MOLECULAR ECOLOGY

Molecular Ecology (2014) doi: 10.1111/mec.12989

Complementarity of statistical treatments to reconstruct worldwide routes of invasion: the case of the Asian ladybird *Harmonia axyridis*

ERIC LOMBAERT,*† ‡ THOMAS GUILLEMAUD,*† ‡ JONATHAN LUNDGREN,§ ROBERT KOCH,¶ BENOÎT FACON,** AUDREY GREZ,† † ANTOON LOOMANS,‡ ‡ THIBAUT MALAUSA,*† ‡ OLDRICH NEDVED,§§ EMMA RHULE,¶¶ ARNSTEIN STAVERLOKK,*** TOVE STEENBERG† † † and ARNAUD ESTOUP**

*Inra, UMR 1355 ISA, Sophia-Antipolis 06903, France, †Université de Nice Sophia Antipolis, UMR ISA, Sophia-Antipolis 06903, France, ‡CNRS, UMR 7254 ISA, Sophia-Antipolis 06903, France, §USDA-ARS, North Central Agricultural Research Laboratory, Brookings, SD 57006, USA, ¶Department of Entomology, University of Minnesota, Saint Paul, MN 55108, USA, **Inra, UMR CBGP (INRA/IRD/CIRAD/Montpellier SupAgro), Montferrier-sur-Lez 34988, France, ††Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, Casilla 2, Correo 15, La Granja, Santiago, Chile, ‡‡National Reference Centre, Netherlands Food and Consumer Product Safety Authority, Wageningen 6706 EA, The Netherlands, §§University of South Bohemia, Ceske Budejovice 37005, Czech Republic, ¶¶Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK, ***Department of Terrestrial Ecology, Norwegian Institute for Nature Research, Trondheim NO-7485, Norway, †††Department of Agroecology, Aarhus University, Slagelse DK-4200, Denmark*

Abstract

Inferences about introduction histories of invasive species remain challenging because of the stochastic demographic processes involved. Approximate Bayesian computation (ABC) can help to overcome these problems, but such method requires a prior understanding of population structure over the study area, necessitating the use of alternative methods and an intense sampling design. In this study, we made inferences about the worldwide invasion history of the ladybird *Harmonia axyridis* by various population genetics statistical methods, using a large set of sampling sites distributed over most of the species' native and invaded areas. We evaluated the complementarity of the statistical methods and the consequences of using different sets of site samples for ABC inferences. We found that the *H. axyridis* invasion has involved two bridgehead invasive populations in North America, which have served as the source populations for at least six independent introductions into other continents. We also identified several situations of genetic admixture between differentiated sources. Our results highlight the importance of coupling ABC methods with more traditional statistical approaches. We found that the choice of site samples could affect the conclusions of ABC analyses comparing possible scenarios. Approaches involving independent ABC analyses on several sample sets constitute a sensible solution, complementary to standard quality controls based on the analysis of pseudo-observed data sets, to minimize erroneous conclusions. This study provides biologists without expertise in this area with detailed methodological and conceptual guidelines for making inferences about invasion routes when dealing with a large number of sampling sites and complex population genetic structures.

Keywords: approximate Bayesian computation, biological invasion, harlequin ladybird, invasion routes, microsatellite

Received 16 June 2014; revision received 18 October 2014; accepted 28 October 2014

Correspondence: Eric Lombaert, Fax: +33 4 92 38 64 01;
E-mail: lombaert@sophia.inra.fr

© 2014 John Wiley & Sons Ltd

Introduction

Inferences about the introduction routes of invasive species make it possible to describe the geographic pathways followed by propagules, between the source and invading populations. This is an important task in the exploration of fundamental eco-evolutionary aspects of colonization success or failure (Keller & Taylor 2008) and for preventing future invasions (Simberloff *et al.* 2013). Historical and observational data for invasive species are often sparse and incomplete. Indirect methods based on molecular genetic markers have therefore been increasingly used for the inference of invasion routes over the last 15 years and have proved effective (Estoup & Guillemaud 2010). One of the main features to have emerged from the growing number of studies based on molecular methods is that the real histories of invasions are often far more complex than initially imagined, particularly due to multiple introductions (e.g. Facon *et al.* 2003; Kolbe *et al.* 2004; Ciosi *et al.* 2008; Darling *et al.* 2008; Yalcindag *et al.* 2012). Multiple population sources and/or stochastic demographic and genetic processes (such as founder effects and/or genetic admixture) may lead to a genetic structure within invaded areas that is difficult to predict. Moreover, invasive populations may themselves serve as a source of colonists invading other areas (i.e. the invasive bridgehead effect, Lombaert *et al.* 2010), thus further complicating the putative invasion routes. As a result, genetic studies of introduction routes of invasive species have become a methodological and analytical challenge in themselves (reviewed in Estoup & Guillemaud 2010).

The approximate Bayesian computation (ABC) method combines the use of simulations and summary statistics (Beaumont *et al.* 2002; Beaumont 2010; Bertorelle *et al.* 2010; Csillery *et al.* 2010). Because ABC is based on simulations, it allows sampling, demographic and genetic stochasticity as well as complex processes (e.g. change in effective population size, admixture events, the involvement of a population that has not been sampled) to be explicitly taken into account. This approach is hence appropriate to make model-based inferences in a Bayesian setting for (very) complex scenarios with intractable likelihoods, such as those relating to the introduction histories of invasive species (see Box 1 for additional details). Thanks to these features and increasing computer power, ABC has been widely used in recent years, to retrace the invasion routes of a growing number of invasive species (e.g. Miller *et al.* 2005; Pascual *et al.* 2007; Lombaert *et al.* 2010; Ascunce *et al.* 2011; Auger-Rozenberg *et al.* 2012; Keller *et al.* 2012; Yalcindag *et al.* 2012; Konecny *et al.* 2013). Studies based on controlled simulated data sets have shown that, in most cases, ABC is more powerful in this context than more traditional population genetics methods, such as the interpretation of neighbour-joining tree or F-statistics analysis (Estoup & Guillemaud 2010; Guillemaud *et al.* 2010; Lombaert *et al.* 2010). However, ABC is not a strict alternative to other population genetics statistical approaches, instead being complementary to such approaches (see Box 1).

The use of the ABC method can be complex and risky, for several reasons. It may be difficult to choose the most appropriate priors and summary statistics (e.g. Templeton 2010; Robert *et al.* 2011), and this approach is also subject to what can be called the 'population sampling curse'. Sampling is a general limitation in population genetic studies. It can have a strong impact on the power of statistical analyses, which depend on sample size and the spatial sampling scheme (Muirhead *et al.* 2008; Schwartz & McKelvey 2009). More generally, the definition of a 'population' is a real issue that must be dealt with (Waples & Gaggiotti 2006; Harwood 2009). Decisions about what and how to sample are even more key issues in ABC analyses of invasion

Box 1. Approximate Bayesian computation inferences on evolutionary models when dealing with a large number of site samples and genetic units: some useful tips

Model selection using approximate Bayesian computation (ABC): principles

The idea behind model selection using ABC is to approximate the complex likelihood of models (i.e. evolutionary scenarios) by massive simulations of data sets (i.e. multilocus genotypes or individual sequences) which are summarized by statistics such as mean number of alleles or F_{ST}. Basically, the (relative) posterior probabilities of the different models are determined by measuring the similarity between the observed data set and the large number of data sets simulated under the different models. Model comparison with ABC can be divided into several steps:
1. Choice of the observed data set to be analysed.

© 2014 John Wiley & Sons Ltd

2. Definition of a finite set of competing evolutionary scenarios, of demographic, historical and genetic parameters and of their prior values.
3. Choice of summary statistics and simulation of a large number of data sets according to each evolutionary scenario, drawing parameters values from their prior distribution.
4. Selection of the simulated data sets that are most similar to the observed data set (based on summary statistics).
5. Computation of posterior probabilities of scenarios through a logistic regression step performed on the subset of simulated data sets retained in the previous step.
6. Quality control of the analysis which includes (i) some computation to evaluate the confidence in scenario choice and (ii) some model checking computation processed on the selected [scenario + parameter posterior distributions] combination to evaluate the fit with the observed data set (e.g. Cornuet *et al.* 2010).

How to deal with a data set including a large number of site samples and genetic structure?

The choice of the observed data set to be analysed is barely mentioned in studies dealing with ABC analyses. This choice is nevertheless crucial especially when the number of site samples is large and when genetic structure occurs in the studied areas. As a matter of fact, the number of genetic units in the data set will strongly impact the heaviness and sometimes feasibility of the ABC analyses (Estoup & Guillemaud 2010). Using Bayesian clustering methods such as STRUCTURE and BAPS represents a rather objective manner to define genetic units that can include several samples and hence substantially reduce the number and complexity of evolutionary scenarios to be compared using ABC. This point has been illustrated in previous studies (e.g. Ascunce *et al.* 2011; Konecny *et al.* 2013) and is central to the present study.

It is worth-stressing, however, that such clustering methods have to be used cautiously because they are based on various and sometimes different model assumptions and they are sensitive to both sampling scheme and size. The objective of inferring a clear-cut 'true' number of population units (i.e. genetic clusters) and partitioning of individuals may thus be awkward and sometimes unrealistic Waples & Gaggiotti 2006; Harwood 2009; Schwartz & McKelvey 2009; Kalinowski 2011). Moreover, each biological model has its own peculiar features, depending on many factors (e.g. migration rates, isolation-by-distance patterns, effective sizes and hierarchical genetic structure) so that a biologically meaningful application of a theoretical population concept may not be always appropriate. It is therefore difficult to propose a simple general methodological framework based for instance on a single clustering method. In practice, we strongly recommend to use several (at least two) independent clustering methods assuming different model assumptions (e.g. possibility of genetic admixture or not, nonspatial vs. spatial models) and possibly to compute such methods at different spatial scales. Results should then be cross-validated and interpreted in the light of what is already known about the species of interest.

Once the number and identity of genetic units in the full data set have been deduced, one may want to pool the site samples belonging to the same genetic unit. Alternatively, one might want to take the opportunity to use site samples as replicates of the same evolutionary history and hence perform replications of ABC analyses. This might represent a more sensible strategy than a simple pooling of sample sites for at least two reasons: (i) clustering methods may have imperfectly identify the real number and identity of the population units so that some of the inferred genetic clusters may be characterized by the presence of residual genetic structure and (ii) doing so represents an indirect way to address the issue of false positive and false negative errors.

How to deal with multiple population sources in invasion scenarios?

The reconstruction of invasion routes is a good example wherein the number of populations can limit the feasibility of ABC analyses. In some cases, even after reducing the number of population units as described above, the number of invasive populations and hence of possible scenarios remains too large to be treated in a single ABC analysis (e.g. a set of one native and six invasive populations corresponds to $6! = 720$ different scenarios, in which each invasive population is successively derived from another invasive population). Historical information, like dates of first observation of the invasive populations, must be used in such cases, to reduce the number of possible scenarios. Such historical information can also be used to define various nested subsets of competing invasion scenarios that are analysed sequentially. The first scenario considers the oldest invasive population as target and determines its invasion route. Step by step, subsequent analyses use the results obtained from the previous analyses, until the most recent invasive populations are considered (Estoup & Guillemaud 2010).

© 2014 John Wiley & Sons Ltd

routes, in which (i) population units must be defined at the outset (What is the target introduced population and what are its putative sources?) and (ii) computational power and scenario complexity limit the number of population samples that can be analysed simultaneously. Widely used population genetics approaches, such as clustering methods, are often used to define population genetic units. Different samples from the same defined genetic units are then pooled together before ABC to infer invasion routes (Box 1, e.g. Auger-Rozenberg *et al.* 2012; Boissin *et al.* 2012; Boubou *et al.* 2012; Keller *et al.* 2012; Konecny *et al.* 2013). This makes sense theoretically, but the methods used to define consistent genetic groups are known to perform poorly for determining the true number and identity of the populations (Pritchard *et al.* 2000; Evanno *et al.* 2005; Waples & Gaggiotti 2006). In practice, processing the data set of interest with several clustering methods assuming different model assumptions might be recommended (Box 1).

To take into account the risk that the sampling scheme may be incomplete (i.e. some important populations may have been missed), ABC users take advantage of the ability of the method to simulate unsampled 'ghost' populations (Beerli 2004; Slatkin 2005). However, only a few take into account the possibility of the clustering of samples being not entirely realistic and the potential qualitative impact on the results of the choice of samples used for ABC analyses (but see Pascual *et al.* 2007; Ascunce *et al.* 2011; Yalcindag *et al.* 2012).

We used a set of statistical methods (including those mentioned above) to make inferences about the invasion history of the Asian ladybird *Harmonia axyridis* from a large number of genotyped samples distributed over most of the areas invaded by this former biocontrol agent. The native area of *H. axyridis* covers a large part of Asia (Kazakhstan, southern Siberia, Mongolia, China, Korea and Japan, reviewed in Brown *et al.* 2011), with two distinct geographic East/West genetic units, as demonstrated by genetic and phenotypic analyses (Dobzhansky 1933; Blekhman *et al.* 2010; Lombaert *et al.* 2011). After a long period of use as a biocontrol agent against aphids during the 20th century, this species has recently become invasive on most continents. Invasive feral populations were first recorded in Eastern (Louisiana, USA, Chapin & Brou 1991) and Western (Oregon, USA, LaMana & Miller 1996) North America, in 1988 and 1991, respectively. Populations were then recorded in Europe (Belgium, Adriaens *et al.* 2003), South America (Argentina, Saini 2004) and Africa (South Africa, Stals 2010) in 2001. This species has spread over large areas and is still spreading in these regions, in which it has become a nuisance (Koch 2003; Roy *et al.* 2012). Lombaert *et al.* (2010) have shown that the two North American outbreaks originated from two independent introductions from the native area. They also found that *H. axyridis* from Eastern North America acted as a bridgehead population, being the source population for the European, South American and African outbreaks, and displaying admixture with a European biocontrol strain in Europe. However, these results were obtained with a limited number of site samples, particularly for the invaded areas (i.e. only one site sample per invaded area was used). This is unfortunate, because the invaded areas are large and their spatial genetic structure remains unknown.

In this study, we first aimed at providing new insight into the invasion routes and processes associated with the worldwide spread of *H. axyridis*, focusing particularly on detection of the presence of bridgehead population(s) (i.e. invasive population(s) serving as a source of colonists invading other areas) and genetic admixture between differentiated sources, two potential important drivers of colonization success. Basically, a bridgehead effect parsimoniously explains the evolution of traits conferring invasiveness in multiple populations, and genetic admixture plays a crucial role in shaping the levels of adaptive genetic variation in introduced populations (Estoup & Guillemaud 2010; Guillemaud *et al.* 2011; Rius & Darling 2014; see also Discussion section). Second, we assessed the complementarity of a number of statistical methods and the consequences of using different sets of site samples for ABC inferences. Third, we aimed to provide biologists without expertise in this field with detailed methodological and conceptual guidelines about how to make inferences about invasion routes, using recently developed Bayesian methods, when dealing with a large number of sampling sites and a complex population genetic structure (hence the extensive Supporting Information section).

Methods

Sampling and genotyping

We sampled *H. axyridis* at a total of 47 sites. Samples were collected from 33 sites in the invaded areas (five of these site samples were previously used by Lombaert *et al.* 2011): nine sites in North America (30–42 individuals per site sample), four sites in South America (30–42 individuals per site sample), 16 sites in Europe (20–42 individuals per site sample) and four sites in Africa (31 individuals per site sample). We also included in our analyses the nine samples collected in the native range (26–36 individuals per site sample) and five European biocontrol site samples (18–29 individuals per site sample) previously used by Lombaert *et al.* (2011). Complete information about the site samples is

© 2014 John Wiley & Sons Ltd

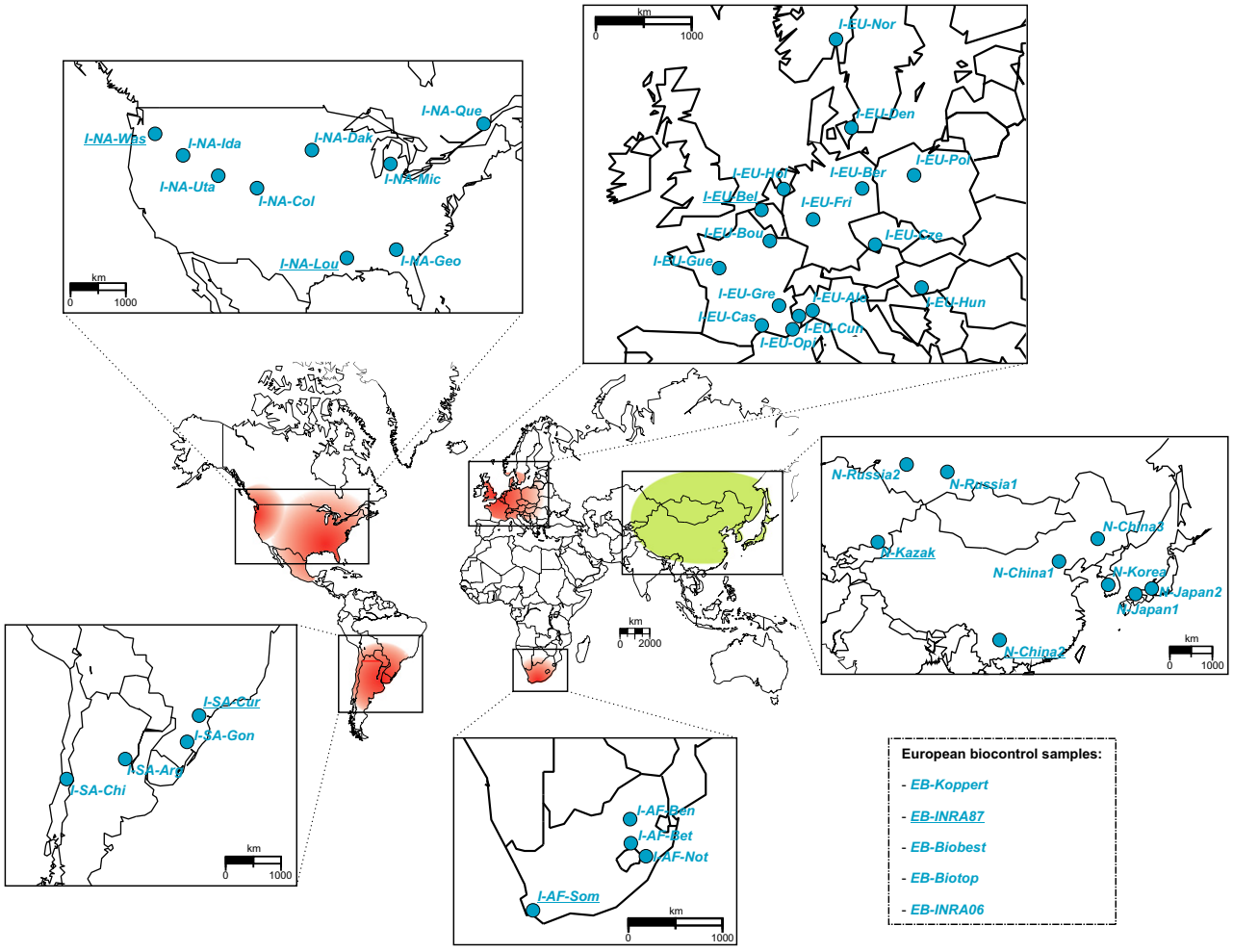


Fig. 1 Geographic locations of the genotyped site samples of *Harmonia axyridis*.
Note: In the central world map, the green and red areas roughly correspond to the native and invasive distribution of the species, respectively. Samples with underline code names correspond to the 'reference' source sample set (see Table 1) previously used in the ABC analyses processed by Lombaert *et al.* (2011). More details about each site samples can be found in Table S1 (Supporting information).

provided in Fig. 1 and Table S1 (Supporting information). Genotyping at 18 microsatellite markers was carried out for all 1442 individuals from all 47 sites as described by Loiseau *et al.* (2009).

Genetic variation within and between site samples, and tree construction

Genetic variation within site samples was quantified by calculating the mean expected heterozygosity H_e (Nei 1987) and the mean allelic richness (*AR*) corrected for 19 individuals by the rarefaction method of Leberg (2002), with FSTAT (version 2.9.3.2, Goudet 2002). Three of the five biocontrol samples were not used in this analysis because they were originally stored dry and amplification was therefore difficult for eight of the 18 microsatellite loci (Lombaert *et al.* 2011).

The level of genetic variation between site samples was measured by calculating pairwise F_{ST} estimates as described by Weir & Cockerham (1984) and carrying out exact tests for population genotypic differentiation (Raymond & Rousset 1995a) for all pairs of site samples, with Genepop (Raymond & Rousset 1995b). As these tests involved nonorthogonal and multiple comparisons, we corrected significance levels by the false discovery rate procedure (Benjamini & Hochberg 1995). We built a neighbour-joining (NJ) tree (Saitou & Nei 1987) with POPULATIONS 1.2.30 software (http://bioinformatics.org/~tryphon/populations/). We used the pairwise genetic distances between populations described by Cavalli-Sforza & Edwards (1967) because it is the most efficient distance for obtaining a correct tree topology with microsatellite markers and it makes no assumption regarding constant population size or

© 2014 John Wiley & Sons Ltd

Table 1 Description of the five different sample sets that were chosen to represent the potential source genetic units in the ABC analyses

	Different sets of samples representative of potential source genetic units				
Potential source genetic unit	'Reference'	'High-F_{ST}'	'Low-F_{ST}'	'Pool-high-F_{ST}'	'Pool-low-F_{ST}'
Eastern Asia	N-China2	N-China1	N-Japan1	N-China2 + N-China1	N-China2 + N-Japan1
Western Asia	N-Kazak	N-Russia1	N-Russia2	N-Kazak + N-Russia1	N-Kazak + N-Russia2
Eastern North America	I-NA-Lou	I-NA-Dak	I-NA-Geo	I-NA-Lou + I-NA-Dak	I-NA-Lou + I-NA-Geo
Western North America	I-NA-Was	I-NA-Ida	I-NA-Was	I-NA-Was + I-NA-Ida	I-NA-Was + I-NA-Ida
European Biocontrol	EB-INRA87	EB-Biobest	EB-Biotop	EB-INRA87 + EB-Biobest	EB-INRA87 + EB-Biotop
South America	I-SA-Cur	I-SA-Arg	I-SA-Gon	I-SA-Cur + I-SA-Arg	I-SA-Cur + I-SA-Gon
Europe	I-EU-Bel	I-EU-Cas	I-EU-Bou	I-EU-Bel + I-EU-Cas	I-EU-Bel + I-EU-Bou
Africa	I-AF-Som	I-AF-not	I-AF-Bet	I-AF-Som + I-AF-not	I-AF-Som + I-AF-bet

Sample names are as in Fig. 1 and Table S1 (Supporting information).

mutation rates among loci (Takezaki & Nei 1996). The robustness of the tree topology was evaluated by carrying out 1000 bootstrap replicates over all loci.

