JOURNAL OF Evolutionary Biology

doi: 10.1111/j.1420-9101.2011.02352.x

Population structure in relation to host-plant ecology and *Wolbachia* infestation in the comma butterfly

U. KODANDARAMAIAH*[1], E. WEINGARTNER*†, N. JANZ*, L. DALÉN† & S. NYLIN*

*Department of Zoology, Stockholm University, Stockholm, Sweden
†Department of Molecular Systematics, Swedish Museum of Natural History, Stockholm, Sweden

Keywords:

host-plant;
phylogeography;
Polygonia c-album;
single refugium;
Wolbachia.

Abstract

Experimental work on *Polygonia c-album*, a temperate polyphagous butterfly species, has shown that Swedish, Belgian, Norwegian and Estonian females are generalists with respect to host-plant preference, whereas females from UK and Spain are specialized on Urticaceae. Female preference is known to have a strong genetic component. We test whether the specialist and generalist populations form respective genetic clusters using data from mitochondrial sequences and 10 microsatellite loci. Results do not support this hypothesis, suggesting that the specialist and generalist traits have evolved more than once independently. Mitochondrial DNA variation suggests a rapid expansion scenario, with a single widespread haplotype occurring in high frequency, whereas microsatellite data indicate strong differentiation of the Moroccan population. Based on a comparison of polymorphism in the mitochondrial data and sequences from a nuclear gene, we show that the diversity in the former is significantly less than that expected under neutral evolution. Furthermore, we found that almost all butterfly samples were infected with a single strain of *Wolbachia*, a maternally inherited bacterium. We reason that indirect selection on the mitochondrial genome mediated by a recent sweep of *Wolbachia* infection has depleted variability in the mitochondrial sequences. We also surmise that *P. c-album* could have expanded out of a single glacial refugium and colonized Morocco recently.

Introduction

Herbivorous insects are strikingly diverse, accounting for a quarter of all described eukaryote species (Bernays, 1998), and the mechanisms underlying their diversification have intrigued biologists for long. The pioneering work of Ehrlich & Raven (1964) on butterfly host-plant coevolution first called attention to the potential role of herbivory in insect diversification. Since then, it has been convincingly demonstrated that intimate herbivore–plant interactions have time and again led to elevated speciation rates (Mitter *et al.*, 1988; Farrell, 1998). Within butterflies, studies have shown that wider host ranges are correlated with enhanced species diversity (Janz *et al.*, 2001, 2006; Weingartner *et al.*, 2006; Nylin & Wahlberg, 2008). Nevertheless, little is known about the mechanistic basis of host-plant-mediated speciation.

One hypothesis – the 'oscillation hypothesis' (Janz & Nylin, 2008) – posits that speciation is driven by successive cycles of expansion of the host-plant repertoire (i.e. broadening of the host range by the addition of novel host-plant species) and the genesis of new species through specialization on different hosts. Although experimental tests of such hypotheses are impractical, investigations into divergences between populations of extant polyphagous species can potentially shed light on incipient speciation. Thus, divergence between populations that differ in host-plant preferences indicates possible segregation of genetic variation between host-plant 'races' (Via, 2001; Drès & Mallet, 2002). In this

Correspondence: Ullasa Kodandaramaiah, Department of Zoology, Stockholm University, 106 91 Stockholm, Sweden.
Tel.: +46 8 164 398; fax: +46 8 167 715;
e-mail: ullasa.kodandaramaiah@zoologi.su.se
[1]Present address: Department of Zoology, University of Cambridge, Cambridge, UK.

regard, studies that have attempted to assess the geographic structuring of host-plant preferences within species are noteworthy. One species that has been extensively investigated with regard to host-plant preference is the nymphalid butterfly *Polygonia c-album* (Nymphalidae: Nymphalinae). Knowledge from past experimental work on this species indicates that female host-plant preference varies across populations and that these differences have a strong genetic component (Nylin, 1988; Janz, 1998; Nygren *et al.*, 2006; Nylin *et al.*, 2009). In this study, we test two competing hypotheses of genetic divergence between populations that have a bearing on our understanding of incipient ecological speciation mediated by host choice.

P. c-album is a widespread Palearctic species distributed in Europe, North Africa and Asia, with minor external morphological differences across its range. Several subspecies of *P. c-album* have been described – e.g. *kultukensis* (Siberia), *hamigera* (South East Russia and Japan), *asakurai* (Taiwan, China and Japan) and *imperfecta* (Morocco). The taxonomic status of *P. interposita* has been disputed. In the study of Tuzov (2000), it was considered a species, whereas Gorbunov (2001) treated it as a subspecies within *P. c-album*. More recently, Churkin (2003) regarded *P. interposita* as a *bona fide* species, based on differences in wing pattern and genitalia. Recent molecular analyses of nuclear genes confirm differences between *P. c-album* and *P. interposita,* although sequences of the mitochondrial gene cytochrome oxidase subunit I (COI) are almost identical (Wahlberg *et al.*, 2009a). In this study, we have considered *P. interposita* as a separate species in accordance with the study of Wahlberg *et al.* (2009a) and thus excluded it from our analysis. The larvae of *P. c-album* feed on foliage of plants in the families Urticaceae, Ulmaceae, Cannabidaceae, Salicaceae, Grossulariaceae, Betulaceae and Corylaceae (Nylin, 1988; Tolman & Lewington, 1997; Savela, 2010).

