Nb. Individuals	Location	First obs. (years)	Mean N_a	H_e	H_o	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

				Mean Na	He	Ho	F_{IS}
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	-	-	-

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

117

118

119   The DNA extraction of each single individual was carried out by thermal shock (Castagnone,
120 Abad, & Castagnone-Sereno, 2005) and 16 microsatellite loci were then amplified as described by
121 Mallez *et al.* (2015; 2013). The locus M26 was excluded due to a lack of amplification for one Chinese
122 site sample. The analyses presented here were thus conducted with 15 microsatellite loci.

123

124 *Descriptive genetic analyses*

125

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

135

136 *Evolutionary relationships between the different site samples*

137

138 We used descriptive genetic methods to (i) determine the level of genetic structure existing in each
139 area under study, (ii) reduce the data and select some site samples to include in the ABC analyses
140 (see below) and (iii) guide the choice of the invasive scenarios to compare in these latter analyses.

141   First, we evaluated the level of genetic differentiation between all site sample pairs with
142 Genepop (Rousset, 2008). Non independent multiple testing implied to adjust the significance level
143 with the sequential Bonferroni method (Sokal & Rohlf, 1995). F_{ST} (Weir & Cockerham, 1984),
144 corrected for null alleles (observed for some of the site samples and loci, see Mallez et al., 2015) and
145 D_{EST} of Jost (2008) between pairs of site samples were computed with FreeNA (Chapuis & Estoup,
146 2007) and SMOGD (Crawford, 2010), respectively.

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

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

165   Finally, a neighbor-joining tree based on Cavalli-Sforza & Edwards' distances (Cavalli-Sforza &
166 Edwards, 1967), corrected for null alleles with FreeNA (Chapuis & Estoup, 2007) was plotted and the
167 mean individual assignment likelihoods (denoted $L_{i \to s}$, Paetkau, Slade, Burden, & Estoup, 2004) of
168 each invasive site sample i to each putative source sample s were computed with GeneClass2 version
169 2.0 (Piry et al., 2004).

170
171 *ABC analyses*
172

173 All ABC model-choice analyses were performed with DIYABC version 2.1.0 (Cornuet et al., 2014). To
174 define the scenarios to compare, we took into account the historical knowledge (i.e., dates of first
175 observation), results from this study and previous analyses (Mallez et al., 2015). Particular site
176 samples representing genetic units inferred from Structure analyses were used for the ABC analyses.
177 As proposed by Lombaert *et al.* (2014), we performed a step-by-step ABC analysis: we first tried to
178 clarify the Asian invasion history (i.e., the oldest invasive area) and then integrated Europe (i.e., the
179 most recent invasive area) into the scenarios. Thus, we first focused on establishing the number of
180 introduction events in Japan and in China, independently (Figures S1 and S2 for a representation of
181 the competing scenarios in each analysis). Then, we studied the relationships between Japan and
182 China (Figures S3 and S4 for a representation of the competing scenarios in each analysis). Finally, we
183 integrated the Portuguese site samples into the inferred Asian invasion scenario to determine their
184 origin among the three possible sources: USA, Japan and China (see Figure 1 for a representation of
185 the competing scenarios in this analysis).
186

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

195
196
197
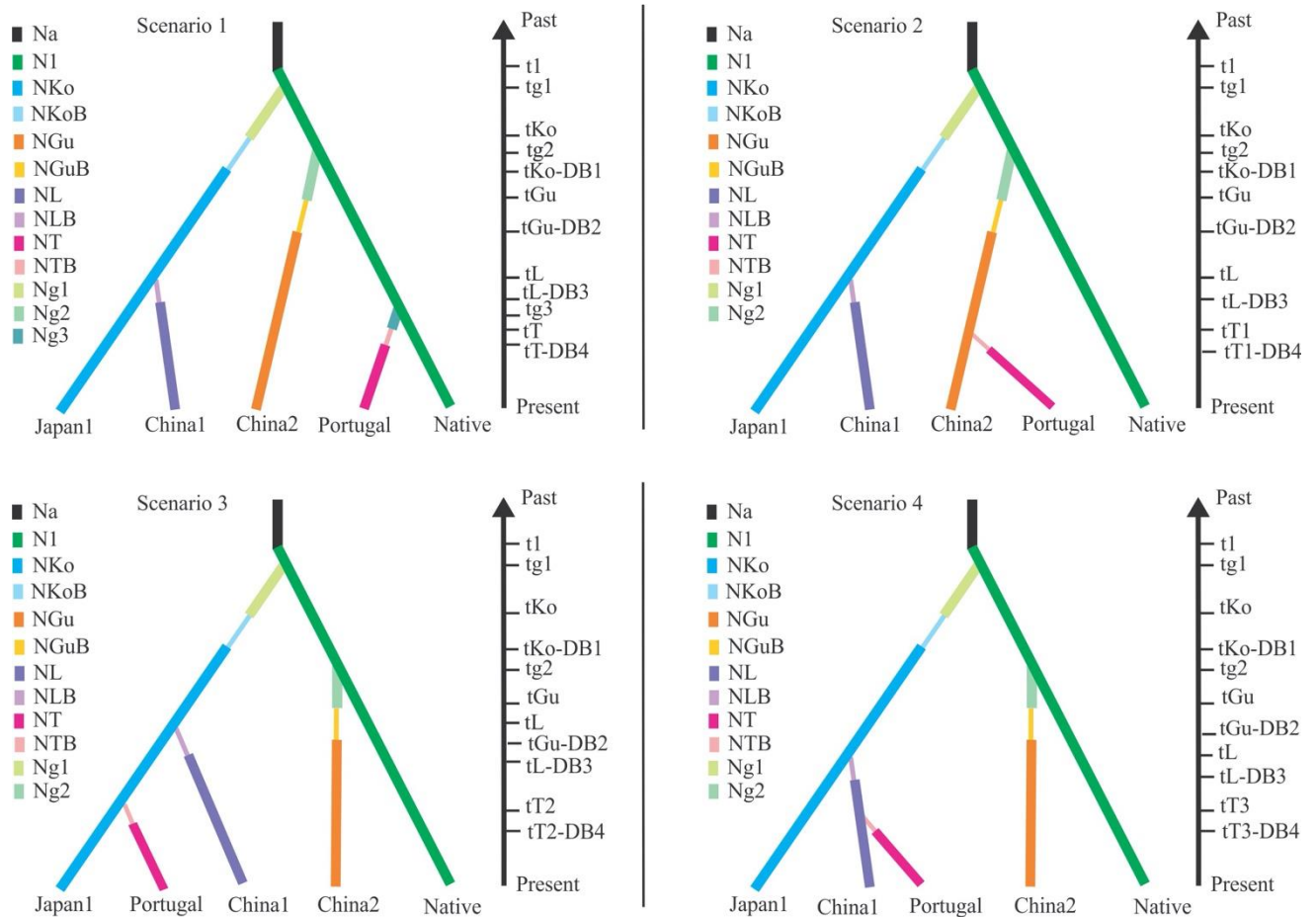
198   For each ABC analysis, we simulated one million datasets per competing scenario. To reduce
199 the datasets, the following summary statistics (SuSts) were used: the mean number of alleles, the
200 mean expected heterozygosity and the mean allelic size variance per population and pairs of
201 populations, the ratio of the number of alleles on the allelic size range (Garza & Williamson, 2001),
202 the mean individual assignment likelihoods of population *i* to population *j* (Pascual et al., 2007) and
203 the shared allele distances between populations (Chakraborty & Jin, 1993). Posterior probabilities
204 were estimated using a polychotomous logistic regression on the 1% of simulated datasets closest to
205 the observed dataset (Cornuet, Ravigne, & Estoup, 2010; J. M. Cornuet et al., 2008). The LDA option,
206 allowing to perform a discriminant analysis on the SuSts before the regression, was used (Cornuet et
207 al., 2014; Estoup et al., 2012) to gain in computation time.
208   To evaluate the robustness of our inferences, we performed the ABC analyses with two
209 different sets of prior distributions (referred to as prior set 1 and prior set 2 in the Results). The
210 generation time of the PWN *in natura* is not well-known due to a lack of direct observations of the
211 PWN inside the tree. Consequently, we tested the effect of choosing various generation times (15, 30
212 and 45 generations per year, Table S1). Thirty generations per year being the most realistic,
213 considering that the PWN reproduces from June to September and estimating an average summer
214 temperature of 25°C where it is currently distributed. We also tested the effect of choosing different

