

RESEARCH ARTICLE

Open Access

# Scent of a break-up: phylogeography and reproductive trait divergences in the red-tailed bumblebee (*Bombus lapidarius*)

Thomas Lecocq<sup>1\*</sup>, Simon Dellicour<sup>2</sup>, Denis Michez<sup>1</sup>, Patrick Lhomme<sup>1</sup>, Maryse Vanderplanck<sup>1</sup>, Irena Valterová<sup>3</sup>, Jean-Yves Rasplus<sup>4</sup> and Pierre Rasmont<sup>1</sup>

## Abstract

**Background:** The Pleistocene climatic oscillations are considered as a major driving force of intraspecific divergence and speciation. During Ice Ages, populations isolated in allopatric glacial refugia can experience differentiation in reproductive traits through divergence in selection regimes. This phenomenon may lead to reproductive isolation and dramatically accentuates the consequences of the climatic oscillations on species. Alternatively, when reproductive isolation is incomplete and populations are expanding again, further mating between the formerly isolated populations can result in the formation of a hybrid zone, genetic introgression or reinforcement speciation through reproductive trait displacements. Therefore changes in reproductive traits driven by population movements during climatic oscillations can act as an important force in promoting pre-zygotic isolation. Notwithstanding, divergence of reproductive traits has not been approached in the context of climatic oscillations. Here we investigate the impact of population movements driven by climatic oscillations on a reproductive trait of a bumblebee species (*Bombus lapidarius*). We characterise the pattern of variation and differentiation across the species distribution (i) with five genes (nuclear and mitochondrial), and (ii) in the chemical composition of male marking secretions (MMS), a key trait for mate attraction in bumblebees.

**Results:** Our results provide evidence that populations have experienced a genetic allopatric differentiation, in at least three main refugia (the Balkans, Centre-Eastern Europe, and Southern Italy) during Quaternary glaciations. The comparative chemical analyses show that populations from the Southern Italian refugium have experienced MMS differentiation and an incipient speciation process from another refugium. The meeting of Southern Italian populations with other populations as a result of range expansion at a secondary contact zone seems to have led to a reinforcement process on local MMS patterns.

**Conclusions:** This study suggests that population movement during Quaternary climatic oscillations can lead to divergence in reproductive traits by allopatric differentiation during Ice Ages and by reinforcement during post-glacial recolonization.

**Keywords:** Phylogeography, Reproductive traits, Genetic differentiation, Bumblebees

\* Correspondence: thomas.lecoq@umons.ac.be

<sup>1</sup>Laboratoire de Zoologie (Research Institute of Biosciences), University of Mons, Place du Parc 20, B-7000 Mons, Belgium

Full list of author information is available at the end of the article

## Background

Historic events such as climatic and topographic changes have triggered current patterns of biodiversity [1,2]. The Pleistocene glaciations, in particular, are considered as a major driving force of intraspecific divergence, diversification and speciation (e.g. [3-5]). The Pleistocene epoch experienced important climatic oscillations [6] that have deeply modified the geographic range of temperate species within the Palaearctic region (e.g. [7]). During the Ice Ages, changes in species distribution occurred as the ice sheets advanced to reach climatically more favourable regions, called refugia [8,9]. At the end of each Ice Age, some populations expanded from these refugia to recolonize previously glaciated areas [5,10].

Geographic isolation between conspecific populations (e.g. between refugia) is recognised to play an important role in allopatric speciation (e.g. [11,12]). When gene flow has been limited, or absent, among populations, different or identical alleles can become fixed depending upon the forces of natural selection and/or genetic drift [13,14]. Thus genetic divergence occurs over a long period, imparting reproductive isolation and possibly leading to speciation [11]. When reproductive isolation is incomplete and formerly isolated lineages meet as a result of range expansion, further mating between the populations produces hybrids, resulting in the formation of a hybrid zone (e.g. [15,16]), genetic introgression (e.g. [17]) or reinforcement speciation through pre-zygotic isolation mechanisms (e.g. [18-20]).

Sexual selection is well recognized for its important role in promoting speciation through its variation among allopatric populations [20-23], either through the local features of a mating signal [24], or different signals in different environments that confer local adaptations [20,25]. Divergence in mating preferences that parallel divergence in mating signals give rise to discrimination against foreign mates, and can ultimately result in reproductive isolation and speciation ([24-27] but see also [28]). This phenomenon may dramatically increase the consequences of climatic oscillations on species. Notwithstanding, the divergence in reproductive traits has not been approached in the context of climatic oscillations and migration of populations. Previous studies have mainly focused on the consequences of climatic oscillations on genetic traits (e.g. [10,15,16]) while the consequences on reproductive traits and comprehensive integrative analyses of reproductive traits, geographical variation, and genetic differentiation are lacking.

Here we investigated the phylogeography, population structure, and geographical variation of a reproductive trait (male marking secretion, MMS) in the red-tailed bumblebee, *Bombus lapidarius*. Bumblebees are an excellent system for exploring historical biogeographic patterns

of Holarctic groups and the evolution of the courtship signal (e.g. [23,29]). These bees live in the coldest areas inhabited by insects and have been able to recolonize areas depopulated by Ice Ages in the last three million years [29]. Their early historical biogeography (from Eocene-Oligocene to Miocene) has already been explored [29] while their recent history (from Pliocene to Holocene) has received comparatively far less attention (e.g. [3]). *B. lapidarius* is a widespread temperate West-Palaearctic bumblebee that includes several subspecies which exhibit different colour patterns (Figure 1A, [30]): (i) *B. lapidarius lapidarius* in the European plains, Balkans and West Anatolia, (ii) *B. lapidarius decipiens* in the Iberian and South Italian peninsulas, (iii) *B. lapidarius caucasicus* and *B. lapidarius eriophorus* in N.E. Anatolia and the Caucasus, and (iv) *B. lapidarius atlanticus* in the Moroccan Atlas.

The courtship signal of bumblebee males includes both behavioural and chemical features (see [31] for a review). Here, we focus on the most commonly studied trait, the male marking secretions (MMS). Most bumblebee males patrol along paths where they scent-mark objects with their MMS that attract conspecific virgin females [32]. The MMS consist of a complex mixture (mainly aliphatic compounds) with several major components and with intraspecific variation (e.g. [33,34]). MMS are produced *de novo* by cephalic labial glands [35] from saturated fatty-acids by the action of species-specific esterases, desaturases and reductases in cephalic labial glands [35,36].

In this study, we used a phylogeographic approach based on five genes (nuclear and mitochondrial) along with comparative chemical analyses of MMS to determine the consequences of historic events on the chemical composition of male courtship signals. Our purposes were (i) to infer the phylogeographic pattern of this widespread European pollinator, (ii) to detect a putative divergence of MMS across its distribution range, and (iii) to compare the phylogeographic pattern and the reproductive trait variation pattern.

## Results

### Genetic divergence and phylogeography of the red-tailed bumblebee

#### Patterns of sequence variation

The final molecular dataset spanned 3933 aligned nucleotides: 1056 bp from *COI* (184 parsimony informative sites [PIS]), 465 bp from *Cytb* (37 PIS), 791 bp from *EF-1 $\alpha$*  F2 copy containing a ~200 bp intron (9 PIS), 711 bp from *LWRh* (2 PIS), and 910 bp from *PEPCK* (11 PIS). Alignments did not require additional indels. *LWRh* was monomorphic across all *B. lapidarius* populations while other loci were polymorphic (Figure 1C). Therefore, we excluded the *LWRh* from the following

