JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.12280

Inferring the mode of colonization of the rapid range expansion of a solitary bee from multilocus DNA sequence variation

S. DELLICOUR*, P. MARDULYN*, O. J. HARDY*, C. HARDY*, S. P. M. ROBERTS† & N. J. VEREECKEN*[1]

*Evolutionary Biology and Ecology, Université Libre de Bruxelles (ULB), Brussels, Belgium
†Centre for Agri-Environmental Research, University of Reading, Reading, UK

Keywords:

coalescence simulations;
Colletes;
phylogeography;
pollen specialization;
range expansion;
solitary bee.

Abstract

Rapid geographic range expansions can have dramatic effects on the distribution of genetic diversity, both within and among populations. Based on field records collected over the past two decades in Western Europe, we report on the rapid geographic range expansion in *Colletes hederae*, a solitary bee species. To characterize how this expansion shaped the distribution of genetic diversity within and among populations, we performed a genetic analysis based on the sequencing of three nuclear loci (RNAp, CAD and WgL). We then simulated the evolution of DNA sequences under a spatially explicit model of coalescence to compare different hypotheses regarding the mode of colonization associated with this rapid expansion and to identify those that are most consistent with the observed molecular data. Our genetic analyses indicate that the range expansion was not associated with an important reduction in genetic diversity, even in the most recently colonized area in the United Kingdom. Moreover, little genetic differentiation was observed among populations. Our comparative analysis of simulated data sets indicates that the observed genetic data are more consistent with a demographic scenario involving relatively high migration rates than with a scenario based on a high reproduction rate associated with few migrants. In the light of these results, we discuss the factors that might have contributed to the rapid geographic range expansion of this pollen-specialist solitary bee species across Western Europe.

Introduction

In the course of their evolution, most species regularly experience major modifications of their geographic range, for example, as a result of climate change, changes in dispersal capabilities, reproduction rate or resource availability, or through the regulating effect of natural enemies (Wallner, 1987; Ferriere *et al.*, 2000; Bowler & Benton, 2005; Phillips *et al.*, 2010). These changes strongly impact on the distribution of genetic diversity within a species, as already highlighted in theoretical studies on range expansion (Ray *et al.*, 2003), range contraction or range shifts (Arenas *et al.*, 2012). In general, theoretical and applied studies of the genetic consequences of a past range modification are expected to benefit greatly from the use of spatially explicit evolutionary models (e.g. Excoffier *et al.*, 2009). However, despite the availability of various spatially explicit models of population evolution (reviewed, e.g., in Ray & Excoffier, 2009; Frank *et al.*, 2011), such models are only used sporadically in population genetic and phylogeographic studies.

Although species range variations certainly occur in bees (Hymenoptera, Apoidea), the dispersal behaviour of individual bees is considered to be limited *a priori*. Indeed, many groups of bees seem to be phylopatric,

Correspondence: Simon Dellicour, Evolutionary Biology and Ecology, Université Libre de Bruxelles (ULB), av. FD Roosevelt 50, B-1050 Brussels, Belgium.Tel.: +3226504525; fax: 3226504525; e-mail: Simon.Dellicour@ulb.ac.be
[1]Present address: Landscape Ecology and Plant Production Systems, Université Libre de Bruxelles (ULB), Campus de la Plaine CP 264/2 Boulevard du Triomphe, B-1050 Brussels, Belgium

that is, they generally remain in or return to the same nesting site across generations (Michener, 1974, 1979, 2007; Westrich, 1990), and their foraging range is generally limited (see Gathmann & Tscharnke, 2002; Zurbuchen *et al.*, 2010; Dorchin *et al.*, 2012 and references therein). Furthermore, distributional data indicate that most bees are not particularly good at bypassing major dispersal barriers (e.g. water barriers such as oceans or mountain ranges) to reach distant habitats without human assistance (Michener, 1979, 2007; Goulson *et al.*, 2011). Reports of long-range jumps in dispersal (100–1000 km) and rapid geographic expansion of bees are indeed relatively scarce (see, e.g., Camargo *et al.*, 1988; Cane, 2003; Fuller *et al.*, 2005; Hinojosa-Díaz *et al.*, 2005; Michener, 2007 and references therein; Zayed *et al.*, 2007; Gibbs & Sheffield, 2009; Vereecken & Barbier, 2009; Strange *et al.*, 2011). In contrast, an increasing number of studies indicate that natural populations of many managed and wild bees, particularly pollen-specialist species, undergo dramatic declines around the world (contraction in range and/or decrease in population sizes), raising worries about the future of the bees' ecosystem services and their contribution to the pollination of cultivated crops (Steffan-Dewenter *et al.*, 2005; Williams, 2005; Biesmeijer *et al.*, 2006; Fitzpatrick *et al.*, 2007; Colla & Packer, 2008; Goulson *et al.*, 2008; Patiny *et al.*, 2009; Potts *et al.*, 2010; Cameron *et al.*, 2011; Gallai *et al.*, 2011; Burkle *et al.*, 2013; Garibaldi *et al.*, 2013).

Among the pollen-specialist bee species that seem to resist the declining trend is *Colletes hederae*, a univoltine species that was described as new to science in 1993 (Schmidt & Westrich, 1993). It is morphologically, genetically and ecologically distinct from its sister species in the monophyletic *C. succinctus* group with different floral choices and geographic distributions (Schmidt & Westrich, 1993; Kuhlmann *et al.*, 2007, 2009), and its flight activity coincides with the long flowering period of its key pollen forage plant, *Hedera helix* (Schmidt & Westrich, 1993; Bischoff *et al.*, 2005; Kuhlmann *et al.*, 2007; Müller & Kuhlmann, 2008; Westrich, 2008). At the time of its description, *C. hederae* was considered to be of very sporadic occurrence in Southern and Western Europe, yet contrary to the decline observed in a wide taxonomic range of pollen-specialist species, multiple independent observations suggest that this species has expanded beyond its initial distribution range in recent years (Vereecken *et al.*, 2009). The first line of evidence comes from the growing number of independent reports on small groups of pollen-laden females initiating their individual nests in hitherto unoccupied habitats (Herrmann, 2007; Frommer, 2010). Perhaps the most convincing line of evidence for the rapid and recent geographic range expansion of *C. hederae* stems from observations made in the UK, where the presence of this species was only confirmed in 2001 along the Dorset coast (Cross, 2002), despite increased search efforts by experienced UK entomologists and apidologists along the UK coasts since 1993, the time when *C. hederae* was described as new to science. Today, *C. hederae* has spread across much of southern England, but also increasingly inland, and is now extremely plentiful in several coastal localities with lots of individual nesting aggregations encompassing tens of thousands of individuals (Fig. S2).

Here, we first provide a detailed and comprehensive report on the recent history of geographic range expansion of *C. hederae* based on field records collected over the past two decades. Because the detailed history of this spread can be inferred from these observations, it offers a unique opportunity to study the mode and genetic consequences of a very recent dispersion, by characterizing the geographic distribution of genetic variation with molecular markers. We therefore investigated sequence data from three independent DNA fragments of samples collected across a large portion of its geographic range, to test whether or not the range expansion involved the migration of a few individuals across relatively long distances and the extent to which this range expansion was associated with a reduction in genetic diversity. Finally, we conducted simulations of DNA sequence evolution under a spatially explicit model of coalescence, to compare the likelihood of different scenarios of dispersion that (i) could explain the observed current distribution of genetic variation and (ii) are consistent with our field records. Finally, the results shed light on the dispersal mechanisms of this pollen-specialist solitary bee species and allow us to discuss the factors that might have contributed to its rapid geographic range expansion across Western Europe.

Materials and methods

Biological records and expansion history in Western Europe

We examined the expansion history of *Colletes hederae* as annual changes in the distribution from 1993 to 2010. The measures of expansion used were (i) the cumulative number of biological records collected each year and (ii) the cumulative number of 25 × 25 km occupied grid cells each year. All biological records were validated and compiled from field records accumulated through consistent sampling effort by the Bees, Wasps, and Ants Recording Society (BWARS, www.bwars.com) and the Observatoire des Abeilles (OA, www.oabeilles.org) through a joint pan-European survey involving both field naturalists and professional bee experts.

Sampling and sequencing

We sampled *C. hederae* in 22 localities covering its 2010 distribution range in Western Europe (Table 1, Fig. 1a). We found no evidence for major stretches of unsuitable

Table 1 Sampling localities and distribution of haplotypes for the three loci used in this study.

Population Locality	Geographical coordinates	*n* (RNAp)	Haplotypes (no. of copies) RNAp	*n* (CAD)	Haplotypes (no. of copies) CAD	*n* (WgL)	Haplotypes (no. of copies) WgL
Bruxelles (Belgium)	50.82°N, 4.32°E	5	14(4), 26	5	1, 3, 4, 5(2)	5	1, 2(3), 3
Genk (Belgium)	51.05°N, 3.73°E	4	14, 19, 34, 41	4	1, 4, 5, 6	4	3(2), 4(2)
Ain (France)	46.20°N, 5.22°E	5	10, 16, 32, 37, 41	5	1(2), 3, 4, 5	4	2, 4(3)
Angervilliers (France)	48.58°N, 2.05°E	4	8, 14, 26, 33	5	1, 3, 5(3)	4	3, 4(3)
Avignon (France)	43.95°N, 4.82°E	9	9, 12, 15, 17, 20, 29, 30, 35, 38	10	1, 2, 4, 5(3), 6(4)	10	3(7), 4(3)
Brest (France)	48.38°N, 4.48°W	2	14(2)	3	1(2), 6	3	2, 3(2)
Etampes (France)	48.43°N, 2.15°E	5	14(2), 15, 24, 41	4	1, 3, 4, 6	5	2, 3(3), 4
Le Nizan (France)	44.47°N, 0.27°W	10	2, 8, 10, 16, 22, 26, 32, 36, 39(2)	10	1(3), 3(3), 4, 6(3)	10	3(2), 4(8)
Vendôme (France)	47.78°N, 1.05°E	4	3, 21, 28, 33	5	4, 5(2), 6(2)	5	2(2), 4(3)
Westhalten (France)	47.95°N, 7.25°E	5	14, 18, 26(2), 32	5	1, 3(2), 4(2)	5	3(2), 4(2), 5
Hamel (France)	50.28°N, 3.08°E	2	3, 26	2	1, 5	2	2, 4
Bad-Dürkheim (Germany)	47.95°N, 8.17°E	5	8(3), 14, 24	5	3(3), 5(2)	5	3, 4(4)
Radolfzell (Germany)	47.73°N, 8.97°E	5	3(2), 14, 19, 42	5	3(3), 6(2)	4	1, 3(3)
Lausanne (Switzerland)	46.52°N, 6.62°E	10	6, 8, 9, 13, 14, 20, 25, 32, 37, 43	10	2, 3(2), 4, 5(2), 6(4)	10	2, 3(3), 4(5), 6
Neuchâtel (Switzerland)	46.98°N, 6.92°E	8	6(2), 8(2), 14, 18, 23, 24	10	1(2), 3(3), 4, 5(2), 6(2)	10	3(7), 4(3)
Guernsey (England)	49.47°N, 2.59°W	4	1, 2(3)	4	1(2), 5(2)	4	4(4)
Dorset (England)	50.60°N, 2.04°W	5	7, 8, 14(2), 28	–	–	5	3(4), 4
Easthbourne (England)	50.76°N, 0.25°E	2	14(2)	2	5(2)	2	4(2)
Oxfordshire (England)	51.76°N, 1.25°W	5	5(2), 14(3)	3	1(3)	5	3, 4(4)
Cornwall (England)	50.11°N, 5.38°W	5	1, 2(2), 5, 14	5	1, 3(2), 5(2)	5	4(5)
Salisbury (England)	51.07°N, 1.80°W	2	14, 27	1	1	4	3(2), 4(2)
Lesyos (Greece)	39.08°N, 26.38°E	12	4, 11(2), 31(2), 40(7)	2	3(2)	14	3(3), 4(11)

