

## ORIGINAL ARTICLE

# Contrasted invasion processes imprint the genetic structure of an invasive scale insect across southern Europe

C Kerdelhué<sup>1</sup>, T Boivin<sup>2</sup> and C Burban<sup>3</sup>

Deciphering the colonization processes by which introduced pests invade new areas is essential to limit the risk of further expansion and/or multiple introductions. We here studied the invasion history of the maritime pine bast scale *Matsucoccus feytaudi*. This host-specific insect does not cause any damage in its native area, but it devastated maritime pine forests of South-Eastern France where it was detected in the 1960s, and since then reached Italy and Corsica. We used population genetic approaches to infer the populations' recent evolutionary history from microsatellite markers and Approximate Bayesian Computation. Consistent with previous mitochondrial data, we showed that the native range is geographically strongly structured, which is probably due to the patchy distribution of the obligate host and the limited dispersal capacity of the scale. Our results show that the invasion history can be described in three successive steps involving different colonization and dispersal processes. During the mid-XXth century, massive introductions occurred from the Landes planted forest to South-Eastern France, probably due to transportation of infested wood material after World War II. Stepping-stone expansion, consistent with natural dispersal, then allowed *M. feytaudi* to reach the maritime pine forests of Liguria and Tuscany in Italy. The island of Corsica was accidentally colonized in the 1990s, and the most plausible scenario involves the introduction of a limited number of migrants from the forests of South-Eastern France and Liguria, which is consistent with an aerial dispersal due to the dominant winds that blow in spring in this region.

*Heredity* (2014) **113**, 390–400; doi:10.1038/hdy.2014.39; published online 21 May 2014

## INTRODUCTION

Species expansions or introductions into new environments can be either due to active dispersal (for example, active flight) or due to passive movements of individuals. Passive mechanisms can involve natural currents, such as winds or water flows that can sometimes carry individuals over long distances, or human activities, when the initial migrants are for instance introduced with transported goods (Kaňuch *et al.*, 2013). Deciphering between natural and human-aided long-distance dispersal events may be difficult, even though anthropogenic dispersal usually occurs from (and to) urbanised areas or regions with active human transport networks such as large harbours, motorways and railways (Robinet *et al.*, 2009; Carrasco *et al.*, 2010; Kaňuch *et al.*, 2013). Invasive species may colonize distant patches and new environments by long-distance dispersal and/or expand into adjacent habitats by regional diffusion. Complex dispersal patterns combining short-distance diffusion with long-distance dispersal are referred to as stratified dispersal, which can lead to greater rates of expansion than that observed in cases of species without long-distance dispersal (Shigesada *et al.*, 1995; Ciosi *et al.*, 2011). Short- and long-range dispersal may be facilitated by biotic (Liebhold and Tobin, 2008) or abiotic dispersal vectors (for example, wind: Ahmed *et al.*, 2009; Reynolds and Reynolds, 2009). Identifying the source populations can bring information about dispersal capacities of the studied species. It can also help designing management efforts to limit further

expansion risks; it allows to understand the evolutionary patterns of introduced populations by comparing ecological characteristics and life history traits, thereby facilitating the prediction of further suitable areas. In some cases, it can help choosing strains of potential auxiliary agents to develop biological control strategies (Estoup and Guillemaud, 2010). Elucidating complex introduction scenarios for species that were introduced in several regions can moreover permit to identify potential 'bridgeheads' (Lombaert *et al.*, 2010), that is invasive regions from which further colonizations are easier, either because of human activities (movements of people or goods for instance) or because of the proximity of other suitable regions.

When the invasion history is complex, it may be extremely difficult to disentangle. In the last decade, molecular markers have been widely used to link the invasive populations to potential sources based on genetic distances or assignment methods (for example, Ciosi *et al.*, 2008; Estoup and Guillemaud, 2010; Kaňuch *et al.*, 2013). The main drawback of the classical analyses performed with molecular data is that they do not take into account the demographic history and the stochasticity involved in introduction scenarios (bottleneck effects, drift, population expansions, admixture from several sources) and do not allow to formally test competing scenarios (Guillemaud *et al.*, 2010). On the contrary, approximate Bayesian computation (Beaumont *et al.*, 2002), which carries out model-based inferences using coalescent theory, allows to avoid these drawbacks and can limit

<sup>1</sup>INRA, UMR CBGP (INRA/IRD/Cirad/Montpellier SupAgro), Campus International de Baillarguet, Montpellier-sur-Lez cedex, France; <sup>2</sup>INRA, UR629 Ecologie des Forêts Méditerranéennes, Site Agroparc, Avignon cedex 9, France and <sup>3</sup>INRA, UMR1202 BIOGECO (INRA/Université de Bordeaux), Cestas cedex, France  
Correspondence: Dr C Kerdelhué, INRA, UMR CBGP (INRA/IRD/Cirad/Montpellier SupAgro), Campus International de Baillarguet, CS 30016, F-34988 Montpellier-sur-Lez cedex, France.

E-mail: Carole.Kerdelhue@supagro.inra.fr

Received 31 May 2013; revised 27 February 2014; accepted 13 March 2014; published online 21 May 2014

misleading biases due to incomplete sampling by including 'ghost' (that is, unsampled) populations (Estoup and Guillemaud, 2010; Guillemaud *et al.*, 2010). This method is particularly adapted to decipher complex introduction scenarios using information from molecular markers, even though it should be used with caution and properly validated (Bertorelle *et al.*, 2010; Robert *et al.*, 2011).

*Matsucoccus feytaudi* Ducasse (Hemiptera: Coccoidea: Margarodidae) is a scale insect strictly associated with the maritime pine *Pinus pinaster* Ait. on which it depends for reproduction and development. The natural range of the so-called maritime pine bast scale is fragmented over Western Europe and Morocco. It is native from Morocco, the Iberian Peninsula and South-Western France, and it colonized the maritime pine stands of South-Eastern France in the mid-XXth century. It later was discovered in Italy (Liguria in the late 1970s and Tuscany in 1999) and in the island of Corsica in 1994 (Fabre, 1980; Jactel *et al.*, 1998; Binazzi, 2005). In its native range, *M. feytaudi* is associated with the Western and Moroccan lineages of its host (Burban and Petit, 2003), which are naturally resistant to the scale, and do not develop any symptoms of decay upon attack by the scale (Harfouche *et al.*, 1995). On the contrary, when it was detected in South-Eastern France, it was already causing heavy damage in pine forests since the late 50s, due to the susceptibility of the local trees that belong to the Eastern lineage of *P. pinaster* (Harfouche *et al.*, 1995; Burbán and Petit, 2003).

Natural dispersal in *M. feytaudi* is mainly due to active male flight and passive transport of both males and first-instar larvae (crawlers), which can be carried by the wind like other wingless small arthropods that make up the 'aerial bioflow' (Reynolds and Reynolds, 2009). Adult females are sessile and do not disperse. Human-aided movements over short or long distances can also occur, due to transportation of infested wood. Short-distance natural gene flow, due to both active male flights and passive wind-assisted migration of males and larvae within continuous (or closely located) maritime pine stands, will result in slow range expansions between contiguous host patches. Genetic diversity can then be maintained along the colonization route, except if a founder effect occurs when a new patch is invaded through stepping-stone dispersal. Long-distance dispersal events, due to human activity or rare events of insect transport by dominant winds over hundreds of kilometres, allow chance colonization of remote hosts. The genetic signatures of such events will mostly depend on the effective number of founders, while anthropogenic vs natural dispersal is likely to result in the colonization of contrasting habitats (active routes of wood exchange vs weakly urbanised areas). These features make *M. feytaudi* a good study system to analyse contrasted colonization scenarios and their imprint on population genetic structure, as its invasion history is likely to include both natural and anthropogenic dispersal, as well as short and long-distance movements of founders.

A first genetic study using mitochondrial markers (Burban *et al.*, 1999) previously identified a strong phylogeographic pattern for the scale, with three allopatric maternal lineages occurring, respectively, in Morocco, Andalusia and Western Europe. Notably, all populations in the colonized range exhibited a same single haplotype that also occurred in most regions of the native Western European lineage, from Portugal to South-Western France. As a consequence, this marker did not allow to point to the precise origins of the South-Eastern French, Italian and Corsican outbreaks, nor to infer dispersal processes. In this study, we took advantage of the development of microsatellite markers for *M. feytaudi* (Kerdelhué and Decroocq, 2006) to explore its nuclear genetic diversity and structure, to identify the origin(s) of the invasive populations and to infer the most likely

dispersal modes acting along the main colonization pathways. We address these issues using a sampling design including the invaded range and the main native areas, with a special sampling effort along the Atlantic coast because mitochondrial data excluded Southern Spain and Morocco as possible sources of introduction. Both classical data analyses and approximate Bayesian computations were conducted to analyse the genetic differentiation of populations, their genetic origin and their historical demographic features.

