Isolation and gene flow: inferring the speciation history of European house mice

Ludovic Duvaux, Khalid Belkhir, Matthieu Boulesteix, Pierre Boursot

MOLECULAR ECOLOGY

Molecular Ecology (2011) **20**, 5248–5264 doi: 10.1111/j.1365-294X.2011.05343.x

Isolation and gene flow: inferring the speciation history of European house mice

LUDOVIC DUVAUX[1], KHALID BELKHIR, MATTHIEU BOULESTEIX[2] and PIERRE BOURSOT
Université Montpellier 2, CNRS UMR 5554, Institut des Sciences de l'Evolution, CC063, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France

Abstract

Inferring the history of isolation and gene flow during species differentiation can inform us on the processes underlying their formation. Following their recent expansion in Europe, two subspecies of the house mouse (*Mus musculus domesticus* and *Mus musculus musculus*) have formed a hybrid zone maintained by hybrid incompatibilities and possibly behavioural reinforcement, offering a good model of incipient speciation. We reconstruct the history of their divergence using an approximate Bayesian computation framework and sequence variation at 57 autosomal loci. We find support for a long isolation period preceding the advent of gene flow around 200 000 generations ago, much before the formation of the European hybrid zone a few thousand years ago. The duration of the allopatric episode appears long enough (74% of divergence time) to explain the accumulation of many post-zygotic incompatibilities expressed in the present hybrid zone. The ancient contact inferred could have played a role in mating behaviour divergence and laid the ground for further reinforcement. We suggest that both subspecies originally colonized the Middle East from the northern Indian subcontinent, *domesticus* settling on the shores of the Persian Gulf and *musculus* on those of the Caspian Sea. Range expansions during interglacials would have induced secondary contacts, presumably in Iran, where they must have also interacted with *Mus musculus castaneus*. Future studies should incorporate this possibility, and we point to Iran and its surroundings as a hot spot for house mouse diversity and speciation studies.

Keywords: allopatry, approximate Bayesian computation, gene flow, isolation, mouse, speciation

Received 16 August 2010; revision received 26 September 2011; accepted 2 October 2011

Introduction

Understanding the conditions under which new species can arise remains a central and debated question. Numerous authors addressed the question, and Mayr (1942) notably postulated that allopatric speciation was the most plausible and common mode of speciation. This has for some time become the 'null hypothesis' (Coyne & Orr 2004) and had the effect to enhance the number of studies on 'nonallopatric speciation' with gene flow between nascent species, with the danger of eluding the question of the exact conditions of gene flow and of what it can tell us about the speciation process (Butlin *et al.* 2008). Recent advances on these matters have resulted from theoretical (Gavrilets 2003; Coyne & Orr 2004; Barton & de Cara 2009) and empirical studies (Via 2001; Stump *et al.* 2005; Roberts *et al.* 2009; Matute 2010; Matute *et al.* 2010). The improvement of methods using coalescent genealogy samplers (Beaumont *et al.* 2002; Becquet & Przeworski 2007; Hey & Nielsen 2007; Kuhner 2009) now allows inferring migration between incipient species. Determining whether a speciation event is fully allopatric over time and space can seem meaningless 'because migrants

Correspondence: Pierre Boursot, Fax: +33(0)467144554; E-mail: pierre.boursot@univ-montp2.fr
Ludovic Duvaux, Fax: +44(0)114 2220002; E-mail: l.duvaux@sheffield.ac.uk
[1]Present addresses: Department of Animal and Plant Sciences, University of Sheffield, Sheffield S102TN, UK.
[2]Laboratoire de Biométrie et Biologie Evolutive, CNRS UMR5558, Université Lyon 1, Villeurbanne, France.

occasionally cross even the most extreme barriers' (Butlin *et al.* 2008) as shown by a growing number of studies from a wide range of plants and animals (Geraldes *et al.* 2008; Niemiller *et al.* 2008; Strasburg & Rieseberg 2008; Faure *et al.* 2009 among others). However, because theory predicts that the fixation of genetic incompatibilities is easier without gene flow (Turelli *et al.* 2001; Barton & de Cara 2009) and because the evolutionary forces at play may differ substantially depending on the migration regime accompanying differentiation, it remains important to characterize this regime and its timing (Won & Hey 2005; Nadachowska & Babik 2009; Li *et al.* 2010), if only to infer historical aspects of a differentiation process, and eventually gain understanding of its geographical onset over time.

The house mouse is our model to attempt to do so here. *Mus musculus* is genetically structured in at least three evolutionary units that we will consider subspecies but are sometimes referred to as full species in the literature. *Mus musculus domesticus* has been described around the Mediterranean and in Western Europe, *Mus musculus musculus* from Eastern Europe to Northern Asia, and *Mus musculus castaneus* in South Asia. The history of their differentiation is mainly documented for old and very recent periods (reviewed in Boursot *et al.* 1993). On the basis of allozyme frequencies and mtDNA variation, the cradle of the species was inferred to lie in a region from the Middle East to northern India (Boursot *et al.* 1996; Din *et al.* 1996; Prager *et al.* 1998). DNA–DNA hybridization and sequencing of a limited number of loci suggest that they diverged between 350 000 and 1 Ma (She *et al.* 1990; Boursot *et al.* 1996; Suzuki *et al.* 2004; Geraldes *et al.* 2008). Only two aspects of their more recent biogeographic history are well documented: one is the Neolithic expansion of *domesticus* in the Levant in association with humans developing agriculture and its later spread to Western and Northern Europe presumably in relation to human trading, as documented in the archaeozoological record (Cucchi *et al.* 2005). Accordingly, mtDNA sequence variation in *domesticus* indicates a major expansion that was hypothesized to correspond to the Neolithic expansion (Gündüz *et al.* 2005; Rajabi-Maham *et al.* 2008). A second aspect is that *musculus* must have simultaneously also extended its range westward as it now occupies central Europe and comes into contact with *domesticus* along a well-studied hybrid zone crossing Europe from the Jutland Peninsula to the Black Sea (reviewed in Boursot *et al.* 1993; Sage *et al.* 1993). Extensive studies of this hybrid zone have revealed the characteristics of a tension zone maintained by a balance between dispersal and selection against hybrids at many loci (Dod *et al.* 1993; Raufaste *et al.* 2005; Macholán *et al.* 2007; Teeter *et al.* 2010). Reduced fertility of hybrids has been documented, especially males (Storchova *et al.* 2004; Britton-Davidian *et al.* 2005); however, this character appears multifactorial and not to be fixed (Vyskocilova *et al.* 2005; Good *et al.* 2008). One gene strongly involved in such a phenotype was recently identified (Mihola *et al.* 2009) and curiously is a major determinant of recombination hot spot location in the genome of both mice and humans (Myers *et al.* 2009; Baudat *et al.* 2010). In addition, there is growing evidence of diverging mating recognition systems between these subspecies (Talley *et al.* 2001; Smadja *et al.* 2004; Bimova *et al.* 2005; Smadja & Ganem 2007; Laukaitis *et al.* 2008). Therefore, these subspecies appear an excellent model of incipient speciation. Reproductive isolation is strong enough that their genomes do not freely admix in the hybrid zone; many postzygotic incompatibility factors have accumulated, and these are more likely to have arisen during long periods of isolation. Behavioural studies suggest a divergence of mating behaviour, which could have occurred in isolation owing to independent sexual selection. However, the suggestion of reinforcement in the present hybrid zone questions the possible role of past episodes of hybridization in this aspect of the divergence of the reproductive system of these subspecies. A better understanding of the history of their differentiation would therefore be valuable. In a recent study, Geraldes *et al.* (2008) fitted an isolation-with-migration model (Nielsen & Wakeley 2001; Hey & Nielsen 2004) to infer the demographic history of *Mus musculus* by accounting for migration after divergence. They inferred significant although moderate gene flow between subspecies and revealed some asymmetries, but the simplicity of the underlying model prevented any detailed interpretation in relation to palaeobiogeography and to the possibility of episodes of isolation and admixture and their duration. A different approach was taken by Pool & Nielsen (2009): based on SNPs distributed along the genome and a small sample of each subspecies, they used the distribution of the sizes of inferred migrant tracts to give evidence for a very recent increase in migration rate between these subspecies, in historical times and presumably in relation to the recent exponential development of human activities. The low density of SNPs prevented inferences about older events though, and the relevance of these very recent conditions to the evolution of reproductive isolation in these mice is not clear.

One possibility to infer complex demographic scenarios is to use an approximate Bayesian computation (ABC) framework. The ABC approaches do not need the computation of likelihood and can thus accommodate complex models. Since their first use with genetic data (Tavare *et al.* 1997; Pritchard *et al.* 1999), they have been continuously improved (Beaumont *et al.* 2002; Wegmann *et al.* 2009; Blum & François 2010; Leuenberger & Wegmann 2010) and claimed to be potentially as efficient as

likelihood-based methods (Excoffier *et al.* 2005; Beaumont 2008). They were successfully used for various inferences relating to invasive expansion (Estoup *et al.* 2004) and other demographic histories (Thornton & Andolfatto 2006; Fagundes *et al.* 2007; Ross-Ibarra *et al.* 2008), radiation and phylogeography (Palero *et al.* 2009), and speciation (Li *et al.* 2010). Because ABC methods can statistically compare different models, they provide a rigorous framework to test biological assumptions that are explicitly formulated and simulated (Estoup *et al.* 2004; Fagundes *et al.* 2007; Beaumont 2008).

This study aims at connecting the inferred demographic and gene flow history during the differentiation of *M. m. domesticus* and *M. m. musculus* with that of the fixation of reproductive incompatibilities as presently observed. For this purpose, we used 57 autosomal loci and ABC methods to compare four possible demographic scenarios and infer the parameters of the best models. On the basis of the history reconstructed in this way, we propose a plausible palaeogeographic model for the differentiation of the house mouse subspecies.