habitat separating the sampling localities; the heterogeneous distribution of the sampling sites across the species range (Fig. 1a) is due to our incomplete knowledge on the location of available populations in certain areas and does not reflect, to the best of our knowledge, an actual variation in population density. Genomic DNA of 111 male and 7 female individuals was extracted using the Qiagen DNeasy® Blood & Tissue kit. Half a thorax per specimen was grounded in the Qiagen ATL buffer and incubated overnight with proteinase K at 56 ° *C*. The remaining DNA extraction steps were conducted as described in the manufacturer's protocol. We sequenced 118 samples of a ~850-base pair (bp)-long fragment of the protein-coding nuclear gene RNAp (*RNA polymerase II*), 105 samples of a ~1000-bp-long fragment of the protein-coding nuclear gene CAD (*Conserved ATPase Domain*), 125 samples of a ~750-bp-long fragment of the WgL (*Wingless*) gene and 32 samples of a ~850-bp-long fragment of the mitochondrial gene COI (*cytochrome oxidase I*). All fragments were PCR-amplified with the TrueStart Hot Start *Taq* DNA polymerase, following the guidelines in the manufacturer's protocol (Fermentas International Inc.). The RNAp fragment was amplified (annealing temperature of 57 °C) using primers Polfor2a and Polrev2a (Danforth *et al.*, 2006), the CAD fragment (annealing temperature of 55 °C) with primers CAD-Mel-for1 (5′-GAR CCY AGY CTC GAT TAY TG-3′) and Ap1098rev2 (Danforth *et al.*, 2006), the WgL fragment (annealing temperature of 63.5 °C) with primers Bee-wg-For2 and Lep-Wg2a-Rev (Almeida & Danforth, 2009) and the COI fragment with primers Jerry and Pat (Simon *et al.*, 1994). The CAD forward primer CAD-Mel-for1 is a modified version of the primer Ap787for2 (Danforth *et al.*, 2006). Using these primers, a portion of one intron was included in the sequenced CAD fragment (~70 bp) and another small intron (70 bp) was included in the sequenced WgL fragment. All the haplotype sequences gathered for this project are available in GenBank under accession numbers JX431564–JX431618.

Data analyses

Sequences were aligned using the MUSCLE algorithm (Edgar, 2004) implemented in CODONCODE ALIGNER (v. 3.7.1.1, Codon Code Corporation). These alignments were checked manually and pruned at both 5′ and 3′ ends. Only one gap of 4 bp was detected, in the CAD data set. This gap was recorded as a single character for the haplotype network inference, summary statistics computations and comparisons with simulated data sets. We used the maximum likelihood method implemented in the software PHASE 2.2.1 (Stephens *et al.*, 2001; Stephens & Donnelly, 2003) to reconstruct the haplotype phase of the few diploid (female) individuals. The algorithm of PHASE is based on the comparison between sequences identified in haploid (or homozygous) individuals and unphased sequences coming from heterozygous individuals. For each locus, we conducted three independent runs of 10 000 iterations, while

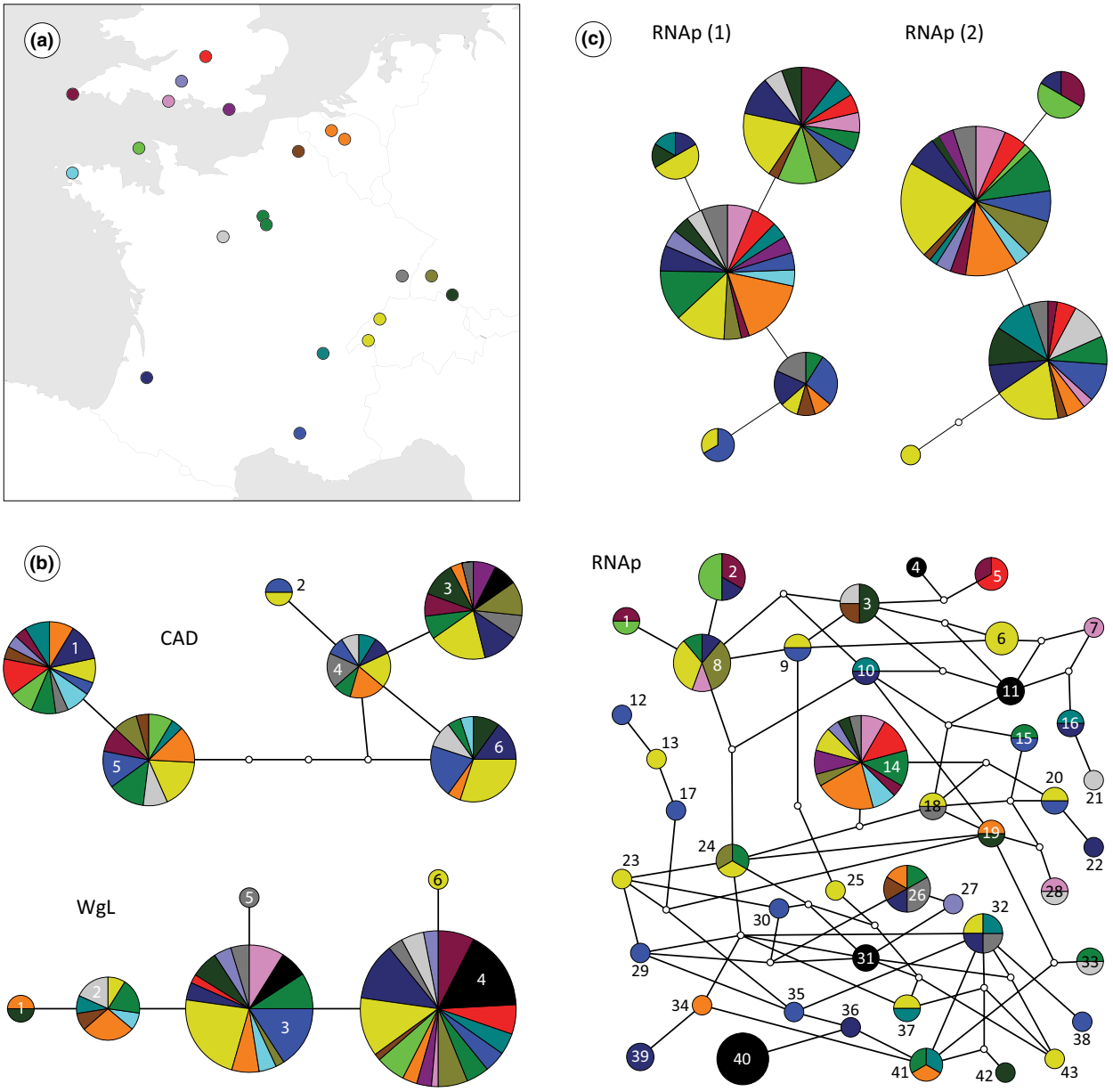

Fig. 1 (a) sampling localities of *Colletes hederae* in Western Europe (all sampling sites except "Lesvos" in Greece which was not included in the coalescent simulations). Each locality is identified by a specific colour. Two pairs of localities were pooled together under the same colour because of their relative geographic proximity: "Angervilliers" has been pooled with "Etampes" (S-Paris, France) and "Lausanne" with "Neuchâtel" (SE-Switzerland). (b) median-joining networks for the three nuclear gene fragments (RNAp, CAD and WgL) used in this study. Each sequenced haplotype is represented by a circle, the size of which is proportional to its overall frequency, and identified by a unique number (see also Table 1). Each line in the network represents a single mutational change. Small white circles indicate intermediate haplotypes not included in our dataset that are necessary to link all observed haplotypes to the network. (c) median-joining networks for two RNAp fragments which do not contain any identified recombination site. RNAp (1) a 221 bp fragment (the first 221 bp on the original RNAp alignment) and RNAp (2) a 287 bp fragment (between nucleotides 511 and 798 of the original RNAp alignment). Haplotype colours correspond to the sampling localities displayed on Fig. 1a.

thinning at every 100 steps and discarding the first 1000 samples as burn-in. Convergence among chains was checked by comparing haplotype reconstructions inferred by the three independent runs. Median-joining networks (Bandelt *et al.*, 1999) were inferred for each gene fragment using the software NETWORK 4.6.6 (available at http://www.fluxus-engineering.com) with $\varepsilon = 0$.

To compare the genetic diversity between the ancestral range (as defined by the 1994 observations) and the newly colonized areas (Fig. 2), sampled populations in Western Europe (i.e. excluding 'Lesvos' in Greece) were arbitrary categorized into three separate groups: group n°1 with all the sampled populations located in the area of origin (i.e. the range before the last recent expansion that begun 20–30 generations ago), group n°2 with sampled populations located in the newly colonized areas in northern France and Belgium and group n°3 with sampled populations located in the most recently colonized area in southern England. We removed the population Lesvos because we had very reliable and complete biogeographical records (field observations) for the colonization of England but not of south-eastern Europe (Fig. S1). Global Φ_{ST} statistics among all sampled populations and AMOVA Φ-statistics (Excoffier *et al.*, 1992) computed for the partition corresponding to the three defined groups of populations were computed using the toolbox SPADS 1.0 (Dellicour & Mardulyn, *in press* available at http://ebe.ulb.ac.be/ebe/Software.html). Statistical significance of all Φ-statistic values was assessed by recalculating these on 10 000 random permutations of the original data sets. We also used SPADS to estimate (i) the relative number of different haplotypes N_H/N_S (i.e. the number of different haplotypes identified in the group divided by the number of sampled sequences in this group), (ii) the allelic richness A_R (El Mousadik & Petit, 1996) and (iii) the nucleotide diversity π (Nei & Li, 1979) within the three defined groups of sampled populations. Although N_H/N_S and A_R are only based on allelic frequencies, π is also based on allelic distances (i.e. the number of mismatches between different haplotypes). In particular, we focus on the variation in these three diversity measures between groups n°1 and n°3. This is because our biogeographical records allow us to state with high confidence that populations in group n°1 are included in the area of origin of the species and that populations in group n°3 were colonized very recently. We have less support for the recent status of the three sampled populations in group n°2 (see Fig. S1).

