MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 6086–6103 doi: 10.1111/mec.12077

Inferences on population history of a seed chalcid wasp: invasion success despite a severe founder effect from an unexpected source population

M.-A. AUGER-ROZENBERG,*[1] T. BOIVIN,†[1] E. MAGNOUX,* C. COURTIN,* A. ROQUES* and C. KERDELHUɇ
*INRA, UR633 Unité de Recherche de Zoologie Forestière, 2163 Avenue de la Pomme de Pin CS 40001 Ardon, F-45075, Orléans, Cedex 2, France, †INRA, UR629 Ecologie des Forêts Méditerranéennes, Site Agroparc, F-84914, Avignon, Cedex 09, France, ‡INRA, UMR CBGP (INRA; IRD; CIRAD; Montpellier Supagro), Campus International de Baillarguet, CS 30016, F-34988, Montferrier-sur-Lez, France

Abstract

Most invasive species established in Europe originate from either Asia or North America, but little is currently known about the potential of the Anatolian Peninsula (Asia Minor) and/or the Near East to constitute invasion sources. Mediterranean forests are generally fragile ecosystems that can be threatened by invasive organisms coming from different regions of the Mediterranean Basin, but for which historical data are difficult to gather and the phylogeographic patterns are still poorly understood for most terrestrial organisms. In this study, we characterized the genetic structure of *Megastigmus schimitscheki*, an invasive seed-feeding insect species originating from the Near East, and elucidated its invasion route in South-eastern France in the mid 1990s. To disentangle the evolutionary history of this introduction, we gathered samples from the main native regions (Taurus Mountains in Turkey, Lebanon and Cyprus) and from the invaded region that we genotyped using five microsatellite markers and for which we sequenced the mitochondrial Cytochrome Oxidase I gene. We applied a set of population genetic statistics and methods, including approximate Bayesian computation. We proposed a detailed phylogeographic pattern for the Near East populations, and we unambiguously showed that the French invasive populations originated from Cyprus, although the available historical data strongly suggested that Turkey could be the most plausible source area. Interestingly, we could show that the introduced populations were founded from an extremely restricted number of individuals that realized a host switch from *Cedrus brevifolia* to *C. atlantica*. Evolutionary hypotheses are discussed to account for this unlikely scenario.

Keywords: biological invasion, Near East, *Megastigmus*, mitochondrial DNA, microsatellite, bottleneck

Received 14 May 2012; revision received 29 August 2012; accepted 4 September 2012

Introduction

Elucidating the routes of introduction of invasive species is essential for the support of effective vigilance against the identified key source populations. It also helps to define the ecological characteristics of introduced populations and then predict the potential extent of their distribution areas (Kolar & Lodge 2001; Estoup & Guillemaud 2010). A recent survey of the alien insects that have established in Europe showed that the arrival of such species has significantly accelerated since the mid-1990s, as a probable result of globalization and the subsequent worldwide movements of goods (Roques 2010a,b). Most invasive species found in Europe originate from Asia and North America, but very rarely from Asia Minor (i.e., Anatolian Peninsula)

Correspondence: M.-A. Auger-Rozenberg, Fax: +33 2 38 41 78 79; E-mail: Marie-Anne.Auger-Rozenberg@orleans.inra.fr

[1] These authors contributed equally to this work.

and/or the Near East. Little is actually known about the potential of these areas to constitute invasion sources. A number of invaders are actually 'hitch-hikers', that is, associated to alien organisms intentionally introduced by humans. One such example is insects associated with cultivated or ornamental plants, in which case their cryptic lifestyle impedes species detection and phytosanitary interventions. Species developing within fruits or seeds are particularly difficult to detect, and may have a high invasive potential (Roques *et al.* 2003).

The insect genus *Megastigmus* (Hymenoptera: Torymidae) is a group of micro-hymenoptera that mostly develop in seeds of various plants (Auger-Rozenberg & Roques 2012). The females directly lay eggs in the host plant, and each larva develops within a single seed until adulthood, affecting the host reproduction. Infested seeds are visually undetectable. Due to seed trade and the increase in international seed exchanges, several species of *Megastigmus* have been introduced accidentally in different parts of the world. When successful, such invasions often have major economic consequences (Roques & Skrzypczynska 2003; Roques 2010a). Previous data on molecular phylogeny of *Megastigmus* spp. showed that the evolution of the conifer-associated wasps is strongly related to host use (Auger-Rozenberg *et al.* 2006). Not surprisingly, in many cases, the *Megastigmus* species were accidentally introduced when their hosts were brought to a new region for cultivation purposes.

The host plant genus *Cedrus* (Pinaceae) includes four coniferous evergreen species, three of which originate from the Mediterranean region (namely *C. atlantica* Carr., *C. libani* A. Richard and *C. brevifolia* (Hook. f.) Henry) and one from the Himalayas (*C. deodara* (Roxburgh) Don). *Cedrus* species have been successfully introduced in many countries outside their natural range (Qiao *et al.* 2007), while native populations have decreased due to human activity (*C. libani* and *C. atlantica*) or to natural decline in population size (*C. brevifolia*, limited to Cyprus) (Bou Dagher-Kharrat *et al.* 2007; Fady *et al.* 2008). Seeds of *C. atlantica* were imported from North Africa to France in the 1860s for reforestation of the Mediterranean zone. The associated seed chalcid *M. pinsapinis* (Hoffmeyer) was probably introduced in France at the end of the nineteenth century with imported seed lots (Fabre *et al.* 2004). Much more recently, the closely related species *M. schimitscheki* Novitzky was detected in 1995 from cedar seeds in Southern France (Fabre *et al.* 2004; Roques *et al.* 2008). Its native range is the Eastern Mediterranean region, where it develops on *C. libani* in the Near East (Turkey, Syria and Lebanon) and *C. brevifolia* in Cyprus (Fabre *et al.* 2004). Although these two cedar species were introduced in many Mediterranean countries as ornamental trees, invasive populations of *M. schimitscheki* have been observed to date solely in South-eastern France. *M. schimitscheki* has progressively invaded most of the planted *C. atlantica* stands, even showing a competitive advantage over its direct competitor *M. pinsapinis*. It is responsible for considerable damage, destroying up to 80% of the seeds in a given locality (Boivin *et al.* 2008).

As *C. libani* seeds used for reforestation of Southern France mainly originated from Turkey in the early 1990s, foresters and entomologists logically suspected a Turkish origin (Boivin *et al.* 2008). Yet, a possibility still remained that the first migrants were imported from Lebanon with seeds of *C. libani* (Fabre *et al.* 2004). On the contrary, a Syrian or Cypriot origin was considered as very unlikely, due to the very restricted seed imports from these countries: *Cedrus* populations are extremely restricted in Syria, where the presence of the wasp is still to be ascertained (Novitzky 1954). Besides, seed importations from Cyprus were not authorized in France (French Ministry of Food, Agriculture & Fishing 2010).

Plausible routes of introduction have traditionally been inferred using direct methods relying on current and historical observations of invasive species and indirect methods relying on patterns of population genetic data (Guillemaud *et al.* 2010). While direct methods are based on historical and observational data that may be incomplete, sparse or misleading, indirect methods using genetic data have frequently shed a critical light on complex and sometimes counterintuitive introduction histories (Estoup & Guillemaud 2010). But stochasticity in the demographic and genetic history occurring during introduction (such as genetic drift and admixture events) may produce complex genetic signals that most indirect methods fail to decipher (Guillemaud *et al.* 2010). Approximate Bayesian computation (ABC) allows the use of molecular data to perform model-based inference in a Bayesian setting (Beaumont *et al.* 2002). The ABC approach has been recently adapted to invasion-associated demographic and evolutionary scenarios (Guillemaud *et al.* 2010; reviewed in Estoup & Guillemaud 2010). ABC is an extension of classical full-likelihood Bayesian analysis, which allows bypassing exact likelihood computation in complex demographic scenarios by using coalescent-based simulations and summary statistics (Csilléry *et al.* 2010).

The aim of the present study was (i) to assess the genetic structure of *M. schimitscheki* populations over their native range, as the phylogeographic patterns of species from the Near East are still poorly described and understood; and (ii) to determine the most plausible demographic scenario(s) describing the introduction history of the invasive French populations. We analysed insects from both native and introduced ranges, gathered in all the regions where *M. schimitscheki* has been mentioned except for Syria. Our goal was also to determine whether genetic data support a bottleneck in population size during the introduction of *M. schimitscheki* in France and if the severity of this bottleneck could be quantified. To achieve these aims, we used two types of molecular markers, namely mitochondrial DNA sequences (mtDNA) and five microsatellite loci. Both markers can now be combined in ABC analyses to more accurately test alternative historical scenarios (Cornuet *et al.* 2010).

Materials and methods

Sample collection, identification and DNA extractions

Samples were collected from 19 sites in the native and introduced ranges of *M. schimitscheki* (Table 1, Fig. 1). The wasps were sampled in the native range by randomly collecting seeds from the ground once the cones had disarticulated, in eleven cedar forests (eight in two regions of Turkey (Western and Eastern Taurus mountains) in 1997, one in Cyprus in 2003 and 2004 and two in Lebanon in 2005 and 2006; difficulties in sampling such material explaining temporal differences of collection dates). In France, we also sampled eight localities during the autumn of 2003. We randomly collected ten cones per tree from five trees per site before cones disarticulated. Wasp-infested seeds were detected by X-ray radiography using a Faxitron-43855® apparatus (15 kV, 3 mA, 3′30″–4′30″) and X-ray sensitive films (Kodak® 'Industrex M'). Infested seeds were placed in individual rearing boxes stored in an outdoor insectary located at INRA, Orléans, France or at INRA, Avignon, France. After emergence and identification, the insects were preserved in 100% ethanol at −20 °C.

As wasps are haplodiploid, only diploid (i.e. female) genotypes were used to assess concordance with the assumptions of further population genetic analyses, as in studies carried out on other species of this insect genus (Nicholls *et al.* 2010). Total genomic DNA was extracted from the entire body of adults following procedures from the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) and eluted in 200 μL of AE buffer, or using the Chelex method (Walsh *et al.* 1991). The numbers of individuals per population used for sequencing (mtDNA) and genotyping (microsatellites) are presented in Table 1.

