The LINNEAN SOCIETY of London BIOLOGICAL Journal of the Linnean Society

Biological Journal of the Linnean Society, 2014, **113**, 136–148. With 4 figures

Signature of post-glacial expansion and genetic structure at the northern range limit of the speckled wood butterfly

JEAN-LUC TISON[1,2]*, VERONICA NYSTRÖM EDMARK[2], EDSON SANDOVAL-CASTELLANOS[2,3], HANS VAN DYCK[4], TOOMAS TAMMARU[5], PANU VÄLIMÄKI[6], LOVE DALÉN[2] and KARL GOTTHARD[3]

[1]*Department of Molecular Biosciences, The Wenner-Gren Institute, Stockholm University, 106 91 Stockholm, Sweden*
[2]*Department of Bioinformatics and Genetics, Swedish Museum of Natural History, SE-10405 Stockholm, Sweden*
[3]*Department of Zoology, Stockholm University, SE-106 91 Stockholm, Sweden*
[4]*Behavioural Ecology and Conservation Group, Biodiversity Research Centre, Earth and Life Institute, Université Catholique de Louvain (UCL), Croix du Sud 4-5, bte. L7.07.04, 1348 Louvain-la-Neuve, Belgium*
[5]*Department of Zoology, Institute of Ecology and Earth Sciences, University of Tartu, Vanemuise 46, EE-51014, Tartu, Estonia*
[6]*Department of Biology, University of Oulu, P.O. Box 3000, FI-90014 Oulu, Finland*

Received 27 February 2014; revised 2 April 2014; accepted for publication 2 April 2014

The post-glacial recolonisation of northern Europe has left distinct signatures in the genomes of many organisms, both due to random demographic processes and divergent natural selection. However, information on differences in genetic variation in conjunction with patterns of local adaptations along latitudinal gradients is often lacking. In this study, we examine genetic diversity and population structure in the speckled wood butterfly *Pararge aegeria* in northern Europe to investigate the species post-glacial recolonisation history and discuss how this may have affected its life-history evolution. We collected 209 samples and analysed genetic variation in nine microsatellite loci. The results demonstrated a more pronounced population structure in northern Europe compared with populations further south, as well as an overall decrease in genetic diversity with latitude, likely due to founder effects during the recolonisation process. Coalescent simulations coupled with approximate Bayesian computation suggested that central Scandinavia was colonised from the south, rather than from the east. In contrast to further south, populations at the northern range margin are univoltine expressing only one generation per year. This suggests either that univoltinism evolved independently on each side of the Baltic Sea, or that bivoltinism evolved in the south after northern Europe was recolonised. © 2014 The Linnean Society of London, *Biological Journal of the Linnean Society*, 2014, **113**, 136–148.

ADDITIONAL KEYWORDS: Bayesian analyses – coalescent simulations – demographic inference – microsatellites – *Pararge aegeria* – population structure – post-glacial recolonisation.

INTRODUCTION

The distribution and abundance of organisms change over time and space. At a large scale, these changes are due to geological events such as vicariance and climate change, and on a smaller scale due to habitat change and adaptation. Among the most important historical factors that have influenced the distribution and abundance of temperate organisms are the glacial cycles (Taberlet *et al.*, 1998; Hewitt, 2001; Slatkin & Excoffier, 2012). During warmer interglacial periods,

*Corresponding author. E-mail: jean-luc.tison@nrm.se

organisms that were constrained in glacial refugia have expanded polewards to recolonise previously glaciated regions (Hewitt, 1999; Stewart *et al.*, 2010). Post-glacial recolonisation events have therefore been recurrent and accompanied by population range shifts or expansions. Accordingly, many temperate species display genetic signatures of post-glacial demographic expansions (Lessa, Cook & Patton, 2003). The genetic consequences of range expansions typically involve changes in both genetic diversity and genetic population structure (Excoffier & Ray, 2008; Excoffier, Foll & Petit, 2009; Arenas *et al.*, 2012). After post-glacial recolonisation events to the north, due to series of founder effects and repeated bottlenecks, populations frequently display a decrease in genetic diversity outwards of the origin and higher genetic differentiation (Hewitt, 2001; Ramachandran *et al.*, 2005). Often this process can reduce the evolutionary potential of the population.

During the post-glacial recolonisation process, organisms are subjected to differential selection because of the latitudinal clinal variation in seasonality (Kivelä, Valimaki & Gotthard, 2013). To deal with such new selective pressures, populations can either respond through adaptation or phenotypic plasticity. Exploring genetic diversity at the population level, especially at the margins of species distributions where range expansions take place, has become increasingly important in the face of anthropogenic environmental change. Current global climate change has already induced significant range expansions in several species of different taxonomic groups and further alterations are predicted but possible population genetic implications are insufficiently understood (Parmesan, 2006; Hellmann *et al.*, 2008; Pöyry *et al.*, 2009).

Butterfly species are suitable models to study evolutionary processes associated with range shifts, as good distribution data are available and their life cycles are well known (Hill, Thomas & Huntley, 1999; Parmesan *et al.*, 1999). The speckled wood butterfly, *Pararge aegeria* (Linnaeus, 1758), is a particularly suitable species in studying recolonisation and range expansions (Hill *et al.*, 2006). Today, the species is found throughout Europe, Asia Minor, North Africa, Russia, and Central Asia. Being a species typically found in woodland habitats (Carter & Hargreaves, 1986), it is likely that *P. aegeria* was confined to refugia in southern Europe and Asia during the Late Pleistocene and expanded northwards during the Holocene (Habel *et al.*, 2013). In recent years, *P. aegeria* has also been observed to expand even further northward. For example, in the United Kingdom, the range of *P. aegeria* has shifted to the north in concert with recent climate change (Parmesan, 2006), leading to a substantial shift in the species northern range margin (Hill *et al.*, 1999; Hughes, Hill & Dytham, 2003). Moreover, the distribution and abundance of *P. aegeria* has increased recently within its European core range, including the Netherlands and Belgium. The species also occurs in Scandinavia and Finland, which constitutes its northern range margin in Europe. In Sweden, recent changes in distribution of *P. aegeria* are well described (Nordström, 1955; Henriksen & Kreutzer, 1982). In central Sweden, the species has been resident for more than 250 years, whereas in southern Sweden there are no records of the species until the 1930s, when recolonisation presumably took place either from Denmark in the south or from a previously undetected local population. The species is absent or at least very rare in 50–100 km wide band in southern Sweden suggesting that the southern and central Swedish distributions are still geographically separated (Eliasson *et al.*, 2005). The observation that the population in central Sweden seems to have been established prior to the one in southern Sweden could indicate that the former was established through recolonisation from the east.

Previous genetic work on *P. aegeria* has shown that populations in central Western Europe display a decrease in genetic variation with latitude attributed to the effect of post-glacial recolonisation (Hill *et al.*, 2006; Vandewoestijne & Van Dyck, 2010; Hill, Griffiths & Thomas, 2011). Despite this, genetic differentiation among populations in central Western Europe as well as at the southern range margin in North Africa is negligible (Habel *et al.*, 2013), likely due to high gene flow among populations (Buckley, Butlin & Bridle, 2012; Habel *et al.*, 2013).