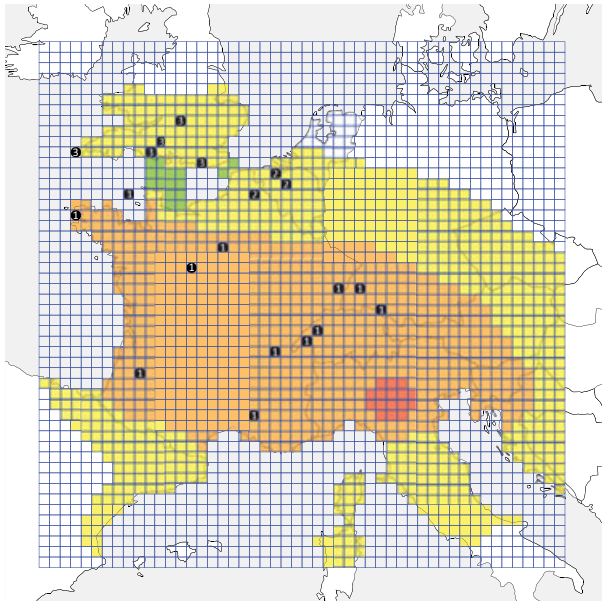

Fig. 2 Grid (30 × 30 km cells) used as a reference for the coalescent simulations performed with PHYLOGEOSIM 1.0. Black circles indicate grid cells with sampled populations, red cells correspond to the fictive area of origin occurring 10 000 generations/years ago, orange cells are the only accessible cells before the last recent expansion (whose beginning was set at 25 generations/years ago according to biogeographical records), yellow and green cells are the newly accessible cells since the beginning of this expansion, with green cells characterized by an effective population size 10% lower than that of other cells. White numbers refer to one of the 3 user-defined group to which each sampled population belongs (see text). The maximal effective population sizes of accessible cells was fixed at Ne = 1000, Ne = 10 000 or Ne = 100 000 depending on the simulation. The orange cells correspond to the only accessible cells for the remaining gene copies when simulations reach 25 generations back in time (ancestral range).

Our RNAp haplotype network displayed a high number of loops (a loop, or cycle, reveals equally parsimonious alternative connections among alleles, that is, uncertainties about the evolutionary relationships in the haplotype network, Fig. 1b). The presence of loops in a haplotype network is most likely explained by two mechanisms: (1) multiple convergent mutations at a single site or (2) one or more recombination events that have occurred within the sequenced DNA fragment. We used the PHI test (Bruen *et al.*, 2006) implemented in the software SPLITSTREES 4.12.3 (Huson & Bryant, 2006) to test for the presence of recombination events. We then estimated the nucleotide position of the inferred recombination events using (i) the methods RDP (Martin & Rybicki, 2000) and MaxChi (Maynard Smith, 1992) implemented in the RDP4 package (Martin *et al.*, 2010) and (ii) the method of Hudson and Kaplan (Hudson & Kaplan, 1985) implemented in the software DNASP 5.0 (Librado & Rozas, 2009). Finally, by taking into account the estimated positions of the recombination sites, we generated median-joining subnetworks devoid of any loop.

Coalescence simulations

To investigate under what conditions a recent and rapid range expansion is able to generate the observed pattern of genetic diversity, computer simulations were run to identify the migration scenario that is most consistent with the observed pattern of genetic variation. In other words, the aim of the comparison between our real data set and the simulated sequences is to identify under which conditions this rapid range expansion could have resulted in the observed geographic distribution of intraspecific genetic diversity.

To simulate the evolution of genetic diversity in a geographic framework, we have superimposed a 2-dimensional grid on the map showing the evolution of the range of *C. hederae*. We used the software PHYLOGEOSIM 1.0 (available, with a detailed manual, at http://ebe.ulb.ac.be/ebe/Software.html) to simulate the evolution of genetic data on this grid, with each cell treated as a separate population (Fig. 2). This was implemented as a two-step process. First, starting from initial values defined by the user, the evolution of effective population sizes in each cell is simulated forward in time (one generation per year) according to the parameters specified *a priori*: short-distance (between adjacent populations) and long-distance (between populations separated by two grid cells) forward migration rates *mf*1 and *mf*2, reproduction rate t_R (population growth rate) and maximum effective size *Ne* of each cell (carrying capacity). This preliminary forward simulation is used to generate backward-in-time migration rates and effective sizes that are required for the main coalescent simulation. Indeed, these parameters may vary between generations during modifications of the species range, and a preliminary forward simulation allows to automatically assign values for these parameters at each generation. The alternative of setting these parameter values manually for the coalescent simulation would be tedious as well as subjective. A similar forward simulation approach was previously developed by Currat *et al.* (2004) and Ray *et al.* (2010) in the context of a geographically explicit model. Second, separately for each locus, a gene genealogy is generated through the simulation of the coalescence process for a user-defined sample that mimics the real data set (i.e. specifying the number of gene copies sampled in each cell), and the evolution of DNA sequences is simulated, using the same DNA sequence length and the same number of mutations as those identified in our real DNA sequence alignments.

Input files for PHYLOGEOSIM were constructed by superimposing a grid with cells of approximately 30 × 30 km to a map of Western Europe (Fig. 2). This grid resolution was chosen so that long-distance migration could be implemented by crossing two grid cells in a single generation. Indeed, the distance spanned by two grid cells is a good approximation of the distance separating the Channel Islands from the coasts of southern England, a distance that had to be crossed by individuals to colonize England. Furthermore, the very few experimental studies documenting bee dispersal distances are consistent with the fact that the reference for long-distance migration should not exceed the equivalent of two grid cells as defined in our study (Kraus *et al.*, 2009; Lepais *et al.*, 2010). For different sets of simulations, the following parameters of the preliminary forward simulations varied: (i) the reproduction rate t_R within each cell, (ii) the two forward migration rates *mf*1 and *mf*2 corresponding to short- and long-distance migration (i.e. between adjacent populations and populations separated by two grid cells), (iii) a matrix of ancestral effective population sizes, (iv) one or more matrices of maximum effective population sizes and finally, (v) the time (in number of generations) at which these matrices occur. Note that, in this study, we did not vary the times at which each matrix occurs. Indeed, as our biogeographical records informed us about the age of the expansion (i.e. we know that this species was first recorded in England in 2002), we constrained our model with this information and did not test different times for the beginning of the range expansion.

Step 1: selection of dispersion scenarios and corresponding simulation parameters

We used our field records to constrain the range of potential scenarios that could explain the sequence data. Based on these biogeographical records, we assumed that the geographic range expansion started from a defined ancestral area (corresponding to the orange cells in Fig. 2) 25 years ago, that is, in a time window of 25 generations, as *C. hederae* is a univoltine species (1 generation/year). Both yellow and green cells in Fig. 2 were accessible at the beginning of this expansion. Because the green cells connect continental Europe to southern England, and include a large area of sea water, their maximum effective population size was set to 10% of the size of the other cells. This reduction in population size results in effect in reducing the level of gene flow between these two geographic entities, relative to that occurring among adjacent cells in other areas of the distribution. To make the simulated data match the observed data, we had to considerably restrict the range of the species starting 10 000 generations ago, going backward in time. Without this strong reduction in range, which allowed the coalescence of the remaining lineages in a relatively short period of time, the simulations generated genealogies with TMRCA values an order of magnitude larger than those observed in our sequence data. Although we arbitrarily located the restricted ancestral range of the species in northern Italy (red cells on Fig. 2), its specific position had no effect on our comparison of scenarios (data not shown).

In the absence of prior information, we tested a large range of short-distance and long-distance forward migration rates (Table 3), which are the proportions of individuals from a cell *A* migrating to an adjacent cell *B* in one generation (see the manual of PHYLOGEOSIM for further details). For each tested couple of short-distance and long-distance forward migration rates, the reproduction rate was adjusted to allow the migration wave to complete the colonization (i.e. to reach the furthest sampled populations) at the sampling generation. The resulting combinations of forward migration rates and reproduction rates correspond to variations in the strengths of isolation by distance and founder effects associated with this geographic expansion. As forward simulations are

stochastic processes, some of them did not allow the migration wave to reach the location of the farthest sampled populations on the grid (UK samples) in the allocated amount of time. These failed forward simulations were systematically discarded. Note that absolute migration rate values have no particular meaning outside the defined spatially explicit model of coalescence and will always directly depend on the selected grid resolution (i.e. the cell dimension). Ultimately, we selected four sets of parameter values to perform the simulations, defining four distinct scenarios consistent with the constraint of a rapid geographic range expansion (see details of parameters in Table 3). The first and fourth scenarios correspond to the two extreme cases of a continuum, with scenario 1 implementing high migration rates associated with a low reproduction rate, and scenario 4, low migration rates associated with a high reproduction rate. The two other scenarios are intermediate cases. It is interesting to notice that long-distance forward migration rate (*fm*2, Table 3) had to be sufficiently high to allow the range expansion to reach the limits of the current distribution in only 25 generations. In the four selected scenarios, this long-distance forward migration rate (*fm*2, Table 3) is only two times lower than the short-distance forward migration rate (*fm*1, Table 3). Because a cell is usually connected to twice more distant cells (through *fm*2) than to adjacent cells (through *fm*1), the probabilities of short- and long-distance migration rates from that cell are similar. This relationship between the short- and long-distance forward migration rates defined the shape of the dispersion considered in the four tested scenarios. These four scenarios were tested under three different maximum cell effective population sizes (*Ne* = 1000; 10 000 and 100 000). However, for a given scenario, the strength of the founder effect, as measured by the products *Ne*fm*1 and *Ne*fm*2 (the product between the maximum cell effective size and the two forward migration rates), was maintained by adjusting the forward migration rate. Forward migration rates are thus lower when *Ne* is larger (see Table 3). As already mentioned, the reproduction rate was every time adjusted to allow the migration wave to complete the colonization, resulting in widely different reproduction rate values among scenarios simulated under different maximum effective sizes.

Step 2: main simulations and comparison with the real data set

We performed a total of 10 000 independent backward simulations for each individual scenario. To take into account the stochastic variation associated with the forward simulation, we reiterated a forward simulation every 100 backward simulations to estimate the backward simulation parameters (backward migration rates and effective population sizes), yielding a total of 100 forward for 10 000 backward simulations. Several summary statistics were automatically computed by PHYLOGEOSIM on each simulated data set: the total number of different haplotypes $N_H tot$, Φ_{ST} (Excoffier *et al.*, 1992) among populations, N_{ST} and G_{ST} (Pons & Petit, 1995, 1996) among populations, IBDSC (isolation by distance slope coefficient; Rousset, 1997) and several genetic diversity statistics estimated within each population and within each defined group of populations: ratios X_H between the number of haplotypes in a user-defined group of populations and the total number of haplotypes, the relative number of different haplotypes N_H/N_S (the number of different haplotypes divided by the number of sampled sequences), the allelic richness A_R (El Mousadik & Petit, 1996), the nucleotide diversities π (Nei & Li, 1979) and the ratio π_R between the nucleotide diversity within a considered user-defined group of populations and the nucleotide diversity within the virtual group formed by all the other populations that are not in this group (Mardulyn *et al.*, 2009). For example, the relative nucleotide diversity of group n°1, π_R(G1), equals π(G1)/[π(G2 $\cup$ G3)]. All these summary statistics are described in Table S1. Combinations of statistics (e.g. the difference between N_{ST} and G_{ST} or the difference between the nucleotide diversities estimated in the two defined groups of populations) were also computed manually. We performed several principal component analyses (PCAs) to determine the best summary statistics for comparing simulated and real data sets (as suggested, e.g., in Veeramah *et al.*, 2012), that is, the most efficient ones to discriminate among data sets simulated under distinct demographic scenarios. The different PCAs were performed using the PCA function available in the R package ade4 (Chessel *et al.*, 2004; Dray *et al.*, 2007).