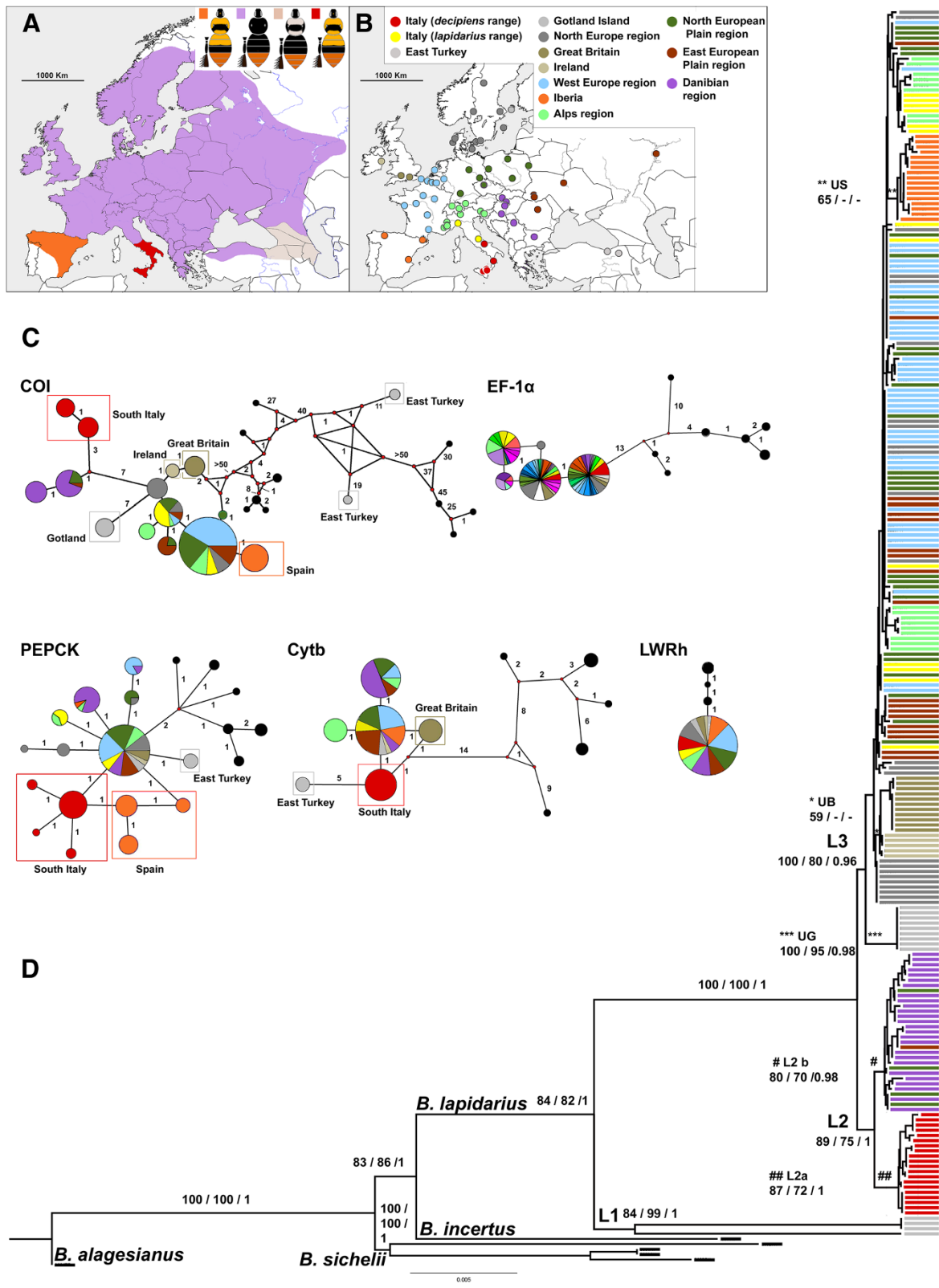


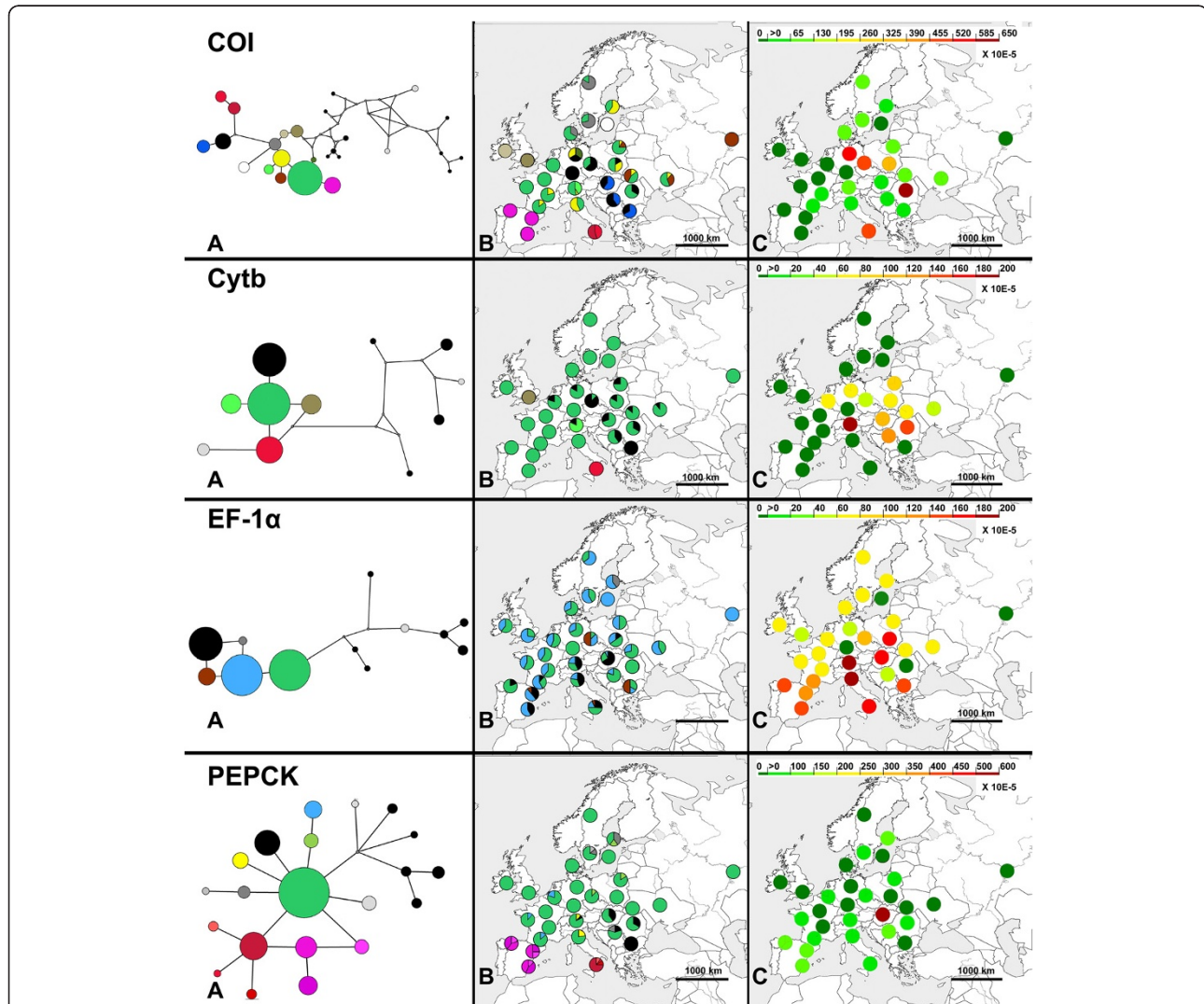
Figure 1 (See legend on next page.)

(See figure on previous page.)

**Figure 1 Distribution and phylogenetic relationship of *B. lapidarius*.** **A:** Colour pattern distribution, all geographic zones are the *B. lapidarius* geographic area; **B:** Map of sampling, colour of points refers to geographic zones. **C:** Median-joining network of haplotypes based on *COI*, *Cytb*, *EF1α*, *PEPCK* and *LWRh*. Circle sizes are proportional to frequencies of haplotypes. Colours of haplotypes refer to geographic area (Figure 1B). Red points on lines, undetected/extinct intermediate haplotype states. Numbers on lines represent the number of mutation(s) between two close haplotypes. Black circles are outgroups. **D:** Neighbour-joining tree based the combined molecular data matrix (*COI*, *Cytb*, *EF1α*, *PEPCK* and *LWRh*), value above branches are Neighbour-joining bootstrap values (only values > 50 are shown)/ maximum likelihood bootstrap values (only values > 70 are shown)/Bayesian posterior probabilities (only values > 0.95 are shown). Colour labels of branches refer to geographic area (Figure 1B).

results. In total, 16 unique haplotypes were observed for *COI*, 7 for *Cytb*, 6 for *EF-1α*, and 16 for *PEPCK* among 244 European and East Turkish individuals (Figure 1C). For geographic groups, haplotype diversity ranged from 0 to 0.8 and nucleotide diversity ranged from 0 to 0.006 (Additional file 1, Figure 2). Our results

displayed endemic haplotypes or haplogroups from (Figures 1C, 2): (i) Gotland Island, Great Britain, Ireland, Southern Italy, East Turkey and Spain for *COI*; (ii) Great Britain, East Turkey, and Southern Italy for *Cytb*; (iii) Southern Italy, East Turkey, and Spain for *PEPCK*.



**Figure 2 Geographic distribution of genetic diversities in *B. lapidarius*.** In the left column median-joining network of haplotypes based on *COI*, *Cytb*, *EF1α*, and *PEPCK* are coloured by haplotypes. In the centre column maps of haplotype frequency for each geographic groups based on *COI*, *Cytb*, *EF1α*, and *PEPCK* have the colours of pie charts as for haplotype colour (left column). In the right column maps of nucleotide diversity for each geographic group based on *COI*, *Cytb*, *EF1α*, and *PEPCK*, Colours of pie charts refer to coloured intensity scale.

### Phylogeny and haplotype relationships

All phylogenetic analyses and MJM networks performed on the same genetic dataset led to quite similar topologies (Figure 1C, see also supplementary trees [TreeBASE: ID S14299]). Phylogenetic analyses based on *EF-1 $\alpha$*  did not show genetic divergence between geographic groups and failed to support strongly the relationship with outgroups. In the same way, despite the distinctive haplotypes of some geographic groups, phylogeny based on *PEPCK* failed to resolve relationships between haplogroups while trees based on *COI* and *Cytb* resolved the relationship between haplogroups. Phylogenetic analyses based on nuclear genes (*EF-1 $\alpha$*  + *PEPCK*) failed to resolve the relationship between genetic groups while analyses based on mitochondrial genes (*COI* + *Cytb*) and all genes combined (*COI* dominated the phylogenetic signal) resolved the relationship between genetic groups. Phylogeny based on all genes combined showed a large-scale structuring in main groups that have a predominantly non-overlapping geographic distribution (Figure 1D). *B. lapidarius* is split into three major lineages: an East Turkey lineage (L1), a Southern Italy and SE European lineage (L2), and a large group including all other European populations (L3) (Figure 1D). The lineage L2 included two sub-lineages: a Southern Italian lineage (L2a) and a SE European lineage (L2b). Inside L3, three distinct geographic groups were distinguished: a Gotland group (UG), a British Isles group (UB), and a Spanish group (US) but they were poorly supported. However there is evidence for divergence between islands, peninsulas and other populations and for maintenance of distinct genetic structure over time. Due to the large genetic divergence of the East Turkey lineage (L1), its very restricted geographic distribution, and the lack of MMS data, we chose to exclude these samples from our analyses of population structure.

### Divergence times

Based on the *COI* data, the estimated divergence between East Turkish (L1) and European populations (L2 + L3) was within a range of 0.49 - 9.56 mya, with a mean of 2.81 mya and median of 1.32 mya (ESS: 222.99) while estimated divergence between European haplogroups (L2 versus L3) was within a range of 0.03-0.96 mya, with a mean of 0.28 mya and median of 0.13 mya (ESS: 231.21). Correspondingly, average simple pairwise divergence (i) between L1 and L2 + L3 was 9.68% or ~ 4.9 mya based on the 2% divergence per million years clock commonly applied to insect mitochondrial DNA [37], and (ii) between L2 and L3 was 1.01% or ~ 0.51 mya.

### Population structure

Geographic structuring among *COI*, *Cytb* and *PEPCK* haplotypes within European populations was significantly

supported by AMOVA results (global  $\Phi_{ST}$  statistic = 0.87 (*COI*), 0.79 (*Cytb*), 0.61 (*PEPCK*), p-value < 0.01). These results suggested a potentially low level of gene flow between several adjacent populations. *EF-1 $\alpha$*  and *LWRh* datasets were discarded from SAMOVA due to their poor genetic structuring (global  $\Phi_{ST}$  statistic = 0.35, p-value < 0.01 for *EF-1 $\alpha$*  and p-value > 0.01 for *LWRh*). In SAMOVA, the  $\Phi_{CT}$  value increased asymptotically with increasing number of groups, levelling out at: seven groups for *COI* ( $\Phi_{CT}$  = 0.87,  $\Phi_{ST}$  = 0.93,  $\Phi_{SC}$  = 0.60) with local  $\Phi_{CT}$  maximum at five groups ( $\Phi_{CT}$  = 0.85,  $\Phi_{ST}$  = 0.94,  $\Phi_{SC}$  = 0.57) and three groups ( $\Phi_{CT}$  = 0.81,  $\Phi_{ST}$  = 0.94,  $\Phi_{SC}$  = 0.66), five groups for *Cytb* ( $\Phi_{CT}$  = 0.89,  $\Phi_{ST}$  = 0.90,  $\Phi_{SC}$  = 0.06), nine groups for *PEPCK* ( $\Phi_{CT}$  = 0.82,  $\Phi_{ST}$  = 0.80,  $\Phi_{SC}$  = -0.09) but above five groups ( $\Phi_{CT}$  = 0.78,  $\Phi_{ST}$  = 0.81,  $\Phi_{SC}$  = 0.13), additional groups represented single sampling sites making little biological sense. All SAMOVA clustering (based on *COI*, *Cytb* and *PEPCK*) split the Southern Italian populations and most of the SE European populations from other populations while only *PEPCK* clustering split Spanish populations from other populations (Additional file 2). Clustering based on mitochondrial genes split some insular populations (Additional file 2). These SAMOVA results were quite congruent to geographic groups observed in MJN and phylogenetic results.