215   samples from the native area (NE5, NE9, NE22), which are the closest genetically to NE2, the native
216   sample selected based on descriptive genetic analyses (see below). Then, we verified the non-
217   overlapping of the 95 % confidence intervals of the posterior probabilities of scenarios within each
218   ABC analysis. We calculated the posterior error rate by performing ABC analyses on 1,000 simulated
219   pseudo-observed datasets. These datasets were simulated by drawing with replacement the scenario
220   identity and the parameter values among the 500 sets of scenarios and parameters values that
221   generated simulated datasets closest to the observed dataset. The posterior error rate corresponds
222   to the proportion of ABC analyses that wrongly identified scenarios and thus informs us on our ability
223   to select the true scenario. Finally, we checked the adequacy between the observed dataset and the
224   selected scenario, i.e., the capacity of this scenario to generate datasets similar to the observed one.
225   Basically, 1,000 new datasets were simulated by drawing the parameter values from the posterior
226   distributions of the selected scenario. Then, each observed SuSt was confronted to the distribution
227   formed by the SuSts obtained from the new simulated datasets, which gave a rejection probability
228   for each observed SuSt. A large number of probabilities being obtained, we adjusted the significance
229   threshold with the false discovery rate method (Benjamini & Hochberg, 1995). For this purpose, two
230   new SuSts were added to the ones used in ABC analyses: the F_{ST} (Weir & Cockerham, 1984) and the
231    $d\mu^2$ distance (Goldstein, Linares, Cavalli-sforza, & Feldman, 1995) between pairs of populations.

232   In addition, as a complementary test, we performed an alternative analysis for each of the
233   scenario comparisons described above using the recent DIYABC Random Forest v1.0 program (Collin
234   et al. 2020). This program implements in the context of ABC a supervised machine-learning algorithm
235   called the Random Forest (Breiman, 2001). Briefly, this non-parametric classification method uses
236   hundreds of bootstrapped decision trees (creating the so-called forest) to perform classification using
237   a set of predictor variables, here the summary statistics. Random Forest is able to learn from a smaller
238   reference table, and it is more robust than polychotomous logistic regression to the choice and
239   number of summary statistics (Pudlo et al. 2016). As a consequence, (i) we simulated 20,000 datasets
240   per competing scenario, and (ii) we used all summary statistics described above, to which we added
241   the F_{ST} (Weir & Cockerham, 1984) and the $d\mu^2$ distance (Goldstein, Linares, Cavalli-sforza, & Feldman,
242   1995) between pairs of populations, the maximum likelihood estimate of admixture proportion for
243   each trio of populations (Choisy et al. 2004) and the axes obtained from a linear discriminant
244   analysis on summary statistics (Estoup et al. 2012). We then grew a classification forest of 1,000
245   trees based on all simulated datasets. The random forest computation applied to the observed
246   dataset provides a classification vote representing the number of times a model is selected among
247   the 1000 trees. The scenario with the highest classification vote was selected as the most likely
248   scenario. We then estimated its posterior probability by way of a second random forest procedure of
249   1,000 trees as described by Pudlo et al. (2016). To evaluate the global performance of our ABC
250   Random Forest scenario choice, we computed the prior error rate based on the available out-of-bag
251   simulations (i.e. simulations that are not used in tree building at each bootstrap).