We combined the selected summary statistics into a chi-square statistic following the formula given below:

$$\chi_i^2 = \sum_j \left(\frac{\left(St_{j,i} - m_j\right)^2}{\sigma_j^2} \right)$$

where $St_{j,i}$ is the jth selected summary statistics estimated on the ith simulated data set (i ranges from 1 to 10 000), and where m_j and σ_j are, respectively, the average and the standard deviations of the jth statistic over the 10 000 simulations generated under the same scenario and set of parameters. For a given scenario and a given set of parameters, we thus obtained a distribution of 10 000 chi-square statistics. We then compared each distribution with the chi-square statistic estimated on the corresponding real data set:

$$\chi_0^2 = \sum_j \left(\frac{\left(St_{j,0} - m_j\right)^2}{\sigma_j^2} \right)$$

where $St_{j,0}$ is the jth selected summary statistics estimated on the real data set. $St_{j,0}$s statistics were estimated using the software SPADS 1.0 (Dellicour & Mardulyn, *in press*), a toolbox computing the same statistics as

PHYLOGEOSIM, but on data sets provided by the user (as opposed to data sets generated by the program itself in PHYLOGEOSIM). This comparison is based on a non-parametric hypothesis test returning a *P*-value equal to the proportion of simulated $\chi_i^2 \geq \chi_0^2$. Note that this *P*-value is highly dependent on the selected set of summary statistics and simply reflects the probability that the associated scenario generated the observed sequences (thus, the sum of probabilities calculated for the alternative scenarios considered can easily exceed 1). Its usefulness is restricted to the comparison of alternative scenarios. *P*-values obtained for each locus under a given set of parameters were then combined using the Fisher's method (Fisher, 1948) to obtain a unique and global *P*-value for each set of simulation parameters.

To evaluate the performance of our approach for comparing demographic scenarios, we conducted a test based on pseudo-observed data sets (pods), which is similar to the procedure suggested to validate a specific approximate Bayesian computation analysis (Wegmann *et al.*, 2009; Bertorelle *et al.*, 2010). More specifically, 1000 additional data sets were simulated under each of the three scenarios that were not identified as the most likely by our analysis. These pods were then compared with the same simulated data that were used to analyse the real data, generated under the scenario previously identified as the most likely (scenario 1 or 2, depending on the combination of parameter values tested; see results), and we recorded the number of times that this validation analysis failed to reject the most likely scenario as having generated the pseudo-observed data. Our aim was to obtain an estimation of the type II error, estimated here as the proportion of pods that are not rejected when compared with data sets simulated under our most likely scenario (combined *P*-value > 0.05). In other words, these estimated proportions correspond to a measure of the rate at which our most likely scenario was chosen when data sets were generated under the alternative scenarios.

Results

Expansion history of *C. hederae* in Western Europe

Our survey of the historical distribution of *Colletes hederae* indicates a clear and recent geographic expansion: both the cumulative number of records and the cumulative number of grid cells occupied by this species in Europe have increased by a 6- to 7-fold factors between 2001 and 2010 (Fig. 3), which also coincides with the time when independent reports have recorded *C. hederae* as new addition to the wild bee fauna of several European countries (Austria, Belgium, The Grand Duchy of Luxembourg, Greece, Serbia, Spain, Switzerland, the Netherlands, The United Kingdom) (Fig. S1, S2; Vereecken *et al.*, 2009).

Back in 1993, *C. hederae* was known from specimens collected in several localities in Europe, including the Channel Islands, only a few tens of kilometres off the UK coasts. Apidologists from the Bees, Wasps and Ants Recording Society (www.bwars.com, which hold a complete database for all UK bee species) have investigated stands of ivy flowers since 1993 along the coasts of southern England, hoping to rapidly add the record of the ivy bee to the fauna of England. Despite intensive year-by-year investigations, the presence of *C. hederae* on the UK coasts was only confirmed in 2001 in a locality that had been thoroughly surveyed in previous years (Cross, 2002). The individual nesting aggregations observed today at these UK localities colonized in 2001 often encompass many thousands of individuals that make them obvious to the general public. No such aggregation has been observed for years in the newly colonized areas of its current distribution.

DNA sequences data set

Surprisingly, we found no sequence variation among the 32 copies of the mitochondrial COI fragment. We therefore focused our comparative analyses on the

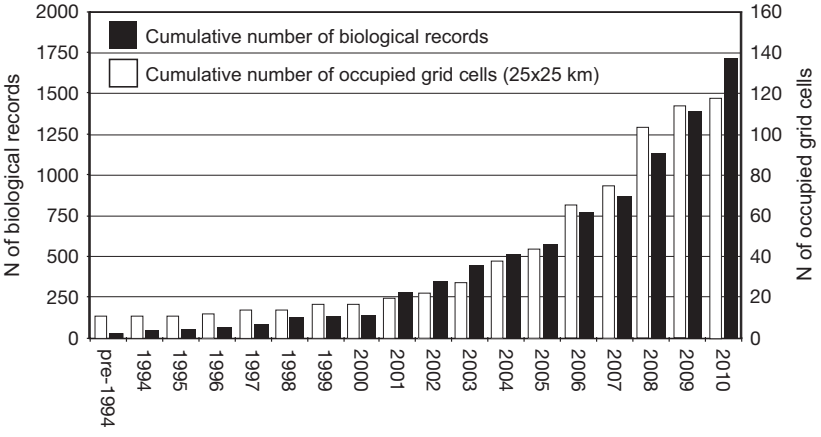

Fig. 3 Geographic range expansion of *Colletes hederae* in Western Europe between 1994 and 2010. The results illustrate a parallel increase in the number of records and the occupied grid-cells since the description of this species in 1993. All biological records were validated and compiled from field records accumulated by both the Bees, Wasps, and Ants Recording Society (BWARS) and the Observatoire des Abeilles (OA) (see methods section for details).

three other DNA sequence alignments, namely RNAp, WgL and CAD, which contained 17 (802 bp), 5 (659 bp) and 4 polymorphic sites (544 bp), respectively. All mutations observed in the coding regions were synonymous: 17 mutations for RNAp, 3 mutations for WgL and 2 mutations for CAD. For WgL and CAD, the remaining mutations, respectively, two and three mutations along with the gap detected for CAD, are located on the sequenced intron. Among the 118 specimens analysed, we identified ten diploid individuals: seven females from Lesvos in Greece and three males from three different UK populations (Cornwall, Dorset and Oxfordshire). Identification of diploidy was simply performed by the detection of ambiguous bases. However, as the probability to observe homozygous sequences at all three loci simultaneously is quite low (especially looking at the RNAp alignment), there is relatively little chance that additional unnoticed diploid males were present in our samples.

In the CAD alignment, diploid individuals were all homozygous, whereas for RNAp and WgL alignments, we found 7 (of 9) and 2 (of 10) heterozygous diploid individuals. The reconstruction of haplotypes from these heterozygous diploid individuals with the PHASE software was facilitated by comparisons with phased haplotypes produced from haploid individuals. All the haplotypes were reconstructed with estimated probabilities > 0.8, and three independent runs launched for each locus confirmed the robustness of our results. Eventually, 6 haplotypes were detected in the CAD and WgL alignments and 43 haplotypes, for only 17 polymorphic sites, in the RNAp alignment.

Although we cannot formally exclude the effects of multiple convergent mutations, recombination events seem to be responsible for the high number of loops observed in the RNAp haplotype network (Fig. 1b), as the PHI test (Bruen *et al.*, 2006) detected statistically significant evidence for recombination (P-value = 0.0002376). However, the RDP (Martin & Rybicki, 2000) or MaxChi (Maynard Smith, 1992) methods did not identify recombination sites within our data set of closely related sequences. We then used the method of Hudson & Kaplan (1985) to define sequence regions within which recombination events have not been detected. We used this partition to generate subnetworks based on regions free of detected recombination sites. A minimum of seven recombination events have been detected in the RNAp alignment, five between the 219th and 510th bp and one between the 672th and 798th bp. Taking this information into account, we managed to generate two haplotype networks without any loop (RNAp(1) and RNAp(2), Fig. 1c) which we used for comparing different expansion scenarios. These two RNAp subnetworks are based on a 221-bp fragment (the first 221 bp of the original alignment) and a 287-bp fragment (located between nucleotides 511 and 798 of the original alignment). RNAp(1) and RNAp(2) alignments both contain four polymorphic sites defining five and four haplotypes, respectively, and display similar levels of polymorphism as those detected for CAD and WgL. To avoid a bias linked to the high polymorphism caused by potential intra-allelic recombinations, these two subnetworks were used in simulation analyses instead of the overall RNAp fragment.

Genetic diversity and structure analyses

Our haplotype networks, along with the geographic distribution of each allele (Fig. 1b,c), show no obvious population genetic structure. However, our global Φ_{ST} statistics (calculated on the 22 populations) and their associated P-value computations did reveal a low but significant genetic structure (significant Φ_{ST} values between 0.086 and 0.284, Table 2). The AMOVA results (Table 2) revealed (i) low but similar estimated Φ_{SC} (measure of the proportion of variation among populations within groups) and Φ_{ST} (measure of the proportion of variation among populations) values and

Table 2 Global Φ_{ST} values among populations, AMOVA Φ-statistics for the partition defined by the three defined groups of populations (see the text), the number of analysed sequences (N_S) in each group and four different diversity measures estimated inside each of these three groups: the number of different haplotypes identified in the group divided by the number of sampled sequences in this group (N_H/N_S), the allelic richness (A_R) and the nucleotide diversity (π). RNAp(1) and (2) correspond to two identified RNAp gene subfragments without any recombination sites.

		AMOVA for the three defined groups of populations			Group n°1: Ancestral populations (orange cells in Fig. 2)				Group n°2: new colonized populations in France and Belgium				Group n°3: new colonized populations in southern England			
Locus	Global Φ_{ST}	Φ_{SC}	Φ_{ST}	Φ_{CT}	N_S	N_H/N_S	A_R	π	N_S	N_H/N_S	A_R	π	N_S	N_H/N_S	A_R	π
CAD	0.086*	0.059*	0.104*	0.048	81	0.012	4.77	0.0040	11	0.076	5.00	0.0039	11	0.046	3.00	0.0034
RNAp	0.150*	0.093*	0.103*	0.011	76	0.461	9.26	0.0053	11	0.546	6.00	0.0043	19	0.421	5.67	0.0050
RNAp(1)	0.096*	0.063	0.081*	0.020	76	0.066	3.77	0.0043	11	0.273	3.00	0.0023	19	0.105	2.00	0.0023
RNAp(2)	0.284*	0.182*	0.173*	−0.011	76	0.053	2.62	0.0023	11	0.182	2.00	0.0015	19	0.158	2.82	0.0019
WgL	0.216*	0.173*	0.243*	0.075	79	0.076	3.08	0.0012	11	0.364	4.00	0.0018	21	0.095	2.00	0.0007

*A significant value for the Φ-statistics (P-value < 0.05 for 10 000 permutations).

