Molecular Ecology (2007) **16**, 3069–3083 doi: 10.1111/j.1365-294X.2007.03336.x

Introduction history of *Drosophila subobscura* in the New World: a microsatellite-based survey using ABC methods

M. PASCUAL,* M. P. CHAPUIS,† F. MESTRES,* J. BALANYÀ,* R. B. HUEY,‡ G. W. GILCHRIST,§ L. SERRA* and A. ESTOUP†

**Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain, †INRA, UMR CBGP (INRA/IRD/Cirad/Montpellier SupAgro), Campus international de Baillarguet, CS 30016, F-34988 Montferrier-sur-Lez Cedex, France, ‡Department of Biology, Box 351800, University of Washington, Seattle, WA 98195-1800, USA, §Department of Biology, Box 8795, College of William and May, Williamsburg, VA 23187, USA*

Abstract

***Drosophila subobscura* is a Palearctic species that was first observed in South and North America in the early 1980s, and that rapidly invaded broad latitudinal ranges on both continents. To trace the source and history of this invasion, we obtained genotypic data on nine microsatellite loci from two South American, two North American and five European populations of *D. subobscura*. We analysed these data with traditional statistics as well as with an approximate Bayesian computation (ABC) framework. ABC methods yielded the strongest support for the scenario involving a serial introduction with founder events from Europe into South America, and then from South America into North America. Stable effective population size of the source population was very large (around one million individuals), and the propagule size was notably smaller for the introduction into South America (i.e. high bottleneck severity index with only a few effective founders) but considerably larger for the subsequent introduction into North America (i.e. low bottleneck severity index with around 100–150 effective founders). Finally, the Mediterranean region of Europe (and most likely Barcelona from the localities so far analysed) is proposed as the source of the New World flies, based on mean individual assignment statistics.**

Keywords: approximate Bayesian computation, colonization, *Drosophila subobscura*, microsatellites, number of founders, sequential invasions

Received 7 October 2007; revision received 14 February 2007; accepted 5 March 2007

Introduction

Recent colonization histories provide the opportunity to investigate the genetic consequences of founder events as well as the evolutionary trajectories of newly established populations in a new environment (Mayr 1954). Replicate colonizations are 'grand experiments in evolution' because they provide remarkable opportunities for evaluating the repeatability of microevolutionary trajectories (Ayala *et al*. 1989). The success of colonizing species may depend on their ability to evolve in response to their new environment. Colonizing species may evolve, both during their initial establishment and during subsequent range expansion, in response to selection pressures (e.g. Lee 2002; Balanyà *et al*. 2006). Genetic drift may also play an important role and even determine colonization success (e.g. Tsutsui *et al*. 2000). Consequently, when a species independently invades several areas, the outcome of the colonization and the convergence of results in the different areas may depend upon the specific genetic pool of the colonizers as well as the population dynamics and selective processes associated with the colonization.

To understand the evolutionary genetics of colonizations, one must identify the most likely source of colonizers, the levels of genetic diversity of both introduced and native populations, the geographical pathways of spread, and the ability of the populations to evolve in novel environments. Highly variable genetic markers such as microsatellites have proven useful in answering these questions (Estoup *et al*. 2004; Miller *et al*. 2005). Standard analytical methodologies, however, often fail to resolve some questions. Approximate Bayesian computations (ABC) are model-based methods that have been used successfully to

Correspondence: Marta Pascual, Fax: +34 93 403 4420;
E-mail: martapascual@ub.edu

describe the recent colonization history of the toad *Bufo marinus* (Estoup *et al*. 2001, 2004), the bird *Zosterops lateralis* (Estoup & Clegg 2003), the corn rootworn *Diabrotica virgifera virgifera* (Miller *et al*. 2005) and *Drosophila melanogaster* (Thornton & Andolfatto 2006). Fully likelihood-inferential methods (e.g. Beaumont 1999) are more powerful but (for numerical reasons) can not yet treat complex evolutionary scenarios. ABC methods do not present such numerical difficulties because they do not require the probability of obtaining a given gene sample configuration to be estimated. Therefore, they have the potential to handle virtually any complex models provided that simulation of data under the model is feasible (Beaumont *et al*. 2002).

Drosophila subobscura colonized the Americas a quarter of a century ago, and this invasion has been described as a large-scale natural experiment with two replicates (Prevosti *et al*. 1988). This Palearctic species is native to a wide range from North Africa to Scandinavia (Krimbas 1993). It was observed for the first time in 1978 in South America (Brncic & Budnik 1980) and in 1982 in North America (Beckenbach & Prevosti 1986). *D. subobscura* rapidly spread over large areas in both South and North America (Prevosti *et al*. 1989; Noor *et al*. 2000), despite facing competition in North America from native flies of the same *obscura* group (Pascual *et al*. 1998).

Drosophila subobscura has proven to be an excellent model to study evolution in nature (Balanyà *et al*. 2006). Within a few years after the introduction was detected, clines for some chromosomal arrangements had already developed and these mimicked those encountered in the Old World (Prevosti *et al*. 1988). Surveys made about two decades after the invasion showed conspicuous wing-size clines coincident with those for Old World flies (Huey *et al*. 2000; Gilchrist *et al*. 2004). The remarkable success and rapid evolution of these invading flies might suggest that the initial invaders contained a high level of genetic diversity, upon which selection could act (Lee 2002). However, both South and North American populations have substantially reduced genetic variation compared to European ones either in mitochondrial (Latorre *et al*. 1986) or nuclear markers (Prevosti *et al*. 1988; Balanyà *et al*. 1994), indicating that a strong bottleneck occurred during the colonization, most probably because the number of founders was small. However, the number of founders varies largely among studies (i.e. between 2 and 150) depending on the markers and methods used for its estimation (Mestres *et al*. 1990; Rozas & Aguadé 1991).

The source of the initial colonizers and the routes of introduction remain unclear. Chromosomal polymorphism data mostly suggests a western Mediterranean origin in Europe (Brncic *et al*. 1981); but the presence of the O_5 arrangement, which is rare in the Mediterranean but is found in New World populations, does not support this hypothesis (Ayala *et al*. 1989). The two colonization events in South and North America do not seem to be independent (i.e. separate invasions from Europe) since they share the same chromosomal arrangements (Balanyà *et al*. 2003), lethal genes (Mestres *et al*. 1990), allozyme alleles (Balanyà *et al*. 1994) and *rp49* haplotypes (Rozas & Aguadé 1991). However, all markers used up to now failed to indicate whether the flies first colonized North or South America, and the first observation dates (i.e. 1978 and 1982) are too close to safely infer which continent was invaded first. Thus, the introduction sequence needs to be thoroughly tested by using highly polymorphic markers and appropriate inferential methods.

The aim of the present study is to answer some of these unresolved questions concerning the colonization history of *D. subobscura* in the Americas. If both colonization events were not independent, which was the sequence of invasion? What was the propagule size in each founding event? Where did the founders come from? We have used traditional statistical treatments and ABC methods on microsatellite data obtained from two South American, two North American and five European populations. We used the ABC methods to infer the order of introductions (e.g. first to South America) and the magnitude of founder events (e.g. bottleneck severity index and effective number of founders). Finally, the question of the origin of American founders was addressed through the use of mean individual assignment statistics.

Materials and methods

Population samples and markers

Drosophila subobscura from South America were collected in November 1999 in La Serena (29.54°S, 71.18°W) and Puerto Montt (41.28°S, 73.00°W), Chile. Fifty individuals were analysed for each population (one first-generation female per isofemale line). The nine microsatellites surveyed in this study corresponded to a subset of those developed by Pascual *et al*. (2000) from the same species: dsub01, dsub02, dsub04, dsub05, dsub18, dsub19, dsub20, dsub21 and dsub27. Note that dsub05, dsub19 and dsub21 loci are X-linked; but the others are autosomal. DNA extraction, microsatellite PCR amplification and allele size determination were processed as described in (Pascual *et al*. 2001). The same loci were previously analysed in five European and two North American populations (Pascual *et al*. 2001).

Demographical model and introduction scenarios

Because *D. subobscura* is well known by naturalists, is highly prolific and has a short generation time (Avelar *et al*. 1987), the first sight dates in South & North America (1978 and 1982) are likely to be close to the actual introduction dates. Therefore, we used these first sight dates to fix the times (translated in number of generations) of the two population split-events characterizing our eight introduction models

(see details below). Preliminary analyses have shown that using different splitting dates (e.g. one to three years earlier than first sight dates) did not change our results due to the large stable effective population sizes of the species relative to the number of generations since introduction events (not shown). In all introduction scenarios, our demographical model was specified by six demographical variable parameters: the generation time (*G*), the stable effective population size (*N0*) which was assumed to be the same in all populations, the effective number of founders in the first and second introduced population (*N1* and *N2*, respectively), and the duration of the bottleneck that occurred during the first and second colonization (*D1* and *D2*, respectively). Because the source and introduced populations are separated by large geographical distances, all populations were assumed to evolve as isolated demes.