## MATERIALS AND METHODS

### Sampling and DNA extraction

Males of the maritime pine bast scale were sampled from 18 localities using traps baited with lures loaded with 50 µg of synthetic pheromone (Jactel *et al.*, 1994). Sampling was conducted from 2004 to 2008, except for one population (1995). Localities were chosen in the native range (Morocco, Spain, Portugal and South-Western France), in the continental invasive range (South-Eastern France as well as Liguria and Tuscany in Italy) and in the invaded island of Corsica (Table 1, Figure 1). For each locality, three traps were placed in maritime pine stands from February to May, the lure being renewed each month. Insects were collected twice a month and immediately stored in 95% ethanol. DNA was extracted from the whole body of each male (21–32 individuals per population), using the GenElute mammalian Genomic DNA miniprep kit (Sigma-Aldrich, Saint-Louis, MS, USA) and eluted in 200 µl of buffer.

### Microsatellite genotyping

Seven microsatellite loci were used to genotype the sampled individuals. Five of these markers, namely Mat211B, Mat234, Mat252, Mat61 and Mat212, are described in Kerdelhué and Decroocq (2006). We added two loci (Mat17 and Mat196) that were developed from the same library as the previous ones. Technical details are given in Supplementary Table S1. Fluorescent PCR products were run and detected on an ABI 3730 automatic sequencer and product sizes were determined using the GENEMAPPER 4.0 software (Applied Biosystems, Carlsbad, CA, USA).

### Data analyses

Allelic richness and frequencies, as well as observed and expected heterozygosities, were calculated for each locus using GENETIX 4.04 (Belkhir *et al.*, 1996–2004). Hardy–Weinberg equilibrium was tested using ARLEQUIN 3.11 (Excoffier *et al.*, 2005) for each locus and population, using 1000 permutation steps and 100 000 steps in the Markov chain. Linkage disequilibrium was tested in each population for all pairs of loci with 10 000 permutations using ARLEQUIN. Null allele frequencies were estimated for each locus using the expectation maximization algorithm performed in the FREEANA package (Chapuis and Estoup, 2007).

**Population genetic structure.** Population structure was first analysed through pairwise  $F_{ST}$  either estimated directly or using the excluding null alleles (ENA) correction implemented in FREEANA to correct for the positive bias induced by the presence of null alleles (Chapuis and Estoup, 2007). The 95% confidence intervals (CIs) were obtained by bootstrapping 1000 times over loci. A neighbour-joining tree of populations was reconstructed using POPULATIONS 1.2.30 (Olivier Langella, <http://bioinformatics.org/~tryphon/populations/>) using Cavalli-Sforza and Edwards chord distance on the genotype data set corrected for null alleles. Bootstrap values were computed by resampling loci and are given as a percentage of 1000 replicates.

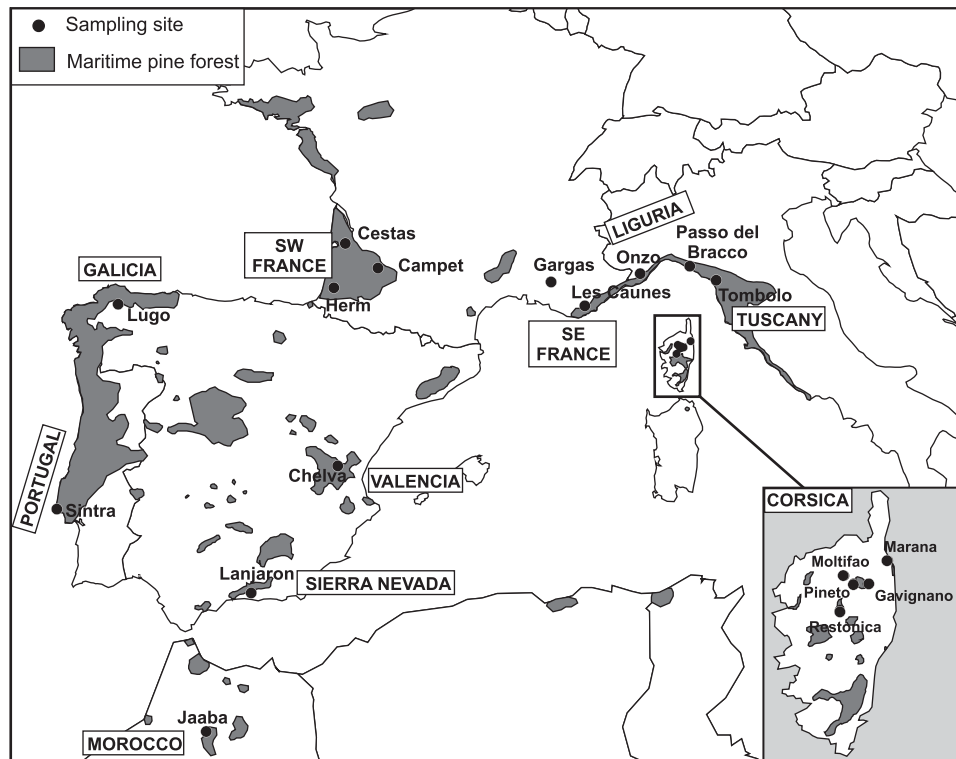
**Test of founder effects.** The program BOTTLENECK 1.2.02 (Piry *et al.*, 1999) was used to detect a potential recent bottleneck in each population. Two mutation models were applied: the strict stepwise mutation model and the two-phase model. Significant deviations in observed heterozygosity over all loci were tested using a non-parametric Wilcoxon test (one-tail test for heterozygote excess) and the mode-shift test.

**Individual assignments.** We assigned individuals to clusters based on their multilocus genotypes using a Bayesian inference method implemented in

**Table 1** Sampling localities, date of collect, number of genotyped individuals per population and indices of population genetics

Country	Region	Sampling site	Latitude	Longitude	Altitude	Date of collection	Native (N) or introduced (I)	N	He	Ho	AR
France	Corsica	Restonica	42°17'35.00"N	9°8'13.58"E	480 m	March–April 2005	I	21	0.36	0.41	2.00
France	Corsica	Gavignano	42°25'21.25"N	9°16'15.45"E	460 m	March–April 2005	I	22	0.37	0.38	2.43
France	Corsica	Moltifao	42°29'11.16"N	9°6'37.80"E	570 m	March–April 2005	I	23	0.35	0.35	2.14
France	Corsica	Marana	42°37'11.08"N	9°28'32.58"E	0 m	March–April 2005	I	24	0.25	0.26	2.00
France	Corsica	Pineto	42°25'44.28"N	9°13'6.81"E	280 m	March–April 2005	I	32	0.37	0.38	2.14
France	PACA	Les Caunes	43°11'49.59"N	6°22'44.92"E	100 m	March 2006	I	23	0.55	0.53	5.57
France	PACA	Gargas	43°54'16.82"N	5°21'26.47"E	300 m	March 1995	I	21	0.52	0.55	5.86
Italy	Liguria (SV)	Onzo	44°3'46.51"N	8°3'12.88"E	250 m	March 2006	I	25	0.53	0.56	5.57
Italy	Liguria (GE)	Passo del Bracco	44°15'32.96"N	9°30'24.25"E	400 m	March 2006	I	25	0.45	0.50	4.14
Italy	Tuscany (PI)	Tombolo	43°36'14.01"N	10°18'6.87"E	0 m	March 2006	I	25	0.30	0.28	2.71
France	Aquitaine	Campet	44°11'118"N	00°11'647"E	150 m	March–April 2008	N	25	0.51	0.50	7.57
France	Aquitaine	Herm	43°45'25.91"N	1°8'23.65"W	40 m	March 2008	N	22	0.49	0.52	7.14
France	Aquitaine	Cestas	44°44'19.36"N	0°46'37.39"W	60 m	March 2005	N	32	0.51	0.51	8.71
Portugal		Sintra	38°47'54.41"N	9°23'17.20"W	200 m	August 2004	N	32	0.52	0.55	5.71
Spain	Galicia	Lugo	43°0'43.51"N	7°33'21.06"W	450 m	March 2008	N	23	0.45	0.45	3.71
Spain	Valencia	Chelva	39°44'51.01"N	0°59'51.64"W	470 m	February 2008	N	25	0.48	0.40	5.86
Spain	Sierra Nevada	Lanjaron	36°56'50"N	3°29'58"W	1300 m	March 2008	N	25	0.58	0.60	6.71
Morocco	Ifrane	Jaaba	33°35'03"N	5°02'59"W	1700 m	April 2008	N	22	0.42	0.30	4.29

Abbreviations: AR, allelic richness (mean number of alleles per locus); He, expected heterozygosity; Ho, observed heterozygosity; N, number of genotyped individuals.