(ii) systematically nonsignificant estimated Φ_{CT} (measure of the proportion of variation among groups) values close to zero. The latter result demonstrates the absence of significant population structure specifically linked to geographic fragmentation between England and the continent. We found no clear pattern of variation in genetic diversity when comparing the different genetic diversity measures estimated within each of the three defined groups of sampled populations (Table 2). When comparing groups n°1 and n°3 (i.e. the area of origin vs. the most recent colonized area), estimated allelic richness A_R and nucleotide diversity π at the three analysed loci were systematically slightly lower in group n°3. On the other hand, the relative number of different haplotypes N_H/N_S was instead higher in group n°3, except for RNAp for which the two values are similar (0.461 for group n°1 against 0.421 for group n°3). Note that due to the small sample size of group n°2, comparisons between this group and the two others need to be interpreted cautiously.

Comparisons with simulated data sets

Principal component analyses of the simulation results allowed us to identify the summary statistics displaying more variability between scenarios (G_{ST}, N_{ST}, Φ_{ST}, X_H(G2), X_H(G3), π(G2), π(G3), π(G1)-π(G2), π(G1)-π(G3), π_R(G1), π_R(G2), π_R(G3); G1, G2 and G3 referring to groups n°1, n°2 and n°3) and those displaying more variability within scenarios ($N_H tot$, π(G1), π_{tot}, N_{ST}-G_{ST}) (Fig. S3 for *Ne* = 1000, Fig. S4 for *Ne* = 10 000 and Fig. S5 for *Ne* = 100 000). Based on this visual comparison, we selected several promising sets of summary statistics for comparing our four alternative scenarios. Several comparison analyses were also conducted using only one of the promising summary statistics. In the end, we identified a set of four statistics as the most discriminant for our model comparison: two statistics measuring the population genetic structure (G_{ST} and N_{ST}; Pons & Petit, 1996) and two statistics measuring the difference in genetic diversity between a newly colonized area and the area of origin π_R(G3) and the difference between N_H(G1)/N_S(G1) and N_H(G3)/N_S(G3). We chose populations from group n°1 (Fig. 2) to represent the area of origin and populations from group n°3 (southern England) to represent a newly colonized area, in agreement with our detailed biogeographical records. The difference between N_H(G1)/N_S(G1) and N_H(G3)/N_S(G3) is the difference between the relative number of different haplotypes identified in groups n°1 and n°3, whereas π_R(G3) is the nucleotide diversity in group n°3 relative to that of the rest of the distribution (group n°1 + group n°2, Mardulyn *et al.*, 2009).

Using these four summary statistics to compare the real data set with the simulated ones, the combined *P*-values over loci (Table 3; see Table S2 for details per locus) reveal that, when considering *Ne* = 1000, scenario n°2 (S2) is clearly more likely than the three alternative scenarios S1, S3 and S4, displaying an associated combined *P*-value close to one. In that case, the second most likely scenario is scenario n°1 (S1a), with a combined *P*-value close to 0.55. On the other hand, when considering *Ne* = 10 000 or *Ne* = 100 000, scenario n°1 (S1b and S1c) becomes the most likely, followed by scenario S2, which is even rejected for *Ne* = 100 000. For the three maximum effective sizes

Table 3 Consistency of the real genetic data set against simulated data sets using PHYLOGEOSIM software for four demographic scenarios (10 000 simulations per set of parameters). *P*-values give the probability that the real data set is consistent with the simulation model, on the basis of four summary statistics (G_{ST}, N_{ST}, π(G1)-π(G3), π_R(G3)) obtained on each of four nuclear genomic regions (CAD, RNAp(1), RNAp(2), WgL).

Scenario	*T*1	*Ne*	*Ne*(barriers)	*fm*1/*fm*2	t_R	*P*-value	Estimation of type II error
S1a	25	1000	100	0.01/0.005	1.5	0.551	0.758
S2a	25	1000	100	0.002/0.001	5	0.978*	–
S3a	25	1000	100	0.0004/0.0002	15	0.055	0.330
S4a	25	1000	100	0.0001/0.00005	25	0.000	0.015
S1b	25	10 000	1000	0.001/0.0005	3	0.972*	–
S2b	25	10 000	1000	0.0002/0.0001	10	0.298	0.641
S3b	25	10 000	1000	0.00004/0.00002	30	0.020	0.074
S4b	25	10 000	1000	0.00002/0.00001	50	0.007	0.027
S1c	25	100 000	10 000	0.0001/0.00005	6	0.427*	–
S2c	25	100 000	10 000	0.00002/0.00001	20	0.033	0.743
S3c	25	100 000	10 000	0.000004/0.000002	60	0.002	0.039
S4c	25	100 000	10 000	0.000002/0.000001	100	–	–

*T*1 is the number of generations since the beginning of expansion, *Ne* is the maximal effective populations sizes of yellow and orange cells (Fig. 2), *Ne*(barriers) is the maximal effective populations sizes of green cells (connection France – UK), *fm*1 is the short-distance forward migration rate and *fm*2 is the long-distance forward migration rate, t_R is the reproduction rate, and the estimation of type II error is the proportion of pods (pseudo-observed data sets generated under the considered scenario), which are not rejected (combined *P*-value > 0.05) when compared with data sets simulated under the most likely scenario (i.e. the scenario associated to the highest *P*-value marked by an asterisk and based on the same *Ne*). *P*-values obtained for each genomic region are detailed the Supporting information (Table S2).

tested, scenarios S3 and S4, characterized by lower migration rates and higher founder effects, are systematically rejected ($P < 0.05$), except for scenario S3a, which is only slightly above the rejection limit ($P = 0.055$). Note that scenario S4c (S4 for *Ne*=100 000) could not be simulated. Indeed, even when increasing the reproduction rate to an extremely high value ($t_R = 100$), the range expansion in the forward simulation never managed to reach the most remote sampled populations. Scenario S4c presents *a priori* the same level of founder effects (i.e. the same *Ne*fm* values), but its parameter values lead to a much slower migration wave than in scenarios S4a and S4b. Overall, these results indicate that an increase in cell effective size further decreases the likelihood of scenarios associated with lower migration rates. Yet, for all effective size values tested, scenarios 1 and 2 are always more likely than scenarios 3 and 4.

Estimation of type II error based on pseudo-observed data sets seems to validate our results further. With the exception of the pseudo-observed data sets simulated under scenario S3a and compared with data sets simulated under scenario S2a, all the type II errors of scenarios S3 and S4 were associated to small probabilities (< 0.05, except for scenario S3b that generated a probability of type II error of 0.074 when compared with data sets simulated under scenario S1b).

Discussion

Evidence for a rapid range expansion without a drastic founder effect

Witnessing and documenting a rapid geographic range expansion of a solitary bee species is a rare phenomenon. Here, our field surveys and observations provide strong evidence that *C. hederae* has undergone a rapid geographic expansion in Western Europe (Figs. 3, S1 and S2). Because colonies of this species are quite conspicuous and portions of its newly colonized range (e.g. southern England) have been extensively scanned by bee experts for several years before its appearance, we are confident that the inference of this range expansion is not an artefact associated to increased search efforts.

Because newly established populations were found to rapidly increase to hundreds, and occasionally to thousands of individuals as observed today in southern England, and given the phylopatric behaviour encountered in many groups of bees, we first hypothesized that *C. hederae* increased its range through the migration of a few individuals across relatively long distances (several hundreds of metres to several kilometres). These migrants would then have reached suitable habitats and would have rapidly established large populations as a result of their high reproduction rate (Bischoff *et al.*, 2005). Strong founder effects would be expected under this hypothesis. However, the absence of a drastic decrease in genetic diversity among populations sampled in the more recently colonized areas (Fig. S1) appears in conflict with this view. Our comparisons with data sets simulated using a spatially explicit coalescence model give more support to a demographic scenario involving relatively high migration rates. It shows that the increase in population size of a newly colonized cell to its maximum density is not only based on the reproduction of the few first individuals that have colonized it, but also based on continuous migration between neighbour cells on the grid. Overall, these results indicate that *C. hederae* is characterized by good dispersal capabilities. Although this may be surprising for such a pollen-specialist species, it is consistent with its recent rapid spread across Europe and with the relative abundance of its host plant in urban, suburban and rural gardens across Western Europe. The recent colonization of southern England by *C. hederae* since 2001 and the lack of a drastic decrease in genetic diversity of the UK populations compared with those sampled in continental Europe highlight the migration capabilities of this bee even further. Contrary to the almost simultaneous arrival of the violet carpenter bee *Xylocopa violacea* in the UK (Peat, 2007) or of the giant Asian resin bee *Megachile sculpturalis* in France (Vereecken & Barbier, 2009) where anthropogenic activities (in particular the import of timber) might have facilitated their establishment, the colonization of southern England and most parts of Western Europe by *C. hederae* during the past decade can reasonably be considered natural.

Among the fifteen males that we analysed from the UK, we identified three diploid individuals. The production of diploid males in Hymenoptera populations can be the result of a loss of allelic diversity at the sex-determining loci (sl-CSD; Zayed, 2009). High frequencies of diploid males in Hymenoptera populations are therefore sometimes interpreted as a sign of inbreeding (Crozier & Page, 1985; Pamilo *et al.*, 1994; Cook & Crozier, 1995; Roubik *et al.*, 1996; Zayed *et al.*, 2004; Zayed & Packer, 2005; but see Paxton *et al.*, 2000), but can also be associated to a population bottleneck or founder event (Ross *et al.*, 1993). The identification of diploid males in southern England could therefore be interpreted as a sign of local founder effects associated with the relatively small number of individuals that have crossed the channel and succeeded in founding new populations in this region. This is not necessarily incompatible with the fact that the overall genetic diversity estimated from the group of populations sampled in southern England is very similar to that of the continent. Indeed, it can be hypothesized that the colonization of England by a small number of individuals has occurred many times and simultaneously (or at least in a relatively short period of time) in different locations. However, although the sampling currently allows comparing the overall genetic diversity among our defined groups of populations (especially between groups n°1 and n°3), the local sampling (number of

sequences available for each sampled locality) remains too small to test further this hypothesis of multiple local founder events. A more thorough sampling of the newly colonized areas in southern England is needed to address this issue.

Drivers and facilitators of range expansion in *C. hederae*

To date, the factors that triggered the geographic range expansion of *C. hederae* have not been accurately pinpointed. Yet, we postulate that climate change, in particular the recurrent occurrence of hotter and longer summers ('heat waves') in Europe over the past two decades (e.g. Schär *et al.*, 2004; Fischer & Schär, 2009; Fischer *et al.*, 2012), might have contributed to the recent range expansion of this solitary bee species. Indeed, its principal forage plant, *H. helix*, is known to reproduce vegetatively in areas characterized by colder climates and that flowering is associated with areas of greater warmth (Iversen, 1944). Hence, an increase in summer temperatures might have contributed to the production of more abundant and longer-lasting flowering stands, which provided *C. hederae* females with a locally abundant source of pollen and nectar.