Materials and methods

Sampling and sequencing

We used ten mice of *M. m. domesticus* origin (from around the Mediterranean basin), nine mice of *M. m. musculus* origin (from Eastern Europe and Japan) and two out-groups (*Mus spretus* and *Mus famulus*). As the genome of *M. m. molossinus* mice (Japanese mice) is mainly of *musculus* origin (Yang *et al.* 2007, 2011), we included them in our *musculus* sample to increase the geographical coverage for this subspecies (Table 1). Most animals used were from wild-derived strains that were partially inbred, although the level of inbreeding varied a lot depending on strain history (number of inbreeding generations and number of founders). We PCR amplified and sequenced both strands of 61 noncoding fragments (Table S1, Supporting information). We chose these amplicons in genomic regions that displayed at least four SNPs in one kilobase in the PERLEGEN SNP database version 4 (from http://mouse.perlegen.com/mouse/, Frazer *et al.* 2007, data now accessible in dbSNP through the PERLEGEN handle). The results of this resequencing experiment describe polymorphism at over 8 million SNPs not only covering the genome for 16 house mouse strains representing mostly laboratory strains of essentially *M. m. domesticus* origin (Yang *et al.* 2007, 2011) but also including one wild-derived representative each of *M. m. domesticus*, *M. m. musculus* and *M. m. castaneus*. We generated *in silico* all possible amplicons of about 1 kb containing at least one of the SNPs reported in this experiment and lying in a noncoding region. We then selected our amplicons among those that contained at least four

Table 1 List of mice and their origin, with the number of haplotypes used in the ABC analyses

	Country	Locality	Strain	Specimen	ABC Chr nb
Mus famulus	India	Kotagiri	FAM	17 728	1
Mus musculus domesticus	Tunisia	Monastir	22MO	7692	1
M. m. domesticus	France	Montpellier	BFM/1	7735	1
M. m. domesticus	Egypt	Cairo	BNC	7659	1
M. m. domesticus	Algeria	Oran	BZO	17 733	1
M. m. domesticus	Spain	Barcelona	DEB	17 658	1
M. m. domesticus	France	Corsica	DFC	7564	2
M. m. domesticus	Italy	Orcetto	DJO	19 194	1
M. m. domesticus	Morocco	Azemmour	DMZ	17 736	1
M. m. domesticus	France	Toulouse	WLA	7739	1
M. m. domesticus	France	Boofzheim	Wild	7185	2
Mus musculus musculus	Austria	Illmitz	MAI	17 669	1
M. m. musculus	Bulgaria	General Toshevo	MBT	17 670	1
M. m. musculus	Czech R	Brno	MCZ	18 147	1
M. m. musculus	Hungary		MHT	18 149	1
M. m. musculus	Lithuania	Vilnius	Wild	10 298	2
M. m. musculus	Poland	Bialowieza	MPB	18 152	1
M. m. musculus	Poland	Warsaw	MPW	17 778	1
M. m. musculus	Japan	Mishima	MOL	7409	1
M. m. musculus	Denmark	Skive	MDS	19 191	1
Mus spretus	Spain	Granada	SEG	17 661	1

ABC, approximate Bayesian computation.

SNPs in this data set. We performed a BLAT search (Kent 2002) of the retained amplicons on build 37 of the house mouse genome to verify the absence of a recent duplication. We retained 57 loci (some of which constituted by several close amplicons whose sequences were concatenated for analysis) distant enough on the genetic map to be considered independent.

We treated the PCRs with the ExoSAP-IT enzyme (Amersham) and sequenced them separately with the two amplification primers using the ABI Prism dGTP BigDye Terminator Ready Reaction Kit (Perkin Elmer). Sequence reactions were purified by ethanol precipitation and run on an ABI Prism 3100 (Perkin Elmer) capillary sequencer. We downloaded from the Web sequence information of a *M. m. molossinus* strain (MS M/Ms strain, http://molossinus.lab.nig.ac.jp/msmdb/cgi-bin/adetail.cgi) and retained regions with sufficient sequence coverage in this experiment. We found no polymorphism between these sequences and those we derived from another *molossinus* strain, MOL (Table 1) so we used the MSM/Ms sequence only for loci where sequencing of MOL failed. When no out-group sequence could be successfully amplified, we downloaded the SNP data between the C57/Bl6J and the Spretus/EiJ strains (http://www.sanger.ac.uk/cgi-bin/modelorgs/mousegenomes/snps.pl) for genomic regions with sequencing coverage of at least 10× in this ongoing experiment.

Alignments and diversity analysis

Sequence chromatograms were interpreted using the CodonCode Aligner software (CodonCode Corporation). The sequences obtained were aligned with FSA version 1.14.2 (Bradley *et al.* 2009). FSA uses an explicit statistical model to produce multiple alignments, which are accompanied by estimates of accuracy for every column of the alignment. To avoid spurious polymorphisms because of misaligned sequences, we suppressed from our final alignments all positions with a score below the maximum value of 9 in the FSA analysis. The major effect of this operation was to remove regions with variable simple repeats, which are essentially difficult to align. Given the large amount of data, this not only prevented manual inspection of the sequences but also provided an objective and repeatable way to remove uncertainties in the alignments. Positions with indels were also automatically suppressed for all subsequent analyses. We reconstructed haplotypes using ISHAPE (Delaneau *et al.* 2007). Descriptive statistics such as θ_π, F_{ST} (sensu Hudson *et al.* 1992) and D_{XY} between subspecies (Nei 1987, p. 276) were calculated from the sequence alignments using the SITES program (Hey & Wakeley 1997, http://genfaculty.rutgers.edu/hey/software#SITES).

Model fitting using ABC

We used ABC to fit and compare different scenarios of subspecies differentiation. The models tested are schematically represented in Fig. 1. They all suppose that an ancestral population of size N_a split at some time T_s in the past to form two populations of sizes N_1 (*M. m. domesticus*) and N_2 (*M. m. musculus*). In model I (for Isolation), there is complete isolation after split, in model F (for gene Flow), there is continuous gene flow (although at potentially asymmetrical rates, $m12$ and $m21$), in model FI, gene flow ceases at time T_m, and in model IF, it only starts at time T_m. Note that the model we denoted F corresponds to the 'classical' IM model often referred to in the literature (Nielsen & Wakeley 2001).

We simulated 6 million data sets for each of the models compared, using a version of the MS program allowing different sample sizes across loci (Hudson 2002; Ross-Ibarra *et al.* 2008), and further modified as described below. For the rejection step, we performed cross-validations using simulated data sets to decide of an appropriate rejection scheme and found that retaining 4000 simulations gave good precision of parameter estimates even after removing the adjusted values falling outside the range of the prior. A logit transform could have avoided this range problem in addition to reducing heteroscedasticity, but such a transform has been shown to introduce biases in posterior estimation (Palero *et al.* 2009), so we only used a log transform to

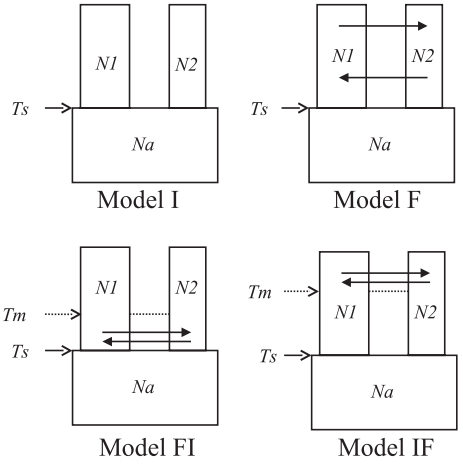

Fig. 1 Schematic representation and acronyms of the speciation models compared. An ancestral population of size N_a splits at time T_s into two daughter populations of sizes $N1$ (*domesticus*) and $N2$ (*musculus*).The daughter populations may exchange genes (indicated by arrows, at rates $m12$ and $m21$, not indicated). In model F, gene flow is continuous from T_s to present, in FI, it stops at time T_m, and in IF, it only starts at time T_m.

address the heteroscedasticity issue. Further, we used the nonlinear heteroscedastic regression method of Blum & François (2010) based on a learning neural network procedure. The neural network is exploited to account for the correlation between statistics and to reduce the distortion of the results because of high dimensionality (Blum & François 2010). Further, it gives a weight to accepted sets of parameter values that declines with the distance of their associated statistics to the observed statistics (Beaumont *et al.* 2002). This neural network approach does not make the assumption of a linear relationship between statistics and parameters and allows heteroscedasticity of residuals. An R script was kindly provided by Michael Blum and is now available inside the R package abc (http://cran.r-project.org/web/packages/abc/index.html). To account for the stochasticity in the learning procedure of the neural network, we averaged regression-adjusted values over 40 successive neural networks. Posterior density estimations were carried out using the locfit R function (Loader 1996). To compare the estimates of equivalent parameters inside a given model (e.g. population sizes of the two subspecies), we generated the posterior distribution of the difference of the two parameters and estimated the quantile of the distribution lying beyond zero.

For model comparisons, we loaded 2 million simulations for each model when comparing four models and 3 million when comparing two models. In all cases, 4000 simulations were retained after the rejection step, the regression step using the neural network was repeated 200 times and posterior probabilities averaged.

Prior model parameter distributions

For convenience, all parameter values were scaled by a factor $N_0 = 200\,000$. Prior distributions of N_a/N_0 were uniform over 0–3, those of N_1/N_0 and N_2/N_0 over 0–2, those of $T_s/4N_0$ and $T_m/4N_0$ over 0–4, and those of $4N_0m_{12}$ and $4N_0m_{21}$ over 0–3. For models using the two time parameters, we first drew the larger parameter value (T_s) in its uniform prior and used it to impose an upper limit to the prior of the second (T_m). Consequently, the priors for times of change of migration regime are not uniform. We did not attempt to estimate locus-specific mutation rate scalars by ABC but rather estimated mutation rates from the substitution rate with an out-group, for each locus separately, and the locus-specific values were used in the simulations (consequently, loci were not anonymous in the simulations). For each locus, we corrected the distance to the out-group by the average of nucleotide diversities estimated in the two subspecies. We used as out-group *M. famulus* whose divergence from *M. musculus* was dated at 3.9×10^6 BP (Suzuki *et al.* 2004; Chevret *et al.* 2005). For some loci, we failed to sequence *M. famulus* and then used *M. spretus* as the out-group, using a divergence time of 1.75×10^6 BP (Suzuki *et al.* 2004; Chevret *et al.* 2005). Mutation rates per generation were computed assuming one generation per year.