Definition and characterization of genetic units for ABC analyses

To define and characterize sensible genetic units for the subsequent ABC analyses, we first assessed the worldwide genetic structure of *H. axyridis*, with the widely used Bayesian clustering method implemented in STRUCTURE v2.3.3 software (Pritchard *et al.* 2000). We chose the admixture model with correlated allele frequencies, and we used the sampling location as prior information (Hubisz *et al.* 2009). We used default values for all the other parameters of the software. Each run consisted of a burn-in period of 10^5 Markov chain Monte Carlo (MCMC) iterations, followed by 10^5 MCMC iterations. We analysed the whole data set, consisting of 47 site samples (totalling 1442 individuals) from both the native and invasive ranges of the species and from the five biocontrol strains. We carried out 20 replicate runs for each prior value of the number *K* of genetic clusters, set between 1 and 15. The STRUCTURE outputs were processed with CLUMPP (Jakobsson & Rosenberg 2007), using the LargeKGreedy algorithm, with 10 000 random permutations. A similarity coefficient (G'-statistic) >90% was used to assign groups of runs to a common clustering pattern, as proposed by Wang *et al.* (2007). The run with the highest likelihood pertaining to the most frequent clustering pattern at each *K* was used for plotting with DISTRUCT (Rosenberg 2004). It is worth-stressing that we used STRUCTURE with the aim of investigating various values of *K* and the sequence of differentiation when *K* increases, rather than to determinate the most likely number of clusters (cf. STRUCTURE is not well designed to estimate the most likely number of clusters; e.g. Waples & Gaggiotti 2006). We, however, determined the highest level of genetic structure by the Δ*K* method (Evanno *et al.* 2005). Bayesian clustering methods are based on various and sometimes different model assumptions and are thus likely to yield different results (Box 1). We hence used the BAPS clustering method (Corander *et al.* 2003) as a complement to the STRUCTURE analysis. BAPS analyses were carried out on groups of individuals (i.e. site samples) rather than individuals, with more simple model assumptions (i.e. no admixture and uncorrelated allele frequencies), and at a different spatial scale (i.e. separate BAPS analyses were carried out for each continent independently). See Appendix S1 (Supporting information) for details. The cross-validation of the results obtained using both STRUCTURE and BAPS allowed us to define the final genetic units for the ABC analyses.

ABC-based inferences about invasion scenarios: overall methodology

We carried out two successive steps of ABC analyses to make inferences about the historical relationships among *H. axyridis* populations between and within the different invaded continents and the native area. It was not possible, in terms of computer capacity, to analyse all the target site samples (i.e. the site samples for which we want to know the origin) together with all the putative source site samples (Box 1). The populations considered in the ABC analyses therefore corresponded to the 15 genetic units identified by the combination of two Bayesian clustering methods (i.e. STRUCTURE and BAPS). All simulations and ABC analyses were carried out in DIYABC v2 software (Cornuet *et al.* 2014).

In all ABC analyses, the statistics used to summarize the data were those used by Lombaert *et al.* (2011).

© 2014 John Wiley & Sons Ltd

These statistics are detailed in Appendix S2 (Supporting information). The ABC analyses were performed with historical, demographic and mutational parameter values drawn from the prior distributions defined from previous studies (Brown *et al.* 2011; Lombaert *et al.* 2011) and described in Table S2 (Supporting information), and by simulating 5×10^5 microsatellite data sets for each competing scenario. For each analysis, we estimated the posterior probabilities of the competing scenarios by polychotomous logistic regression (Cornuet *et al.* 2008) on the 1% of the simulated data sets closest to the observed data set. To substantially reduce the dimension of the set of explanatory variable, we used summary statistics transformed by linear discriminant analysis (LDA; Estoup *et al.* 2012). The selected scenario was that with the highest posterior probability value. When an admixed scenario was selected, we estimated the posterior distributions of the admixture rate parameter by local linear regression on the 1% of the 5×10^5 simulated data sets closest to the observed data set (Cornuet *et al.* 2008). The robustness and relevance of our inferences were assessed with methods based on the analysis of pseudo-observed simulated data sets (Cornuet *et al.* 2010; Robert *et al.* 2011): (i) type I error rate (risk to exclude the focal scenario when it is the true one) and type II error rate (risk to select the focal scenario when it is false) were calculated, to evaluate the robustness of scenario choice, and (ii) posterior model checking was performed on the final scenario of every analysis, to evaluate the goodness of fit between the inferred evolutionary history and the real data. These analyses are detailed in Appendix S2 (Supporting information).

First set of ABC analyses: origin of target genetic units

We first used eight relatively old and established genetic units (i.e. first observation of *H. axyridis* in 2001 or before) as a putative source of each of the 10 target genetic units analysed in this first set of ABC analyses (see Results section for more details about the description of the genetic units). The potential source genetic units included the two native genetic units (West and East) identified by Lombaert *et al.* (2011), the European biocontrol genetic unit and the ones associated with each of the five invasive site samples used in the previous studies by Lombaert *et al.* (2010, 2011) (see Fig. 1). In every analysis, we only used putative source genetic units observed before or, at the latest, in the same year as the target genetic unit. Because of the large number of genetic unit, we excluded any genetic unit from the same continent as a putative source in this first set of analyses for Europe, South America and Africa. The earlier dates of first observation in North America allowed us, however, to include genetic units from the same continent in this specific area within the first set of analyses.

In all competing scenarios, an introduction event corresponds to a simple divergence event from a single or two (in case of an admixed origin) source genetic unit (s) followed by a period at low effective size (bottleneck event) predating a demographic stabilization at a higher effective size. Biocontrol genetic units are characterized by lower effective sizes compared to wild populations. No migration between any pair of genetic units was assumed. The competing scenarios only differ by the identity of the source(s) of the target introduced population. See Fig. S1 (Supporting information) for an illustration, Table S2 (Supporting information) for prior distributions and Appendix S2 (Supporting information) for additional details.

Second set of ABC analyses: relationships between target genetic units at the intracontinental scale

In the second set of ABC analyses, we defined specific competing invasion scenarios, to test the historical independence of two target genetic units from the same continent for which the same extra-continental sources were identified in the first set of ABC analyses. Our aim here was to determine whether the presence of several genetic units within a continent was due to multiple independent intercontinental introductions or to the secondary intracontinental foundation of these genetic units. See the results section and Appendix S2 (Supporting information) for details.

Consequences of using different sets of site samples for ABC inferences

We tested the impact on all our ABC inferences of using different samples to represent each of the target genetic units. We selected two site samples from each genetic unit. If at least three sites had been sampled for a given target genetic unit, we chose (i) the site sample having the lowest mean F_{ST} value with the other site samples from the same genetic unit (i.e. the 'most representative' site sample) and (ii) the one with the highest mean F_{ST} value (i.e. the 'least representative' site sample).

We also assessed the impact of using different sets of site samples within each source genetic unit. In a first sample set (referred to hereafter as the 'reference' set), we used the same site samples as in the study by Lombaert *et al.* (2011) (Table 1). In the second and third sample sets, we used the site samples displaying the highest and lowest F_{ST} values with the 'reference'

© 2014 John Wiley & Sons Ltd

sample (as defined above) from the same predefined genetic unit, respectively. The second and third sample sets are thus referred to as the 'high-F_{ST}' and 'low-F_{ST}' sets, respectively. Finally, to evaluate the effect of pooling different samples from the same genetic unit, we tested two additional sets of source samples constructed by pooling two site samples together: the first set corresponded to the pooling of the 'reference' and 'high-F_{ST}' site samples (hereafter referred to as the 'pool-high-F_{ST}' set), and the second corresponded to the pooling of the 'reference' and 'low-F_{ST}' site samples (hereafter referred to as the 'pool-low-F_{ST}' set).

Overall, because we considered five different sample sets representing the different source genetic units (Table 1) to analyse 15 target site samples (from the 10 target genetic units; see Results section and Table 2), we performed a total of 75 independent analyses in the first set of ABC analyses (see Appendix S2, Supporting information for details). Likewise, the second set consisted of 10 independent analyses (see Results section and Table 3).

Results

Genetic variation and tree construction

The complete worldwide data set, including a total of 1442 individuals from 47 site samples (Table S1, Supporting information), displayed substantial polymorphism, with a mean of 14.94 alleles per locus, over all samples. Allelic richness (and heterozygosity H_e) ranged from 2.37 alleles per locus (H_e = 0.20) in a biocontrol sample (EB-INRA06) to 6.48 (H_e = 0.56) in a native sample (N-Japan1). See Fig. S2 (Supporting information) for a concise presentation of diversity measurements for each site sample.

We found statistically significant genotypic differentiation, in 1045 of a total of 1081 pairwise comparisons between site samples (Table S3, Supporting information). For 34 of the 36 pairs of site samples displaying no significant genotypic differentiation, both samples of the pair were located on the same continent. The two remaining pairs involved the Chilean sample (I-SA-Chi), which was not significantly different from two North American samples (I-NA-Geo and I-NA-Mic). As previously described by Lombaert *et al.* (2011), there was a low level of genetic differentiation in the native area, with a mean pairwise F_{ST} of 0.013. The level of genetic differentiation between all invasive samples was moderate, with a mean F_{ST} of 0.052. Within-continent mean F_{ST} values differed substantially: 0.004, 0.023, 0.040 and 0.091 for Africa, North America, Europe and South America, respectively. European biocontrol samples displayed a high degree of differentiation from all feral (i.e. native and invasive) site samples (mean F_{ST} between biocontrol and feral samples = 0.219, Table S3, Supporting information), but the lowest F_{ST} values were systematically obtained with European invasive site samples.

In the unrooted NJ tree, the position of most of the site samples was geographically consistent (Fig. S3, Supporting information). Some distinctive and interesting patterns emerge from the tree (see legend of Fig. S3, Supporting information for details). However, the low bootstrap values at most tree nodes make it difficult to draw any robust conclusions about the relationships between the different groups observed in the tree.

Definition and characterization of genetic units for ABC analyses

STRUCTURE analyses of the worldwide *H. axyridis* data set yielded consistent results (i.e. low variance and high overall similarity coefficient) over the 20 runs tested, especially from K = 1 to K = 5. The natural logarithm of the likelihood of the data $\ln P(X|K)$ was highest for K = 14, but began to level off well before this K value was reached, and the ΔK statistic was higher for K = 2 (Fig. S4, Supporting information).

Detailed STRUCTURE clustering results (Fig. 2) gave qualitative outcomes consistent with F-statistics and NJ tree analyses and with previous knowledge about the invasion history of *H. axyridis* (Lombaert *et al.* 2010, 2011). European biocontrol samples were clearly differentiated from the Asian, American and African samples at K = 2, whereas most site samples from European invasive populations displayed admixture between the two clusters. This pattern is consistent with most European invasive populations being admixtures between European biocontrol and Eastern North America (ENA) clusters. This admixed origin was previously demonstrated with a single European sample from Belgium (I-EU-Bel, Lombaert *et al.* 2010). One notable exception is the sample from the south of France with a very high degree of ancestry from the biocontrol cluster (I-EU-Opi). The 3rd cluster (appearing at K = 3) distinguished between invasive ENA clusters from the native area and invasive Western North America (WNA) clusters and confirmed the link between the ENA outbreak and South America, Africa and Europe (e.g. Lombaert *et al.* 2010). A new cluster emerges in South America at K = 4 (except for the Chilean sample, which did not at any point differ significantly from ENA samples) and in Africa at K = 6. The 5th cluster distinguished WNA from the native area, and two admixed ENA/WNA samples could be identified (I-NA-Col and I-NA-Uta). Interestingly, the STRUCTURE pattern at K = 5 indicates a moderate to high level of WNA ancestry in Europe:

© 2014 John Wiley & Sons Ltd

© 2014 John Wiley & Sons Ltd

Table 2 Results of the first set of ABC analyses processed to make inferences about the origin of different target invasive genetic units

Continent	Target genetic unit	Target site sample	Number of scenarios	Most likely scenario (i.e. origin) when using different sets of samples representative of source genetic units (with its posterior probability)					Final selected scenario
				'Reference'	'High-F_{ST}'	'Low-F_{ST}'	'Pool-high-F_{ST}'	'Pool-low-F_{ST}'	
Africa	South Africa	I-EU-Bet	28	ENA (0.55)	ENA (0.76)	ENA (0.70)	ENA (0.68)	ENA (0.77)	ENA (5/5)
		I-EU-Som	28	ENA (0.71)	ENA (0.64)	ENA (0.89)	ENA (0.73)	ENA (0.88)	ENA (5/5)
South America	Brazil	I-SA-Cur	28	ENA (0.41)	ENA + AF (0.49)	ENA + AF (0.42)	ENA (0.45)	ENA (0.44)	ENA (3/5)
		I-SA-Gon	28	ENA (0.81)	ENA (0.40)	ENA (0.50)	ENA (0.77)	ENA (0.83)	ENA (5/5)
	Argentina	I-SA-Arg	28	ENA (0.48)	WNA + AF (0.59)	ENA (0.57)	ENA + WNA (0.41)	ENA (0.43)	ENA (3/5)
	Chile	I-SA-Chi	28	ENA (0.51)	ENA (0.41)	ENA (0.58)	ENA (0.71)	ENA (0.71)	ENA (5/5)
Europe	Western Europe	I-EU-Bel	28	EBC + ENA (0.78)	EBC + ENA (0.51)	EBC + ENA (0.72)	EBC + ENA (0.65)	EBC + ENA (0.70)	EBC + ENA (5/5)
		I-EU-Cas	28	EBC + ENA (0.54)	EBC + AF (0.63)	EBC + ENA (0.76)	EBC + ENA (0.41)	EBC + ENA (0.79)	EBC + ENA (4/5)
	South of France	I-EU-Opi	28	EBC (0.88)	EBC + WNA (0.46)	EBC (0.72)	EBC (0.98)	EBC (0.97)	EBC (4/5)
	Italy	I-EU-Cun	28	WNA (0.69)	WNA (0.78)	WNA (0.72)	WNA (0.97)	WNA (0.90)	WNA (5/5)
		I-EU-Ale	28	WNA (0.66)	WNA (0.58)	WNA (0.53)	WNA (0.75)	WNA (0.89)	WNA (5/5)
	Eastern Europe	I-EU-Hun	28	EBC + ENA (0.42)	EBC + ENA (0.44)	EBC + ENA (0.37)	EBC + SA (0.41)	EBC + ENA (0.54)	EBC + ENA (4/5)
		I-EU-Nor	28	ENat + EBC (0.24)	EBC + WNA (0.67)	EBC + WNA (0.50)	EBC + WNA (0.66)	EBC + WNA (0.55)	EBC + WNA (4/5)
North America	Colorado	I-NA-Col	10	ENA + WNA (0.36)	ENA + WNA (0.83)	ENA + WNA (0.55)	ENA + WNA (0.77)	ENA + WNA (0.81)	ENA + WNA (5/5)
	Utah	I-NA-Uta	10	ENA + WNA (0.77)	ENA + WNA (0.42)	ENA + WNA (0.84)	ENA + WNA (0.68)	ENA + WNA (0.81)	ENA + WNA (5/5)

ENat, Eastern native area; ENA, Eastern North America; WNA, Western North America; SA, South America; AF, Africa; EBC, European Biocontrol.
The most likely scenario [i.e. the source genetic unit(s)] with its posterior probability between brackets is given for each target invasive site sample (names as in Fig. 1 and Table S1). For each target site sample, the last column indicates the scenario that was found most frequently among the five ABC analyses (the frequency is given between brackets). More detailed results are provided in Table S4.

Table 3 Results of the second set of ABC analyses processed to make inferences about the intracontinental relationships between a subset of target invasive genetic unit

Continent (target genetic units)	Number of competing scenarios	Source sample set	Selected scenario	Posterior probability of selected scenario with 95% credibility intervals between brackets
South America (Brazil/ Argentina/Chile)	5	'Reference'	Two independent ENA introductions: [Chile] + [Brazil, Argentina]	0.747 [0.733, 0.762]
		'High-F_{ST}'	Two independent ENA introductions: [Chile] + [Brazil, Argentina]	0.788 [0.778, 0.798]
		'Low-F_{ST}'	Two independent ENA introductions: [Chile] + [Brazil, Argentina]	0.692 [0.676, 0.708]
		'Pool-high-F_{ST}'	Two independent ENA introductions: [Chile] + [Brazil, Argentina]	0.877 [0.867, 0.886]
		'Pool-low-F_{ST}'	Two independent ENA introductions: [Chile] + [Brazil, Argentina]	0.674 [0.661, 0.688]
Europe (East Europe/West Europe/Italy)	4	'Reference'	Two independent WNA introductions: [East Europe] + [Italy]/One ENA introduction: [West Europe]	0.521 [0.510, 0.532]
		'High-F_{ST}'	Two independent WNA introductions: [East Europe] + [Italy]/One ENA introduction: [West Europe]	0.737 [0.724, 0.749]
		'Low-F_{ST}'	Two independent WNA introductions: [East Europe] + [Italy]/One ENA introduction: [West Europe]	0.508 [0.498, 0.518]
		'Pool-high-F_{ST}'	Two independent WNA introductions: [East Europe] + [Italy]/One ENA introduction: [West Europe]	0.524 [0.507, 0.542]
		'Pool-low-F_{ST}'	Two independent WNA introductions: [East Europe] + [Italy]/One ENA introduction: [West Europe]	0.646 [0.634, 0.657]

ENA, Eastern North America; WNA, Western North America. See main text (Methods sections) and Table 1 for details regarding the different source sample sets.

high in Italy and moderate in several of the most Eastern site samples. The genetic structuring of the native area into two clusters (previously identified by Lombaert *et al.* 2011) appeared at $K = 8$. Finally, a last cluster corresponding to the invasive Italian site samples appeared at $K = 9$.

The results of the intracontinental BAPS clustering analyses were similar to those of the STRUCTURE analyses (Appendix S1, Supporting information). The main, but slight, differences were the clearer East/West geographic pattern in Europe with BAPS and the identification as genetic unit of the I-NA-Col site sample in North America with STRUCTURE. Capitalizing on the above STRUCTURE and BAPS results, we grouped the 47 site samples into 15 genetic units for ABC analyses: two in Asia, four in North America, three in South America, one in South Africa, four in Europe and one European biocontrol (see Fig. 2). We then defined the 8 source and 10 target genetic units as well as the various sample sets used in the ABC analyses (see Table 1, Table 2 and Appendix S2, Supporting information).

First set of ABC analyses: origin of target genetic units

The results of this first set of ABC analyses are summarized in Table 2 (see more detailed results in Table S4, Supporting information). We provide here the posterior probabilities of competing scenarios estimated on the 1% of the simulated data sets closest to the observed data set, but results using another threshold of 0.5% were qualitatively identical (results not shown). Our evaluation of confidence in scenario choice for three ABC analyses (Appendix S2, Supporting information) revealed moderately high type I errors (0.39, 0.20 and 0.13, Table S5, Supporting information), but markedly low mean type II errors (0.035, 0.051 and 0.014, Table S5, Supporting information), suggesting that our model choice analyses were reliable overall. To check the adequacy of the chosen evolutionary scenarios and associated posterior distributions of parameters to the data, we performed model checking analyses. They indicated that the data simulated under the chosen model and

© 2014 John Wiley & Sons Ltd

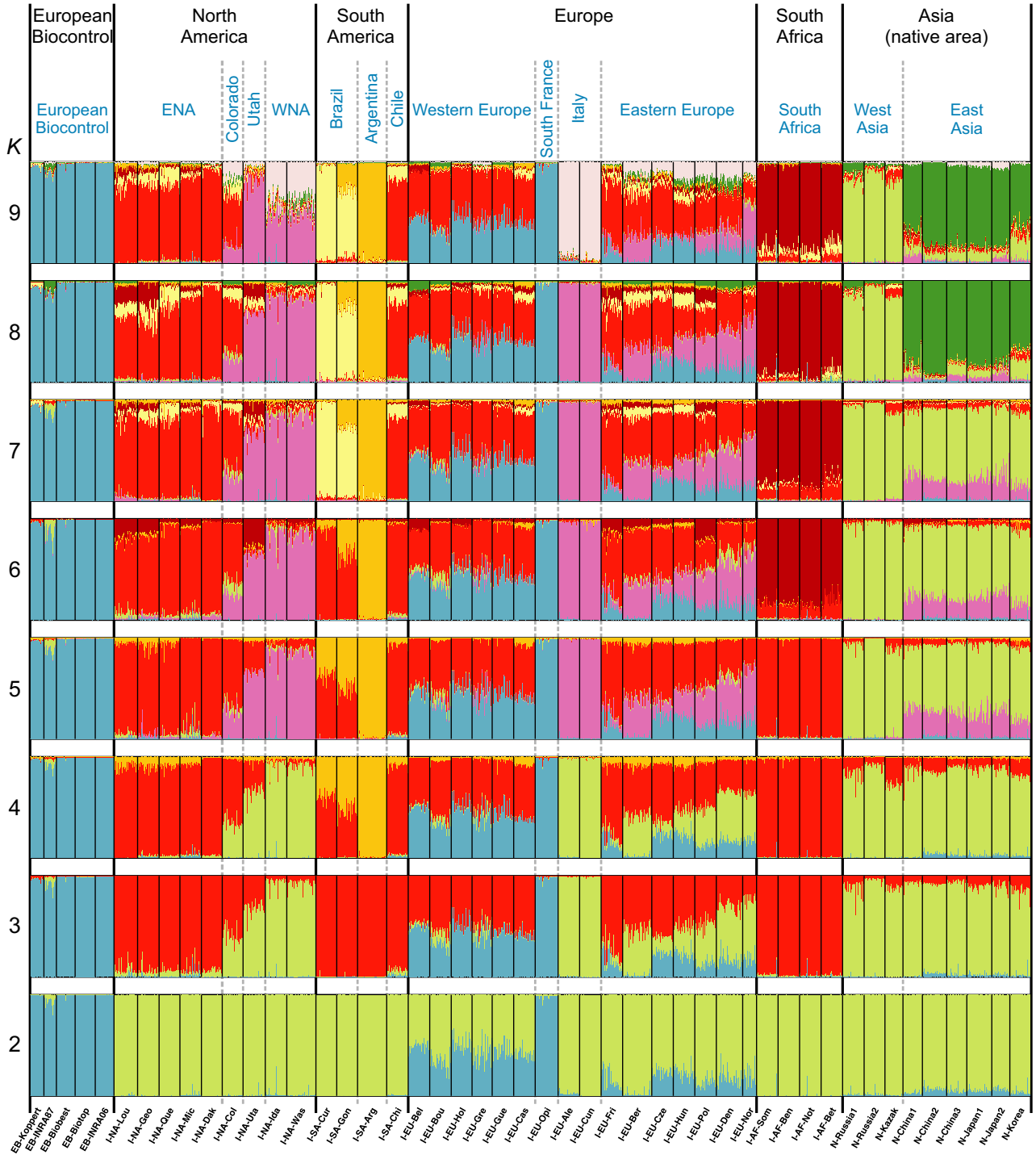


Fig. 2 Genetic clustering of the site samples of *Harmonia axyridis*.
Note: Ancestry estimation based on the Bayesian individual clustering method implemented in STRUCTURE is given for K = 2–9 genetic clusters (left). Each vertical line represents an individual and each colour represents a genetic cluster. Individuals are grouped by site sample (names at the bottom of the figure are as in Fig. 1 and Table S1, Supporting information). The black text at the top of the figure groups the samples by continent. The blue text at the top of the figure represents the fifteen final genetic units defined from both STRUCTURE and BAPS (see Appendix S1, Supporting information) analyses and thereafter used in the ABC analyses.

© 2014 John Wiley & Sons Ltd

posteriors fitted well the observed genetic data for *H. axyridis* (Table S6, Supporting information).