P. aegeria has a complex life cycle, where variation in seasonal factors (e.g. photoperiod and temperature) lead to the expression of different developmental pathways (Shreeve, 1986; Nylin, Wickman & Wiklund, 1989; Gotthard, 2004). Three major developmental pathways are known: (1) direct development through the larval and pupal stage into reproductive adult within the same season, (2) winter diapause in the pupal stage, and (3) winter diapause in the larval stage (Aalberg Haugen, Berger & Gotthard, 2012; Gotthard, 2004; Gotthard & Berger, 2010; Nylin *et al.*, 1989; Shreeve, 1986; Van Dyck & Wiklund, 2002; Wiklund & Friberg, 2011). The expression of these different developmental pathways varies with latitude; populations at the northernmost range margin on both sides of the Baltic Sea display winter diapause in the pupal stage, and are generally univoltine with only one generation per year (Seppänen, 1969; Nylin *et al.*, 1989; Gotthard & Berger, 2010; Aalberg Haugen *et al.*, 2012). Populations further south display all three developmental pathways and have at least two generations per year, i.e. are bi- or multivoltine (Shreeve, 1986; Nylin *et al.*,

1989; Gotthard & Berger, 2010; Aalberg Haugen *et al.*, 2012). While the life history of *P. aegeria* in Scandinavia is well known, its recolonisation history and the genetic relationships among populations in relation to latitudinal variation in life histories remains unresolved (Weingartner, Wahlberg & Nylin, 2006).

In this study, we analysed microsatellite diversity in *P. aegeria* populations in northern Europe (Fig. 1). The overall aim was to investigate the species post-glacial recolonisation history and to compare genetic diversity and population structure in relation to latitude and voltinism. Specifically, we tested the two principal hypotheses that central Scandinavia was either recolonised from east (i.e. Finland) or alternatively from the south (i.e. Denmark). Secondly, we expected to find a decreasing pattern of genetic diversity with latitude reflecting the species post-glacial recolonisation history. Furthermore, the presence of local adaptations in life history could indicate a low degree of gene flow among differently adapted populations. We therefore expected a relatively high genetic differentiation among bi- and univoltine populations.

MATERIAL AND METHODS

Sampling and DNA analysis

Individuals for DNA analysis were wild caught from Belgium ($N = 16$), the Netherlands ($N = 8$), Estonia ($N = 11$), Latvia ($N = 1$) pooled with Estonia samples afterwards, Denmark ($N = 19$), Finland ($N = 40$), and six localities in Sweden (Skåne, $N = 23$; Öland, $N = 17$; Gotland, $N = 23$; Småland, $N = 4$; Stockholm, $N = 20$; Sundsvall, $N = 27$) between 1984 and 2011 (Fig. 1). DNA was extracted from one to three legs from each individual using the Molestrips DNA tissue kit (GeneMole). Nine microsatellite loci developed for *P. aegeria* (*Pae2*, *Pae3*, *Pae4*, *Pae7*, *Pae11*, *Pae16*, *Pae17*, *Pae19* and *Pae20*; Abdoullaye *et al.*, 2010) were subsequently amplified in three multiplexes using fluorescently labelled primers (Applied Biosystems, Table 1). Some of these loci contain single base pair insertions and deletions in the flanking regions leading to one-nucleotide step deviations from the repeat motif (Abdoullaye *et al.*, 2010). Polymerase chain reactions (PCRs) were performed in 12-μl volumes, each containing 1–2 μl DNA extract, 0.1–

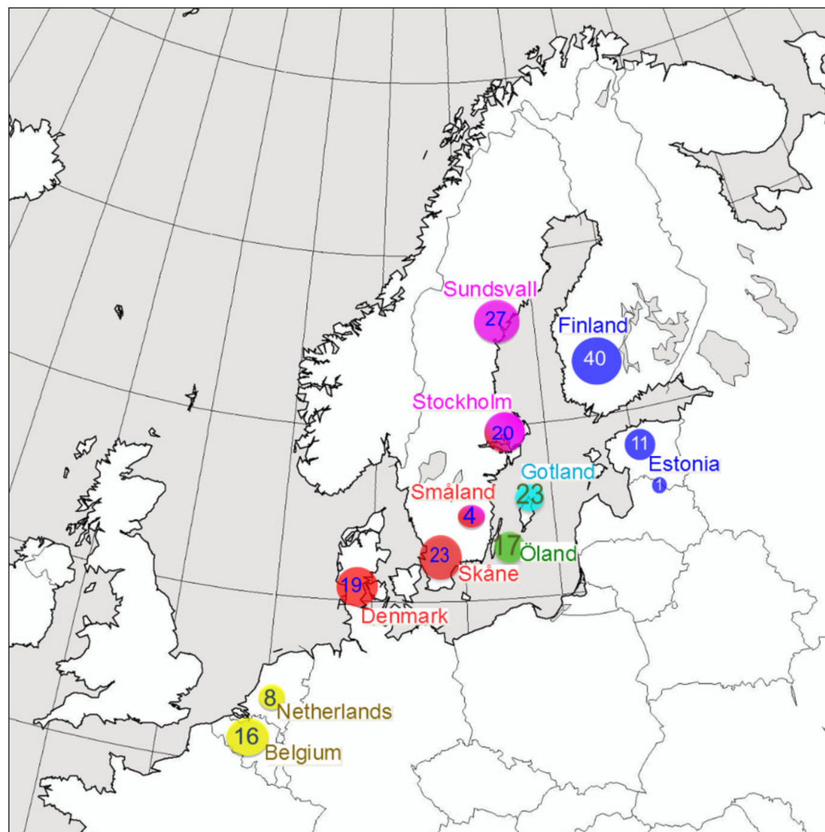

Figure 1. Geographic distribution of the 209 samples of *Pararge aegeria* included in the study. The locations are illustrated with coloured circles matching the colours of the six clusters identified by STRUCTURE software presented in Figure 4. The numbers indicate the sampling sizes.

© 2014 The Linnean Society of London,

Table 1. Conditions used for PCR amplification, and levels of genetic diversity per locus. The primer annealing temperature T_a is given, as well as the primer concentrations (in μM). Microsatellite loci were pooled in different multiplexes (A–C). Genetic diversity indices include the number of alleles (*nA*), observed heterozygosity (H_o), expected heterozygosity (H_e), polymorphism information content (PIC) and allelic richness (*AR*)

Locus	Multiplex	T_a	Primer conc. [μM]	Sample sizes	(*nA*)	H_o	H_e	PIC	*AR*
Pae2	A	58 °C	0.3	205	15	0.737	0.833	0.812	4.11
Pae3	A	58 °C	0.3	206	15	0.738	0.806	0.78	3.90
Pae11	A	58 °C	0.2	206	23	0.699	0.853	0.837	4.32
Pae17	A	58 °C	0.2	206	19	0.66	0.768	0.744	3.74
Pae4	B	60 °C	0.2	205	25	0.805	0.872	0.859	4.51
Pae19	B	60 °C	0.2	206	9	0.107	0.108	0.107	1.33
Pae20	B	60 °C	0.2	206	6	0.311	0.357	0.33	2.00
Pae16	C	58 °C	0.4	205	15	0.351	0.475	0.447	2.43
Pae7	C	58 °C	0.1	205	13	0.512	0.577	0.493	2.46
Mean					*15.6*	*0.547*	*0.628*	*0.601*	*3.20*