MtDNA sequencing and analysis

We used polymerase chain reactions (PCR) to amplify a segment of cytochrome oxidase I (COI) using the primers 1775-COI-F (Clyde, 5′-CGAATAAATAATA-TAAGATTTTG-3′), and 2773-COI-R (Bonnie, 5′-GGAT AATCTCTATATCGACGAGGTAT-3′) (Scheffer & Grissell 2003). All PCR products were purified with QIAquick PCR purification kit (Qiagen) or genelute PCR clean-up kit (Sigma, Steinheim, Germany). Purified PCR products were directly sequenced with Bonnie and an internal primer C1-J-2183 (Jerry, 5′-CAACATTTATT TTGATTTTTTGG-3′, Simon *et al.* 1994). Sequencing was performed using the big-dye terminator sequencing kit (PE Applied Biosystems) and carried out with an ABI 3100 automatic sequencer.

One hundred and thirty six individuals were sequenced in both directions. Sequences were aligned using Clustal W (Thompson *et al.* 1994) as implemented in BioEdit 7.05 (Hall 1999). No insertion, deletion or stop codons were observed. All sequences were truncated to the same length (730 bp).

A 95% confidence statistical parsimony network was constructed using TCS 1.21 (Clement *et al.* 2000). Number of haplotypes, haplotype diversity and nucleotide diversity were obtained using DNASP 5.0 (Librado & Rozas 2009). Neutrality of the data was tested for geographic populations using Fu's *Fs* (Fu 1997) and R2 (Ramos-Onsins & Rozas 2002) statistics available in DnaSP. A simulated annealing procedure implemented in SAMOVA 1.0 (Spatial Analysis of MOlecular VAriance, Dupanloup *et al.* 2002) was used within the native area to define groups of samples that are geographically homogeneous and differentiated from each other. The GPS coordinates of each sampled locality in the Near East were used as geographic information. The simulations were conducted with K (number of groups) ranging from two to ten and repeating each simulation annealing process 100 times. The optimal number of groups supposedly corresponds to the value of K for which the F_{CT} value (the proportion of genetic variation among groups) reaches a plateau (Dupanloup *et al.* 2002).

Microsatellite genotyping and analysis

Samples were genotyped at five microsatellite loci previously developed for *M. spermotrophus* (Boivin *et al.* 2003). Either the forward or reverse primer set was 5′ end-labelled with a fluorescent dye (6-FAM, HEX-Sigma, or NED-Applied Biosystems) for laser-beam detection of the PCR products that were sized on an ABI PRISM^R 3100 Genetic Analyzer using the GeneScan software (Applied Biosystems). Genotypes that could not be reliably scored were re-amplified at least once, and if necessary were excluded as missing data. The data set finally contained 340 individuals from 19 sites.

File format conversions were done using Convert 1.31 (Glaubitz 2004). The same software was used to

Table 1 Sample information including the sampled localities and hosts, site codes, coordinates, sample size (N), Haplotype frequencies, nucleotide diversity (π ± SD), mean expected heterozygosity (He), mean observed heterozygosity (Ho), allelic richness adjusted to a minimum sample size of 10 (AR) and private allelic richness (PAR) estimated after a rarefaction procedure

Country	Locality	Site code	Geographic coordinates	Host species	N (mtDNA)	Haplotypes abundance	π	N (microsatellites)	He	Ho	AR	PAR
Intro. area												
France	Ménerbes	1	43°48′98″N 5°13′15″E	*C. atlantica*	5	5 HT1	0	14	0.379	0.443	2.14	0.00
	Vénasques	2	43°58′57″N 5°12′28″E	*C. atlantica*	5	5 HT1	0	30	0.489	0.480	2.63	0.00
	Mont Ventoux	3	44°07′45″N 5°15′36″E	*C. atlantica*	8	8 HT1	0	30	0.536	0.560	2.59	0.00
	Ventouret	4	44°08′29″N 5°23′10″E	*C. atlantica*	7	7 HT1	0	25	0.514	0.440	2.67	0.00
	Sault	5	44°06′12″N 5°25′25″E	*C. atlantica*	6	6 HT1	0	6	—	—	—	—
	St Michel l'Observatoire	6	43°53′32″N 5°43′46″E	*C. atlantica*	7	7 HT1	0	25	0.331	0.304	2.33	0.03
	St Pierre d'Argençon	7	44°51′66″N 5°70′00″E	*C. atlantica*	5	5 HT1	0	25	0.402	0.384	2.59	0.00
	Sisteron	8	44°14′13″N 5°55′10″E	*C. atlantica*	7	7 HT1	0	29	0.486	0.545	2.55	0.00
	Total								0.466	0.457	2.50	0.00
Origin area												
West Turkey	Arpaçik	9	37°49′00″N 29°14′00″E	*C. libani*	6	1 HT1, 3 HT2, 2 HT4	0.00210 ± 0.00087	13	0.593	0.619	6.01	0.45
	Konak	10	37°17′00″N 29°04′00″E	*C. libani*	6	4 HT2, 1 HT3, 1 HT6	0.00137 ± 0.00061	7	—	—	—	—
	Sevindik	11	36°32′00″N 29°46′00″E	*C. libani*	7	4 HT2, 1 HT3, 1 HT5, 1 HT10	0.00117 ± 0.00039	10	0.704	0.460	8.00	0.70
	Aykiriçay	12	36°27′00″N 30°10′00″E	*C. libani*	6	5 HT2, 1 HT5	0.00046 ± 0.00029	14	0.652	0.624	6.66	0.23
	Total								0.677	0.577	6.89	0.46
East Turkey	Sultandagi	13	38°32′00″N 31°08′00″E	*C. libani*	7	1 HT1, 6 HT2	0.00117 ± 0.00081	16	0.728	0.644	7.33	0.78
	Kapidag	14	38°05′00″N 30°42′00″E	*C. libani*	6	4 HT2, 1 HT3, 1 HT7	0.00091 ± 0.00038	16	0.795	0.768	8.05	1.04

Table 1 *continued*

Country	Locality	Site code	Geographic coordinates	Host species	N (mtDNA)	Haplotypes abundance	π	N (microsatellites)	He	Ho	AR	PAR
	Y. Gokdere	15	37°42′00″N 30°49′00″E	*C. libani*	6	5 HT2, 1 HT8	0.00046 ± 0.00029	4	—	—	—	—
	Abanoz	16	36°20′00″N 32°56′00″E	*C. libani*	13	2 HT1, 8 HT2, 2 HT3, 1 HT9	0.00176 ± 0.00058	24	0.816	0.717	8.46	0.96
	Total								0.711	0.809	7.95	0.93
Cyprus	Paphos	17	34°59′00″N 32°41′00″E	*C. brevifolia*	7	7 HT1	0	29	0.685	0.655	5.57	0.24
Lebanon	Tannourine	18	34°11′00″N 35°55′09″E	*C. libani*	19	2 HT2, 17 HT11	0.00054 ± 0.00031	20	0.554	0.560	4.18	0.51
	Maasser Ech′ Chouff	19	33°40′00″N 35°40′00″E	*C. libani*	3	3 HT11	0	3	—	—	—	—

calculate the numbers and allelic frequencies of private and shared alleles among populations. Mean observed and expected heterozygosities (H_O and H_E, respectively) were computed over all loci for each sampling site in Genetix 4.05.2 (Belkhir *et al.* 2004) (Table 1). We used FSTAT 2.9.3 (Goudet 2002) to calculate allelic richness (AR) for each population and each main region, corrected for differences in sample size. To calculate the richness of private alleles, we used the rarefaction procedure in HP-RARE (Kalinowski 2005), which compensates for the increase in likelihood of sampling rare alleles as sample size increases.

Hardy-Weinberg equilibrium (HWE) was tested using Arlequin 3.11 (Excoffier *et al.* 2005) for each locus and population, with a significance level of 95%, using 1000 permutation steps and 100 000 steps in the Markov chain. Linkage-disequilibrium (LD) was also tested in each population for all pairs of loci with 10 000 permutations. Sequential Bonferroni corrections (Rice 1989) for multiple comparisons were applied for both tests.

FreeNA (Chapuis & Estoup 2007) was used to estimate the null allele (NA) frequencies for each locus in each population according to the Expectation Maximization algorithm described by Dempster *et al.* (1977). Pairwise F_{ST} (Weir & Cockerham 1984) were estimated using the Excluding Null Alleles (ENA) correction implemented in FreeNA (Chapuis & Estoup 2007). The 95% confidence intervals of corrected pairwise F_{ST} values were obtained by bootstrapping 5000 times over loci. A population-based neighbour-joining (NJ) tree was reconstructed using Populations 1.2.30 software (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 2000 replicates. Pairwise F_{ST} and phylogenetic trees of populations were assessed after exclusion of the four populations for which the sampling size was insufficient, namely Konak (Pop10), Y. Gokdere (Pop15), Maasser Ech′ Chouff (Pop19) and Sault (Pop5).