We considered eight introduction scenarios which differed in the order of introduction and on whether founder events occurred:

Scenario 1: serial introductions with founder events from Europe into South America in 1978, and then from South America into North America in 1982 (EU → SA → NA).
Scenario 2: same as scenario 1 without founder events.
Scenario 3: serial introductions with founder events from Europe into North America in 1978, and then from North America into South America in 1982 (EU → NA → SA).
Scenario 4: same as scenario 3 without founder events.
Scenario 5: independent introductions with founder events from Europe into South America in 1978, and then from Europe into North America in 1982 (EU → SA and then EU → NA).
Scenario 6: same as scenario 5 without founder events.
Scenario 7: independent introductions with founder events from Europe into North America in 1978, and then from Europe into South America in 1982 (EU → NA and then EU → SA).
Scenario 8: same as scenario 7 without founder events.

The exact geographical origin of European colonizers remains uncertain, although most chromosomal polymorphism data suggested a western Mediterranean origin (Brncic *et al*. 1981), an area represented by samples from Barcelona and Montpellier in our study. Moreover, assignment methods suggest Barcelona as the most likely origin among our available European samples (see Results). Collecting records in the New World suggest that Puerto Montt and Bellingham are the best available samples to represent the initial introduced populations in South and North America, respectively (Brncic *et al*. 1981; Beckenbach & Prevosti 1986). Hence, all scenarios were treated using three sets of samples: (i) a single population sample set that includes the most likely source and introduced populations (i.e. Barcelona for Europe, Puerto Montt for South America and Bellingham for North America); (ii) a first set of pooled populations which includes Montpellier + Barcelona as the European source, Puerto Montt + Serena as the South American sample, and Bellingham + Fort Bragg as the North American sample; and (iii) a second set of pooled samples that includes all European samples (i.e. Aarhus + Lille + Montpellier + Barcelona + Málaga) as the European source, Puerto Montt + Serena as the South American sample, and Bellingham + Fort Bragg as the North American sample.

Parameter inferences using approximate Bayesian computation

Calculation of the probability distribution of the demographical and mutational parameters for the full genetic data is numerically difficult for complex demographical histories (see Stephens 2003). To surmount these difficulties, we used an ABC method based on summary statistics to infer posterior distributions of variable parameters without explicit likelihood calculations (Beaumont *et al*. 2002).

We used a time-continuous approximation of the coalescence process (Hudson 1990) to simulate genetic datasets under a given introduction model. Within- and between-population genetic variation of the three populations was summarized with five different types of statistics: the mean number of alleles (*A*) per locus and population, the mean expected heterozygosity (*H*; Nei 1978), the mean ratio of the number of alleles over the range of allelic sizes per population expressed in base pairs (*M*; Garza & Williamson 2001), F_{ST} between pairs of sampled populations (Weir 1996), and the mean individual assignment likelihood of individuals collected in population *i* and assigned into population *j* ($L_{i\to j}$; cf. formula 9 in Rannala & Mountain 1997). Because all introduction models include three populations we had a total of 18 summary statistics (3 *A*, 3 *H*, 3 *M*, 3 F_{ST} and 6 $L_{i\to j}$).

Briefly, the ABC method involves two successive steps (see Beaumont *et al*. 2002). The first is a rejection step. It consists of accepting only sets of parameter values drawn in prior distributions that give values of summary statistics computed from simulated data sets close to those computed from the observed data set (i.e. our target summary statistics). A Euclidian distance is computed between the normalized summary statistics of the observed and simulated datasets. Iteration is accepted when the Euclidian distance is lower than a given threshold. The second step is a local linear regression adjustment, which attempts to model the relationship between the parameter values and the summary statistics in the vicinity of the target summary statistics and thereby to correct the accepted parameter values.

For the rejection step, we set a tolerance threshold δ to be the quantile $p_\delta = 10^{-3}$ of the empirical distribution function of the simulated Euclidian distance values. In agreement with Beaumont *et al*. (2002), preliminary analyses showed similar estimation results when using different tolerance threshold corresponding to $p_\delta = 5 \times 10^{-3}$ and $p_\delta = 2 \times 10^{-4}$ (results not shown). To avoid storing a large number of outputs, the normalizing factors and the critical quantile, δ, were first calculated from 10^6 iterations for a given introduction model. Simulations were then run, keeping only those outputs with summary statistics within the tolerance threshold (i.e. with a Euclidian distance < δ) until 10 000 sets of parameter values were accepted. The regression step was then processed on the 10 000 accepted values with all parameters values transformed on a log scale to reduce inequality of variances among parameters in the regression (Estoup *et al*. 2004). Adjusted values were then back-transformed taking the exponential for all parameters to express posterior densities on a normal scale.

Similar estimation results were obtained when using a weighted Euclidian distance as proposed by Hamilton *et al*. (2005a) and/or a log-tangent transformation of parameters as proposed by Hamilton *et al*. (2005b; results not shown).

We used personal programs for all above computation and the locfit function (Loader 1996) implemented in version 2.2.1 of the R package (Ihaka & Gentleman 1996) for representation of posterior distributions of parameters.

Comparison of introduction scenarios using approximate Bayesian computation

A similar ABC framework was used to discriminate among our eight introduction scenarios. This step was processed before estimating posterior densities of variable parameters, the latter inferences being made only under the most likely of our eight scenarios. Prior probability for each introduction scenario was set to be equal (i.e. 1/8). The empirical distribution function of multiscenario Euclidian distances (δ) was computed as described in the previous section from a total of 8×10^7 δ-values corresponding to 10^7 values for each scenario. The lowest 100 or 500 of those 8×10^7 δ-values were used to estimate the posterior probability of each introduction scenario, as the proportion of time each scenario was represented within this subset of n_δ 'best' Euclidian distances. The precision of the posterior probability estimation of scenarios is expected to decrease when δ-values and hence n_δ increase. However, a large variance of those estimations is also expected for a too small number of retained δ-values. Because, we found that such a variance became small for $n_\delta > 40$ (results not shown), we will only present posterior probability estimations for $n_\delta = 100$ and $n_\delta = 500$.

Prior distributions of parameters

The prior distributions we used for inferences are given in Table 1. Because *D. subobscura* has four to six generations per year in nature (Begon 1976), we chose a uniform prior distribution bounded between these two values (*G*). *G* was considered here as a nuisance parameter (i.e. a variable parameter that is included in the simulation process because a certain amount of uncertainty — expressed through a prior distribution — exits on it, but we do not intend to make specific inferences on it).The prior distribution for the long-term stable effective population size (*N0*) was set as a loguniform, bounded between 200 000 and 2×10^6 diploid individuals. This range includes estimation values obtained for European localities with different methodologies (Mestres & Serra 1991; Pascual *et al*. 2001). The priors

Table 1 Prior distributions for the variable demographical and marker parameters describing the introduction models of *D. subobscura* in the New World

	Parameter	Distribution	Range of supported value (2.5% and 97.5% quantiles)
Demographical parameters	*N0*	logUniform[2×10^5, 2×10^6]	21 1700–1 883 880
	N1	logUniform[2, 500]	2.29–435
	N2	logUniform[2, 500]	2.29–435
	G	Uniform[4, 6]	4.05–5.95
	D1	Uniform[2, 5]	2.07–4.93
	D2	Uniform[2, 5]	2.07–4.93
Marker parameters	μ_{rn}	Gamma(3, 321 930)	1.9×10^{-6}–2.3×10^{-5}
	σ^2	Exp(0.36)	0.009–1.334
	μ_{id}	Exp(2.5×10^{-8})	6.3×10^{-10}–9.3×10^{-8}

Note: *N0* = stable effective population size; *N1* and *N2* = effective number of founders during the first and second introduction step, respectively; *G* = number of generations per year. *D1* and *D2* = duration of the bottleneck during the first and second introduction step, respectively; μ_{rn} = mutation rate for the repeat motif core sequence; μ_{id} = mutation rate for insertion-deletions in the flanking regions. Exp: exponential distribution; σ^2 = variance of the geometric distribution of repeat number change.

for the effective numbers of founding individuals (*N1* and *N2*) were set as loguniform bounded between 2 and 500 diploid individuals, a range suggested when comparing European and American populations (Mestres *et al*. 1990; Rozas & Aguade 1991). Balanced sex ratios could be assumed (Pascual *et al*. 2004) so that the drawn effective population size values were multiplied by 3/4 for the three X-linked microsatellites. It is worth noting that the durations of the bottleneck (*D1* and *D2*) and the effective numbers of individuals during this period (*N1* and *N2*) cannot be individually identified in the likelihood, so that the marginal distribution of *N1* and *N2* strongly depends on the prior on *D1* and *D2*, respectively. This has two main consequences. First, the prior distributions on *D1* and *D2* have to be carefully chosen to get sensible estimation of *N1* and *N2*. Because *D. subobscura* is a prolific species (Avelar *et al*. 1987), the duration of the bottleneck in a newly established population is likely to be short in a favourable environment. In fact the species spread and became abundant across a large latitudinal gradient of *c*. 1500 km in only 1.5 years in South America (Ayala *et al*. 1989). Hence we assumed that population bottlenecks (*D1* and *D2*) lasted only a few generations, and so we used a uniform prior distribution bounded between two and five generations. Second, we considered a combined parameter called the bottleneck severity index computed as $K1 = D1/N1$ and $K2 = D2/N2$ for the first and subsequent introduced population, respectively (Wright *et al*. 2005).