We hypothesize that several factors might have contributed to the demographic changes and the rapid geographic expansion of *C. hederae*. First, features of the host plant upon which females of *C. hederae* strongly rely for pollen, the flowering ivy (*H. helix*), presumably facilitated the colonization of novel habitats across Western Europe while minimizing competition for pollen resources with other bees. Indeed, *H. helix* is very widespread, locally abundant and produces large numbers of flowers in both urban and peri-urban habitats across Europe (Grivet & Petit, 2002). The distribution of *C. hederae* is much wider than that of e.g. *C. halophilus*, one of its sister species in the *C. succinctus* group: the latter species is strictly associated with salt marshes of the Atlantic coasts of France and the North Sea where females provision their brood cells primarily with pollen of *Aster tripolium* (Asteraceae), a more ephemeral and geographically restricted source of pollen in July–August for various other wild bee species (Kuhlmann *et al.*, 2007; Rooijakkers & Sommeijer, 2009; Sommeijer *et al.*, 2009, 2012). Moreover, the very late flowering period (August–November) of the flowering ivy makes it a very poorly exploited pollen source by most wild bee species. Second, the local abundance of the flowering ivy and its long flowering period (6–8 weeks), particularly when summers are warm and the early autumn is mild (Bischoff *et al.*, 2005), allow females of *C. hederae* to collect larger quantities of pollen over a longer period and to complete sometimes up to eighteen brood cells, whereas a completed nest of *C. halophilus* generally consists of five or six brood cells (O'Toole & Raw, 1991).

These factors might have been acting in concert, enabling a spectacular range expansion and leading to the recent burst of biological records over the past decade (2000-2010). We predict that further northward and eastward expansion of *C. hederae* in Europe will probably be limited by the ivy's inability to produce large flowering stands for long periods in areas where summers are wetter and where temperatures in early autumn are harsher (i.e. colder) (see also Roberts *et al.*, 2011). Another potentially important consequence of the recent and rapid geographic range expansion of *C. hederae* is that populations on the invasion front have apparently escaped from their specific natural enemies such as the cleptoparasitic solitary bee species *Epeolus fallax* (N.J. Vereecken, unpublished data) or the brood parasitic meloid beetle *Stenoria analis* (Coleoptera, Meloidae) (Vereecken & Mahé, 2007; Vereecken *et al.*, 2010) that are regularly found in southern regions of Europe. This expansion of *C. hederae* into areas devoid of specialist enemies may have contributed to the rapid development of dense populations of *C. hederae* in recently colonized habitats, for example, in the southern UK populations where we have observed the largest known populations of *C. hederae* encompassing tens of thousands of individuals and no specialist nest parasite in recent years. The absence of natural enemies might in turn have favoured the migration of individuals to neighbouring suitable habitats to minimize local frequency-dependent density and competition as observed in other groups of organisms (Hamilton & May, 1977; Clobert *et al.*, 2001; Matthysen, 2005; Ronce, 2007; Kim *et al.*, 2009; Nowicki & Vrabec, 2011).

Pollen specialization can be viewed *a priori* as an apparent obstacle to geographic range expansion as it constrains the ability of the migrating individuals to locate and exploit their specific host plants in novel habitats. Yet, our study shows that *C. hederae* has succeeded in expanding its geographic range in Western Europe in regions where populations of its host plant capable of supporting colonies of this solitary bee were already present. Similar cases have been reported from the USA where some adventive solitary bees native to Europe are also pollen specialists [*Chelostoma* on *Campanula* (Campanulaceae), *Hoplitis anthocopoides* on *Echium* (Boraginaceae), *Lithurgus chrysurus* on *Centaurea* (Asteraceae)], illustrating that although a generalist diet can facilitate the colonization of novel habitats (Giles & Ascher, 2006), pollen-specialist solitary bee species are also capable of invading new territories while keeping their floral preferences unchanged, particularly if their specific pollen hosts are locally abundant (Matteson *et al.*, 2008). In the specific case of *C. hederae*, our analyses show that this rapid range expansion was made possible by important levels of migration occurring between newly colonized territories and areas in which the species is already established and

demonstrates the strong dispersal capabilities of this specialist solitary bee. Because this range expansion was not associated with a drastic decrease in genetic diversity, the newly established populations probably maintain their evolutionary potential, for example, to cope with environmental variations (Soulé, 1987).

Largely different levels of polymorphism among loci

The polymorphism levels reported in this study were strikingly different among loci, with the COI locus displaying a single allele over the entire species range, the CAD and Wg loci displaying each 6 alleles and the RNAp locus displaying as many as 43 different alleles. Although this variation can be explained to some extent by the stochastic nature of the coalescent process, assuming all 3 loci are unlinked, by the occurrence of recombination events in the RNAp fragment and by the smaller effective size of the mitochondrial locus, natural selection cannot be ruled out as another possible cause. Indeed, evidence that natural selection can influence the pattern of variation in the mitochondrial genome was presented recently (Ballard & Whitlock, 2004; Bazin *et al.*, 2006). Adaptive selection could at least partially explain the absence of polymorphism found in the COI sequences.

A common assumption of phylogeographic methods is that the examined loci are neutral and that the observed pattern of variation is essentially created by demographic history. Without this crucial assumption, the interpretation of many phylogeographic data sets would be erroneous, because selection strongly impacts the shape of gene genealogies (e.g. Avise, 2000). In the case of this study, however, we believe this assumption is less important. Indeed, the investigated period of time spans only two decades, which is much too short to account for the shape of gene genealogies characterizing the studied loci; that is, these genealogies were generated over a much longer period of time. When comparing different colonization scenarios that have occurred in a 20-year period, the analyses rely essentially on the geographic distribution of the different alleles and their frequencies, not on the shape of genealogies. Even if a few alleles are currently subject to selection pressures, the majority of those are likely to be influenced only by dispersal and reproduction. Moreover, the weak signal of population structure found across the range of the species suggests that adaptive selection, if present, played only a minor role in shaping the geographic distribution of alleles. Its presence would actually result in underestimating migration in this species, and correcting for this effect would only strengthen our conclusion of high dispersal capabilities. Finally, the COI locus, for which suspicion of adaptive selection is highest, was excluded from the analyses, due to its lack of polymorphism.

The use of a spatially explicit model of coalescence

A spatially explicit model of coalescence appears essential to simulate the evolution of molecular data in a population or species going through a geographic range expansion, certainly for the goal of investigating the impact of this range expansion on the geographic distribution of genetic diversity. Such a spatially explicit model was defined here by simply superimposing a grid on the map of the species range, in which each cell is defined as a separate population within which a coalescent event can occur. Although this feature of the model makes it much more realistic, its demographic component is defined by a relatively small number of parameters. In addition to the size of the grid, we have only defined the following demographic parameters: initial and maximum effective population sizes per cell, short- and long-distance migration rates and a single reproduction rate. By combining different values of these parameters in an extensive simulation analysis, we have explored a large portion of the space of possible scenarios that could have resulted in the geographic expansion observed in the field. This exploration allowed us to identify the mode of colonization that was the most compatible with our sequence data, as well as to exclude scenarios that were clearly incompatible.

Although using an explicitly geographic model of coalescence offers a natural way to implement a range expansion, possibly associated to a pattern of isolation by distance, it necessarily leads to an increase in model complexity that results in slower computation simulations. However, when comparing demographic hypotheses, a large number of simulations are needed, to explore the space of all combinations of parameter values as thoroughly as possible. The longer computation time needed for this explicit geographic model prevented us from using an approximate Bayesian computation (ABC) framework (e.g. Beaumont *et al.*, 2002) that would have been better suited for exploring that space, despite recent advances decreasing the required number of simulations (e.g. Beaumont *et al.*, 2009; Wegmann *et al.*, 2009). Clearly, a compromise needs to be reached in population genetic studies between model complexity and accuracy of the method selected to compare evolutionary models. In the context of this study, a model sufficiently complex to explicitly describe the geographic expansion of the studied bee seemed unavoidable.

Despite recent works in the context of ABC methods (e.g. Joyce & Marjoram, 2008; Nunes & Balding, 2010; Blum *et al.*, 2012; Fearnhead & Prangle, 2012), the choice and number of summary statistics to use for the comparison of simulated and real data sets outside an ABC framework still remains a challenging issue for which there is a lack of theoretical knowledge. The use of the PCA here, as has already been suggested by others (e.g. by Veeramah *et al.*, 2012), to explore the value

of each summary statistic in discriminating among different scenarios of expansion, has been proved to be very efficient. Even if simply based on a visual inspection of the PCA graph, this approach allowed to identify summary statistics that vary mostly among scenarios rather than within scenarios. In addition, the calculation of a chi-square statistic combining all summary statistics into a single value greatly facilitated the comparison between the simulated and real data. Note that the estimated *P*-values reported remain closely associated to a given set of summary statistics. A complementary approach to cross-validate the results implemented in this study has therefore consisted in testing distinct sets of statistics that had been identified as promising by the PCA analysis. Furthermore, a validation procedure such as that performed in this study for evaluating the probability of a type II error helps to increase our confidence in the results.

Authors' contributions

SD, NJV and PM designed research; NJV and SPMR collected material; CH, SD and NJV performed laboratory analyses; SD, NJV, OJH and PM analysed data; SD, NJV, OJH, SPMR and PM wrote the paper.

Acknowledgments

We thank two anonymous reviewers for providing helpful comments on a previous version of this manuscript. We are grateful to J. D'Haeseleer, J.-Y. Cretin, E. Dufrêne, L. Guilbaud, V. Le Féon, F. Dittlo, P. Bourlet, P. Straub, J. Adams, R. Burger, M. Herrmann, C. Praz, G. Lemoine, K. Standfuss, C. Clee, M. Jenner, M. Smith and J. King for their invaluable help in collecting specimens. Computational resources have been provided by the High Performance Computing Centre co-funded by ULB and VUB (HPC cluster 'Hydra'). We wish to thank R. Leplae from the ULB/VUB Computing Centre for his precious help and J. Vleminckx for his useful advices on data analyses. This research project was funded by the Belgian FRS-FNRS through a FRFC project, a PhD grant to SD, as well as a postdoctoral grant and several travel grants to NJV (2008-2012).

References

Almeida, E.A.B. & Danforth, B.N. 2009. Phylogeny of colletid bees (Hymenoptera: Colletidae) inferred from four nuclear genes. *Mol. Phylogenet. Evol.* **50**: 290–309.

Arenas, M., Ray, N., Currat, M. & Excoffier, L. 2012. Consequences of range contractions and range shifts on molecular diversity. *Mol. Biol. Evol.* **29**: 207–218.

Avise, J.C. 2000. *Phylogeography: The history and formation of species Cambridge*. Harvard University Press, Massachusetts.