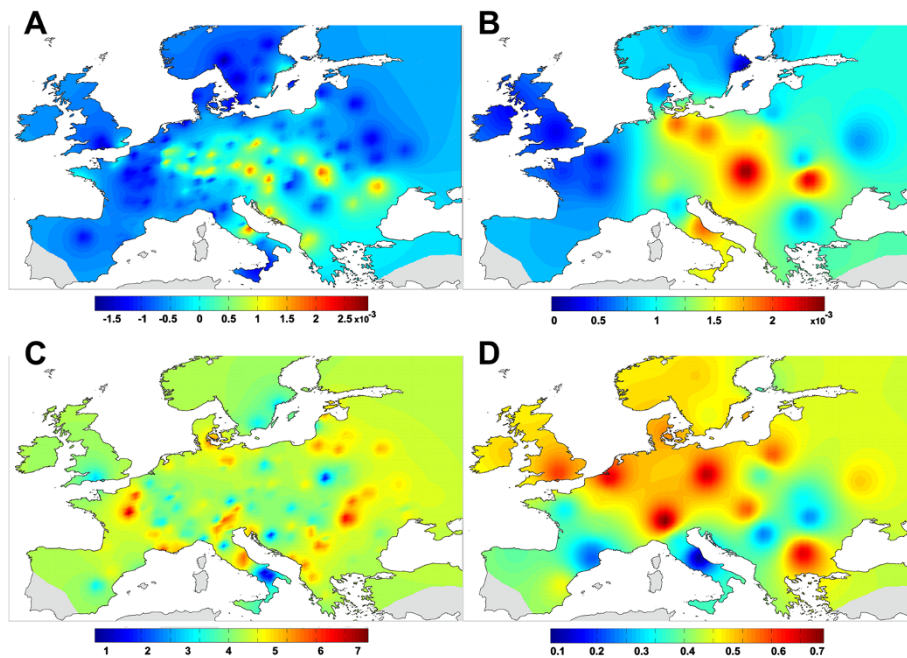
### Inter-individual genetic distance and genetic variability

The interpolation map based on multi-loci genetic distances showed larger local divergence in DNA sequences in sympatric areas between L2b and L3, and L2a and L3 (Central-Eastern Europe and Central Italy, respectively, Figure 3A). Nucleotide and haplotype diversities calculated for each geographic group (Additional file 1, Figure 2) showed: (i) a higher diversity for *COI* in South Italy and Central-Eastern Europe ( $h$  from 0.521 to 0.8,  $\pi$  from 0.0048 to 0.0063), (ii) higher diversities for *Cytb* in Central-Eastern Europe ( $h$  from 0.467 to 0.667,  $\pi$  from 0.001 to 0.00171), (iii) a higher diversity for *EF1 $\alpha$*  in Central-Eastern Europe and Italy ( $h$  from 0.511 to 0.733,  $\pi$  from 0.00161 to 0.00188), (iv) higher diversities for *PEPCK* in Central-Eastern Europe, South Italy and Spain ( $h$  from 0.456 to 0.8,  $\pi$  from 0.00055 to 0.00132). Interpolation maps based on averaged genetic diversity for all markers confirmed the trend of a larger diversity in South Italy and Central-Eastern Europe (Figure 3B).

### MMS variation of the red-tailed bumblebee

#### MMS composition

55 compounds were detected in the MMS of *B. lapidarius* (Additional file 3, Additional file 4). The five main compounds detected were aliphatic compounds: hexadec-9-en-1-ol ( $C_{16}H_{32}O$ ), hexadecane-1-ol ( $C_{16}H_{34}O$ ), hexadecenoic acid ( $\Delta 9$  and  $\Delta 7$ ,  $C_{16}H_{30}O_2$ ), hexadecenyl hexadecenoate



**Figure 3** Divergences, diversity and variability in genetic and MMS in *B. lapidarius*. **A:** Interpolation maps of genetic distance. **B:** Interpolation maps of nucleotide diversity. **C:** Interpolation maps of MMS distance. **D:** Interpolation maps of MMS variability.

( $C_{32}H_{60}O_2$ ), and hexadecyl hexadecenoate ( $C_{32}H_{62}O_2$ ). Hexadec-7-enoic acid was only detected in South Italian populations while hexadec-9-enoic acid was detected in other populations. The North Italian populations displayed specific compounds hexadecenal and hexadecanal. The MMS compositions of *B. lapidarius lapidarius* were similar to the previous studies [32,38,39] except for two minor components (tetradecanol and ethyl hexadecenoate).

#### MMS divergences

The nMDS (stress value = 0.13) and perMANOVA analyses separated the European populations into two main groups: Southern Italian populations (group G1) and all other European populations (group G2) (perMANOVA: DF = 1, F = 49.93, p-value < 0.01; Figure 4A). G1 corresponded to the lineage L2a while G2 corresponded to the lineages L2b + L3 (Figure 4B). The IndVal method reveals 18 significant indicator compounds for G1 and seven for G2.

#### Inter-individual MMS distance and MMS variability

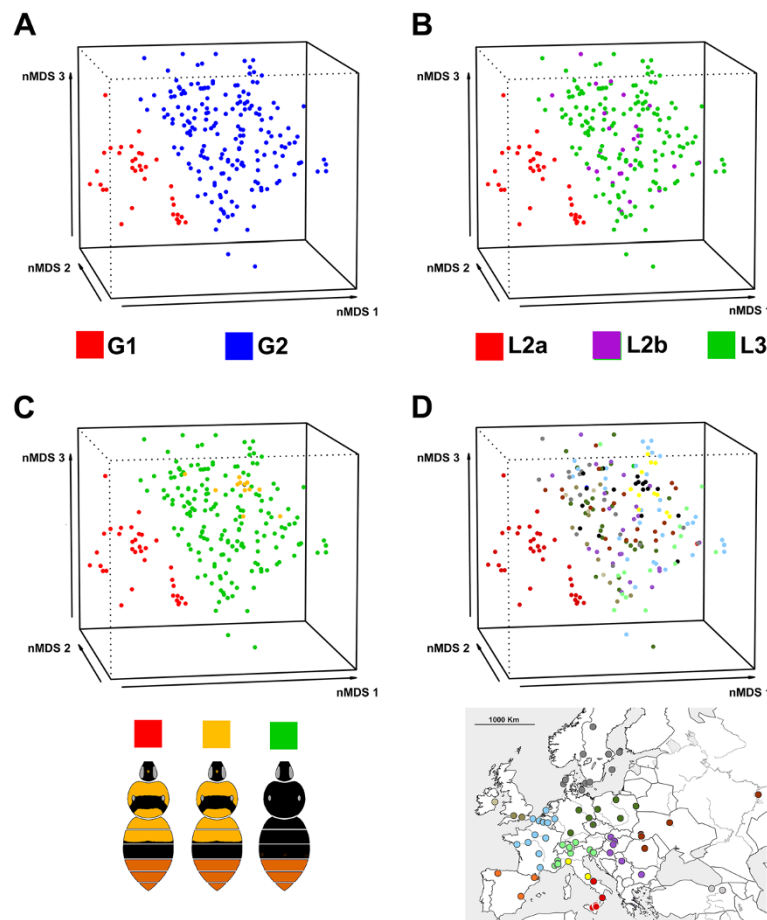
The interpolation map based on the MMS distance showed larger local divergence in MMS composition in the Alps, Central Italy, Eastern Europe, and Western France (Figure 3C). The MMS variability interpolation map (Figure 3D) revealed (i) the lowest variability in Central Italy, (ii) low variability in West Pyrenees and Carpathian region, and (iii) high variability in NE

Europe, the Alps and Central Europe. Comparison of the MMS variability showed significant differences between all geographic groups (AMOVA of multivariate dispersion p-value < 0.01, permutation test of multivariate dispersion p-value < 0.01) and between pairwise geographic groups confirmed these observations (Additional file 5).

#### Comparative statistical analyses between geographic origin, colour pattern, genetic and MMS

Mantel tests showed significant positive low correlations between (i) genetic distances *versus* geographic distances (Mantel's  $r = 0.18$ , p-value < 0.01), (ii) MMS distances *versus* geographic distances (Mantel's  $r = 0.13$ , p-value < 0.01), (iii) colour distances *versus* geographic distances (Mantel's  $r = 0.32$ , p-value < 0.01), (iv) genetic distances *versus* MMS distances (Mantel's  $r = 0.18$ , p-value < 0.01), (v) genetic distances *versus* colour distances (Mantel's  $r = 0.41$ , p-value < 0.01), and (vi) MMS distances *versus* colour distances (Mantel's  $r = 0.19$ , p-value < 0.01). Partial Mantel tests showed significant positive low correlations between (i) genetic distance *vs.* MMS distance (Mantel's  $r = 0.17$ , p-value < 0.01), (ii) genetic distances *versus* colour distances (Mantel's  $r = 0.38$ , p-value < 0.01), (iii) MMS distances *versus* colour distances (Mantel's  $r = 0.16$ , p-value < 0.01).

Our correlation analysis showed no linear correlation between genetic diversity and MMS variability (p-value = 0.25).



**Figure 4 nMDS analysis of MMS.** Three first axes of the nMDS ordination plot based on Bray-Curtis distances calculated on a male marking secretion matrix of *B. lapidarius* (Stress value = 0.13,  $R^2 = 0.86$ ). **A:** Colour of individuals in nMDS based on groups detected in perMANOVA. **B:** Colour of individuals in nMDS based on genetic lineages (see Figure 1D). **C:** Colour of individuals in nMDS based on colour pattern (in red *B. lapidarius decipiens* South Italy). **D:** Colour of individuals in nMDS based on geographic areas (see Figure 1B).

## Discussion

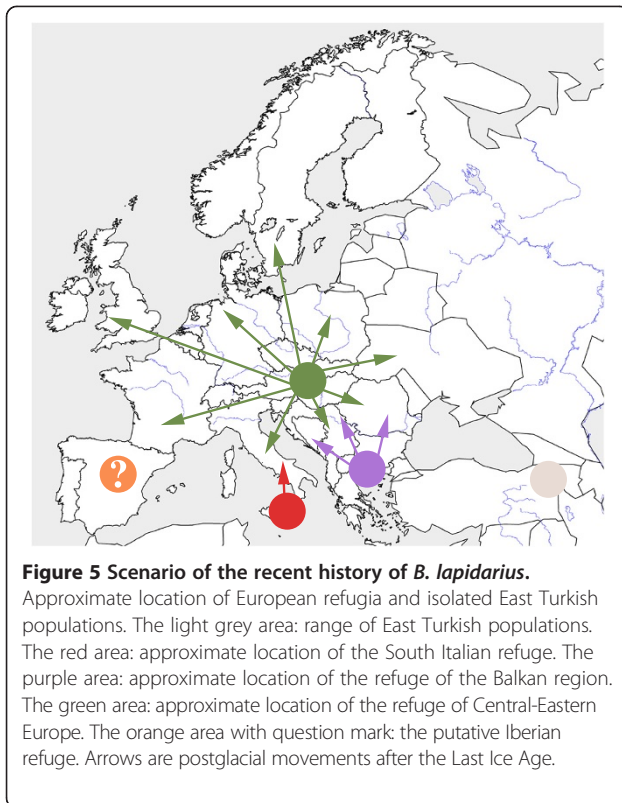
### Phylogeography of the red-tailed bumblebee

Phylogeographic structure is usually consistent with long-term isolation in multiple refugia during climatic oscillations (e.g. [5]). In this way, the Quaternary glacial events have been shown to explain the intraspecific genetic pattern of many European species [4,15,16]. During Ice Ages, the ranges of most European temperate species contracted to the southern peninsulas of Iberia, Italy and Balkans (classic theory: e.g. [5]). Glaciers, European mountains or repeated marine-flooding during the Pleistocene have constituted major vicariant barriers between these southern refugia leading to allopatric genetic divergence in many species (e.g. [16,40]). Interglacial and postglacial recolonizations of Europe have arisen from these Mediterranean glacial refugia and principally from the Balkans [16,41]. Therefore, it is to be expected that populations of temperate species surviving in Mediterranean refugia display higher genetic variability in the southern peninsulas [41].

Here, our results (trees, networks, Mantel tests, and SAMOVA) detect an intraspecific phylogeographic structure within *B. lapidarius* with three major lineages (L1, L2, and L3; Figure 1; Additional file 2). The roughly estimated molecular clock analyses date divergences between European lineages (L2 and L3) to the Quaternary while the divergence of East Turkish populations takes place earlier (Pliocene). Therefore, the divergence of European lineages can be considered as a consequence of allopatric differentiation during Ice Ages.

A previous study based on morphology has hypothesized that *B. lapidarius* has survived up to the last Ice Age in Iberia, Italy, Balkans, and Turkey prior to a post-glacial recolonization of Europe from the Balkans [42]. Our results suggest an alternative scenario of the recent history of *B. lapidarius* (Figure 5).