The presence at several *D. subobscura* microsatellite loci differing by both even and uneven allele sizes indicates the occurrence of two types of mutational events: some change in number of repeat units in the microsatellite core sequence and uneven insertion/deletion (most likely of one nucleotide length) within the amplified fragment. Both mutational processes have been modelled. Prior information regarding the mutation rate for dinucleotide repeats was formalized using data observed in *D. melanogaster* (9.3×10^{-6}, Schug *et al*. 1998). Such a low mean mutation rate seems to also hold for *D. subobscura* (Pascual *et al*. 2000). To allow for variation of mutation rate across loci, we used a gamma ($\alpha = 3$, $\lambda = 321\,930$) distribution for drawing single locus mutation rates μ_{rn}. A nonbounded GSM mutation model was assumed; for each locus, the variance of the geometric distribution was randomly drawn from an exponential distribution with a mean of 0.36 (Estoup *et al*. 2001). Uneven insertion/deletion events within the amplified fragment were modelled by changing the allele size by +/− one nucleotide at a rate μ_{id}. In *Drosophila* sequences the ratio of insertion or deletion to substitution is 0.145 (Petrov & Hartl 1998). Because point mutation rates (i.e. insertion + deletion + substitution) are estimated as 10^{-9} per nucleotide (e.g. Lewin 1994), we estimated the rate of insertion-deletion as 2.5×10^{-8} in a fragment of 200 base pairs (which approximately correspond to the average length of amplified fragments). Hence we used an exponential distribution of mean 2.5×10^{-8} for drawing single locus insertion-deletion mutation rates μ_{id}.

Sensitivity to priors

We assayed the sensitivity of our inferences to demographical priors by assuming different priors on effective population sizes and bottleneck severity. Prior set 1 corresponds to the standard priors described in Table 1. Prior set 2 assays a uniform distribution bounded between 200 000 and 2×10^6 diploid individuals for *N0*, and between 2 and 500 individuals for *N1* and *N2*. We tested sensitivity of our inferences to mutation model by assuming a strict stepwise mutation model (SMM) for all markers (prior set 3) or a GSM with constraints on allele size by imposing reflecting boundaries to an allele size range of $k = 30$ possible continuous allelic states (e.g. Feldman *et al*. 1997; prior set 4). To test for sensitivity of our inferences to mutation rates, we assumed a more diffuse prior for μ_{rn} by using an exponential ($\alpha = 1$, $\lambda = 107\,310$) distribution for drawing single locus mutation rates of repeat numbers (prior set 5), or the absence of uneven insertion/deletion within the amplified fragments (i.e. $\mu_{id} = 0$; prior set 6).

We assessed effects of these different priors both for our inferences on the posterior probability of each introduction scenario (5×10^6 instead of 10^7 iterations per scenario, hence taking the lowest 50 or 250 δ-values) and on posterior distributions of demographical parameters under the most likely scenario (from 5000 instead of 10 000 accepted values, with $p_\delta = 10^{-3}$). These simulations were processed only using the single population sample set that included the most likely source and introduced populations.

Traditional approaches for retracing introduction history

We used a neighbour joining (NJ) tree representation of relationships between populations as a traditional way to make inferences on population introduction histories (e.g. Caracristi & Schlötterer 2003; Colautti *et al*. 2005). To construct this tree, we used the chord distance of Cavalli-Sforza & Edwards (1967). The robustness of the tree topology was evaluated by carrying out 1000 bootstrap replicates over loci. All estimations were done using the software package POPULATIONS (Langella 2002).

Identification of European source populations

Individual assignment statistics (e.g. Rannala & Mountain 1997) are traditionally used to identify the source of introduced populations (Davies *et al*. 1999). However, the behaviour of such statistics remains to be assessed when the introduced population has endured a strong bottleneck during colonization. This question was tackled by running

Table 2 F_{ST} values and genetic differentiation significance between populations

	South America		North America		Europe			
	PM	LS	BE	FB	AA	LI	MO	BA
LS	0.007							
BE	0.016*	0.018*						
FB	0.000	0.017*	0.003					
AA	0.087*	0.098*	0.118*	0.104*				
LI	0.084*	0.100*	0.115*	0.095*	0.005*			
MO	0.087*	0.100*	0.114*	0.097*	0.006	0.003		
BA	0.079*	0.093*	0.108*	0.091*	0.004	0.001	0.001	
MA	0.084*	0.098*	0.114*	0.094*	0.009*	0.006*	0.004*	0.000

Note: LS = La Serena, PM = Puerto Montt, BE = Bellingham, FB = Fort Bragg, AA = Aarhus, LI = Lille, MO = Montpellier, BA = Barcelona and MA = Málaga. *$P < 0.05$.

computer simulations based on the coalescent process (Hudson 1990). The demographical model and parameter values were chosen to fit the introduction situation and biological model studied in the present paper: a separation time between the source and introduced populations of 100 generations, a stable effective population size of 10^6 diploid individuals in the source population, a founding propagule of 10 individuals with a duration of five generations vs. no founder event, a mutation rate of 9.3×10^{-6}, and a GSM model with a variance equal to 0.36. The model included a single introduced population and two potential source populations, only one being the actual source population. The divergence times between the two potential source populations were chosen so that their level of differentiation, as measured by the mean F_{ST} computed over 10 000 iterations, was low (i.e. between 0.0005 and 0.022). Sample sizes were 50 diploid individuals for each population and nine dinucleotide microsatellite loci. Using formula 9 in Rannala & Mountain (1997), we computed for each iteration the mean individual assignment likelihoods of individuals collected in the introduced population and assigned into the actual source population ($L_{i\to as}$) and into the non actual source population ($L_{i\to ns}$). The actual source population was considered to be identified when $L_{i\to as} > L_{i\to ns}$. The proportion of assignment to the actual source population was computed as the number of times $L_{i\to as} > L_{i\to ns}$ over 10 000 iterations. Because we ran simulations for demographical models both with and without founder event, we could assessed to which extent this demographical event affected the proportion of assignment to the actual source population.

Results

Genetic variation within and between populations

In South American populations, allele number was significantly reduced compared to European populations (Wilcoxon signed rank test, $P < 0.01$). In fact, 75.4% of the alleles present in Europe are absent in South America. Most of the absent alleles are in low frequency in Europe, but 34.4% of those alleles having a frequency higher than 0.1 in Europe were missing as well (see supplementary data). Mean allele number did not differ significantly between the two American hemispheres ($P > 0.9$). Moreover, most of the alleles (77%) present in the New World were common to both colonized areas, while 20% were found in only one locality and always at low frequency (≤ 0.035). Mean expected heterozygosity was significantly reduced in South America when compared to Europe ($P < 0.05$), but not when compared to North America ($P > 0.5$). F_{ST} values were larger between South American populations than between North American populations. All comparisons including La Serena showed higher differentiation, probably because it is the lowest latitude population in the southern hemisphere and is subjected to harsh dry conditions. Nonetheless F_{ST} values between American and European populations were much larger than those within and between the American continents (Table 2).