Summary statistics for ABC inference

We used fifteen different summary statistics but avoided some that could be too much biased by our sampling scheme (species-wide rather than population oriented), such as statistics depending on the allele frequency spectrum (e.g. Tajima's D, Przeworski 2002; Ptak & Przeworski 2002; Arunyawat *et al.* 2007). We used the means and variances of the following statistics: the number of sites with fixed derived alleles in one population yet polymorphic in the other (sx_jf_i), the number of sites with a derived allele fixed in one population and the ancestral allele fixed in the other (statistics sf_i and sf_j), and the number of sites with shared derived allele (ss). We also considered mean F_{ST}, mean D_A, mean D_{XY} (Nei 1987, p. 276) and mean nucleotide diversities (π).

Posterior predictive simulations

To check the validity of Bayesian inferences, posterior predictive simulations are widely used (Gelman *et al.* 1996; Thornton & Andolfatto 2006; Becquet & Przeworski 2009; Beaumont *et al.* 2010), the main idea being 'that if the model fits the data, replicated data from the model should look similar to the observed data' (Ingvarsson 2008). For each model retained, we ran 10 000 simulations for 57 independent loci using parameter values sampled from the posterior joint distribution. We computed the distribution of the means and variances of five statistics (π_{dom}, π_{mus}, D_{XY}, D_A, F_{ST}) over these simulations. These distributions were then used to obtain a two-sided posterior predictive P-value that is simply the probability of obtaining the observed statistic or a more extreme value under the estimated model. Finally, a correction for multiple tests was applied using false discovery rate (FDR) as implemented in the qvalue R package (Storey & Tibshirani 2003). We also performed goodness of fit for each locus independently, following the same principle.

Robustness and power of model comparison tests

To evaluate the ability of the method to discriminate between the two favoured models (F and IF, Fig. 1), we performed a receiver operating characteristic (ROC) analysis (Bazin *et al.* 2010). Simulations under the two

models (10 000 each, with parameter values drawn from their joint posterior distribution given the real data) were used as real data and submitted to a comparison of the two models by ABC. We counted the rate of true and false positives (cases where IF was favoured when the data were simulated under IF and F, respectively) for any threshold posterior probability value and plotted one against the other on a ROC graph, using the R ROCR package (Sing *et al.* 2005). Such a graph thus allows reading variations across posterior probability threshold of the power (rate of true positives) and robustness (rate of false positives) of the discrimination method used (see Bazin *et al.* 2010 for more details concerning the use of ABC model choice procedure as a classifier).

Accuracy of parameter estimates

To evaluate the accuracy of the estimation of a parameter of interest (the time of onset of gene flow, T_m, in the IF model), we generated simulated data sets under the IF model. We used all combinations of ten values of T_s (from $0.4N0$ to $4N0$, with $N_0 = 200\,000$) and nine values of the ratio T_m/T_s (from 0.1 to 0.9). For each of these 90 combinations, we produced 100 simulated data sets (drawing all other parameter values in the posterior distributions given the real data), from which we measured the proportion of times the true value of T_m was included in the 90% credibility interval, as well as the average unsigned deviation of the modal estimates of T_m to the true value (using the ABC inference procedure under model IF for each simulated data set).

Data used for ABC inferences

We removed variable CpGs from the sequence alignments because they can be a source of homoplasy mimicking recombination and violating the infinite site model used in the simulations, and we then performed the four gamete test (Hudson & Kaplan 1985) to keep only blocks free of apparent recombination (IMGC software, Woerner *et al.* 2007).

For samples derived from laboratory maintained stocks, we only retained one randomly chosen haplotype per sample to avoid correlations owing to inbreeding. Both haplotypes were kept in the case of wild-caught animals and laboratory-raised animals that had a comparable heterozygosity on the basis of our sequence results (see Table 1 for the number of haplotypes retained per sample).

Furthermore, we removed from the data set some loci for samples MDS and MOL, for the following reasons. The MDS strain was derived from a locality in Northern Jutland, rather close to the natural hybrid zone with *M. m. domesticus*, and could thus possess alleles of *domesticus* origin at some loci, because of recent gene flow across the hybrid zone. In fact, on the basis of 67 independent SNPs across the genome, that are diagnostic between the subspecies in this hybrid zone, the MDS sample was estimated to be 89% *musculus* (data not shown). As we are here interested in the long-term history of differentiation of these subspecies, rather than the consequences of introgression across this very recent hybrid zone, we wanted to avoid loci for which MDS possessed an allele clustering with *domesticus* alleles, while our other *musculus* samples far from the hybrid zone did not. We found only one such locus among the 57 studied (Loc32), and it was removed from the data set for the MDS sample. We also used the MOL strain, derived from Japan. This strain is related to the MSM/Ms laboratory strain that was shown to be 98.6% *musculus* in a genome-wide survey using a high-density SNP array (Yang *et al.* 2011). A small proportion of its genome however appears derived from *M. m. castaneus*. Therefore, for the MOL sample, we excluded from the data set two loci (Loc02 and Loc50) for which it did not cluster with other *musculus* samples, but rather stood alone externally to the *domesticus* and *musculus* branches of the tree (not shown).

Assessing the consequences of locus choice

As mentioned above, the loci studied were chosen as showing at least four SNPs (in 1-kb fragments) in the Perlegen resequencing study (Frazer *et al.* 2007), which involved representatives of the three subspecies of house mouse. This could have biased the choice in favour of genomic regions with high mutation rates, but this is not an issue since locus-specific mutation rates are considered in our simulations. However, this could have biased the choice in favour of loci showing deep coalescence times between these representatives of the subspecies. To evaluate this, we retrieved the Sanger Center genome sequencing data of the wild-derived laboratory strains WSB/EiJ (*M. m. domesticus*), PWK/PhJ (*M. m. musculus*) and Spretus/EiJ (*M. spretus*) (http://www.sanger.ac.uk/cgi-bin/modelorgs/mousegenomes/snps.pl). Note that in the Perlegen experiment, wild-derived *M. m. domesticus* was also represented by WSB/EiJ and *M. m. musculus* by PWD/PhJ, the origin of which is close to that of PWK/PhJ (Gregorová & Forejt 2000), while *M. spretus* was not represented. We generated the distribution of the ratio of the distance between WSB/EiJ and PWK/PhJ, divided by the average of their distances to Spretus/EiJ to account for variations of mutation rates across genomic regions (distances were just the number of differing SNPs based on the Sanger data), for all 1-kb windows of noncoding

DNA across all autosomes. We then compared it to the distribution only taking into account windows that had four SNPs or more in the Perlegen data set (among which our study loci were chosen). We found the latter distribution to be skewed towards higher values as compared to the autosome-wide distribution, suggesting that on average, segments with more than four SNPs in the Perlegen experiment had older coalescence times of the *M. m. domesticus* and *M. m. musculus* strains, taking the divergence with *M. spretus* as a reference. Note that we excluded from the above distributions a number of windows: those having no SNPs in the Sanger data (and thus an undefined ratio), those with a ratio equal to 2 (indicating a null branch leading to Spretus/EiJ) and those lying in genomic regions where either WSB/EiJ or PWK/PhJ appeared to be alien origin (from another subspecies), according to the high-density SNP study of Yang et al. (2011, http://msub.csbio.unc.edu/PhylogenyTool.html).

Accounting for locus choice bias in ABC simulations

We attempted to reproduce in our simulations the bias induced by the locus choice process mentioned above in the following way. Our objective was to mimic the skewed distribution of the ratio of the distance between WSB/EiJ (*M. m. domesticus*) and PWK/PhJ (*M. m. musculus*) over their distance to Spretus/EiJ (*M. spretus*), which we observed among 1-kb windows with four SNPs or more in the Perlegen data set, windows among which the study loci were chosen. For each locus, we multiplied this distribution by the average distance between *M. spretus* and the house mouse (measured from our sequences) to obtain the distribution of the distance between *domesticus* and *musculus*, which we then aimed at obeying in our simulations of the given locus. In each simulation, we checked whether among all possible simulated pairs of haplotypes including one *domesticus* and one *musculus*, at least one pair had a distance falling in the range of a randomly chosen bin of this distribution, which was fractionated into 10 bins. If such was the case, the simulation was kept for the ABC analysis, after removing one of the pairs of haplotypes obeying the criterion, because the simulations contained one more *domesticus* and one more *musculus* than the real data set. If not, the simulation was retried (with the same set of model parameters values), until the criterion was obeyed, or until 1000 unsuccessful tries had been performed. Although for a single locus, the criterion was obeyed most of the time within 1000 tries, all loci had to obey the criterion for a simulation set to be accepted with a given set of parameters drawn from the priors. This resulted in a combined rate of failure too high for the process to be feasible in a practical calculation time. However, dropping simulation sets in failure would have induced a risk of unintentionally modifying the prior distributions and thus biasing the ABC analyses. We therefore adopted a compromise consisting in accepting simulations with no more than three loci failing to obey the bias-mimicking criterion, and we checked *a posteriori* the shape of the resulting prior distributions. We performed the simulations described here after modifying the code of the MS program (Hudson 2002; Ross-Ibarra *et al.* 2008).

Results

Divergence and diversity patterns

We sequenced an average of 870 bp for 61 noncoding autosomal fragments in a sample of the two subspecies (10 *M. m. domesticus* and nine *M. m. musculus*, Table 1) as well as two out-groups, *M. spretus* and *M. famulus*. This represents a total of about 48 kb of aligned sequences. All fragments appeared unique in a BLAT search against the complete mouse genome. Some fragments predicted to be very close to each other in the genome, and which did not present any sign of recombination, were concatenated (loc05-06 and loc19-22), so the final data set consisted of 57 independent loci. Nucleotide diversity is higher in *musculus* (0.0023 and 0.0027 for median of π *and* θw, respectively) than in *domesticus* (0.0016 and 0.0018, see Table S2, Supporting information for details). These higher values in *musculus* appeared to result from sharing of haplotypes with *domesticus*, a possible consequence of gene flow. For this reason, we were unable to perform tests of selection based on polymorphism and divergence, such as the HKA test, which would be meaningless in the presence of gene flow. On the basis of our data, the differentiation between subspecies was high (median $F_{ST} = 0.55$), but there was a high variance with values ranging from 0 to 1 (Table S2, Supporting information).