The four lineages seem to arise from four regions: East Turkey, South Italy, the Balkan region, and Central-Eastern Europe (C.E. Europe). (i) East Turkey hosts largely differentiated and long-term divergent populations (Figure 1; see



**Figure 5 Scenario of the recent history of *B. lapidarius*.**

Approximate location of European refugia and isolated East Turkish populations. The light grey area: range of East Turkish populations. The red area: approximate location of the South Italian refuge. The purple area: approximate location of the refuge of the Balkan region. The green area: approximate location of the refuge of Central-Eastern Europe. The orange area with question mark: the putative Iberian refuge. Arrows are postglacial movements after the Last Ice Age.

molecular clock results). These populations do not contribute to the European genetic pool. (ii) South Italy has acted as a refugium as shown by its large genetic diversity (Figure 3B) and its endemic haplotypes gathered in one distinct lineage L2a (Figures 1, 2 and 3B, Additional file 1). (iii) The Balkan region has acted as a refugium according to its specific haplotypes gathered in one distinct lineage L2b (Figure 1, Additional file 1). The low genetic diversity of the Balkan region could be due to repeated bottlenecks of populations in this region and/or a prolonged bottleneck during past glacial oscillations (e.g. [4,43]). (iv) C.E. Europe has a large genetic diversity (Figure 3B) and many distinct haplotypes gathered in one specific lineage L3 (Figures 1, 2 and 3B, Additional file 1). Even if the genetic diversity is increased by admixture with an allopatric lineage L2b (Figures 1, 3A; see in other species [43]), C.E. Europe hosts the largest diversity of specific haplotypes (Figures 1, 2, 3B). Therefore, this region seems to have been a major glacial refuge for *B. lapidarius*. This evidence fits well with the theory of cryptic refugia in Central, Northern or Eastern Europe, an alternative to the classic theory, observed in other temperate species (review in [44]).

The status of Iberia remains confused according to the slight differentiation/low diversity in mitochondrial markers and the higher differentiation/high diversity in nuclear markers of Iberian populations (similar to South Italian, East Turkish or South East European populations;

see Figure 2). The poorly resolved relationship with the lineage L3 (Figure 1) makes it impossible to know if Iberia acted as a refugium (with a strong bottleneck) or has been secondarily colonized by populations from C.E. Europe.

European refugia have probably been isolated by Alpine glaciers, the Alps, Carpathians, Dinaric Alps or repeated marine-flooding of the Crati-Sibari plain (North Calabria, [45]) during the Pleistocene period leading to allopatric divergences of European lineages. In contrast, the frequent drops in the level of the Adriatic Sea during the Quaternary Ice Ages [46] probably allowed gene flow between the Balkans and Italian Peninsula leading to a close relationships between these lineages as observed in rodents (e.g. [47]) and in turtles (e.g. [48]). According to our results, most of Europe has been recolonized from the C.E. Europe at the end of the last Ice Age (Figure 1). The C.E. European lineage has experienced allopatric differentiation (Figure 1C-D; SAMOVA results) probably by reduction of the gene flow [49] resulting from natural sea barriers separating the British Isles and Gotland. The Balkan lineage also dispersed from its refugium and largely mixed with the C.E. European lineage in Central Europe (Figures 1, 3A), a common suture zone in many species (review in [50]). In contrast the South Italian lineage did not contribute to the postglacial colonization, and geographically slightly overlapped the C.E. European lineage in a narrow area in Central Italy (a similar contact zone exists for amphibians and reptiles [51,52]). This was most likely caused by the invasion of North Italy by C.E. European lineage and a competitive exclusion process [53,54] between lineages in this region.

#### MMS differentiation

Reproductive traits are generally assumed to be mainly shaped by (i) intraspecific interactions to maximize encounter rates among conspecific mates (sexual selection; [55,56]) and (ii) interspecific interactions to maintain isolation barriers and decrease the likelihood of hybridization events among syntopic sister species [21,57,58]. This results in a stabilizing selection on reproductive traits to promote a species-specific signal (e.g. [21,58]). Across species distribution, local reproductive trait variations (e.g. in moths: [59], in flies: [60], in solitary bees: [28], in birds: [61]), promoted by selection for specific optimized reproductive traits [58,62], can appear due to changes in factors that affect communication systems: (i) mutation of genes involved in reproductive traits (e.g. [62,63]), (ii) intraspecific interactions like local preferences of the receivers [23,64], (iii) the presence/abundance of species with a similar courtship signal which would result in selection for releasers with the most distinct, optimized reproductive traits (e.g. [58]). These changes in factors that affect communication can happen through allopatric differentiation as observed in



insular bumblebees where genetic divergence (fostered by geographic isolation) and local specific sexual selection (female preferences, e.g. [34]) lead to MMS divergence from other populations [23]. This reproductive trait differentiation can persist in secondary sympatry if the differentiation has led to the establishment of a reproductive (pre-zygotic) isolation barrier (allopatric hypothesis, see [65]). Alternatively, reproductive trait divergence can happen when formerly isolated lineages meet as a result of range expansion at a secondary contact zone (reinforcement hypothesis; see [25,66,67]). Reinforcement leads to local adaptations in secondary contact zones (e.g. [68,69]).

Our statistical analyses (Figure 4, perMANOVA results) support MMS quantitative as well as qualitative differentiations of Southern Italian populations from other European populations of *B. lapidarius*. These qualitative changes (e.g. from hexadec-9-enoic acid in Southern Italian individuals to hexadec-7-enoic acid in other individuals) have probably arisen from enzyme change (e.g. from  $\Delta 9$ -desaturase to  $\Delta 7$ -desaturase acting on palmitic acid) by mutations of enzyme genes (e.g. [62]) or by activation of a non-functional enzyme gene transcript present in a common ancestor, as observed in moths [64].

The MMS differentiation is poorly explained by geographic distance but displays an obvious geographic pattern (Figure 4, Mantel tests results). The MMS divergence is also weakly correlated with the genetic distance but strongly matches the divergence of the lineage L2a (Figure 4B; Mantel tests results). This suggests that the South Italian MMS differentiation is related to the divergence of the South Italian lineage. Moreover, the lack of gene flow without an obvious geographic barrier between Northern and Southern Italy (Figures 1, 2, 3A, but see [30,42] for putative hybrid records) suggests a reproductive isolation fostered by species intrinsic isolation mechanisms, probably by MMS differentiation leading to pre-zygotic isolation barriers between populations. Therefore we expect that the MMS divergence has appeared subsequent to (i) allopatric divergence during Ice Ages and persistence during current interglacial period (allopatric hypothesis) or (ii) a reinforcement process during the range expansion and the admixture with the C.E. lineage (reinforcement hypothesis). The Southern Italian MMS differentiation seems due to allopatric divergence during glacial isolation rather than reinforcement in secondary contact zone. This is because the Southern Italian MMS pattern is spread over the whole area of the Southern Italian lineage while reinforcement should mainly lead to local adaptations in secondary contact zones as observed in flies [68].

Our statistical analyses demonstrate that the intrapopulation MMS variability is not correlated to the intrapopulation genetic diversity. The geographic areas of

high/low intrapopulation MMS variability do not match with areas of high/low intrapopulation genetic diversity (Figure 3B-D). The MMS variability map shows the lowest intrapopulation MMS variability in the contact zone between C.E. European and Southern Italian lineages. We hypothesise that this local decrease of intrapopulation MMS variability could be due to the intensification of the stabilizing selection on MMS for local specific optimized reproductive traits further to new interaction between closely related taxa (lineage L2a and L3) during the post-glacial range expansion. This local reinforcement process is also supported by the divergence of Northern Italian populations of C.E. European lineage from other C.E. European lineage populations in MMS composition which increases the MMS divergence with Southern Italian populations (Additional file 3).

The comparisons of the phylogeographic patterns, MMS pattern distribution and MMS variability distribution suggest that the MMS divergence leading to pre-zygotic isolation between Southern Italian and C.E. European lineages has been fostered by past allopatry between refugia and persists in the current interglacial period (allopatric hypothesis). The post-glacial range expansion of these lineages differentiated in MMS seems to have led to a reinforcement process in the secondary contact zone. Further studies in contact zones (e.g. testing the fitness of hybrids, bioassays) are needed to check this hypothesis.

#### Colour pattern variation

The genetic distances and lineages only partially reflect the colour patterns [30] (Figure 1, see also Mantel test results). The colour pattern *B. lapidarius caucasicus* fits with L1 while other colour forms (*B. lapidarius decipiens* and *B. lapidarius lapidarius*) spread in L2 and L3 (Figure 1). The MMS distance is also poorly correlated to the colour pattern distance. The MMS and colour patterns are only congruent for the South Italy individuals (Figure 4C). This incongruence between colour patterns, MMS and phylogeny should be a consequence of local Müllerian mimicry processes (e.g. convergence of Spanish *B. lapidarius* to the Pyrenean colour pattern [70]) as observed in other bumblebees [71,72].

#### Speciation

The consequence of genetic and MMS divergences between populations fostered by species range contraction during climatic oscillations can range from simple regional variations (e.g. [73]), to a speciation processes between allopatric refugia (e.g. [16,74]). Several pieces of evidence strongly suggest an incipient speciation between Southern Italian and other European lineages of *B. lapidarius* contrary to the current literature (e.g. [30]): (i) the concordance of genetic divergence in tree topologies

derived from mitochondrial and nuclear genes [4,75], (ii) the persistence of the genetic divergence through time despite the current secondary contact zone [75], (iii) the divergences in chemical reproductive traits (e.g. qualitative changes or double-bond position shifts) which can lead to discrimination by receivers and to pre-mating isolation [21] as observed in moths (e.g. [64]) and in bumblebees [76]. Therefore, Southern Italy could be a speciation centre for *B. lapidarius* as observed in amphibians [52,77,78] or in reptiles (e.g. [48,51,79]). In the same way, the large mitochondrial and nuclear divergences observed in East Turkish *B. lapidarius* strongly suggest that they are a different species (MMS data not available). In contrast, the speciation process has not happened in other lineages according to current gene flow and the lack of MMS and morphological differentiations. However, these speciation hypotheses need to be checked with further integrative taxonomy, including morphologic, chemical, ethologic and genetic criteria (e.g. [80]) based on an adapted species concept [81] in order to assess the taxonomic status within the *B. lapidarius* complex.

## Conclusions

Quaternary climatic oscillations seem to have led to the European *B. lapidarius* population contraction in three main refugia. These refugia were localized inside Mediterranean peninsulas (Southern Italy and the Balkan) and in Central-Eastern Europe. Inside the Southern Italian refugium, populations have experienced allopatric MMS differentiation and an incipient speciation process from other lineages. At the end of Ice Age, the range expansion of lineages have led to secondary contact zone between the C.E. Europe lineage and Southern Italy lineage where a reinforcement process on local MMS patterns seems to take place.

This study suggests that the population movement during Quaternary climatic oscillations can lead to divergence in reproductive traits by allopatric differentiation during Ice Age and by reinforcement during post-glacial recolonization.

## Methods

### Sampling

259 males of *B. lapidarius* covering its whole European distribution were collected (Figure 1A, Additional file 6). Males were killed by freezing at  $-20^{\circ}\text{C}$ . The MMS were extracted in 400  $\mu\text{l}$  of hexane from dissected cephalic labial glands or entire cut heads [82]. Bodies without heads were preserved in 99% ethanol for DNA extraction. All samples were stored at  $-40^{\circ}\text{C}$  until analyses. Populations from the Caucasus (*B. lapidarius eriophorus*) and Morocco (*B. lapidarius atlanticus*) were not collected. Closely related species were also collected as outgroup [83] (see Additional file 6).

## Genetic analyses

### Gene selection

We sequenced five genes commonly used to study inter-specific and intraspecific relationships (e.g. [80,83,84]): mitochondrial cytochrome oxidase 1 (*COI*), cytochrome b (*Cytb*), nuclear protein-coding genes long-wavelength rhodopsin copy 1 (*LWhH*), elongation factor-1 alpha F2 copy (*EF-1 $\alpha$* ), and phosphoenolpyruvate carboxykinase (*PEPCK*). We used a multi-gene approach to check congruence of phylogenetic and phylogeographic patterns inferred from independent genes.