The high proportion of alleles common to both colonized areas, in combination with the low pairwise F_{ST} values between populations from the New World, suggests that the introductions of *D. subobscura* into North and South America were not independent events, but rather correspond to serial introduction events [i.e. Europe → (South America or North America) → (North America or South America)]. The NJ tree (Fig. 1) confirmed the nonindependence of North and South American populations, because all four New World populations grouped together. The much lower genetic variability in American relative to European populations suggests that a severe bottleneck occurred at least during the first introduction event. The similar level of within-population variability for South and North American populations suggests a bottleneck of low severity for the second introduction event (South America

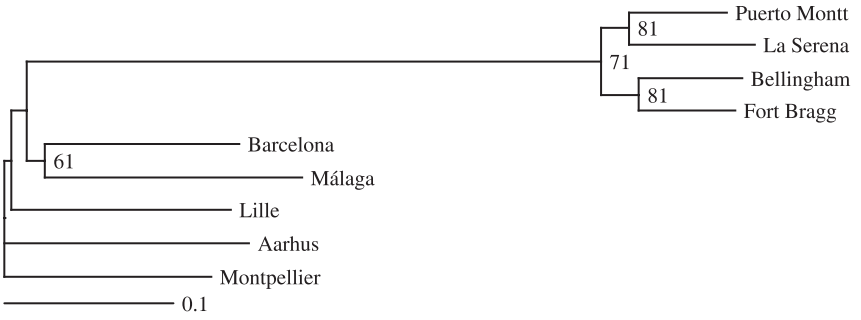


Fig. 1 Neighbour-joining tree based on the chord distance of Cavalli-Sforza & Edwards (1967). Note: Bootstrap values computed over 1000 replications are given as percentage (only values > 50% are shown).

→ North America or North America → South America). However, none of the traditional treatments used here, including the construction of a NJ tree, allows differentiating between a South America → North America and a North America → South America introduction scenario.

Identification of a European source population

None of the above traditional treatments suggested the likely European source of the New World populations. In particular, the low bootstrap values among European populations, reflecting the low genetic differentiation between European populations (i.e. $0.000 < F_{ST} < 0.009$; Table 2), makes such inferences uncertain when based on a NJ tree (Fig. 1). Methods based on mean individual assignment statistics turned out to be more efficient in this context. Computer simulations showed a high probability that the mean individual assignment likelihood of introduced individuals is larger when referring to the actual source population ($L_{i\to as}$) than to the non actual source population ($L_{i\to ns}$), even for very low level of differentiation between the actual and nonactual source populations (e.g. $P = 0.99$ and $P = 0.90$ for $F_{ST} = 0.01$, when the introduced population endured no founder event and a strong founder event, respectively; Fig. 2). This probability was always lower when the introduced population endured a strong founder event than when no founder event occurred. This is due to a larger variance of assignment likelihood values in the founder event scenario, so that, even if mean and modal values remain similar for both scenarios, the probability that by chance $L_{i\to as} < L_{i\to ns}$ is larger under the scenario with strong founder event (see framed distributions of $L_{i\to as}$ and $L_{i\to ns}$ in Fig. 2). However, the proportion of correct assignment remains high under the founder event scenario, even for low F_{ST} values ($P > 0.95$ for $F_{ST} > 0.012$).

Applying this approach to our microsatellite dataset, we found that mean individual assignment likelihoods of New World individuals were the highest when referring to the population from Barcelona, followed by Montpellier (Table 3). A North European origin (i.e. Aarhus or Lille) was particularly unlikely as compared to a southwestern

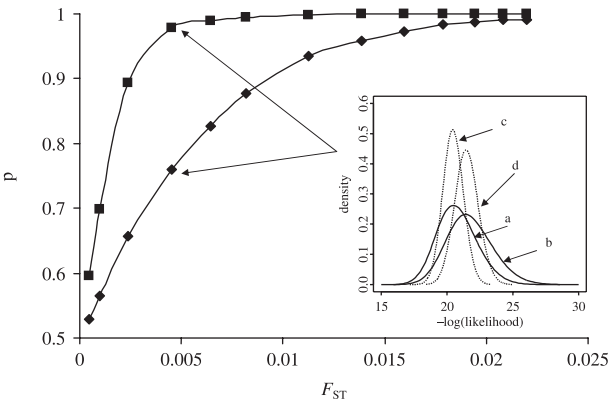


Fig. 2 Proportion of assignment of an introduced population to the actual source population. Note: p = proportion of times the introduced population has been assigned to its actual source population according to assignment likelihood statistics ($L_{i\to s}$). F_{ST} = mean F_{ST} values between the two possible source populations. The introduced population endured a strong founder event (diamonds) or no founder event (squares) immediately after introduction. Likelihood distributions for assignment to the correct and incorrect source population is given for mean $F_{ST} = 0.0045$. Curves a and b = $L_{i\to s}$ distributions for assignment to the actual and non actual source population for an introduction with founder event, respectively. Curves c and d = $L_{i\to s}$ distributions for assignment to the actual and non actual source population for an introduction without founder event, respectively. All values computed over 10 000 iterations (see text for details on the simulation process).

Mediterranean one (i.e. Barcelona or Montpellier). Hence, in further treatments we considered that, among the populations sampled herein, Barcelona is the most representative single population source of the New World populations.

Inferences using approximate Bayesian computation

Our ABC framework allowed discrimination among our eight introduction scenarios (Table 4). Posterior probabilities clearly rejected scenarios assuming independent introductions from Europe as well as scenarios assuming no founder events (posterior probabilities < 0.01 or < 0.002

Table 3 Mean individual assignment likelihood of introduced populations to different European populations

		Potential European source populations (North → South geographical gradient)				
		Aarhus	Lille	Montpellier	Barcelona	Málaga
Introduced population	Puerto Montt (Nc = 48)	22.055 (2.638)	20.282 (2.244)	19.588 (2.197)	18.853 (2.014)	19.849 (2.480)
	SA (Nc = 94)	22.324 (2.949)	20.683 (2.509)	19.804 (2.319)	19.097 (2.176)	20.044 (2.327)
	Bellingham (Nc = 39)	20.684 (2.832)	19.955 (2.284)	18.543 (2.185)	17.859 (1.676)	19.304 (1.946)
	NA (Nc = 68)	21.168 (3.061)	20.254 (2.327)	19.011 (2.235)	18.353 (1.904)	19.549 (2.005)

Note: $-\log_{10}$ of the mean individual likelihood values indicated; standard deviations between parentheses; Nc = number of individuals with genotypes completed at all nine loci. Only Puerto Montt and Belligham were assayed alone because according to collection data, they are our best available samples to represent the initial introduced populations for South and North America, respectively. SA = pool of South American individuals (Puerto Montt + Serena), and NA = pool of North American individuals (Bellingham + Fort Bragg).

Table 4 Comparison of introduction scenarios

		Posterior probabilities					
		Sample set 1		Sample set 2		Sample set 3	
Scenario #	Introduction events	$n_\delta = 100$	$n_\delta = 500$	$n_\delta = 100$	$n_\delta = 500$	$n_\delta = 100$	$n_\delta = 500$
1	EU → SA → NA with founder events	0.93	0.856	0.82	0.736	0.78	0.744
2	EU → SA → NA without founder events	< 0.01	< 0.002	< 0.01	< 0.002	< 0.01	< 0.002
3	EU → NA → SA with founder events	0.07	0.144	0.18	0.264	0.22	0.256
4	EU → NA → SA without founder events	< 0.01	< 0.002	< 0.01	< 0.002	< 0.01	< 0.002
5	EU → SA + EU → NA with founder events	< 0.01	< 0.002	< 0.01	< 0.002	< 0.01	< 0.002
6	EU → SA + EU → NA without founder events	< 0.01	< 0.002	< 0.01	< 0.002	< 0.01	< 0.002
7	EU → NA + EU → SA with founder events	< 0.01	< 0.002	< 0.01	< 0.002	< 0.01	< 0.002
8	EU → NA + EU → SA without founder events	< 0.01	< 0.002	< 0.01	< 0.002	< 0.01	< 0.002

Note: Prior probability was the same for each scenario and hence equal to 1/8 = 0.125. Posterior probabilities were computed from the n_δ smallest Euclidian distances among 8×10^7 values (10^7 values per scenario; see text for details). EU: Europe, SA: South America, NA: North America. Population acronyms as in Table 2. ALL = all European population samples. Sample set 1: EU = BA, SA = PM, NA = BE. Sample set 2: EU = BA + MO, SA = PM + LS, NA = BE + FB. Sample set 3: EU = ALL, SA = PM + LS, NA = BE + FB.

depending on the number of lowest Euclidian distances, n_δ, considered). For both $n_\delta = 100$ and $n_\delta = 500$, the strongest support was obtained for the scenario 1 (serial introductions with founder events from Europe into South America, and then from South America into North America). Depending on the population sampling set considered, posterior probabilities ranged from 0.78 to 0.93 for this scenario when $n_\delta = 100$ and from 0.736 to 0.856 when $n_\delta = 500$.

Because scenario 1 was judged superior, we will present posterior distributions of parameters only for this scenario.