We also performed assignment tests using all genotyped individuals, i.e., also including the populations with a low sample size. We first used the program Structure 2.3.1 (Pritchard *et al.* 2000) to estimate the number of genetic clusters actually represented in the data set to disentangle the genetic structure over the sampled area. We used 100 000 burn-in steps followed by 100 000 MCMC simulation steps with a model allowing admixture. To assess the consistency of results, we performed 25 independent runs for each value of *K* ranging from 1 to 10. A similar analysis was then conducted within each of the two identified groups, namely the native range (11 localities in the Near East) and the invasive range (eight

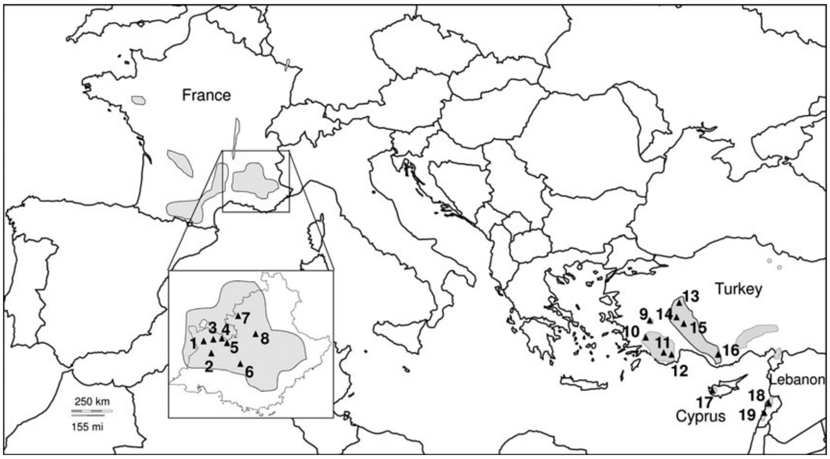


Fig. 1 Sample locations of 19 *M. schimitscheki* populations used in this study. The shaded zone corresponds to the distribution areas of the *Cedrus* host species (from Fady *et al.* 2008).

populations from France) with $K = 1$–8 and $K = 1$–6, respectively, using the same options. The most likely number of clusters was determined using the method described in Evanno *et al.* (2005).

Subsequently, we assigned the invasive French individuals to the four Near East main regions, namely West Turkey, East Turkey, Cyprus and Lebanon (that also proved to correspond to different genetic clusters, see results) using Geneclass 2.0 (Piry *et al.* 2004). The Bayesian assignment test was based on microsatellite signature using the standard criterion described by Rannala & Mountain (1997). We calculated the mean multilocus assignment likelihood for each French individual (i) to each of the four putative sources (the groups of Near East native populations) (s), expressed as $-\log(L_{i\rightarrow s})$. Then, we calculated the mean of this statistic per introduced population. The deduced most probable source was identified as the group with the lowest $-\log(L_{i\rightarrow s})$ values for a given French population (Ciosi *et al.* 2008).

Approximate Bayesian Computation analysis of introduction routes

An ABC approach was further developed using DIYABC 1.0 (Cornuet *et al.* 2010) with both mitochondrial and microsatellite markers, to generate probabilistic estimations of competing introduction scenarios. DIYABC provides the opportunity to explicitly include unsampled populations that potentially contribute to an invasion scenario (see Miller *et al.* 2005; Guillemaud *et al.* 2010; Lombaert *et al.* 2011). Basically, a scenario is quantified by a historical model describing how the sampled populations are connected to their common ancestor through a succession in time of events (population divergence, discrete changes in effective population size, admixture and sampling) and interevent periods (Cornuet *et al.* 2010).

We did not consider the Lebanese cluster in the scenarios, as all results suggested that it is not a plausible source for the French populations (see results). For the sake of simplicity, native population samples were grouped according to regions (West Turkey, East Turkey and Cyprus). To determine the invasion routes, we first considered only the French site where *M. schimitscheki* was first observed in Europe, namely Mont Ventoux (Pop3). Then, as the French populations were shown to be structured in two genetic groups (see F_{ST} and Structure results), we also conducted an ABC analysis in which the French sample was St Michel that mostly correspond to the second cluster, to determine whether this population could originate from a different source. *M. schimitscheki* displays a generation time of 2 years (Fabre *et al.* 2004), we thus assumed 0.5 generation per year in all models.

A two-step analysis procedure was followed to make the ABC analysis computationally feasible. The first step consisted in modelling the population structure in the Near East (West Turkey, East Turkey and Cyprus) assuming a common unsampled ancestral population and including the possibility of genetic admixture between them (10 competing scenarios, Fig. 2A). 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 consisted in modelling the establishment of the French invasive population (Mont Ventoux or St Michel) by simulating a new reference table taking into account the scenario that had the highest significant probability value in the first step. The same priors of the scenario parameters were used at both first and second steps, so that the posterior distributions of parameters from the first step were not used as prior in the second one. Sampling of *M. schimitscheki* populations could not be exhaustive in the native area, as sampling is not straightforward there.

Recent work suggests that if the actual source population has not been sampled and if native populations are genetically structured, studies based on molecular markers may generate incorrect conclusions about the source of an introduced population (Lombaert *et al.* 2011). Following the suggestions of these authors, we thus integrated the effects of unsampled native source populations on DIYABC inferences by building 10 competing scenarios in which the introduced population could originate from an unsampled population and allowing admixture to occur (Fig. 2B). A bottleneck in population size at introduction was included in all models.

Genetic variation within and between populations was summarized using a set of statistics conventionally used in ABC analyses (Cornuet *et al.* 2008, 2010). For microsatellites, we used the mean Garza-Williamson's M index as a one sample summary statistic, and used pairwise F_{ST} values and the mean classification index for two sample summary statistics. For both one and two sample summary statistics, we used the mean number of alleles per locus, the mean genic diversity and the mean allelic size variance. The maximum likelihood of admixture was also used. For mtDNA sequences, we used the number of distinct haplotypes, the number of segregating sites, the mean number of pairwise differences, the variance of the number of pairwise differences for one sample summary statistics, and the mean of within sample pairwise differences and the mean of between sample pairwise differences for two sample summary statistics.

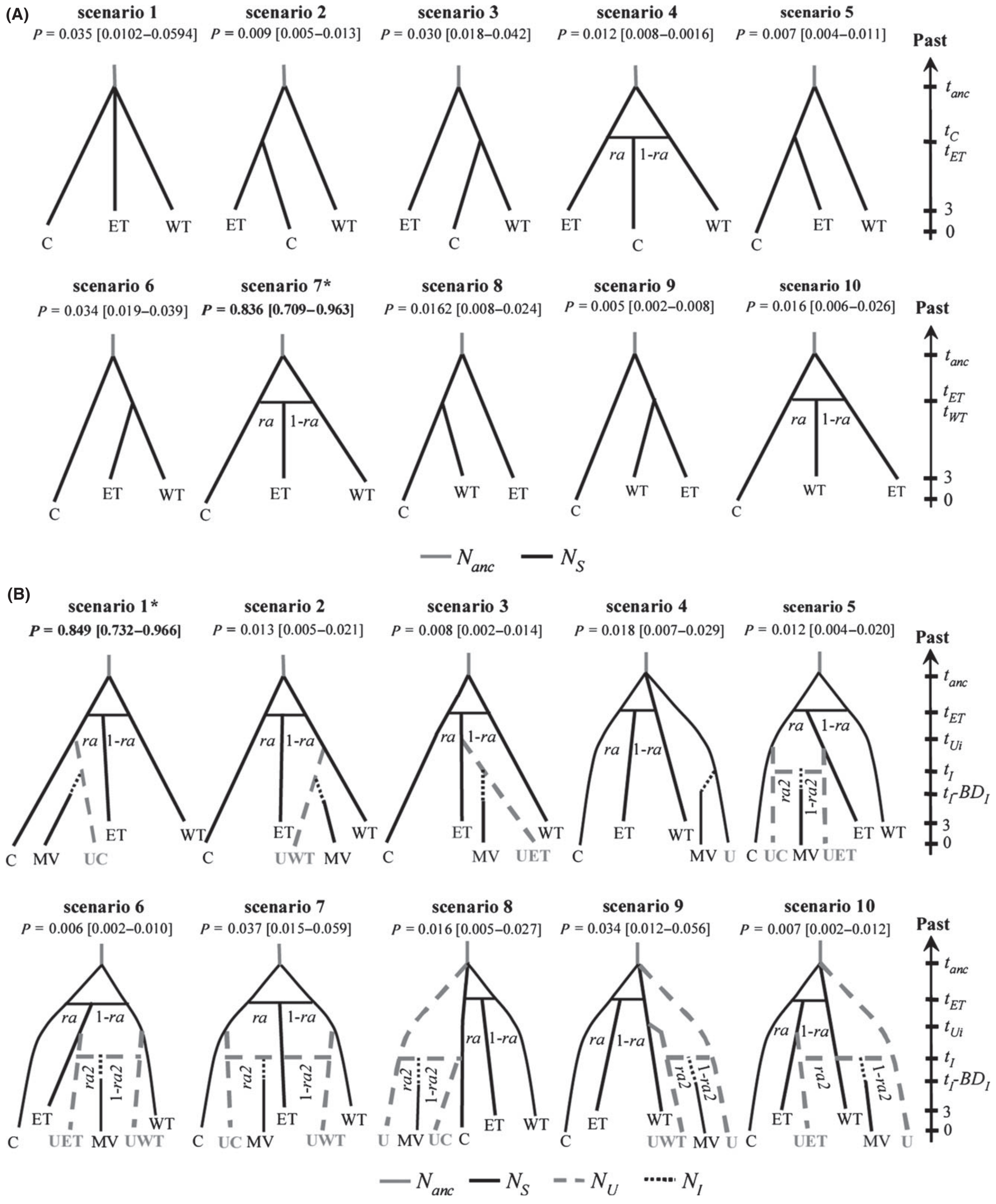
For each of the two steps of the ABC analyses, 5 000 000 genetic data sets were simulated using the coalescent approach implemented in DIYABC, providing 500 000 simulations for each scenario. Parameters of genetic data sets were drawn from their prior distributions, which were kept deliberately broad as no prior information was known. For both steps, robustness of inferences was evaluated by running two analyses with two different sets of priors (Table S1, Supporting Information), and by estimating for both sets of priors the posterior probabilities of the competing scenarios using a polychotomous logistic regression on the 1% of simulated data sets closest to the observed data (Cornuet *et al.* 2010). The selected scenario was that with the highest significant probability value with a nonoverlapping 95% confidence interval. Confidence in scenario choice was evaluated by computing type I and type II errors in the selection of scenarios. Posterior distributions of demographic parameters under the selected scenario were estimated using a local linear regression on the 5000 simulated data sets (0.1%) closest to our real data (Beaumont *et al.* 2002). Because the effective number of founders of the French invasive population (N_I) may partly depend on the duration of the bottleneck period during introduction (BD_I), bottleneck severity at introduction was assessed from posterior distributions as a composite demographic parameter expressed as $\log10(BD_I/N_I)$ (Lye *et al.* 2011). Bottleneck duration was bounded to a maximum of five generations after introduction because invasive populations of *M. schimitscheki* displayed relatively high population growth rates (Fabre *et al.* 2004; Boivin *et al.* 2008).