All ABC analyses for the African target genetic unit gave similar results, regardless of the samples used as the target or source populations: the ENA genetic unit was unambiguously the source of the African genetic unit (Table 2). The results were less clear-cut for the three South American target genetic units. However, an ENA origin was considered to be the most likely scenario overall for each target genetic unit (i.e. Chile, Brazil and Argentina; Table 2). We obtained a small number of discrepant results for the I-SA-Cur and I-SA-Arg target samples, probably due to the high type I error risks associated with other scenarios. For example, we considered the Brazilian sample I-SA-Cur to have probably originated from the ENA genetic unit, despite two of the five ABC analyses indicating possible admixture between the ENA and African genetic units. This admixture scenario is genealogically similar to the simpler scenario of an ENA origin, because the African genetic unit itself originates from the ENA genetic unit (Table 2). It is therefore not surprising that the observed posterior probabilities were rather low (below 0.5) and similar for the three most likely scenarios in this case (Table S4, Supporting information). Moreover, analyses of the I-SA-Gon sample, from the same Brazilian genetic unit, systematically suggested an ENA origin (Table 2). Similar arguments led us to draw the same conclusion (i.e. an ENA origin) for the Argentinian genetic unit represented by the I-SA-Arg sample.

Three of the four European target genetic units gave clear-cut ABC results (Table 2) consistent with NJ and STRUCTURE analyses (Fig. S3, Supporting information and Fig. 2). We found that the WNA genetic unit was the source of the Italian target genetic unit (I-EU-Ale and I-EU-Cun samples), whereas the south of France genetic unit (I-EU-Opi) was probably directly derived from the European biocontrol strain. Finally, the two samples used to determine the origin of the Western European genetic unit (I-EU-Bel and I-EU-Cas) clearly suggested a scenario involving an admixture between the ENA genetic unit and the European biocontrol strain. By contrast, the ABC analyses of the two samples representing the Eastern European genetic unit (I-EU-Hun and I-EU-Nor) gave different results. Depending on the target sample and the putative source sample set, various admixture scenarios (involving the Eastern native, ENA, WNA, South American or European biocontrol genetic units) were selected, often with low to moderate posterior probabilities (Table 2). This makes it difficult to draw any firm conclusions about this specific invasive genetic unit at this point. The possibility of admixture between three different sources (ENA, WNA and European Biocontrol) suggested by the STRUCTURE analyses (e.g. $K = 7$; Fig. 2) was not tested formally in these ABC analyses.

Regarding the two North American target genetic units (i.e. the I-NA-Col and I-NA-Uta samples), we obtained the highest posterior probability for an admixture scenario, with the ENA and WNA genetic units as parental sources, regardless of the set of site samples used as putative sources. This finding is consistent with the observed NJ tree topology (Fig. S3, Supporting information) and STRUCTURE results (Fig. 2). ABC estimates of the genetic contribution (i.e. admixture rate) of the ENA parental source were also consistent with geographic patterns (Fig. 1), as we obtained a higher rate for the more eastern Colorado sample I-NA-Col (mean of 71% over the five ABC analyses) than for the more western Utah sample I-NA-Uta (mean of 36% over the five ABC analyses), with almost nonoverlapping distributions of admixture rates between the two site samples.

Second set of ABC analyses: relationships between target genetic units at the intracontinental scale

As mentioned above, the invasion histories of South America and Eastern Europe remained unclear. The first set of ABC analyses indicated that all three South American target genetic units had the ENA genetic unit as their source. We investigated whether these target genetic units were established independently from the ENA source genetic unit (i.e. two or three ENA introductions into South America) or whether their establishment was not independent (i.e. a single ENA introduction responsible for all the South American genetic units), by defining a second set of ABC analyses including five competing scenarios (see Appendix S2, Supporting information for details). Whatever the set of source samples used, we found that the Chilean genetic unit was independent of the other two South American genetic units (Argentina and Brazil). By contrast, the South American genetic units from Argentina and Brazil both originated from the same introduction event from the ENA source genetic unit (Table 3). The low type I and II error rates indicated that this result was robust (Table S5, Supporting information) and there was a good fit between the model and the observed data (see model checking analyses summarized in Table S7, Supporting information). This scenario is consistent with the observed NJ tree topology (Fig. S3, Supporting information) and with STRUCTURE results (e.g. Argentina and Brazil appeared to be associated with the same private cluster at $K = 4$; Fig. 2). Thus, the *H. axyridis* populations in South America originated from two independent introductions from Eastern North America.

© 2014 John Wiley & Sons Ltd

The first set of ABC analyses and the STRUCTURE analyses indicated that several genetic units within Europe may have originated from similar sources. In particular, the Italian and Eastern European genetic units both have a WNA origin (at least partially for the Eastern European genetic units), whereas the Western and Eastern European genetic units both originate at least partly from the ENA and biocontrol genetic units. We defined a second set of ABC analyses with four competing scenarios, corresponding to all the possible combinations of one single or two independent introductions from ENA and WNA (see Appendix S2, Supporting information for details). Whatever the set of putative source samples used, we found that a single introduction from ENA and two independent introductions from WNA were involved in Europe (Table 3). However, despite a good fit between the selected scenario and the data (model checking analyses, Table S7, Supporting information), type I and type II error rates were found to be high (0.400 and 0.136, respectively; Table S5, Supporting information), suggesting that these results should be interpreted with caution.

Discussion

New insights into the worldwide population structure and invasion routes of H. axyridis

Previous population genetic studies of the highly invasive species *H. axyridis* have focused on a limited number of site samples. In their study, Lombaert *et al.* (2011) used a total of 19 site samples, only five of which were sampled from invaded areas. Here, we have added another 28 invasive site samples, increasing the number of individuals from 561 to 1442. Through this major additional sampling effort and the use of different methodological tools, we provide new insight into the worldwide spatial structure and invasion routes of *H. axyridis* populations (summarized in Fig. 3).

North America was the first continent to be invaded, with two independent introductions from the Asian native area (Lombaert *et al.* 2010), one in the East, first observed in 1988 (ENA outbreak), and one in the West, first observed in 1991 (WNA outbreak). *H. axyridis* has since spread throughout North America and is now present in almost all states and jurisdictions of the USA, Canada and Mexico (Koch *et al.* 2006; Brown *et al.* 2011). Our survey of the genetic structure of North American *H. axyridis* populations is highly consistent with a spatial expansion of these two initial invasive populations, with no additional introduction event. A contact zone between both outbreaks with substantial genetic admixture was identified in Utah and Colorado. This result is consistent with historical spatial establishment data for this species in North America (Koch *et al.* 2006) and raises new unresolved questions about the evolutionary and practical consequences of such genetic admixture between two already successful invasive populations.

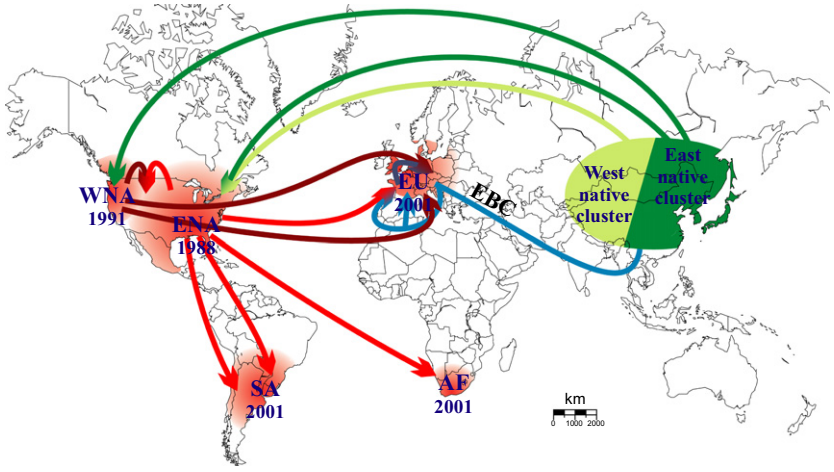


Fig. 3 Worldwide invasion scenario of *Harmonia axyridis*.
Note: Most likely scenario of invasions of *H. axyridis* into Eastern North America (ENA), Western North America (WNA), South America (SA), Europe (EU) and Africa (AF) deduced from a set of complementary population genetics analyses. For each outbreak, the arrow indicates the most likely invasion pathway. Bicolour arrows indicate admixture events. Years of first observation of invasive populations are indicated for each continent. Initially collected from the native area in 1982, the European biocontrol strain (EBC) is represented by a blue arrow. The ranges of the two native genetic units are roughly drawn and coloured in clear and dark green. The invasion routes of the most early WNA and ENA outbreaks were inferred by Lombaert *et al.* (2011).

© 2014 John Wiley & Sons Ltd

The previously identified predominant role of North America in the worldwide invasion was confirmed by our results. We found that the Eastern North America (ENA) outbreak was the probable source of at least four introductions into other continents. One ENA introduction was responsible for the African outbreak. In South America, ladybirds from the ENA outbreak have independently founded at least two invasive populations. The population found in Brazil/Argentina has been described before (Lombaert *et al.* 2010), but the Chilean invasive population was found to have originated from a second independent introduction from ENA never before described. The previously described introduction of ENA propagules into the Western part of Europe (Belgium, France, Holland) followed by admixture with the European biocontrol strain (Lombaert *et al.* 2010) was confirmed by our analyses. We found that the Western North America (WNA) population was the source of at least two independent introductions in Europe. One of these introductions led to the establishment of a feral population in northern Italy, whereas the second probably resulted in admixture with the Western European invasive population in the more recently invaded eastern and northern parts of the continent (Germany, Poland, Czech Republic, Hungary, Denmark, Norway). The inferred history of this species in Eastern Europe should be considered with caution, as indicated by the high type II error rates (Table S5, Supporting information) and the greater heterogeneity of STRUCTURE Q-plots than for all other target genetic units (Fig. 2). Additional introduction events may have been involved in this area. Further analyses with additional competing scenarios, samples and/or markers are required for a finer assessment of the invasion history of *H. axyridis* in Eastern Europe.

We identified a single population in southeast of France that appeared to have originated exclusively from the European biocontrol (EBC) strain introduced into Europe in 1982. At first sight, this is surprising, because EBC individuals have long been thought to be unable to survive in the wild (Ferran *et al.* 1997; Turgeon *et al.* 2011). This feral population was first observed in 2005 in the town of Opio (France), in an area in which many attempts at introduction for biocontrol purposes occurred in the 1990s. *H. axyridis* has since repeatedly been observed in this area, in which it appears to have established a feral population. This population does not seem to have expanded spatially, unlike most of the other known *H. axyridis* outbreaks. This locally established population attests to the ability of the European biocontrol strain to found small overwintering populations in the wild, in an area with clement winters. It may also account for this species occasionally being recorded in France for brief periods during the 1990s, before the introduction and expansion of the highly invasive Eastern North America population (Brown *et al.* 2008).

Two invasive bridgehead populations and several genetic admixture situations

The identification of bridgehead populations is of prime importance in the understanding of the evolutionary processes at work during a successful invasion. As a matter of fact, the invasive bridgehead scenario is evolutionary parsimonious (Estoup & Guillemaud 2010; Guillemaud *et al.* 2011): a single evolutionary shift in a single introduced population (the bridgehead) is required, whereas multiple changes are required in the case of multiple introduced populations, which must independently evolve traits conferring invasiveness. From an evolutionary biology perspective, the invasive bridgehead scenario is thus fundamentally different from scenarios in which the invasive populations originate directly from the native area, and this will have important consequences for the understanding of the whole story as well as for the prediction of the future distribution of the species. Our results highlight the complexity of the worldwide invasion history of *H. axyridis* with the identification of two independent bridgehead invasive populations which have been the source of at least six successful secondary introductions on three other continents. The findings reported here confirm the previously described bridgehead status of the ENA *H. axyridis* population (Lombaert *et al.* 2010), but the role of the WNA population as a second bridgehead population was previously unknown. After several attempts at the acclimation of *H. axyridis* for biocontrol purposes during the 20th century (Krafsur *et al.* 1997), the WNA and ENA outbreaks correspond to the only two populations known to have been established from the native area and to have spread. The eco-evolutionary characteristics of the ENA outbreak have been studied in detail (e.g. Labrie *et al.* 2006; Facon *et al.* 2011; Turgeon *et al.* 2011; Tayeh *et al.* 2012, 2013), whereas those of the WNA outbreak have yet to be investigated. Comparisons of the life history traits of the two North American bridgehead populations with those of noninvasive populations (native or biocontrol populations) might prove a fruitful source of knowledge about the respective role of adaptation and chance in invasion success. It would be particularly interesting to determine whether the deleterious alleles that contribute to inbreeding depression have been purged in the WNA population, as in the ENA population (Facon *et al.* 2011).

Genetic admixture events between differentiated sources are known to play a crucial role in shaping the levels of genetic variation in introduced populations.

© 2014 John Wiley & Sons Ltd

They can produce new recombinant genotypes, counterbalance bottlenecks and promote high levels of genetic diversity (e.g. Kolbe *et al.* 2004; Bossdorf *et al.* 2005). They may also directly increase or decrease the mean fitness of individuals in a population, depending on the importance of both genetic load and local adaptations (Lynch 1991; Edmands 1999; Rius & Darling 2014). Admixture has therefore been identified as one of the key factors underlying invasion success, through its effects on the process of adaptation following establishment (Wares *et al.* 2005; Facon *et al.* 2006; Keller & Taylor 2008; Rius & Darling 2014). In this study, we identified several situations of genetic admixture in North America and Europe (Fig. 3). Besides, the ENA outbreak was previously shown to have probably originated from admixture between the two native genetic units (Lombaert *et al.* 2011). Other admixture events have probably already occurred or will occur in the near future, due to the presence of several expanding independent outbreaks on several continents, such as South America and Europe. It has been shown that the ENA propagules introduced into Western Europe may have benefited from admixture with the European biocontrol strain (Turgeon *et al.* 2011). On the other hand, no heterosis or outbreeding depression has been observed in laboratory crosses between individuals from various invasive *H. axyridis* populations (Tayeh *et al.* 2013). However, the full consequences of admixture in the wild remain to be explored, and the involvement of the WNA population in several admixture situations in North America and Europe has not yet been investigated.

Various methods providing congruent and complementary results

Despite the impressive complexity of the inferred invasion routes, most of the population genetics methods that we used provided congruent results. Basic measurements and representations of genetic variation within and between samples (e.g. allelic richness, F_{ST}, NJ tree) and the use of Bayesian clustering methods implemented in STRUCTURE or BAPS software provided a first set of meaningful qualitative insights into the invasion routes of *H. axyridis*. Clustering methods also made it possible to classify samples into sensible genetic units. The subsequent use of ABC methods then made it possible to carry out rigorous tests for evolutionary relationships between these predefined genetic units. We argue that the combined use of several methods, as in this study, is crucial, given the known difficulties in making robust inferences about complex genetic structure from data for a large set of site samples (e.g. Evanno *et al.* 2005; Waples & Gaggiotti 2006; Kalinowski 2011).

ABC is a powerful method, but it can give questionable results if misused. It is important therefore to check the quality of the results obtained thoroughly (Bertorelle *et al.* 2010; Cornuet *et al.* 2010; Robert *et al.* 2011). Our estimates of type II and to a lesser extent type I error rates in several ABC analyses suggest that, overall, we had sufficient power to discriminate between the sets of scenarios studied. Interestingly, we found that, despite the robustness of our first set of ABC analyses, the choice of data sets for the analysis did, in some cases, affect the selection of the most likely scenario. This was particularly true for analyses of the origin of the Eastern European genetic unit, and to a lower extent for analyses focusing on the population genetic units of Brazil, Argentina, Western Europe and the south of France. However, in all the replicate analyses carried out, a single scenario predominated, and this scenario was, in each case, consistent with previous population genetics analyses. Closer analysis of the cases in which the 'wrong scenario' was selected indicated that type I errors may have been responsible for these incorrect selections. For example, when the ENA-origin scenario was simulated for the analysis of the Brazilian site sample I-SA-Cur, most of the type I errors (the overall type I error was 0.39, Table S5, Supporting information) were associated with a scenario identifying the source as the African genetic unit (the type I error associated with this scenario was 0.19) or an admixture between the ENA and African genetic units (the type I error associated with this scenario was 0.07). Consistent with these findings, these two scenarios were those systematically presenting the highest posterior probabilities, together with the ENA-origin scenario, in analyses of real data sets (Table S4, Supporting information), leading to erroneous conclusions in some cases. We hence show here that the choice of site samples can have non-negligible consequences for the conclusions drawn from a single ABC analysis. The common practice of pooling-differentiated site samples may also give misleading results, as illustrated by the erroneous result obtained with the 'pool-high-F_{ST}' sample set for analysis of the invasion history of the Argentinian sample I-SA-Arg (Table 2). We argue that, besides standard quality controls based on the analysis of pseudo-observed data sets, the use of different sample sets and comparisons of the results obtained represent a sensible solution to minimize the misinterpretation of ABC analyses.

Conclusion

This study shows that multiple invasion bridgeheads, introductions and admixtures have served as corner-

© 2014 John Wiley & Sons Ltd

stone events in the invasion history of *H. axyridis*. This further strengthens the idea that such processes might be of general importance in invasion successes. The actual invasion history of *H. axyridis* may be even more complex than that inferred here. Although intensive, our sampling scheme remains incomplete, and we therefore cannot exclude the possibility of other as yet unelucidated introduction events. Overall, our findings confirm that accidental introductions have probably played a major role in the current distribution of *H. axyridis* (Koch 2003; Brown *et al.* 2011), as the invasion history described here does not correspond to knowledge of the history of biocontrol attempts. Our findings also show that this species is a very good model for exploring the role of multiple introduction and admixture events in the evolutionary potential of invasive species and the potential occurrence of evolutionary and/or environmental shifts in bridgehead populations (Wares *et al.* 2005; Facon *et al.* 2006). Finally, from a broader methodological perspective, this study highlights the importance of coupling ABC methods with more traditional approaches and suggests that carrying out independent ABC analyses on several sample sets, when possible, constitutes a complementary approach to standard quality controls based on the analysis of pseudo-observed data sets, to minimize erroneous model choices and conclusions.

Acknowledgements

We thank J.-M. Cornuet, A. Dehne-Garcia, M. Kenis, L. J. Lawson Handley, R. Ware and I. A. Zakharov for fruitful discussions. We also thank our colleagues N. Berkvens, C. Borges, P. Brown, A. Coeur d'Acier, P. De Clercq, I. Goryacheva, L. Hautier, W. Jones, E. Jousselin, G. Kergoat, S. Kögel, J. Li, A. Migeon, H. Pang, G. J. Prinsloo, C. L. Raak-van den Berg, R. Stals, V. Ternois, C. E. Thomas, J. van Lenteren and S. Wang for additional *H. axyridis* samples. We also thank A. Blin and A. Auguste for technical and administrative assistance. This work was funded by grants from the French ANR Biodiversite #ANR-06-BDIV-008-01 and the French Agropolis Fondation (RTRA, Montpellier, BIOFIS project #1001-001).

References

Adriaens T, Branquart E, Maes D (2003) The Multicoloured Asian Ladybird *Harmonia axyridis* Pallas (Coleoptera: Coccinellidae), a threat for native aphid predators in Belgium? *Belgian Journal of Zoology*, **133**, 195–196.

Ascunce MS, Yang CC, Oakey J *et al.* (2011) Global invasion history of the fire ant *Solenopsis invicta*. *Science*, **331**, 1066–1068.

Auger-Rozenberg MA, Boivin T, Magnoux E *et al.* (2012) Inferences on population history of a seed chalcid wasp: invasion success despite a severe founder effect from an unexpected source population. *Molecular Ecology*, **21**, 6086–6103.

Beaumont MA (2010) Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 379–406.

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

Beerli P (2004) Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Molecular Ecology*, **13**, 827–836.

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

Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology*, **19**, 2609–2625.

Blekhman AV, Goryacheva II, Zakharov IA (2010) Differentiation of *Harmonia axyridis* Pall. according to polymorphic morphological traits and variability of the mitochondrial COI gene. *Moscow University Biological Sciences Bulletin*, **65**, 174–176.

Boissin E, Hurley B, Wingfield MJ *et al.* (2012) Retracing the routes of introduction of invasive species: the case of the *Sirex noctilio* woodwasp. *Molecular Ecology*, **21**, 5728–5744.

Bossdorf O, Auge H, Lafuma L *et al.* (2005) Phenotypic and genetic differentiation between native and introduced plant populations. *Oecologia*, **144**, 1–11.

Boubou A, Migeon A, Roderick GK *et al.* (2012) Test of colonisation scenarios reveals complex invasion history of the red tomato spider mite *Tetranychus evansi*. *PLoS ONE*, **7**, 13.

Brown PMJ, Adriaens T, Bathon H *et al.* (2008) *Harmonia axyridis* in Europe: spread and distribution of a non-native coccinellid. *BioControl*, **53**, 5–21.

Brown P, Thomas C, Lombaert E *et al.* (2011) The global spread of *Harmonia axyridis* (Coleoptera: Coccinellidae): distribution, dispersal and routes of invasion. *BioControl*, **56**, 623–641.

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

Chapin J, Brou V (1991) *Harmonia axyridis* (Pallas), the third species of the genus to be found in the United States (Coleoptera: Coccinellidae). *Proceedings of the Entomological Society of Washington*, **93**, 630–635.

Ciosi M, Miller NJ, Kim KS *et al.* (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.

Corander J, Waldmann P, Sillanpaa MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics*, **163**, 367–374.

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, Ravigne V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, **11**, 401.

Cornuet J-M, Pudlo P, Veyssier J *et al.* (2014) DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide poly-

© 2014 John Wiley & Sons Ltd

morphism, DNA sequence and microsatellite data. *Bioinformatics*, **30**, 1187–1189.

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

Darling JA, Bagley MJ, Roman J, Tepolt CK, Geller JB (2008) Genetic patterns across multiple introductions of the globally invasive crab genus *Carcinus*. *Molecular Ecology*, **17**, 4992–5007.

Dobzhansky T (1933) Geographical variation in lady-beetles. *The American Naturalist*, **67**, 97–126.

Edmands S (1999) Heterosis and outbreeding depression in interpopulation crosses spanning a wide range of divergence. *Evolution*, **53**, 1757–1768.

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

Estoup A, Lombaert E, Marin JM *et al.* (2012) Estimation of demo-genetic model probabilities with Approximate Bayesian Computation using linear discriminant analysis on summary statistics. *Molecular Ecology Resources*, **12**, 846–855.

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.

Facon B, Pointier J-P, Glaubrecht M *et al.* (2003) A molecular phylogeography approach to biological invasions of the New World by parthenogenetic Thiarid snails. *Molecular Ecology*, **12**, 3027–3039.

Facon B, Genton BJ, Shykoff J *et al.* (2006) A general eco-evolutionary framework for understanding bioinvasions. *Trends in Ecology & Evolution*, **21**, 130–135.

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

Ferran A, Giuge L, Brun J, Gambier J, Kabiri F (1997) Coccinelle *Harmonia axyridis* pallas: mise au point sur son introduction et son utilisation en lutte biologique. *Adalia*, **36**, 21–24.

Goudet J (2002) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3.2). Updated from Goudet (1995).

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.

Guillemaud T, Ciosi M, Lombaert E, Estoup A (2011) Biological invasions in agricultural settings: insights from evolutionary biology and population genetics. *Comptes Rendus Biologies*, **334**, 237–246.

Harwood TD (2009) The circular definition of populations and its implications for biological sampling. *Molecular Ecology*, **18**, 765–768.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

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

Kalinowski ST (2011) The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. *Heredity*, **106**, 625–632.

Keller SR, Taylor DR (2008) History, chance and adaptation during biological invasion: separating stochastic phenotypic evolution from response to selection. *Ecology Letters*, **11**, 852–866.

Keller SR, Gilbert KJ, Fields PD, Taylor DR (2012) Bayesian inference of a complex invasion history revealed by nuclear and chloroplast genetic diversity in the colonizing plant, *Silene latifolia*. *Molecular Ecology*, **21**, 4721–4734.