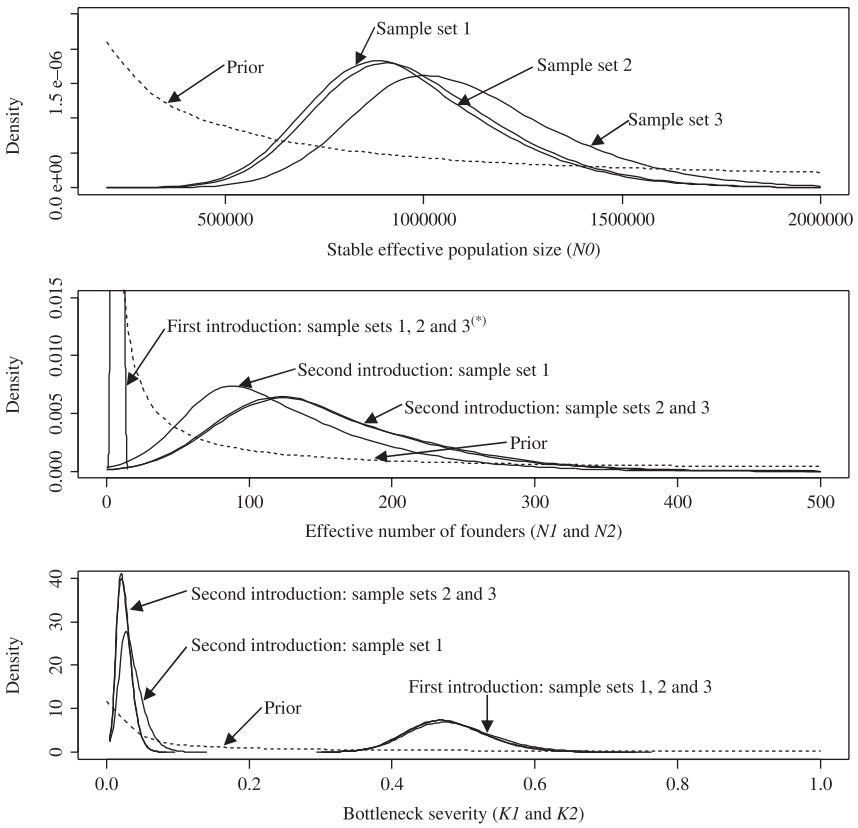


Fig. 3 Prior and posterior density curves for stable population size (*N0*), number of founders in the first (*N1*) and second (*N2*) introduction and bottleneck severity during the first (*K1*) and second (*K2*) introduction. Note: The short dashed and solid lines correspond to the prior and posterior density curves, respectively. Sample set 1 = most likely source and introduced population samples (i.e. Barcelona for Europe, Puerto Montt for South America and Bellingham for North America); Sample set 2 = Montpellier + Barcelona as the European source, Puerto Montt + Serena as the South American sample, and Bellingham + Fort Bragg as the North American sample; Sample set 3 = all European samples (i.e. Aarhus + Lille + Montpellier + Barcelona + Málaga) as the European source, Puerto Montt + Serena as the South American sample, and Bellingham + Fort Bragg as the North American sample. All prior and posterior densities are based on 100 000 and 10 000 values, respectively. (*)Highest density values for the number of founders for the first introduction reach *c*. 0.16 for all sample sets.

Note that posterior distributions were however similar when scenario 3 (serial introductions with founder events from Europe into North America, and then from North America into South America) was used instead of scenario 1 (results not shown). Appendix I shows that none of the observed summary statistics are outliers in their corresponding marginal distributions estimated through simulations, with quantile values ranging from 0.052 to 0.921. This indicates that the posterior distributions of parameters provide sensible values of all summary statistics when compared to observed values.

Figure 3 shows that the posterior density curves of the stable effective population size (*N0*), the number of founders (*N1* and *N2*), and the bottleneck severity (*K1* and *K2*) differ noticeably from the priors. This means that the genetic data contain substantial information for those parameters. Posterior distributions obtained for the single or pooled population sampling sets were similar for *N1* and *K1*, and gave (only) slightly larger values for *N2* and *N0* (lower for *K2*), when pooled samples are considered (Fig. 3). Therefore, mode, mean and quantiles values of posteriors are detailed in Table 5 for the sample set 1 only (Barcelona → Puerto Montt → Bellingham). Posterior distributions support large *N0* values around one million of individuals for mean, mode and 50% quantile values of the source population, with 5% and 95% quantiles around 6×10^5 and 1.4×10^6, respectively. The bottleneck severity was more than one order of magnitude larger for the first introduced population in South America than in the subsequent introduced population in North America. In agreement with this, very small numbers of founders were supported for the first introduction event in South America (*N1*: 50% quantile value of only seven effective individuals), with a weak dispersion of values (5% and 95% quantiles of 4 and 11, respectively). A much larger founding cohort was supported for the second introduction event, into North America (*N2*: mean, mode and 50% quantile values around 100–150 individuals depending on the sample set considered), with 5% and 95% quantiles around 40 and 250, respectively.

The posterior density curves of the marker parameters did not differ noticeably from the priors (Table 5). Thus, the genetic data contained little information on the mutation rates for the number of repeats and insertion-deletions in the flanking regions, and on the variance of the geometric distribution of mutation sizes.

Robustness of inferences

The assumption of different priors for demographical or marker parameters did not change the conclusion that the most supported model is scenario 1 (posterior probabilities

Table 5 Mode, mean and quantile values of the priors and posteriors for the demographical and marker parameters under *D. subobscura* introduction scenario 1

		Mean	Mode	Q5%	Q50%	Q95%
N0	Prior	781 840	200 000	224 290	633 830	1 778 460
	Posterior	952 000	885 230	616 470	921 290	1 392 450
N1	Prior	89.8	2.0	2.6	31.4	378.0
	Posterior	7.3	7.1	3.8	7.2	11.2
N2	Prior	89.8	2.0	2.6	31.4	378.0
	Posterior	121.4	89.3	43.1	106.3	245.2
K1	Prior	0.315	0.004	0.009	0.107	1.313
	Posterior	0.484	0.474	0.393	0.480	0.590
K2	Prior	0.315	0.004	0.009	0.107	1.313
	Posterior	0.035	0.027	0.013	0.032	0.066
μ_{rn}	Prior	9.3×10^{-6}	7.3×10^{-6}	2.5×10^{-6}	8.3×10^{-6}	2×10^{-5}
	Posterior	9.2×10^{-6}	6.8×10^{-6}	2.6×10^{-6}	8.2×10^{-6}	1.9×10^{-5}
μ_{id}	Prior	2.5×10^{-8}	1.4×10^{-8}	1.3×10^{-9}	1.7×10^{-8}	7.5×10^{-8}
	Posterior	2.7×10^{-8}	1.6×10^{-8}	1.1×10^{-9}	1.9×10^{-8}	8.1×10^{-8}
σ^2	Prior	0.360	0.000	0.019	0.249	1.074
	Posterior	0.360	0.000	0.020	0.255	1.073

Note: Results for the sample set 1 only (Barcelona → Puerto Montt → Bellingham). *K1* = *D1*/*N1* and *K2* = *D2*/*N2* measure the bottleneck severity for the first and second introduced population, respectively. Q5%, Q50% and Q95% = 5%, 50%, and 95% quantile values, respectively. All values are estimated from 100 000 and 10 000 values for priors and posteriors, respectively.

between 0.84 and 0.96 for $n_\delta = 50$; Table 6). With regards to demographical parameters estimation, Table 7 shows that the assumption of different priors had limited effect on the posterior distributions for bottleneck severities (*K1* and *K2*), the number of founders in South America (*N1*), and to a lesser extent in North America (*N2*). On the other hand, non-negligible effect could be observed for the stable effective population size (*N0*), especially when a SMM was assumed (cf. the support for higher values considerably increased). Using a different prior for μ_{rn} and assuming a different mutation model confirm that the genetic data contained little information on mutational processes (i.e. similar prior and posterior distributions were obtained; results not shown).

Hence, although the effect of some prior assumptions on posterior distributions of at least one demographical parameter (*N0*) is not negligible, we found that variation between estimated posterior distributions are limited, indicating resilience of our inferences to changes in demographical and marker priors, at least to those tested here. In particular, results remained in agreement with the general conclusions that: (i) scenario 1 is the most supported scenario; (ii) stable effective population sizes are very large in *D. subobscura* (in the order of one million individuals); and (iii) bottleneck severity is more than one order of magnitude larger for the first introduced population in South America than in the subsequent introduced population in North America due to a propagule size notably smaller in South America (i.e. only a few effective founders) than in North America (i.e. around 100–150 effective founders).