252

253

254

255

256   **RESULTS**

257

258   *Descriptive genetic analyses*

259

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

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

273

274   *Relationships between site samples within geographical areas*

275

276   The analysis of the American site samples led to the general observation of a strong genetic structure
277   among PWN populations in USA, in good agreement with data from a previous sampling in the same
278   area (Mallez et al., 2015). Indeed, almost all the American site samples appeared significantly
279   differentiated after correction for multiple tests (Fisher method of probability combination on loci, p
280    $< 10^{-2}$). Only one test, between NC2 and GA3, was not significant (Fisher method, $p = 0.11$). F_{ST},
281   corrected for null alleles, reached high values and ranged from 0.036 to 0.76 (Table S2) and D_{EST}
282   ranged from 0.0041 to 0.329 (Table S2). This strong genetic differentiation pattern was supported by
283   the Structure analysis, which revealed biological relevant clustering of individuals within site samples
284   for high values of K, although the ΔK inferred $K = 2$ as the main structure (Figure S5). Different
285   clustering patterns were identified with Clumpp, but they exhibited similar results, i.e., a strong
286   genetic structure (Figure S5). In Japan, all site samples were also significantly differentiated, except
287   the three undifferentiated ones mentioned above (Fisher method, $p < 10^{-5}$) with F_{ST} corrected for null
288   alleles ranging from 0.627 to 0.995 and D_{EST} ranging from 0.0042 to 0.0938 (Table S2). In China, all
289   site samples were undifferentiated except Guangdong that appeared significantly differentiated from
290   the others (Fisher method, $p < 10^{-5}$) with a high mean corrected F_{ST} equal to 0.997 and a lower mean
291   D_{EST} equal to 0.106 (Table S2). This result was confirmed by the Bayesian clustering analysis, which
292   inferred a relevant clustering of individuals for $K = 2$, Guangdong being pulled apart from the others
293   Chinese site samples (Figure S6). In Portugal/Madeira, all site samples were undifferentiated except
294   the site sample 128S, that was significantly different from the others (Fisher method, $p < 10^{-5}$) with a
295   mean corrected F_{ST} equal to 0.341 and a mean D_{EST} equal to 0.0005 (Table S2). In view of the presence
296   of undifferentiated site samples within Japan, China and Portugal/Madeira, only one of these site
297   samples was conserved in each invaded area for simplification. The conserved site samples were the
298   following: Jap212, Laoshan, TR1 and Mad23PC for Japan, China, Portugal and Madeira, respectively.

299

300   *Relationships between site samples among geographical areas*

301

302   All the site samples from different invaded areas appeared significantly differentiated between them
303   and from the American site samples (Fisher method, $p < 10^{-5}$). NE2 was the American site sample with
304   the lowest differentiation from site samples from invaded areas based on the D_{EST} (mean value =
305   0.034). Another American site sample, NE9, was selected based on the corrected F_{ST} (mean value =
306   0.469), NE2 being the second lowest one with this measure (mean value = 0.541). The neighbor-
307   joining tree confirmed that the site sample NE2 is the closest site sample to all site samples from the
308   invaded areas and that the Portuguese site samples are closer to a Japanese site sample than a
309   Chinese or an American site samples (Figure 2).

310

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.
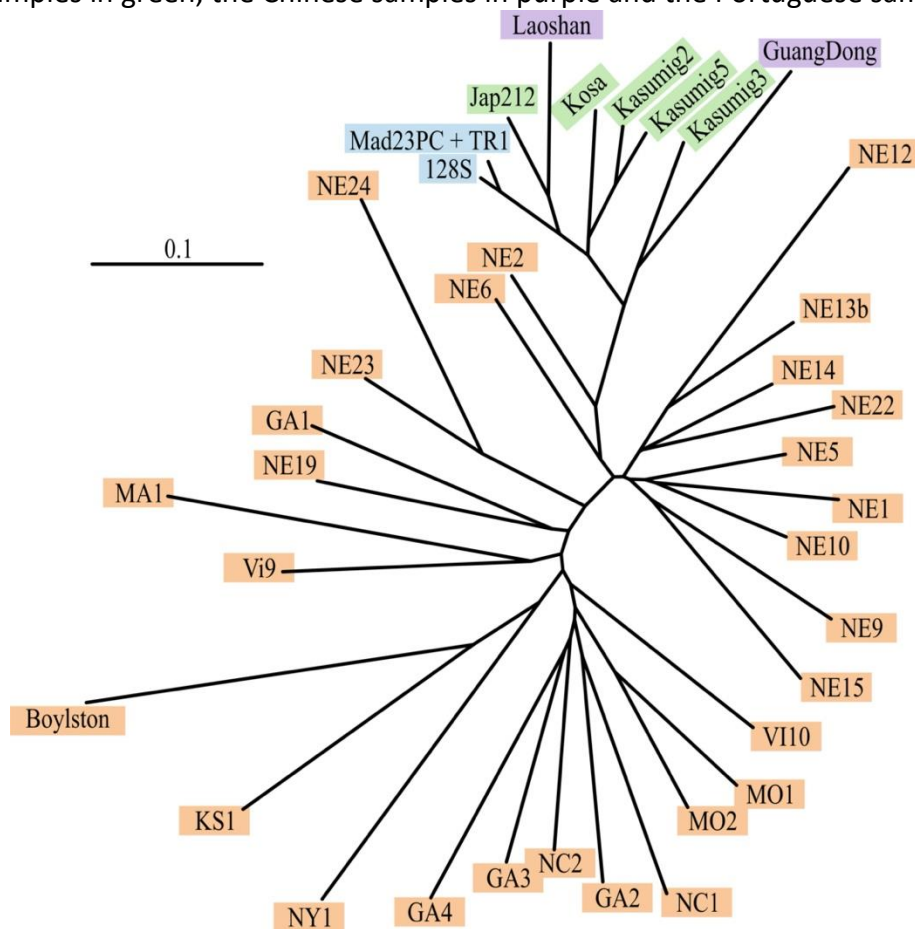
For the Bayesian clustering analysis, the ΔK inferred K=2 as the main structure (Figure S7A), in which all the invasive site samples pulled apart from the native American site samples (Figure 3). From K=3, different clustering patterns were observed (Figures 3 and S7B), in congruence with the sub-structure 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).