Koch RL (2003) The multicolored Asian lady beetle, *Harmonia axyridis*: a review of its biology, uses in biological control, and non-target impacts. *Journal of Insect Science*, **3**, 1–16.

Koch RL, Venette RC, Hutchison WD (2006) Invasions by *Harmonia axyridis* (Pallas) (Coleoptera: Coccinellidae) in the Western Hemisphere: implications for South America. *Neotropical Entomology*, **35**, 421–434.

Kolbe JJ, Glor RE, Schettino LRG *et al.* (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, **431**, 177–181.

Konecny A, Estoup A, Duplantier JM *et al.* (2013) Invasion genetics of the introduced black rat (*Rattus rattus*) in Senegal, West Africa. *Molecular Ecology*, **22**, 286–300.

Krafsur ES, Kring TJ, Miller JC *et al.* (1997) Gene flow in the exotic colonizing ladybeetle *Harmonia axyridis* in North America. *Biological Control*, **8**, 207–214.

Labrie G, Lucas E, Coderre D (2006) Can developmental and behavioral characteristics of the multicolored Asian lady beetle *Harmonia axyridis* explain its invasive success? *Biological Invasions*, **8**, 743–754.

LaMana ML, Miller JC (1996) Field observations on *Harmonia axyridis* Pallas (Coleoptera: Coccinellidae) in Oregon. *Biological Control*, **6**, 232–237.

Leberg PL (2002) Estimating allelic richness: effects of sample size and bottlenecks. *Molecular Ecology*, **11**, 2445–2449.

Loiseau A, Malausa T, Lombaert E, Martin JF, Estoup A (2009) Isolation and characterization of microsatellites in the harlequin ladybird, *Harmonia axyridis* (Coleoptera, Coccinellidae), and cross-species amplification within the family Coccinellidae. *Molecular Ecology Resources*, **9**, 934–937.

Lombaert E, Guillemaud T, Cornuet JM *et al.* (2010) Bridgehead effect in the worldwide invasion of the biocontrol harlequin ladybird. *PLoS ONE*, **5**, e9743.

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.

Lynch M (1991) The genetic interpretation of inbreeding depression and outbreeding depression. *Evolution*, **45**, 622–629.

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 *et al.* (2008) Identifying the source of species invasions: sampling intensity vs. genetic diversity. *Molecular Ecology*, **17**, 1020–1035.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

© 2014 John Wiley & Sons Ltd

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.

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

Raymond M, Rousset F (1995a) An exact test for population differentiation. *Evolution*, **49**, 1280–1283.

Raymond M, Rousset F (1995b) Genepop (version. 1.2), a population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rius M, Darling JA (2014) How important is intraspecific genetic admixture to the success of colonising populations? *Trends in Ecology & Evolution*, **29**, 233–242.

Robert CP, Cornuet JM, Marin JM, Pillai NS (2011) Lack of confidence in approximate Bayesian computation model choice. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 15112–15117.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Roy HE, Adriaens T, Isaac NJB *et al.* (2012) Invasive alien predator causes rapid declines of native European ladybirds. *Diversity and Distributions*, **18**, 717–725.

Saini E (2004) Presencia de *Harmonia axyridis* (Pallas) (Coleoptera: coccinellidae) en la provincia de Buenos aires. Aspectos biologicos y morfologicos. *RIA*, **33**, 151–160.

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

Schwartz MK, McKelvey KS (2009) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics*, **10**, 441–452.

Simberloff D, Martin JL, Genovesi P *et al.* (2013) Impacts of biological invasions: what's what and the way forward. *Trends in Ecology & Evolution*, **28**, 58–66.

Slatkin M (2005) Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. *Molecular Ecology*, **14**, 67–73.

Stals R (2010) The establishment and rapid spread of an alien invasive lady beetle: *Harmonia axyridis* (Coleoptera: Coccinellidae) in southern Africa, 2001–2009. *IOBC/WPRS Bulletin*, **58**, 125–132.

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, **144**, 389–399.

Tayeh A, Estoup A, Laugier G *et al.* (2012) Evolution in biocontrol strains: insight from the harlequin ladybird *Harmonia axyridis*. *Evolutionary Applications*, **5**, 481–488.

Tayeh A, Estoup A, Hufbauer RA *et al.* (2013) Investigating the genetic load of an emblematic invasive species: the case of the invasive harlequin ladybird *Harmonia axyridis*. *Ecology and Evolution*, **3**, 864–871.

Templeton AR (2010) Coherent and incoherent inference in phylogeography and human evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 6376–6381.

Turgeon J, Tayeh A, Facon B *et al.* (2011) Experimental evidence for the phenotypic impact of admixture between wild and biocontrol Asian ladybird (*Harmonia axyridis*) involved in the European invasion. *Journal of Evolutionary Biology*, **24**, 1044–1052.

Wang S, Lewis CM, Jakobsson M *et al.* (2007) Genetic variation and population structure in Native Americans. *Plos Genetics*, **3**, 2049–2067.

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1439.

Wares JP, Hughes AR, Grosberg RK (2005) Mechanisms that drives evolutionary change: insights from species introductions and invasions. In: *Species Invasions: Insights into Ecology, Evolution and Biogeography* (eds Sax DF, Stachowicz JJ, Gaines SD), pp. 229–257. Sinauer Associates Inc, Sunderland, Massachusetts.

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

Yalcindag E, Elguero E, Arnathau C *et al.* (2012) Multiple independent introductions of *Plasmodium falciparum* in South America. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 511–516.

E.L., A.E. and T.G. conceived and designed the experiments. E.L. performed the experiments. E.L., A.E. and T.G. analysed the data. E.L., A.E., J.L., R.K., B.F., A.G., A.L., T.M., O.N., E.R, A.S. and T.S. coordinated the worldwide sampling. E.L., A.E. and T.G. wrote the manuscript. All authors have revised and approved the final manuscript.

Data accessibility

Sampling locations and microsatellite data: Dryad doi:10.5061/dryad.80kg5.

Supporting information

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

Appendix S1 Analyses of *Harmonia axyridis* genetic structure with the clustering method implemented in BAPS.

Appendix S2 Retracing the routes of invasion using approximate Bayesian computation (ABC): a methodological guideline.

Fig. S1 Graphic representation of four ABC scenarios among the 28 compared to infer the origin of the Western European target genetic unit.

Fig. S2 Genetic diversity in the biocontrol, invasive and native site sample of *Harmonia axyridis*.

Fig. S3 Neighbor-joining tree for the studied *Harmonia axyridis* site samples based on the distance of Cavalli-Sforza & Edwards (1967, *American Journal of Human Genetics*, 19:233–257).

© 2014 John Wiley & Sons Ltd

Fig. S4 Likelihood of the multilocus genetic data as a function of the number of genetic clusters among all 47 site samples of *Harmonia axyridis*.

Table S1 Biocontrol, invasive and native site samples of *Harmonia axyridis*.

Table S2 Prior distributions of demographic, historical and genetic parameters used in all ABC analyses processed to retrace the worldwide routes of invasion of *H. axyridis*.

Table S3 Pairwise estimates of FST between all pairs of *Harmonia axyridis* site samples.

Table S4 Detailed results of the first set of ABC analyses processed to make inferences about the origin of invasive genetic units.

Table S5 Evaluation of confidence in scenario selection using ABC.

Table S6 Model checking results for the first set of ABC analyses (i.e. origin of invasive genetic units)

Table S7 Model checking results for the second set of ABC analyses (i.e. intracontinental relationship between target invasive genetic units).

© 2014 John Wiley & Sons Ltd

Supporting Information

MS Title: Complementarity of statistical treatments to reconstruct worldwide routes of invasion: the case of the Asian ladybird *Harmonia axyridis*.

Authors: Eric Lombaert, Thomas Guillemaud, Jonathan Lundgren, Robert Koch, Benoît Facon, Audrey Grez, Antoon Loomans, Thibaut Malausa, Oldrich Nedved, Emma Rhule, Arnstein Staverlokk, Tove Steenberg and Arnaud Estoup.

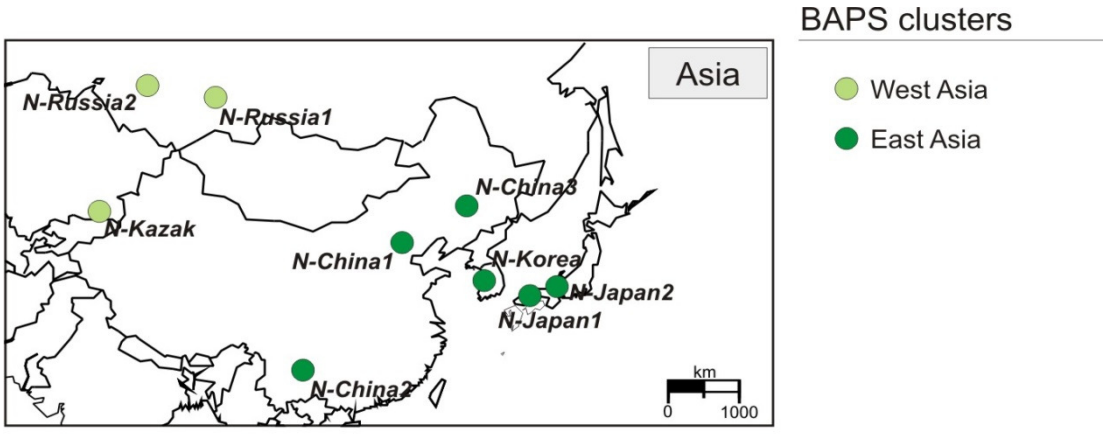
Appendix S1 – Analyses of Harmonia axyridis genetic structure with the clustering method implemented in BAPS

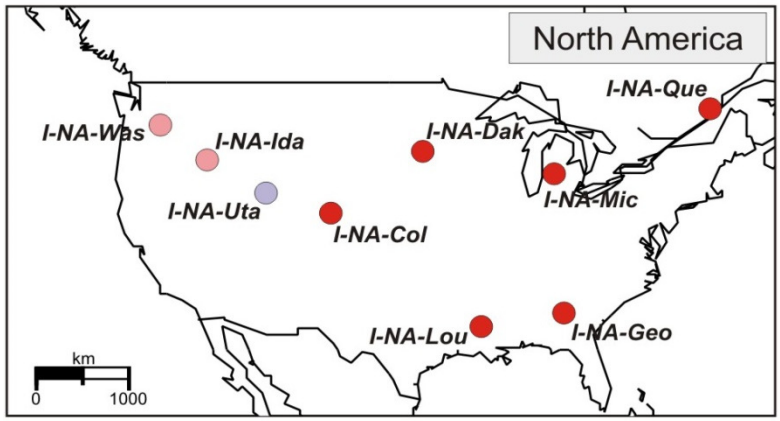
Motivation and methods:

In order to retrace the invasion history of *Harmonia axyridi*s (HA), we used an approximate Bayesian computation framework (ABC, Beaumont *et al.* 2002). Because it was not computationally possible to conduct ABC analyses including all target site samples with all putative source site samples (see Box 1), we first grouped site samples into genetic units using two Bayesian clustering methods: STRUCTURE (Pritchard *et al.* 2000) and BAPS (Corander *et al.* 2003). Such genetic units represent the basic elements of our ABC analyses. The STRUCTURE analysis has been detailed in the main text. In the present appendix section we present the details of our BAPS analysis. We used the software BAPS 5.2 (Corander *et al.* 2004) to conduct five analyses considering each continent independently (i.e. Asia, North America, South America, Europe and Africa). The aim was to obtain clear-cut clusters within each continent by analysing group of individuals assuming a model without admixture as implemented in BAPS. For each of the five continental analyses, we conducted a series of 20 replicate runs with the upper limit for the number of clusters set at the number of site samples in the target continent. The consistency of each inferred genetic unit was probed in light of worldwide F_{ST}-based, NJ tree-based and STRUCTURE-based analyses (see main text).

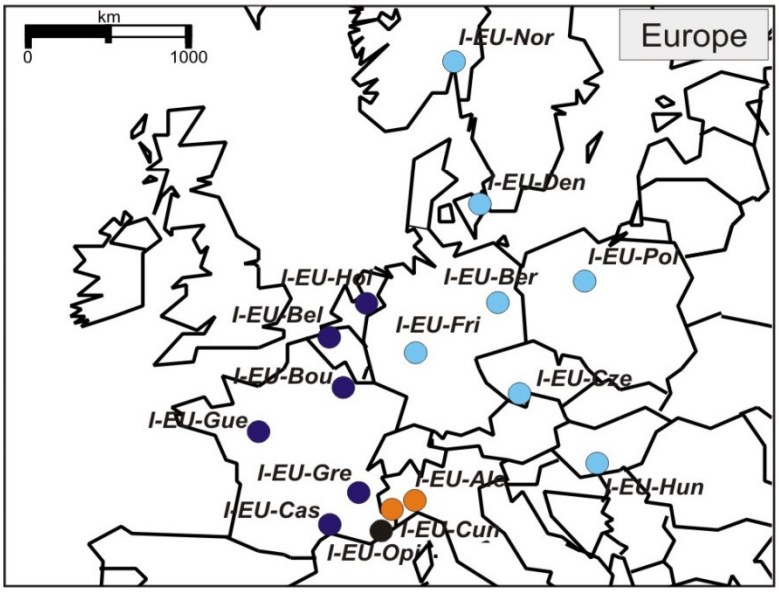
Results:

Results obtained from the BAPS analyses were highly consistent with those of the STRUCTURE analysis (see Fig. 2 of main text). Results for Asia were the same as those of Lombaert *et al.* (2011) with two clusters showing a clear east/west geographical pattern. In North America, three BAPS clusters were inferred, one including six Eastern site samples (the ENA cluster), one including two Western site samples (the WNA cluster) and one with a single geographically intermediate site sample (I-NA-Uta). In South America, three BAPS clusters were inferred each being associated with a single country: Brazil (two site samples), Argentina and Chile. In Europe, four BAPS clusters were inferred. One is formed by a single site sample (I-EU-Opi) located in the South of France, while another one is formed by the two Italian site samples. A third BAPS European cluster was formed by six samples located in Western Europe (France, Belgium and Holland). A fourth BAPS European cluster included seven samples located in more eastern areas (Germany, Denmark, Czech Republic, Hungary, Poland and Norway). Finally, only one BAPS cluster was identified in Africa.
The following maps summarize the genetic clusters inferred from BAPS analyses for each continent. Site sample names are as in Fig. 1 and Table S1:

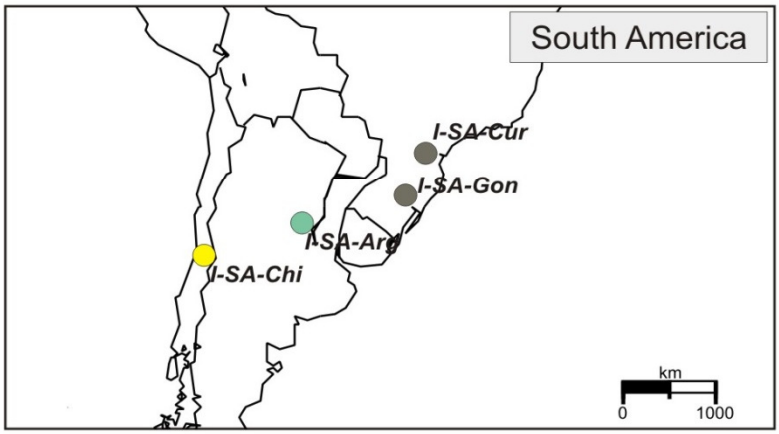




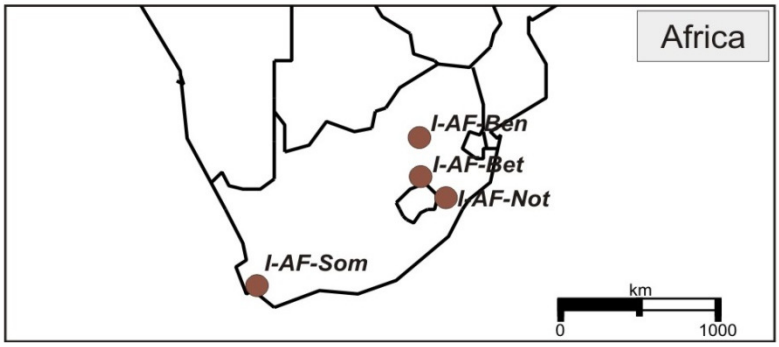
- ENA
- Utah
- WNA



- Western Europe
- South France
- Italy
- Eastern Europe



- Brazil
- Argentina
- Chile



- South Africa

References

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

Corander J, Waldmann P, Marttinen P, Sillanpaa MJ (2004) BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics* **20**, 2363-2369.

Corander J, Waldmann P, Sillanpaa MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics* **163**, 367-374.

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.

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

Appendix S2 - Retracing the routes of invasion using approximate Bayesian computation (ABC): a methodological guideline

1/ 1st set of ABC analyses: inferences about the origin of target invasive genetic units

Summary statistics, prior distributions and simulations

The nature and the number of summary statistics used to summarise the data have a crucial role in the power and reliability of an ABC analysis. The choice of the summary statistics has to be made on the basis of the addressed question(s). In our case, the choice was based on various patterns expected to occur in the specific case of invasion processes and were the same as in Lombaert *et al.* (2011). First, invasion may generate strong demographic bottleneck leading to reduction in genetic diversity. Statistics such as number of alleles, heterozygosity, range of allelic size or the Garza-Williamson M (Garza & Williamson 2001) statistics are informative in this context. Second, inter-population statistics, including F_{ST} (Weir & Cockerham 1984), number of alleles, heterozygosity, range of allelic size and mean individual assignment likelihoods (Pascual *et al.* 2007), are useful to identify population source of invasive populations. Finally, as genetic admixture was one of the focal interests of our ABC analyses, the maximum likelihood estimate of admixture proportion (Cornuet *et al.* 2008) was added to our set of summary statistics.

The prior distributions of demographic, historical and genetics parameters were defined from previous studies (see Brown *et al.* 2011; Lombaert *et al.* 2011) and are detailed in Table S2.

Simulations were performed within a coalescent framework using a combination of a generation by generation algorithm (for small effective size) and a continuous time algorithm (for large effective size). See Cornuet *et al.* (2008) for more details. All simulations and analyses were performed with the software DIYABC v2 (Cornuet *et al.* 2014).

To estimate ABC scenario probabilities, we used linear discriminant analysis (LDA) transformation on summary statistics before computing logistic regression (Estoup *et al.* 2012; Cornuet *et al.* 2014). More specifically, we used LDA to transform the set of usually large number J of summary statistics into (X - 1) independent variables maximizing the differences among the X compared scenarios. This allows reducing the dimension of the set of explanatory variables from J non-independent to (X - 1) independent variables, and this whatever the value of J. A major practical advantage of using LDA-transformed summary statistics is that it substantially decreases the dimension of explanatory variables making computation of scenario probability much faster and sometimes simply feasible when the memory space is not large enough to compute the matrix of second partial derivatives of the likelihood. This allows larger data-scenarios settings to be analysed. Another important advantage of LDA-transformation of raw summary statistics when processing model choice is that it reduces the potential difficulties associated with the curse of dimensionality (due to a too large number of summary statistics) and avoids correlation among explanatory variables (i.e. multi-co-linearity) during the regression step (Estoup *et al.* 2012).

Definition of source and target genetic units

The STRUCTURE and BAPS clustering analyses were to a large extent congruent with one another. We based the definition of genetic units for the ABC analyses on the results of both Bayesian clustering analyses (see main text, Fig. 2 and Appendix S1). In the following, the BAPS/STRUCTURE clusters represent the genetic units of our ABC analyses.

We used pairwise F_{ST} estimations between the site samples within each genetic unit to define five different sample sets representing the different sources of invasive genetic units considered in the ABC analyses (see main text and Table 1). Regarding the target invasive genetic units (i.e. the

invasive genetic units for which we want to know the origin), we did not reanalyse the origin of the Eastern USA (ENA) and Western USA (WNA) genetic units associated with the I-NA-Lou and I-NA-Was site samples respectively. As a matter of fact, these ancient genetic units (first observation of the outbreaks in 1988 and 1991, respectively) have been thoroughly described in previous studies (Lombaert *et al.* 2010; Lombaert *et al.* 2011). We thus had a total of 10 target invasive genetic units (Table 2): five were represented by a single site sample (I-NA-Col and I-NA-Uta in North America, I-SA-Arg and I-SA-Chi in South America, and I-EU-Opi in Europe), two by two site samples (Brazil in South America with I-SA-Cur and I-SA-Gon, and Italy in Europe with I-EU-Ale and I-EU-Cun) and three by more than two samples. For each of these last three genetic units, we chose two site samples based on the minimum and maximum mean intra-units F_{ST} values: I-EU-Bel and I-EU-Cas for Western Europe, I-EU-Hun and I-EU-Nor for Eastern Europe, and I-AF-Bet and I-AF-Som in Africa. Altogether, this represented a total of 15 target site samples (Table 2).

Definition of competing scenarios

Because we considered five different sample sets representing the different putative source genetic units (Table 1) to analyse 15 target site samples, we performed a total of 75 independent ABC analyses. In each ABC analysis, we fixed the evolutionary history of every putative source genetic units according to the results of Lombaert *et al.* (2011). We compared introduction scenarios which differed by the target genetic unit originating from one of the source genetic units or from an admixture between two of them. Except in the case of central North American target genetic units (Colorado and Utah), we did not use in this first set of ABC analyses any source genetic units located in the same continent. Given the dates of first observation of each outbreak and the associated priors (Table S2), all analyses consisted of 28 competing scenarios (cf. seven putative sources; see Fig. S1 for an illustration), except when the target site sample was I-NA-Col or I-NA-Uta for which the number of competing scenarios was ten (cf. four putative sources in these cases).

Example 1:
There were ten competing scenarios for the analysis of the origin of the central North American site sample I-NA-Col (Colorado, USA, first observation in 1999). Four source genetic units were used and their evolutionary history was based on previous knowledge: there were two diverging native genetic units (East and West), one Eastern North American genetic unit (ENA) originating from an admixture between both native genetic units, and one Western North American genetic unit (WNA) originating from the Eastern native genetic unit. The ten competing scenarios were thus:

- Scenario 1 = East native genetic unit origin.
- Scenario 2 = West native genetic unit.
- Scenario 3 = ENA genetic unit.
- Scenario 4 = WNA genetic unit.
- Scenario 5 = admixture between East native and West native genetic units.
- Scenario 6 = admixture between East native and ENA genetic units.
- Scenario 7 = admixture between East native and WNA genetic units.
- Scenario 8 = admixture between ENA and West native genetic units.
- Scenario 9 = admixture between ENA and WNA genetic units.
- Scenario 10 = admixture between West native and WNA genetic units.

Example 2:
There were 28 competing scenarios for the analysis of the origin of the European site sample I-EU-Bel (Belgium, first observation in 2001). Seven source genetic units were used and their evolutionary history was based on previous knowledge: there were two diverging native genetic units (East and West), one ENA genetic unit originating from an admixture between both native genetic units, one WNA genetic unit originating from the Eastern native genetic unit, one European Biocontrol genetic unit (EBC) originating from the Eastern native genetic unit, one South American genetic unit (SA) originating from the ENA genetic unit and one African genetic unit (AF) originating from the ENA genetic unit. The 28 competing scenarios were thus:

- Scenario 1 = East native genetic unit origin (e.g. Fig. S1).