Results

Mitochondrial DNA data

A total of 11 haplotypes was detected (GenBank accession numbers JX516077–JX516087). No population had a statistically significant Fu's *Fs* value, but three Turkish populations had a significantly negative R2 value (Kapi-

Fig. 2 Graphic representation and posterior probabilities of the competing scenarios considered in the first (A) and second (B) steps of the approximate Bayesian computation (ABC) analysis. *The best scenario was the scenario with the highest posterior probability (*P*) and which 95% CI did not overlap with those of other competing scenarios. Times (t_{anc}, t_C, t_{ET}, t_{WT}, t_{Ui} and t_I) are translated into numbers of generations running back in time from 2003 assuming 0.5 generation per year. Time 0 is the sampling year 2003 and time 3 is the sampling year 1997. All parameters (N_{anc}, N_S, N_U, N_I, t_{anc}, t_C, t_{ET}, t_{WT}, t_{Ui}, t_I, BD_I, and *ra*) with associated prior distributions are described in Table S1 (Supporting Information). For all scenarios, populations were assumed to be isolated from each other, with no exchange of migrants. (A) Assessment of the genetic relationships of the three most likely sources of introduction in the native area. Native populations of Cyprus (C), West Turkey (WT) and East Turkey (ET) had a stable effective population size N_S and merged t_{anc} generations ago into an unsampled ancestral population of stable effective population size N_{anc}. Native C, ET or WT populations merged into another native population t_C, t_{ET} or t_{WT} generations ago (respectively), with $t_{anc} > t_C$, t_{ET} and t_{WT}. The possibility of genetic admixture at a rate *ra* between C, WT and ET populations was considered in four scenarios. (B) assessment of the source of the French invasive population. The French invasive population was founded at Mont Ventoux (MV) t_I generations ago from unsampled native populations (UC, UET and UWT). Introduction events were followed by bottleneck period of BD_I generation (s) involving a potentially small constant number of founders (N_I), which then reached a larger stable effective population size (N_S). Relationships between native populations [Cyprus (C), West Turkey (WT) and East Turkey (ET)] were consistent with the results of the first step of the ABC procedure. Unsampled native populations (UC, UET and UWT) merged into the sampled native populations (C, ET and WT, respectively) at time t_{Ui}, with $t_{Ui} \leq t_{anc}$ and t_{ET}, $t_{Ui} \geq t_I$, and i referring to as UC, UET or UWT.



dag, Konak and Sevindik; $P < 10^{-5}$). The overall haplotype (h) and nucleotide (π) diversities were estimated to 0.689 and 0.00263, respectively. Haplotype diversities within populations ranged from 0 to 0.733, 10 of the 19 sites having a haplotype diversity of 0 (all populations from Cyprus, Lebanon and France, Table 1).

The geographic distribution of the 11 haplotypes is shown on Fig. 3A, B. There were three major haplotypes, found in at least 20 individuals. The main haplotype HT1 was found in 61 individuals (all French, all Cypriot and four Turkish individuals). HT2 was the most common in Turkey (41 individuals), and HT11 was found only in Lebanon (20 individuals). Two haplotypes, namely HT3 and HT5, were found in Turkey and shared by fewer individuals. All other haplotypes were unique and found only in Turkey.

The haplotype network (Fig. 3C) had a maximum of five mutational steps between haplotypes. The three main haplotypes described earlier differed from each other by one or two mutation(s) only. The Turkish HT2 was surrounded by sets of closely related rarer haplotypes also restricted to Turkey. Note that the Lebanese haplotype was the most distant from HT1, the only haplotype found in the introduced range.

The SAMOVA was conducted on individuals of the native range. The F_{CT} values reached a plateau when K was set to 3 ($F_{CT} = 0.71465$; $P = 0.00196$). SAMOVA group 1 was located in Cyprus (Pop17), SAMOVA group 2 was located in Lebanon (Pop18 and 19) and SAMOVA group 3 corresponded to the Turkish populations (Pop9 to Pop16).

Microsatellite data

Genetic diversity and null alleles. Multilocus genotypes were obtained with high genotyping success (97.6%). Except for assignment tests that were run on all individuals, analyses were performed on the 15 populations with at least 10 individuals. The estimated genetic variability across the five loci for each population is summarized in Table 1. Allelic frequencies per population are shown in Table S2 (Supporting Information). The observed and expected heterozygosities (H_O and H_E) ranged from 0.46 to 0.82 in Near East populations, and from 0.30 to 0.54 in France.

Significant LD was found between some pairs of loci in some populations, but a given pair of loci was never in significant LD more than three times. We thus considered that the five loci were independent. The average proportion of NA for each locus was below 3%, except for MS2-162 (almost 14%). After Bonferroni correction, the HWE tests revealed that eight of the 75 populations × locus combinations showed significant departures. Two-thirds of the significant tests concerned the locus MS2-162, which may be due to the presence of NA.

Consequently, we performed the following analyses both with and without MS2-162 to achieve unbiased and robust results. As the results were similar in both cases, we only report those concerning the five loci data set.

Population structure and assignment tests. The matrix of pairwise F_{ST} obtained with the ENA correction is given in Table 2. The global measure of genetic differentiation was quite high, with $F_{ST} = 0.209$ (CI = 0.16–0.28).

The unrooted NJ tree clearly separated the introduced populations from the native ones with high bootstrap value (Fig. 4). Four subclades could be further identified within the native populations, corresponding to the main regions, that is, West Turkey (Pop9, 11 and 12), East Turkey (Pop13, 14 and 16), Cyprus (Pop17) and Lebanon (Pop18). Interestingly, the Cyprus population was closest to the introduced clade, reflecting the mitochondrial results.

We further used Structure on different data sets to infer the population genetics structure through individual assignments.

Global data set: Following the method of Evanno *et al.* (2005), the optimal number of clusters was 2. In all 25 runs, the two groups corresponded to the invasive *vs.* the native populations, with mostly well-assigned individuals. Some Cypriot individuals appeared admixed between both clusters (Fig. 5A). Near East data set: The method of Evanno *et al.* (2005) suggested that the most likely number of clusters within the native area was $K = 3$. The 25 runs raised very similar results, with the three identified clusters corresponding to the sampled countries, thus reflecting the SAMOVA mitochondrial results. As ΔK reached a maximum for $K = 3$ but formed a wide peak, with relatively high values between $K = 2$ and $K = 4$, we also explored the genetic structures corresponding to these values of K. Interestingly, for $K = 2$, one cluster was formed by the Lebanese populations while the other one grouped all Turkish and Cypriot individuals. Concerning $K = 4$, the results revealed a genetic structure corresponding to the main regions, that is, Lebanon, Cyprus, and Eastern and Western Taurus mountains in Turkey. The results for the three values of K are shown in Fig. 5B. Invasive data set: The structure results clearly suggested $K = 2$ as the most likely number of clusters in France (Fig. 5C). The results were consistent across runs, and showed one main cluster grouping populations Pop1 to Pop4 (including the region of first detection of the wasp, Mont Ventoux) and one cluster mostly corresponding to Pop5 to Pop8 (including the differentiated population of St Michel).

The assignment tests obtained with GeneClass allow ascertaining population membership of individuals. The mean $-\mathrm{Log}(L_{i\rightarrow s})$ values obtained for each French population are presented in Table 3. The group with the lowest $-\mathrm{Log}(L_{i\rightarrow s})$ values for each invasive population was always Cyprus. On the other hand, Lebanon was always the least plausible source for French populations (highest $-\mathrm{Log}(L_{i\rightarrow s})$).

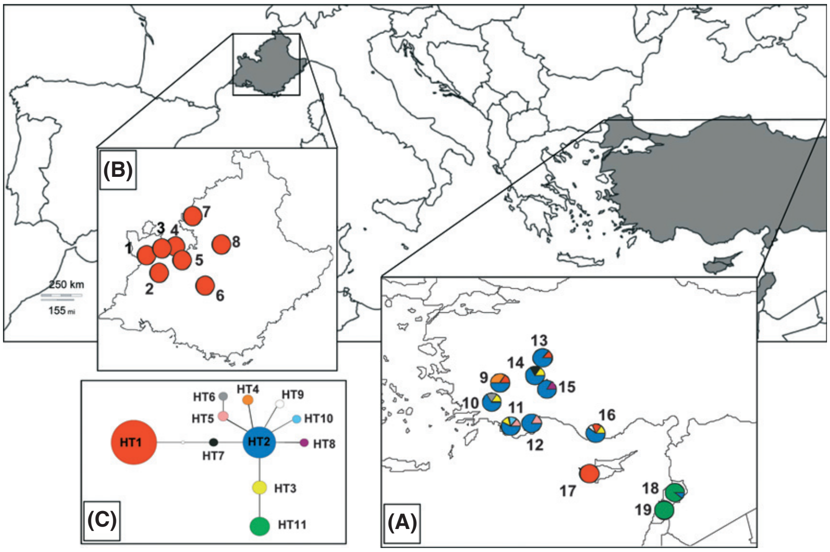


Fig. 3 Geographical distribution of *M. schimitscheki* mitochondrial haplotypes and parsimony network. (A) Native range in the Near East, (B) Introduction range in France, (C) Haplotype network of the 11 haplotypes.