**Figure 1** Map of the sampled localities (black dots). The shaded area represents the main distribution of the host plant *Pinus pinaster*.

STRUCTURE 2.3.3 (Pritchard *et al.*, 2000). We used 100 000 burn-in steps followed by 100 000 Markov Chain Monte Carlo (MCMC) simulation steps with a model allowing admixture. This analysis was first run on the whole data set (18 populations), the number of clusters ( $K$ ) varying from 1 to 10. It was then run on subsets of the data containing (1) only the native populations,  $K$  ranging from 1 to 8; (2) only the invasive populations,  $K = 1$  to  $K = 8$ ; and (3) each of the two clusters identified within the invasive range,  $K = 1$  to  $K = 5$

(see Results). The optimal number of clusters ( $K$ ) represented by the data was determined with the method described in Evanno *et al.* (2005), implemented in STRUCTURE HARVESTER (Earl and vonHoldt, 2012). We also examined the curve of  $\text{Log } P(X|K)$  and examined the results obtained for different values of  $K$  to detect the most stable features. To assess the consistency of results, we performed 20 independent runs for each value of  $K$ . The results were graphically displayed using DISTRICT 1.1 (Rosenberg, 2004).

**Approximate Bayesian computation analysis of introduction routes.** An approximate Bayesian computation (ABC) approach was developed to obtain probabilistic estimations of competing introduction scenarios of *M. feytaudi* in South-Eastern France, Liguria, Tuscany and Corsica from the native areas. All analyses and computations were developed using DIYABC 1.0 (Cornuet *et al.*, 2010). For each tested scenario, genetic variation within and between populations was summarized using a set of statistics conventionally used in ABC analyses. We used the mean number of alleles per locus, the mean genetic diversity and the mean allelic size variance for each population as well as pairwise  $F_{ST}$  values and the mean classification index between all pairs of populations. For three sample statistics, we used the maximum likelihood of admixture.

We built the different scenarios using Portugal, Galicia and South-Western France as the most plausible source populations in the native area, as all results ( $F_{ST}$ , genetic distances) as well as previous mitochondrial data (Burban *et al.*, 1999) suggested that other Spanish and Moroccan populations could be excluded as potential sources. As the pest scale occurred as early as the 50s in South-Eastern France, we hypothesized that this region could only originate from the native area. In Liguria, *M. feytaudi* was observed in the late 70s, and we therefore allowed an origin either from the native area or from South-Eastern France. As the pest was detected recently in Corsica (1994) and Tuscany (1999), we considered the native area, South-Eastern France and Liguria as potential sources in both cases. We chose one sampled locality per region each time several sites were studied and were not genetically differentiated (Lombaert *et al.*, 2010; 2011). This was the case in South-Western France (localities Cestas, Herm and Campet), South-Eastern France and Western Italy (Les Caunes, Gargas, Onzo) and Corsica (five sampled localities) (see Results). We chose Cestas in South-Western France because it corresponded to the highest sampling size in this region; nevertheless, choosing Campet or Herm did not change the results (data not shown). In the invasive range, we chose the localities that were closest to the first historical record of the pest in the region, namely Les Caunes in South-Eastern France and Pineto in Corsica. Moreover, the introduced population was modelled in each scenario as originating from an unsampled ('ghost') population merging into the sampled source population, taking into account the possibility of incomplete sampling in the tested source areas (Lombaert *et al.*, 2011). Introduction events were followed by a bottleneck period involving a

potentially small constant number of founders, followed by a population expansion leading to a larger stable effective population size. For all models, we assumed no regular exchange of migrants between populations, but admixture was allowed.

In order to make the ABC approach computationally feasible, we performed five serial nested analyses involving the successive *M. feytaudi* outbreaks (South-Eastern France, Liguria, Tuscany and Corsica). A new reference table taking into account the most likely scenario established in the previous step was simulated in each analysis. The same priors of the scenario parameters were used at each step, so that the posterior distributions of parameters from a given step were not used as prior in the next one.

All the tested scenarios are shown in Supplementary Figure S1. The first step (ABC1) consisted in modelling the population structure in the native area (Sintra, Lugo and Cestas) assuming a common unsampled ancestral population (10 competing scenarios). Bottlenecks were not included in the models because we expected any bottleneck event to have happened too far in the past to be detectable at the time the populations were sampled. The second step (ABC2) consisted in modelling the establishment of the South-Eastern French invasive population (Les Caunes) from the scenario that had the highest significant probability value in ABC1 (10 competing scenarios). The third step (ABC3) was built to test the origins of the Ligurian outbreak in Italy (Passo del Bracco) based on the scenarios selected previously (six competing scenarios). Finally, the fourth and fifth steps consisted in modelling the introduction event in Corsica and in Tuscany, respectively (ABC4 and ABC5, 10 competing scenarios each). A bottleneck in population size at introduction was included in all models of ABC2, ABC3, ABC4 and ABC5.

For each step of the ABC analyses, 3 000 000–5 000 000 genetic data sets were simulated using the coalescent approach implemented in DIYABC, providing 500 000 simulations for each scenario. Parameters of genetic data sets were drawn from their previous distributions (Table 2). At each step, a scenario was selected if it had the highest posterior probability value (estimated using a polychotomous logistic regression on the 1% of simulated data sets closest to the observed data), 95% CI did not overlap with that of any other scenario (Cornuet *et al.*, 2010). Confidence in scenario selection was further evaluated by computing type I and type II errors from DIYABC outputs. Posterior distributions of demographic parameters under the selected invasion scenarios in ABC2, ABC3 and ABC4 were estimated using a local linear

**Table 2** Prior distributions of demographic, historic and mutation parameters used in the ABC analyses

Parameter	Definition	ABC analysis	Distribution (interval)
$N_{anc}$	Stable effective population size of an unsampled ancestor	All	Uniform (100–20 000)
$N_U$	Stable effective population size of an unsampled population merging into an unsampled ancestor	All	Uniform (100–20 000)
$N_{U_i}$	Stable effective population size of an unsampled population merging into a possible source <i>i</i>	All	Uniform (100–20 000)
$N_s$	Stable effective population size of a sampled population	All	Uniform (100–20 000)
$N_i$	Effective number of founders during an invasion step	ABC2 ABC3, 4 and 5	Uniform (10–10 000) Loguniform (2–1000)
$ra$	Rate of admixture (only for scenarios with admixture)	All	Uniform (0.01–0.99)
$t_{anc}, t_n$	Divergence time in native source populations	All	Uniform (100–10 000)
$t_{U_i}$	Foundation time of an unsampled population merging into a possible source <i>i</i>	All	Uniform (20–100)
$t_i$	Foundation time of an invasive population	ABC2 ABC3 ABC4 ABC5	Uniform (50–100) Uniform (26–50) Uniform (11–21) Uniform (10–26)
$BD_i$	Bottleneck duration	All	Uniform (1–5)
Mean $\mu$	Mean mutation rate	All	Uniform ( $10^{-5}$ – $10^{-3}$ )
Mean $P$	Mean of the geometric distribution of the number of repeats	All	Uniform (0.1–0.3)
Mean $\mu_{SNI}$	Mean single nucleotide insertion/deletion mutation rate	All	Uniform ( $10^{-8}$ – $10^{-4}$ )

Graphical representations of the use of these parameters are presented in Supplementary Figure S1. A generalized stepwise mutation model (GSM) was used with a mean mutation rate (Mean  $\mu$ ) and a mean parameter of the geometric distribution (Mean  $P$ ) of the length in number of repeats of mutation events and  $\mu_{SNI}$  was the mean insertion or deletion of single nucleotide to the microsatellite sequence across loci. Each locus had a possible range of 40 contiguous allelic states and was characterized by individual  $\mu_{loc}$  drawn from a gamma (mean = mean  $\mu$  and shape = 2),  $P_{loc}$  drawn from a gamma (mean = mean  $P$  and shape = 2) and  $\mu_{SNI_{loc}}$  drawn from a gamma (mean = mean  $\mu_{SNI}$  and shape = 2) distributions.

regression on 1% of the simulated data sets closest to our real data (Beaumont *et al.*, 2002). The precision of parameter estimations was assessed by computing the relative median of the absolute error (RMAE) on 500 pseudo-observed data sets simulated under each best invasion scenario (a low RMAE value indicates that the parameter can be reliably estimated, Cornuet *et al.*, 2010). Because overlapping 95% CI of posterior probabilities did not allow to decide between two scenarios in ABC5 (see Results, Table 3), a joint parameter estimation for these two scenarios was performed. Following Lye *et al.* (2011), bottleneck severity at introduction was estimated as a composite demographic parameter expressed as  $\log_{10}(BD_I/N_I)$ , where  $N_I$  is the effective number of founders and  $BD_I$  the duration of the bottleneck. In all ABC steps, bottleneck duration was bounded to a maximum of five generations after introduction because *M. feytaudi* populations generally display high population growth rates, and was previously observed to reach outbreak levels in only a few years (Jactel *et al.*, 1998).