0.4 μM of each primer, and 6 μl of 2× Qiagen Multiplex PCR Master Mix (Qiagen). PCR thermal cycling conditions were: 15 min denaturation at 95 °C, 30–35 cycles of 30 s. denaturation at 94 °C, 90 s. annealing at 58 °C or 60 °C and 60 s. extension at 72 °C, followed by a single 30–45 min extension step at 60 °C (Table 1). The PCR products were run on an ABI 3130xl sequencer and the fragments were sized with GeneScan 500 LIZ size standard (Applied Biosystems) and analysed using GeneMapper v4.0 (Applied Biosystems). All microsatellite amplifications and genotyping were performed at the Swedish Museum of Natural History. The precise sampling location of the samples and the genotypes identified in this study are available in Supporting Information (Table S5).

Recolonisation history

To infer the population history of *P. aegeria* we used an approximate Bayesian computation (ABC) approach (Csillery *et al*., 2010; Aeschbacher, Futschik & Beaumont, 2013) implemented in DIYABC v2.0 (Cornuet *et al*., 2014). As the main aim of these analyses was to investigate the origin of *P. aegeria* in mainland Scandinavia, we excluded the island populations of Öland and Gotland from these simulations. Thus, four populations identified in the population structure analysis (see below) were included in the ABC analyses. Three different recolonisation scenarios were evaluated (Fig. 2). The first scenario was designed to fit with a recolonisation from the south (i.e. Benelux) along both the eastern and western sides of the Baltic Sea, with successive colonisation of Latvia, Estonia and Finland on one route, and Denmark, South Scandinavia and central Scandinavia on the other route. In the second scenario, the Central Scandinavian population originated from the east. The third scenario was based on the hypothesis that Scandinavia was recolonised in two waves: one ancient establishment of the central Scandinavian population followed by a second, more recent wave of expansion into south Scandinavia. The analyses were set up following established guidelines from Beaumont (2010) and Csillery *et al*. (2010). Specifics of the parameter's priors of each scenario, including effective population size and time to splitting or merging of the populations can be found in Supporting Information (Table S1). ABC analysis employs a rejection procedure for selecting the simulations that produce datasets more similar to the observed one, as a way to approximate the posterior sample of the parameters conditioned to the data. The similarity among observed and simulated datasets is assessed by means of Euclidean distances taken from several summary statistics that are calculated and normalised from the datasets (Storz & Beaumont, 2002). The employed summary statistics (specific for the microsatellite data) were: the mean number of alleles across loci (*A*), the mean genetic diversity across loci (G), the mean allele size variance (*V*), the mean Garza–Williamson's M index (Garza & Williamson, 2001), the F_{ST} between two samples, the shared allele distance between two samples (DAS) and the $(\delta\mu)^2$ distance between two samples (Goldstein *et al*., 1995), see Supporting Information (Table S2). The mutation model used, as well as prior ranges and distributions for each parameter, are given in the Table S1 of the Supporting Information. For each scenario, one million datasets were simulated, and were subsequently added into a reference table. Following this, the normalised Euclidean distances between the

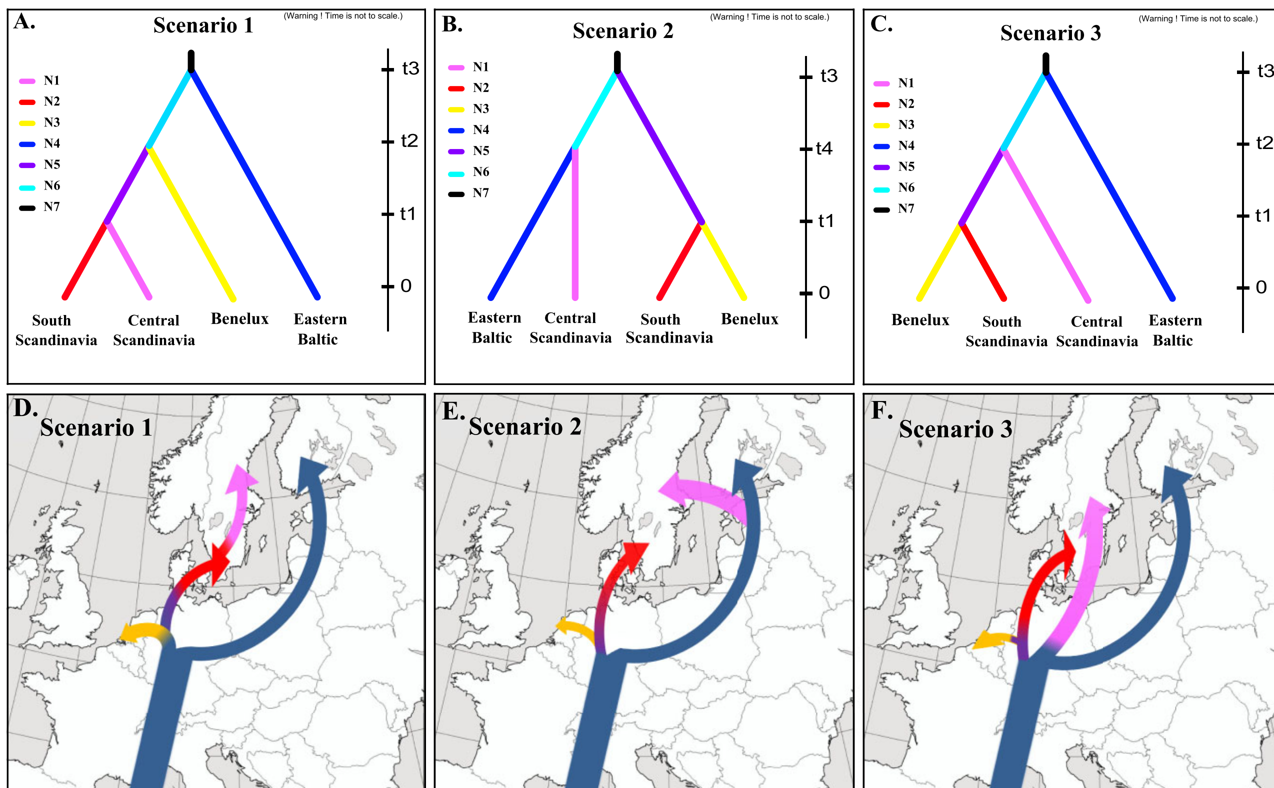

Figure 2. Schematic representations of three scenarios (A–F) to test recolonisation events for population differentiation in the north of Europe using the ABC approach (DIYABC v2.0). The populations mentioned correspond to the clusters identified by the software STRUCTURE, and the colours of the current populations branches match the colours presented in Figures 1 and 4. Ti corresponds to the time of event, in numbers of generation. Several conditions were considered: $T_2 > T_1$, $T_3 \geq T_2$, $T_3 > T_4$. N corresponds to effective population sizes of each group.

simulated datasets and the observed dataset were computed and the 1‰ closest simulated datasets were used to estimate the relative posterior probability (with 95% credible intervals) of each scenario, using the two algorithms implemented in the software DIYABC: the direct approach and the logistic regression. For estimating the involved demographic parameters, 10^6 additional simulations were run for the scenario that obtained the largest support in the scenarios comparison. For this analysis the program applied a local linear regression algorithm (Beaumont, Zhang & Balding, 2002) to the 1% of the closest simulations to the observed data. Bias, accuracy and precision of this procedure of parameter estimation was assessed by means of a cross-validation analysis, also implemented in DIYABC, in which 500 pseudo observed datasets (PODs) were employed to play the role of ‘real’ datasets. In a similar way, the statistical power for choosing each of the three recolonisation scenarios, was investigated by means of 500 PODs simulated for each model. This allows to evaluate the confidence in our model choice procedure, and was performed also with both the direct approach and the logistic regression (Fagundes *et al.*, 2007).