### DNA preparation, amplification and sequencing

Total DNA was extracted using a QIAGEN DNeasy<sup>®</sup> Tissue Kit (Quiagen Inc., Valencia, CA). Legs were removed from the specimen, crushed using liquid nitrogen, and digested (four hours in proteinase K at  $56^{\circ}\text{C}$ ). Polymerase chain reaction (PCR) amplifications were carried out for all genes and all samples using primer pair Apl2013/ApH2931 [84] for *COI*, CB1/CB2 [85] for *Cytb*, LWRhF/LWRhR [86] for *LWRh*, F2-ForH/F2-RevH2 [87] for *EF-1 $\alpha$* , and FHv4/RHv4 [83] for *PEPCK*. PCR amplifications were carried out by initial denaturing for three minutes at  $94^{\circ}\text{C}$ , 35 (*COI*, *LWRh*, *EF-1 $\alpha$* ) or 40 (*Cytb* and *PEPCK*) cycles of one minute denaturing at  $94^{\circ}\text{C}$ , one minute annealing at  $51^{\circ}\text{C}$  (*COI*),  $50^{\circ}\text{C}$  (*Cytb*),  $60^{\circ}\text{C}$  (*LWRh*),  $54^{\circ}\text{C}$  (*EF-1 $\alpha$* ) or  $48.5^{\circ}\text{C}$  (*PEPCK*), two minutes elongation at  $72^{\circ}\text{C}$  and a final extension for ten minutes at  $72^{\circ}\text{C}$ . Voucher specimens and PCR products used in molecular investigation were deposited at the University of Mons (Belgium). Genes were sequenced with an ABI 3730XL sequencer (Applied Biosystems, Foster City, CA, USA), with ABI 3730 DNA analyzer or by GENOSCOPE (*Centre National de Séquençage*; Evry, France). Both strands of each PCR product were sequenced. Consensus sequences were computed with CodonCode Aligner 3.0.1. There is no uncertainty in the consensus sequences. The bumblebee origin of each sequence was checked with BLAST 2.2.20 [88]. The alignment was performed by MAFFT ver.6. (using FFT-NS-2 algorithms, default parameters) [89]. The data matrix was computed on Mesquite 2.74 (build 486) [90]. Translation to proteins (using the *Drosophila* mitochondrial DNA genetic code or Universal genetic code) was performed on Mesquite. Sequences were deposited in GenBank [GenBank: KC915396 to KC916646] (Additional file 6).

### Haplotype relationships

We performed phylogeographic and phylogenetic analyses to investigate the relationships between haplotypes and to infer the phylogenetic lineages. A test of saturation was applied to each gene fragment in PAUP\* 4.0b 10 [91].

Haplotype networks were computed using the median-joining method (MJM, [92]) in Network 4.6.1.0

(www.fluxus-engineering.com) for each gene. The median-joining method uses a maximum parsimony approach to search for all the shortest phylogenetic trees for a given data set [92]. To reconstruct the network, we weighted transversions twice as high as transitions.

Phylogenetic analyses were carried out using distance (neighbour joining [NJ]), maximum likelihood (ML) and Bayesian (MB) methods. Trees were rooted with species from outgroup. For each approach, we analysed (i) each gene individually, (ii) all mitochondrial genes combined (*COI* + *Cytb*), (iii) all nuclear genes combined (*EF-1 $\alpha$*  + *LWRh* + *PEPCK*), and (iv) all genes combined. We assessed the level of incongruence in phylogenetic reconstructions among genes by comparing congruence and incongruence of well-supported clades across NJ, ML and MB trees of each gene. Overall, the gene trees were not strongly conflicting with each other and well-supported clades were similar across all trees (see supplementary trees at [TreeBASE ID: S14299]). We did not perform the incongruence length difference test [93] because its utility in evaluating homogeneity test has been questioned (e.g. [94]).

NJ analyses were performed using PAUP\*. We used jModeltest [95] using the Akaike information criteria corrected for small sample sizes (AICc, [96]) to determine the best fitting substitution models (GTR + G). The robustness of the NJ trees was assessed by bootstrap resampling (10,000 random replications, [97]).

ML analyses were conducted in GARLI 2.0 [98]. Each gene was partitioned to explore the best substitution model: (i) *EF-1 $\alpha$*  into two exons and one intron, (ii) *LWRh* into three exons and two introns, (iii) *PEPCK* into two exons and two introns, and (iv) *COI*, *Cytb* and each nuclear exon by base positions (1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup>). The best fitting substitution models were chosen with jModeltest using the AICc for each dataset. The chosen models were: (i) for *COI*: TIM1 + G (1<sup>st</sup>), F81 (2<sup>nd</sup>) and TPM1uf + G (3<sup>rd</sup>); (ii) for *Cytb*: TrN + I + G (1<sup>st</sup>), F81 (2<sup>nd</sup>) and HKY + G (3<sup>rd</sup>); (iii) for *EF-1 $\alpha$*  exon 1: F81 (1<sup>st</sup> and 2<sup>nd</sup>) and HKY (3<sup>rd</sup>); (iv) for *EF-1 $\alpha$*  intron: HKY; (v) for *EF-1 $\alpha$*  exon 2: F81 (1<sup>st</sup>), K80 (2<sup>nd</sup>) and TPM3 (3<sup>rd</sup>); (vi) for *LWRh* exon 1: F81 (1<sup>st</sup> and 2<sup>nd</sup>) and K80 (3<sup>rd</sup>); (vii) for *LWRh* introns 1 and 2: F81; (viii) for *LWRh* exon 2: F81 (1<sup>st</sup>), JC (2<sup>nd</sup>) and K80 (3<sup>rd</sup>); (ix) for *LWRh* exon 3: JC (1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup>); (x) for *PEPCK* introns 1 and 2: HKY; (xi) for *PEPCK* exon 1: TrN (1<sup>st</sup>), JC (2<sup>nd</sup>) and TrNef (3<sup>rd</sup>); (xii) for *PEPCK* exon 2: F81 + I (1<sup>st</sup>), K80 (2<sup>nd</sup>) and JC (3<sup>rd</sup>). A random starting tree and the automated stopping criterion (stop when the ln score remained constant for 20,000 consecutive generations) were used. Ten independent runs in GARLI were carried out; the topology and  $-\ln(L)$  were identical among replicates. The highest likelihood of one of those runs was retained. Statistical confidence in nodes was evaluated

using 10,000 non-parametric bootstrap replicates [97] using the automated stopping criteria set at 10,000 generations. Topologies with bootstrap values  $\geq 70\%$  were considered well supported [99].

MB analyses were carried out using MrBayes 3.1.2 [100]. The model selection process was the same as that for ML analyses. Selected models which are not implemented in MrBayes were substituted by the closest overparameterized model [101]. The TIM1, TIM1uf, and TrN were replaced by the GTR model whereas the TPM3 and TrNef were replaced by the SYM model. The proportion of invariant sites (I) and gamma distributed rates (G) defined in jModeltest were conserved in all models. Five independent analyses were carried out (30 million generations, four chains with mixed-models, default priors, saving trees every 100 generations). The analyses were stopped after checking convergence between runs using the average standard deviation of split frequencies and by plotting likelihood values across generations using Tracer 1.4 [102]. The first three million generations were discarded as burn-in. The phylogeny and posterior probabilities were then estimated from the remaining trees and a majority-rule 50% consensus tree was constructed. Topologies with posterior probabilities  $\geq 0.95$  were considered as well supported [103].

#### Estimating divergence time

Following the approach of Duennes *et al.* [104], we analyzed the *COI* dataset in BEAST v1.7.2 [105] to roughly estimate the time to most recent common ancestor of *B. lapidarius* haplogroups. Using the GTR + G model selected by jModeltest, we ran Markov chain Monte Carlo simulations with the coalescent constant population size tree model and the strict clock model. Prior, the relaxed uncorrelated lognormal molecular clock model [106] was implemented to assess the clock-like nature of the data. The sampled marginal posterior probability (PP) distribution of both standard deviation and coefficient of variation of substitution rates among tree branches included 0 in the 95% HPD, indicating that there was no strong evidence for substantial rate heterogeneity among lineages [105]. We specified a range of possible substitution rates which include extreme rate of insect mitochondrial genes recorded in the literature (see [104]) using a flat prior ranging from  $1 \times 10^{-9}$  to  $1 \times 10^{-7}$  substitutions site<sup>-1</sup> and year<sup>-1</sup>. Simulations were run for 30 million generations, sampling every 1000 generations. Four independent runs were assessed in Tracer v1.4.1 [102] to confirm convergence, determine burn-in, and examine the effective sample size of all posterior parameters. Log files from each run were combined in LogCombiner v1.6.1 [102] for final parameter estimates.

### **Identifying population structure**

Population structure was assessed for each gene by estimating “global”  $\Phi_{ST}$  statistic on populations (i.e. only AMOVA  $\Phi_{ST}$  estimator when considering only one group of populations [107]) using ARLEQUIN 3.5 [108] with 100,000 permutations. When global  $\Phi_{ST}$  statistic was significant for one gene, we analyzed the population structure using the spatial AMOVA procedure (SAMOVA) [109] implemented in SPADS 1.0 [110]. This procedure assigns populations to groups based on geographical vicinity and sequence similarity. The most likely structure corresponds to the partition of populations that maximized among-group variation measured by the AMOVA  $\Phi_{ST}$  statistic [107]. We performed ten independent repetitions of 10,000 simulated annealing steps for  $K$  values varying from two to 45.

### **Estimating genetic diversity**

Genetic diversity was estimated with the number of different haplotypes, the haplotype diversity ( $h$ , [111]), and the nucleotide diversity ( $\pi$ , [112]) using DnaSP [113] within defined geographic groups and for each gene separately. The defined geographic groups gathered samples from the same geographic regions (see Additional file 6).

### **Cartography of inter-individual genetic distance and intra-population genetic diversity**

We used an extension of the method developed by Miller [114] to represent spatial variations of genetic patterns (distance and diversity). The method of Miller [114] is based on a connectivity network (e.g. a Delaunay triangulation) built from the sampling localities. Inter-individual genetic distances are first estimated and assigned to landscape coordinates at midpoints of each Delaunay triangulation edge. An interpolation procedure (inverse distance-weighted interpolation, [115,116]) is then used to infer genetic distances at locations on a uniformly spaced grid. Here, we also used this method to represent the spatial variation of any kind of distance and measures of diversity. The interpolation procedure of measures of diversity is based on (diversity) values directly estimated at each sampling point rather than on (distance) values assigned at midpoints of each edge of a connectivity network. Interpolation surfaces were generated with the MATLAB (The MathWorks, Inc) functions GDisPAL for “genetic distance patterns across landscapes” and GDivPAL for “genetic diversity patterns across landscapes” [110]. All generated surfaces were finally superimposed on a map of the distribution area of the geographic region studied. We generated two surfaces based (i) inter-individual genetic distance measured as DNA sequence mismatches averaged over the four polymorphic loci (i.e. averaged  $p$ -distances over loci), and (ii) on genetic diversity measured with the nucleotide diversity estimated within each geographic group

(see Additional file 1) and averaged over the four polymorphic loci. Both the inter-individual genetic distance and nucleotide diversity were computed with SPADS [110]. The geographic coordinates of geographic groups were the barycenter between all sampling places included in geographic groups. All surfaces were generated using a distance weighting parameter  $a = 2$  and distance surfaces were both based on a Delaunay triangulation connectivity network. For the interpolation parameter  $a$ , we selected a value that allowed to generate interpolation surfaces displaying local details. As advised by Miller et al [114], we performed the distance interpolations using (i) residual distances derived from the linear regression of genetic vs. geographical distances [117] and (ii) MMS vs. geographic distances in order to account for potential correlation between genetic and geographic distances.