We performed a model checking analysis for the model selected in ABC4. Its goodness-of-fit was assessed from a principal component analysis in the space of summary statistics, by assessing the location of 5000 points simulated from the posterior predictive distribution relative to the one corresponding to the observed data (Cornuet *et al.*, 2010). In order to avoid an overestimation of scenario fit to our data, we used different summary statistics for model checking than for computations of parameter posterior distributions (Cornuet *et al.*, 2010), that is, the mean Garza-Williamson's M index for each population, the mean allele size variance, the shared allele distance and the distance  $(\delta\mu)^2$  between all pairs of populations.

## RESULTS

### Microsatellite and population characteristics

The total number of alleles per locus varied from 5 in locus Mat17 to 46 in Mat196. For each population, observed and expected heterozygosities and mean number of alleles per locus are given in Table 1.

All indices were lowest in Corsica and Tuscany. Allelic frequencies are shown in Supplementary Table S2.

For each locus, estimates of null allele frequencies were below 8% in at least 16 out of the 18 populations. They were above 10% in only eight cases, namely Mat211B in Tombolo (15%) and Chelva (32%), Mat252 in Lugo (16%) and Chelva (10%), Mat17 in Gavignano (11%), Mat196 in Lanjaron (11%) and Jaaba (32%), and Mat212 in Moltifao (10%). After correction for multiple comparisons, all populations were in Hardy–Weinberg equilibrium for all loci except in 3 out of 126 tests (Mat211B in Chelva and Mat196 in Lanjaron and Jaaba). Note that the two highest rates of null allele frequencies (32%) corresponded to deviations from Hardy–Weinberg equilibrium. No pairs of loci were in significant linkage disequilibrium in more than two populations, except for the pairs Mat243–Mat17 and Mat17–Mat196 that were in linkage disequilibrium in three and four populations, respectively. Hence, the microsatellite loci used were considered independent.

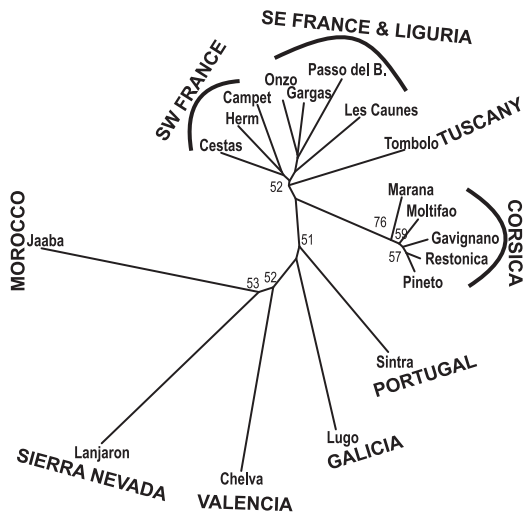
### Population genetic structure

The matrices of pairwise  $F_{ST}$  obtained with and without applying the ENA correction for the presence of null alleles are given in Supplementary Table S3. These indices were significant for most pairwise comparisons, except within South-Western France, within South-Eastern France + Onzo (Western Italy) and within Corsica. The populations from Morocco and from Southern Spain (Lanjaron) were the most differentiated from all others. The phylogenetic tree of populations (Figure 2) clearly showed a high differentiation of Moroccan and Iberian populations (Portugal, Galicia, Valencia and Andalusia), which is consistent with the high  $F_{ST}$  values found

**Table 3** Description of the five ABC analyses aiming at reconstructing the invasion routes of *Matsucoccus feytaudi* in South-Eastern France, Italy (Liguria and Tuscany) and Corsica

Analysis and target event	Considered source area (+ admixture between all population pairs)	Number of competing scenarios	Selected scenario	Posterior probability of selected scenario (95% CI)	Type I/type II errors
ABC1: structure in the native range	(1) South-Western France (2) Portugal (3) Galicia (4) Unsampled ancestor	10	All native populations originated from a common unsampled ancestor	0.8564 (0.8219–0.8910)	0.071/0.028
ABC2: South-Eastern French outbreak	(1) South-Western France (2) Portugal (3) Galicia (4) Unsampled population	10	Introduction from South-Western France	0.8277 (0.6521–1)	0.062/0.045
ABC3: First Italian outbreak (Liguria)	(1) South-Western France (2) South-Eastern France (3) Unsampled population	6	Introduction from South-Eastern France	0.7982 (0.6968–0.9478)	0.111/0.024
ABC4: Corsican outbreak	(1) South-Western France (2) South-Eastern France (3) Liguria (4) Unsampled population	10	Admixture between South-Eastern France and Liguria	0.7544 (0.6193–0.8895)	0.074/0.051
ABC5: Second Italian outbreak (Tuscany)	(1) South-Western France (2) South-Eastern France (3) Liguria (4) Unsampled population	10	Introduction from Liguria Admixture between South-Eastern France and Liguria	0.4707 (0.2725–0.6688) 0.3940 (0.3190–0.4690)	- -

Posterior probabilities were calculated by polychotomous logistic regression on the simulations corresponding to the 1% smallest Euclidean distances. In each ABC analysis, the 95% CI of the selected scenarios never overlapped with those of competing scenarios. Type I error is the proportion of cases in which the scenario considered is not selected, although being actually the true one. Type II error is the proportion of cases in which the scenario considered is selected but is not the true one. The samples used in the analyses were as follows (see Table 1): South-Western France = Cestas; Portugal = Sintra; Galicia = Lugo; South-Eastern France = Les Caunes; Liguria = Passo del Bracco; Tuscany = Tombolo; Corsica = Pineto. Scenarios involving an unsampled population as a potential source were included in all 'outbreak' analyses to make the ABC sensitive to errors concerning the source. In each ABC analysis, selection of the most likely scenario was based on its posterior probability and confidence in scenario choice was evaluated by computing type I and type II errors.



**Figure 2** Neighbour-joining tree of populations based on Cavalli-Sforza and Edwards' chord distances derived from allelic frequencies of the 7 microsatellite loci.

between these localities and all others. The Corsican populations formed a monophyletic clade with very short branches. The three South-Western French populations grouped together, and the same was true for the four populations from South-Eastern France and Liguria.

#### Test of founder effects

With few exceptions, no sign of bottleneck was detected in any continental population (Morocco, Spain, Portugal, South-Western and South-Eastern France, Italy), except for Lugo (Spain), Les Caunes (France) and Passo del Bracco (Italy), where only the Wilcoxon test under the infinite allele model (IAM) hypothesis was significant. On the contrary, populations from Corsica all experienced a severe bottleneck; in most cases, the three tests proved significant (Wilcoxon under IAM and two-phase model, and the mode-shift test). Yet, in Marana, only the shift test was significant, and in Gavignano only the two Wilcoxon tests were significant.

#### Individual assignments

All genotyped individuals were analysed using the Bayesian method implemented in STRUCTURE with the hypothetical number of clusters  $K$  ranging from 1 to 10. The Evanno's method clearly showed that  $\Delta K$  was maximal for  $K=3$  (Supplementary Figure S2). In all the 20 runs, the populations from Corsica formed a separate cluster. In 15 runs, the other two clusters were Continental France and Italy vs the Iberian peninsula and Morocco; in the other 5 runs, they corresponded to Continental France, Italy, Sintra and Lugo vs Chelva, Lanjaron and Morocco. Both types of results are shown in Figure 3a. As  $\text{Log } P(X|K)$  rather reached a plateau for  $K=5$ , we also examined the results obtained for  $K=4-7$ . A number of solutions existed for each value of  $K$ , but the populations of Corsica always grouped together, and so did the populations of South-Western France and South-Eastern France, yet with admixed individuals when  $K$  increased; the main differences between runs concerned the grouping of the Iberian and Moroccan populations (Supplementary Figure S2).