Standard population genetic parameters

To investigate if putative identical genotypes were present in our sample, we calculated the probability of finding two samples with identical genotypes (Taberlet & Luikart, 1999) prior to further population genetic analyses using CERVUS v.3.0.3 (Kalinowski, Taper & Marshall, 2007). The probability of identity was calculated using the equation described by Paetkau and Strobeck (Paetkau & Strobeck, 1994). Null alleles, large allelic dropout and scoring errors due to stutter peaks were investigated using MICROCHECKER v.2.2.0.3 (Dapporto *et al.*, 2011). Each microsatellite locus was tested for departure from the Hardy–Weinberg equilibrium for each geographical region separately (Guo, 2012). Linkage disequilibrium was tested using the likelihood-ratio test between each pair of loci for each geographical region

separately, as implemented in ARLEQUIN v3.5.1 (Excoffier & Lischer, 2010). Analyses of Hardy–Weinberg equilibrium were performed using a Markov chain with 10^7 steps and 10^6 dememorization steps. The test for linkage disequilibrium was computed using two initial conditions followed by 1000 permutations. The inbreeding coefficient, F_{IS} for each population was calculated using FSTAT 2.3.2 (Goudet, 2001). Genetic diversity was estimated as the number of alleles per locus (*nA*), observed heterozygosity (H_o), the unbiased expected heterozygosity (H_e) and the polymorphism information content (*PIC*) (Botstein *et al.*, 1980). The *PIC* is based on expected heterozygosity and the number of alleles per site and is representative of the diversity found at each locus. These values were estimated using ARLEQUIN v3.5.1. Allelic richness (*AR*) was calculated using FSTAT (Goudet, 2001).

Population structure

To estimate the most probable number of populations (*K*) in the dataset, a Bayesian clustering analysis was performed using STRUCTURE 2.3.4 (Pritchard, Stephens & Donnelly, 2000). Markov Chain Monte Carlo (MCMC) methods were conducted using the admixture and correlated allele frequency models but without using the sampling location as a prior (Hubisz *et al.*, 2009). The burn-in length was set at 10^4 steps, followed by 10^5 MCMC steps. Five independent runs were conducted for each value of *K* ranging between 1 and 11 to test the consistency of estimates of $P(X|K)$. The *K* value where the likelihood plateaued was chosen as the most probable number of populations (Pritchard *et al.*, 2000). To distinguish between $K = 5$ and $K = 6$ we ran an additional five independent runs for each *K*, where the burn-in length was set at 10^5 steps followed by 10^6 MCMC rep steps. We also inferred the best estimates of *K* as implemented in STRUCTURE harvester (Earl & Vonholdt, 2012) using the method of Evanno, Regnaut & Goudet (2005). We used CLUMPP v1.1.2 (Jakobsson & Rosenberg, 2007) to combine the outputs of the ten iterations from STRUCTURE for *the best K* using 10 000 random permutations of the computationally faster 'Greedy' algorithm. A visual output of the CLUMPP results was generated using the cluster visualization program DISTRUCT (Rosenberg, 2004). Genetic differentiations among the six populations identified as clusters in STUCTURE, measured as pairwise F_{ST} values, were calculated in ARLEQUIN and were tested for significance using 10^4 permutations (Excoffier & Lischer, 2010). The same analysis was also conducted among the 11 sampling locations.

Table 2. Posterior probabilities with 95% confidence intervals (in brackets) of each scenario in the ABC analyses. The posterior probabilities were measured using the 5×10^3 and 5×10^5 closest datasets for the direct approach and the logistic regression, respectively

Scenario	Posterior probabilities	
	Direct approach	Logistic regression
Sc1	46.0% (36.2–55.7)	92.7% (90.9–94.5)
Sc2	11.5% (05.2–17.7)	0.2% (0.0–0.3)
Sc3	42.6% (32.9–52.3)	7.2% (5.4–9.0)

RESULTS

Individual identity

All samples, except one, were successfully extracted, amplified and genotyped. The probability of identity (*PI*) across the nine microsatellites loci was calculated at 6.37×10^{-9}. The probability of identity for siblings (PI_{sib}), which represents the upper limit on the possible range of *PI* in a population, was estimated at 9.99×10^{-4}. In our dataset, we found two matching pairs of genotypes with $PI = 1.20 \times 10^{-6}$ ($PI_{sib} = 5.20 \times 10^{-3}$) and $PI = 2.26 \times 10^{-8}$ ($PI_{sib} = 2.80 \times 10^{-3}$), respectively. These pairs of samples were from Finland and were considered as likely being duplicates (i.e. the same specimen sampled twice). Therefore only one of each genotype was retained for further population genetic analysis. Thus, the final dataset was consisted of 206 genotypes.

Recolonisation history

The ABC performed in DIYABC v2.0 led to a successful testing of the three scenarios of recolonisation. The PCA performed on the first 10^5 simulated datasets of the reference table confirmed that our model was able to produce datasets similar to the observed one, as our observed dataset was surrounded by simulated datasets. The 1‰ closest simulated datasets and the posterior probability of the scenario choice analysis (Table 2) allowed us to reject scenario 2 (i.e. that the central Scandinavian population originated from Finland). However, we could not disentangle scenarios 1 and 3 (i.e. single or double wave of expansion from the south). Nonetheless, these results suggest that all of Scandinavia was colonised from the south (Benelux). The posterior probability densities of the effective population sizes and split times were wide and not very informative, see Supporting Information (Fig. S1). Overall, the statistical power to select the right scenario was high using both the direct approach (80.0%, 88.8% and 69.6% for scenarios 1, 2 and 3, respectively), and the

Table 3. Levels of genetic diversity per identified population, shown as the number of alleles (*nA)*, mean number of alleles per locus, the allelic richness, the observed heterozygosity (H_o), and the expected heterozygosity (H_e) as level of polymorphism index. The sample size (N) and the inbreeding coefficient F_{IS} are also given. An asterisk (*) indicates comparisons based only on those loci that were polymorphic

Population	N	*nA*	mean *nA*	*AR*	H_o	H_e	F_{IS}
Central Scandinavia	51	78	8.67	6.20	0.561	0.585	0.040
Gotland	23	56	6.22	5.74	0.580	0.569	−0.019
Öland	17	19	2.11	2.11	0.441*	0.389*	−0.139
South Scandinavia	42	74	8.22	6.66	0.622	0.636	0.023
Benelux	24	95	10.56	9.15	0.588	0.711	0.176
Eastern Baltic	49	64	7.11	5.31	0.519	0.601	0.138
Mean		*64.3*	*7.15*	*5.86*	*0.552*	*0.555*	
s.d.		*29.9*	*2.87*	*2.28*	*0.064*	*0.169*	

logistic regression (85.2%, 95.0% and 82.6% for scenarios 1, 2 and 3, respectively).