Simulation of the effects of locus choice

We have seen (Materials and methods) that the protocol used to choose the loci studied here could have introduced a bias towards loci with a high divergence between the representatives of *M. m. domesticus* and *M. m. musculus* in the data set used to select them (the Perlegen resequencing data set) as compared to their divergence to *M. spretus*. We therefore attempted to reproduce such a bias in our simulations (see Materials and methods for details). However, due to calculation time limitations, we had to reject many simulation sets, in which some of the loci failed to fit our bias criterion, despite 1000 attempts for each locus (in fact, we rejected

those for which more than three loci failed, see Materials and methods). This raises two potential issues. First, we wanted to check that despite this limitation, we had been able to correctly respect the desired distribution of this divergence of the *domesticus–musculus* pair. Figure S1 (Supporting information) shows that such is the case, as for all loci, there was good concordance between the aimed and realized distributions. Second, we were concerned that the rejection of many simulations failing to respect this criterion might have affected our priors on the distribution of the parameters of the models. This effect can be appreciated on Fig. S2 (Supporting information) in the case of 10 000 simulations under model F. There appears to be no effect either on the migration parameters ($m12$ and $m21$) or on the population size parameters for *domesticus* (N_1) and *musculus* (N_2), whose prior distributions in the retained simulation sets appears uniform. However, an effect can be seen on divergence time (T_s) and size of the ancestral population (N_a), with a deficit in low values as compared to the initial uniform distribution. The effect is however slight, and most of the range of parameter values of the prior remain reasonably covered.

Comparison of models

We used the ABC framework to compare several explicit scenarios of speciation with different patterns of gene flow since the initial divergence of *M. m. domesticus* and *M. m. musculus* (Fig. 1, see Materials and methods for the nomenclature of models). When comparing the four models, we found very low posterior probabilities for two of them: the I model with complete isolation since divergence (PP = 0, Table 2) and the FI model in which gene flow occurs immediately after divergence and then stops (PP = 0.008). The model with recent gene flow following initial isolation had a higher probability (model IF, PP = 0.697) than that with constant gene flow since divergence (model F, PP = 0.295).

Posterior predictive simulations and robustness test

To check whether the results of model comparisons were affected by either bad fit of the models to the data or some strong outlier loci, we performed posterior predictive simulations (Gelman *et al.* 1996; Ingvarsson 2008; Beaumont *et al.* 2010) for the two most probable models, F and IF, using the means and variances of five statistics (π_{dom}, π_{mus}, D_{XY}, D_A, F_{ST}). As seen on Fig. S3 (Supporting information), there is a generally good agreement between observed and simulated data sets. Despite this overall positive result, it is worth raising the question of the quality of the model to account for data at individual loci, since strong outliers could affect model fitting, and especially parameter estimations. We could see two major sources of potential outliers. First, we did not estimate mutation rate scalars within the ABC framework but simply using a crude estimate of the divergence with the out-groups for each locus, but such estimations could be poor for some loci. Second, some loci could be influenced by selection affecting their coalescent and/or migration patterns. We therefore performed posterior predictive simulations for each locus separately and for both models to detect potential outlier loci *a posteriori*. We used the total number of segregating sites in the whole sample (S) and F_{ST} as potential indicators of outlier loci due to badly estimated mutation rate or divergent selection, respectively (Beaumont 2005). Nine loci were found as potential outliers for either of the models, most often for both (Table S3, Supporting information). Note that most outliers show possible deviation for S only, suggesting that poor mutation rate estimation is a more likely source of bad fit than selection on this set of loci. However, only one locus (loc45) remained significant (for the S statistics) after applying a correction for multiple tests (Storey & Tibshirani 2003). We therefore omitted this locus to redo the model comparisons and to estimate model parameters on a 56-locus data set. The comparison between the F and IF models produced results similar to those on all loci (PP $\approx$ 0.7 for the IF model, Table 2). On this basis, it could be said that the IF model is more probable than the F model given our data, but to assess the power and robustness of the approach, we conducted a ROC analysis, the results of which are presented in Fig. S4 (Supporting information). Although the power appears moderate since at the posterior probability threshold used here (0.7), the rate of true positives is only about 40%, it appears robust since we find only 5.2% false positives. We can thus be reasonably confident in the result of our model comparison and retain the IF model as more probable.

Table 2 Posterior probabilities of compared models, with uncertainty (median absolute deviation) calculated over 200 independent neural networks

	Model			
Data set	I	F	FI	IF
57 loci	0 (0)	0.295 (0.035)	0.008 (0.002)	0.697 (0.0034)
56 loci		0.302 (0.107)		0.698 (0.107)

Parameter estimations

In line with the observation of Becquet & Przeworski (2009) that few unusual loci do not influence drastically demographic inferences, the results using 56 loci after

removing one outlier were very close to those with all 57 loci, so we will only use the former here, and present the results of the IF model, which we have shown to be the best among those tested (Table 3 and Fig. 2). The effective population size of *M. m. domesticus* is estimated to be larger than that of *M. m. musculus* (around 96 000 and 50 000, respectively, for modal estimates), and this difference is significant ($P = 0.044$ from the

Table 3 Parameter estimates and 90% credibility intervals for the 56-locus data set under the IF model, with times in generations

Parameter	Median	Mode	HPD90low	HPD90high
*N*1	95 616	96 010	66 190	128 234
*N*2	47 804	49 656	24 529	81 212
N_a	314 321	280 848	46 104	552 740
*m*12	2.71E-07	2.75E-07	0.90E-07	6.96E-07
2*N*1*m*12	0.064	0.050	0.018	0.122
*m*21	11.47E-07	9.81E-07	5.36E-07	23.62E-07
2*N*2*m*21	0.120	0.107	0.050	0.205
T_m	359 526	202 888	67 530	1 197 514
T_s	1 562 146	1 283 108	516 003	2 705 606
$(T_s - T_m)/T_s$	0.740	0.744	0.398	0.930

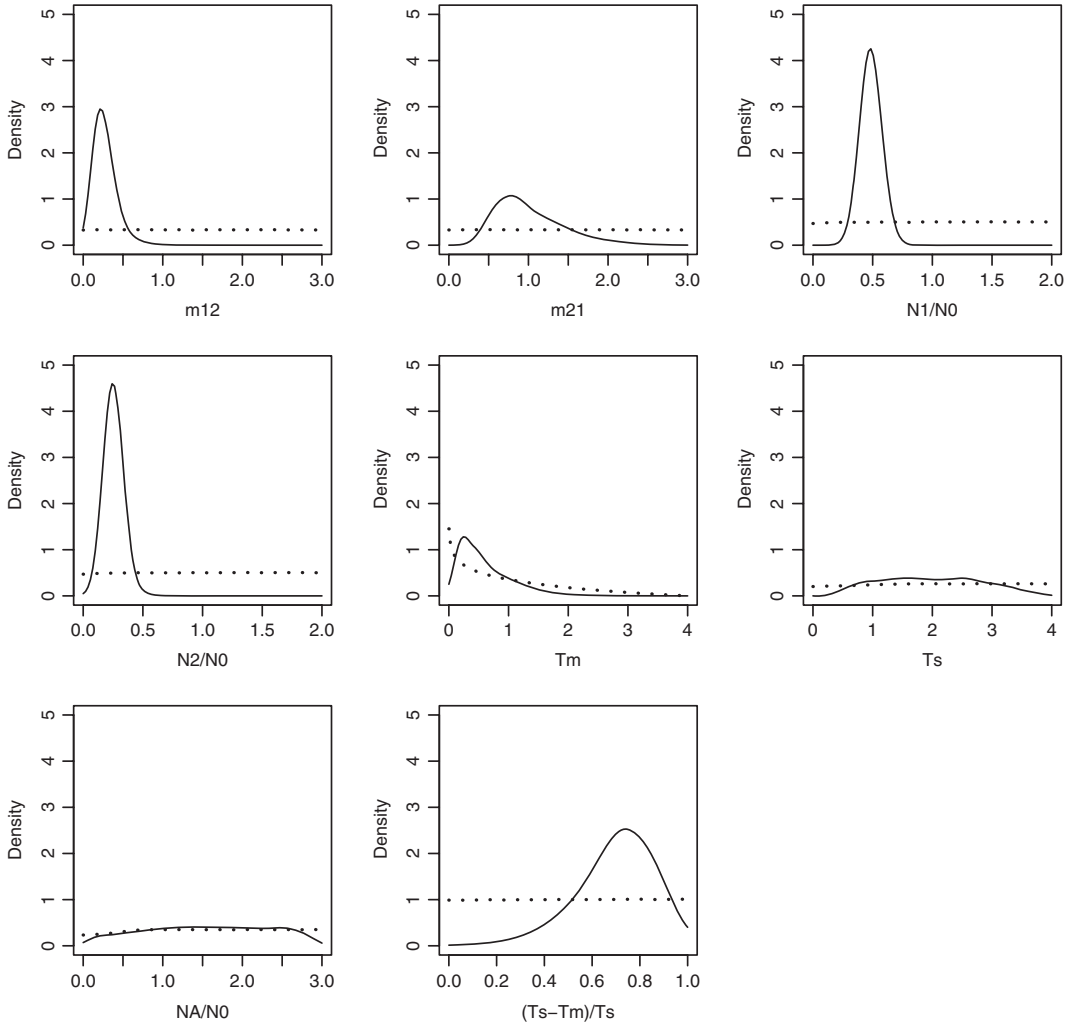

Fig. 2 Prior (dotted curves) and posterior (continuous curves) distributions of the parameters of the IF model, as well as of the proportion of time in isolation $(T_s - T_m)/T_s$.

posterior distribution of the difference). We also estimate significantly asymmetric gene flow ($P = 0.024$) with an approximately three times higher rate into *musculus*. The onset of migration was estimated with a mode around 200 000 generations ago but with a rather wide credibility interval. The analysis of the quality of this estimation we conducted (Fig. S5, Supporting information) indicates a satisfactory result in the range of T_s inferred from the real data (see discussion for such estimates), as in this range, real values were included in the HPD90 interval more than 90% of the time, and the precision appeared reasonable. The posterior distributions of split time and ancestral population size were too flat to allow meaningful inferences. However, some information appeared to exist to infer the distribution of the fraction of divergence time spent in isolation $(T_s - T_m)/T_s$, with a modal estimate of 0.744.