Ballard, J.W.O. & Whitlock, M.C. 2004. The incomplete natural history of mitochondria. *Mol. Ecol.* **13**: 729–744.

Bandelt, H.J., Forster, P. & Röhl, A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**: 37–48.

Bazin, E., Glemin, S. & Galtier, N. 2006. Population size does not influence mitochondrial genetic diversity in animals. *Science* **312**: 570–572.

Beaumont, M.A., Zhang, W. & Balding, D.J. 2002. Approximate Bayesian computation in population genetics. *Genetics* **162**: 2025–2035.

Beaumont, M.A., Cornuet, J.M., Marin, J.M. & Robert, C.P. 2009. Adaptive approximate Bayesian computation. *Biometrika* **96**: 983–990.

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

Biesmeijer, J.C., Roberts, S.P.M., Reemer, M., Ohlemüller, R., Edwards, M., Peeters, T. *et al.* 2006. Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science* **313**: 351–354.

Bischoff, I., Eckelt, E. & Kuhlmann, M. 2005. On the biology of the Ivy-bee *Colletes hederae* Schmidt & Westrich, 1993 (Hymenoptera, Apidae). *Bonn. Zool. Beitr.* **53**: 27–35.

Blum, M.G.B., Nunes, M.A., Prangle, D. & Sisson, S.A. 2012. A comparative review of dimension reduction methods in approximate Bayesian computation. *Stat. Sci.* **28**: 189–208.

Bowler, D.E. & Benton, T.G. 2005. Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. *Biol. Rev.* **80**: 205–225.

Bruen, T.C., Philippe, H. & Bryant, D. 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**: 2665–2681.

Burkle, L.A., Marlin, J.C. & Knight, T.M. 2013. Plant-pollinator interactions over 120 years: loss of species, co-occurrence and function. *Science* **339**: 1611–1615.

Camargo, J.M.F., Moure, J.S. & Roubik, D.W. 1988. *Melipona yucatanica* new species (Hymenoptera: Apidae: Meliponinae): stingless bee dispersal across the Caribbean arc and post-Eocene vicariance. *Pan-Pac. Entomol.* **64**: 147–157.

Cameron, S.A., Lozier, J.D., Strange, J.P., Koch, J.B., Cordes, N., Solter, L.F. *et al.* 2011. Patterns of widespread decline in North American bumble bees. *Proc. Natl. Acad. Sci. USA* **108**: 662–667.

Cane, J.H. 2003. Exotic nonsocial bees (Hymenoptera: Apiformes) in North America: ecological implications. In: *Crops, Whence Pollinators of the Future? (eds Strickler K and Cane JH)* (F. Non-native, ed), pp. 113–126. Thomas Say Publications in Entomology, Proceedings of the Entomological Society of America, Lanham, MD, USA.

Chessel, D., Dufour, A.B. & Thioulouse, J. 2004. The ade4 package-I- One-table methods. *R News* **4**: 5–10.

Clobert, J., Danchin, E., Dhondt, A.A. & Nichols, J.D. 2001. *Dispersal*. Oxford University Press, Oxford, UK.

Colla, S.R. & Packer, L. 2008. Evidence for decline in eastern North American bumblebees (Hymenoptera: Apidae), with special focus on *Bombus affinis* Cresson. *Biodivers. Conserv.* **17**: 1379–1391.

Cook, J.M. & Crozier, R.H. 1995. Sex determination and population biology in the Hymenoptera. *Trends Ecol. Evol.* **10**: 281–286.

Cross, I.C. 2002. *Colletes hederae* Schmidt & Westrich (Hym., Apidae) new to mainland Britain with notes on its ecology in Dorset. *Entomol. Mon. Mag.* **138**: 201–203.

Crozier, R.H. & Page, R.E. 1985. On being the right size: male contributions and multiple mating in social Hymenoptera. *Behav. Ecol. Sociobiol.* **18**: 105–115.

Currat, M., Ray, N. & Excoffier, L. 2004. SPLATCHE: a program to simulate genetic diversity taking into account environmental heterogeneity. *Mol. Ecol. Notes* **4**: 139–142.

Danforth, B.N., Fang, J. & Sipes, S. 2006. Analysis of family-level relationships in bees (Hymenoptera: Apiformes) using 28S and two previously unexplored nuclear genes: CAD and RNA polymerase II. *Mol. Phylogenet. Evol.* **39**: 358–372.

Dorchin, A., Filin, I., Izhaki, I. & Dafni, A. 2012. Movement patterns of solitary bees in a threatened fragmented habitat. *Apidologie* **44**: 90–99.

Dray, S., Dufour, A.B. & Chessel, D. 2007. The ade4 package - II: Two-table and K-table methods. *R News* **7**: 47–52.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with improved accuracy and speed. *Nucleic Acids Res.* **32**: 1792–1797.

El Mousadik, A. & Petit, R.J. 1996. Chloroplast DNA phylogeography of the argan tree of Morocco. *Mol. Ecol.* **5**: 547–555.

Excoffier, L., Smouse, P.E. & Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479–491.

Excoffier, L., Foll, M. & Petit, R.J. 2009. Genetic consequences of range expansions. *Annu. Rev. Ecol. Evol. Syst.* **40**: 481–501.

Fearnhead, P. & Prangle, D. 2012. Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. *J. Roy. Stat. Soc. B Met.* **74**: 419–474.

Ferriere, R., Belthoff, J.R., Olivieri, I. & Krackow, S. 2000. Evolving dispersal: where to go next? *Trends Ecol. Evol.* **15**: 5–7.

Fischer, E.M. & Schär, C. 2009. Future changes in daily summer temperature variability: driving processes and role for temperature extremes. *Clim. Dyn.* **33**: 917–935.

Fischer, E.M., Rajczak, J. & Schär, C. 2012. Changes in European summer temperature variability revisited. *Geophys. Res. Lett.* **39**: L19702.

Fisher, R.A. 1948. Questions and answers #14. *Am. Stat.* **2**: 30–31.

Fitzpatrick, U., Murray, T.E., Paxton, R.J., Breen, J., Cotton, D., Santorum, V. *et al.* 2007. Rarity and decline in bumblebees — a test of causes and correlates in the Irish fauna. *Biol. Conserv.* **136**: 185–194.

Frank, B.M., Piccolo, J.J. & Baret, P.V. 2011. A review of ecological models for brown trout: towards a new demogenetic model. *Ecol. Freshw. Fish* **20**: 167–198.

Frommer, U. 2010. Beobachtungen zum Ausbreitungsmodus der Efeu-Seidenbiene *Colletes hederae* Schmidt & Westrich, 1993 (Hymenoptera, Apidae) in Hessen und die Bedeutung des bluhenden Efeus (*Hedera helix* L.). *Hessische Faunistische Briefe* **29**: 1–20.

Fuller, S., Schwarz, M. & Tierney, S. 2005. Phylogenetics of the allodapine bee genus *Braunsapis*: historical biogeography and long-range dispersal over water. *J. Biogeogr.* **32**: 2135–2144.

Gallai, N., Vaissiere, B.E., Potts, S.G. & Salles, J. 2011. Assessing the monetary value of global crop pollination services. In: *Natural Capital: Theory and practice of mapping ecosystem services* (P. Kareiva, H. Tallis, T.H. Ricketts, G.C. Daily, S. Polasky, eds), Oxford University Press, Oxford, UK.

Garibaldi, L.A., Steffan-Dewenter, I., Winfree, R., Aizen, M.A., Bommarco, R., Cunningham, S.A. *et al.* 2013. Wild pollinators enhance fruit set of crops regardless of honey bee abundance. *Science* **339**: 1608–1611.

Gathmann, A. & Tscharnke, T. 2002. Foraging range of solitary bees. *J. Anim. Ecol.* **71**: 757–764.

Gibbs, J. & Sheffield, C.S. 2009. Rapid range expansion of the wool-carder bee, *Anthidium manicatum* (Linnaeus) (Hymenoptera: Megachilidae), in North America. *J. Kansas Entomol. Soc.* **82**: 21–29.

Giles, V. & Ascher, J.S. 2006. A survey of the bees of the Black Rock Forest preserve, New York (Hymenoptera: Apoidea). *J. Hymenopt. Res.* **15**: 208–231.

Goulson, D., Lye, G.C. & Darvill, B. 2008. Decline and conservation of bumble bees. *Annu. Rev. Entomol.* **53**: 191–208.

Goulson, D., Kaden, J.C., Lepais, O., Lye, G.C. & Darvill, B. 2011. Population structure, dispersal and colonization history of the garden bumblebee *Bombus hortorum* in the Western Isles of Scotland. *Conserv. Genet.* **12**: 867–879.

Grivet, D. & Petit, R.J. 2002. Phylogeography of the common ivy (*Hedera* sp.) in Europe: genetic differentiation through space and time. *Mol. Ecol.* **11**: 1351–1362.

Hamilton, W.D. & May, R.M. 1977. Dispersal in stable habitats. *Nature* **269**: 578–581.

Herrmann, M. 2007. Ausbreitungswelle der Efeu-Seidenbiene (*Colletes hederae*) in Baden-Württemberg (Hymenoptera, Apidae) und die Erschließung eines ungewöhnliches Nisthabitates. *Mitteilungen des entomologischen Vereins Stuttgart* **42**: 96–98.

Hinojosa-Díaz, I.A., Yáñez-Ordóñez, O., Chen, G., Peterson, A.T. & Engel, M.S. 2005. The North American invasion of the giant resin bee (Hymenoptera: Megachilidae). *J. Hymenopt. Res.* **14**: 69–77.

Hudson, R.R. & Kaplan, N.L. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**: 147–164.

Huson, D.H. & Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**: 254–267.

Iversen, J. 1944. *Viscum*, *Hedera* and *Ilex* as climatic indicators. A contribution to the study of past-glacial temperature climate. *Geol. Fören. Stock. För.* **66**: 463–483.

Joyce, P. & Marjoram, P. 2008. Approximately sufficient statistics and Bayesian computation. *Stat. Appl. Genet. Mol.* **7**, Article 26.

Kim, S.-Y., Torres, R. & Drummond, H. 2009. Simultaneous positive and negative density-dependent dispersal in a colonial bird species. *Ecology* **90**: 230–239.

Kraus, F.B., Wolf, S. & Moritz, R.F.A. 2009. Male flight distance and population substructure in the bumblebee *Bombus terrestris*. *J. Anim. Ecol.* **78**: 247–252.

Kuhlmann, M., Else, G.R., Dawson, A. & Quicke, D.L.J. 2007. Molecular, biogeographical and phenological evidence for the existence of three Western European sibling species in the *Colletes succinctus* group (Hymenoptera: Apidae). *Org. Divers. Evol.* **7**: 155–165.

Kuhlmann, M., Almeida, E.A.B., Laurenne, N. & Quicke, D.L.J. 2009. Molecular phylogeny and historical biogeography of the bee genus *Colletes* Latreille, 1802 (Hymenoptera: Apiformes: Colletidae), based on mitochondrial COI and nuclear 28S sequence data. *Insect Syst. Evol.* **40**: 291–318.