Genetic diversity, tests of HW genotypic proportions and linkage disequilibrium

The nine microsatellite loci had a high level of allelic diversity, 15.56 on average (range: 6–25, see Table 1). The average allelic richness per population was 5.86 (range: 2.11–9.15). The mean expected heterozygosity was 0.56 (C.I: 0.23–0.71), and the mean observed heterozygosity was 0.55 (C.I.: 0.44–0.59). The levels of genetic diversity for each identified population are presented in Table 3 (see Supporting Information Table S3 for area-specific information). The Öland population had a particularly low diversity ($AR = 1.61$, $H_o = 0.44$). Moreover, a decrease of allelic richness with latitude was detected when Öland was excluded (simple regression, $r^2 = 0.80$, $P = 0.0005$; Fig. 3). Twelve tests out of 99 showed deviations from the Hardy–Weinberg equilibrium. However, when sequential Bonferroni correction was applied to correct for multiple comparisons, none of the loci displayed a significant deviation from Hardy–Weinberg equilibrium (Rice, 1989). Linkage disequilibrium was observed in 29 out of 396 pairwise comparisons. Although slightly higher than expected by chance (19.8 with $P \leq 0.05$), no pattern of significant linkage disequilibrium was observed after Bonferroni correction, and there was no consistency between populations, suggesting the absence of physical linkage among the loci.

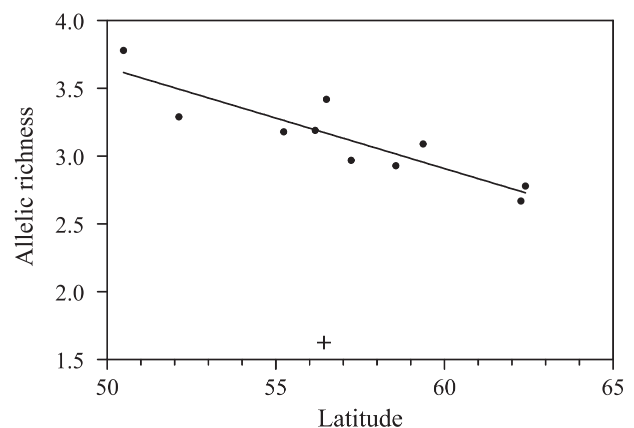

Figure 3. Allelic richness compared to latitude for each geographical region. The Öland value (+) was considered an outlier and was removed from the linear regression analysis.

Population structure

The Bayesian model-based clustering analyses performed in STRUCTURE indicated the presence of six distinct genetic clusters: $K = 6$; ln $P(X|K) = -5283.8$; and $\Delta K = 12.8$ (Fig. 4 and Supporting Information, Fig. S2). These six genetic clusters corresponded to the following sampling regions: [1] central Scandinavia (Sundsvall and Stockholm), [2] Gotland, [3] Öland, [4] South Scandinavia (Skåne-Denmark), [5] Benelux (the Netherlands and Belgium), [6] East Baltic (Finland, Estonia and Latvia). Samples from Småland ($N = 4$) seemed to be an admixture of central Scandinavian and southern Scandinavian genotypes (Fig. 4). F_{ST} values were statistically significant between most sampling locations for most of the pairwise comparisons (49 out of 55 comparisons were significant) and ranged from 0 to 0.37 (see Table 4 and Supporting Information, Table S4). In particular, Öland was highly differentiated from all other populations ($F_{ST} = 0.17$–0.37).

DISCUSSION

In this study, we used microsatellite markers to assess the genetic variability and genetic differentiation in *Pararge aegeria* in northern Europe, as well as

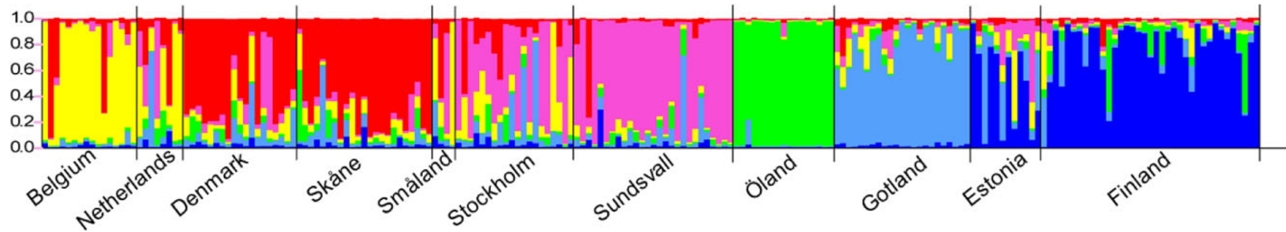

Figure 4. Population structure graphically displayed by DISTRUCT and estimated using the software STRUCTURE 2.3.4, assuming six genetic clusters (K = 6). The vertical bars represent inferred proportion of ancestry for each individual multi-locus genotype for each of the six clusters.

Table 4. Pairwise F_{ST} values per identified population. The F_{ST} values are shown in bold above the diagonal, and the corresponding *p*-value are given below the diagonal

	Central Scandinavia	Gotland	Öland	South Scandinavia	Benelux	Eastern Baltic
Central Scandinavia		**0.061**	**0.177**	**0.036**	**0.045**	**0.079**
Gotland	< 0.01		**0.197**	**0.041**	**0.076**	**0.077**
Öland	< 0.01	< 0.01		**0.166**	**0.213**	**0.160**
South Scandinavia	< 0.01	< 0.01	< 0.01		**0.025**	**0.059**
Benelux	< 0.01	< 0.01	< 0.01	< 0.01		**0.080**
Eastern Baltic	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	

to reconstruct the species post-glacial recolonisation history. Our results suggested that the post-glacial colonisation of central Scandinavia most likely took place from the south, via Denmark, rather than from the east. In addition, we found a significant decrease in genetic variation with latitude, as well as a marked genetic differentiation among locations.