### **MMS analyses**

#### **Chemical analyses of MMS**

We analyzed the MMS of 226 individuals (Additional file 6). The composition of MMS was determined by gas chromatography-mass spectrometry (GC/MS) on a Finigan GCQ with a DB-5 ms non-polar capillary column (5% phenyl (methyl) polysiloxane stationary phase; 30 m  $\times$  0.25 mm  $\times$  0.25  $\mu$ m) and an ion trap in electron impact mode “full scan (300-600)”. We used a splitless injection mode (220°C) and helium as carrier gas (50 cm/s). The temperature programme of the column was set to 70°C for two minutes and then increased at a rate of 10°C/min to 320°C. The temperature was then held at 320°C for five minutes. Compounds were identified in Xcalibur™ using their mass spectra compared to those at National Institute of Standards and Technology library (NIST, U.S.A) using NIST MS Search 2.0. The double bond positions were determined i) from mass spectra of dimethyl disulphide adducts of unsaturated components (reaction time: four hours) and ii) by chemical ionization with acetonitrile as a reaction gas. The products were analyzed by GC/MS using the same temperature programme as for the original extracts. An ion trap GC/MS instrument (Varian Saturn, 2000) was used for chemical ionization. The compound relative quantification of MMS was estimated using a gas chromatograph Shimadzu GC-2010 with a SLB-5 ms non-polar capillary column (5% diphenyl/95% dimethyl siloxane; 30 m  $\times$  0.25 mm  $\times$  0.25  $\mu$ m) and a flame ionization detector. The chromatographic conditions were the same as above. The absolute amounts of compounds were quantified using GCsolution Postrun (Shimadzu Corporation) with automatic peak detection and noise measurement. Relative amounts (RA in%) of compounds in each sample were calculated by dividing the absolute amounts of compounds by the total absolute amount of compounds in each sample. We did not use

any correction factor to calculate the RA of individual compounds. All compounds for which RA were recorded as less than 0.1% for all specimens were discarded [118]. The data matrix for each species was elaborated with the relative proportion of each compound for each individual [82] using GCAaligner 1.0 [119] before final check.

#### **Comparative statistical analyses based on MMS**

All statistical analyses were performed using R [120] to detect MMS differentiations between populations. Data consisting of the relative proportion of all compounds were transformed ( $\log(x-1)$ ) to reduce the large difference of abundance between highly and slightly concentrated compounds and standardized (mean = 0, standard deviation = 1) to reduce the sample concentration effect [82]. We compared the profile for each sample (ungrouped) with non-metric multidimensional scaling (nMDS) ordination using a Bray-Curtis similarity matrix, three dimensions and 50 runs (R-package *ecodist*, [121]). When we detected MMS differentiations between population groups in nMDS, we assessed results by performing permutation-based version of the multivariate analysis of variance (perMANOVA) using the Bray-Curtis similarity matrix and 10,000 permutations (R-package *vegan*, [122]). Like conventional analyses of variances, the perMANOVA calculates an F statistic by taking the ratio of among group sums of squares to within group sums of squares. The perMANOVA is robust to violations of multivariate normality but requires homogeneity of variances. When we tested more than two groups and the returned p-value was significant ( $p < 0.01$ ), multiple pairwise comparisons were conducted and p-values were adjusted using Bonferroni's correction to avoid increases of type error I due to multiple testing. Prior to perMANOVA we checked its assumption (similar variance homogeneity between groups) by performing a distance-based test for multivariate homogeneity of group dispersions (MHGH) for a one-way ANOVA design [123]. The procedure first calculates the Euclidean distances between MMS composition and respective geographic group centroids on Pearson r correlation matrices (R- package *vegan*, [122]). To test if one group was more variable than the other, the magnitudes of these distances were then compared between groups using ordinary ANOVA. Moreover, a permutation test was run to generate a permutation distribution of F under the null hypothesis of no difference in dispersion between groups (R- package *vegan*, [122]).

#### **Specific MMS compounds**

To determine compounds specific to groups defined by nMDS and confirmed by perMANOVA, we used the indicator value (IndVal) method [124]. A high value is obtained when the compound is specific and regular to a

particular group compared to the whole set of observations. The statistical significance of a compound as an indicator at the 0.05 level was evaluated using a randomization procedure.

#### **Chemical variability within populations**

We compared variability of MMS between each geographic group (pairwise analyses) using a distance-based test for MHGH for a one-way ANOVA design [123].

#### **Cartography of inter-individual MMS distance and intra-population MMS variability**

We used the extended Miller's [114] method (i.e. GDisPAL and GDivPAL functions, see above) to represent spatial variations of MMS patterns (distance and variability). We generated two different surfaces: (i) on MMS distance (Bray-Curtis similarity matrix with dataset transformed and standardized), (ii) on MMS variability (the mean Euclidean distance between MMS composition of each sample and its respective geographic group centroids (computed on Bray-Curtis similarity matrix with dataset transformed and standardized)). MMS variability was estimated within each geographic group (see Additional file 1). The geographic coordinates of geographic groups were the same as genetic analyses. All surfaces were generated using a distance weighting parameter  $a = 2$  and distance surfaces were both based on a Delaunay triangulation connectivity network. Furthermore, we performed the distance interpolations using residual distances derived from the linear regression of MMS *versus* geographical distances in order to account for potential correlation between MMS and geographical distances.

#### **Comparative statistical analyses between geographic origin, colour patterns, genetic and MMS**

We investigated correlations between genetic divergence, colour patterns, MMS differentiation, and geographic distance by performing six Mantel tests [125] with 10,000 random permutations in R (R-package *vegan*): (i) genetic distances *versus* geographic distances, (ii) MMS distances *versus* geographic distances, (iii) colour distances *versus* geographic distances, (iv) genetic distances *versus* MMS distances, (v) genetic distances *versus* colour distances, and (vi) MMS distances *versus* colour distances. Geographic distances between all samples were computed with Geographic Distance Matrix Generator version 1.2.3 [126]. Genetic distances were calculated in the same way as for the cartography of inter-individual genetic distance (see above) and were thus based on DNA sequence mismatches averaged over the four polymorphic loci (i.e. averaged p-distances over loci) computed with SPADS [113]. The MMS distance matrix was the individual-by-individual Bray-Curtis distance matrix based on MMS compounds. The colour distance matrix compared the

colour pattern between all individuals and was generated as followed: when two individuals shared the same colour pattern, their colour distance was set to “0” and otherwise it was set to “1”.

We also performed three partial Mantel tests based on 10,000 random permutations controlling for spatial autocorrelation between (i) genetic distances *versus* MMS distances, (ii) genetic distances *versus* colour distances, and (iii) MMS distances *versus* colour distances.

We used a generalized linear model [127] to estimate the potential linear correlation between genetic diversity and MMS variability. To estimate diversities we gathered samples from the same geographic groups (Additional file 6) and excluded geographic groups with only one sample. Genetic diversity used was the nucleotide diversity estimated within each geographic group and averaged across loci. MMS variability was the mean Euclidean distance between MMS composition of each sample and its respective geographic group centroids (computed on Bray-Curtis similarity matrix with dataset transformed and standardized).

#### Data availability

The genetic datasets supporting the results of this article are available at TreeBase <http://purl.org/phylo/treebase/phylo/study/TB2:S14299>. Genbank accession numbers are listed in the Additional file 6 - Table of sampling. The MMS dataset supporting the results of this article is included within the article and its Additional file 3.

#### Additional files

**Additional file 1: MMS variability, nucleotide diversity and haplotype diversity observed among geographic groups.**

Geographic group: see Additional file 1, MMS variability: Euclidean distance between MMS composition of each sample and its respective geographic group centroids.

**Additional file 2: The results from the SAMOVA analyses. K: number of genetic clusters.** Number in gene column is the group of each geographic group.

**Additional file 3: MMS data matrix (relative amounts of each compound) of *B. lapidarius*.**

**Additional file 4: List of the identified compounds in European *B. lapidarius*.** Molecular weight [MW (m/z)], median [Med (%)], first and fourth quartiles [Q1 (%) and Q2 (%)], minimum and maximum [Min (%) and Max (%)] of the 55 identified compounds. Unknown x are undetermined compounds.

**Additional file 5: The results of AMOVA of multivariate dispersion test between each geographic group.** Values are the p-values.

**Additional file 6: Table of sampling.** Sample Code: sample labels used in analyses and supplementary tree, Taxa: Name of taxa, Geographic groups: Groups of sampling place, *COI*, *Cytb*, *EF-1a*, *LWRh*, and *PEPCK* are the Genbank accession numbers for each sample.

#### Abbreviations

AMOVA: Analysis of molecular variance; C.E. Europe: Central and Eastern Europe; *h*: Haplotype diversity; L1: East Turkey lineage; L2: Southern Italy (L2a) and South Eastern European (L2b) lineages; L3: C.E. European lineage;

LWRh: Nuclear protein-coding genes long-wavelength rhodopsin copy 1; MB: Bayesian; MHGH: Multivariate homogeneity of group dispersions; ML: Maximum likelihood; MMS: Male marking secretion; NJ: Neighbour joining; perMANOVA: Permutation-based version of the multivariate analysis of variance; SAMOVA: Spatial AMOVA; *n*: Nucleotide diversity.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

TL and SD conceived the study and interpreted the results. The experiments were designed by TL. TL, SD, MV, and IV analysed the data. TL, SD, DM, PL, IV, J-YR, and PR contributed to materials/analysis tools. TL wrote the manuscript. All authors finalised the manuscript. All authors read and approved the final manuscript.

#### Acknowledgements

We are grateful to P. Mardulyn for his hospitality at the *Université Libre de Bruxelles* (Brussels, Belgium) for the molecular analyses. We gratefully acknowledge B. Cederberg (Uppsala, Sweden), A. Cetkovic (Belgrade, Serbia), M. Cornalba (Pavia, Italy), T. De Meulemeester (Leiden, Netherlands), M. Franzen (Halle, Germany), A. Jenic (Ljubljana, Slovenia), A. Lachaud (Nantes, France), R. Moerman (Brussels, Belgium), O. Ponchau (Mons, Belgium), and M. Terzo (Mons, Belgium) for providing samples or for their help and support in sampling. The authors acknowledge the *Fonds pour la recherche dans l'industrie et l'agriculture* (FRIA) and *Fonds pour la Formation à la Recherche Fondamentale et Collective* (FNRS, FRFC 2.4613.10) for financial support. TL, PL and SD are FRIA grant holders. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 244090, STEP Project (Status and Trends of European Pollinators, [www.step-project.net](http://www.step-project.net)). MV is funded by the FNRS. Partial financial support (IV) was provided by the Academy of Sciences of the Czech Republic (subvention for development of research organization RVO: 61388963). This project was supported by the network *Bibliothèque du Vivant* funded by the CNRS, the *Muséum National d'Histoire Naturelle* (MNHN) and the *Institut National de la Recherche Agronomique* (INRA), and technically supported by the *Genoscope*. Special thanks are also due to N. Brasero (Mons, Belgium) and S. Lambert (Mons, Belgium) for helpful comments and discussions on earlier versions of this manuscript. We also thank three anonymous referees for their constructive and useful comments. Authors acknowledge D. Baldock (Milford, UK), the *Centre de Langues Vivantes* (University of Mons), David Notton (the Natural History Museum, London, UK), Stuart P.M. Roberts (University of Reading, Reading, UK) for correcting the English. Computational resources have been provided by the Interuniversity Scientific Computing Facility center (iSCF) of the University of Namur and the High Performance Computing Centre of the University of Brussels (HPC cluster “Hydra”).