The same method was then used to analyse only the individuals from the native range, with  $K=1-8$ . In that case,  $\Delta K$  was highest for

$K=2$  and  $K=4$ , but with very low values. Consistently, the results obtained across different runs were quite variable. For  $K=2$ , the two clusters obtained in a majority of the runs (11 out of 20) were (South-Western France, Lugo and Sintra) vs (Chelva, Lanjaron and Morocco). In the other runs, the populations from South-Western France were always grouped in the same cluster. The position of the other populations varied. For  $K=4$ , the main grouping (10 runs) was (South-Western France) vs (Lugo and Sintra) vs (Chelva and Lanjaron) vs (Morocco). These results are shown in Figure 3b. Interestingly, South-Western France corresponded to a well-defined cluster in all runs, and the populations from Portugal and Galicia (Lugo and Sintra) fell within the same cluster in 16 out of 20 runs. As  $\text{Log } P(X|K)$  rather reached a plateau for  $K=5$ , we also examined the results obtained for  $K=5$  and  $K=6$ , which raised the same major conclusions (Supplementary Figure S2).

Concerning the colonization range, the results clearly pointed to  $K=2$  as the optimal number of clusters. All runs showed that one cluster grouped all Corsican populations while the other grouped South-Eastern France and Italy (Figure 3c). We also examined the genetic structure suggested for  $K=3$  and  $K=4$ , as they corresponded to high values of  $\text{Log } P(X|K)$ . When  $K$  was set to 3, the main solution (17 runs) was (Corsica)–(South-Eastern France and Liguria)–(Tuscany). For  $K=4$ , the main result was that the individuals of South-Eastern France and Liguria were mostly admixed between two groups (Supplementary Figure S2). We then looked for a potential sub-structure within both groups. There was only one cluster in Corsica, while there were two groups in the continental invasive populations, namely (South-Eastern France and Liguria) vs the population from Tuscany (Tombolo) (Figure 3d). Increasing the number of clusters in this region resulted in obtaining mostly admixed individuals (Supplementary Figure S2).

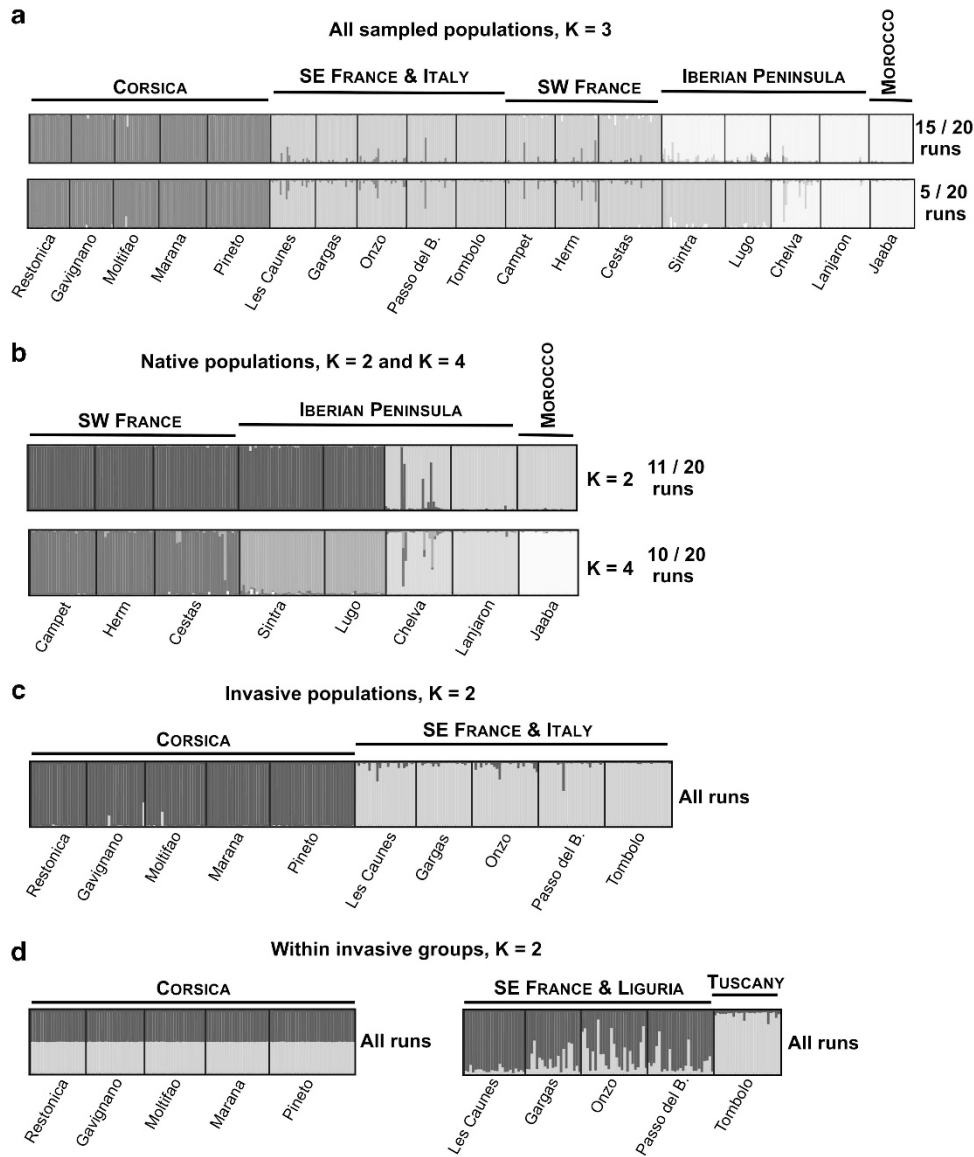
#### ABC analysis of introduction routes

In the first step of the ABC procedure (ABC1), the results unambiguously pointed to a simple scenario where all native populations originated from a common unsampled ancestor with no admixture (Table 3; Figure 4). No particular phylogenetic structure was thus chosen. The choice of this scenario was tentatively supported by a high posterior probability ( $P=0.8564$ , 95% CI never overlapping those of the other competing scenarios, Supplementary Figure S1) and low values of both type I and type II errors (Table 3).

In ABC2, there was unambiguous evidence for an introduction of South-Eastern France from the South-Western French area without any admixture (Table 3; Figure 4). This scenario was supported by a strong posterior probability ( $P=0.8277$ , 95% CI never overlapping those of the competing scenarios, Supplementary Figure S1) and low values of both type I and type II errors (Table 3).

Concerning the origin of the Ligurian outbreak in Italy (ABC3), the invasive South-Eastern French area was found the most likely source of introduction, without any admixture (Table 3; Figure 4). The corresponding scenario was supported by a strong posterior probability ( $P=0.7982$ , 95% CI never overlapping those of the competing scenarios, Supplementary Figure S1) and low values of both type I and type II errors (Table 3).

Concerning Corsica (ABC4), the results suggested that invasion resulted from an admixture between the South-Eastern French and the Ligurian populations (Table 3; Figure 4). This was supported by a strong posterior probability ( $P=0.7544$ , 95% CI never overlapping those of the competing scenarios, Supplementary Figure S1) and low values of type I and type II errors (Table 3). Concerning the origin of the Tuscan population (ABC45), there were overlapping 95% CI



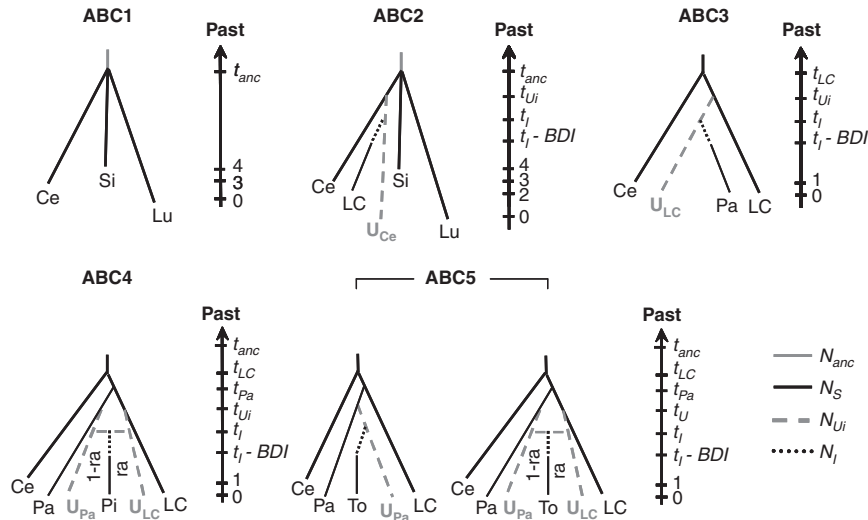
**Figure 3** Estimated population genetic structure obtained with the Bayesian analysis implemented in Structure. (a) The two most frequent results obtained from 20 runs for all sampled populations,  $K=3$  clusters; (b) Native populations,  $K=2$  and  $K=4$ ; (c) Invasive populations,  $K=2$ ; (d) Within Corsica ( $K=2$ ) and within continental invasive range ( $K=2$ ).

between the probability of an invasion from Liguria alone ( $P=0.4707$ ) and that of an admixed origin from South-Eastern France and Liguria ( $P=0.3940$ ) (Supplementary Figure S1). This did not allow to determine which of these two scenarios was the most probable (Figure 4).