Latitudinal gradient of genetic diversity

Within northern Europe, a significant decline in genetic variation with latitude was observed. This is in agreement with Vandewoestijne & Van Dyck (2010), suggesting that the post-glacial recolonisation of *Pararge aegeria* in Europe led to a gradual loss of genetic variation with increasing distance from the glacial refugium. This form of 'southern richness and northern purity' is a recurrent pattern among temperate species (Hewitt, 2004) and implies that recolonisations took place through a series of founder events as species expanded to the north. Most previous studies have focused on comparisons between southern refugia and northern previously glaciated regions (Lessa *et al*., 2003; Hewitt, 2004). It is therefore interesting that we found such a clear pattern on a comparatively limited geographical scale (from Belgium to central Sweden), where local populations displayed a gradually decreasing diversity towards the north (Fig. 3). The result of a latitudinal decrease of genetic diversity also seems to indicate that central Scandinavia was recolonised from one refugium. Indeed, in case of admixture in central Scandinavia of two recolonising populations with different refugial origins, a higher local genetic diversity than predicted from latitude alone should have been observed (Hewitt, 1999; Waits *et al*., 2000; Hill *et al*., 2011).

Population structure in the northern range margin

Recent studies have reported none or very little genetic differentiation among *P. aegeria* populations within central Europe and North Africa (Abdoullaye *et al*., 2010; Vandewoestijne & Van Dyck, 2010; Habel *et al*., 2013). In contrast, we found significant genetic differentiation among the populations in northern Europe, as indicated by the pairwise F_{ST} values (Table 4) and the exploration of population structure (Fig. 4). The cluster analysis indicated the existence of six genetically distinct populations: Benelux, southern Scandinavia, central Scandinavia, East Baltic, as well as each of the Baltic islands Öland and Gotland. The two islands were found to be particularly differentiated from the other populations, likely due to their relative isolation. Interestingly, the Gotland population had a high genetic diversity comparable with other populations at the same latitude, which could imply that the population was established a long time ago. A second explanation for the high diversity on Gotland could be that the island was

colonised from two different sources, leading to an admixture effect (Hewitt, 1999; Hill *et al*., 2011). However, our current genetic data does not have the statistical power required to test this hypothesis. In contrast to Gotland, the Öland population displayed very low genetic variation coupled with a high genetic differentiation. We suggest that this could be either because Öland population was established by a very small number of founders, or because one or several population bottlenecks have taken place after the initial colonisation.

It is intriguing that an island population so close to the mainland (Öland; c. 5 km) has such low genetic diversity. This suggests that even short stretches of open water could constitute considerable barriers to gene flow in *P. aegeria*. This is in clear contrast to the extensive gene flow among *P. aegeria* populations inhabiting oases in North Africa separated by up to hundreds of kilometres, where no genetic differentiation was detected (Habel *et al*., 2013). The geographic difference may simply arise from climatic factors; the Baltic Sea is cold after winter and thus effectively prevents individual movements when the first *P. aegeria* generation emerges. This offers interesting scope for future work on latitudinal variation in dispersal and functional connectivity (Van Dyck & Wiklund, 2002; Vandewoestijne & Van Dyck, 2011). Another explanation for a more pronounced population differentiation at the northern margin of temperate species is the post-glacial recolonisation events by themselves. Indeed, the range expansions may constitute a series of founder events, which may result in an increasing divergence among populations as the distance from the glacial refugium increases (Klopfstein, Currat & Excoffier, 2006; Eckert, Samis & Lougheed, 2008; Excoffier & Ray, 2008; Slatkin & Excoffier, 2012). This effect could also be further exacerbated if suitable habitats at the northern range margin are less abundant and more fragmented (Excoffier *et al*., 2009). Together, the apparently low gene flow among populations with the decreasing level of genetic diversity towards the north of the range implies that populations at the northern edge-of-range may have only limited potential to respond adaptively to environmental changes such as the ongoing climate change.

Recolonisation history

The coalescent-based analyses and approximate Bayesian computation (ABC) indicated that after the last Ice Age, northern Europe was recolonised along two routes, one on each side of the Baltic Sea. Thus, the population in central Scandinavia appears to originate from a colonisation event across the strait between Denmark and Sweden. Interestingly, historical records suggest that *P. aegeria* was absent from Denmark as well as southern Sweden until it reappeared in the early 20th Century (Nordström, 1955) whereas it has been present in central Sweden as long as there has been entomological records. Our coalescent-based ABC analysis could not resolve if the reappearance of *P. aegeria* in southern Sweden represents a second colonisation from Denmark, or if the population originates from a local, undetected source population (scenarios 1 and 3, respectively). However, based on the historical records that describe a nearly simultaneous, but sequential, appearance in Denmark and southern Sweden, it seems likely that this event represents a second recolonisation wave from further south (Nordström, 1955; Henriksen & Kreutzer, 1982). To differentiate between scenarios 1 and 3 in an ABC framework, it will be necessary to obtain a higher differentiation among the summary statistics employed in the analyses. This could be achieved either by increasing the number of microsatellite markers, or by analysing different types of genetic markers, evolving at different rates.

Voltinism

Based on the clustering analyses, we found no apparent relationship between voltinism and population genetic structure. Indeed, the populations that display bivoltinism (Benelux, South Scandinavia, Gotland and Öland) did not cluster together compared to the univoltine populations (central Scandinavia and East Baltic). The result from the ABC analysis, suggesting a recolonisation along two routes on each side of the Baltic Sea, could indicate that univoltinism evolved independently on both sides of the Baltic Sea. The evolution of univoltinism is necessarily due to evolutionary changes in the photoperiod dependent reaction norms that underlie the facultative induction of direct *vs*. diapause development (often expressed as the critical day length). Latitudinal adaptation in this reaction norm is indeed present in the Swedish distribution of *P. aegeria* so that individuals from northern Sweden enter diapause in daylengths where south Swedish individuals develop directly (Nylin, Wickman & Wiklund, 1995, Aalberg Haugen and Gotthard, submitted manuscript). However, the hypothesis of parallel evolution of univoltinsm on each side of the Baltic Sea relies on the assumption that the source population in the south was originally multivoltine. An alternative explanation is that the ancestral Pleistocene populations were univoltine and that southern populations evolved multivoltinism only after the initial recolonisation of northern Europe. In our view this second explanation appears less likely given the observation that *P. aegeria* populations at the

northern range margin can be experimentally induced to become bivoltine (Nylin *et al*., 1995; Aalberg Haugen *et al*., 2012). This implies that the genetic basis for multivoltinism, at least to some extent, existed in the ancestral southern populations.

In general, the demonstrated relatively high degree of isolation among local populations in northern Europe may have facilitated the evolution of local adaptations such as the life-cycle regulation underlying variation in voltinism (Nylin *et al*., 1989, 1995; Aalberg Haugen *et al*., 2012). However, it should be noted that we cannot exclude the possibility that gene flow between the univoltine populations in Finland and central Scandinavia has led to adaptive introgression of univoltinism from one population to the other, since a recent simulation study has showed that admixture versus pure divergence events generally require large data sets, i.e. 20 microsatellite loci (Sousa *et al*., 2012). An analysis of additional microsatellites markers or genome-wide single nucleotide polymorphisms could provide a way to investigate if such admixture between Finland and central Scandinavia has occurred The latter could potentially also provide an opportunity to identify genomic regions under divergent selection among differently adapted populations.

Overall, this study revealed a complex genetic structure for *P. aegeria* in northern Europe. Our results demonstrated an interesting pattern of a more pronounced population structure and a lower genetic variability at the northern range margin of *P. aegeria*, as compared to central Western Europe as well as at the species southern range margin in North Africa (Habel *et al*., 2013). We suggest that this pattern primarily is a consequence of the species post-glacial recolonisation of central and Northern Europe, which appears to have taken place along two routes, on each side of the Baltic Sea.