Table 2 Pairwise F_{ST} divergence between 15 different geographical populations of *M. schimitscheki.* Populations no. 5, 10, 15 and 19 were excluded because of their weak sample sizes. Pairwise F_{ST} matrix was obtained using all microsatellite loci after applying the Excluding Null Alleles correction for null alleles using FreeNA. $^{NS}F_{ST}$ values nonsignificantly different from 0 ($P > 0.05$)

		Invasive range							Native range							
Populations		1	2	3	4	6	7	8	9	11	12	13	14	16	17	18
Ménerbes	1															
Vénasques	2	0.13														
Mont Ventoux	3	0.13	0.09													
Ventouret	4	0.07	0.02^{NS}	0.06												
St Michel l'Obs.	6	0.30	0.25	0.29	0.19											
St Pierre d'Argençon	7	0.16	0.11	0.18	0.05	0.24										
Sisteron	8	0.06	0.07	0.07	0.00^{NS}	0.15	0.06									
Arpaçik	9	0.35	0.27	0.31	0.25	0.30	0.30	0.28								
Sevindik	11	0.34	0.26	0.25	0.22	0.29	0.27	0.26	0.06							
Aykiriçay	12	0.31	0.24	0.27	0.22	0.27	0.26	0.25	0.00^{NS}	0.02^{NS}						
Sultandagi	13	0.30	0.23	0.23	0.20	0.29	0.25	0.23	0.11	0.02^{NS}	0.07					
Kapidag	14	0.28	0.22	0.20	0.18	0.29	0.25	0.22	0.13	0.04	0.08	0.01^{NS}				
Abanoz	16	0.27	0.22	0.23	0.19	0.27	0.24	0.22	0.10	0.04	0.07	0.04	0.02			
Paphos	17	0.20	0.16	0.11	0.14	0.32	0.24	0.17	0.22	0.17	0.19	0.13	0.11	0.14		
Tannourine	18	0.48	0.41	0.37	0.39	0.49	0.45	0.42	0.30	0.21	0.26	0.19	0.19	0.22	0.26	

Approximate Bayesian Computation analysis of introduction routes. The first step of the ABC procedure aimed at assessing the genetic relationships of the three most likely sources of introduction in the native area (West Turkey, East Turkey and Cyprus), as previous analyses allowed to exclude Lebanon as a putative source. The results unambiguously pointed to a scenario where the East Turkish group is an admixture of West Turkey and Cyprus (scenario 7, see Fig. 2A). The choice of this scenario was patently supported by high posterior probabilities using two alternative sets of priors (0.836 and 0.803) and low type I and type II errors using prior set 1 (Table 4).

The second step of the ABC treatment aimed at defining the source of the French invasive populations. When using Mont Ventoux as the tested invasive population, there was unambiguous evidence for an introduction from Cyprus without any admixture with any of the Turkish groups (scenario 1, Fig. 2B). The choice of this scenario was highly supported by strong posterior probabilities using two alternative sets of priors (0.849 and 0.794) and by noticeably low type I and type

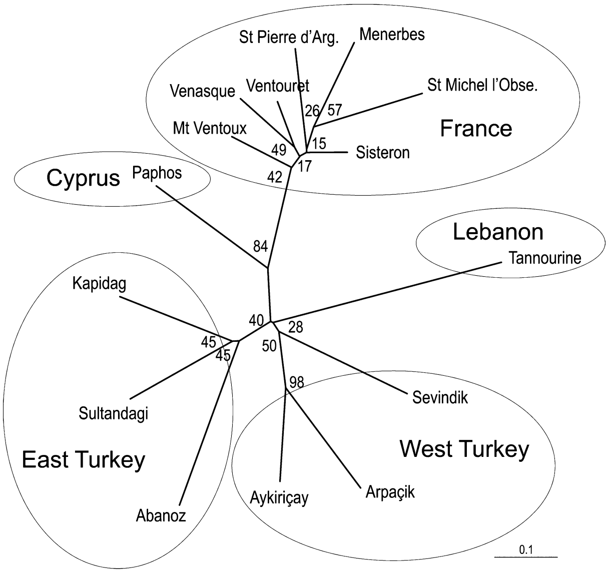


Fig. 4 Unrooted neighbour-joining tree based on Cavalli-Sforza & Edwards (1967) chord distance. Bootstrap values indicated at each node are percentage support based on 2000 bootstrap iterations.

II errors using prior set 1 (Table 4). Consistently, when using St Michel l'Observatoire as the tested invasive population, an introduction from Cyprus only was still considered as the most plausible scenario, which was supported by strong posterior probabilities using the same two alternative sets of priors (P = 0.798 [0.604–0.992], and P = 0.755 [0.632–0.921]) and low type I and type II errors using prior set 1 (0.078 and 0.109, respectively). Because mtDNA haplotype frequencies were found identical in France and Cyprus, we also tested for their possible overarching effect on ABC inferences by performing a similar step 2 ABC analysis using microsatellite data only and prior set 1. The best scenario was still an introduction from Cyprus only (P = 0.756 [0.632–0.880]; type I error = 0.081 and type II error = 0.131), which indicated consistency with our previous results and no overarching effect of mtDNA frequencies.

We used a local linear regression to estimate the posterior distributions of all parameters of the scenario selected in the second step analysis using Mont Ventoux (scenario 1, Table S1, Supporting Information). Only the posterior distributions of the effective number of founders during an introduction step (N_I), the stable effective population size (N_S) and the admixture rate (*ra*) showed distinct modes, which were different from the distribution of the prior values (Table S1 & Fig. S1, Supporting Information). This suggested that these posterior values were informative as being derived from the genetic data rather than being driven by the prior distributions (Cornuet *et al.* 2010). DIYABC estimated the median rate of admixture from Cyprus into Eastern Turkey at 0.23 [0.12–0.61]. The median stable effective population size in the native area (N_S) was estimated at 50 100 [21 100–89 900], while the number of founders of the invasive population (N_I) was estimated to at the rather low value of 9.3 [4.4–101] (Table S1, Supporting Information). Computing posterior distributions of the effective number of founders of the French invasive population (N_I) and the duration of the bottleneck period during introduction (BD_I) provided an additional strong support for a particularly severe bottleneck in population size during the establishment of the French population (Fig. 6).

Discussion

Population genetic diversity and genetic structure in the native range

Very few studies have analysed the phylogeographic history of terrestrial organisms in the Near East, especially for insects (Horn *et al.* 2006; Bodur *et al.* 2007; Simonato *et al.* 2007; Stone *et al.* 2007; Nicholls *et al.* 2010). The present study demonstrates that Near East can be the source of successful forest invasive species and provides critical knowledge of the biogeographical structure of Anatolia.

Across its native range, *M. schimitscheki* exhibited a rather high genetic diversity and a strong genetic structure. The native area consisted of three or four genetic clusters identified with microsatellites, corresponding to a geographical pattern, that is, West Turkey, East Turkey, Cyprus and Lebanon (the differentiation within Turkey being weakest based on Structure results). Interestingly, the results obtained with mtDNA were quite similar, corresponding to the three sampled countries.

First, results showed an isolation of Lebanese populations, with a major endemic haplotype, and a high level of genetic differentiation observed with microsatellites. *Cedrus libani* from Lebanon and Turkey are genetically strongly differentiated and could be regarded as separate subspecies that probably arose from one refugium in Lebanon and multiple refugia in Turkey after the last Quaternary glacial cycle (Fady *et al.* 2008). The phylogeographic structure of its associated wasp suggests very similar genetic patterns, probably resulting from historical host—parasite association (Auger-Rozenberg *et al.* 2006). Yet, the Lebanese population of *M. schimitscheki* is genetically depauperate when compared to the Turkish populations (only two mtDNA haplotypes and a relatively low allelic richness). In Lebanon, the isolated populations may thus result from relict populations that did not expand after the last glacial maximum, as observed for the circum Mediterranean bark beetle *Tom-*

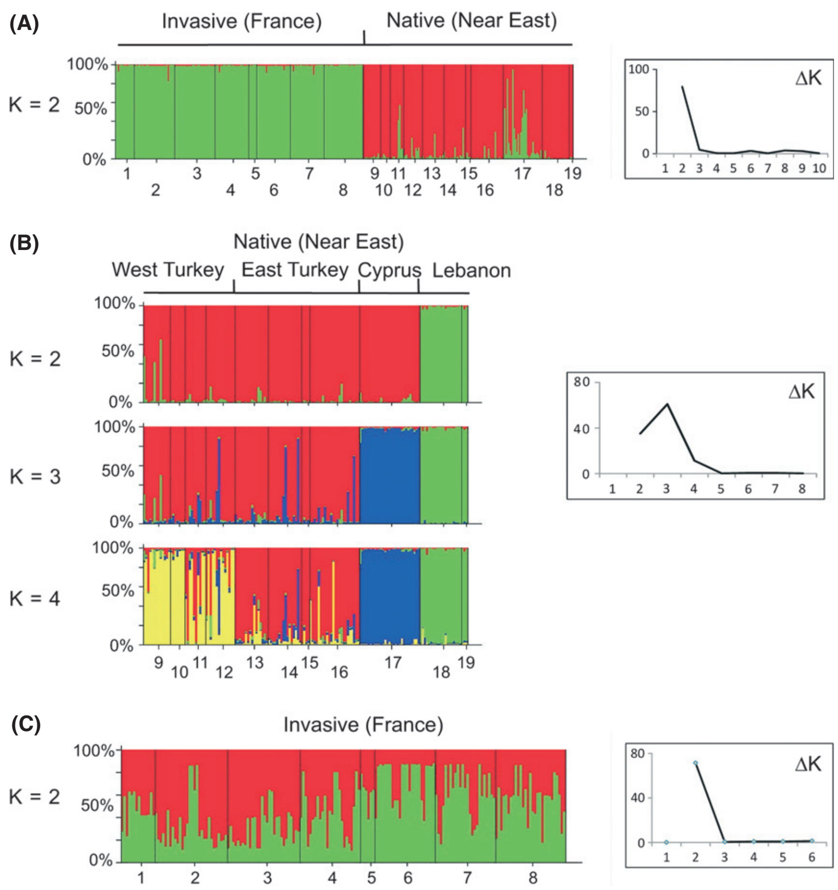


Fig. 5 Graphical representations of nuclear genetic clusters for samples from the global data set, the native and the introduction ranges, respectively, inferred from STRUCTURE simulations, with the corresponding curves of Evanno's DeltaK. Each individual is depicted by a vertical line that is partitioned into *K* coloured sections. (A) Assignment of the 340 individuals from the global data set to *K* = 2 clusters. (B) Assignment of the 156 native individuals to *K* = 2, *K* = 3 and *K* = 4 clusters; (C) Assignment of the 184 introduced French individuals to *K* = 2 clusters.

icus destruens (Horn *et al*. 2006). These populations may well have experienced genetic drift due to the geographical isolation, fragmentation and drastic reduction of its host populations from anthropogenic activities (Fady *et al*. 2008).