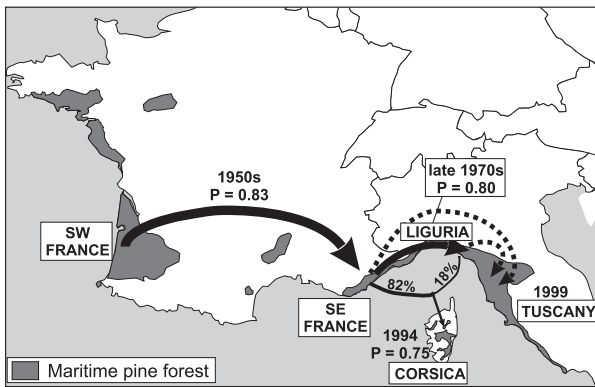
We thus developed a model checking approach only for the final scenario concerning the colonization of Corsica. The target point corresponding to the real data set was located within the principal component analysis points simulated from the posterior predictive distribution, which indicated a good fit of the model (Supplementary Figure S3). The most plausible invasion scenario of *M. feytaudi* in Southern Europe is summarized in Figure 5.

We then used a local linear regression to estimate the posterior distributions of all the parameters of the selected scenarios, except for ABC1.  $N_I$  was estimated at 883 (Q2.5% = 392, Q97.5% = 1369; RMAE = 0.179) in ABC2 (invasion of South-Eastern France); at 302

(Q2.5% = 14, Q97.5% = 799; RMAE = 0.162) in ABC3 (invasion of Liguria); and at 52 (Q2.5% = 4, Q97.5% = 175; RMAE = 0.131) in ABC4 (invasion of Corsica). In this scenario, the rate of admixture  $ra$  between South-Eastern France and Liguria was estimated at 0.82 (Q2.5% = 0.37, Q97.5% = 1; RMAE = 0.147). In ABC5 (invasion of Tuscany),  $N_I$  was estimated at 411 (Q2.5% = 29, Q97.5% = 1043) and  $ra$  at 0.33, but they were not considered fully reliable due to larger computed RMAE values (0.399 and 0.418, respectively). Computations of the bottleneck severity parameter from posterior distributions provided interesting supports for a minute, a moderate and a strong bottleneck severity during the invasions of South-Eastern France, Liguria and Corsica, respectively (Supplementary Figure S4). We did not find any significant information in the genetic data on the foundation time of each invasive population (RMAE values reaching 0.337, 0.414, 0.275 and 0.466 in ABC2, ABC3, ABC4 and ABC5, respectively).



**Figure 4** Graphical representation of the scenarios selected in each of the five ABC analyses conducted on the invasion route of *M. feytaudi* in South-Eastern France, Italy and Corsica. ABC1: population structure in the native area; ABC2: colonization of South-Eastern France; ABC3: colonization of Liguria (Italy); ABC4: colonization of the island of Corsica; ABC5: colonization of Tuscany (Italy). Note that it was not possible to distinguish between two scenarios in ABC5.  $N_{anc}$ =stable effective population size of an unsampled ancestor in the native area (number of diploid individuals).  $N_{U_i}$ =stable effective population size of an unsampled population merging into a possible source  $i$  at time  $t_{U_i}$ .  $N_S$ =stable effective population size of sampled populations of *M. feytaudi* in both native and invaded areas.  $t_i$ =foundation date of invasive populations of *M. feytaudi*.  $N_i$ =effective number of founders during an introduction step lasting  $BD_i$  generation(s).  $ra$ =rate of admixture (only for scenarios with admixture).  $t_{anc}$ =dates of ancestral divergences in native populations.  $t_{LC}$  and  $t_{Pa}$ =foundation dates of LC and Pa populations (respectively). For all scenarios, populations were assumed to be isolated from each other, with no exchange of migrants. Times ( $t_{anc}$ ,  $t_i$ ,  $t_{U_i}$ ,  $t_{LC}$ ,  $t_{Pa}$  and  $t_i$ ) were translated into numbers of generations running back in time from time 0 (sampling year 2008) by assuming one generation per year (note that time is not to scale). All parameters with associated prior distributions are described in Table 2. Posterior probabilities and type I and II errors of each selected scenario are presented in Table 3. Populations: Ce=Cestas (native, France); Si=Sintra (native, Portugal); Lu=Lugo (native, Galicia); LC=Les Caunes (invasive, France); Pa=Passo del Bracco (invasive, Liguria); To=Tombo (invasive, Tuscany); Pi=Pineto (invasive, Corsica);  $U_i$ =unsampled population merging into a possible source  $i$ .



**Figure 5** Graphical representation of the invasion scenario of *M. feytaudi* in southern Europe, including the four outbreaks in South-Eastern France, Liguria, Tuscany and Corsica, deduced from analyses based on approximate Bayesian computation (ABC). For each outbreak, the arrows indicate the most likely invasion pathways and the associated posterior probability value ( $P$ ). The dates of introduction, based on historical data, are indicated for each region. The thickness of the arrows represents the estimated numbers of founders.

## DISCUSSION

### A strong genetic structure among regions

Consistent with the fragmentation of the distribution range and the limited dispersal capacity of the studied species, most of the sampled localities were genetically significantly structured. The polymorphic nuclear markers used here allowed to better describe the population

genetic patterns as compared with the mitochondrial data studied earlier (Burban *et al.*, 1999). Microsatellite data confirmed the strong differentiation of populations from Morocco and Andalusia that were already identified as originating from divergent refugia and also allowed to identify some differentiation within the native ‘Western lineage’ and within the invasive range. All results were consistent and showed that the Corsican populations formed a homogeneous and differentiated cluster with a restricted genetic diversity. The invasive populations found in South-Eastern France and Italy also formed a cluster and were genetically very close to the native French populations sampled in the Aquitaine region. Finally, all localities from the Iberian Peninsula and Morocco were differentiated from each other and from the rest of the distribution range. Such a strong geographic structure can allow to precisely infer the colonization processes, but it can also lead to the selection of inaccurate scenarios if crucial samples are lacking in the data set, that is, if the actual source population was missed (Barrès *et al.*, 2012). Accounting for unsampled ‘ghost’ populations when building the set of scenarios to be compared in the ABC analysis was of utmost importance (Guillemaud *et al.*, 2010). In addition, the simulation of such ‘ghost’ populations allowed dealing with the existence of several slightly differentiated samples in source areas (for example, Cestas, Campet and Herm sample sites in South-Western France in this study), which were not all used to make the ABC analyses feasible (Lombaert *et al.*, 2011).

### Contrasting colonization processes in different invaded regions

Historical records suggest three steps in the colonization process of Eastern maritime pine forests: (i) introduction in South-Eastern France; (ii) expansion along the Mediterranean coast through Liguria



up to Tuscany; (iii) introduction in Corsica, but the origins and dispersal processes were still unclear. The most likely sources and dispersal modes were inferred from molecular markers using both classical population genetics and ABC approaches. In addition, the posterior probability of the selected scenario was particularly high in each ABC step of this study, and both type I and II error rates (that is, the proportions of times that a selected scenario did not have the higher posterior probability while being the true scenario and that a selected scenario had the higher posterior probability while being a wrong scenario) were always very low. Finally, the model checking procedure implemented in DIYABC and used in the final ABC4 step indicated the goodness-of-fit of the selected historical model to our genetic data (Supplementary Figure S3). In spite of a relatively low number of nuclear markers, we were thereby able to highlight drastically different colonization processes and to accurately estimate key demographic parameters, which suggest that the number of markers used was probably sufficient.

*Continental invasive populations: accidental human-aided introductions followed by natural dispersion.* South-Eastern populations were founded by individuals originating from the Aquitaine region in South-Western France, located several hundreds of kilometres apart but in the same country. Harmful invasions of pests usually originate from native sources located further away, for example, in different continents (Ciosi *et al.*, 2008; Lombaert *et al.*, 2011), or follow the introduction of their host plant (for example, the oak gall wasp *Andricus kollari*, Stone *et al.*, 2007). The rapid colonization of Europe by the horse-chestnut leafminer is one striking example of an invasive species that suddenly expanded from a geographically close region, namely the Balkans (Valade *et al.*, 2009). In that particular case though, the expansion of the moth was due to the massive plantation of its preferred host as ornamental tree in European cities. The case of *M. feytaudi* is unique in that its invasive populations originated from the same country, and its host tree was already naturally present in the invaded range. *Matsucoccus feytaudi* was probably introduced soon after (or even during) World War II, as damages were detected in the 50s and the local maritime pines are highly susceptible to the pest (Harfouche *et al.*, 1995). In Aquitaine, maritime pine mostly occurs in a large plantation forest set up for wood production purposes in the mid XIXth century. All results show that genetic differentiation between the source and the invasive area is negligible.