Discussion

Before we comment on the biological implications of our results, we must examine methodological issues related to the complex inference methods used here.

Methodological issues of the ABC approach

Coalescent genealogy samplers are known to be efficient in reconstructing population history when many loci are analysed, even if on a limited number of samples per population (Hudson & Coyne 2002; Donnelly & Tavare 2003; Felsenstein 2006; Kuhner 2009). We used a relatively large number of loci as compared to previous studies (56 after removing a potential outlier) and will comment later on the quality of the sampling. The approach we used makes the assumptions that populations are panmictic, that the loci are independent and evolve neutrally, that the mutation model is adequate and that no unsampled population influences the genetic make-up of the sampled populations. As these methods of analysis are relatively new, little is yet known on their performance in the face of violations of these assumptions, but a recent study has shown the likelihood-based IMa approach to perform satisfactorily in the face of moderate violations of most of these assumptions (Strasburg & Rieseberg 2010). Taking all these aspects in turn, panmixia is unlikely to hold given our sampling scheme, but we avoided statistics that would be too much biased by population structure. The independence of loci is obviously not an issue in our case. Divergent selection can also contribute to bias split time estimate, yet multiple sweeps are needed (no significant biases were observed with one selected locus in Becquet & Przeworski 2009), and they must be recent and strong (Strasburg & Rieseberg 2010). We performed a goodness of fit for each locus and removed an outlier, but this did not affect much the estimates of the parameters of the models. Concerning the mutation model, we were careful in eliminating regions of uncertain alignments, and variable CpGs, thus minimizing the causes of deviation from the infinite site model. One serious source of bias in our results could come from a role played by the third house mouse subspecies not sampled here, *M. m. castaneus,* since Geraldes *et al.* (2008) found evidence of gene flow between this subspecies and at least *M. m. musculus*. Gene flow from an unsampled population could lead to a bimodal posterior distribution of the divergence time if this unsampled population is very divergent from the studied populations (Strasburg & Rieseberg 2010), but this is unlikely in the case of house mice, which appear to have diverged from their common ancestor in a short time. Finally, a potential bias could have been introduced in our analyses by poor estimates of mutation rates. We chose not to estimate mutation rate scalars per locus in the ABC framework by fear of overparameterization, and we estimated rates by direct comparison with the out-groups. Overall, we found a geometric mean of $\mu = 3.4 \times 10^{-9}$ per bp for all loci, which is consistent with values reported in the literature (Keightley & Eyre-Walker 2000; Geraldes *et al.* 2008; Liu *et al.* 2008), and we attempted to point and eliminate loci with potentially poor estimates by performing a goodness of fit on all loci.

The choice of loci and its consequences

In fact, the most serious concern we had about the validity of our approach resulted from our choice of loci. Although we found a way to simulate the resulting bias, apparently relatively faithfully, it remains true that our inferences do not rely on a random subset of the genome. It must therefore be kept in mind that the history we can infer here may be partially that of the study loci rather than that of a representative random sample of the genome.

We note though that it is unclear what strategy should be used to reach such an ideal sample of loci, especially when the number of loci is relatively limited (which despite our efforts is still the case in our study), and because different genomic regions are differently affected by evolutionary forces such as recombination and selection. To date, studies similar to the present one rely on a sample of loci often much smaller than that used here, and the criteria to choose loci are generally not described. The tendency to avoid loci with little variation must have cryptically imprinted the make-up of most available sequence polymorphism data sets. It could however be that choosing loci the way we did

allowed us to reveal aspects that would have been invisible from a 'random' sample. Some authors have relied on the peculiar characteristics of some genomic regions and taken advantage of them to make otherwise difficult inferences (for instance, Huff *et al.* (2010) used regions flanking polymorphic transposable elements, which have deep coalescence times, to infer ancestral human population size).

An undesirable consequence of our locus choice procedure would be if it had enriched in loci strongly influenced by selection, but this is unlikely since the simple choice criterion we applied is far from extreme. In fact, 20% of noncoding 1-kb windows genome-wide obeyed the choice criterion, and such a high proportion of the genome is unlikely to be globally submitted to strong selection pressures.

Isolation in allopatry and the accumulation of genetic incompatibilities

A previous study based on a smaller number of loci but a larger sample had rejected a model of differentiation in complete isolation (Geraldes *et al.* 2008), and we here confirm this conclusion, but we show that the differentiation process likely started with an initial phase of isolation. In the comparison between our F and IF models (Table 2), the latter was more probable (a conclusion that resisted our analysis of robustness), which is notable since the rejected model had lower complexity, and thus benefited from a better exploration of the parameter space since equal numbers of simulations entered the comparison procedure for each model.

The existence of an initial phase of isolation could be critical to account for the accumulation of genetic incompatibilities in a context of allopatric divergence. Several observations point to the existence of multiple incompatibilities between these subspecies. The analysis of multilocus patterns of allele frequency change across the hybrid zone has led to rough estimates of the number of incompatibilities, under a simple model of underdominance though, not accounting for epistasis (estimates ranging from 43 to 120 loci, Raufaste *et al.* 2005; Macholán *et al.* 2007). Studies of hybrid male sterility in laboratory crosses have shown the implication of multiple loci with epistatic interactions (Storchova *et al.* 2004; Britton-Davidian *et al.* 2005; Vyskocilova *et al.* 2005; Good *et al.* 2008; White *et al.* 2011). Orr (1995) and Orr & Turelli (2001) noted that under a simple model of accumulation of new mutations with epistatic interactions, the mean number of Dobzhansky–Muller incompatibilities (DMI) between two allopatric species is expected to grow as the square of their divergence time. An estimate of the fraction of time the subspecies remained in isolation since their divergence could thus be used to approximately quantify the proportion of incompatibilities that had accumulated before secondary contact. Taking the square of the modal estimate of $(T_s - T_m)/T_s$ (0.744, Table 3), we find that 55% (90 HPD 16–86%) of the incompatibilities could already have accumulated when secondary contact occurred. We also used the simple calculation of net nucleotide divergence to estimate T_s and found a geometric mean of 0.79 MY across our 56 loci (using a calibration time of 3.9 MY for the divergence with *M. famulus*, Suzuki *et al.* 2004; Chevret *et al.* 2005). Using the modal estimate of T_m at 0.20 MY (Table 3), this gives an estimated 56% accumulation of incompatibilities, in good agreement with the ABC estimate. Note though that our choice of loci could have led to an overestimation of divergence time. Previous estimates we could compare it to were based on different data sets, ranging from DNA–DNA hybridization to sequence data of a few nuclear or mitochondrial fragments and to the use of different and uncertain absolute calibration points (no more uncertain than those we used here though, She *et al.* 1990; Suzuki *et al.* 2004; Chevret *et al.* 2005). It is therefore difficult to synthesize all these inferences except by noticing that none of the intervals proposed spanned values higher than about 1 MY. There is thus no obvious discrepancy with the value derived from our 56 loci reported above. None of these estimates account for migration after divergence (since our attempt and that of Geraldes *et al.* 2008 failed), and thus, all may be underestimates (including ours based on net divergence). If such was the case, it would however reinforce the conclusion that a significant fraction of the evolution of these subspecies occurred in allopatry before they started exchanging genes.

Secondary gene flow and reinforcement

The data analysed here appeared to contain some information on the timing of migration, and modal estimates point to an onset around 200 000 BP, assuming one generation a year. Even the lower bound (*c.* 67 000 BP) is much more ancient than the formation of the European hybrid zone, which is at most a few thousand years old. Despite the great uncertainty on this time of first secondary contact, the fact that it appears ancient as compared to the formation of the European hybrid zone is interesting in relation to an aspect of speciation other than the accumulation of post-zygotic incompatibilities discussed up to now. The mate recognition systems of *M. m. musculus* and *M. m. domesticus* are known to be somewhat divergent, and based on odorant cues, particularly in urine or saliva (Laukaitis *et al.* 1997; Talley *et al.* 2001; Smadja & Ganem 2002, 2005, 2008; Bimova *et al.* 2005). Interestingly, there is growing evidence that

reinforcement is taking place in the European hybrid zone concerning the ability to discriminate and choose conspecifics for reproduction, based on urinary odorant cues (Smadja & Ganem 2005, 2008; Bímová *et al.* 2011). The ancient and potentially prolonged contacts that we infer here might have participated in the initiation of the divergence of the mate recognition systems through a similar process, especially if many of the post-zygotic incompatibilities were already in action, as we inferred above. Further, it is interesting to consider that some characteristics of reinforcement now seen in the European hybrid zone might have an origin much older than the very recent formation of the zone and might segregate in natural populations since the times of an ancient and prolonged (or repeated) interaction.