Table 3 Mean per population of individual assignment likelihoods for each French locality ($L_{i\to s}$ values expressed on a −log scale), to potential source populations corresponding to the four Near East groups. In bold, the sample with the lowest −Log($L_{i\ s}$) values

	Native populations: Near East groups			
Introduced populations	West Turkey	East Turkey	Cyprus	Lebanon
1-Ménerbes	10.807	9.733	**6.227**	15.534
2-Vénasques	10.613	9.460	**6.386**	15.911
3-Mont Ventoux	11.588	9.769	**6.094**	16.054
4-Ventouret	10.660	9.526	**6.580**	16.141
5-Sault	9.745	8.772	**6.792**	16.669
6-St Michel l'Observatoire	8.397	9.204	**7.410**	13.843
7-St Pierre d'Argençon	9.714	9.001	**7.508**	16.142
8-Sisteron	10.672	9.356	**6.512**	15.847

Cyprus was the second-most differentiated population. F_{ST} values between Cyprus and Turkey were high, but still lower that those between Turkey and Lebanon or Cyprus and Lebanon. The special feature of the Cypriot population of *M. schimitscheki* is that it is exclusively associated with *C. brevifolia*, an endemic Cypriot taxon that probably diverged several million years ago from *C. libani* (Qiao *et al*. 2007). *C. brevifolia* is also found in small and fragmented populations but shows a high level of diversity. Its range is currently restricted to a single forest (Paphos) with patchy nonuniform human-exploited populations. Seed dispersal and the original genetic pool can partly explain the high genetic diversity of the cedar species in this small area (Eliades *et al*. 2011). The single mtDNA haplotype can reflect, as in Lebanon, a demographic bottleneck resulting in strong genetic drift due to drastic anthropogenic impacts on host population size.

Within Turkey, mtDNA haplotypes were all closely related. The absence of deep phylogeographic isolation suggests that Turkish *M. schimitscheki* populations originate from the same refugium or the same group of refugial areas. The several haplotypes indicate a probable recent diversification centre in that region. The significantly negative values of R2 for some Turkish populations

Table 4 Confidence in scenario selection obtained from approximate Bayesian computation (ABC) analyses on Near East and French populations of the invasive *M. schimitscheki*

ABC analysis	Population structure in the Near East native area (step 1)	Foundation of the French invasive population (step 2)
Number of competing scenarios	10	10
Selected scenario*	East Turkey is an admixture of West Turkey and Cyprus	Cyprus is the source of introduction
Posterior probability [95% CI] with prior set 1†	0.836 [0.709–0.963]	0.849 [0.732–0.966]
Posterior probability [95% CI] with prior set 2†	0.803 [0.732–0.874]	0.794 [0.695–0.893]
Type I error for scenario choice‡	0.11	0.04
Type II error for scenario choice‡	0.038	0.008

*See Fig. 2 for a graphical representation of the scenarios.

†The 95% confidence intervals of the selected scenarios never overlapped those of other competing scenarios.

‡Type I and type II errors were estimated using prior set 1. 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.

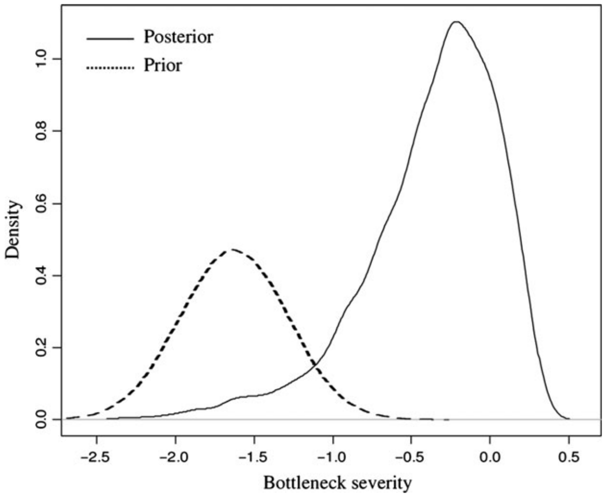


Fig. 6 Bottleneck severity during the introduction of *M. schimitscheki* in South-eastern France. Bottleneck severity (black line) was expressed as the log10 of the ratio between the duration of the bottleneck during introduction (BD_I, in number of generations) and the effective number of founders (N_I) of the invasive population to take into account the potential dependence of BD_I on N_I. Values superior to −0.5 reflect a particularly severe bottleneck. Best estimates of bottleneck severity occur where the posterior probability density function peaks.

(even if the Fu's *Fs*, better for large samples (Ramos-Onsins & Rozas 2002) did not reject the null hypothesis of constant size) and the typical star-shape topology of Turkish haplotypes suggest a genetic structure influenced by a rapid range or demographic expansion, probably also linked to the extensive national reforestation programmes in Turkey (Kaya & Raynal 2001). Microsatellite markers characterized two differentiated population groups corresponding to the Western and Eastern Taurus Mountains, when four clusters were fixed for the assignment tests. High gene flow apparently occurs at a regional scale (within mountainous ranges) as observed in another *Megastigmus* species (Nicholls *et al.* 2010) but is more restricted at relatively larger scales (between regions). This is consistent with the flight capacities of *Megastigmus* species over several dozens of kilometres (Jarry *et al.* 1997). Moreover, ABC unexpectedly pointed to a high admixture rate (23%) from Cyprus to the Eastern Taurus. This suggests recurrent unidirectional gene flow between Cyprus and the closest continental populations assisted by dominant northwards-blowing winds because the distance between Cyprus and Eastern Taurus (approximately 280 km) exceeds the active flight abilities of the wasp. Similar patterns were found in the eastern pine processionary moth *Thaumetopoea wilkinsoni* (Simonato *et al.* 2007; Kerdelhué *et al.* 2009), in which mitochondrial markers showed a strong differentiation between Cyprus and Turkey while microsatellites and AFLPs suggested relatively high levels of gene flow. The direction of male exchanges was unfortunately not estimated here.

Incompatibility between historic and genetic information

The collection and use of both historical and observational data are generally a prerequisite to understanding the invasion processes. In this study, we were in the exceptional position of having a good knowledge about the place of first introduction of the wasp in France, and about the time of introduction. Additionally, official routes of insect transport via international seed trades could be relatively well identified as authorized importations of *C. libani* seeds in France for reforestation were restricted to a limited number of South-eastern Turkish cedar stands, near the sampled Abanoz area (French Ministry of Food, Agriculture & Fishing 2010). But isolated *C. libani* and *C. brevifolia* seed imports from Lebanon and/or Cyprus (respectively) could not be formally excluded at this period, especially for clandestine ornamental purposes.

The use of both classical population genetics and ABC approaches allowed us to unambiguously identify Cyprus (rather than Turkey) as the source area of the introduced *M. schimitscheki* population, suggesting that the genetic and historical data were not compatible. The results were consistent whatever the French population considered (Mont Ventoux or St Michel l'Observatoire). Three main sources of error may negatively influence phylogeographical reconstructions of invasions (Muirhead *et al.* 2008): (i) insufficient numbers of individuals sampled in putative source populations, (ii) an incomplete list of putative source populations surveyed, and (iii) a low spatial genetic structure in the native area. In this study, we were able to counteract these three sources of errors. First, the approach developed here combined genetic information from both microsatellites and mitochondrial DNA based on representative samplings among the known distribution of *C. libani* and *C. brevifolia* (Fig. 1). Second, our ABC design ensured more accuracy in source detection as it allowed considering scenarios including putative unsampled populations in the native area, especially possible populations in Syria. Moreover, the rejection of scenarios involving admixture between two native populations supported the hypothesis of an introduction from a single source. Third, we found evidence of significant spatial genetic structure within the native area with a mean F_{ST} of 0.14 (Fig. 4). We thus provide strong evidence of the misleading potential of historical information in retracing invasion routes. Our conclusions about an unexpected scenario recall previous studies on different invasive species, for instance the western corn rootworm *Diabrotica virgifera* (Ciosi *et al.* 2008) or the harlequin ladybird *Harmonia axyridis* (Lombaert *et al.* 2010), in which invasion scenarios were found to be much more complex than previously believed. Acquiring correct information on the source of introduced individuals will help to target the critical native populations for further comparative studies aiming at exploring the ecological and evolutionary components of such an invasion success.

The Structure analysis revealed that the French populations were split into two clusters. As ABC inferences defined Cyprus as a unique source of invasion when we used reference populations from either of these two groups (namely Mont Ventoux or St Michel), these two French genetic clusters may either result from two distinct introductions from the same source, or from a genetic signature of the post-introduction expansion of the wasp. Although our ABC design did not formally test for a single versus multiple introduction(s) from Cyprus, several elements led us to favour the post-introduction scenario. First, allelic frequencies indicated that all French populations shared the same alleles, with the exception of one rare allele found at St Michel l'Observatoire (see Table S2, Supporting Information). Second, South-eastern French cedar stands were routinely surveyed since the late 70s to monitor the already established seed-wasp *M. pinsapinis* (Fabre *et al.* 1999). This allowed detection of *M. schimitscheki* for the first time in France on Mont Ventoux in 1995, following its spatial expansion towards the south-east and identifying the St Michel l'Observatoire population as a colonization front in 2002 (T. Boivin, personal observations). Third, cedar stands are of various sizes and patchily distributed across South-eastern France, and interpatch distances are frequently greater than the wasp's active flight abilities (Roques *et al.* 2008). Hence, post-introduction expansion may have involved a combination of short-distance diffusion within large and/or adjacent cedar stands and a stepping-stone pattern between distant cedar patches in the landscape. Populations that are growing in effective population size are expected in the short term to show low genetic drift, relatively stable allele frequencies, an excess of rare alleles, and homozygosity at neutral loci, but they may also show founder effects and isolation by distance along spatial expansion axes (Excoffier *et al.* 2009). This is consistent with our data, and work is currently in progress to further explore this hypothesis.