Table 6 Robustness on prior choice for discriminating among introduction scenarios

	Posterior probabilities					
	Scenario 1		Scenario 3		Scenario 2, 4–8	
Prior set	$n_\delta = 50$	$n_\delta = 250$	$n_\delta = 50$	$n_\delta = 250$	$n_\delta = 50$	$n_\delta = 250$
1	0.94	0.872	0.06	0.128	< 0.02	< 0.004
2	0.84	0.816	0.16	0.184	< 0.02	< 0.004
3	0.88	0.808	0.12	0.192	< 0.02	< 0.004
4	0.96	0.872	0.04	0.128	< 0.02	< 0.004
5	0.88	0.820	0.12	0.180	< 0.02	< 0.004
6	0.92	0.832	0.08	0.168	< 0.02	< 0.004

Note: Prior set 1: standard priors as described in Table 1. Prior set 2: uniform distributions bounded between 200 000 and 2×10^6 diploid individuals for *N0*, and between 2 and 500 for *N1* and *N2*. Prior set 3: stepwise mutation model (SMM) for all markers. Prior set 4: GSM with constraints on allele size (i.e. 30 possible continuous allelic states). Prior set 5: individual locus mutation rates of repeat numbers μ_{rn} drawn in an exponential (1, 107 310). Prior set 6: insertion-deletion mutation rate in flanking regions assumed to be equal to zero ($\mu_{id} = 0$). Posterior probabilities were computed from the n_δ smallest Euclidian distances among 4×10^7 values (5×10^6 values per scenario). Prior probability was the same for each scenario and hence equal to 1/8 = 0.125.

Discussion

The colonization of North and South America by *D. subobscura* does not correspond to independent colonizations.

Table 7 Robustness on prior choice for inferences on demographical parameters

	Prior set	Mean	Mode	Q5%	Q50%	Q95%
N0	1	952 000	885 230	616 470	921 290	1 392 450
	2	1 015 070	922 560	658 090	983 780	1 474 700
	3	1 615 720	1 673 900	1 134 880	1 628 490	1 999 850
	4	1 182 310	1 086 390	746 410	1 148 940	1 725 780
	5	1 120 600	1 004 346	628 450	1 086 300	1 716 710
	6	1 069 820	972 500	672 790	1 032 240	1 597 790
N1	1	7.3	7.1	3.8	7.2	11.2
	2	8.5	8.5	4.5	8.5	12.5
	3	7.1	6.8	3.6	6.9	10.9
	4	7.6	7.4	3.9	7.5	11.8
	5	7.5	7.3	3.9	7.4	11.7
	6	7.7	7.4	4.0	7.6	11.7
N2	1	121.4	89.3	43.1	106.3	245.2
	2	135.5	110.8	52.8	127.6	246.3
	3	118.9	86.3	42.6	104.2	242.6
	4	115.5	84.8	41.5	102.6	232.8
	5	112.7	88.8	39.8	99.1	228.5
	6	124.9	91.5	45.1	108.6	256.4
K1	1	0.484	0.474	0.393	0.480	0.590
	2	0.458	0.448	0.376	0.454	0.554
	3	0.488	0.476	0.395	0.483	0.595
	4	0.481	0.474	0.389	0.478	0.585
	5	0.478	0.469	0.387	0.474	0.583
	6	0.479	0.470	0.387	0.475	0.484
K2	1	0.035	0.027	0.013	0.032	0.066
	2	0.033	0.026	0.014	0.030	0.063
	3	0.037	0.028	0.014	0.034	0.069
	4	0.034	0.027	0.013	0.032	0.064
	5	0.037	0.029	0.014	0.034	0.069
	6	0.036	0.029	0.014	0.033	0.067

Note: Posterior distributions have been estimated under introduction scenario 1 and different prior sets. Prior set acronyms as in Table 6 legend. All values are estimated using the sample set 1, from 5000 accepted values.

The ABC framework gave a strong posterior support to the scenario of serial introductions (with founder events) from Europe into South America, and then from South America into North America. This support was robust to different prior assumptions. This result underlines the potential of ABC methods to discriminate between complex evolutionary scenarios, a feature previously illustrated by Miller *et al*. (2005) in a similar context of introduction routes reconstruction for the pest *Diabrotica virgifera*.

In agreement with a non-independence of colonization events, we found that the number of microsatellite alleles shared between both colonized areas is high and that nonshared alleles are always at low frequency. Previous studies, had already noted the striking similarity of North and South American *D. subobscura* populations (Prevosti *et al*. 1988; Mestres *et al*. 1990; Rozas & Aguade 1991; Balanyà *et al*. 1994). However, all of them failed to infer the colonization sequence probably because of the lower level of polymorphism of the markers used in those studies, the low bottleneck severity of the second colonization event and a limited discrimination power of the methods used so far to tackle this question.

Most chromosomal arrangements data supported a western Mediterranean origin of the New World populations of *D. subobscura* (Brncic *et al*. 1981). However, the presence of the O_5 arrangement, which is rare in the Mediterranean but found in New World populations, did not support this hypothesis (Ayala *et al*. 1989). This apparent incompatibility may be explained by the high migration rates between *D. subobscura* ancestral populations (Pascual *et al*. 2001) and the quick changes in frequency on chromosomal arrangements due to selection pressures (Balanyà *et al*. 2006). In the present microsatellite-based study, treatments based on mean individual assignment statistics identified the western Mediterranean (specifically Barcelona among our samples) as the most likely source of the New World *D. subobscura*. We evaluated the robustness of such inference via computer simulations. These simulations showed that in similar introductions with low level of differentiation between potential source

populations, there is a high probability that the mean individual assignment likelihood of introduced individuals would be the largest when referring to the actual source population. Although this probability was always lower if the introduced population endured a strong founder event, it remained high under the founder event scenario even for low F_{ST} values, and hence for F_{ST} values similar to those observed between European populations of *D. subobscura*.

Traditional population genetics treatments gave useful — but only qualitative — insights into demographical parameters of interest (large stable effective population size, but how large? Small number of founders for the first introduction event, but how small? Larger number of founders for the second introduction event, but how large?). In contrast, ABC treatments provided robust quantitative insights into such parameters. ABC posterior distributions suggested a bottleneck severity more than one order of magnitude larger in the first introduced population than in the subsequent introduced population. The propagule size was only approximately seven effective founders for South America (*N1*; 5% and 95% quantiles of 4 and 12, respectively), whereas it reached 100–150 effective founders for North America. The estimated values of *N1* are very similar to the effective number of *D. subobscura* colonists (4–11 individuals) previously estimated by computer simulations using MULTSIM (Noor *et al*. 2000), in a comparison of populations from North America and Europe (Pascual *et al*. 2001). They are also in close agreement with previous estimates obtained from data on chromosomal polymorphism comparing populations from South America and Europe (10–15 individuals; (Brncic *et al*. 1981) and from restriction site polymorphism in the *rp49* region (4–6 individuals; (Rozas & Aguade 1991), comparing populations from North America, South America and Europe. The similar number of founders estimated when comparing either North America or South America with Europe reflects the low bottleneck severity of the second colonization event. Consequently, estimating drift pulse due to founder events in North America alone will be similar to measuring such a pulse in South America alone.

The relevance of the number of founders in shaping diversity has been previously underlined in *D. buzzatii* populations from Europe and Australia, colonized from South America 200 and 65 years ago, respectively, where the loss of allele variation in the latter was smaller due to the larger number of founders (Frydenberg *et al*. 2002). A large number of founders increases the probability of a successful invasion, both by reducing demographical stochasticity (Rouget & Richardson 2003) and by providing a larger pool of selectable genetic variation (Lee 2002). Multiple colonizations from different sources (Caracristi & Schlötterer 2003; Colautti *et al*. 2005) that would increase genetic variation in the colonizing area would also have profound implications as to the probability of the successful establishment and spread of invasive species. Nonetheless, a single propagule of large size and/or multiple invasions are not necessary for a species to successfully invade a new area (Roderick & Navajas 2003). In agreement with this, the *D. subobscura* propagule size was only approximately seven effective individuals in South America and the species does not appear to have made multiple invasions from Europe as no additional Palearctic inversions have ever been detected in intensive surveys made subsequent to the initial one in 1981 (Balanyà *et al*. 2003).