329 **Figure 3: Genetic structure of the pinewood nematode samples used in this study.** Bar plots of the
330 coefficients of co-ancestry obtained in STRUCTURE analyses with several values of *K*. Each bar
331 corresponds to one individual nematode and each cluster is represented with a particular color. The
332 major clustering patterns are shown (see Figure S7B for the minor ones, when applicable).
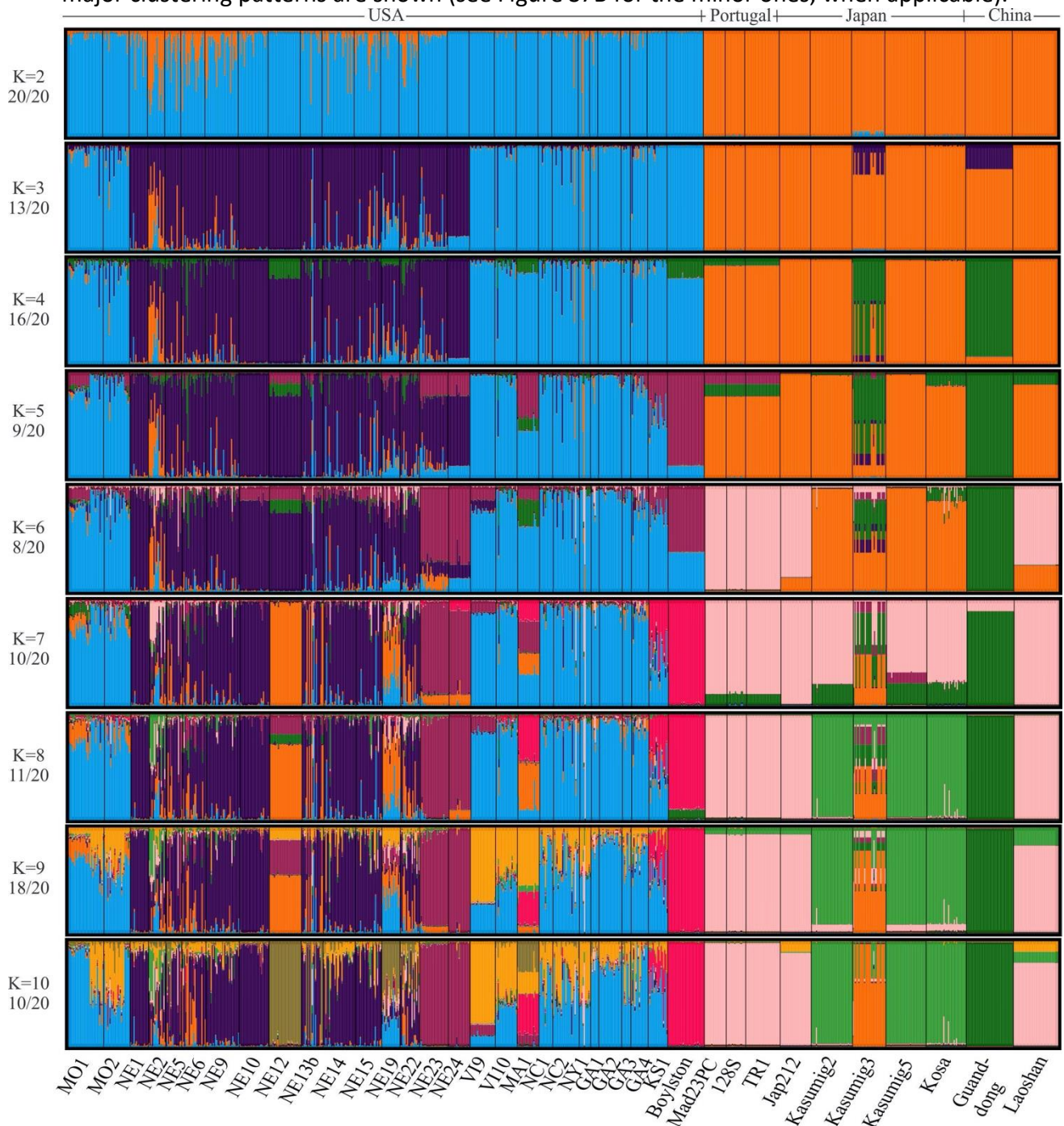
333
334
335 Summarizing and considering the results presented above, the site sample NE2 was the closest
336 site sample of the native area to all invasive site samples but Guangdong. Thus, it was used to
337 represent the native area in the competing scenarios of the ABC analyses. In view of the incomplete
338 sampling of the native area (no sample from Canada was obtained) and the strong genetic
339 structuration observed in the USA, it is unlikely that NE2 is a sample of the true source population.
340 An unsampled "ghost" population diverging from NE2 was thus included in all scenarios (see Figures
341 1, S1, S2, S3, S4 and Table S1). We also considered the following relevant genetic units to characterize
342 the invaded areas in the following ABC analyses:
343 - Japanese genetic unit 1, hereafter called Japan1, represented by Kosa, the closest sample to the
344 introduction point in Japan and considered coming from the native area (Japan being the first
345 known outbreak area of the PWN);

346 - Japanese genetic unit 2, hereafter called Japan2, represented by Kasumig3;

347 - Chinese genetic unit 1, hereafter called China1, represented by Laoshan;

348 - Chinese genetic unit 2, hereafter called China2, represented by Guangdong;

349 - Portuguese genetic unit, hereafter called Portugal, represented by TR1.

350

351 *ABC analyses*

352

353 In Japan, the results clearly indicated that a multiple introduction scenario is more likely than a single
354 introduction scenario with posterior probabilities over 0.95 for prior sets 1 and 2 and non-overlapping
355 confidence intervals, Japan2 being thus independent from Japan1 (Table 2). A similar result was
356 observed in China: China1 was independent from China2 with posterior probabilities of 0.94 and 0.70
357 for prior sets 1 and 2 respectively and non-overlapping confidence intervals, suggesting two
358 introduction events in China (Table 2). The ABC analyses carried out between site samples from Japan
359 and China revealed that: (i) Japan1 was probably the source of China1 and (ii) China2 is independent
360 from Japan1 (high posterior probabilities, results robust to prior set changes, Table 2). The analysis
361 between China2 and Japan2 gave inconsistent results, relative to the prior set used (Table 2). With
362 the prior set 1, the scenario describing an admixture event was selected but the posterior probability
363 was low, and a high posterior error rate was observed (Table 2). On the contrary, with the prior set
364 2, the scenario describing two independent events of introduction was selected with once again a
365 low posterior probability and a high posterior error rate value (Table 2). Consequently, the origin of
366 Japan2 remains unclear.