- Scenario 2 = West native genetic unit.
- Scenario 3 = EBC genetic unit.
- Scenario 4 = ENA genetic unit.
- Scenario 5 = WNA genetic unit (e.g. Fig. S1).
- Scenario 6 = SA genetic unit.
- Scenario 7 = AF genetic unit.
- Scenario 8 = admixture between East native and West native genetic units.
- Scenario 9 = admixture between East native and EBC genetic units.
- Scenario 10 = admixture between East native and ENA genetic units.
- Scenario 11 = admixture between East native and WNA genetic units (e.g. Fig. S1).
- Scenario 12 = admixture between East native and SA genetic units.
- Scenario 13 = admixture between East native and AF genetic units.
- Scenario 14 = admixture between West native and EBC genetic units.
- Scenario 15 = admixture between West native and ENA genetic units.
- Scenario 16 = admixture between West native and WNA genetic units.
- Scenario 17 = admixture between West native and SA genetic units.
- Scenario 18 = admixture between West native and AF genetic units.
- Scenario 19 = admixture between EBC and ENA genetic units.
- Scenario 20 = admixture between EBC and WNA genetic units (e.g. Fig. S1).
- Scenario 21 = admixture between EBC and SA genetic units.
- Scenario 22 = admixture between EBC and AF genetic units.
- Scenario 23 = admixture between ENA and WNA genetic units.
- Scenario 24 = admixture between ENA and SA genetic units.
- Scenario 25 = admixture between ENA and AF genetic units.
- Scenario 26 = admixture between WNA and SA genetic units.
- Scenario 27 = admixture between WNA and AF genetic units.
- Scenario 28 = admixture between SA and AF genetic units.

To provide further practical insights on our ABC treatments, we present below the way we coded the above scenarios 1 and 11 (see also Fig. S1) within the dedicated window of the graphical interface of the software DIYABC version 2 (Cornuet *et al.* 2014):

```
scenario 1:
N N N3 N N N N N N N NA
7 sample 1
7 sample 2
58 sample 3
7 sample 4
7 sample 5
5 sample 6
4 sample 7
5 sample 8
t8-DB8 VarNe 8 NF8
t8 VarNe 8 N
t8ge merge 1 8
t7-DB7 VarNe 7 NF7
t7 merge 4 7
t6-DB6 VarNe 6 NF6
t6 merge 4 6
t5-DB5 VarNe 5 NF5
t5 VarNe 5 N
t5ge merge 1 5
t4-DB4 VarNe 4 NF4
t4 split 4 9 10 ra4
t4ge merge 1 9
t4gw merge 2 10
t3 VarNe 3 N
t3ge merge 1 3
ta merge 11 1
ta merge 11 2
```

```
scenario 11:
N N N3 N N N N N N N NA N
7 sample 1
7 sample 2
58 sample 3
7 sample 4
7 sample 5
5 sample 6
4 sample 7
5 sample 8
t8-DB8 VarNe 8 NF8
t8 split 8 12 5 ra8
t8ge merge 1 12
t7-DB7 VarNe 7 NF7
t7 merge 4 7
t6-DB6 VarNe 6 NF6
t6 merge 4 6
t5-DB5 VarNe 5 NF5
t5 VarNe 5 N
t5ge merge 1 5
t4-DB4 VarNe 4 NF4
t4 split 4 9 10 ra4
t4ge merge 1 9
t4gw merge 2 10
t3 VarNe 3 N
t3ge merge 1 3
ta merge 11 1
ta merge 11 2
```

With the following conditions:

```
ta>=t3ge
ta>=t4gw
ta>=t4ge
ta>=t5ge
ta>=t8ge
t3ge>=t3
t4gw>=t4
t4ge>=t4
t5ge>=t5
t8ge>=t8
```

And with the following site samples (here the "reference" source sample set, Table 1) in the data file:

Sample 1: East native (here N-China2)
Sample 2: West native (here N-Kazak)
Sample 3: EBC (here EB-INRA87)
Sample 4: ENA (here I-NA-Lou)
Sample 5: WNA (here I-NA-Was)
Sample 6: SA (here I-SA-Cur)
Sample 7: AF (here I-AF-Som)
Sample 8: Western Europe (I-EU-Bel)

Prior distributions of demographic parameters are described in Table S2.

Quality control of data analysis

Confidence in scenario choice (type I and type II error rates)

We evaluated the ability of ABC to correctly select the true scenario by analyzing pseudo-observed datasets (pods) simulated from the various competing scenarios with the same number of loci and individuals as in the real dataset. For one ABC analysis, 100 pods of each scenario were simulated using parameter values drawn from the same probability distributions as the priors (Table S2). The posterior probability of each competing scenario was estimated for each pod using the 1% closest datasets. We then focused on the scenario selected with the real dataset (i.e. the scenario with the highest posterior probability among the set of compared scenarios) to compute type I and II error rates. Type I error rate is the risk to exclude the focal scenario when it is the true one and type II error rate is the risk to select the focal scenario when it is false (see Cornuet *et al.* 2010 for details).

Because we focused on the scenario selected with the real dataset, the most important information we want here corresponds to an estimation of the risk that we erroneously selected this scenario. This is measured by the type II errors. By contrast, the type I errors are not informative to evaluate the above risk but are useful for interpreting discrepancies between replicated ABC analyses based on different sample sets.

The first set of ABC analyses consisted in a total of 75 independent analyses (Table 2). Because type I and type II error rates are computationally intensive, we performed such computations on a selection of three of these 75 analyses: (i) the analysis of the Brazilian I-SA-Cur site sample using the "reference" source sample set (final selected scenario: Eastern North America origin, Table 2), (ii) the analysis of the Belgian I-EU-Bel site sample using the "reference" source sample set (final selected scenario: Eastern North America and European biocontrol admixture origin, Table 2), and (iii) the analysis of the Italian I-EU-Cun site sample using the "reference" source sample set (final selected scenario: Western North America origin, Table 2). The three analyses were chosen because the final selected scenarios were very different. See Table S5 for results.

Model checking

ABC methods for scenario comparison provide posterior probabilities of scenarios and associated posterior distributions of parameters without any "goodness of fit" information. We therefore used the model checking option of DIYABC on the final scenario inferred for each ABC analysis to determine whether this scenario matches well with the observed genetic data (Cornuet *et al.* 2010). The principle is as follows: if an invasion scenario fits the observed data correctly, then data simulated under this model with parameters drawn from their posterior distribution should be close to the observed data (Gelman *et al.* 1995). The lack of fit of the model to the data with respect to the posterior predictive distribution can be measured by determining the frequency at which the observed summary statistics are extreme with respect to the test statistic (here, our simulated summary statistics) distribution; hence defining a tail-area probability or P-value, for each test statistics. For each model checking analysis, we proceeded as follows: from the $5x10^5$ datasets simulated under the selected scenario, we obtained a "posterior sample" of $5x10^3$ values of the posterior distributions of parameters through a rejection step based on Euclidian distances and a linear regression post-treatment (Beaumont *et al.* 2002). We simulated 10^4 datasets with parameter values drawn from this "posterior sample". Our set of test statistics included the summary statistics used for ABC analysis and two previously unused statistics: the shared allele distances (Chakraborty & Jin 1993) and $(\delta\mu)^2$ distances (Goldstein *et al.* 1995) between each population pair. We did this to reduce the conservative bias associated with the use of summary statistics previously selected for ABC analysis as test statistics (Cornuet *et al.* 2010). Each observed test statistic was compared with the 10^4 simulated test statistics, and its p-value was calculated. See Table S6 for results.

2/ 2nd set of ABC analyses: inferences about the intra-continental relationships between a subset of target invasive genetic units sharing similar extra-continental origin

Summary statistics, prior distributions and simulation

The statistics employed to summarize the data and parameter prior distributions were the same as in the first set of ABC analyses. Simulations and analyses were also performed with the coalescent based software DIYABC v2 (Cornuet *et al.* 2014).

Definition of competing scenarios

The second set of ABC analysis aimed at deciphering the historical relationships between different genetic units from the same continent that shared the same extra-continental origin according to the first set of ABC analyses. The main question that we addressed was: how many independent introductions occurred? As done so far, we fixed in each ABC analysis the evolutionary history of every putative source genetic units according to the results of Lombaert et al. (2011) and the first set of ABC analyses. We then compared several invasion scenarios which differ in the number of independent introductions at the origin of the target genetic units.

In the case of South America (SA), an Eastern North American (ENA) origin was inferred in the first set of ABC analyses for each of the three target genetic units (i.e. Brazil, Argentina and Chile; Table 2). We hence designed an ABC analysis with the following fixed history: two diverging native genetic units (East and West) and an Eastern North American genetic unit (ENA) originating from an admixture between both native genetic units. We then compared five scenarios corresponding to different levels of dependency between the three South American genetic units:

- Scenario 1: all SA genetic units derived from a single ENA introduction.
- Scenario 2: the SA genetic units derived from 2 independent ENA introductions = one at the origin of the Chilean genetic unit and one at the origin of both the Argentinean and the Brazilian genetic units.
- Scenario 3: the SA genetic units derived from 2 independent ENA introductions = one at the origin of the Brazilian genetic unit and one at the origin of both the Argentinean and the Chilean genetic units.
- Scenario 4: the SA genetic units derived from 2 independent ENA introductions = one at the origin of the Argentinean genetic unit and one at the origin of both the Brazilian and the Chilean genetic units.
- Scenario 5: all SA genetic units derived from independent ENA introductions.

In the case of Europe (EU), several genetic units appeared to share some common extra-continental sources according to the first set of ABC analyses (Table 2) as well as to the STRUCTURE analysis (Fig. 2). Eastern Europe originated from a double admixture between ENA, European Biocontrol (EBC) and Western North America (WNA). Western Europe originated from an admixture between ENA and EBC, and Italy originated from WNA. We hence formalized an ABC analysis with the following fixed history: two diverging native genetic units (East and West), one ENA genetic unit originating from an admixture between both native genetic units, one WNA genetic unit originating from the Eastern native genetic unit and one EBC genetic unit originating from the Eastern native genetic unit. We then compared four scenarios:

- Scenario 1: the EU genetic units derived from two independent ENA introductions (one in Eastern Europe and the other one in Western Europe) and from two independent WNA introductions (one in Eastern Europe and the other one in Italy).
- Scenario 2: the EU genetic units derived from a single ENA introduction (at the origin of both the Eastern Europe and the Western Europe genetic units) and from two independent WNA introductions (one in Eastern Europe and the other one in Italy).

- Scenario 3: the EU genetic units derived from two independent ENA introductions (one in Eastern Europe and the other one in Western Europe) and from a single WNA introduction (at the origin of both the Eastern Europe and the Italian genetic units).
- Scenario 4: the EU genetic units derived from a single ENA introduction (at the origin of both the Eastern Europe and the Western Europe genetic units) and from a single WNA introduction (at the origin of both the Eastern Europe and the Italian genetic units).

For both SA and EU, we performed five independent ABC analyses with each of the five sets of source site samples (Table 1). The method for the ABC analyses was the same than in the 1st set of ABC analyses (see main text) except that we simulated 10^6 microsatellite datasets for each competing scenario.

Quality control of data analysis

Confidence in scenario choice (type I and type II error rates)

The procedure was similar to that described above for the 1st set of ABC analyses except that error rates were based on the analysis of 500 pseudo-observed datasets (pods) per scenario. For both continents (i.e. SA and EU) type I and type II error rates were computed on the analyses done with the "reference" source sample set. See Table S5 for results.

Model checking

The procedure was similar to that described above for the 1st set of ABC analyses except that 10^6 datasets were simulated under the selected scenario, and the "posterior sample" was thus obtained from 10^4 values of the posterior distributions of parameters through the rejection step. See Table S7 for results.

References

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

Brown P, Thomas C, Lombaert E*, et al.* (2011) The global spread of *Harmonia axyridis* (Coleoptera: Coccinellidae): distribution, dispersal and routes of invasion. *Biocontrol* **56**, 623-641.

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

Cornuet J-M, Pudlo P, Veyssier J*, et al.* (2014) DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. 10.1093/bioinformatics/btt1763.

Cornuet JM, Ravigne V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics* **11**, 401.

Cornuet JM, Santos F, Beaumont M, Robert C, Marin JM, Balding D, Guillemaud T, Estoup A (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* **24**, 2713-2719.

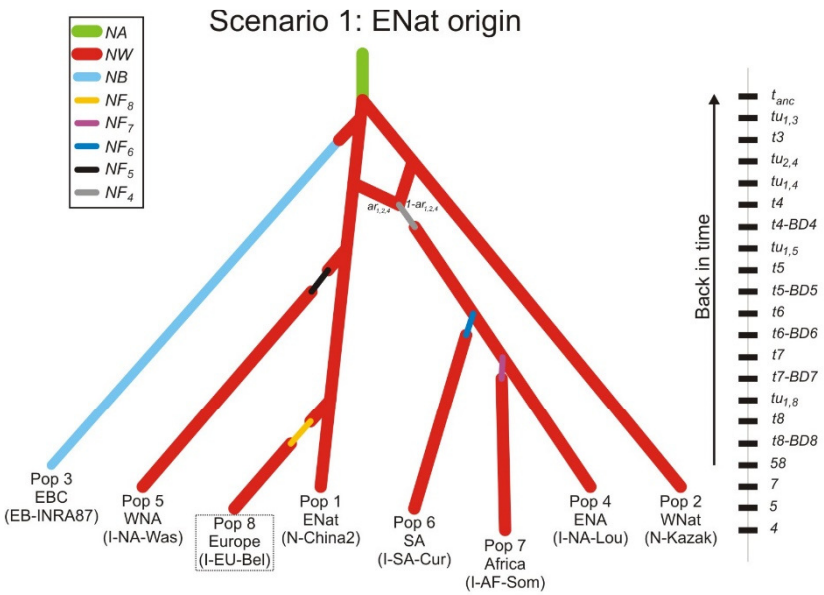
Estoup A, Lombaert E, Marin JM*, et al.* (2012) Estimation of demo-genetic model probabilities with Approximate Bayesian Computation using linear discriminant analysis on summary statistics. *Molecular Ecology Resources* **12**, 846-855.

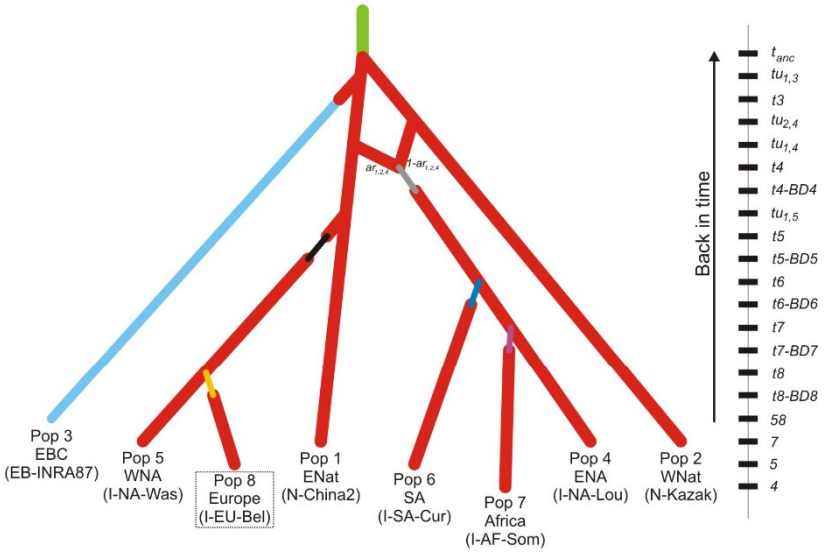
Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* **10**, 305-318.

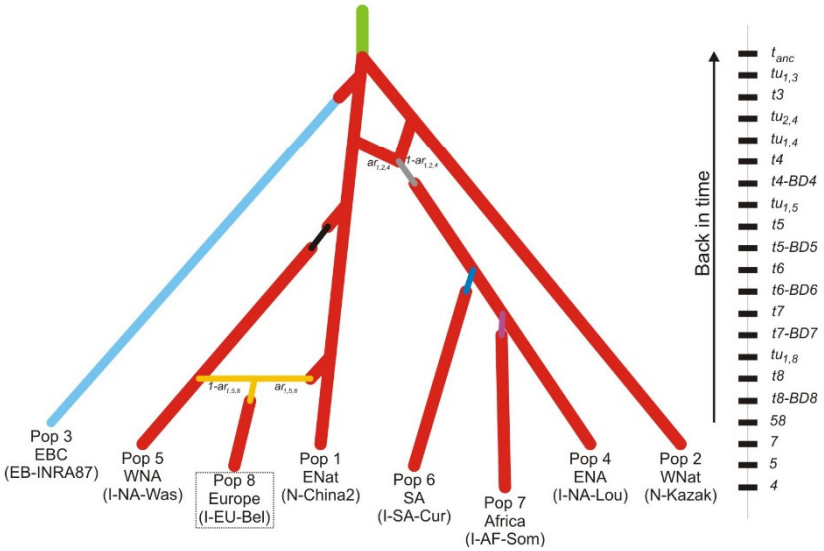
Gelman A, Carlin JB, Stern HS, Rubin DB (1995) *Bayesian Data Analysis* Chapman and Hall, New York.

Goldstein DB, Linares AR, Cavalli-sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**, 463-471.

Lombaert E, Guillemaud T, Cornuet JM*, et al.* (2010) Bridgehead effect in the worldwide invasion of the biocontrol harlequin ladybird. *Plos One* **5**, e9743.
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.
Nei M (1987) *Molecular Evolutionary Genetics* Columbia University Press, New York.
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.
Weir BS, Cockerham C (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.


Scenario 1: ENat origin
NA
NW
NB
NF8
NF7
NF6
NF5
NF4
Back in time
tanc
tu1,3
t3
tu2,4
tu1,4
t4
t4-BD4
tu1,5
t5
t5-BD5
t6
t6-BD6
t7
t7-BD7
tu1,8
t8
t8-BD8
58
7
5
4
Pop 3 EBC (EB-INRA87)
Pop 5 WNA (I-NA-Was)
Pop 8 Europe (I-EU-Bel)
Pop 1 ENat (N-China2)
Pop 6 SA (I-SA-Cur)
Pop 7 Africa (I-AF-Som)
Pop 4 ENA (I-NA-Lou)
Pop 2 WNat (N-Kazak)


Scenario 5: WNA origin
Back in time
tanc
tu1,3
t3
tu2,4
tu1,4
t4
t4-BD4
tu1,5
t5
t5-BD5
t6
t6-BD6
t7
t7-BD7
t8
t8-BD8
58
7
5
4
Pop 3 EBC (EB-INRA87)
Pop 5 WNA (I-NA-Was)
Pop 8 Europe (I-EU-Bel)
Pop 1 ENat (N-China2)
Pop 6 SA (I-SA-Cur)
Pop 7 Africa (I-AF-Som)
Pop 4 ENA (I-NA-Lou)
Pop 2 WNat (N-Kazak)


Scenario 11: admixture ENat + WNA
Back in time
tanc
tu1,3
t3
tu2,4
tu1,4
t4
t4-BD4
tu1,5
t5
t5-BD5
t6
t6-BD6
t7
t7-BD7
tu1,8
t8
t8-BD8
58
7
5
4
Pop 3 EBC (EB-INRA87)
Pop 5 WNA (I-NA-Was)
Pop 8 Europe (I-EU-Bel)
Pop 1 ENat (N-China2)
Pop 6 SA (I-SA-Cur)
Pop 7 Africa (I-AF-Som)
Pop 4 ENA (I-NA-Lou)
Pop 2 WNat (N-Kazak)

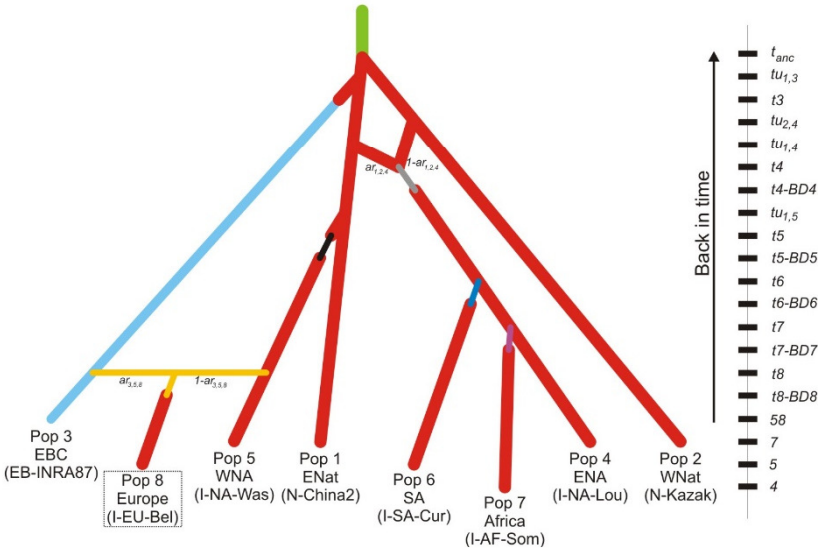

Scenario 20: admixture EBC + WNA
Back in time
tanc
tu1,3
t3
tu2,4
tu1,4
t4
t4-BD4
tu1,5
t5
t5-BD5
t6
t6-BD6
t7
t7-BD7
t8
t8-BD8
58
7
5
4
Pop 3 EBC (EB-INRA87)
Pop 8 Europe (I-EU-Bel)
Pop 5 WNA (I-NA-Was)
Pop 1 ENat (N-China2)
Pop 6 SA (I-SA-Cur)
Pop 7 Africa (I-AF-Som)
Pop 4 ENA (I-NA-Lou)
Pop 2 WNat (N-Kazak)

Figure S1: Graphic representation of four ABC scenarios among the 28 compared to infer the origin of the Western European target genetic unit.
Note: The Western European target genetic unit is represented by the site sample I-EU-Bel (framed sample) and putative sources are those of the "reference" set (see Table 1). Site sample code names are as in Fig. 1 and Table S1. All parameters with associated prior distributions are described in Table S2. Time 0 is the sampling year 2010 and time 58 is the sampling year 1987 (we consider 2.5 generations per year). Time is not at scale. Scenario 1 corresponds to an Eastern native origin (Pop 1) of the target genetic unit (Pop 8). Scenario 5 corresponds to a Western North American origin (Pop 5). In scenario 11, the target genetic unit is the result of an admixture between individuals from the Eastern native area (Pop 1) at a rate $ar_{1,5,8}$ and from the Western North American area (Pop 5) at a rate $1 - ar_{1,5,8}$. In scenario 20, the target genetic unit is the result of an admixture between individuals from the European biocontrol genetic unit (Pop 3) at a rate $ar_{3,5,8}$ and from the Western North American area (Pop 5) at a rate $1 - ar_{3,5,8}$. The scenarios presented here are part of the first set of ABC analyses (see Appendix S2 for more information).