Our ABC estimations showed that the stable effective population size is very large for European *D. subobscura* (in the order of one million individuals). As a consequence, genetic variation in genes that may be involved in adaptive processes is expected to be high in European populations, even for slowly evolving genes. Therefore, even a very low number of founder individuals from Europe is expected to provide a substantial amount of genetic variation in the colonized areas, and thus to foster the invasion success of this species in America. In agreement with this, a severe bottleneck event did not preclude heterozygosity to remain relatively high in both South and North America at microsatellite loci. Moreover, when a species colonizes a novel area the estimated effective number of colonizers can be a nonrandom fraction of the original cohort of colonists. In particular, the more heterozygous individuals in that cohort may have greater chances to reproduce (Grant 2002). *D. subobscura* individuals from the Mediterranean area are highly heterozygous (Pascual *et al*. 2001) consequently, in spite of the small number of colonizers from that area, invasion success could be granted since there would be enough variability in which selection could act.

No other species of the *obscura* group are present in South America (Prevosti *et al*. 1989). Thus invading *D. subobscura* probably did not face interspecific competition with native *Drosophila*, increasing the chances of reproducing and quickly expanding in that new area. As (Brncic & Budnik 1987) pointed out, the success of the colonization of *D. subobscura* can be attributed mostly to the clear differences in seasonality in relation to native *Drosophila* species in South America. On the other hand, other species of the *obscura* group are present in North America. Laboratory experiments demonstrated that these species could act as competitors (Pascual *et al*. 1998), and therefore could slow down the rate of expansion as well as affect the evolution of adaptive traits (Gilchrist *et al*. 2004). However, this does not seem to be the case since the number of effective founders reaching North America from South America was large enough (100–150 effective individuals) to maintain existing genetic variation and hence to have a similar potential for adaptation in both hemispheres.

Our findings demonstrate the utility of using highly variable markers and ABC methods to discriminate between

complex invasion scenarios and give useful quantitative insights into demographical parameters of interest. These markers and methods may prove particularly useful in understanding and tracing the evolutionary history of the ever-growing number of invasive species.

Acknowledgements

This work was financially supported by grant CGL2006-13423 from Ministerio de Ciencia y Tecnologia to LS, grant 2005SGR-00995 from Generalitat de Catalunya, an ANR grant # BLAN-0196-01 to AE, NSF grant DEB 9981598 to RBH, and NSF grant DEB0242313 to GWG. AE thanks Mark Beaumont and Jean-Marie Cornuet for stimulating discussions on ABC methods.

References

Avelar T, Rocha Pité MT, Matos M (1987) Variation of some demographic parameters in *Drosophila simulans*, *D. subobscura* and *D. phalerata* (*Diptera*, *Drosophilidae*) throughout the year under semi-natural conditions in Portugal. *Acta Oecologica*, **8**, 347–356.

Ayala FJ, Serra L, Prevosti A (1989) A grand experiment in evolution: the *Drosophila subobscura* colonization of the Americas. *Genome*, **31**, 246–255.

Balanyà J, Segarra C, Prevosti A, Serra L (1994) Colonization of America by *Drosophila subobscura*: the founder event and a rapid expansion. *Journal of Heredity*, **85**, 427–432.

Balanyà J, Serra L, Gilchrist GW *et al*. (2003) Evolutionary pace of chromosomal polymorphism in colonizing populations of *Drosophila subobscura*: an evolutionary time series. *Evolution*, **57**, 1837–1845.

Balanyà J, Oller JM, Huey RB, Gilchrist GW, Serra L (2006) Global genetic change tracks global climate warming in *Drosophila subobscura*. *Science*, **313**, 1773–1775.

Beaumont MA (1999) Detecting population expansion and decline using microsatellites. *Genetics*, **153**, 2013–2029.

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

Beckenbach AT, Prevosti A (1986) Colonization of North America by the European species, *Drosophila subobscura* and *D. ambigua*. *American Midland Naturalist*, **115**, 10–18.

Begon M (1976) Temporal variations in the reproductive condition of *Drosophila obscura* Fallen and *Drosophila subobscura* Collin. *Oecologia*, **23**, 31–47.

Brncic D, Budnik M (1980) Colonization of *Drosophila subobscura* Collin in Chile. *Drosophila Information Service*, **55**, 20.

Brncic D, Budnik M (1987) Some interactions of the colonizing species *Drosophila subobscura* with local *Drosophila* fauna in Chile. *Genética Ibérica*, **39**, 249–267.

Brncic D, Prevosti A, Budnick M, Monclus M, Ocaña J (1981) Colonization of *Drosophila subobscura* in Chile. — I. First population and cytogenetic studies. *Genetica*, **56**, 3–9.

Caracristi G, Schlötterer C (2003) Genetic differentiation between American and European *Drosophila melanogaster* populations could be attributed to admixture of African alleles. *Molecular Biology and Evolution*, **20**, 792–799.

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

Colautti RI, Manca M, Viljanen M *et al*. (2005) Invasion genetics of the Eurasian spiny waterflea: evidence for bottlenecks and gene flow using microsatellites. *Molecular Ecology*, **14**, 1869–1879.

Davies N, Villablanca FX, Roderick GK (1999) Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics. *Trends in Ecology and Evolution*, **14**, 17–21.

Estoup A, Clegg SM (2003) Bayesian inferences on the recent island colonization history by the bird *Zosterops lateralis lateralis*. *Molecular Ecology*, **12**, 657–674.

Estoup A, Wilson IJ, Sullivan C, Cornuet JM, Moritz C (2001) Inferring population history from microsatellite and enzyme data in serially introduced cane toads, *Bufo marinus*. *Genetics*, **159**, 1671–1687.

Estoup A, Beaumont M, Sennedot F, Moritz C, Cornuet JM (2004) Genetic analysis of complex demographic scenarios: spatially expanding populations of the cane toad, *Bufo marinus*. *Evolution*, **58**, 2021–2036.

Feldman MW, Bergman A, Pollock DD, Goldstein DB (1997) Microsatellite genetic distances with range constraints: analytic description and problems of estimation. *Genetics*, **145**, 207–216.

Frydenberg J, Pertoldi C, Dahlgaard J, Loeschcke V (2002) Genetic variation in original and colonizing *Drosophila buzzati* populations analysed by microsatellite loci isolated with a new PCR screening method. *Molecular Ecology*, **11**, 181–190.

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

Gilchrist GW, Huey RB, Balanyà J, Pascual M, Serra L (2004) A time series of evolution in action: latitudinal cline in wing size in South American *Drosophila subobscura*. *Evolution*, **58**, 768–780.

Grant PR (2002) Founder effects and silvereyes. *Proceedings of the National Academy of Sciences USA*, **99**, 7818–7820.

Hamilton G, Currat M, Ray N, Heckel G, Beaumont MA, Excoffier L (2005a) Bayesian estimation of recent migration rates after spatial expansion. *Genetics*, **170**, 409–417.

Hamilton G, Stoneking M, Excoffier L (2005b) Molecular analysis reveals tighter social regulation of immigration in patrilocal populations than in matrilocal populations. *Proceedings of the National Academy of Sciences USA*, **102**, 7476–7480.

Hudson RR (1990) Gene genealogies and the coalescent process. In: *Oxford Surveys in Evolutionary Biology* (eds Futuyama DJ, Antonovics J), pp. 1–44. Oxford University Press, Oxford.

Huey RB, Gilchrist GW, Carlson ML, Berrigan D, Serra L (2000) Rapid evolution of a geographic cline in size in an introduced fly. *Science*, **287**, 308–309.

Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. *Journal of Computational Graphics and Statistics*, **5**, 299–314.

Krimbas CB (1993) *Drosophila subobscura. Biology, Genetics and Inversion Polymorphism*. Verlag Dr, Kovac, Hamburg.

Langella O (2002) Populations. http://bioinformatics.org/project/?group_id=84.

Latorre A, Moya A, Ayala FJ (1986) Evolution of mitochondrial DNA in *Drosophila subobscura*. *Proceedings of the National Academy of Sciences USA*, **83**, 8649–8653.

Lee CE (2002) Evolutionary genetics of invasive species. *Trends in Ecology and Evolution*, **17**, 386–391.

Lewin B (1994) *Genes IV*, Ed. 4. John Willey & Sons, New York.

Loader CR (1996) Local likelihood density estimation. *Annals of Statistics*, **24**, 1602–1618.

Mayr E (1954) Changes of genetic environment and evolution. In: *Evolution as a Process* (eds Huxley J, Hardy AC, Ford EB), pp. 157–180. Allen & Unwin, London.

Mestres F, Serra L (1991) Lethal allelism in *Drosophila subobscura*: difficulties in the estimation of certain population parameters. *Journal of Zoological Systematics and Evolutionary Research*, **29**, 264–279.

Mestres F, Pegueroles G, Prevosti A, Serra L (1990) Colonization of America by *Drosophila subobscura*: lethal genes and the problem of the O_5 inversion. *Evolution*, **44**, 1823–1836.