#### Author details

<sup>1</sup>Laboratoire de Zoologie (Research Institute of Biosciences), University of Mons, Place du Parc 20, B-7000 Mons, Belgium. <sup>2</sup>Evolutionary Biology and Ecology, Université Libre de Bruxelles, Avenue FD Roosevelt 50, B-1050 Brussels, Belgium. <sup>3</sup>Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flamingovo nám 2, CZ-166 10 Prague, Czech Republic. <sup>4</sup>Institut national de la recherche agronomique, UMR 1062 Centre de Biologie pour la Gestion des Populations, CS 30 016, F-34988 Montpellier/Lez cedex, France.

Received: 14 June 2013 Accepted: 26 November 2013

Published: 2 December 2013

#### References

- Wallace A: **The cause of an ice age.** *Nature* 1896, **53**:220–221.
- Molengraaff G, Weber M: **On the relation between the pleistocene glacial period and the origin of the Sunda Sea (Java- and South China-Sea), and its influence on the distribution of coralreefs and on the land- and freshwater fauna.** *Proceeding Sect Sci* 1921, **23**:395–439.
- Reinig WF: *Die Holarktis. Ein Beitrag zur diluvialen und alluvialen Geschichte der zirkumpolaren Faunen- und Florengebiete.* Jena, Germany: Verlag von Gustav Fischer; 1937.

4. Avise JC: *Phylogeography: the history and formation of species*. Cambridge, MA: Harvard University Press; 2000.
5. Hewitt GM: **Genetic consequences of climatic oscillations in the quaternary**. *Philos Trans R Soc B* 2004, **359**:183–195.
6. Andersen BG, Borns HWJ: *The ice age world: an introduction to quaternary history and research with emphasis on North America and Northern Europe during the last 2.5 million*. Oxford, UK: Oxford University Press; 1994.
7. Bennett KD: *Evolution and ecology: the pace of life*. Cambridge, UK: Cambridge University Press; 1997.
8. Zagwijn WH: **The beginning of the ice age in Europe and its major subdivisions**. *Quat Sci Rev* 1992, **11**:583–591.
9. Stewart JR, Lister AM, Barnes I, Dalén L: **Refugia revisited: individualistic responses of species in space and time**. *Proc R Soc B Biol Sci* 2010, **277**:661–671.
10. Taberlet P: **Biodiversity at the intraspecific level: the comparative phylogeographic approach**. *J Biotechnol Genome Anal Chang Face Biotechnol* 1998, **64**:91–100.
11. Coyne JA, Orr HA: *Speciation*. Sunderland, MA: Sinauer; 2004.
12. Sobel JM, Chen GF, Watt LR, Schemske DW: **The biology of speciation**. *Evolution* 2010, **64**:295–315.
13. Blondel J: *Biogéographie: approche écologique et évolutive*. Paris: Mason; 1995.
14. Mayr E: *Systematics and the origin of species*. New York: Columbia University Press; 1942.
15. Hewitt GM: **Some genetic consequences of ice ages, and their role in divergence and speciation**. *Biol J Linn Soc* 1996, **58**:247–276.
16. Hewitt GM: **Speciation, hybrid zones and phylogeography - or seeing genes in space and time**. *Mol Ecol* 2001, **10**:537–549.
17. Excoffier L, Foll M, Petit RJ: **Genetic consequences of range expansions**. *Annu Rev Ecol Syst* 2009, **40**:481–501.
18. Bush GL: **Sympatric speciation in animals: new wine in old bottles**. *Trends Ecol Evol* 1994, **9**:285–288.
19. Foster SA: **The geography of behaviour: an evolutionary perspective**. *Trends Ecol Evol* 1999, **14**:190–195.
20. Boughman JW: **How sensory drive can promote speciation**. *Trends Ecol Evol* 2002, **17**:571–577.
21. Paterson HEH: *Evolution and the recognition concept of species*. Baltimore, MD: The Johns Hopkins University Press; 1993.
22. Gee JM: **IEG and behavioral response to conspecific and heterospecific calls in hybridizing avian species, California and Gambel's quail**. *Integr Comp Biol* 2005, **45**:1000.
23. Lecocq T, Vereecken NJ, Michez D, Dellicour S, Lhomme P, Valterová I, Rasplus J-Y, Rasmont P: **Patterns of genetic and reproductive traits differentiation in mainland vs. Corsican populations of bumblebees**. *PLoS One* 2013, **8**:e65642.
24. Ptáček MB: **The role of mating preferences in shaping interspecific divergence in mating signals in vertebrates**. *Behav Process* 2000, **51**:111–134.
25. Nosil P, Vines TH, Funk DJ: **Perspective: reproductive isolation caused by natural selection against immigrants from divergent habitats**. *Evolution* 2005, **59**:705–719.
26. West-Eberhard MJ: **Sexual selection, social competition, and speciation**. *Q Rev Biol* 1983, **58**:155–183.
27. Panhuis TM, Butlin R, Zuk M, Tregenza T: **Sexual selection and speciation**. *Trends Ecol Evol* 2001, **16**:364–371.
28. Vereecken NJ, Mant J, Schiestl FP: **Population differentiation in female sex pheromone and male preferences in a solitary bee**. *Behav Ecol Sociobiol* 2007, **61**:811–821.
29. Hines HM: **Historical biogeography, divergence times, and diversification patterns of bumble bees (Hymenoptera: Apidae: Bombus)**. *Syst Biol* 2008, **57**:58–75.
30. Reinig WF: **On the variation of *Bombus lapidarius* L. and its cuckoo, *Psithyrus rupestris* Fabr., with notes on mimetic similarity**. *J Genet* 1935, **30**:321–356.
31. Baer B: **Bumblebees as model organisms to study male sexual selection in social insects**. *Behav Ecol Sociobiol* 2003, **54**:521–533.
32. Bergström G, Svensson BG, Appelgren M, Groth I: **Complexity of bumble bee marking pheromones: biochemical, ecological and systematical interpretations**. *Syst Assoc* 1981, **19**:175–183.
33. Žáček P, Kalinová B, Šobotník J, Hovorka O, Ptáček V, Coppée A, Verheggen F, Valterová I: **Comparison of age-dependent quantitative changes in the male labial gland secretion of *Bombus terrestris* and *Bombus lucorum***. *J Chem Ecol* 2009, **35**:698–705.
34. Coppée A: ***Bombus terrestris* (L. 1758): A complex species or a species complex? - Intraspecific pheromonal and genetic variations of *Bombus terrestris* (L.), Impacts on the speciation**. PhD thesis. Université de Mons, Laboratoire de Zoologie; 2010.
35. Žáček P, Prchalová-Hornáková D, Tykva R, Kindl J, Vogel H, Svatoš A, Pichová I, Valterová I: **De Novo biosynthesis of sexual pheromone in the labial gland of bumblebee males**. *ChemBioChem* 2013, **14**:361–371.
36. Luxová A, Valterová I, Stránský K, Hovorka O, Svatoš A: **Biosynthetic studies on marking pheromones of bumblebee males**. *Chemoecology* 2003, **13**:81–87.
37. Brower AVZ: **Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution**. *Proc Natl Acad Sci USA* 1994, **91**:6491–6495.
38. Svensson BG, Bergström G: **Volatile marking secretions from the labial gland of north European *pyrobombus* D. T. males (Hymenoptera, Apidae)**. *Insectes Soc* 1977, **24**:213–224.
39. Bergman P: *Chemical communication in bumblebee premating behaviour*. PhD thesis. Göteborg University, Department of Chemical Ecology; 1997.
40. Cucci L: **Raised marine terraces in the northern Calabrian Arc (southern Italy): a 600 kyr-long geological record of regional uplift**. *Ann Geophys* 2004, **47**:1391–1406.
41. Hewitt GM: **Post-glacial re-colonization of European biota**. *Biol J Linn Soc* 1999, **68**:87–112.
42. Reinig WF: **Bastardierungs zonen und mischpopulationen bei Hummeln (*Bombus*) und Schmarotzerhummeln (*Psithyrus*) (Hymenopt., Apidae)**. *Mittl Mue Entomol Ges* 1969, **59**:1–89.
43. Petit RJ, Aguinagalde I, De Beaulieu J-L, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Müller-Starck G, Demesure-Musch B, Palmé A, Martín JP, Rendell S, Vendramin GG: **Glacial refugia: hotspots but not melting pots of genetic diversity**. *Science* 2003, **300**:1563–1565.
44. Stewart JR, Lister AM: **Cryptic northern refugia and the origins of the modern biota**. *Trends Ecol Evol* 2001, **16**:608–613.
45. Martini I, Sagri M, Colella A: **Neogene-Quaternary basins of the inner Apennines and Calabria arc**. In *Anat an Orogen Apennines Adjac Mediterr basins*. Edited by Martini I. Dordrecht, Netherlands: Kluwer Academic Publishers; 2001:375–400.
46. Strahler A, Strahler A: *Elements of physical geography*. 3rd edition. New York, NY: John Wiley & Sons; 1989.
47. Michaux JR, Libois R, Filippucci M-G: **So close and so different: comparative phylogeography of two small mammal species, the yellow-necked fieldmouse (*Apodemus flavicollis*) and the woodmouse (*Apodemus sylvaticus*) in the western Palearctic region**. *Heredity* 2005, **94**:52–63.
48. Fritz U, D'Angelo S, Pennisi MG, LoValvo M: **Variation of Sicilian pond turtles, *Emys trinacris* - what makes a species cryptic?** *Amphibia-Reptilia* 2006, **27**:513–529.
49. Mac Arthur R, Wilson E: *The theory of island biogeography*. Princeton, NJ: Princeton University Press; 1967.
50. Hewitt GM: **The genetic legacy of the quaternary ice ages**. *Nature* 2000, **405**:907–913.
51. Joger U, Fritz U, Guicking D, Kalyabina-Hauf S, Nagy ZT, Wink M: **Phylogeography of western Palaeartic reptiles - spatial and temporal speciation patterns**. *Zool Anz* 2007, **246**:293–313.
52. Mattoccia M, Marta S, Romano A, Sbordoni V: **Phylogeography of an Italian endemic salamander (genus *Salamandrina*): glacial refugia, postglacial expansions, and secondary contact**. *Biol J Linn Soc* 2011, **104**:903–922.
53. Gause GF: **Experimental studies on the struggle for existence: 1. Mixed population of two species of yeast**. *J Exp Biol* 1932, **9**:389–402.
54. Hardin G: **The competitive exclusion principle**. *Science* 1960, **131**:1292–1297.
55. Andersson M: *Sexual selection*. Princeton, USA: Princeton University Press; 1994.
56. Wyatt TD: *Pheromones and animal behaviour: communication by smell and taste*. Cambridge, UK: Cambridge University Press; 2003.
57. Loftus-Hills JJ, Littlejohn MJ: **Reinforcement and reproductive character displacement in *Gastrophryne carolinensis* and *G. olivacea* (Anura: Microhylidae): a reexamination**. *Evolution* 1992, **46**:896–906.
58. Symonds MRE, Moussalli A, Elgar MA: **The evolution of sex pheromones in an ecologically diverse genus of flies**. *Biol J Linn Soc* 2009, **97**:594–603.
59. Clearwater JR, Foster SP, Muggleston SJ, Dugdale JS, Priesner E: **Intraspecific variation and interspecific differences in sex pheromones of sibling**