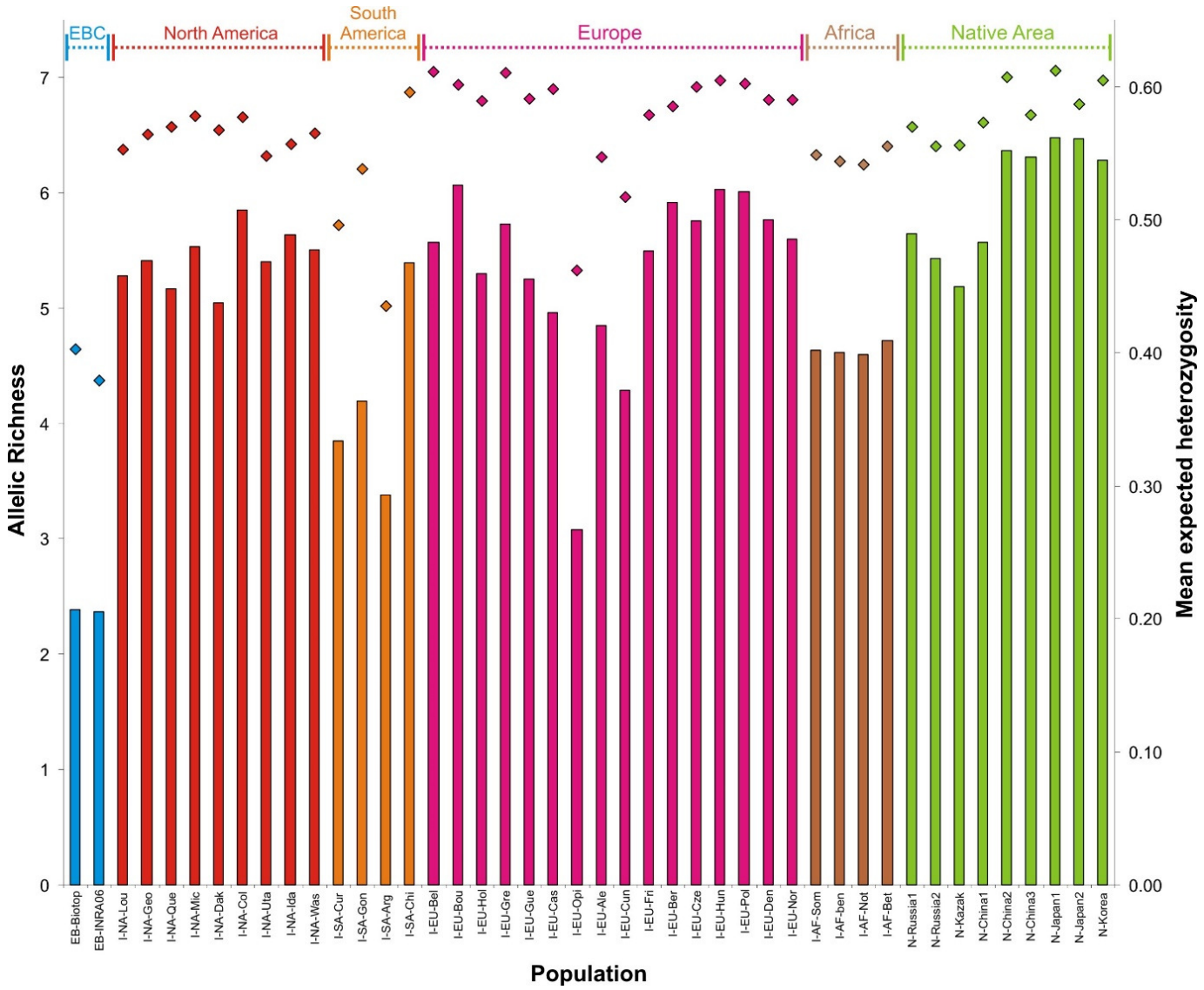


Figure S2: Genetic diversity in the biocontrol, invasive and native site sample of *Harmonia axyridis*.
Mean expected heterozygosity (diamonds) and average genetic diversity estimated as allelic richness at 18 microsatellite loci corrected for 19 individuals (bars) are shown. Names of site samples given at bottom are as in Fig. 1 and Table S1. Colors correspond to the continental origin or status (for native site samples and EBC = European biocontrol).

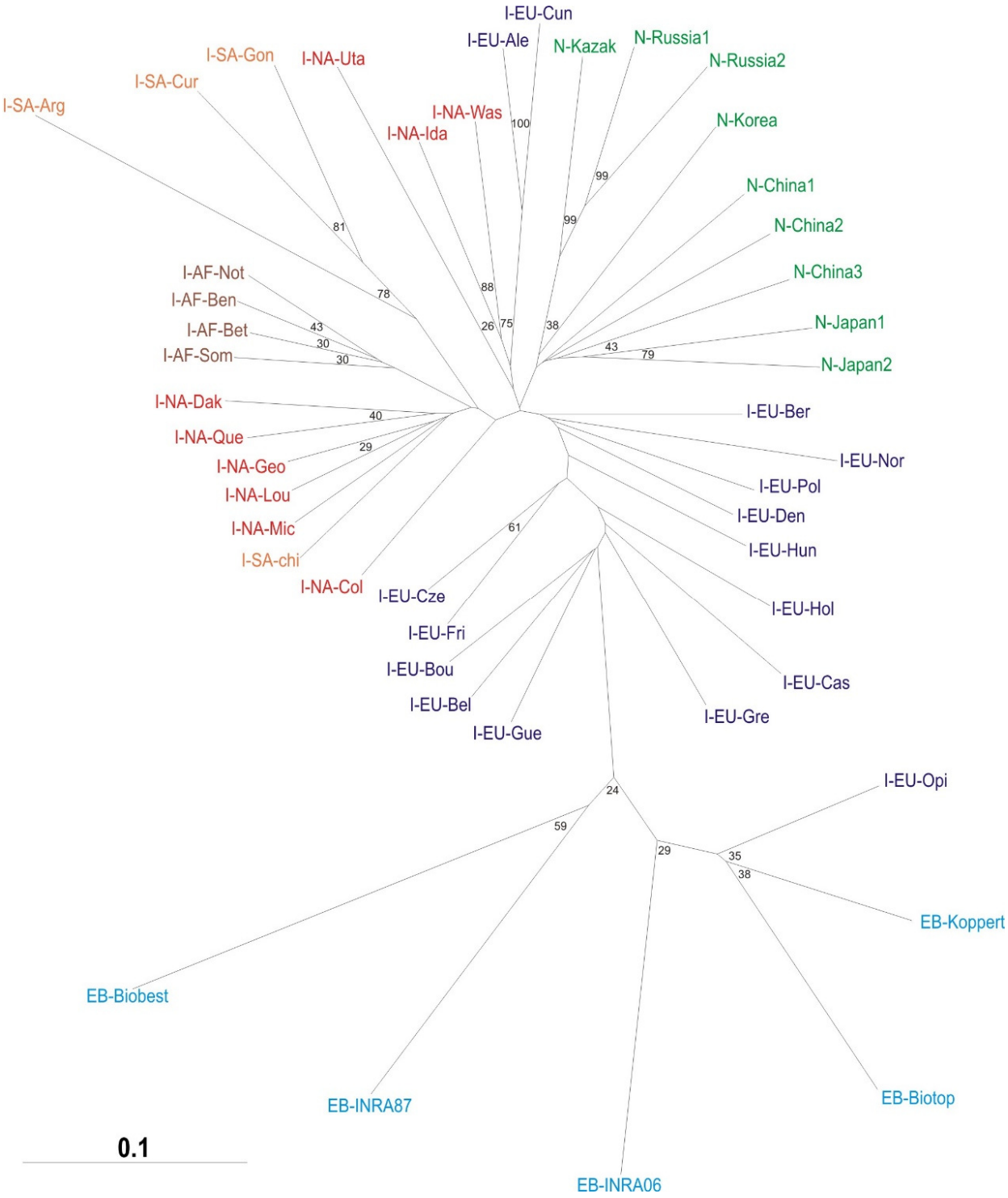


Figure S3: Neighbor-joining tree for the studied *Harmonia axyridis* site samples based on the distance of Cavalli-Sforza & Edwards (1967, *American Journal of Human Genetics*, 19:233-257).
The position on the tree of most of the site samples was geographically consistent. Some distinctive patterns emerge from the tree. In North America, the nine samples constitute two groups, with two intermediate samples (I-NA-Col and I-NA-Uta). These two groups clearly correspond to an east versus west geographical pattern (Fig. 1). In South America, the Chilean sample grouped within the set of Eastern North American samples, whereas the Argentinean and Brazilian samples grouped together, albeit on long branches. In Europe, most samples (13 of 16) were located in a similar position in the tree, between all the other feral and biocontrol samples. One sample from the South of France (I-EU-Opi) grouped with the biocontrol samples, and the two Italian samples (I-EU-Cun and I-EU-Ale) grouped with the most western North American site samples. Finally, all African samples are grouped together.
Note: code names of site samples are as in Fig. 1 and Table S1. Different colors correspond to different continents or status: native area in green, North America in red, South America in orange, Africa in brown, Europe in purple and European biocontrol in blue. Bootstrap values calculated over 1,000 replications are given as percentages (only values > 20% are shown).

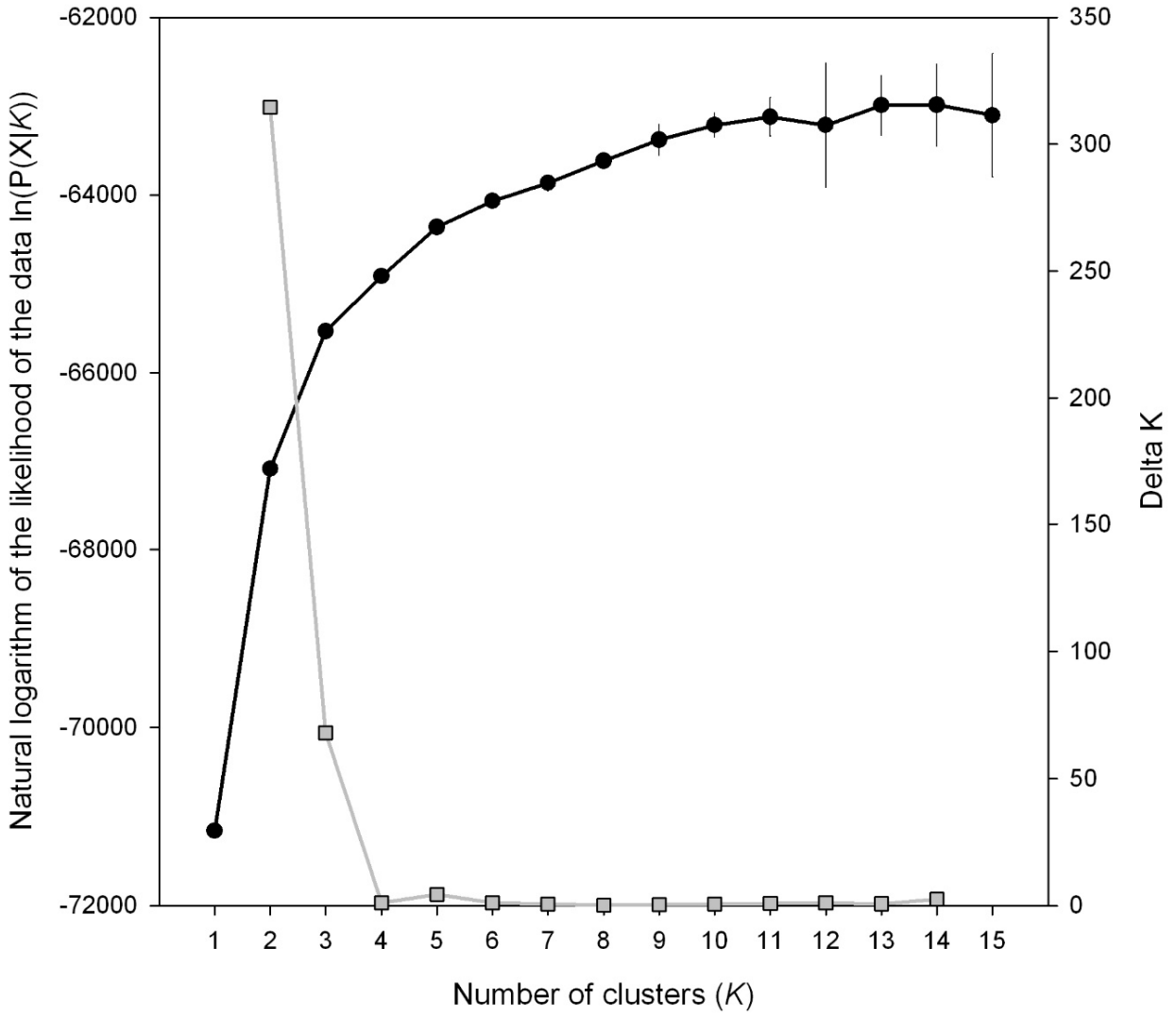


Figure S4: Likelihood of the multilocus genetic data as a function of the number of genetic clusters among all 47 site samples of *Harmonia axyridis*.
The mean (± SD) natural logarithm of the likelihood of the data (LnP(X|*K*)) calculated over 20 replicated STRUCTURE runs is given for each value of the putative number of clusters (*K*), from 1 to 15 (black circles). Grey squares are the delta *K* values computed as in Evanno *et al.* (2005, *Molecular Ecology*, 14:2611-2620). Note that the highest value of *K* in our analyses was 15 because results suggested that no further relevant genetic information was found in our dataset for higher values of *K*: (i) the Evanno delta *K* value was the highest for *K*=2, (ii) ln(P(X|*K*)) values started to level off around *K*=10, (iii) standard deviation of ln(P(X|*K*)) became very high above *K*=10, and (iv) similarity coefficient between runs started to decrease strongly above *K*=6.

Site sample name	Sampling location and historical information	Geographic coordinates	Sampling date (month-year)	Number of genotyped individuals
EB-Koppert (*)	European Biocontrol: rearing stock, Koppert biofactory Historically: strain obtained by Koppert from EB-Biotop in 1997	-	07-2003	20
EB-INRA87 (*)	European Biocontrol: rearing stock, INRA laboratory History: descendent from a population sampled in China in 1982	-	04-1987	18
EB-Biobest (*)	European Biocontrol: rearing stock, Ghent University laboratory Historically: strain obtained by Biobest from EB-Biotop in 1997	-	04-2007	27
EB-Biotop (*)	European Biocontrol: rearing stock, Biotop biofactory Historically: strain obtained by Biotop from EB-INRA87 in 1995	-	11-2007	29
EB-INRA06 (*)	European Biocontrol: rearing stock, INRA laboratory Historically: direct descendant of EB-INRA87	-	11-2006	27
I-NA-Lou (*)	North American invasive area: Joyce, Louisiana, USA	31.94°N 92.60°W	11-2007	34
I-NA-Geo	North American invasive area: Plains, Georgia, USA	32.03°N 84.39°W	09-2007	31
I-NA-Que	North American invasive area: Quebec, Quebec, Canada	46.82°N 71.25°W	06-2007	30
I-NA-Mic	North American invasive area: Lowel, Michigan, USA	42.97°N 85.35°W	10-2007	31
I-NA-Dak	North American invasive area: Brookings, South Dakota, USA	44.35°N 96.81°W	09-2006	30
I-NA-Col	North American invasive area: Fort Collins, Colorado, USA	40.57°N 105.08°W	09-2008	30
I-NA-Uta	North American invasive area: Logan, Utah, USA	41.74°N 111.83°W	06-2007	32
I-NA-Ida	North American invasive area: Caldwell, Idaho, USA	43.65°N 116.68°W	07-2007	31
I-NA-Was (*)	North American invasive area: Sunnyside, Washington, USA	46.32°N 120.01°W	09-2007	42
I-SA-Cur (*)	South American invasive area: Curitiba, Brazil	25.45°S 49.24°W	02-2008	30
I-SA-Gon	South American invasive area: Bento Gonçalves, Brazil	29.16°S 51.53°W	02-2008	30
I-SA-Arg	South American invasive area: Rafaela, Argentina	31.25°S 61.48°W	01-2006	42
I-SA-Chi	South American invasive area: Huechuraba, Chile	33.37°S 70.63°W	05-2010	31
I-EU-Bel (*)	European invasive area: Ghent, Belgium	51.05°N 3.71°E	10-2007	31
I-EU-Bou	European invasive area: Boult-aux-Bois, France	49.43°N 4.84°E	01-2007	30
I-EU-Hol	European invasive area: Apeldoorn, Holland	52.10°N 6.02°E	11-2008	31
I-EU-Gre	European invasive area: Grenoble, France	45.19°N 5.74°E	01-2008	29
I-EU-Gue	European invasive area: Le Guedeniau, France	47.50°N 0.02°W	11-2008	31
I-EU-Cas	European invasive area: Castries, France	43.68°N 3.99°E	10-2008	31
I-EU-Opi	European invasive area: Opio, France	43.67°N 6.99°E	09-2006	33
I-EU-Ale	European invasive area: Alessandria, Italia	44.86°N 8.60°E	08-2009	31

Site sample name	Sampling location and historical information	Geographic coordinates	Sampling date (month-year)	Number of genotyped individuals
I-EU-Cun	European invasive area: Cuneo, Italia	44.47°N 7.57°E	08-2009	31
I-EU-Fri	European invasive area: Friedberg, Germany	50.33°N 8.76°E	10-2008	31
I-EU-Ber	European invasive area: Berlin, Germany	52.52°N 13.41°E	07-2008	42
I-EU-Cze	European invasive area: Ceske Budejovice, Czech Republic	48.97°N 14.47°E	10-2007	31
I-EU-Hun	European invasive area: Bataszek, Hungary	46.15°N 18.69°E	10-2008	31
I-EU-Pol	European invasive area: Torun, Poland	53.01°N 18.60°E	06-2009	31
I-EU-Den	European invasive area: Copenhagen, Denmark	55.69°N 12.55°E	01-2008	38
I-EU-Nor	European invasive area: Oslo, Norway	59.95°N 10.73°E	07-2008	20
I-AF-Som (*)	African invasive area: Somerset West, South Africa	34.03°S 18.83°E	05-2008	31
I-AF-Ben	African invasive area: Benoni, South Africa	26.17°S 28.33°E	04-2008	31
I-AF-Not	African invasive area: Nottingham road, South Africa	29.37°S 29.83°E	05-2008	31
I-AF-Bet	African invasive area: Bethlehem, South Africa	28.25°S 28.32°E	05-2008	31
N-Russia1 (*)	Native area: Abakan, Khakassia, Russia	53.73°N 91.46°N	10-2007	31
N-Russia2 (*)	Native area: Novosibirsk, Novosibirsk Oblast, Russia	55.04°N 82.93°E	10-2007	30
N-Kazak (*)	Native area: Almaty, Oblys d'Almaty, Kazakhstan	43.24°N 76.95°E	10-2007	26
N-China1 (*)	Native area: Beijing, China	40.24°N 116.23°E	05-2007	28
N-China2 (*)	Native area: Shilin city, Yunnan Province, China	24.90°N 103.35°E	08-2007	35
N-China3 (*)	Native area: Changchun city, Jilin Province, China	43.88°N 125.31°E	11-2006	29
N-Japan1 (*)	Native area: Fuchu, Japan	34.57°N 133.24°E	09-2005	36
N-Japan2 (*)	Native area: Kyoto, Japan	35.01°N 135.77°E	08-2008	26
N-Korea (*)	Native area: Daejeon, South Korea	36.37°N 127.35°E	11-1998	30
TOTAL				1442

Table S1: Biocontrol, invasive and native site samples of *Harmonia axyridis*.
In the "Site sample name" column, "*" indicates that the corresponding site sample was previously used by Lombaert *et al.* (2011, *Molecular Ecology*, 20:4654-4670). See Fig. 1 for a map of sampling localities.

Parameters	Distribution	Mean	Median	Mode	Quantile 2.5%	Quantile 97.5%
NW; NA	Uniform [100 – 20,000]	10,050	10,050	NA	590	19,510
NB	Loguniform [10 – 1,000]	213	97	10	11	891
NF_i	Loguniform [2 – 1,000]	154	39	2	2	850
BD_i	Uniform [0 – 5]	2.5	2.5	NA	0	5
$ar_{j,k,i}$	Uniform [0.1 – 0.9]	0.5	0.5	NA	0.12	0.88
t_i	Uniform [x_i – x_i+5]	DV	DV	NA	DV	DV
$tu_{n,i}$	Loguniform [t_i – 3000]	DV	DV	DV	DV	DV
t_{anc}	Uniform [100 – 3000]	1,550	1,550	NA	172	2,928
mean μ	Uniform [10^{-5} – 10^{-3}]	$5.0x10^{-4}$	$5.0x10^{-4}$	NA	$3.4x10^{-5}$	$9.8x10^{-4}$
mean *P*	Uniform [0.1 – 0.3]	0.2	0.2	NA	0.10	0.30
mean μSNI	Uniform [10^{-8} – 10^{-4}]	$5.0x10^{-5}$	$5.0x10^{-5}$	NA	$2.5x10^{-6}$	$9.7x10^{-5}$

Table S2: Prior distributions of demographic, historical and genetic parameters used in all ABC analyses processed to retrace the worldwide routes of invasion of *H. axyridis*.