The nuclear microsatellite data suggest a bottleneck at introduction of very low severity (Supplementary Figure S4), supporting the hypothesis of a high number of individuals reaching South-Eastern France ( $N_I = 883$ , RMAE = 0.179) and founding the first historical invasive populations in spite of a gap in host distribution between the source and the invaded regions. This result is consistent with the absence of mitochondrial polymorphism described in South-Eastern France in spite of a low bottleneck intensity, as most populations from Aquitaine already exhibited only that particular mitochondrial haplotype. In contrast, Iberian populations exhibited higher polymorphism as expected in glacial refugia (Burban *et al.*, 1999).

In other insect invasion systems, similar estimations of bottleneck severity revealed that moderate or minute bottlenecks generally result from large numbers of founders introduced at once or from multiple introductions from one or several sources (Lombaert *et al.*, 2011; Lye *et al.*, 2011). In *M. feytaudi*, whether this occurred at once or over repeated introduction events cannot be deciphered. The mechanism underlying the observed long-distance dispersal event could be a natural transport by air current, or passive transport of invading individuals with wood movements, as observed in other forest pests

(Robinet *et al.*, 2009; Carter *et al.*, 2010). Invasive species transported by man or good exchanges are more likely to experience repeated introductions, and the founder effect in the introduced range will then be weak. The low intensity of the founder event and the origin of the invasion (the largest maritime pine production planted forest) are rather consistent with human-aided long-distance dispersal through the transportation of infested (but symptom-free) wood, probably in order to bring material for reconstruction during or after World War II. A high wood demand and the absence of symptoms in infested but resistant stands from South-Western France probably impeded risk detection and favoured repeated transportations of the pest. The scales could found viable populations in South-Eastern France because suitable hosts were present in the large natural stands of the Maures and Estérel ranges, as well as in planted areas nearby. Outbreaks were probably accelerated by the high susceptibility of local maritime pines (Harfouche *et al.*, 1995). The genetic results and historical records suggest that the invasive scale insects then expanded eastwards to Liguria over several decades, without significant genetic differentiation, nor loss of genetic diversity, from the originally invaded area. This pattern may be supported by the moderate bottleneck severity at introduction in Liguria (Supplementary Figure S4) that was assessed from the moderate estimates of effective population size ( $N_I = 302$ ). More recently, an outbreak was discovered in Tuscany where maritime pine is quite fragmented (Binazzi, 2005). Genetic data suggested that this population was significantly differentiated and could originate either from South-Eastern France or from Liguria, as the ABC analyses could not decide between both scenarios. We thus can confidently conclude that once introduced accidentally from South-Western France, the scale expanded its range through natural, gradual, short-distance dispersal along more or less continuous host forests, and finally reached the fragmented edge of the host distribution either through a stepping-stone colonization from South-Eastern France, or through a local expansion from the closest populations in Liguria. Such stratified dispersal is usually observed for insects expanding into a fragmented habitat, for example, colonizing a patchy host (Ciosi *et al.*, 2011; Gilioli *et al.*, 2013). Fully analysing and understanding the local expansion patterns would require the development of genetic and modelling approaches at regional and landscape scales (Etherington, 2011) to take into account habitat characteristics and connectivity. Reliable historical surveys and data mining should then be used to validate the results (Gilioli *et al.*, 2013). Such analyses fall beyond the goal of the present paper and would represent interesting perspectives.

*Colonization of Corsica: a rare event of long-distance, wind-borne dispersal.* The patterns observed in Corsica were drastically different, as the invasive populations there showed signs of a strong founder effect with a relatively low number of founders ( $N_I = 52$ , Supplementary Figure S4). Moreover, all the populations sampled in Corsica had the same allelic pools and were genetically not differentiated, suggesting one unique colonization event followed by a step-by-step local expansion between neighbouring host patches, as was shown for the Cedar seed wasp *Megastigmus schimitscheki* (Auger-Rozenberg *et al.*, 2012). Similar findings of severe bottlenecks at introduction in different species suggested that successful invasions can also result from a very small number of original migrants, which is sometimes considered as a characteristic of successful invasive species (Kaňuch *et al.*, 2013). Diverse genetic and/or ecological mechanisms may circumvent the loss of genetic variation occurring during introduction events (Lye *et al.*, 2011; Auger-Rozenberg *et al.*,

2012). It is worth noting that the scale was first observed in a forest patch far from the main communication routes (harbours and main roads) (Jactel *et al.*, 1998). The long-distance colonization event that led to the colonization of the island is thus likely to be independent from any human activity and could correspond to an accidental transport of larvae from the continent due to the main winds. The scale insects seem to have been transported mainly from the shores of both South-Eastern France and Liguria (Table 3; Figure 4), but with a much higher contribution of the French source area to the genetic pool of the Corsican populations, as the admixture rate was estimated to 82%. This scenario is fully consistent with the action of the dominant wind, namely the Mistral, which is classically observed in spring (that is, when larvae are available) in this region of the Mediterranean Sea. It is known to extend from the French coasts as far as few hundred kilometres, eventually reaching Corsica (Jansá, 1987). It is associated with outflows blowing from Liguria and the Gulf of Genoa, which results in the ‘Genoa cyclone’, that is, turning winds between Corsica and the coast of northern Italy (Salameh *et al.*, 2007). The genetic data are consistent with the observation of particles emitted both from the French shore and Italy and found over the western Mediterranean area (Salameh *et al.*, 2007). These wind data thus suggest that the introduction of *M. feytaudi* in Corsica is most likely due to the accidental, passive wind transportation of larvae during such a strong Mistral event in spring. The colonization of the northern Mediterranean coast was a pre-requisite for a successful invasion of Corsica, South-Eastern France acting as a ‘bridgehead’, that is, as a new potential source population to reach new territories (Lombaert *et al.*, 2010). A similar situation was observed for the bush cricket *Metrioptera roeselii* that crossed the Baltic Sea once introduced along the coast (Kaňuch *et al.*, 2013). Larvae may be regularly transported from the French and Italian coasts to Corsica, but the probability of settlement is rather low, because the larvae need a suitable host to survive and found an invasive population. Actually, although the insect has been present in South-Eastern France since the 50s, it was detected in Corsica only in the mid 1990s. Once introduced, it could expand in a diffusive manner and reach the existing maritime pine stands in the island, similarly to its expansion to Italy once introduced in South-Eastern France.

## CONCLUSION

To summarize, the present study provided for the first time strongly supported hypotheses describing the invasion routes and dispersal processes of a major pest responsible for the decline of maritime pine forests in South-Eastern France, Italy and Corsica. The studied biological model is unique in that it originated from a relatively local origin and invaded previously unoccupied patches of its native host. Interestingly, we showed that the colonization history of the maritime pine scale involved drastically different processes, namely passive long-distance human-assisted, stratified or passive long-distance natural dispersal. More, we found that the first invaded region constituted a ‘bridgehead’ (*sensu* Lombaert *et al.*, 2010) from which wind-borne larvae could accidentally reach an island, Corsica. The relictual maritime pine stands of Algeria and Tunisia, which are genetically close to South-Eastern French, Corsican and Italian populations of this tree species (Burban and Petit, 2003), are so far free of this major pest. As they proved to be highly susceptible in provenance trials (Harfouche *et al.*, 1995) strict targeted monitoring and management policies should now be developed to prevent new worldwide introductions of *M. feytaudi*.

## DATA ARCHIVING

Microsatellite genotype data have been deposited at Dryad, doi:10.5061/dryad.bt29j. DNA sequences have been deposited in GenBank, accession numbers KJ508822 and KJ508823.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

This work benefited from the efficient help of many collaborators for sampling. We thank M Branco, O Cerati, M Chèdeville, G Cortini, M Hachimi, M Hautcloq, J Hodar, F Lombardero, H Mas i Gisbert, P Ménassieu, J Mirault, F Pennacchio, H Ramzi, J Regad, F-X Saintonge, L Tastevin, J Thévenet, I Van Halder and C Vidal. We also thank Jonathan Descat for his help with DNA extraction and microsatellite genotyping. The genotyping and sequencing were performed at the Genotyping and Sequencing facility of Bordeaux (grants from the Conseil Régional d’Aquitaine n° 20030304002FA and 20040305003FA and from the European Union, FEDER n° 2003227). We are grateful to Franck Salin for his help and advice during laboratory work. This research was supported by the Collectivité Territoriale de Corse (Convention n°B20050177).