367

368

369 **Table 2:** Results of the ABC analyses carried out to infer the history of the pinewood
370 nematode invasion in Asia. Posterior probabilities and their 95% confidence intervals
371 (CI) were obtained using a polychotomous logistic regression (J. M. Cornuet et al.,
372 2010; J. M. Cornuet et al., 2008) and the LDA option (J. M. Cornuet et al., 2014;
373 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]	
Native → China2 → Japan2	0.104 [0.099 ; 0.110]	0.513
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.

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 also 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 once more 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).

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]**	
Native → Japan1 → China1 ; Native → China2 → Portugal	0.013 [0.000 ; 0.040]	0.231
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]**	
Native → Japan1 → China1 ; Native → China2 → Portugal	0.0210 [0.000 ; 0.067]	0.235
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]**	
Native → Japan1 → China1 ; Native → China2 → Portugal	0.039 [0.000 ; 0.118]	0.158
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]**	
Native → Japan1 → China1 ; Native → China2 → Portugal	0.077 [0.023 ; 0.130]	0.134
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]	

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

411
412

413   The posterior error rates appeared quite low for all the analyses (except the Japan2 – China2
414 analysis, Tables 2 and 3), demonstrating that the competing scenarios were actually distinguishable.
415 In addition, simulations from the selected scenarios and the posterior distributions of the parameters
416 were able to produce data close to the observed ones: none of the observed SuSts was in the extreme
417 5% tails of the distribution of simulated SuSts after correction (Table S6). Finally, the alternative
418 analyses performed using the ABC Random Forest algorithm confirmed all previous results involving
419 the main native sample NE2 (Table S7).

420
421 **DISCUSSION**
422
423 The main objective of this study was to infer the invasion routes of the PWN using the ABC method.
424 We demonstrated the existence of multiple independent introductions from the native area to Japan,
425 China and Europe. This inference has not been possible using descriptive population genetics
426 methods performed in a previous study (Mallez et al., 2015). Conversely, the ABC method helped
427 clarify the worldwide history of the PWN invasion even though the post-hoc analyses tempered the
428 conclusions, highlighting the importance of assessing the confidence in the results. The discrepancies
429 observed between descriptive genetic methods and the ABC method are worth investigating and are
430 discussed here.

431
432 *Invasion routes of the PWN and the origin of Portuguese populations*
433
434 Despite the low genetic diversity generally present in the PWN populations, we were able to reliably
435 select scenarios (among the ones compared) with high posterior probabilities and good quality
436 analyses in most cases. The most probable worldwide invasion history of the PWN identified in this
437 study (Figure 4) includes multiple introductions in Asia and an American origin of the European
438 samples.
439

440   **Figure 4: Worldwide invasion routes of the pinewood nematode as inferred from this study.** The
441   year of first infestation for every invaded area is indicated in parenthesis.
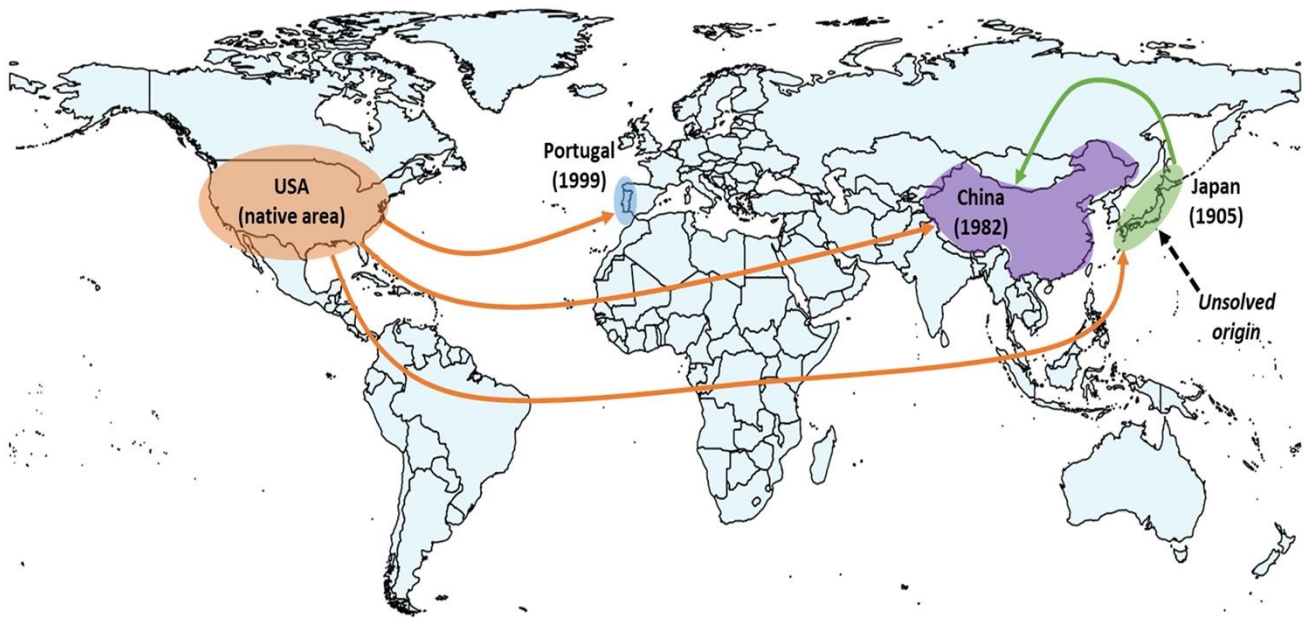
442
443
444   The analyses allowed us to reveal the existence of multiple independent events of
445   introduction in China and Japan. Two events of introduction were inferred in both China (one from
446   the USA and one from Japan) and Japan (one from the USA and one with an unknown origin).
447   Especially, these results strengthen the hypothesis of two events of introduction in China, the first
448   one near Nanjing in 1982 and the second one near Hong-Kong in 1988, already suggested by modeling
449   approaches in Robinet et al (2009; but see Cheng, Cheng, Xu, & Xie, 2008). Concerning Europe, the
450   main ABC analyses of the PWN microsatellites data designated the USA as the most probable source
451   of the Portuguese populations and excluded Japan and China as possible sources, contrary to what is
452   often presented in the literature based on genetic distance trees (Figueiredo et al., 2013; Fonseca et
453   al., 2012; Metge & Burgermeister, 2008; Valadas et al., 2012; Valadas, Laranjo, Mota, & Oliveira, 2013;
454   Vieira, Burgermeister, Mota, Metge, & Silva, 2007). This result was quite robust to changes in the
455   parameters' priors and to the posterior model checking analyses, but it depended on the samples
456   considered in the native area, which may be considered a problem. Interestingly, our results displayed
457   a pattern similar to the one obtained by Guillemaud *et al.* (2010). These authors found a negative
458   correlation between the genetic distance of the samples considered to the most probable source and
459   the probability of an independent scenario when this scenario was true. In other words, when two
460   invasive populations originate from two independent introductions from a native population, the
461   more distant is the native sample from the true source, the more probable it is to erroneously point
462   to a scenario with two serial introductions. Therefore the "mixed" results obtained by changing the
463   sample of the native area are expected if the true scenario includes an American origin of the
464   European samples. Overall, the ABC method allowed us to clarify the history of the PWN invasion in
465   Europe with good statistical confidence.
466   This study highlights the central role of the native area in the worldwide history of invasion of
467   the PWN, adding up three as the minimum number of independent introductions from the USA. Only
468   one detectable event of introduction seems to have occurred in Portugal. Considering the volume of
469   global trade between Portugal and the worldwide infested areas, this result shows that quarantine
470   and inspection efforts have been globally successful and efficient to date to minimize the number of
471   independent outbreaks and confirms the importance of prevention and pro-active actions to limit
472   new introductions (Simberloff & Rejmanek, 2011). At the European level, ecological and pedo-climatic