Amount of gene flow

The episode of gene flow that we inferred was relatively moderate and did not correspond to extensive admixture given our estimates of modal $2Nm$ (0.05 in *domesticus* and 0.107 in *musculus*), even considering the upper bounds of the credibility intervals (0.122 and 0.205). Our estimates are comparable to those derived by Geraldes *et al.* (2008) based on eight loci and a larger sample. However, this consistent result may appear contradictory with another study. Using a genome-wide SNP data set on a small sample of both subspecies, Pool & Nielsen (2009) inferred very high and extremely recent gene flow between them (in historical times) and even suggested that it was intense enough to eventually lead to admixture if persisting on the long term. The approach they took is completely different from ours and has much more power to detect recent gene flow (despite small sample size, but because it involves a genome scan) and to estimate its intensity because it does not rely on estimates of allele frequencies, necessarily poorly estimated on samples of the size used in either study, but rather on the distribution of the sizes of introgression chromosome blocks. Very recent gene flow certainly exists, owing to the exponential increase in human activities and the close association of mice with humans, and has been sporadically invoked in the literature, when a general phylogeographic structure appeared to characterize subspecies apart from rare clear outliers (Yonekawa *et al.* 1988; Terashima *et al.* 2006; Bonhomme *et al.* 2007). It is difficult to draw quantitative conclusions on the basis of such anecdotal

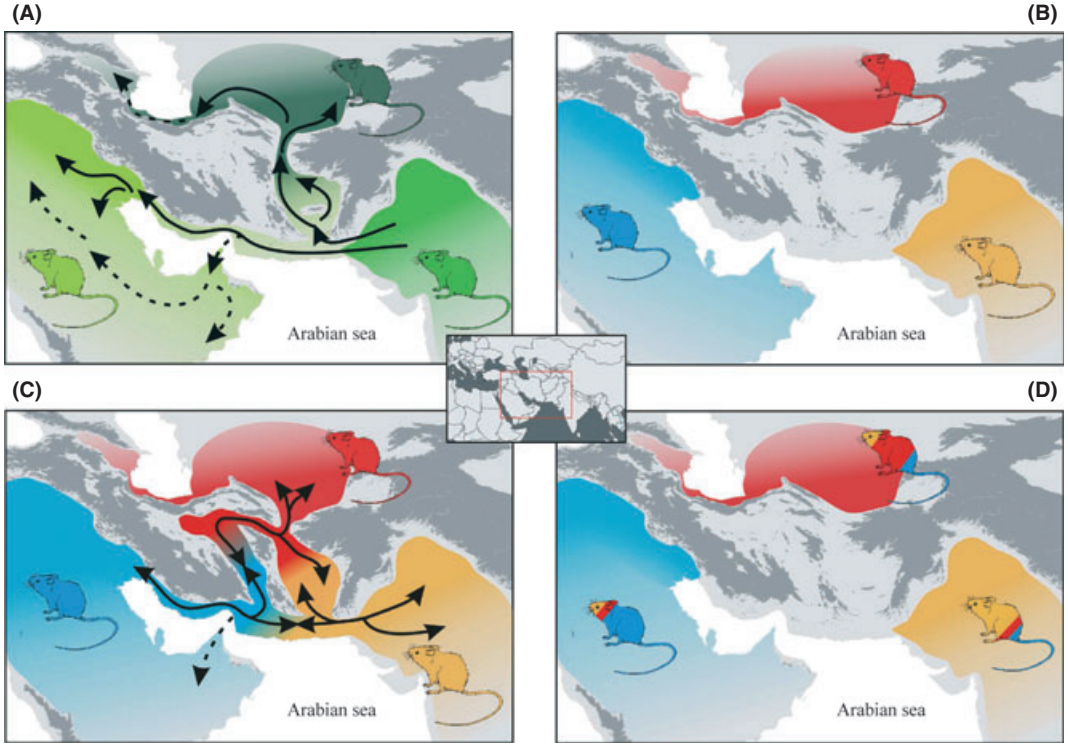

Fig. 3 Hypothetical history of differentiation and gene flow in house mice. (A) initial colonization of the Middle East from the Indian cradle. (B) Phase of isolation and divergence, blue = *Mus musculus domesticus*, red = *Mus musculus musculus*, yellow = *Mus musculus castaneus*. (C) Expansion, secondary contact and gene flow during interglacials. (D) Isolation during glacial periods. Phases C and D could have occurred several times, the first episode starting around 200 000 BP, and the present interglacial situation corresponding to type C. White areas indicate the sea, light grey elevations up to 1500 m. and dark grey higher elevations. Distribution areas are coloured according to taxonomy and mice according to predicted genetic make-up.

observations, but the inference of Pool & Nielsen (2009) is intriguing when compared to ours or that of Geraldes *et al.* (2008). It could be that some of this recent gene flow is driven by selection in relation to the change of ecological niche of house mice when they became commensal with human during the Neolithic or more recently as the commensal niche changed with rapidly changing human activities in the industrial era. In fact, it has been reported that *M. m. domesticus* acquired resistance to anticoagulants used as rodenticides through very recent selective introgression of a chromosome fragment from *M. spretus,* which is naturally resistant to these poisons (Song *et al.* 2011). Because it was genome-wide, the approach taken by Pool & Nielsen (2009) would have been quite efficient at detecting similar cases between the two subspecies. Recent adaptive introgression is expected to drive large alien genomic blocks to high frequencies, making them detectable even on a limited sample, which accounts well for the findings of these authors. Our approach based on a few loci scattered across the genome is of course totally inefficient at capturing the effects of such a phenomenon that must concern a small fraction of the genome.

Biogeographic scenario

The demographic inferences we performed here only make full sense if they can be reconciled with geography. Phases of differentiation without gene flow must have corresponded to distinct distribution areas, while phases of gene flow necessarily implied at least adjacent, if not overlapping, distributions. We here attempt to reconstruct a plausible palaeogeographic scenario compatible with such constraints and summarize it in Fig. 3. Following previous authors (reviewed in Boursot *et al.* 1993), we will consider an origin in the northern Indian subcontinent most likely. Given their present distributions, *M. m. domesticus* and *M. m. musculus* must thus have at some stage experienced a westward colonization from there, through Iran and into new territories. White *et al.* (2009), using a dense SNP data set on one representative of each of the three subspecies, found that genomic regions for which the most probable phylogeny had *domesticus* as an out-group to the two other subspecies were the most abundant across the genome. This may suggest that *domesticus* diverged first from the ancestral population. However, our scenario does not critically depend on the order in which *domesticus* and *musculus* migrated to the west from the ancestral population. Given the geography of Iran, a likely colonization route for *domesticus* goes along the coast of the Persian Gulf into Mesopotamia. Crossing the Strait of Hormuz into the Arabian Peninsula must have also been possible when sea level was low. Prager *et al.* (1998) have described an anciently derived mtDNA lineage, sister of the *domesticus* lineage (and baptized *gentilulus*) and only found in the Arabian Peninsula (also later found to occur in Madagascar, Duplantier *et al.* 2002). Although Prager *et al.* (1998) considered the existence of this old lineage as evidence that the Arabic Peninsula may be at the origin of the differentiation of the house mouse, we believe that this is rather the sign of an ancient isolation of this population from the rest of the species. The colonization route taken by the proto-*musculus* population can be logically thought to have brought it across the Iranian plateaus to the north of Iran, along the coast of the Caspian Sea, and possibly all the way to the Caucasus because present-day populations of these regions clearly belong to this subspecies and potentially represent part of its ancestral range (Prager *et al.* 1998; Milishnikov *et al.* 2004; Darvish *et al.* 2006). The distribution ranges of these proto-subspecies must have been sufficiently disjoint to allow the long phase of isolation that we inferred. Nevertheless, they later had the possibility to exchange genes, presumably during Pleistocene interglacial periods (see Lisiecki & Raymo 2005 for Pleistocene palaeoclimatic data) such as the two to three such events documented by global palaeoclimatic data during the last 250 000 BP (see for example Fig. 2b,c in Augustin *et al.* 2004). However, the geography of Iran and its mountain ranges point to few possible places for such exchanges if our inferences of the ancestral ranges of the subspecies are correct. Although some contacts between *domesticus* and *musculus* in the Caucasus cannot be ruled out since they have been shown to presently occur (Orth *et al.* 1996; Milishnikov *et al.* 2004), a probable place for such exchanges is Iran. It is however difficult to conceive that a contact in such a place did not also involve the third subspecies, *M. m. castaneus*. The present-day genetic composition of Iranian mice is yet poorly described. Populations west of the Zagros Mountains clearly belong to *domesticus* (Baines & Harr 2007; Geraldes *et al.* 2008; Rajabi-Maham *et al.* 2008) and those from the extreme north-east resemble *musculus*, but there appears to be a transition between *musculus* in the extreme north-east and *castaneus*-like populations further south (Darvish *et al.* 2006). Therefore, if admixture between *domesticus* and *musculus* occurred in Iran, the process most likely also implied *castaneus*. The models we built here for ABC inference may thus be incomplete in failing to take into account such a possibility. Future studies should clearly include the three subspecies, and Iran appears a hot spot of house mouse diversity, perhaps holding important clues of our understanding of the evolution of reproductive isolation between the three house mouse subspecies.

Acknowledgements

We are highly indebted to members of the GEPV laboratory in Lille (Xavier Vekemans, Vincent Castric and Camille Roux) for invaluable help in the implementation of the ABC method and for sharing experience and unpublished software to calculate summary statistics. Michael Blum provided much advice and unpublished code for the implementation of the neural network regression method. We also greatly benefited from the meetings organized by GDR1928 'Population Genomics'. Pierre Caminade provided efficient help in sample preparation and sequencing experiments. Laurence Meslin was of great help to produce Fig. 3. Comments from the referees greatly contributed to improve the analyses. LD was supported by a PhD fellowship from the French Ministère de l'Enseignement Supérieur et de la Recherche. The project was funded by grants ANR-05-BLAN-0088 and ANR-06-BDIV-003-06 to PB. Sequencing was partly done on the platform of IFR119 'Montpellier Environnement Biodiversité'. This is contribution ISEM-2011-121.

References

Arunyawat U, Stephan W, Stadler T (2007) Using multilocus sequence data to assess population structure, natural selection, and linkage disequilibrium in wild tomatoes. *Molecular Biology and Evolution*, **24**, 2310–2322.

Augustin L, Barbante C, Barnes PRF *et al.* (2004) Eight glacial cycles from an Antarctic ice core. *Nature*, **429**, 623–628.

Baines JF, Harr B (2007) Reduced X-linked diversity in derived populations of house mice. *Genetics*, **175**, 1911–1921.

Barton NH, de Cara MAR (2009) The evolution of strong reproductive isolation. *Evolution*, **63**, 1171–1190.

Baudat F, Buard J, Grey C *et al.* (2010) PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science*, **327**, 836–840.

Bazin E, Dawson KJ, Beaumont M (2010) Likelihood-free inference of population structure and local adaptation in a Bayesian hierarchical model. *Genetics*, **185**, 587–602.