ACKNOWLEDGEMENTS

This work was supported by grants from the Swedish Research Council to KG and LD, as well as the strategic research programme EKOKLIM at Stockholm University, and by a research grant (ARC-grant 10/15-031 of the Académie Louvain) to HVD. TT was supported by the Estonian Research Council (IUT20-33) and by the European Union through the European Regional Development Fund (Center of Excellence FIBIR). We thank M. Olofsson, S. Vandewoestijne and several Finnish entomologists for help in collecting butterflies and three anonymous reviewers for constructive comments on a previous version of the manuscript. The authors have no conflicts of interest to declare.

REFERENCES

Aalberg Haugen IM, Berger D, Gotthard K. 2012. The evolution of alternative developmental pathways: footprints of selection on life-history traits in a butterfly. *Journal of Evolutionary Biology* **25:** 1377–1388.

Abdoullaye D, Acevedo I, Adebayo AA, Behrmann-Godel J, Benjamin RC, Bock DG, Born C, Brouat C, Caccone A, Cao L-Z, Casado-Amezúa P, Catanéo J, Correa-Ramirez MM, Cristescu ME, Dobigny G, Egbosimba EE, Etchberger LK, Fan B, Fields PD, Forcioli D, Furla P, Garcia de Leon FJ, García-Jiménez R, Gauthier P, Gergs R, González C, Granjon L, Gutiérrez-Rodríguez C, Havill NP, Helsen P, Hether TD, Hoffman EA, Hu X, Ingvarsson PK, Ishizaki S, Ji H, Ji XS, Jimenez ML, Kapil R, Karban R, Keller SR, Kubota S, Li S, Li W, Lim DD, Lin H, Liu X, Luo Y, Machordom A, Martin AP, Matthysen E, Mazzella MN, McGeoch MA, Meng Z, Nishizawa M, O'Brien P, Ohara M, Ornelas JF, Ortu MF, Pedersen AB, Preston L, Ren Q, Rothhaupt K-O, Sackett LC, Sang Q, Sawyer GM, Shiojiri K, Taylor DR, Van Dongen S, Van Vuuren BJ, Vandewoestijne S, Wang H, Wang JT, Wang L, Xu X-L, Yang G, Yang Y, Zeng YQ, Zhang Q-W, Zhang Y, Zhao Y, Zhou Y; Molecular Ecology Resources Primer Development Consortium. 2010. Permanent genetic resources added to *Molecular Ecology Resources* database 1 August 2009–30 September 2009. *Molecular Ecology Resources* **10:** 232–236. doi: 10.1111/j.1755-0998.2009.02796.x.

Aeschbacher S, Futschik A, Beaumont MA. 2013. Approximate Bayesian computation for modular inference problems with many parameters: the example of migration rates. *Molecular Ecology* **22:** 987–1002.

Arenas M, Ray N, Currat M, Excoffier L. 2012. Consequences of range contractions and range shifts on molecular diversity. *Molecular Biology and Evolution* **29:** 207–218.

Beaumont MA. 2010. Approximate Bayesian computation in evolution and ecology. In: Futuyma DJ, Shafer HB, Simberloff D, eds. *Annual review of ecology, evolution, and systematics, vol 41*. Palo Alto: Annual Reviews, 379–406.

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

Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of a genetic-linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32:** 314–331.

Buckley J, Butlin RK, Bridle JR. 2012. Evidence for evolutionary change associated with the recent range expansion of the British butterfly, Aricia agestis, in response to climate change. *Molecular Ecology* **21:** 267–280.

Carter D, Hargreaves B. 1986. *A field guide to caterpillars of butterflies and moths in Britain and Europe*. London: William Collins & Sons Ltd.

Cornuet JM, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, Marin JM, Estoup A. 2014. DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single

nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* **8:** 1187–1189.

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

Dapporto L, Schmitt T, Vila R, Scalercio S, Biermann H, Dinca V, Gayubo SF, Gonzalez JA, Lo Cascio P, Dennis RLH. 2011. Phylogenetic island disequilibrium: evidence for ongoing long-term population dynamics in two Mediterranean butterflies. *Journal of Biogeography* **38:** 854–867.

Earl DA, Vonholdt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4:** 359–361.

Eckert C, Samis K, Lougheed S. 2008. Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Molecular Ecology* **17:** 1170–1188.

Eliasson CU, Ryrholm N, Holmer M, Jilg K, Gärdefors U. 2005. *Nationalnyckeln till Sveriges flora och fauna. Fjärilar. Hesperiidae – Nymphalidae*. Uppsala: ArtDatabanken, SLU.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14:** 2611–2620.

Excoffier L, Foll M, Petit RJ. 2009. Genetic consequences of range expansions. *Annual Review of Ecology Evolution and Systematics* **40:** 481–501.

Excoffier L, Lischer H. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10:** 564–567.

Excoffier L, Ray N. 2008. Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution* **23:** 347–351.

Fagundes NJR, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L. 2007. Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 17614–17619.

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

Goldstein DB, Linares AR, Cavallisforza LL, Feldman MW. 1995. An evaluation of genetic distances for use with micrasatellite loci. *Genetics* **139:** 463–471.

Gotthard K. 2004. Growth strategies and optimal body size in temperate Pararginii butterflies. *Integrative and Comparative Biology* **44:** 471–479.

Gotthard K, Berger D. 2010. The diapause decision as a cascade switch for adaptive developmental plasticity in body mass in a butterfly. *Journal of Evolutionary Biology* **23:** 1129–1137.

Goudet J. 2001. *FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3)*. Available at: http://www2.unil.ch/popgen/softwares/fstat.htm. Updated from Goudet (1995).

Guo Q. 2012. Incorporating latitudinal and central-marginal trends in assessing genetic variation across species ranges. *Molecular Ecology* **21:** 5396–5403.

Habel JC, Husemann M, Schmitt T, Dapporto L, Rodder D, Vandewoestijne S. 2013. A forest butterfly in Sahara desert oases: isolation does not matter. *Journal of Heredity* **104:** 234–247.

Hellmann JJ, Pelini SL, Prior KM, Dzurisin JD. 2008. The response of two butterfly species to climatic variation at the edge of their range and the implications for poleward range shifts. *Oecologia* **157:** 583–592.

Henriksen HJ, Kreutzer IB. 1982. *Skandinaviens dagsommerfugle i naturen*. Odense: Skandinavisk Bogforlag.

Hewitt GM. 1999. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68:** 87–112.

Hewitt GM. 2001. Speciation, hybrid zones and phylogeography – or seeing genes in space and time. *Molecular Ecology* **10:** 537–549.

Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **359:** 183–195.

Hill JK, Griffiths HM, Thomas CD. 2011. Climate change and evolutionary adaptations at species' range margins. *Annual Review of Entomology* **56:** 143–159.

Hill JK, Hughes CL, Dytham C, Searle JB. 2006. Genetic diversity in butterflies: interactive effects of habitat fragmentation and climate-driven range expansion. *Biology Letters* **2:** 152–154.