473 conditions favorable to the PWN are encountered in many countries (Robinet, Van Opstal, Baker, &
474 Roques, 2011). However, the invasion is currently limited to the Iberian Peninsula (plus Madeira
475 Island). This reinforces the idea that the quarantine regulation measures have been efficient to date
476 to protect Europe from the entry of new invasive PWN populations from abroad, and from a large
477 dissemination of the PWN from the original Portuguese outbreak.
478
479 *Different methods with different results*
480
481 In the case of the PWN invasion, the results obtained from descriptive genetic approaches pointed
482 out different putative origins for the European populations of PWN (Mallez et al., 2015): American
483 based on F_{ST} values (Weir & Cockerham, 1984) and mean individual assignment likelihoods (Paetkau
484 et al., 2004), and Japanese based on Cavalli-Sforza and Edwards' distances (Cavalli-Sforza & Edwards,
485 1967) and Bayesian clustering (Pritchard et al., 2000). The ABC analyses, quantitative and probability-
486 based, allowed us to improve greatly the history of the PWN invasion. Our study highlights
487 discrepancies between the ABC methodology and some of the descriptive approaches, namely the
488 Bayesian clustering of individuals (Pritchard et al., 2000) and distance genetic trees (Saitou & Nei,
489 1987). To our knowledge, such discrepancies have not been observed yet, although a simulation study
490 showed that classic clustering methods such as Structure analyses (Pritchard et al., 2000) may be
491 erroneously interpreted in the context of invasion route inference (Lombaert, Guillemaud, & Deleury,
492 2018). The origin of the discrepancies we observed is not completely clear to us, but hypotheses may
493 be proposed.
494   They may have arisen from the very low level of genetic diversity observed in the PWN
495 populations, which might disturb the analyses of the descriptive genetic methods and thus their
496 outcomes. In an invasion context, a low genetic diversity is usually the translation of strong founder
497 effects, which implies a strong genetic drift. Genetic drift, in turn, is a stochastic process whose
498 consequences are random. Employing methods such as ABC that take these stochastic effects into
499 account may thus be more reliable than other methods that do not, such as Bayesian clustering or
500 distance-based methods (Guillemaud et al., 2010). In these conditions of low genetic diversity, we
501 can expect descriptive methods to lead to variable results when applied to various realizations of the
502 same historical scenario. We may expect the low level of genetic diversity to produce a genetic
503 structure pattern and/or relationship between site samples that we misinterpret (as in Lombaert et
504 al., 2018). For instance, if the native area is weakly diversified so that it exhibits a few very frequent
505 alleles, it is probable that two independent introductions from this native area (native → invasive 1
506 and native → invasive 2) lead to samples closer to each other than to their native area. When
507 analyzed, such results would suggest serial introductions (native → invasive 1 → invasive 2, because
508 invasive1 and 2 are closer to each other than to native) instead of the true independent introduction
509 scenario, because the interpretation is binary and not based on probability computation. Using a
510 model-based stochastic method, such as ABC, a genetic structure with invading samples closer to
511 each other than to the native area may lead to the choice of independent introduction scenario
512 because it is probable to get such structure with this scenario when the diversity is low.
513   The observed discrepancies may thus just be the consequence of the absence of statistical
514 support associated with descriptive genetic methods permitting to quantitatively base the choice of
515 evolutionary scenarios (criticism presented in Estoup & Guillemaud, 2010).
516   Carrying out simulations may help to disentangle the effect of the genetic diversity in the
517 native area on the outcomes of clustering methods and ABC analyses, a work partly performed by
518 Lombaert *et al.* (2018) who did not use ABC analyses. More precisely, analyzing datasets, simulated
519 under chosen scenarios and known levels of genetic diversity, may allow to assess the behavior of
520 these analyses in relation to the level of genetic diversity. These simulations would thus be useful to
521 (i) determine the real impact of the genetic diversity on the outcomes and (ii) verify that the
522 interpretations made from these methods in the biological invasion framework are appropriate.
523

530

531 **DATA ACCESSIBILITY STATEMENT**
532 Complete dataset (microsatellite), STRUCTURE outputs and ABC configuration files were deposited at
533 Zenodo: 10.5281/zenodo.4019342

534

535 **AUTHOR CONTRIBUTIONS**
536 SM and CC performed microsatellite genotyping. SM performed the analyses, with input from EL. SM,
537 EL, PCS and TG interpreted the results. SM wrote the manuscript, with guidance from PCS and TG,
538 and input from EL. All authors have read and approved the final manuscript.