Lepais, O., Darvill, B.E.N., O'Connor, S., Osborne, J.L., Sanderson, R.A., Cussans, J. *et al.* 2010. Estimation of bumblebee queen dispersal distances using sibship reconstruction method. *Mol. Ecol.* **19**: 819–831.

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

Mardulyn, P., Mikhailov, Y. & Pasteels, J.-M. 2009. Testing phylogeographic hypotheses in a Euro-Siberian cold-adapted leaf beetle with coalescent simulations. *Evolution* **63**: 2717–2729.

Martin, D. & Rybicki, E. 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**: 562–563.

Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D. & Lefeuvre, P. 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**: 2462–2463.

Matteson, K.C., Ascher, J.S. & Langellotto, G.A. 2008. Bee richness and abundance in New York city urban gardens. *Ann. Entomol. Soc. Am.* **101**: 140–150.

Matthysen, E. 2005. Density-dependent dispersal in birds and mammals. *Ecography* **28**: 403–416.

Maynard Smith, J. 1992. Analyzing the mosaic structure of genes. *J. Mol. Evol.* **34**: 126–129.

Michener, C.D. 1974. *The Social Behavior of the Bees*. Harvard University Press, Cambridge.

Michener, C.D. 1979. Biogeography of the bees. *Ann. Mo. Bot. Gard.* **66**: 277–347.

Michener, C.D. 2007. *The Bees of the World*. 2nd edn. Johns Hopkins University Press, Baltimore.

Müller, A. & Kuhlmann, M. 2008. Pollen hosts of western palaearctic bees of the genus *Colletes* (Hymenoptera: Colletidae): the Asteraceae paradox. *Biol. J. Linn. Soc.* **95**: 719–733.

Nei, M. & Li, W.H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* **76**: 5269–5273.

Nowicki, P. & Vrabec, V. 2011. Evidence for positive density-dependent emigration in butterfly metapopulations. *Oecologia* **167**: 657–665.

Nunes, M.A. & Balding, D.J. 2010. On optimal selection of summary statistics for approximate bayesian computation. *Stat. Appl. Genet. Mol. Biol.* **9**, Article 34.

O'Toole, C. & Raw, A. 1991. *Bees of the World*. Facts on File, New York & Oxford.

Pamilo, P., Sundstrom, L., Fortelius, W. & Rosengren, R. 1994. Diploid males and colony-level selection in Formica ants. *Ethol. Ecol. Evol.* **6**: 221–235.

Patiny, S., Rasmont, P. & Michez, D. 2009. A survey and review of the status of wild bees in the West-Palaearctic region. *Apidologie* **40**: 313–331.

Paxton, R.J., Thorén, P.A., Gyllenstrand, N. & Tengö, J. 2000. Microsatellite DNA analysis reveals low diploid male production in a communal bee with inbreeding. *Biol. J. Linn. Soc.* **69**: 483–502.

Peat, L. 2007. Xylocopa violacea excitement. *BWARS Newsletter (Autumn)* **3**: 26–27.

Phillips, B.L., Brown, G.P. & Shine, R. 2010. Life-history evolution in range-shifting populations. *Ecology* **91**: 1617–1627.

Pons, O. & Petit, R.J. 1995. Estimation, variance and optimal sampling of genetic diversity. I. Haploid locus. *Theor. Appl. Genet.* **90**: 462–470.

Pons, O. & Petit, R.J. 1996. Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* **144**: 1237–1245.

Potts, S.G., Biesmeijer, J.C., Kremen, C., Neumann, P., Schweiger, O. & Kunin, W.E. 2010. Global pollinator declines: trends, impacts and drivers. *Trends Ecol. Evol.* **25**: 345–353.

Ray, N. & Excoffier, L. 2009. Inferring past demography using spatially explicit population genetic models. *Hum. Biol.* **81**: 141–157.

Ray, N., Currat, M. & Excoffier, L. 2003. Intra-deme molecular diversity in spatially expanding populations. *Mol. Biol. Evol.* **20**: 76–86.

Ray, N., Currat, M., Foll, M. & Excoffier, L. 2010. SPLATCHE2: a spatially explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics* **26**: 2993–2994.

Roberts, S.P.M., Potts, S.G., Biesmeijer, J.C., Kuhlmann, M., Kunin, W.E. & Ohlemüller, R. 2011. Assessing continental-scale risks for generalist and specialist pollinating bee species under climate change. *BioRisk* **6**: 1–18.

Ronce, O. 2007. How does it feel to be like a rolling stone? Ten questions about dispersal evolution. *Annu. Rev. Ecol. Evol. Syst.* **38**: 231–253.

Rooijakkers, E.F. & Sommeijer, M.J. 2009. Gender-specific brood cells in the solitary bee *Colletes halophilus* (Hymenoptera; Colletidae). *J. Insect Behav.* **22**: 492–500.

Ross, K.G., Vargo, E.L., Keller, L. & Trager, J.C. 1993. Effect of a founder event on variation in the genetic sex-determining system of the fire ant Solenopsis invicta. *Genetics* **135**: 843–854.

Roubik, D.W., Weigt, L.A. & Bonilla, M.A. 1996. Population genetics, diploid males, and limits to social evolution of euglossine bees. *Evolution* **50**: 931–935.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**: 1219–1228.

Schär, C., Vidale, P.L., Lüthi, D., Frei, C., Häberli, C., Liniger, M.A. *et al.* 2004. The role of increasing temperature variability in European summer heatwaves. *Nature* **427**: 332–336.

Schmidt, K. & Westrich, P. 1993. *Colletes hederae* n.sp., eine bisher unerkannte, auf Efeu (*Hedera*) spezialisierte Bienenart (Hymenoptera: Apoidea). *Entomol. Z.* **103**: 89–93.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. & Floors, P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* **87**: 651–701.

Sommeijer, M.J., Rooijakkers, E.F., Jcobusse, C. & Kerkvliet, J.D. 2009. Larval food composition and food plants of the solitary bee *Colletes halophilus* (Hymenoptera: Colletidae). *J. Apicult. Res.* **48**: 149–155.

Sommeijer, M.J., Neve, J. & Jacobusse, C. 2012. The typical development cycle of the solitary bee *Colletes halophilus*. *Entomologische Berichten* **72**: 52–58.

Soulé, M.E. 1987. *Viable Populations for Conservation*. Cambridge University Press, Cambridge, UK.

Steffan-Dewenter, I., Potts, S.G. & Packer, L. 2005. Pollinator diversity and crop pollination services are at risk. *Trends Ecol. Evol.* **20**: 651–652.

Stephens, M. & Donnelly, P. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* **73**: 1162–1169.

Stephens, M., Smith, N.J. & Donnelly, P. 2001. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**: 978–989.

Strange, J.P., Koch, J.B., Gonzalez, V.H., Nemelka, L. & Griswold, T. 2011. Global invasion by *Anthidium manicatum* (Linnaeus) (Hymenoptera: Megachilidae): assessing potential distribution in North America and beyond. *Biol. Invasions* **13**: 2115–2133.

Veeramah, K.R., Wegmann, D., Woerner, A., Mendez, F.L., Watkins, J.C., Destro-Bisol, G. *et al.* 2012. An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. *Mol. Biol. Evol.* **29**: 617–630.

Vereecken, N.J. & Barbier, E. 2009. Premières données sur la présence de l'abeille asiatique *Megachile* (*Callomegachile*) *sculpturalis* Smith (Hymenoptera, Megachilidae) en Europe. *OSMIA* **3**: 4–6.

Vereecken, N.J. & Mahé, G. 2007. Larval aggregations of the blister beetle *Stenoria analis* (Schaum) (Coleoptera: Meloidae) sexually deceive patrolling males of their host, the solitary bee *Colletes hederae* Schmidt & Westrich (Hymenoptera: Colletidae). *Ann. Soc. Entomol. Fr.* **43**: 493–496.

Vereecken, N.J., Schwenninger, H., Gogala, A. & Roberts, S.P.M. 2009. Mise à jour de la distribution géographique de l'abeille du lierre, *Colletes hederae* Schmidt & Westrich (Hymenoptera, Colletidae) en Europe. *OSMIA* **3**: 2–3.

.Vereecken, N.J., Dufrêne, P., Lückmann, J., Devalez, J., Raemakers, I. & Bologna, M. 2010. Synthèse des données biogéographiques de *Stenoria analis* (Schaum) (Coleoptera, Meloidae) en France et dans les régions voisines. *OSMIA* **4**: 1–4.

Wallner, W.E. 1987. Factors affecting insect population dynamics: differences between outbreak and non-outbreak species. *Annu. Rev. Entomol.* **32**: 317–340.

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

Westrich, P. 1990. *Die Wildbienen Baden-Wurttembergs*. E. Ulmer, Stuttgart, Germany.

Westrich, P. 2008. Flexibles Pollensammelverhalten der ansonsten streng oligolektischen Seidenbiene *Colletes hederae* Schmidt & Westrich (Hymenoptera: Apidae). *Eucera* **1**: 17–29.

Williams, P. 2005. Does specialization explain rarity and decline among British bumblebees? A response to Goulson *et al*. *Biol. Conserv.* **122**: 33–43.

Zayed, A. 2009. Bee genetics and conservation. *Apidologie* **40**: 237–262.

Zayed, A. & Packer, L. 2005. Complementary sex determination substantially increases extinction proneness of haplodiploid populations. *Proc. Natl. Acad. Sci. USA* **102**: 10742–10746.

Zayed, A., Roubik, D.W. & Packer, L. 2004. Use of diploid male frequency data as an indicator of pollinator decline. *P. Roy. Soc. Lond. B-Bio.* **271**: S9–S12.

Zayed, A., Constantin, S.A. & Packer, L. 2007. Successful biological invasion despite a severe genetic load. *PLoS ONE* **2**: e868.

Zurbuchen, A., Landert, L., Klaiber, J., Müller, A. & Dorn, S. 2010. Maximum foraging ranges in solitary bees: only few individuals have the capability to cover long foraging distances. *Biol. Conserv.* **143**: 669–676.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Figure S1 Evolution of the distribution of records for *Colletes hederae* in Europe from the time of its description (-1994), through 2002, 2006 and 2010.

Figure S2 Biannual changes in the distribution of *Colletes hederae* in the United Kingdom since the confirmation of its presence in 2001 along the Dorset coast (see black arrow indicating its entry point in the UK).

Figure S3 Principal component analyses (PCA) of 24 summary statistics per locus computed on 1000 simulated data sets for each of four demographic scenarios with a cell effective population size of *Ne* = 1000.

Figure S4 Principal component analyses (PCA) of 24 summary statistics per locus computed on 1000 simulated data sets for each of four demographic scenarios with a cell effective population size of *Ne* = 10 000.

Figure S5 Principal component analyses (PCA) of 24 summary statistics per locus computed on 1000 simulated data sets for each of four demographic scenarios with a cell effective population size of *Ne* = 100 000.

Table S1 list, brief description and reference of the different summary statistics computed on each simulated dataset.

Table S2 Consistency of the real genetic data set against simulated data sets using PHYLOGEOSIM software for four demographic scenarios (10 000 stimulations per set of parameters).

Received 30 April 2013; accepted 14 October 2013