Note: *NW*, *NA* and *NB* = stable effective population size (number of diploid individuals) of wild (native or invasive: *NW*), ancestral (*NA*) or biocontrol (*NB*) populations; NF_i = effective number of founders during an introduction step lasting BD_i generation(s) for invasive population *i*; $ar_{j,k,i}$ = genetic contribution (admixture rate) of population *j* when the invasive population *i* is formed by the admixture between the putative source populations *j* and *k*; t_i = introduction time of invasive populations *i* with bounds x_i fixed from dates of first observation of established population (as in Brown *et al.* 2011, *Biocontrol*, 56:623-641); $tu_{n,i}$ = merging time of the source unsampled native population into the sampled native population *n*, when this native population is or is supposed to be the source of the invasive or biocontrol population *i* (see Lombaert *et al.* 2011, Molecular Ecology, 20:4654-4670); t_{anc} = merging time of the two native populations into an ancestral unsampled one (with condition $tu_j \leq t_{anc}$). All times were expressed in numbers of generation assuming 2.5 generations per year and running back in time. The microsatellite loci were assumed to follow a generalized stepwise mutation model (Estoup *et al.* 2002, *Molecular Ecology*, 11:1591-1604) with three parameters: the mean mutation rate (mean μ), the mean parameter of the geometric distribution (mean *P*) of length in terms of the number of repeats of mutation events and the mean mutation rate for single nucleotide instability (mean μSNI). Each locus has a possible range of 40 contiguous allelic states and is characterized by individual μ_{loc} drawn from a Gamma (mean = mean μ and shape = 2), P_{loc} drawn from a Gamma (mean = mean *P* and shape = 2) and μSNI_{loc} drawn from a Gamma (mean = mean μSNI and shape = 2) distribution. For a loguniform[x;y] distribution, Log(x) and Log(y) are the bounds of a uniform distribution. All prior quantities presented were computed from 10^6 values. NA = not applicable; DV = can take different values. See Fig. S1 for a graphical representation of several evolutionary scenarios with associated parameters considered in the ABC analyse

SITE SAMPLE	EB-Koppert	EB-INRA87	EB-Biobest	EB-Biotop	EB-INRA06	I-NA-Lou	I-NA-Geo	I-NA-Que	I-NA-Mic	I-NA-Dak	I-NA-Col	I-NA-Uta	I-NA-Ida	I-NA-Was	I-SA-Cur	I-SA-Gon	I-SA-Arg	I-SA-Chi	I-EU-Bel	I-EU-Bou	I-EU-Hol	I-EU-Gre	I-EU-Gue	I-EU-Cas	I-EU-Opi	I-EU-Ale	I-EU-Cun	I-EU-Fri	I-EU-Ber	I-EU-Cze	I-EU-Hun	I-EU-Pol	I-EU-Den	I-EU-Nor	I-AF-Som	I-AF-ben	I-AF-Not	I-AF-Bet	N-Russia1	N-Russia2	N-Kazak	N-China1	N-China2	N-China3	N-Japan1	N-Japan2	
EB-INRA87	0,140																																														
EB-Biobest	0,312	0,243																																													
EB-Biotop	0,062	0,165	0,332																																											
EB-INRA06	0,174	0,231	0,381	0,218																																										
I-NA-Lou	0,207	0,188	0,335	0,224	0,255																																										
I-NA-Geo	0,198	0,180	0,335	0,223	0,245	**-0,001**																																									
I-NA-Que	0,203	0,183	0,339	0,219	0,252	**0,003**	**0,002**																																								
I-NA-Mic	0,191	0,175	0,323	0,211	0,239	**0,002**	**0,000**	0,004																																							
I-NA-Dak	0,214	0,191	0,336	0,234	0,272	0,012	0,011	0,008	0,003																																						
I-NA-Col	0,190	0,172	0,336	0,214	0,242	0,010	**0,012**	0,009	**0,007**	0,014																																					
I-NA-Uta	0,196	0,214	0,366	0,218	0,247	0,056	0,053	0,046	0,059	0,067	0,048																																				
I-NA-Ida	0,206	0,185	0,321	0,220	0,256	0,023	0,033	0,020	0,027	0,036	0,012	0,045																																			
I-NA-Was	0,197	0,184	0,325	0,206	0,239	0,023	0,030	0,023	0,030	0,035	0,017	0,043	**0,005**																																		
I-SA-Cur	0,248	0,279	0,418	0,281	0,312	0,064	0,057	0,063	0,057	0,074	0,068	0,115	0,111	0,107																																	
I-SA-Gon	0,213	0,232	0,393	0,240	0,273	0,038	0,032	0,035	0,038	0,056	0,046	0,068	0,076	0,064	0,034																																	
I-SA-Arg	0,299	0,312	0,439	0,323	0,350	0,144	0,131	0,146	0,136	0,163	0,145	0,160	0,154	0,158	0,174	0,095																															
I-SA-Chi	0,178	0,164	0,320	0,207	0,226	0,008	**0,003**	0,008	**0,003**	0,014	0,010	0,056	0,027	0,026	0,058	0,037	0,147																															
I-EU-Bel	0,081	0,097	0,230	0,120	0,160	0,041	0,039	0,039	0,033	0,047	0,038	0,063	0,050	0,053	0,101	0,073	0,166	0,033																													
I-EU-Bou	0,112	0,132	0,294	0,148	0,165	0,028	0,024	0,027	0,021	0,034	0,022	0,058	0,037	0,037	0,093	0,064	0,156	0,019	0,008																													
I-EU-Hol	0,074	0,114	0,267	0,119	0,152	0,052	0,046	0,049	0,045	0,059	0,049	0,075	0,056	0,055	0,122	0,083	0,182	0,038	**0,001**	**0,005**																												
I-EU-Gre	0,078	0,113	0,274	0,122	0,145	0,048	0,040	0,043	0,038	0,053	0,034	0,056	0,058	0,051	0,096	0,068	0,171	0,032	**0,001**	**0,006**	**0,006**																												
I-EU-Gue	0,072	0,126	0,275	0,123	0,151	0,049	0,042	0,047	0,039	0,053	0,047	0,068	0,062	0,060	0,104	0,074	0,168	0,036	**0,002**	**0,005**	**0,002**	**0,002**																										
I-EU-Cas	0,097	0,131	0,252	0,135	0,189	0,050	0,045	0,052	0,044	0,062	0,056	0,068	0,063	0,062	0,104	0,080	0,174	0,043	0,012	0,026	0,015	0,015	0,018																								
I-EU-Opi	0,032	0,138	0,268	0,099	0,155	0,221	0,214	0,211	0,202	0,225	0,203	0,202	0,212	0,209	0,271	0,235	0,311	0,191	0,094	0,129	0,090	0,094	0,089	0,107																							
I-EU-Ale	0,210	0,180	0,334	0,231	0,240	0,046	0,043	0,037	0,049	0,053	0,039	0,058	0,022	0,022	0,134	0,087	0,182	0,039	0,065	0,051	0,059	0,062	0,067	0,077	0,217																						
I-EU-Cun	0,222	0,208	0,361	0,226	0,262	0,048	0,050	0,044	0,051	0,062	0,043	0,057	0,026	0,022	0,139	0,088	0,175	0,050	0,078	0,063	0,072	0,071	0,077	0,087	0,231	**0,003**																					
I-EU-Fri	0,144	0,159	0,308	0,165	0,212	0,018	0,009	0,012	0,010	0,014	0,016	0,046	0,031	0,026	0,065	0,043	0,137	0,013	0,014	0,009	0,022	0,016	0,020	0,026	0,162	0,044	0,053																				
I-EU-Ber	0,177	0,167	0,306	0,196	0,240	0,009	0,014	0,012	0,007	0,017	0,008	0,049	0,018	0,020	0,063	0,051	0,144	0,009	0,034	0,027	0,047	0,037	0,044	0,041	0,188	0,044	0,045	0,011																			
I-EU-Cze	0,116	0,134	0,280	0,152	0,191	0,021	0,013	0,018	0,011	0,019	0,013	0,042	0,028	0,028	0,070	0,044	0,128	0,015	0,006	**0,005**	0,014	0,008	0,009	0,018	0,132	0,040	0,048	**-0,002**	0,011																	
I-EU-Hun	0,126	0,125	0,286	0,159	0,183	0,013	0,013	0,012	0,012	0,025	0,009	0,044	0,023	0,021	0,070	0,039	0,126	0,009	0,010	0,006	0,014	0,012	0,011	0,019	0,141	0,031	0,040	0,006	0,009	**0,000**																
I-EU-Pol	0,146	0,150	0,299	0,173	0,198	0,022	0,016	0,016	0,011	0,022	0,009	0,042	0,022	0,021	0,071	0,048	0,133	0,008	0,020	0,015	0,031	0,022	0,026	0,039	0,165	0,035	0,044	0,005	0,008	0,004	**0,004**														
I-EU-Den	0,131	0,140	0,303	0,155	0,202	0,024	0,025	0,018	0,021	0,027	0,008	0,041	0,020	0,015	0,092	0,053	0,145	0,020	0,019	0,011	0,020	0,016	0,022	0,030	0,146	0,036	0,038	0,009	0,011	**0,003**	**0,001**	0,010													
I-EU-Nor	0,141	0,147	0,310	0,171	0,228	0,044	0,045	0,041	0,032	0,044	0,031	0,059	0,035	0,032	0,107	0,076	0,168	0,028	0,030	0,037	0,032	0,035	0,039	0,033	0,155	0,051	0,058	0,025	0,013	0,020	0,012	0,009	0,010											
I-AF-Som	0,209	0,188	0,337	0,221	0,286	0,023	0,021	0,024	0,021	0,027	0,017	0,079	0,038	0,037	0,089	0,072	0,172	0,023	0,052	0,037	0,065	0,059	0,066	0,066	0,227	0,057	0,069	0,023	0,019	0,029	0,031	0,029	0,033	0,054										
I-AF-ben	0,221	0,208	0,350	0,243	0,281	0,020	0,014	0,022	0,017	0,027	0,027	0,069	0,040	0,037	0,081	0,059	0,150	0,025	0,054	0,041	0,066	0,063	0,065	0,059	0,236	0,061	0,067	0,022	0,024	0,030	0,030	0,032	0,033	0,052	**0,007**									
I-AF-Not	0,208	0,202	0,344	0,228	0,274	0,025	0,026	0,027	0,024	0,033	0,031	0,066	0,047	0,045	0,084	0,071	0,167	0,032	0,052	0,038	0,069	0,063	0,059	0,060	0,221	0,071	0,078	0,027	0,026	0,033	0,031	0,032	0,036	0,051	0,009	**0,001**								
I-AF-Bet	0,193	0,200	0,351	0,219	0,269	0,017	0,013	0,017	0,011	0,019	0,013	0,055	0,035	0,035	0,064	0,048	0,144	0,018	0,044	0,027	0,056	0,048	0,046	0,054	0,214	0,059	0,068	0,013	0,015	0,019	0,019	0,017	0,023	0,040	**0,005**	**0,001**	**0,000**							
N-Russia1	0,202	0,178	0,313	0,226	0,262	0,036	0,032	0,041	0,036	0,038	0,025	0,081	0,033	0,030	0,122	0,080	0,168	0,033	0,054	0,043	0,062	0,058	0,068	0,069	0,215	0,050	0,058	0,037	0,039	0,031	0,038	0,033	0,034	0,056	0,031	0,039	0,049	0,042						
N-Russia2	0,213	0,174	0,312	0,235	0,277	0,041	0,049	0,051	0,048	0,047	0,035	0,095	0,036	0,038	0,143	0,103	0,185	0,045	0,063	0,053	0,075	0,072	0,080	0,082	0,219	0,055	0,063	0,050	0,045	0,043	0,047	0,045	0,043	0,068	0,041	0,055	0,060	0,058	**-0,006**					
N-Kazak	0,208	0,182	0,325	0,224	0,271	0,028	0,034	0,026	0,032	0,029	0,020	0,060	0,018	0,019	0,116	0,077	0,175	0,026	0,048	0,044	0,062	0,057	0,066	0,065	0,214	0,036	0,043	0,030	0,027	0,029	0,030	0,024	0,025	0,041	0,034	0,042	0,047	0,043	**0,004**	**0,003**				
N-China1	0,211	0,179	0,319	0,233	0,261	0,029	0,036	0,032	0,029	0,040	0,022	0,058	0,012	0,026	0,114	0,087	0,174	0,033	0,050	0,044	0,064	0,056	0,068	0,060	0,214	0,037	0,037	0,036	0,024	0,031	0,038	0,031	0,032	0,052	0,039	0,046	0,053	0,047	0,033	0,035	0,024			
N-China2	0,168	0,149	0,305	0,190	0,216	0,017	0,024	0,021	0,023	0,032	0,008	0,040	0,009	0,008	0,096	0,061	0,155	0,020	0,035	0,025	0,041	0,033	0,044	0,044	0,179	0,026	0,027	0,021	0,015	0,017	0,011	0,016	0,006	0,025	0,032	0,030	0,031	0,027	0,021	0,027	0,015	0,010		
N-China3	0,182	0,165	0,309	0,196	0,240	0,012	0,018	0,017	0,020	0,022	0,011	0,047	0,010	0,009	0,097	0,061	0,150	0,016	0,036	0,025	0,042	0,035	0,046	0,045	0,194	0,029	0,028	0,017	0,013	0,016	0,018	0,017	0,011	0,035	0,030	0,035	0,038	0,032	0,021	0,027	0,013	0,009	0,002	
N-Japan1	0,170	0,164	0,308	0,203	0,225	0,016	0,023	0,022	0,022	0,033	0,016	0,044	0,013	0,011	0,097	0,062	0,161	0,018	0,034	0,025	0,043	0,035	0,046	0,047	0,185	0,029	0,032	0,025	0,016	0,018	0,018	0,020	0,013	0,033	0,034	0,035	0,040	0,035	0,020	0,026	0,015	0,007	0,000	0,005	
N-Japan2	0,182	0,161	0,318	0,214	0,231	0,022	0,029	0,024	0,027	0,033	0,012	0,048	0,014	0,016	0,101	0,065	0,155	0,025	0,035	0,024	0,045	0,037	0,045	0,057	0,188	0,031	0,035	0,027	0,026	0,020	0,021	0,022	0,018	0,045	0,042	0,044	0,044	0,039	0,019	0,023	0,012	0,007	0,006	**0,004**	**0,002**	
N-Korea	0,181	0,157	0,319	0,205	0,230	0,034	0,031	0,033	0,029	0,037	0,012	0,054	0,029	0,019	0,103	0,066	0,153	0,024	0,044	0,034	0,058	0,034	0,057	0,059	0,184	0,039	0,043	0,029	0,023	0,021	0,024	0,021	0,017	0,038	0,030	0,040	0,050	0,039	0,015	0,022	0,014	0,018	0,008	0,012	0,005	0,008

Table S3: Pairwise estimates of FST between all pairs of *Harmonia axyridis* site samples. Site sample names are as in Fig.1 and Table S1. FST in blue and bold typeface indicates non significant pairwise differentiation, as assessed using Fisher's exact test with correction for multiple comparisons.

Continent	Target genetic unit	Target site sample	Number of competing scenarios	Source sample set	Selected scenario (cf. highest probability)	2nd scenario	3rd scenario	4th scenario	5th scenario
North America	Colorado	I-NA-Col	10	"Reference"	ENA + WNA 0.358 [0.343,0.372]	ENA 0.261 [0.248,0.274]	ENA + WNat 0.176 [0.166,0.185]	ENat + ENA 0.150 [0.142,0.159]	WNat + WNA 0.021 [0.019,0.023]
				"High-F_{ST}"	ENA + WNA 0.834 [0.822,0.845]	ENat + ENA 0.082 [0.074,0.090]	ENat + WNA 0.029 [0.025,0.032]	ENA 0.014 [0.012,0.016]	ENA + WNA 0.013 [0.011,0.015]
				"Low-F_{ST}"	ENA + WNA 0.551 [0.530,0.573]	ENA 0.303 [0.284,0.322]	ENat + ENA 0.107 [0.099,0.115]	ENA + WNat 0.037 [0.033,0.041]	WNat + WNA 0.000 [0.000,0.001]
				"Pool-high-F_{ST}"	ENA + WNA 0.774 [0.756,0.791]	ENat + ENA 0.097 [0.088,0.107]	ENA 0.090 [0.079,0.102]	ENA + WNat 0.021 [0.018,0.024]	ENat + WNA 0.008 [0.006,0.009]
				"Pool-low-F_{ST}"	ENA + WNA 0.805 [0.783,0.827]	ENA 0.163 [0.142,0.185]	ENat + ENA 0.021 [0.018,0.023]	ENA + WNat 0.009 [0.007,0.010]	WNat + WNA 0.001 [0.001,0.001]
	Utah	I-NA-Uta	10	"Reference"	ENA + WNA 0.768 [0.748,0.788]	WNA 0.077 [0.063,0.090]	ENat + WNA 0.047 [0.041,0.054]	ENat + ENA 0.041 [0.036,0.046]	WNat + WNA 0.031 [0.027,0.036]
				"High-F_{ST}"	ENA + WNA 0.419 [0.387,0.450]	WNA 0.352 [0.319,0.385]	ENat + WNA 0.141 [0.127,0.155]	ENat 0.033 [0.029,0.037]	ENat + ENA 0.028 [0.024,0.032]
				"Low-F_{ST}"	ENA + WNA 0.839 [0.820,0.858]	WNA 0.060 [0.047,0.073]	ENat + WNA 0.057 [0.048,0.066]	ENat + ENA 0.025 [0.021,0.028]	ENA 0.011 [0.009,0.013]
				"Pool-high-F_{ST}"	ENA + WNA 0.685 [0.660,0.710]	WNA 0.118 [0.100,0.137]	ENat + WNA 0.111 [0.098,0.123]	ENat + ENA 0.053 [0.046,0.059]	ENat 0.023 [0.020,0.026]
				"Pool-low-F_{ST}"	ENA + WNA 0.811 [0.781,0.841]	WNA 0.123 [0.095,0.150]	ENat + WNA 0.036 [0.030,0.043]	ENat + ENA 0.019 [0.016,0.023]	ENat 0.006 [0.005,0.007]
South America	Brazil	I-SA-Cur	28	"Reference"	ENA 0.414 [0.380,0.448]	ENA + AF 0.359 [0.327,0.391]	AF 0.203 [0.178,0.227]	WNat + ENA 0.006 [0.005,0.007]	ENA + WNA 0.005 [0.004,0.006]
				"High-F_{ST}"	ENA + AF 0.486 [0.446,0.527]	ENA 0.337 [0.299,0.376]	AF 0.150 [0.127,0.173]	ENA + Europe 0.005 [0.004,0.006]	ENA + WNA 0.005 [0.003,0.006]
				"Low-F_{ST}"	ENA + AF 0.422 [0.393,0.451]	AF 0.334 [0.306,0.362]	ENA 0.237 [0.215,0.260]	WNA + AF 0.001 [0.001,0.002]	ENat + AF 0.001 [0.001,0.002]
				"Pool-high-F_{ST}"	ENA 0.447 [0.402,0.492]	ENA + AF 0.333 [0.294,0.373]	AF 0.197 [0.165,0.228]	ENA + WNA 0.005 [0.004,0.006]	ENat + AF 0.003 [0.003,0.004]
				"Pool-low-F_{ST}"	ENA 0.437 [0.401,0.473]	ENA + AF 0.354 [0.321,0.387]	AF 0.179 [0.155,0.202]	ENA + WNA 0.008 [0.007,0.010]	ENat + ENA 0.005 [0.004,0.006]

Continent	Target genetic unit	Target site sample	Number of competing scenarios	Source sample set	Selected scenario (cf. highest probability)	2nd scenario	3rd scenario	4th scenario	5th scenario
		I-SA-Gon	28	"Reference"	ENA 0.807 [0.783,0.830]	ENA + AF 0.114 [0.097,0.131]	AF 0.045 [0.037,0.054]	ENA + WNA 0.013 [0.010,0.016]	ENat + ENA 0.010 [0.008,0.012]
				"High-F_{ST}"	ENA 0.403 [0.365,0.441]	ENA + AF 0.265 [0.234,0.297]	AF 0.248 [0.214,0.281]	WNat + ENA 0.022 [0.018,0.026]	WNat + AF 0.013 [0.010,0.015]
				"Low-F_{ST}"	ENA 0.504 [0.459,0.548]	ENA + AF 0.384 [0.341,0.427]	AF 0.077 [0.063,0.090]	ENA + WNA 0.021 [0.016,0.025]	ENat + ENA 0.004 [0.003,0.005]
				"Pool-high-F_{ST}"	ENA 0.767 [0.743,0.791]	ENA + AF 0.159 [0.139,0.179]	AF 0.039 [0.033,0.045]	WNat + ENA 0.016 [0.013,0.019]	WNat + AF 0.007 [0.006,0.009]
				"Pool-low-F_{ST}"	ENA 0.825 [0.797,0.854]	ENA + AF 0.122 [0.099,0.145]	AF 0.037 [0.028,0.045]	WNat + AF 0.006 [0.004,0.007]	WNat + ENA 0.005 [0.003,0.006]
	Argentina	I-SA-Arg	28	"Reference"	ENA 0.482 [0.427,0.538]	ENA + WNA 0.161 [0.128,0.194]	ENat + ENA 0.105 [0.083,0.127]	AF 0.083 [0.058,0.107]	ENA + AF 0.076 [0.055,0.097]
				"High-F_{ST}"	WNA + AF 0.591 [0.541,0.641]	ENA + WNA 0.340 [0.295,0.386]	ENat + AF 0.013 [0.010,0.017]	AF 0.010 [0.006,0.013]	ENat + ENA 0.008 [0.007,0.010]
				"Low-F_{ST}"	ENA 0.574 [0.532,0.615]	ENA + AF 0.148 [0.119,0.177]	AF 0.100 [0.078,0.123]	ENA + WNA 0.097 [0.079,0.115]	ENat + ENA 0.033 [0.026,0.039]
				"Pool-high-F_{ST}"	ENA + WNA 0.411 [0.364,0.459]	WNA + AF 0.261 [0.216,0.307]	AF 0.130 [0.094,0.165]	ENA 0.109 [0.087,0.131]	ENA + AF 0.048 [0.035,0.062]
				"Pool-low-F_{ST}"	ENA 0.430 [0.385,0.474]	ENA + WNA 0.204 [0.170,0.239]	AF 0.145 [0.113,0.177]	ENA + AF 0.118 [0.094,0.141]	WNA + AF 0.065 [0.050,0.080]
	Chile	I-SA-Chi	28	"Reference"	ENA 0.509 [0.470,0.548]	ENA + AF 0.187 [0.162,0.212]	WNat + ENA 0.174 [0.148,0.199]	WNat + AF 0.036 [0.030,0.042]	ENA + Europe 0.032 [0.026,0.039]
				"High-F_{ST}"	ENA 0.406 [0.368,0.445]	ENA + Europe 0.277 [0.241,0.313]	ENA + AF 0.134 [0.115,0.153]	Europe + AF 0.037 [0.030,0.043]	EBC + ENA 0.036 [0.028,0.044]
				"Low-F_{ST}"	ENA 0.579 [0.556,0.602]	ENA + AF 0.289 [0.268,0.309]	ENA + Europe 0.076 [0.067,0.085]	AF 0.024 [0.021,0.027]	Europe + AF 0.012 [0.011,0.014]
				"Pool-high-F_{ST}"	ENA 0.713 [0.683,0.742]	ENA + AF 0.206 [0.180,0.231]	ENA + Europe 0.041 [0.033,0.048]	AF 0.012 [0.009,0.014]	WNat + ENA 0.008 [0.007,0.009]
				"Pool-low-F_{ST}"	ENA 0.713 [0.686,0.741]	ENA + AF 0.191 [0.168,0.214]	ENA + Europe 0.048 [0.039,0.056]	AF 0.018 [0.015,0.022]	ENA + WNA 0.008 [0.006,0.010]

Continent	Target genetic unit	Target site sample	Number of competing scenarios	Source sample set	Selected scenario (cf. highest probability)	2nd scenario	3rd scenario	4th scenario	5th scenario
Europe	West Europe	I-EU-Bel	28	"Reference"	EBC + ENA 0.780 [0.729,0.831]	EBC + AF 0.150 [0.104,0.197]	EBC + SA 0.031 [0.018,0.044]	ENat + EBC 0.025 [0.016,0.033]	WNat + EBC 0.010 [0.007,0.014]
				"High-F_{ST}"	EBC + ENA 0.508 [0.433,0.582]	EBC + AF 0.321 [0.244,0.397]	ENat + EBC 0.134 [0.090,0.177]	EBC + SA 0.017 [0.009,0.026]	WNat + EBC 0.014 [0.009,0.019]
				"Low-F_{ST}"	EBC + ENA 0.720 [0.677,0.762]	EBC + AF 0.199 [0.158,0.239]	ENat + EBC 0.058 [0.043,0.072]	EBC + SA 0.016 [0.011,0.021]	WNat + EBC 0.007 [0.005,0.008]
				"Pool-high-F_{ST}"	EBC + ENA 0.651 [0.604,0.699]	EBC + AF 0.278 [0.230,0.326]	EBC + SA 0.035 [0.025,0.045]	ENat + EBC 0.029 [0.022,0.037]	WNat + EBC 0.004 [0.003,0.005]
				"Pool-low-F_{ST}"	EBC + ENA 0.698 [0.651,0.744]	EBC + AF 0.148 [0.113,0.183]	ENat + EBC 0.132 [0.099,0.165]	EBC + SA 0.016 [0.011,0.021]	WNat + EBC 0.002 [0.001,0.003]
		I-EU-Cas	28	"Reference"	EBC + ENA 0.541 [0.485,0.598]	EBC + SA 0.274 [0.219,0.328]	EBC + AF 0.169 [0.136,0.203]	WNat + EBC 0.005 [0.003,0.007]	ENat + EBC 0.004 [0.003,0.006]
				"High-F_{ST}"	EBC + AF 0.623 [0.568,0.673]	EBC + ENA 0.340 [0.290,0.391]	EBC + SA 0.029 [0.020,0.038]	ENat + EBC 0.004 [0.003,0.005]	EBC + WNA 0.003 [0.002,0.004]
				"Low-F_{ST}"	EBC + ENA 0.761 [0.727,0.795]	EBC + AF 0.173 [0.143,0.203]	EBC + SA 0.063 [0.049,0.076]	ENat + EBC 0.002 [0.002,0.003]	WNat + EBC 0.000 [0.000,0.000]
				"Pool-high-F_{ST}"	EBC + ENA 0.410 [0.356,0.464]	EBC + AF 0.388 [0.330,0.445]	EBC + SA 0.200 [0.155,0.244]	ENat + EBC 0.001 [0.001,0.002]	EBC + WNA 0.000 [0.000,0.001]
				"Pool-low-F_{ST}"	EBC + ENA 0.792 [0.759,0.826]	EBC + AF 0.122 [0.097,0.147]	EBC + SA 0.084 [0.064,0.104]	ENat + EBC 0.001 [0.001,0.001]	EBC + WNA 0.000 [0.000,0.000]
	South France	I-EU-Opi	28	"Reference"	EBC 0.880 [0.865,0.896]	ENat + EBC 0.049 [0.042,0.057]	EBC + WNA 0.022 [0.018,0.026]	WNat + EBC 0.019 [0.015,0.022]	EBC + SA 0.016 [0.012,0.019]
				"High-F_{ST}"	EBC + WNA 0.456 [0.413,0.499]	EBC 0.266 [0.224,0.308]	WNat + EBC 0.115 [0.097,0.132]	ENat + EBC 0.105 [0.092,0.119]	EBC + ENA 0.028 [0.023,0.034]
				"Low-F_{ST}"	EBC 0.722 [0.687,0.758]	WNat + EBC 0.104 [0.086,0.121]	ENat + EBC 0.069 [0.059,0.080]	EBC + ENA 0.048 [0.040,0.057]	EBC + WNA 0.032 [0.027,0.038]
				"Pool-high-F_{ST}"	EBC 0.981 [0.978,0.984]	EBC + WNA 0.008 [0.006,0.009]	ENat + EBC 0.004 [0.004,0.005]	WNat + EBC 0.003 [0.003,0.004]	EBC + SA 0.002 [0.002,0.003]
				"Pool-low-F_{ST}"	EBC 0.965 [0.960,0.971]	EBC + WNA 0.015 [0.012,0.018]	ENat + EBC 0.011 [0.009,0.013]	EBC + SA 0.003 [0.002,0.004]	WNat + EBC 0.003 [0.002,0.004]