539

540 **CONFLICT OF INTEREST DISCLOSURE**
541 The authors of this article declare that they have no financial conflict of interest with the content of
542 this article. TG is co-founder of PCI and TG and EL are recommenders for PCI Evol Biol, PCI Ecol and
543 PCI Zool.

544

545

546 **REFERENCES**
547 Abelleira, A., Picoaga, A., Mansilla, J. P., & Aguin, O. (2011). Detection of *Bursaphelenchus xylophilus*,
548   causal agent of pine wilt disease on *Pinus pinaster* in Northwestern Spain. *Plant Disease, 95*(6),
549   776.
550 Ascunce, M. S., Yang, C. C., Oakey, J., Calcaterra, L., Wu, W. J., Shih, C. J., Shoemaker, D. (2011). Global
551   Invasion History of the Fire Ant Solenopsis invicta. *Science, 331*(6020), 1066-1068.
552   doi:10.1126/science.1198734
553 Barres, B., Carlier, J., Seguin, M., Fenouillet, C., Cilas, C., & Ravigne, V. (2012). Understanding the
554   recent colonization history of a plant pathogenic fungus using population genetic tools and
555   Approximate Bayesian Computation. *Heredity, 109*(5), 269-279. doi:10.1038/hdy.2012.37
556 Beaumont, M. A., Zhang, W. Y., & Balding, D. J. (2002). Approximate Bayesian computation in
557   population genetics. *Genetics, 162*(4), 2025-2035.
558 Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., & Bonhomme, F. (1996-2004). GENETIX 4.05, logiciel
559   sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations,
560   Interactions, CNRS UMR 5000, Université de Montpellier Montpellier (France).
561 Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful
562   approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological,*
563   *57*(1), 289-300.
564 Boissin, E., Hurley, B., Wingfield, M. J., Vasaitis, R., Stenlid, J., Davis, C., . . . Slippers, B. (2012).
565   Retracing the routes of introduction of invasive species: the case of the Sirex noctilio
566   woodwasp. *Molecular Ecology, 21*(23), 5728-5744. doi:10.1111/mec.12065
567 Boucher, A. C., Mimee, B., Montarry, J., Bardou-Valette, S., Belair, G., Moffett, P., & Grenier, E. (2013).
568   Genetic diversity of the golden potato cyst nematode Globodera rostochiensis and
569   determination of the origin of populations in Quebec, Canada. *Molecular Phylogenetics and*
570   *Evolution, 69*(1), 75-82. doi:10.1016/j.ympev.2013.05.020
571 Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. doi:10.1023/A:1010933404324
572 Castagnone, C., Abad, P., & Castagnone-Sereno, P. (2005). Satellite DNA-based species-specific
573   identification of single individuals of the pinewood nematode *Bursaphelenchus xylophilus*

574   (Nematoda : Aphelenchoididae). *European Journal of Plant Pathology, 112*(2), 191-193.
575   doi:10.1007/s10658-004-0580-2

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

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

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

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

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

588 Ciosi, M., Miller, N. J., Kim, K. S., Giordano, R., Estoup, A., & Guillemaud, T. (2008). Invasion of Europe
589   by the western corn rootworm, *Diabrotica virgifera virgifera*: multiple transatlantic
590   introductions with various reductions of genetic diversity. *Molecular Ecology, 17*(16), 3614-
591   3627.

592 Collin, F.-D., Durif, G., Raynal, L., Lombaert, E., Gautier, M., Vitalis, R., … Estoup, A. (2020) Extending
593   Approximate Bayesian Computation with Supervised Machine Learning to infer demographic
594   history from genetic polymorphisms using DIYABC Random Forest. Manuscript submitted for
595   publication.

596 Cornuet, J.-M., Santos, F., Beaumont, M. A., Robert, C. P., Marin, J.-M., Balding, D. J., . . . Estoup, A.
597   (2008). Inferring population history with DIY ABC: a user-friendly approach to approximate
598   Bayesian computation. *Bioinformatics, 24*(23), 2713-2719.
599   doi:10.1093/bioinformatics/btn514

600 Cornuet, J. M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., . . . Estoup, A. (2014).
601   DIYABC v2.0: a software to make approximate Bayesian computation inferences about
602   population history using single nucleotide polymorphism, DNA sequence and microsatellite
603   data. *Bioinformatics, 30*(8), 1187-1189. doi:10.1093/bioinformatics/btt763

604 Cornuet, J. M., Ravigne, V., & Estoup, A. (2010). Inference on population history and model checking
605   using DNA sequence and microsatellite data with the software DIYABC (v1.0). *Bmc
606   Bioinformatics, 11*. doi:40110.1186/1471-2105-11-401

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

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

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

613 Estoup, A., Lombaert, E., Marin, J.-M., Guillemaud, T., Pudlo, P., Robert, C. P., & Cornuet, J.-M. (2012).
614   Estimation of demo-genetic model probabilities with Approximate Bayesian Computation
615   using linear discriminant analysis on summary statistics. *Molecular Ecology Resources, 12*(5),
616   846-855. doi:10.1111/j.1755-0998.2012.03153.x