A successful invasion despite a severe founder effect and a host shift

While there was a strong historical support for a single introduction area of *M. schimitscheki* at Mont Ventoux in the early 90s, whether the invasion success of this species was subsequent to the introduction of a few or a large number of founders was still a critical question to address using molecular markers. Our genetic analyses converged to an introduction from a single source resulting in a severe founder event: (i) genetic diversity in France was drastically lower than in each of the populations of the native area, and (ii) ABC provided statistical support for a very low number of founders from one unique population source in Cyprus. We estimated that bottleneck intensity was particularly severe during the introduction of *M. schimitscheki* to France (log10 $(BD_I/N_I) > -0.5$). Other studies using the same estimation of bottleneck severity revealed either particularly severe bottlenecks as in the bumblebee *Bombus subterraneus* introduced to New Zealand (Lye *et al.* 2011), or moderate to minute bottlenecks as in *B. terrestris* (Lye *et al.* 2011) and *H. axyridis* during its worldwide invasion (Facon *et al.* 2011; Lombaert *et al.* 2011).

Invasion success is generally thought to be positively correlated to the number of founders, a concept referred to as 'propagule pressure' (Colautti *et al.* 2006). Successful invasions through multiple introductions are commonplace, partly explaining how high genetic diversity can be restored in invasive populations, especially when source populations are genetically divergent (Pascual *et al.* 2007; Lawson Handley *et al.* 2011). Yet, in agreement with the present study, successful and widespread invasions can also be founded by a surprisingly small number of individuals (<10) (Grevstad 1999; Pascual *et al.* 2007; Zayed *et al.* 2007). There is evidence for the ability of some successful invasive species to circumvent the loss of genetic variation associated with bottlenecks during introductions. In the Argentine ant (*Linepithema humile*), founder effects following invasion towards North America resulted in lower intraspecific aggression that allowed the formation of successful supercolonies (Tsutsui *et al.* 2000). In *H. axyridis*, bottlenecks of intermediate intensity could also facilitate invasions by purging deleterious mutations underlying inbreeding depression (Facon *et al.* 2011). More specifically to haplodiploid organisms, a genetic load that is hidden in heterozygous female diploids should be expressed and thus purged by selection in the haploid males (Werren 1993). Hence, haplodiploidy in *M. schimitscheki* might have limited adverse effects of inbreeding depression during the first steps of the invasion in France.

From an ecological point of view, there may be explanations of invasion success even after a severe founder effect that may alter adaptive potential. Resource availability and natural enemies are among the main drivers of the growth rate of introduced populations (Shea & Chesson 2002), and some adequate life-history traits may be a key to establishment and persistence in the new environment (Sakai *et al.* 2001). Phenotypic plasticity is one of those traits allowing invaders to cope with a variety of biotic and abiotic environmental characteristics in a new area (Lavergne & Molofsky 2007). *M. schimitscheki* shifted successfully from its native hosts *C. libani* and *C. brevifolia* to the local host *C. atlantica*, which does not occur in Cyprus but is by far the major *Cedrus* species occurring in the southern French landscape. Natural crosses between *C. atlantica* and *C. brevifolia* produce fertile hybrids (Fady *et al.* 2003), suggesting phylogenetic proximity. In France, *C. atlantica* may thus have represented a directly exploitable resource for the seed chalcid, which became extremely competitive on this new host, although they did not co-evolve. Moreover, *M. schimitscheki* exhibits both an earlier phenology and a higher realized fecundity than its closely related competitor, *M. pinsapinis*, conferring to the former a competitive advantage through enhanced seed resource pre-emption (Boivin *et al.* 2008). Another factor contributing to the success of *M. schimitscheki* may be the fact that *Cedrus atlantica* is a masting species (Krouchi *et al.* 2004), which means that seed production is highly variable from year to year (Kelly 1994). The fluctuating resource availability theory of invasion states that pulses of resources, that is, resources becoming sporadically highly abundant or scarce during variable intervals of time, are likely to increase invasion potential through fluctuations in the intensity of competition, which may limit or prevent competitive exclusion (Davis *et al.* 2000). Finally, no specialist natural enemies of *M. schimitscheki* known in the native area have been detected in South-eastern France since 1995 (T. Boivin, personal observation). The invaded French cedar stands are then likely to provide enemy escape opportunities that conferred a strong advantage to the wasp's growth rate in its new environment (Shea & Chesson 2002). Finally, arrhenotokous parthenogenesis is a partially asexual reproduction mode that may favour population establishment during the first steps of invasion (Sakai *et al.* 2001; Fabre *et al.* 2004).

Risk of invasion bridgehead in West-European and North African Mediterranean forests

Mediterranean forests deserve specific attention as they are characterized by remarkable features that make them naturally and aesthetically attractive, but also quite fragile due to rapid environmental changes (Scarascia-Mugnozza *et al.* 2000). Cedars are expected to play a fundamental role in the adaptation of Mediterranean forests to climate change as they are not only relatively resistant to drought and fire but also efficient soil conservation and erosion control species (Quézel & Médail 2003). For more than a decade, French *C. atlantica* products have become particularly attractive for both forestry and seed suppliers as reflected by the considerable exports of French seeds within and outside the European Union (French Ministry of Food, Agriculture & Fishing 2010). However, quality control and phytosanitary regulation of collected South-eastern French *C. atlantica* seeds are almost inexistent, mainly due to the need of sophisticated and expensive screening methods such as X-ray radiography to detect insects within seeds. Consequently, commercial French cedar seed lots can contain up to 20% of seeds infested by *M. schimitscheki* (T. Boivin, personal observation). Many worldwide invasions have been shown to involve a particularly successful invasive population, which serves as a source for colonists for distant new territories, a phenomenon called the invasive bridgehead effect (Lombaert *et al.* 2010). To our knowledge, South-eastern France is the only area of occurrence of *M. schimitscheki* outside of its native range. The inferred history of these populations shows that it can be a very successful invader even after a strong founder effect and a necessary shift to a different host plant. In agreement with other studies (Grevstad 1999; Frankham 2005; Zayed *et al.* 2007), we report here an interesting case study of persistence and spread of an invasive population that could be nearly independent of propagule size as occurring despite a particularly severe bottleneck at introduction. We emphasize that chance events, ecological traits of exotic species and the properties of invaded ecosystems may also be of primary importance to explain invasion success. Beyond the goal of the present paper, further work on the processes of the spatio-temporal expansion of *M. schimitscheki* in South-eastern France is now required to apprehend the magnitude and the potential ecological consequences of such invasion. Furthermore, in a context of increasing demands of South-eastern French *C. atlantica* seeds in Southern Europe or in already threatened northern African cedar stands, we should point out the increased risks of new introductions, from the invasive populations of southern France rather than from the native area of the pest.

Acknowledgements

The authors are indebted to Jean-Pierre Fabre, Alain Chalon and Nabil Nemer who supplied seed samples for analysis. We thank Solen Boivin, Emeline Lhuillier, Marie Tardy, Philippe Lorme, Patrick Pineau and Olivier Denux for technical assistance. We gratefully acknowledge Thomas Guillemaud for insightful discussion on the ABC procedure, Jérôme Rousselet and Magally Torres-Leguizamon for their help in assessing and interpreting some analyses, Bruno Fady and Jean-Noel Candau for discussion on *Cedrus* biology and insect population dynamics, and Géraldine Roux, Christian Burban, Graham Stone and two anonymous reviewers for useful comments and language revision. We also greatly acknowledge support from the French 'Agence Nationale de la Recherche' (MacBi-ANR-2010-INTB-1705), and from the French Ministry of Ecology and sustainable development (Biological invasions program 2003–2006, *INVAGRAINES*).

References

Auger-Rozenberg M-A, Roques A (2012) Seed wasp invasions promoted by unregulated seed trade affect vegetal and animal biodiversity. *Integrative Zoology*, **7**, 228–246.

Auger-Rozenberg M-A, Kerdelhué C, Magnoux E, Turgeon J, Rasplus J-Y, Roques A (2006) Molecular phylogeny and evolution of host-plant use in conifer seed chalcids in the genus *Megastigmus* (Hymenoptera: Torymidae). *Systematic Entomology*, **31**, 47–64.

Beaumont MA, Zhang WY, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France).

Bodur C, Kence M, Kence A (2007) Genetic structure of honeybee, *Apis mellifera* L. (Hymenoptera: Apidae) populations of Turkey inferred from microsatellite analysis. *Journal of Apicultural Research*, **46**, 61–67.

Boivin S, Kerdelhué C, Rozenberg MA, Cariou ML, Roques A (2003) Characterization of microsatellite loci in a seed chalcid, *Megastigmus spermotrophus* (Hymenoptera: Torymidae). *Molecular Ecology Notes*, **3**, 363–365.

Boivin T, Rouault G, Chalon A, Candau JN (2008) Differences in life history strategies between an invasive and a competing resident seed predator. *Biological Invasions*, **10**, 1013–1025.

Bou Dagher-Kharrat M, Mariette S, Lefèvre F *et al.* (2007) Geographical diversity and genetic relationships among *Cedrus* species estimated by AFLP. *Tree Genetics and Genomes*, **3**, 275–285.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics*, **19**, 233–257.

Chapuis M-P, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, **24**, 621–631.

Ciosi M, Miller NJ, 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. *Molecular Ecology*, **17**, 3614–3627.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Colautti RI, Grigorovich I, Macisaac HJ (2006) Propagule pressure: a null model for biological invasions. *Biological Invasions*, **8**, 1023–1037.

Cornuet JM, Santos F, Beaumont MA *et al.* (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics*, **24**, 2713–2719.

Cornuet JM, 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**, 11.

Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, **25**, 410–418.

Davis MA, Grime JP, Thompson K (2000) Fluctuating resources in plant communities: a general theory of invasibility. *Journal of Ecology*, **88**, 528–534.

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Serie B*, **39**, 1–38.

Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of population. *Molecular Ecology*, **11**, 2571–2581.

Eliades N-G, Gailing O, Leinemann L, Fady B, Finkeldey R (2011) High genetic diversity and significant population structure in *Cedrus brevifolia* Henry, a narrow endemic Mediterranean tree from Cyprus. *Plant Systematics and Evolution*, **294**, 185–198.

Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology*, **19**, 4113–4130.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics online*, **1**, 47–50.

Excoffier L, Foll M, Petit RJ (2009) Genetic consequences of range expansions. *Annual Review of Ecology, Evolution and Systematics*, **40**, 481–501.

Fabre JP, Mouna M, Du Merle P, Benhalima S (1999) Le point sur certains ravageurs du cèdre de l'atlas en Afrique du Nord, en France et en Europe. *Forêt Méditerranéenne*, **20**, 203–221.

Fabre J-P, Auger-Rozenberg M-A, Chalon A, Boivin S, Roques A (2004) Competition between exotic and native insects for seed resources in trees of a Mediterranean forest ecosystem. *Biological Invasions*, **6**, 11–22.

Facon B, Hufbauer RA, Tayeh A *et al.* (2011) Inbreeding depression is purged in the invasive insect *Harmonia axyridis*. *Current Biology*, **21**, 424–427.

Fady B, Lefèvre F, Reynaud M *et al.* (2003) Gene flow among different taxonomic units: evidence from nuclear and cytoplasmic markers in *Cedrus* plantation forests. *Theoretical and Applied Genetics*, **107**, 1132–1138.

Fady B, Lefèvre F, Vendramin G, Ambert A, Régnier C, Bariteau M (2008) Genetic consequences of past climate and human impact on eastern Mediterranean *Cedrus libani* forests. *Implications for their conservation. Conservation Genetics*, **9**, 85–95.

Frankham R (2005) Resolving the genetic paradox in invasive species. *Heredity*, **94**, 385.

French Ministry of Food, Agriculture and Fishing (2010) Statistiques annuelles sur les récoltes et flux de graines forestières. Note de service DGPAAT/SDFB/N2010-3037, 15 septembre 2010.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Glaubitz JG (2004) CONVERT: a user friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Molecular Ecology Notes*, **4**, 309–310.

Goudet J (2002) FSTAT, a program to estimate and test gene diversities and fixation indices Version 2.9.3.2. Available from http://www2.unil.ch/popgen/softwares/fstat.htm.

Grevstad F (1999) Experimental invasions using biological control introductions: the influence of release size on the chance of population establishment. *Biological Invasions*, **1**, 313–323.

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.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Horn A, Roux-Morabito G, Lieutier F, Kerdelhué C (2006) Phylogeographic structure and past history of the circum-Mediterranean species *Tomicus destruens* Woll. *(Coleoptera: Scolytinae)*. *Molecular Ecology*, **15**, 1603–1615.

Jarry M, Candau JN, Roques A, Icart B (1997) Impact of emigrating seed chalcid, *Megastigmus spermotrophus* Wachtl (Hymenoptera: Torymidae), on seed production in a Douglas-fir seed orchard in France and modelling of orchard invasion. *Canadian Entomologist*, **129**, 7–19.

Kalinowski ST (2005) HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. *Molecular Ecology Notes*, **5**, 187–189.

Kaya Z, Raynal DJ (2001) Biodiversity and conservation of Turkish forests. *Biological Conservation*, **97**, 131–141.

Kelly D (1994) The evolutionary ecology of mast seeding. *Trends in Ecology & Evolution*, **9**, 465–470.

Kerdelhué C, Zane L, Simonato M *et al.* (2009) Quaternary history and contemporary patterns in a currently expanding species. *BMC Evolutionary Biology*, **9**, 1–14.

Kolar CS, Lodge DM (2001) Progress in invasion biology: predicting invaders. *Trends in Ecology & Evolution*, **16**, 199–204.

Krouchi F, Derridj A, Lefèvre F (2004) Year and tree effect on reproductive organisation of *Cedrus atlantica* in a natural forest. *Forest Ecology and Management*, **197**, 181–189.

Lavergne S, Molofsky J (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences, USA*, **104**, 3883–3888.

Lawson Handley LJ, Estoup A, Evans D *et al.* (2011) Ecological genetics of invasive alien species. *BioControl*, **56**, 409–428.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Lombaert E, Guillemaud T, Cornuet JM, Malausa T, Facon B, Estoup A (2010) Bridgehead effect in the worldwide invasion of the Biocontrol harlequin ladybird. *PLoS One*, **5**, 7.

Lombaert E, Guillemaud T, Thomas CE *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*. *Molecular Ecology*, **20**, 4654–4670.

Lye GC, Lepais O, Goulson D (2011) Reconstructing demographic events from population genetic data: the introduction of bumblebees to New Zealand. *Molecular Ecology*, **20**, 2888–2900.

Miller N, Estoup A, Toepfer S *et al.* (2005) Multiple transatlantic introductions of the western corn rootworm. *Science*, **310**, 992.

Muirhead JR, Gray DK, Kelly DW, Ellis SM, Heath DD, Macisaac HJ (2008) Identifying the source of species invasions: sampling intensity vs. genetic diversity. *Molecular Ecology*, **17**, 1020–1035.

Nicholls JA, Fuentes-Utrilla P, Hayward A *et al.* (2010) Community impacts of anthropogenic disturbance: natural enemies exploit multiple routes in pursuit of invading herbivore hosts. *BMC Evolutionary Biology*, **10**, 322.

Novitzky S (1954) Beiträge zur Kenntnis der Nadelholz-Samenzerstörer. I. Zedernsamen-Chalcidier—Megastigmus schimitscheki nov. spec. *Zeitung angewante Entomologie*, **29**, 218–224.

Pascual M, Chapuis MP, Mestres F *et al.* (2007) Introduction history of *Drosophila subobscura* in the New World: a microsatellite-based survey using ABC methods. *Molecular Ecology*, **16**, 3069–3083.

Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, **95**, 536–539.

Pritchard JK, Stephens M, Donnely P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Qiao CY, Ran JH, Li Y, Wang X-Q (2007) Phylogeny and biogeography of *Cedrus* (Pinaceae) inferred from sequences of seven paternal chloroplast and maternal mitochondrial DNA regions. *Annals of Botany*, **100**, 573–580.

Quézel P, Médail F (2003) *Ecologie et biogéographie des forêts du bassin méditerranéen*. Elsevier, Paris.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences, USA*, **94**, 9197–9201.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

Roques A (2010a) Alien forest insects in a warmer world and a globalised economy: Impacts of changes in trade, tourism and climate on forest biosecurity. *New Zealand Journal of Forestry Science*, **40**(Suppl), S77–S94.

Roques A (2010b) Taxonomy, time and geographic patterns. Chapter 2. *BioRisk*, **4**, 11–26.

Roques A, Skrzypczynska M (2003) Seed-infesting chalcids of the genus *Megastigmus* Dalman, 1820 (Hymenoptera: Torymidae) native and introduced to the West Palearctic region: taxonomy, host specificity and distribution. *Journal of Natural History*, **37**, 127–238.

Roques A, Sun JH, Auger-Rozenberg M-A, Hua O (2003) Potential invasion of China by exotic pests associated with tree seeds. *Biodiversity and Conservation*, **12**, 2195–2210.

Roques L, Auger-Rozenberg M-A, Roques A (2008) Modelling the impact of an invasive insect via reaction-diffusion. *Mathematical Biosciences*, **216**, 47–55.

Sakai AK, Allendorf FW, Holt JS *et al.* (2001) The population biology of invasive species. *Annual Review of Ecology and Systematics*, **32**, 305–332.

Scarascia-Mugnozza G, Oswald H, Piussi P, Radoglou K (2000) Forests of the Mediterranean region: gaps in knowledge and research needs. *Forest Ecology and Management*, **132**, 97–109.

Scheffer SJ, Grissell EE (2003) Tracing the geographical origin of *Megastigmus transvaalensis* (Hymenoptera: Torymidae): an African wasp feeding on a South American plant in North America. *Molecular Ecology*, **12**, 415–421.

Shea K, Chesson P (2002) Community ecology theory as a framework for biological invasions. *Trends in Ecology & Evolution*, **17**, 170–176.

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.

Simonato M, Mendel Z, Kerdelhué C *et al.* (2007) Phylogeography of the pine processionary moth *Thaumetopoea wilkinsoni* in the Near East. *Molecular Ecology*, **16**, 2273–2283.

Stone G, Challis RJ, Atkinson RJ *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*. *Molecular Ecology*, **16**, 2768–2781.

Thompson JD, Higgins DG, Gibson TJ (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

Tsutsui ND, Suarez AV, Holway DA, Case TJ (2000) Reduced genetic variation and the success of an invasive species. *Proceedings of the National Academy of Sciences, USA*, **97**, 5948–5953.

Walsh PS, Metzger DA, Higuchi R (1991) Chelex-100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques*, **10**, 506–513.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the analysis of population-structure. *Evolution*, **38**, 1358–1370.

Werren JH (1993) The evolution of inbreeding in a haplodiploid organism. In: *The Natural History of Inbreeding and Outbreeding* (ed. Thornhill NW), pp. 42–94. University of Chicago Press, Chicago, Illinois.

Zayed A, Constantin SA, Packer L (2007) Successful biological invasion despite a severe genetic load. *PLoS One*, **2**, e868.

Conceived and designed the experiments: M.-A.A.R., C.K., T.B. and A.R. Performed the experiments: .M-A.A.R., E.M., and C.C. Analyzed the data: M.-A.A.R., C.K. and T.B. Wrote the paper: M.-A.A.R., C.K. and T.B. All authors read and approved the final manuscript.

Data accessibility

Microsatellite and mtDNA data: DRYAD entry doi:10.5061/dryad.tj37s. GenBank accession numbers: JX516077–JX516087.

Supporting information

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

Fig. S1 Posterior distributions of demographic parameter estimates for the invasion scenario of *M. schimitscheki* at Mont Ventoux, South-eastern France (scenario 7, ABC step 2).

Table S1 Prior distribution and posterior estimates of demographic, historic and mutation parameters used in ABC analyses of *M. schimitscheki* introduction routes.

Table S2 Allele frequencies of the 19 studied populations of *M. schimitscheki* sampled in both invasive and native areas (Pop1-Pop8 and Pop9-Pop19, respectively).