Beaumont MA (2005) Adaptation and speciation: what can Fst tell us? *Trends in Ecology & Evolution*, **20**, 435–440.

Beaumont MA (2008) Joint determination of topology, divergence time and immigration in population trees. In: *Simulation, Genetics and Human Prehistory* (eds Matsumura S, Forster P, Renfrew C), pp. 134–154. McDonald Institute Monographs: Cambridge McDonald Institute for Archeological Research, Cambridge, UK.

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

Beaumont MA, Nielsen R, Robert CP *et al.* (2010) In defence of model-based inference in phylogeography. *Molecular Ecology*, **19**, 436–446.

Becquet C, Przeworski M (2007) A new approach to estimate parameters of speciation models with application to apes. *Genome Research*, **17**, 1505–1519.

Becquet C, Przeworski M (2009) Learning about modes of speciation by computational approaches. *Evolution*, **63**, 2547–2562.

Bimova B, Karn RC, Pialek J (2005) The role of salivary androgen-binding protein in reproductive isolation between two subspecies of house mouse: *Mus musculus* musculus and *Mus musculus* domesticus. *Biological Journal of the Linnean Society*, **84**, 349–361.

Bímová BV, Macholán M, Baird SJE *et al.* (2011) Reinforcement selection acting on the European house mouse hybrid zone. *Molecular Ecology*, **20**, 2403–2424.

Blum MGB, François O (2010) Non-linear regression models for approximate Bayesian computation. *Statistics and Computing*, **20**, 63–73.

Bonhomme F, Rivals E, Orth A *et al.* (2007) Species-wide distribution of highly polymorphic minisatellite markers suggests past and present genetic exchanges among house mouse subspecies. *Genome Biology*, **8**, 1034–1043.

Boursot P, Auffray J-C, Britton-Davidian J, Bonhomme F (1993) The evolution of the house mice. *Annual Review of Ecology and Systematics*, **24**, 119–152.

Boursot P, Din W, Anand R *et al.* (1996) Origin and radiation of the house mouse: mitochondrial DNA phylogeny. *Journal of Evolutionary Biology*, **9**, 391–415.

Bradley RK, Roberts A, Smoot M *et al.* (2009) Fast statistical alignment. *PLoS Computational Biology*, **5**, e1000392.

Britton-Davidian J, Fel-Clair F, Lopez J, Alibert P, Boursot P (2005) Postzygotic isolation between the two European subspecies of the house mouse: estimates from fertility patterns in wild and laboratory-bred hybrids. *Biological Journal of the Linnean Society*, **84**, 379–393.

Butlin RK, Galindo J, Grahame JW (2008) Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363**, 2997–3007.

Chevret P, Veyrunes FR, Britton-Davidian J (2005) Molecular phylogeny of the genus Mus (Rodentia: Murinae) based on mitochondrial and nuclear data. *Biological Journal of the Linnean Society*, **84**, 417–427.

Coyne JA, Orr AH (2004) *Speciation*. Sinauer Associates, Sunderland, Massachusetts.

Cucchi T, Vigne J-D, Auffray J-C (2005) First occurrence of the house mouse (*Mus musculus* domesticus Schwarz & Schwarz, 1943) in the Western Mediterranean: a zooarchaeological revision of subfossil occurrences. *Biological Journal of the Linnean Society*, **84**, 429–445.

Darvish J, Orth A, Bonhomme F (2006) Genetic transition in the house mouse, *Mus musculus* of Eastern Iranian Plateau. *Folia Zoologica*, **55**, 349–357.

Delaneau O, Coulonges C, Boelle P-Y *et al.* (2007) ISHAPE: new rapid and accurate software for haplotyping. *BMC Bioinformatics*, **8**, 205.

Din W, Anand R, Boursot P *et al.* (1996) Origin and radiation of the house mouse: clues from nuclear genes. *Journal of Evolutionary Biology*, **9**, 519–539.

Dod B, Jermiin LS, Boursot P *et al.* (1993) Counterselection on sex chromosomes in the *Mus musculus* European hybrid zone. *Journal of Evolutionary Biology*, **6**, 529–546.

Donnelly P, Tavare S (2003) Coalescents and genealogical structure under neutrality. *Annual Review of Genetics*, **29**, 401–421.

Duplantier J-M, Orth A, Catalan J, Bonhomme F (2002) Evidence for a mitochondrial lineage originating from the Arabian peninsula in the Madagascar House Mouse (*Mus musculus*). *Heredity*, **89**, 154–158.

Estoup A, Beaumont MA, Sennedot F *et al.* (2004) Genetic analysis of complex demographic scenarios: spatially

expanding populations of the cane toad, *Bufo marinus*. *Evolution*, **58**, 2021–2036.
Excoffier L, Estoup A, Cornuet J-M (2005) Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics*, **169**, 1727–1738.
Fagundes NJR, Ray N, Beaumont M *et al.* (2007) Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences*, **104**, 17614–17619.
Faure B, Jollivet D, Tanguy A, Bonhomme F, Bierne N (2009) Speciation in the deep sea: multi-locus analysis of divergence and gene flow between two hybridizing species of hydrothermal vent mussels. *PLoS ONE*, **4**, e6485.
Felsenstein J (2006) Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? *Molecular Biology and Evolution*, **23**, 691–700.
Frazer KA, Eskin E, Kang HM *et al.* (2007) A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature*, **448**, 1050–1055.
Gavrilets S (2003) Perspective: models of speciation: what have we learned in 40 years? *Evolution*, **57**, 2197–2215.
Gelman A, Meng X-L, Stern H (1996) Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, **6**, 733–760.
Geraldes A, Basset P, Gibson B *et al.* (2008) Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Molecular Ecology*, **17**, 5349–5363.
Good JM, Handel MA, Nachman MW (2008) Asymmetry and polymorphism of hybrid male sterility during the early stages of speciation in house mice. *Evolution*, **62**, 50–65.
Gregorová S, Forejt J (2000) PWD/Ph and PWK/Ph inbred mouse strains of *Mus m. musculus* subspecies – a valuable resource of phenotypic variations and genomic polymorphisms. *Folia Biologica*, **46**, 31–41.
Gündüz I, Rambau RV, Tez C, Searle JB (2005) Mitochondrial DNA variation in the western house mouse (*Mus musculus* domesticus) close to its site of origin: studies in Turkey. *Biological Journal of the Linnean Society*, **84**, 473–485.
Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–760.
Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences*, **104**, 2785–2790.
Hey J, Wakeley J (1997) A coalescent estimator of the population recombination rate. *Genetics*, **145**, 833–846.
Hudson RR (2002) Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics*, **18**, 337–338.
Hudson RR, Coyne JA (2002) Mathematical consequences of the genealogical species concept. *Evolution*, **56**, 1557–1565.
Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.
Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. *Genetics*, **132**, 583–589.
Huff CD, Xing J, Rogers AR, Witherspoon D, Jorde LB (2010) Mobile elements reveal small population size in the ancient ancestors of *Homo sapiens*. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 2147–2152.
Ingvarsson PK (2008) Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. *Genetics*, **180**, 329–340.
Keightley PD, Eyre-Walker A (2000) Deleterious mutations and the evolution of sex. *Science*, **290**, 331–333.
Kent WJ (2002) BLAT – the BLAST-like alignment tool. *Genome Research*, **12**, 656–664.
Kuhner MK (2009) Coalescent genealogy samplers: windows into population history. *Trends in Ecology & Evolution*, **24**, 86–93.
Laukaitis CM, Crister ES, Karn RC (1997) Salivary androgen-binding protein (ABP) mediates assortative mate selection in *Mus musculus*. *Evolution*, **51**, 2000–2005.
Laukaitis CM, Heger A, Blakley TD *et al.* (2008) Rapid bursts of androgen-binding protein (Abp) gene duplication occurred independently in diverse mammals. *BMC Evolutionary Biology*, **8**, 46.
Leuenberger C, Wegmann D (2010) Bayesian computation and model selection without likelihoods. *Genetics*, **184**, 243–252.
Li J-W, Yeung CKL, Tsai P-W *et al.* (2010) Rejecting strictly allopatric speciation on a continental island: prolonged postdivergence gene flow between Taiwan (Leucodioptron taewanus, Passeriformes Timaliidae) and Chinese (L. canorum canorum) hwameis. *Molecular Ecology*, **19**, 494–507.
Lisiecki LE, Raymo ME (2005) A Pliocene-Pleistocene stack of 57 globally distributed benthic delta O-18 records. *Paleoceanography*, **20**, PA1003.
Liu YH, Takahashi A, Kitano T *et al.* (2008) Mosaic genealogy of the *Mus musculus* genome revealed by 21 nuclear genes from its three subspecies. *Genes & Genetic Systems*, **83**, 77–88.
Loader CR (1996) Local likelihood density estimation. *Annals of Statistics*, **24**, 1602–1618.
Macholán M, Munclinger P, Sugerková M *et al.* (2007) Genetic analysis of autosomal and X-linked markers across a mouse hybrid zone. *Evolution*, **61**, 746–771.
Matute DR (2010) Reinforcement of gametic isolation in Drosophila. *PLoS Biology*, **8**, e1000341.
Matute DR, Butler Ia, Turissini DA, Coyne JA (2010) A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science (New York, N.Y.)*, **329**, 1518–1521.
Mayr E (1942) *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. Harvard University Press, Cambridge, Massachusetts.
Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J (2009) A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science*, **323**, 373–375.
Milishnikov AN, Lavrenchenko LA, Lebedev VS (2004) Origin of the house mice (superspecies complex *Mus musculus* sensu lato) from the Transcaucasia region: a new look at dispersal routes and evolution. *Russian Journal of Genetics*, **40**, 1011–1026.
Myers S, Bowden R, Tumian A *et al.* (2009) Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science*, **327**, 876–879.
Nadachowska K, Babik W (2009) Divergence in the face of gene flow: the case of two newts (Amphibia: Salamandridae). *Molecular Biology and Evolution*, **26**, 829–841.
Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov Chain Monte Carlo approach. *Genetics*, **158**, 885–896.