Hill JK, Thomas CD, Huntley B. 1999. Climate and habitat availability determine 20th century changes in a butterfly's range margin. *Proceedings of the Royal Society B-Biological Sciences* **266:** 1197–1206.

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

Hughes CL, Hill JK, Dytham C. 2003. Evolutionary trade-offs between reproduction and dispersal in populations at expanding range boundaries. *Proceedings of the Royal Society B-Biological Sciences* **270:** S147–S150.

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

Kalinowski ST, Taper ML, Marshall TC. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* **16:** 1099–1106.

Kivelä SM, Valimaki P, Gotthard K. 2013. Seasonality maintains alternative life-history phenotypes. *Evolution* **67:** 3145–3160.

Klopfstein S, Currat M, Excoffier L. 2006. The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution* **23:** 482–490.

Lessa EP, Cook JA, Patton JL. 2003. Genetic footprints of demographic expansion in North America, but not

Amazonia, during the Late Quaternary. *Proceedings of the National Academy of Sciences of the United States of America* **100:** 10331–10334.

Nordström F. 1955. *The distribution of fennoscandian butterflies* [in Swedish]. Lund: Gleerup.

Nylin S, Wickman PO, Wiklund C. 1989. Seasonal plasticity in growth and development of the speckled wood butterfly, *Pararge aegeria* (Satyrinae). *Biological Journal of the Linnean Society* **38:** 155–171.

Nylin S, Wickman PO, Wiklund C. 1995. Life-cycle regulation and life-history plasticity in the speckled wood butterfly – are reaction norms predictable. *Biological Journal of the Linnean Society* **55:** 143–157.

Paetkau D, Strobeck C. 1994. Microsatellite analysis of genetic-variation in black bear populations. *Molecular Ecology* **3:** 489–495.

Parmesan C. 2006. Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology Evolution and Systematics* **37:** 637–669.

Parmesan C, Ryrholm N, Stefanescu C, Hill JK, Thomas CD, Descimon H, Huntley B, Kaila L, Kullberg J, Tammaru T, Tennent WJ, Thomas JA, Warren M. 1999. Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature* **399:** 579–583.

Pöyry J, Luoto M, Heikkinen RK, Kuussaari M, Saarinen K. 2009. Species traits explain recent range shifts of Finnish butterflies. *Global Change Biology* **15:** 732–743.

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

Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences of the United States of America* **102:** 15942–15947.

Rice WR. 1989. Analyzing tables of statistical tests. *Evolution* **43:** 223–225.

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

Seppänen EJ. 1969. Suurperhostemme talvehtimisasteet. *Annales Entomologici Fennici* **35:** 129–152.

Shreeve T. 1986. The effect of weather on the life-cycle of the speckled wood butterfly pararge aegeria. *Ecological Entomology* **11:** 325–332.

Slatkin M, Excoffier L. 2012. Serial founder effects during range expansion: a spatial analog of genetic drift. *Genetics* **191:** 171–181.

Sousa VC, Beaumont MA, Fernandes P, Coelho MM, Chikhi L. 2012. Population divergence with or without admixture: selecting models using an ABC approach. *Heredity* **108:** 521–530.

Stewart JR, Lister AM, Barnes I, Dalen L. 2010. Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B-Biological Sciences* **277:** 661–671.

Storz JF, Beaumont MA. 2002. Testing for genetic evidence of population expansion and contraction: an empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evolution* **56:** 154–166.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7:** 453–464.

Taberlet P, Luikart G. 1999. Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society* **68:** 41–55.

Van Dyck H, Wiklund C. 2002. Seasonal butterfly design: morphological plasticity among three developmental pathways relative to sex, flight and thermoregulation. *Journal of Evolutionary Biology* **15:** 216–225.

Vandewoestijne S, Van Dyck H. 2010. Population genetic differences along a latitudinal cline between original and recently colonized habitat in a butterfly. *PLoS ONE* **5:** e13810.

Vandewoestijne S, Van Dyck H. 2011. Flight morphology along a latitudinal gradient in a butterfly: do geographic clines differ between agricultural and woodland landscapes? *Ecography* **34:** 876–886.

Waits L, Taberlet P, Swenson JE, Sandegren F, Franzen R. 2000. Nuclear DNA microsatellite analysis of genetic diversity and gene flow in the Scandinavian brown bear (Ursus arctos). *Molecular Ecology* **9:** 421–431.

Weingartner E, Wahlberg N, Nylin S. 2006. Speciation in *Pararge* (Satyrinae : Nymphalidae) butterflies – North Africa is the source of ancestral populations of all *Pararge* species. *Systematic Entomology* **31:** 621–632.

Wiklund C, Friberg M. 2011. Seasonal development and variation in abundance among four annual flight periods in a butterfly: a 20-year study of the speckled wood (*Pararge aegeria*). *Biological Journal of the Linnean Society* **102:** 635–649.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Figure S1. Posterior distributions for the parameters of the ABC analysis including demographic parameters and microsatellites mutation model parameters. N(*i*) corresponds to effective population sizes. T(*i*) corresponds to the time of event, in numbers of generations. For posteriors regarding mutation were: the mean mutation rate (μmic), the shape parameter of the gamma distribution describing the heterogeneity of rates among individual loci (pmic) and the single insertion nucleotide rate (snimic). The punctual estimation of each parameter (the median) is indicated in square brackets ([]).

Figure S2. Best estimate of K using Evanno's method (ΔK).
Table S1. Prior distribution of parameters used in our ABC analysis. Population size parameters are in units of population effective size (N_i) and time parameters (ti) are in units of generations. Gamma distribution shape parameter = 2. For the microsatellite, the mutation models priors were the mutation rate (μ_{mic}), the parameter determining the shape of the gamma distribution of individual loci mutation rate (*P*) and the single insertion nucleotide rate (SNI).
Table S2. Comparison of summary statistics for the observed dataset and simulated posterior datasets used in the ABC analyses. The employed summary statistics for the microsatellite data were: the mean number of alleles across loci (*A*), the mean genetic diversity (*G*), the mean Garza-Williamson's *M* index, the mean allele size variance (*V*), the F_{ST} between two samples, the shared allele distance between two samples (DAS) and the $(\delta\mu)^2$ distance between two samples. The *p*-values for each scenario were obtained from the probability *Prob* ($S_{simul} < S_{obs}$) reported here. The *Prob* ($S_{simul} < S_{obs}$) were computed from 1000 datasets simulated from the posterior distribution of parameters obtained. The populations CSc, SSc, Ben and EBa correspond to central Scandinavia, southern Scandinavia, Benelux and Eastern Baltic, respectively.
Table S3. Level of polymorphism per area. An asterisk (*) indicates comparisons based only on the six loci that were polymorphic.
Table S4. Pairwise F_{ST} per area. The F_{ST} values are indicated above the diagonal with the corresponding *p*-value below the diagonal. The F_{ST} values with a p-value < 0.05 are marked in bold.
Table S5. Genotypes for the nine microsatellites with individual names for each sampling region, and specific location of the samples included in this study.