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

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

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

625   Fonseca, L., Cardoso, J. M. S., Lopes, A., Pestana, M., Abreu, F., Nunes, N., . . . Abrantes, I. (2012). The
626   pinewood nematode, *Bursaphelenchus xylophilus*, in Madeira Island. *Helminthologia, 49*(2),
627   96 - 103.
628   Fontaine, M. C., Gladieux, P., Hood, M. E., & Giraud, T. (2013). History of the invasion of the anther
629   smut pathogen on *Silene latifolia* in North America. *New Phytologist, 198*(3), 946-956.
630   doi:10.1111/nph.12177
631   Fraimout, A., Debat, V., Fellous, S., Hufbauer, R. A., Foucaud, J., Pudlo, P., … Estoup, A. (2017).
632   Deciphering the Routes of invasion of *Drosophila suzukii* by Means of ABC Random Forest.
633   *Molecular Biology and Evolution*, *34*(4), 980–996. doi:10.1093/molbev/msx050
634   Futai, K. (2013). Pinewood nematode, *Bursaphelenchus xylophilus*. *Annual Review of Phytopathology,*
635   *51*, 61-83.
636   Garza, J. C., & Williamson, E. G. (2001). Detection of reduction in population size using data from
637   microsatellite loci. *Molecular Ecology, 10*(2), 305-318. doi:10.1046/j.1365-294X.2001.01190.x
638   Goldstein, D. B., Linares, A. R., Cavalli-sforza, L. L., & Feldman, M. W. (1995). An evaluation of genetic
639   distances for use with microsatellite loci. *Genetics, 139*(1), 463-471.
640   Goudet, J. (2002). Fstat, a program to estimate and test gene diversities and fixation indices. Retrieved
641   from http://www2.unil.ch/popgen/softwares/fstat.htm
642   Guillemaud, T., Beaumont, M. A., Ciosi, M., Cornuet, J.-M., & Estoup, A. (2010). Inferring introduction
643   routes of invasive species using approximate Bayesian computation on microsatellite data.
644   *Heredity, 104*(1), 88 - 99. doi:doi:10.1038/hdy.2009.92
645   Guillemaud, T., Blin, A., Le Goff, I., Desneux, N., Reyes, M., Tabone, E., . . . Lombaert, E. (2015). The
646   tomato borer, Tuta absoluta, invading the Mediterranean Basin, originates from a single
647   introduction from Central Chile. *Scientific Reports, 5*, 8371. doi:10.1038/srep08371
648   Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for
649   dealing with label switching and multimodality in analysis of population structure.
650   *Bioinformatics, 23*(14), 1801 - 1806.
651   Jost, L. (2008). G(ST) and its relatives do not measure differentiation. *Molecular Ecology, 17*(18), 4015-
652   4026. doi:10.1111/j.1365-294X.2008.03887.x
653   Kelager, A., Pedersen, J., & Bruun, H. (2013). Multiple introductions and no loss of genetic diversity:
654   invasion history of Japanese Rose, Rosa rugosa, in Europe. *Biological Invasions, 15*(5), 1125-
655   1141. doi:10.1007/s10530-012-0356-0
656   Kiritani, K., & Morimoto, N. (2004). Invasive insect and nematode pests from North America *Global*
657   *Environmental Research, 8*(1), 75 - 88.
658   Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). - Clumpak: a
659   program for identifying clustering modes and packaging population structure inferences
660   across K. - *15*(- 5), - 1191.
661   Lombaert, E., Guillemaud, T., Cornuet, J. M., Malausa, T., Facon, B., & Estoup, A. (2010a). Bridgehead
662   Effect in the Worldwide Invasion of the Biocontrol Harlequin Ladybird. *Plos One, 5*(3).
663   doi:e974310.1371/journal.pone.0009743
664   Lombaert, E., Guillemaud, T., Cornuet, J. M., Malausa, T., Facon, B., & Estoup, A. (2010b). Bridgehead
665   Effect in the Worldwide Invasion of the Biocontrol Harlequin Ladybird. *Plos One, 5*(3), e9743.
666   doi:10.1371/journal.pone.0009743
667   Lombaert, E., Guillemaud, T., & Deleury, E. (2018). Biases of STRUCTURE software when exploring
668   introduction routes of invasive species. *Heredity, 120*(6), 485 - 499. doi:doi:10.1038/s41437-
669   017-0042-1
670   Lombaert, E., Guillemaud, T., Lundgren, J., Koch, R., Facon, B., Grez, A., Estoup, A. (2014).
671   Complementarity of statistical treatments to reconstruct worldwide routes of invasion: the
672   case of the Asian ladybird *Harmonia axyridis*. *Molecular Ecology, 23*, 5979 - 5997.
673   Mack, R. N., Simberloff, D., Mark Lonsdale, W., Evans, H., Clout, M., & Bazzaz, F. A. (2000). Biotic
674   invasions : Causes, Epidemiology, Global Consequences, and control. *Ecological Applications,*
675   *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.

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. doi:10.1093/bioinformatics/btv684

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

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. doi:10.1111/mec.15071

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.

776 Viglierchio, D. R., & Schmitt, R. V. (1983). On the methodology of nematode extraction from field
777   samples : Baermann funnel modifications. *Journal of Nematology, 15*(3), 438 - 444.
778 Vitousek, P. M., Dantonio, C. M., Loope, L. L., & Westbrooks, R. (1996). Biological invasions as global
779   environmental change. *American Scientist, 84*(5), 468-478.
780 Walker, B. H., & Steffen, W. (1997). An overview of the implications of global change for natural and
781   managed terrestrial ecosystems. *Conservation Ecology [online]*. Retrieved from
782   http://www.consecol.org/vol1/iss2/art2/
783 Wan, X. W., Liu, Y. H., & Zhang, B. (2012). Invasion History of the Oriental Fruit Fly, Bactrocera dorsalis,
784   in the Pacific-Asia Region: Two Main Invasion Routes. *Plos One, 7*(5).
785   doi:10.1371/journal.pone.0036176
786 Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the analysis of population structure.
787   *Evolution, 38*(6), 1358-1370.
788 Wilcove, D. S., Rothstein, D., Dubow, J., Phillips, A., & Losos, E. (1998). Quantifying threats to imperiled
789   species in the United States. *BioScience, 48*(8), 607-615. doi:10.2307/1313420