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

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

Noor MF, Pascual M, Smith KR (2000) Genetic variation in the spread of *Drosophila subobscura* from a nonequilibrium population. *Evolution*, **54**, 696–703.

Pascual M, Serra L, Ayala FJ (1998) Interspecific laboratory competition of the recently sympatric species *Drosophila subobscura* and *Drosophila pseudoobscura*. *Evolution*, **52**, 269–274.

Pascual M, Schug MD, Aquadro CF (2000) High density of long dinucleotide microsatellites in *Drosophila subobscura*. *Molecular Biology and Evolution*, **17**, 1259–1267.

Pascual M, Aquadro CF, Soto V, Serra L (2001) Microsatellite variation in colonizing and Palearctic populations of *Drosophila subobscura*. *Molecular Biology and Evolution*, **18**, 731–740.

Pascual M, Mestres F, Serra L (2004) Sex-ratio in natural and experimental populations of *Drosophila subobscura* from North America. *Journal of Zoological Systematics and Evolutionary Research*, **42**, 33–37.

Petrov DA, Hartl DL (1998) High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species group. *Molecular Biology and Evolution*, **15**, 293–302.

Prevosti A, Ribó G, Serra L *et al*. (1988) Colonization of America by *Drosophila subobscura*: experiment in natural populations that supports the adaptive role of chromosomal-inversion polymorphism. *Proceedings of the National Academy of Sciences USA*, **85**, 5597–5600.

Prevosti A, Serra L, Aguadé M *et al*. (1989) Colonization and establishment of the Palearctic species *Drosophila subobscura* in North and South America. In: *Evolutionary Biology of Transient Unstable Populations* (ed. Fontdevila A), pp. 114–129. Springer-Verlag, Berlin.

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

Roderick G, Navajas M (2003) Genes in new environments: Genetics and evolution in biological control. *Nature Reviews Genetics*, **4**, 889–899.

Rouget M, Richardson DM (2003) Inferring process from pattern in plant invasions: a semimechanistic model incorporating propagule pressure and environmental factors. *American Naturalist*, **162**, 713–724.

Rozas J, Aguadé M (1991) Using restriction-map analysis to characterize the colonization process of *Drosophila subobscura* on the American continent. — I. rp49 region. *Molecular Biology and Evolution*, **8**, 447–457.

Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TF, Aquadro CF (1998) The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. *Molecular Biology and Evolution*, **15**, 1751–1760.

Stephens M (2003) Inference under the coalescent. In: *Handbook of Statistical Genetics* (eds Balding DJ, Bishop M, Cannings C), pp. 213–238. Wiley, Chichester.

Thornton K, Andolfatto P (2006) Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. *Genetics*, **172**, 1607–1619.

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

Weir BS (1996). *Genetic Data Analysis II*. Sinauer, Sunderland, Massacusetts.

Wright SI, Bi IV, Schroeder SG *et al*. (2005) The effect of artificial selection on the maize genome. *Science*, **308**, 1310–1314.

M. Pascual research focuses on molecular population studies of marine and model organisms. M. P. Chapuis is a post-doctoral researcher mainly studying population genetics and evolution of density-dependent phenotypic traits in outbreeding insects. F. Mestres research interest is the knowledge of the genetic structure of *Drosophila subobscura* based in different genetic markers (lethal genes, chromosomal inversions and nucleotide sequences). J. Balanyà studies the long term changes and the adaptive significance of the chromosomal inversion polymorphism of *D. subobscura*. R. Huey is an evolutionary physiologist interested in ecological and evolutionary responses of *Drosophila* and lizards to temperature change. G. W. Gilchrist is an evolutionary biologist who studies evolutionary responses to environmental change in time and space. L. Serra research focuses on the genetic basis of thermal adaptation and contingency versus adaptation in patterns of geographic variability of *D. subobscura*. A. Estoup's current research focuses mainly on the evolutionary biology of invading species.

Supplementary information

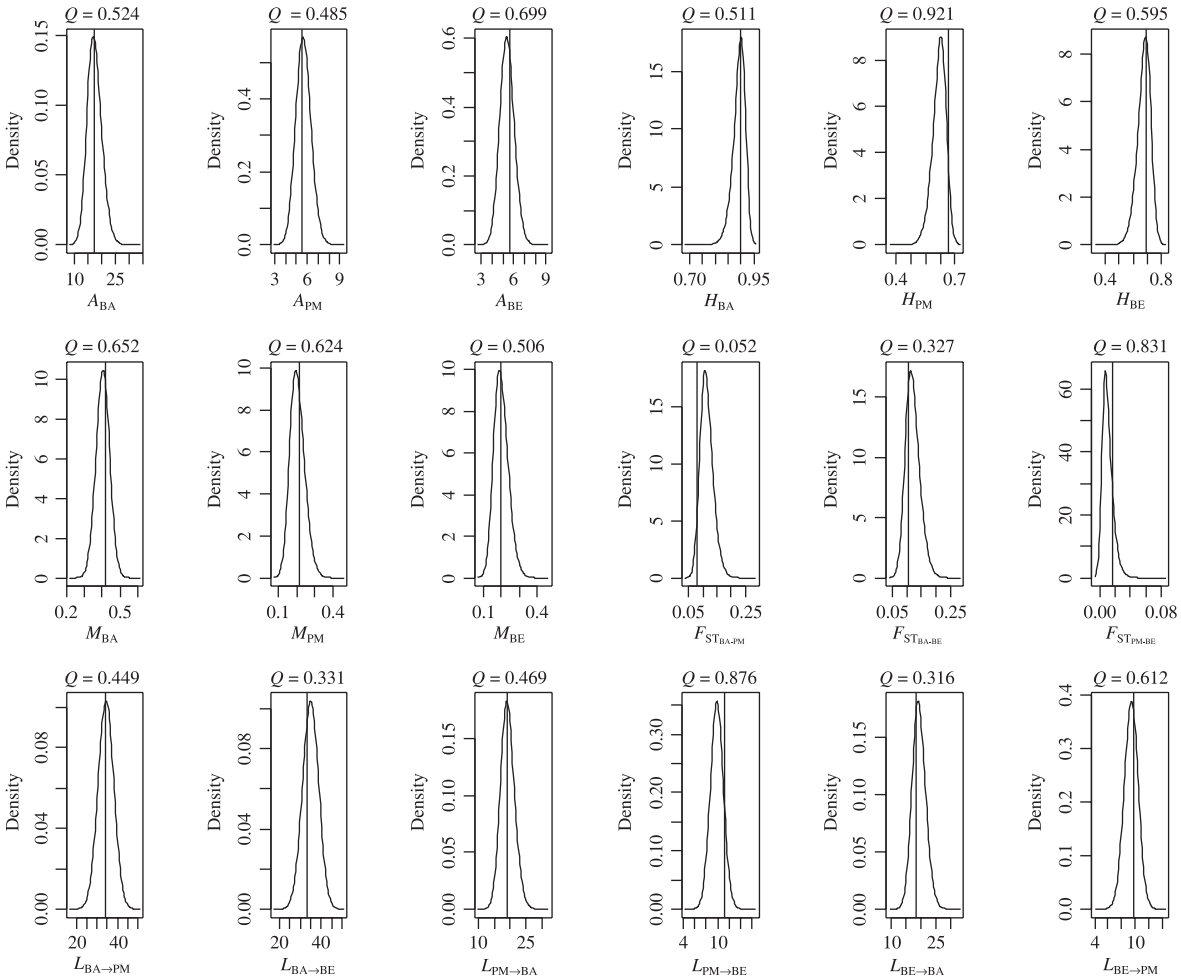
The following supplementary material is available for this article:

Supplementary data Sample size in number of genes (n), number of alleles (A) and allele frequencies for each of the nine microsatellite loci genotyped in two South American, two North American and five European populations of *D. subobscura*.

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2007.03336.x
(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Appendix I Marginal distributions of summary statistics



Note: Marginal distributions were obtained by running simulations under the scenario 1 and using the sample set 1 (i.e. Barcelona for Europe, Puerto Montt for South America and Bellingham for North America). 10^6 sets of parameter values were randomly drawn with replacement among the 10 000 accepted sets of parameter values (i.e. with a Euclidian distance < δ) adjusted through the local linear regression step. A = mean number of alleles per locus and population, H = mean expected heterozygosity, M = mean ratio of the number of alleles over the range of allelic sizes, F_{ST} = genetic *differentiation* between pairs of populations, $L_{i\to j}$ = mean assignment likelihood of individuals collected in population i and assigned into population j. Vertical lines correspond to the observed values of the summary statistics. Q = quantile values of the observed summary statistics in the marginal distributions. BA = Barcelona, PM = Puerto Montt, BE = Bellingham.