Niemiller ML, Fitzpatrick BM, Miller BT (2008) Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae: Gyrinophilus) inferred from gene genealogies. *Molecular Ecology*, **17**, 2258–2275.

Orr AH (1995) The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics*, **139**, 1805–1813.

Orr AH, Turelli M (2001) The evolution of postzygotic isolation: accumulating dobzhansky-muller incompatibilities. *Evolution*, **55**, 1085–1094.

Orth A, Lyapunova E, Kandaurov A *et al.* (1996) L'espèce polytypique *Mus musculus* en transcaucasie. *Comptes Rendus des de l'Academie des Sciences*, **319**, 435–441.

Palero F, Lopes J, Abello P *et al.* (2009) Rapid radiation in spiny lobsters (Palinurus spp) as revealed by classic and ABC methods using mtDNA and microsatellite data. *BMC Evolutionary Biology*, **9**, 263.

Pool JE, Nielsen R (2009) Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics*, **181**, 711–719.

Prager EM, Orrego C, Sage RD (1998) Genetic variation and phylogeography of central Asian and other house mice, including a major new mitochondrial lineage in Yemen. *Genetics*, **150**, 835–861.

Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, **16**, 1791–1798.

Przeworski M (2002) The signature of positive selection at randomly chosen loci. *Genetics*, **160**, 1179–1189.

Ptak SE, Przeworski M (2002) Evidence for population growth in humans is confounded by fine-scale population structure. *Trends in Genetics*, **18**, 559–563.

Rajabi-Maham H, Orth A, Bonhomme F (2008) Phylogeography and postglacial expansion of *Mus musculus* domesticus inferred from mitochondrial DNA coalescent, from Iran to Europe. *Molecular Ecology*, **17**, 627–641.

Raufaste N, Orth A, Belkhir K *et al.* (2005) Inferences of selection and migration in the Danish house mouse hybrid zone. *Biological Journal of the Linnean Society*, **84**, 593–616.

Roberts RB, Ser JR, Kocher TD (2009) Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. *Science*, **326**, 998–1001.

Ross-Ibarra J, Wright SI, Foxe JP *et al.* (2008) Patterns of polymorphism and demographic history in natural populations of *Arabidopsis lyrata*. *PLoS ONE*, **3**, e2411.

Sage RD, Atchley WR, Capanna E (1993) House mice as models in systematic biology. *Systematic Biology*, **42**, 523–561.

She JX, Bonhomme F, Boursot P, Thaler L, Catzeflis F (1990) Molecular phylogenies in the genus Mus: comparative analysis of electrophoretic, scnDNA hybridization, and mtDNA RFLP data. *Biological Journal of the Linnean Society*, **41**, 83–103.

Sing T, Sander O, Beerenwinkel N, Lengauer T (2005) ROCR: visualizing classifier performance in R. *Bioinformatics (Oxford, England)*, **21**, 3940–3941.

Smadja C, Ganem G (2002) Subspecies recognition in the house mouse: a study of two populations from the border of a hybrid zone. *Behavioral Ecology*, **13**, 312–320.

Smadja C, Ganem G (2005) Asymmetrical reproductive character displacement in the house mouse. *Journal of Evolutionary Biology*, **18**, 1485–1493.

Smadja C, Ganem G (2007) Divergence of odorant signals within and between the two European subspecies of the house mouse. *Behavioral Ecology*, **19**, 223–230.

Smadja C, Ganem G (2008) Divergence of odorant signals within and between the two European subspecies of the house mouse. *Behavioral Ecology*, **19**, 223–230.

Smadja C, Catalan J, Ganem G (2004) Strong premating divergence in a unimodal hybrid zone between two subspecies of the house mouse. *Journal of Evolutionary Biology*, **17**, 165–176.

Song Y, Endepols S, Klemann N *et al.* (2011) Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Current Biology*, **21**, 1296–1301.

Storchova R, Gregorova S, Buckiova D *et al.* (2004) Genetic analysis of X-linked hybrid sterility in the house mouse. *Mammalian Genome*, **15**, 515–524.

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 9440–9445.

Strasburg JL, Rieseberg LH (2008) Molecular demographic history of the annual sunflowers *Helianthus annuus* and H-petiolaris – large effective population sizes and rates of long-term gene flow. *Evolution*, **62**, 1936–1950.

Strasburg JL, Rieseberg LH (2010) How robust are ''isolation with migration'' analyses to violations of the IM model? A simulation study *Molecular Biology and Evolution*, **27**, 297–310.

Stump AD, Fitzpatrick MC, Lobo NF *et al.* (2005) Centromere-proximal differentiation and speciation in *Anopheles gambiae*. *Proceedings of the National Academy of Sciences*, **102**, 15930–15935.

Suzuki H, Shimada T, Terashima M, Tsuchiya K, Aplin K (2004) Temporal, spatial, and ecological modes of evolution of Eurasian Mus based on mitochondrial and nuclear gene sequences. *Molecular Phylogenetics and Evolution*, **33**, 626–646.

Talley HM, Laukaitis CM, Karn RC (2001) Female preference for male saliva: implications for sexual isolation of *Mus musculus* subspecies. *Evolution*, **55**, 631–634.

Tavare S, Balding DJ, Griffiths RC, Donnelly P (1997) Inferring coalescence times from DNA sequence data. *Genetics*, **145**, 505–518.

Teeter KC, Thibodeau LM, Gompert Z *et al.* (2010) The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution*, **64**, 472–485.

Terashima M, Furusawa S, Hanzawa N *et al.* (2006) Phylogeographic origin of Hokkaido house mice (*Mus musculus*) as indicated by genetic markers with maternal, paternal and biparental inheritance. *Heredity*, **96**, 128–138.

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.

Turelli M, Barton NH, Coyne JA (2001) Theory and speciation. *Trends in Ecology & Evolution*, **16**, 330–343.

Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution*, **16**, 381–390.

Vyskocilova M, Trachtulec Z, Forejt J, Pialek J (2005) Does geography matter in hybrid sterility in house mice? *Biological Journal of the Linnean Society*, **84**, 663–674.

Wegmann D, Leuenberger C, Excoffier L (2009) Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics*, **182**, 1207–1218.

White MA, Ane C, Dewey CN, Larget BR, Payseur BA (2009) Fine-scale phylogenetic discordance across the house mouse genome. *PLoS Genetics*, **5**, e1000729.

White MA, Steffy B, Wiltshire T, Payseur BA (2011) Genetic dissection of a key reproductive barrier between nascent species of house mice. *Genetics*, **189**, 289–304.

Woerner AE, Cox MP, Hammer MF (2007) Recombination-filtered genomic datasets by information maximization. *Bioinformatics*, **23**, 1851–1853.

Won Y-J, Hey J (2005) Divergence population genetics of chimpanzees. *Molecular Biology and Evolution*, **22**, 297–307.

Yang H, Bell TA, Churchill GA, Pardo-Manuel de Villena F (2007) On the subspecific origin of the laboratory mouse. *Nature Genetics*, **39**, 1100–1107.

Yang H, Wang JR, Didion JP *et al.* (2011) Subspecific origin and haplotype diversity in the laboratory mouse. *Nature Genetics*, **43**, 648–655.

Yonekawa H, Moriwaki K, Gotoh O *et al.* (1988) Hybrid origin of Japanese mice ''*Mus musculus* molossinus'': evidence from restriction analysis of mitochondrial DNA. *Molecular Biology and Evolution*, **5**, 63–78.

Data accessibility

DNA sequences: EMBL accession numbers HE588206–HE589444.

Final DNA sequence assembly: uploaded in Supporting information.

Supporting information

Data S1 Aligned sequences used for the ABC inferences. For each sequence we give the locus name, the sample name, and an arbitrary haplotype name.

Fig. S1 Efficiency of the simulation of the consequences of locus choice. Each graph corresponds to one of the 57 loci. Histograms represent the distribution of the distance between one *domesticus* and one *musculus*, in number of substitutions. Black histogram: distribution genome-wide in 1 kb windows of non-coding DNA, estimated from complete sequences of strains WSB/EiJ and PWK/PhJ. Red histogram: same but from windows with ≥4 SNPs in the Perlegen experiment. Dotted black histogram: Distribution obtained from the simulations, aimed at following the red distribution. All distributions were calculated after scaling by the divergence between the house mouse and *M. spretus* to account for differences of mutation rates between loci.

Fig. S2 Distributions of the priors realized after discarding simulations not obeying the criterion to simulate locus choice bias. Three histograms are superimposed here. The red one is the prior realized using all simulation sets (expected to follow the uniform prior). The white histogram shows the prior actually used after keeping only simulation sets with no more than three loci not obeying the criterion. The grey histogram is the distribution for simulation sets retained despite having up to three loci not obeying the criterion.

Fig. S3 Results of the posterior predictive simulations for the 57 locus dataset. Each graph shows the distribution of one statistic in the simulated dataset, under the F (continuous curve) or IF (dotted curve) model. The vertical line indicates the value of the statistics from the real data. The *P*-values for both models are given on each graph.

Fig. S4 Result of the ROC analysis for the comparison of the F and IF models. The full and open red dots correspond to posterior probabilities of 0.7 and 0.5 for IF, respectively.

Fig. S5 Estimation of the quality and precision of the estimation of parameter *Tm* in the IF model. Each square in these plots represents the results of 100 simulations for one value of *Tm/Ts* and one of *Ts*. The left graph shows the proportion of times the true value of *Tm* was included in the HPD90 interval, according to the color chart. The graph on the right gives the color-coded absolute deviation between the true value of *Tm* and its modal estimate from the ABC analysis. The thick black lines delineate the region of parameter space broadly corresponding to the posterior distribution obtained from our data.

Table S1 Coordinates of the 61 amplicons in mouse genome sequence Build37, and sequences of the PCR primers used.

Table S2 Descriptive statistics for the 57 locus dataset.

Table S3 Tests of goodness of fit of individual loci to the F and IF models. *P*-values result from locus-specific posterior predictive simulations. They are signed to indicate the direction of deviation from the median of the distribution. $*P < 0.05$; $**P < 0.01$, $***P < 0.001$, 1: outlier locus for just one model, 2: outlier locus for both models.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.