- species in *Ctenopseustis obliquana* complex. *J Chem Ecol* 1991, **17**:413–429.
60. Blyth JE, Lachaise D, Ritchie MG: Divergence in multiple courtship song traits between *Drosophila santomea* and *D. yakuba*. *Ethology* 2008, **114**:728–736.
61. Förschler MI, Kalko EKV: Geographical differentiation, acoustic adaptation and species boundaries in mainland citril finches and insular Corsican finches, superspecies *Carduelis citrinella*. *J Biogeogr* 2007, **34**:1591–1600.
62. Löfstedt C: Moth pheromone genetics and evolution. *Philos Trans R Soc B* 1993, **340**:167–177.
63. Takanashi T, Huang Y, Takahasi KR, Hoshizaki S, Tatsuki S, Ishikawa Y: Genetic analysis and population survey of sex pheromone variation in the adzuki bean borer moth, *Ostrinia scapularis*. *Biol J Linn Soc* 2005, **84**:143–160.
64. Roelofs WL, Liu W, Hao G, Jiao H, Rooney AP, Linn CE Jr: Evolution of moth sex pheromones via ancestral genes. *Proc Natl Acad Sci USA* 2002, **99**:13621–13626.
65. Templeton AR: Mechanisms of speciation: a population genetics approach. *Annu Rev Ecol Syst* 1981, **12**:23–48.
66. Dobzhansky T: Speciation as a stage in evolutionary divergence. *Am Nat* 1940, **74**:312–321.
67. Noor MAF: Reinforcement and other consequences of sympatry. *Heredity* 1999, **83**:503–508.
68. Higgie M, Chenoweth S, Blows MW: Natural selection and the reinforcement of mate recognition. *Science* 2000, **290**:519–521.
69. Lukhtanov VA: Dobzhansky's rule and reinforcement of prezygotic reproductive isolation in zones of secondary contact. *Biol Bull Rev* 2011, **1**:2–12.
70. Delmas R: Contribution à l'étude de la faune française des Bombinae (Hymenoptera, Apoidea, Bombidae). *Ann Soc Entomol Fr* 1976, **12**:247–290.
71. Williams PH: The distribution of bumblebee colour patterns worldwide: possible significance for thermoregulation, crypsis, and warning mimicry. *Biol J Linn Soc* 2007, **92**:97–118.
72. Hines HM, Williams PH: Mimetic colour pattern evolution in the highly polymorphic *Bombus trifasciatus* (Hymenoptera: Apidae) species complex and its comimics. *Zool J Linn Soc* 2012, **166**:805–826.
73. Tóth M, Löfstedt C, Blair BW, Cabello T, Farag AI, Hansson BS, Kovalev BG, Maini S, Nesterov EA, Pajor I, Sazonov AP, Shamshev IV, Subchev M, Szöcs G: Attraction of male turnip moths *Agrotis segetum* (Lepidoptera: Noctuidae) to sex pheromone components and their mixtures at 11 sites in Europe, Asia, and Africa. *J Chem Ecol* 1992, **18**:1337–1347.
74. Santucci F, Emerson BC, Hewitt GM: Mitochondrial DNA phylogeography of European hedgehogs. *Mol Ecol* 1998, **7**:1163–1172.
75. Avise JC: What is the field of biogeography, and where is it going? *Taxon* 2004, **53**:893–898.
76. Coppée A, Mathy T, Cammaerts M-C, Verheggen FJ, Terzo M, Iserbyt S, Valterová I, Rasmont P: Age-dependent attractivity of males' sexual pheromones in *Bombus terrestris* (L.) [Hymenoptera, Apidae]. *Chemoecology* 2011, **21**:75–82.
77. Canestrelli D, Nascetti G: Phylogeography of the pool frog *Rana (Pelophylax) lessonae* in the Italian peninsula and Sicily: multiple refugia, glacial expansions and nuclear-mitochondrial discordance. *J Biogeogr* 2008, **35**:1923–1936.
78. Canestrelli D, Sacco F, Nascetti G: On glacial refugia, genetic diversity, and microevolutionary processes: deep phylogeographical structure in the endemic newt *Lissotriton italicus*. *Biol J Linn Soc* 2012, **105**:42–55.
79. Nagy ZT, Joger U, Guicking D, Amann T, Wink M: Phylogeography of the European whip snake, *Coluber (Hierophis) viridiflavus* Lacépède, 1789, inferred from nucleotide sequences of the mitochondrial cytochrome b gene and ISSR genomic fingerprinting. *Biota* 2002, **3**:109–118.
80. Lecocq T, Lhomme P, Michez D, Dellicour S, Valterová I, Rasmont P: Molecular and chemical characters to evaluate species status of two cuckoo bumblebees: *Bombus barbutellus* and *Bombus maxillosus* (Hymenoptera, Apidae, Bombini). *Syst Entomol* 2011, **36**:453–469.
81. De Queiroz K: Species concepts and species delimitation. *Syst Biol* 2007, **56**:879–886.
82. De Meulemeester T, Gerbaux P, Boulvin M, Coppée A, Rasmont P: A simplified protocol for bumble bee species identification by cephalic secretion analysis. *Insectes Soc* 2011, **58**:227–236.
83. Cameron SA, Hines HM, Williams PH: A comprehensive phylogeny of the bumble bees (*Bombus*). *Biol J Linn Soc* 2007, **91**:161–188.
84. Pedersen BV: European bumblebees (Hymenoptera: Bombini) - phylogenetic relationships inferred from DNA sequences. *Insect Syst Evol* 2002, **33**:361–386.
85. Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P: Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann Entomol Soc Am* 1994, **87**:652–701.
86. Mardulyn P, Cameron SA: The major opsin in bees (Insecta: Hymenoptera): a promising nuclear gene for higher level phylogenetics. *Mol Phylogenet Evol* 1999, **12**:168–176.
87. Hines HM, Cameron SA, Williams PH: Molecular phylogeny of the bumble bee subgenus *Pyrobombus* (Hymenoptera: Apidae: *Bombus*) with insights into gene utility for lower-level analysis. *Invertebr Syst* 2006, **20**:289–303.
88. Zhang Z, Schwartz S, Wagner L, Miller W: A greedy algorithm for aligning DNA sequences. *J Comput Biol* 2000, **7**:203–214.
89. Katoh K, Misawa K, Kuma K-I, Miyata T: MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002, **30**:3059–3066.
90. Maddison W, Maddison D: *Mesquite: a modular system for evolutionary analysis*. 2007. Version 2.75. 2011. Available from <http://mesquiteproject.org>.
91. Swofford DL: *PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods)*. Version 4. Sunderland: Sinauer Associates; 2001.
92. Bandelt H-J, Forster P, Röhl A: Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 1999, **16**:37–48.
93. Farris JM, Källersjö M, Kulge AG, Bult AC: Testing significance of incongruence. *Cladistics* 1994, **10**:315–319.
94. Barker FK, Lutzoni FM: The utility of the incongruence length difference test. *Syst Biol* 2002, **51**:625–637.
95. Posada D: jModelTest: phylogenetic model averaging. *Mol Biol Evol* 2008, **25**:1253–1256.
96. Hurvich CM, Tsai C-L: Regression and time series model selection in small samples. *Biometrika* 1989, **76**:297–307.
97. Felsenstein J: Phylogenies and the comparative method. *Am Nat* 1985, **125**:1–15.
98. Zwickl DJ: *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criteria*. PhD thesis. University of Texas, School of Biological Sciences; 2006.
99. Hillis DM, Bull JJ: An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst Biol* 1993, **42**:182–192.
100. Ronquist F, Huelsenbeck JP: MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* 2003, **19**:1572–1574.
101. Huelsenbeck JP, Rannala B: Frequentist properties of bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Syst Biol* 2004, **53**:904–913.
102. Rambaut A, Drummond AJ: *Tracer version 1.4*. <http://beast.bio.ed.ac.uk/Tracer>.
103. Wilcox TP, Zwickl DJ, Heath TA, Hillis DM: Phylogenetic relationships of the dwarf boas and a comparison of bayesian and bootstrap measures of phylogenetic support. *Mol Phylogenet Evol* 2002, **25**:361–371.
104. Duennes MA, Lozier JD, Hines HM, Cameron SA: Geographical patterns of genetic divergence in the widespread Mesoamerican bumble bee *Bombus ephippiatus* (Hymenoptera: Apidae). *Mol Phylogenet Evol* 2012, **64**:219–231.
105. Drummond AJ, Rambaut A: BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007, **7**:214.
106. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A: Relaxed phylogenetics and dating with confidence. *PLoS Biol* 2006, **4**:699–710.
107. Excoffier L, Smouse PE, Quattro JM: Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 1992, **131**:479–491.
108. Excoffier L, Lischer HEL: Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 2010, **10**:564–567.
109. Dupanloup I, Schneider S, Excoffier L: A simulated annealing approach to define the genetic structure of populations. *Mol Ecol* 2002, **11**:2571–2581.
110. Dellicour S, Mardulyn P: SPADS 1.0: a toolbox to perform spatial analyses on DNA sequence datasets. *Mol Ecol Resour* 2013. in press.
111. Nei M, Tajima F: Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. *Genetics* 1983, **105**:207–217.
112. Nei M: *Phylogeography: the history and formation of species*. New York, NY: Columbia University Press; 1987.
113. Rozas J, Rozas R: DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. *Comput Appl Biosci* 1997, **13**:307–311.



114. Miller MP: **Alleles In Space (AIS): computer software for the joint analysis of interindividual spatial and genetic information.** *J Hered* 2005, **96**:722–724.
115. Watson DF, Phillips GM: **A refinement of inverse distance weighted interpolation.** *Geo-processing* 1985, **2**:315–327.
116. Watson DF: *Contouring: a guide to the analysis and display of spatial data.* New York, NY: Pergamon Press; 1992.
117. Manni F, Guérard E, Heyer E: **Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm.** *Hum Biol* 2004, **76**:173–190.
118. Terzo M, Valterová I, Rasmont P: **Atypical secretions of the male cephalic labial glands in bumblebees: the case of *Bombus (Rhodobombus) mesomelas* Gerstaecker(Hymenoptera, Apidae).** *Chem Biodivers* 2007, **4**:1466–1471.
119. Dericour S, Lecocq T: **GALIGNER 1.0: an alignment program to compute a multiple sample comparison data matrix from large eco-chemical datasets obtained by gas chromatography.** *J Sep Sci* 2013, **36**:3206–3209.
120. R Development Core Team: *R: A language and environment for statistical computing.* Vienna, Austria; 2012.
121. Goslee SC, Urban DL: **The ecodist package for dissimilarity-based analysis of ecological data.** *J Stat Softw* 2007, **22**:1–19.
122. Oksanen FJ, Blanchet G, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H: *Vegan: community ecology package,* R package version 2.0-2. 2011. <http://CRAN.R-project.org/package=vegan>.
123. Anderson MJ: **Distance-based tests for homogeneity of multivariate dispersions.** *Biometrics* 2006, **62**:245–253.
124. Dufrêne M, Legendre P: **Species assemblages and indicator species: the need for a flexible asymmetrical approach.** *Ecol Monogr* 1997, **67**:345–366.
125. Mantel N: **The detection of disease clustering and a generalized regression approach.** *Cancer Res* 1967, **27**:209–220.
126. Ersts PJ: *Geographic distance matrix generator(version 1.2.3).* American Museum of Natural History, Center of Biodiversity and Conservation; 2012. Available from [http://biodiversityinformatics.amnh.org/open\\_source/gdmg](http://biodiversityinformatics.amnh.org/open_source/gdmg).
127. Dobson AJ: *An introduction to generalized linear models.* 2nd edition. Boca Raton, Florida: CRC Press; 2001.

doi:10.1186/1471-2148-13-263

**Cite this article as:** Lecocq *et al.*: Scent of a break-up: phylogeography and reproductive trait divergences in the red-tailed bumblebee (*Bombus lapidarius*). *BMC Evolutionary Biology* 2013 **13**:263.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