- Ahmed S, Compton SG, Butlin RK, Gilmartin PM (2009). Wind-borne insects mediate directional pollen transfer between desert fig trees 160 kilometers apart. *Proc Natl Acad Sci USA* **106**: 20342–20347.
- Auger-Rozenberg M-A, Boivin T, Magnoux E, Courtin C, Roques A, Kerdelhué C (2012). Inferences on population history of a seed chalcid wasp: invasion success despite a severe founder effect from an unexpected source population. *Mol Ecol* **21**: 6086–6103.
- Barrès B, Carlier J, Seguin M, Fenouillet C, Cilas C, Ravigné V (2012). Understanding the recent colonization history of a plant pathogenic fungus using population genetic tools and Approximate Bayesian Computation. *Heredity* **109**: 269–279.
- Beaumont MA, Zhang W, Balding DJ (2002). Approximate Bayesian computation in population genetics. *Genetics* **162**: 2025–2035.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996–2004). GENETIX 4.05, logiciel sous Windows™ pour la génétique des populations. *Laboratoire Génome, Populations, Interactions, CNRS UMR 5000*. Université de Montpellier II: Montpellier (France).
- Bertorelle G, Benazzo A, Mona S (2010). ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Mol Ecol* **19**: 2609–2625.
- Binazzi A (2005). La cocciniglia del pino marittimo in Italia—Strategie di contenimento del *Matsucoccus feytaudi* Ducasse e orientamenti per gli interventi di recupero ambientale delle pinete di *Pinus pinaster* Aiton. *APAT Rapporti* **55/2005**: 75–88.
- Burban C, Petit RJ, Carcreff E, Jactel H (1999). Rangeland variation of the maritime pine bast scale *Matsucoccus feytaudi* Duc. (Homoptera: Matsucoccidae) in relation to the genetic structure of its host. *Mol Ecol* **8**: 1593–1602.
- Burban C, Petit RJ (2003). Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. *Mol Ecol* **12**: 1487–1495.
- Carrasco LR, Mumford JD, MacLeod A, Harwood T, Grabenweger G, Leach AW *et al.* (2010). Unveiling human-assisted dispersal mechanisms in invasive alien insects: Integration of spatial stochastic simulation and phenology models. *Ecol Model* **221**: 2068–2075.
- Carter M, Smith M, Harrison R (2010). Genetic analyses of the Asian longhorned beetle (Coleoptera, Cerambycidae, *Anoplophora glabripennis*), in North America, Europe and Asia. *Biol Invasions* **12**: 1165–1182.
- Chapuis M-P, Estoup A (2007). Microsatellite null alleles and estimation of population differentiation. *Mol Biol Evol* **24**: 621–631.
- Ciosi M, Miller J, Kim KS, 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. *Mol Ecol* **17**: 3614–3627.
- Ciosi M, Miller NJ, Toepfer S, Estoup A, Guillemaud T (2011). Stratified dispersal and increasing genetic variation during the invasion of Central Europe by the western corn rootworm, *Diabrotica virgifera virgifera*. *Evol Appl* **4**: 54–70.
- Cornuet J-M, Ravigné 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.
- Earl DA, vonHoldt BM (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Cons Genet Res* **4**: 359–361.
- Estoup A, Guillemaud T (2010). Reconstructing routes of invasion using genetic data: why, how and so what? *Mol Ecol* **19**: 4113–4130.
- Etherington TR (2011). Python based GIS tools for landscape genetics: visualising genetic relatedness and measuring landscape connectivity. *Methods Ecol Evol* **2**: 52–55.
- Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**: 2611–2620.

- Excoffier L, Laval G, Schneider S (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinformatics Online* **1**: 47–50.
- Fabre J-P (1980). Mortalité dans les peuplements de pin maritime à la suite de l'introduction de *Matsucoccus feytaudi* Duc. en Italie. *Italia Forestale e Montana* **35**: 40–42.
- Gilioli G, Pasquali S, Tramontini S, Riolo F (2013). Modelling local and long-distance dispersal of invasive chestnut gall wasp in Europe. *Ecol Model* **263**: 281–290.
- Guillemaud T, Beaumont MA, Ciosi M, Cornuet JM, Estoup A (2010). Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity* **104**: 88–99.
- Harfouche A, Baradat P, Durel CE (1995). Variabilité intraspécifique chez le pin maritime (*Pinus pinaster* Ait) dans le sud-est de la France. I. Variabilité des populations autochtones et des populations de l'ensemble de l'aire de l'espèce. *Ann For Sci* **52**: 307–328.
- Jactel H, Mégnassieu P, Ceria A, Burban C, Regad J, Normand S *et al.* (1998). Une pullulation de la cochenille *Matsucoccus feytaudi* provoque un début de dépérissement du pin maritime en Corse. *Rev For Fr* **50**: 33–45.
- Jactel H, Mégnassieu P, Lettere M, Mori K, Einhorn J (1994). Field response of maritime pine scale, *Matsucoccus feytaudi* Duc. (Homoptera: Margarodidae), to synthetic sex pheromone stereoisomers. *J Chem Ecol* **20**: 2159–2170.
- Jansá A (1987). Distribution of the Mistral: a satellite observation. *Meteorol Atmos Phys* **36**: 201–214.
- Kañuch P, Berggren Å, Cassel-Lundhagen A (2013). Colonization history of *Metrioptera roeselii* in northern Europe indicates human-mediated dispersal. *J Biogeogr* **40**: 977–987.
- Kerdelhué C, Decroocq S (2006). Characterization of 8 new microsatellite loci in the invading maritime pine bast scale *Matsucoccus feytaudi* (Hemiptera: Coccoidea: Margarodidae). *Mol Ecol Notes* **6**: 1168–1170.
- Liebold AM, Tobin PC (2008). Population ecology of insect invasions and their management. *Annu Rev Entomol* **53**: 387–408.
- Lombaert E, Guillemaud T, Cornuet J-M, Malausa T, Facon B, Estoup A (2010). Bridgehead effect in the worldwide invasion of the biocontrol harlequin ladybird. *PLoS ONE* **5**: e9743.
- Lombaert E, Guillemaud T, Thomas CE, Lawson Handley LJ, Li J, Wang S *et al.* (2011). Inferring the origin of populations introduced from a genetically structured native range by approximate Bayesian computation: case study of the invasive ladybird *Harmonia axyridis*. *Mol Ecol* **20**: 4654–4670.
- Lye GC, Lepais O, Goulson D (2011). Reconstructing demographic events from population genetic data: the introduction of bumblebees to New Zealand. *Mol Ecol* **20**: 2888–2900.
- Piry S, Luikart G, Cornuet J-M (1999). BOTTLENECK: a computer program for detecting recent reduction in the effective population size using allele frequency data. *J Heredity* **90**: 502–503.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Reynolds AM, Reynolds DR (2009). Aphid aerial density profiles are consistent with turbulent advection amplifying flight behaviours: abandoning the epithet 'passive'. *Proc R Soc Lond B* **276**: 137–143.
- Robert CP, Cornuet J-M, Marin JM, Pillai NS (2011). Lack of confidence in approximate Bayesian computation model choice. *Proc Natl Acad Sci USA* **108**: 15112–15117.
- Robinet C, Roques A, Pan H, Fang G, Ye J, Zhang Y *et al.* (2009). Role of human-mediated dispersal in the spread of the pinewood nematode in China. *PLoS ONE* **4**: e4646.
- Rosenberg NA (2004). DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes* **4**: 137–138.
- Salameh T, Drobinski P, Menut L, Bessagnet B, Flamant C, Hodzic A *et al.* (2007). Aerosol distribution over the western Mediterranean basin during a Tramontane/Mistral event. *Ann Geophys* **25**: 2271–2291.
- Shigesada N, Kawasaki K, Takeda Y (1995). Modeling stratified diffusion in biological invasions. *Am Nat* **146**: 229–251.
- Stone GN, Challis RJ, Atkinson RJ, Csóka G, Hayward A, Melika G *et al.* (2007). The phylogeographical clade trade: tracing the impact of human-mediated dispersal on the colonization of northern Europe by the oak gallwasp *Andricus kollari*. *Mol Ecol* **16**: 2768–2781.
- Valade R, Kenis M, Hernandez-Lopez A, Augustin S, Mari Mena N, Magnoux E *et al.* (2009). Mitochondrial and microsatellite DNA markers reveal a Balkan origin for the highly invasive horse-chestnut leaf miner *Cameraria ohridella* (Lepidoptera, Gracillariidae). *Mol Ecol* **18**: 3458–3470.

Supplementary Information accompanies this paper on Heredity website (<http://www.nature.com/hdy>)