Continent	Target genetic unit	Target site sample	Number of competing scenarios	Source sample set	Selected scenario (cf. highest probability)	2nd scenario	3rd scenario	4th scenario	5th scenario
	Italy	I-EU-Cun	28	"Reference"	WNA 0.692 [0.654,0.730]	WNat + WNA 0.161 [0.135,0.188]	ENat + WNA 0.117 [0.100,0.134]	EBC + WNA 0.010 [0.007,0.012]	ENA + WNA 0.008 [0.007,0.010]
				"High-F_{ST}"	WNA 0.779 [0.747,0.811]	ENat + WNA 0.092 [0.076,0.108]	ENA + WNA 0.083 [0.067,0.100]	WNA + AF 0.015 [0.012,0.019]	WNA + SA 0.012 [0.009,0.015]
				"Low-F_{ST}"	WNA 0.725 [0.696,0.753]	WNat + WNA 0.123 [0.106,0.141]	ENat + WNA 0.084 [0.073,0.094]	ENA + WNA 0.040 [0.034,0.046]	WNA + AF 0.013 [0.011,0.016]
				"Pool-high-F_{ST}"	WNA 0.973 [0.968,0.977]	ENat + WNA 0.015 [0.012,0.018]	WNat + WNA 0.007 [0.005,0.008]	ENA + WNA 0.003 [0.002,0.003]	WNA + SA 0.001 [0.001,0.002]
				"Pool-low-F_{ST}"	WNA 0.897 [0.882,0.912]	WNat + WNA 0.050 [0.041,0.059]	ENat + WNA 0.034 [0.029,0.040]	ENA + WNA 0.013 [0.011,0.016]	WNA + AF 0.002 [0.002,0.003]
		I-EU-Ale	28	"Reference"	WNA 0.659 [0.629,0.690]	ENat + WNA 0.154 [0.136,0.171]	WNat + WNA 0.093 [0.081,0.106]	EBC + WNA 0.062 [0.051,0.073]	ENat 0.014 [0.012,0.016]
				"High-F_{ST}"	WNA 0.584 [0.549,0.619]	EBC + WNA 0.176 [0.145,0.208]	ENat + WNA 0.124 [0.110,0.138]	ENA + WNA 0.048 [0.041,0.055]	WNA + AF 0.027 [0.023,0.031]
				"Low-F_{ST}"	WNA 0.527 [0.485,0.569]	ENat + WNA 0.341 [0.305,0.376]	ENat 0.050 [0.042,0.058]	WNat + WNA 0.022 [0.018,0.026]	EBC + WNA 0.022 [0.016,0.027]
				"Pool-high-F_{ST}"	WNA 0.753 [0.723,0.783]	ENat + WNA 0.115 [0.099,0.131]	EBC + WNA 0.093 [0.073,0.113]	WNat + WNA 0.017 [0.014,0.020]	ENA + WNA 0.007 [0.006,0.008]
				"Pool-low-F_{ST}"	WNA 0.889 [0.875,0.903]	ENat + WNA 0.061 [0.053,0.070]	WNat + WNA 0.016 [0.013,0.019]	EBC + WNA 0.015 [0.011,0.018]	ENA + WNA 0.010 [0.009,0.012]
	East Europe	I-EU-Hun	28	"Reference"	EBC + ENA 0.423 [0.369,0.478]	EBC + SA 0.284 [0.229,0.339]	ENat + ENA 0.054 [0.044,0.065]	ENat + SA 0.050 [0.040,0.061]	EBC + AF 0.028 [0.021,0.035]
				"High-F_{ST}"	EBC + ENA 0.440 [0.360,0.521]	EBC + WNA 0.334 [0.244,0.423]	EBC + AF 0.115 [0.083,0.146]	ENat + EBC 0.035 [0.024,0.047]	EBC + SA 0.022 [0.012,0.031]
				"Low-F_{ST}"	EBC + ENA 0.374 [0.329,0.420]	EBC + AF 0.327 [0.280,0.375]	ENat + EBC 0.172 [0.133,0.212]	EBC + SA 0.042 [0.031,0.052]	ENat + AF 0.018 [0.014,0.022]
				"Pool-high-F_{ST}"	EBC + SA 0.414 [0.335,0.493]	EBC + ENA 0.382 [0.314,0.450]	EBC + AF 0.074 [0.053,0.096]	ENat + EBC 0.056 [0.037,0.075]	EBC + WNA 0.050 [0.030,0.070]
				"Pool-low-F_{ST}"	EBC + ENA 0.541 [0.481,0.601]	EBC + SA 0.285 [0.226,0.343]	EBC + AF 0.081 [0.061,0.101]	EBC + WNA 0.031 [0.019,0.043]	ENat + EBC 0.024 [0.016,0.032]

Continent	Target genetic unit	Target site sample	Number of competing scenarios	Source sample set	Selected scenario (cf. highest probability)	2nd scenario	3rd scenario	4th scenario	5th scenario
		I-EU-Nor	28	"Reference"	ENat + EBC 0.240 [0.208,0.271]	ENat + WNA 0.238 [0.213,0.262]	ENat 0.179 [0.159,0.199]	EBC + WNA 0.153 [0.129,0.176]	ENat + ENA 0.027 [0.023,0.031]
				"High-F_{ST}"	EBC + WNA 0.668 [0.600,0.736]	WNA 0.121 [0.090,0.152]	ENA + WNA 0.077 [0.057,0.097]	WNat + WNA 0.046 [0.034,0.058]	WNA + AF 0.038 [0.028,0.049]
				"Low-F_{ST}"	EBC + WNA 0.498 [0.430,0.566]	ENat + EBC 0.382 [0.320,0.445]	EBC + ENA 0.037 [0.026,0.048]	ENat + WNA 0.019 [0.014,0.024]	ENat 0.013 [0.009,0.016]
				"Pool-high-F_{ST}"	EBC + WNA 0.664 [0.613,0.714]	ENat + WNA 0.088 [0.071,0.105]	ENat + EBC 0.063 [0.048,0.079]	ENat 0.033 [0.026,0.041]	ENA + WNA 0.031 [0.025,0.038]
				"Pool-low-F_{ST}"	EBC + WNA 0.548 [0.493,0.602]	ENat + EBC 0.258 [0.214,0.303]	ENat + WNA 0.102 [0.083,0.121]	ENat 0.043 [0.035,0.052]	WNA 0.013 [0.010,0.016]
Africa	South Africa	I-EU-Bet	28	"Reference"	ENA 0.554 [0.500,0.608]	ENA + SA 0.333 [0.279,0.387]	SA 0.062 [0.046,0.079]	ENA + WNA 0.022 [0.017,0.027]	WNat + ENA 0.010 [0.008,0.012]
				"High-F_{ST}"	ENA 0.755 [0.717,0.794]	ENA + SA 0.122 [0.091,0.152]	ENA + WNA 0.052 [0.041,0.063]	ENat + ENA 0.026 [0.021,0.031]	ENA + Europe 0.017 [0.012,0.022]
				"Low-F_{ST}"	ENA 0.700 [0.661,0.739]	ENA + SA 0.189 [0.154,0.224]	ENA + WNA 0.044 [0.034,0.053]	SA 0.037 [0.028,0.047]	ENat + ENA 0.010 [0.008,0.012]
				"Pool-high-F_{ST}"	ENA 0.684 [0.629,0.739]	ENA + SA 0.298 [0.244,0.352]	SA 0.012 [0.008,0.016]	ENA + Europe 0.003 [0.002,0.003]	ENA + WNA 0.001 [0.001,0.002]
				"Pool-low-F_{ST}"	ENA 0.769 [0.736,0.803]	ENA + SA 0.171 [0.141,0.201]	SA 0.049 [0.037,0.060]	ENA + WNA 0.005 [0.004,0.006]	ENA + Europe 0.002 [0.002,0.003]
		I-EU-Som	28	"Reference"	ENA 0.709 [0.664,0.754]	ENA + SA 0.241 [0.197,0.286]	SA 0.017 [0.012,0.022]	WNat + ENA 0.014 [0.011,0.017]	ENA + WNA 0.009 [0.007,0.011]
				"High-F_{ST}"	ENA 0.637 [0.599,0.675]	WNat + ENA 0.108 [0.091,0.125]	ENat + ENA 0.070 [0.059,0.081]	ENA + WNA 0.066 [0.054,0.077]	ENA + SA 0.046 [0.033,0.058]
				"Low-F_{ST}"	ENA 0.890 [0.869,0.911]	ENA + SA 0.084 [0.065,0.103]	SA 0.013 [0.009,0.016]	ENA + WNA 0.005 [0.004,0.006]	WNat + ENA 0.004 [0.003,0.004]
				"Pool-high-F_{ST}"	ENA 0.736 [0.688,0.784]	ENA + SA 0.221 [0.175,0.267]	SA 0.028 [0.019,0.037]	WNat + ENA 0.006 [0.004,0.007]	ENA + WNA 0.004 [0.003,0.005]
				"Pool-low-F_{ST}"	ENA 0.881 [0.861,0.901]	ENA + SA 0.089 [0.072,0.107]	SA 0.017 [0.013,0.022]	ENA + WNA 0.004 [0.003,0.005]	WNat + ENA 0.003 [0.003,0.004]

Table S4: Detailed results of the first set of ABC analyses processed to make inferences about the origin of invasive genetic units.
Note: We present in the table the most likely scenario (i.e. the one with the highest probability) and the four next scenarios according to their probabilities for each target genetic unit (names are as in Fig. 1 and Table S1) and for each putative source sample set (see Table 1). Posterior probabilities are given with 95% credibility intervals between brackets. ENat = Eastern native genetic unit; ENA = Eastern North America; WNA = Western North America; SA = South America; AF = Africa; EBC = European Biocontrol.

ABC analyses	Target genetic unit(s)	Number of competing scenarios	Source sample set	Selected scenario	Type I error	Type II error mean (min - max)
Set 1	Brazil (I-SA-Cur sample)	28	"Reference"	ENA	0.39	0.035 (0.000 - 0.410)
	West Europe (I-EU-Bel sample)	28	"Reference"	EBC + ENA	0.20	0.051 (0.000 - 0.610)
	Italy (I-EU-Cun sample)	28	"Reference"	WNA	0.13	0.014 (0.000 - 0.110)
Set 2	Brazil / Argentina / Chile	5	"Reference"	Two ENA introductions: [Chile] + [Brazil, Argentina]	0.26	0.070 (0.032 - 0.104)
	West Europe / East Europe / Italy	4	"Reference"	- Two WNA introductions: [East Europe] + [Italy] - One ENA introduction: [West Europe]	0.40	0.136 (0.046 - 0.256)

Table S5: Evaluation of confidence in scenario selection using ABC.
Note: Three and two analyses (defined by the target genetic unit) are presented for the first and second set of ABC analyses respectively. Set 1 corresponds to the analyses of the origin of the target invasive genetic units. Set 2 corresponds to the analyses of the intracontinental relationships between target invasive genetic units sharing similar extracontinental origin. See Appendix S2 for details.

						Model checking on selected scenario		
Continent	Target genetic unit	Target site sample	Number of competing scenarios	Source sample set	Selected scenario	Total number of test statistics (*ts*)	Number (and proportion) of *ts* with : $P < 0.05$ or $P > 0.95$	Number (and proportion) of *ts* with : $P < 0.01$ or $P > 0.99$
North America	Colorado	I-NA-Col	10	"Reference"	ENA + WNA	107	0 (0.000)	0 (0.000)
				"High-F_{ST}"	ENA + WNA	107	0 (0.000)	0 (0.000)
				"Low-F_{ST}"	ENA + WNA	107	0 (0.000)	0 (0.000)
				"Pool-high-F_{ST}"	ENA + WNA	107	0 (0.000)	0 (0.000)
				"Pool-low-F_{ST}"	ENA + WNA	107	0 (0.000)	0 (0.000)
	Utah	I-NA-Uta	10	"Reference"	ENA + WNA	107	0 (0.000)	0 (0.000)
				"High-F_{ST}"	ENA + WNA	107	0 (0.000)	0 (0.000)
				"Low-F_{ST}"	ENA + WNA	107	0 (0.000)	0 (0.000)
				"Pool-high-F_{ST}"	ENA + WNA	107	0 (0.000)	0 (0.000)
				"Pool-low-F_{ST}"	ENA + WNA	107	0 (0.000)	0 (0.000)
South America	Brazil	I-SA-Cur	28	"Reference"	ENA	279	11 (0.039)	0 (0.000)
				"High-F_{ST}"	ENA + AF	279	21 (0.075)	3 (0.011)
				"Low-F_{ST}"	ENA + AF	279	5 (0.018)	0 (0.000)
				"Pool-high-F_{ST}"	ENA	279	12 (0.043)	0 (0.000)
				"Pool-low-F_{ST}"	ENA	279	8 (0.029)	0 (0.000)
		I-SA-Gon	28	"Reference"	ENA	279	7 (0.025)	0 (0.000)
				"High-F_{ST}"	ENA	279	21 (0.075)	3 (0.011)
				"Low-F_{ST}"	ENA	279	8 (0.029)	1 (0.004)
				"Pool-high-F_{ST}"	ENA	279	22 (0.079)	1 (0.004)
				"Pool-low-F_{ST}"	ENA	279	8 (0.029)	1 (0.004)
	Argentina	I-SA-Arg	28	"Reference"	ENA	279	11 (0.039)	0 (0.000)
				"High-F_{ST}"	WNA + AF	279	15 (0.054)	1 (0.004)
				"Low-F_{ST}"	ENA	279	12 (0.043)	1 (0.004)
				"Pool-high-F_{ST}"	ENA + WNA	279	21 (0.075)	2 (0.007)
				"Pool-low-F_{ST}"	ENA	279	22 (0.079)	1 (0.004)
	Chile	I-SA-Chi	28	"Reference"	ENA	279	9 (0.032)	0 (0.000)
				"High-F_{ST}"	ENA	279	19 (0.068)	3 (0.011)
				"Low-F_{ST}"	ENA	279	12 (0.043)	1 (0.004)
				"Pool-high-F_{ST}"	ENA	279	7 (0.025)	1 (0.004)
				"Pool-low-F_{ST}"	ENA	279	13 (0.047)	2 (0.007)
Europe	West Europe	I-EU-Bel	28	"Reference"	EBC + ENA	279	15 (0.054)	0 (0.000)
				"High-F_{ST}"	EBC + ENA]	279	26 (0.093)	2 (0.007)
				"Low-F_{ST}"	EBC + ENA	279	7 (0.025)	0 (0.000)
				"Pool-high-F_{ST}"	EBC + ENA	279	15 (0.054)	2 (0.007)
				"Pool-low-F_{ST}"	EBC + ENA	279	15 (0.054)	0 (0.000)

Continent	Target genetic unit	Target site sample	Number of competing scenarios	Source sample set	Selected scenario	Total number of test statistics (*ts*)	Number (and proportion) of *ts* with : $P < 0.05$ or $P > 0.95$	Number (and proportion) of *ts* with : $P < 0.01$ or $P > 0.99$
		I-EU-Cas	28	"Reference"	EBC + ENA	279	6 (0.022)	0 (0.000)
				"High-F_{ST}"	EBC + AF	279	9 (0.032)	3 (0.011)
				"Low-F_{ST}"	EBC + ENA	279	3 (0.011)	2 (0.007)
				"Pool-high-F_{ST}"	EBC + ENA	279	8 (0.029)	1 (0.004)
				"Pool-low-F_{ST}"	EBC + ENA	279	6 (0.021)	0 (0.000)
	South France	I-EU-Opi	28	"Reference"	EBC	279	23 (0.082)	2 (0.007)
				"High-F_{ST}"	EBC + WNA	279	7 (0.025)	1 (0.004)
				"Low-F_{ST}"	EBC	279	10 (0.036)	6 (0.022)
				"Pool-high-F_{ST}"	EBC	279	19 (0.068)	4 (0.014)
				"Pool-low-F_{ST}"	EBC	279	13 (0.047)	7 (0.025)
	Italy	I-EU-Cun	28	"Reference"	WNA	279	4 (0.014)	0 (0.000)
				"High-F_{ST}"	WNA	279	15 (0.054)	1 (0.004)
				"Low-F_{ST}"	WNA	279	9 (0.032)	1 (0.004)
				"Pool-high-F_{ST}"	WNA	279	20 (0.072)	1 (0.004)
				"Pool-low-F_{ST}"	WNA	279	2 (0.007)	0 (0.000)
		I-EU-Ale	28	"Reference"	WNA	279	4 (0.014)	0 (0.000)
				"High-F_{ST}"	WNA	279	7 (0.025)	1 (0.004)
				"Low-F_{ST}"	WNA	279	1 (0.004)	0 (0.000)
				"Pool-high-F_{ST}"	WNA	279	9 (0.032)	1 (0.004)
				"Pool-low-F_{ST}"	WNA	279	4 (0.014)	1 (0.004)
	East Europe	I-EU-Hun	28	"Reference"	EBC + ENA	279	6 (0.022)	0 (0.000)
				"High-F_{ST}"	EBC + ENA	279	13 (0.047)	2 (0.007)
				"Low-F_{ST}"	EBC + ENA	279	3 (0.011)	0 (0.000)
				"Pool-high-F_{ST}"	EBC + SA	279	19 (0.068)	3 (0.011)
				"Pool-low-F_{ST}"	EBC + ENA	279	4 (0.014)	1 (0.004)
		I-EU-Nor	28	"Reference"	ENat + EBC	279	1 (0.004)	0 (0.000)
				"High-F_{ST}"	EBC + WNA	279	18 (0.065)	3 (0.011)
				"Low-F_{ST}"	EBC + WNA	279	6 (0.022)	0 (0.000)
				"Pool-high-F_{ST}"	EBC + WNA	279	12 (0.043)	1 (0.004)
				"Pool-low-F_{ST}"	EBC + WNA	279	12 (0.043)	0 (0.000)
Africa	South Africa	I-EU-Bet	28	"Reference"	ENA	279	4 (0.014)	0 (0.000)
				"High-F_{ST}"	ENA	279	20 (0.072)	3 (0.011)
				"Low-F_{ST}"	ENA	279	6 (0.022)	0 (0.000)
				"Pool-high-F_{ST}"	ENA	279	4 (0.014)	1 (0.004)
				"Pool-low-F_{ST}"	ENA	279	13 (0.047)	1 (0.004)
		I-EU-Som	28	"Reference"	ENA	279	11 (0.039)	0 (0.000)
				"High-F_{ST}"	ENA	279	22 (0.079)	4 (0.014)
				"Low-F_{ST}"	ENA	279	8 (0.029)	0 (0.000)
				"Pool-high-F_{ST}"	ENA	279	18 (0.065)	1 (0.004)
				"Pool-low-F_{ST}"	ENA	279	6 (0.022)	0 (0.000)

Table S6: Model checking results for the first set of ABC analyses (i.e. origin of invasive genetic units).
Note: the number and proportion of observed test statistics which significantly (at 5% and at 1%) lay in the tails of the probability distribution of statistics calculated from posterior simulations are given in the last two columns. The false discovery rate associated with multiple comparisons has not been corrected here. If we do so, as advised by Cornuet *et al.* (2010, *BMC Bioinformatics*, 11:401), not any observed statistics remains out of the posterior ranges. Site sample names are as in Fig. 1 and Table S1. See Appendix S2 for details regarding methods.

				Model checking on selected scenario		
Continent (target genetic units)	Number of competing scenarios	Source sample set	Selected scenario	Total number of test statistics (*ts*)	Number (and proportion) of *ts* with : $P < 0.05$ or $P > 0.95$	Number (and proportion) of *ts* with : $P < 0.01$ or $P > 0.99$
South America (Brazil / Argentina / Chile)	5	"Reference"	Two independent ENA introductions: [Chile] + [Brazil, Argentina]	101	1 (0.010)	0 (0.000)
		"High-F_{ST}"	Two independent ENA introductions: [Chile] + [Brazil, Argentina]	101	1 (0.010)	0 (0.000)
		"Low-F_{ST}"	Two independent ENA introductions: [Chile] + [Brazil, Argentina]	101	1 (0.010)	0 (0.000)
		"Pool-high-F_{ST}"	Two independent ENA introductions: [Chile] + [Brazil, Argentina]	101	2 (0.020)	0 (0.000)
		"Pool-low-F_{ST}"	Two independent ENA introductions: [Chile] + [Brazil, Argentina]	101	1 (0.010)	0 (0.000)
Europe (East EU / West EU / Italy)	4	"Reference"	- Two independent WNA introductions: [East Europe] + [Italy] - One ENA introduction: [West Europe]	333	10 (0.030)	0 (0.000)
		"High-F_{ST}"	- Two independent WNA introductions: [East Europe] + [Italy] - One ENA introduction: [West Europe]	333	2 (0.006)	0 (0.000)
		"Low-F_{ST}"	- Two independent WNA introductions: [East Europe] + [Italy] - One ENA introduction: [West Europe]	333	8 (0.024)	0 (0.000)
		"Pool-high-F_{ST}"	- Two independent WNA introductions: [East Europe] + [Italy] - One ENA introduction: [West Europe]	333	3 (0.009)	0 (0.000)
		"Pool-low-F_{ST}"	- Two independent WNA introductions: [East Europe] + [Italy] - One ENA introduction: [West Europe]	333	4 (0.012)	0 (0.000)

Table S7: Model checking results for the second set of ABC analyses (i.e. intracontinental relationship between target invasive genetic units).
Note: the number and proportion of observed test statistics which significantly (at 5% and at 1%) lay in the tails of the probability distribution of statistics calculated from posterior simulations are given in the last two columns. The false discovery rate associated with multiple comparisons has not been corrected here. If we do so, as advised by Cornuet *et al.* (2010, *BMC Bioinformatics*, 11:401), not any observed statistics remains out of the posterior ranges. See Appendix S2 for details regarding methods.