Nylin (1988) compared the host preferences of spring generation females from the UK and Swedish populations. He found that the UK females were specialized to a stronger degree on *Urtica dioca* (Urticaceae) compared with *Salix caprea* (Salicaeae), whereas the Swedish females were more general in preference, a result that was further corroborated in the study of Nylin & Janz (1993). Janz (1998) and Nygren *et al.* (2006) also showed, using reciprocal crosses, that the female preference trait was linked to the X chromosome. The X chromosome has also been implicated in female host choice in *Papilio* butterflies (Thompson, 1988; Scriber & Lederhouse, 1992). Nygren *et al.* (2006) and Nylin *et al.* (2009) reported that Spanish populations strongly preferred Urticaceae, whereas the Belgian and Swedish populations were more general in host-plant choice. Furthermore, Estonian females are generalists, similar to Swedish females.

Experiments have shown that Urticalean rosids (Urticaceae, Ulmaceae, Cannabidaceae) support faster larval growth rates in this species (Nylin, 1988; Janz *et al.*, 1994). Nygren *et al.* (2006) surmise that the geographic variation in female host preference could be linked to patterns of voltinism, and this was christened the 'time constraint hypothesis' in the study of Nylin *et al.*, 2009. Accordingly, specialization on urticalean rosids is selected for in the time-stressed bi- or trivoltine southern populations, whereas the northern univoltine populations face relaxed selection. A competing hypothesis is that the specialist and generalist traits evolved separately in different populations, and the current pattern we observe is determined by the phylogeographic history of, and geneflow between, populations – for instance, specialist and generalist populations could have survived in two different glacial refugia, followed by post-glacial expansions from each. Thus, this hypothesis predicts that the generalist and specialist populations form respective genetic clusters. We here test this hypothesis based on molecular data from mitochondrial (henceforth mtDNA) and microsatellite (henceforth nDNA) markers from nine populations, including two specialist (UK and Spain) and two generalist (Sweden and Belgium) populations.

The goal of this study was to understand the population genetic and phylogeographic structure of populations in this species. In addition to being a framework to test the above-stated hypothesis, such data are crucial to understand the evolution of host-plant use in this model species and ultimately shed light on the mechanistic basis of host-plant-mediated ecological speciation. We found that results from mtDNA and nDNA were not congruent and therefore also tested for *Wolbachia* infestation in the species. Bacteria in the genus *Wolbachia* are commonly occurring cytoplasmic endosymbionts found in 20 to >66% of all insect species (Werren & Windsor, 2000; Hilgenboecker *et al.*, 2008). Being maternally inherited, they also have the potential to alter the contemporary mitochondrial genetic structure of populations. We have used molecular assays to ascertain the presence of *Wolbachia* and strain diversity in *P. c-album*.

Materials and methods

Sampling and DNA extraction

We obtained 85 specimens of *P. c-album* from nine populations in Europe, Africa and Asia (Appendix 1). The sampling mainly represents populations in Spain, Morocco, UK, Sweden and Russia, but also Macedonia, France, Belgium and Finland. Most samples were collected by us or colleagues, and some were bought from commercial suppliers. DNA was preserved either by desiccation or by immersing two legs in alcohol. Extractions were made from two legs of each individual using the QIAGEN (Hilden, Germany) DNEasy Extraction Kit following the manufacturer's protocol.

Sequencing

The mitochondrial gene COI, the 'barcoding gene' (Hebert *et al.*, 2003), is known to be variable enough within nymphalid species (Wahlberg & Saccheri, 2007; Vandewoestijne *et al.*, 2004) for phylogeographic analyses. The two primer pairs, LCO-HCO and Jerry-Pat, were used to amplify 1450 bp of COI (see Kodandaramaiah & Wahlberg, 2007 for primer sequences). The PCR protocol was as follows: 95 °C for 5 min, followed by 40 cycles of 94 °C for 30 s, 50 °C for 30 s, 72 °C for 1 min and a final extension period of 72 °C for 10 min. Sequencing was carried out in a Beckman Coulter (Bromma, Sweden) CEQ 8000 capillary sequencer using forward primers. In the cases where the sequence quality in the second half was not acceptable, they were resequenced with the reverse primer. Chromatograms were checked and aligned visually in BioEdit v7.0.5.3 (Hall, 1998). Appendix 1 lists GenBank accession numbers of all sequences.

Microsatellites

Ten variable microsatellite loci, the development of which has been described in the study of S. Nylin, E. Weingartner, N. Janz and U. Kodandaramaiah (submitted), were utilized in the study. The PCR cycling profile consisted of the following steps: (i) 95 °C (15 min); (ii) 30 cycles of annealing temperature (30 s), 72 °C (30 s) and 95 °C (30 s); (iii) annealing temperature (1 min); and (iv) final extension of 30 min at 72 °C. The annealing temperature was 50 °C, except for '*Polalb 11*' for which it was 60 °C. DNA amplifications were carried out with 10-μL PCRs that included 1.0 μL DNA template, 1 × PCR buffer, 40 μM dNTP, 0.5 units of HotStart Taq (Qiagen) and 0.1–0.25 μM primer (adjusted according to the binding efficiency of respective primer pairs). Forward primers were dye-labelled, and amplified products were electrophoretically separated on a Beckman Coulter CEQ 8000 capillary sequencer (Bromma, Sweden).

Sequence analyses

A statistical parsimony network (Templeton *et al.*, 1992) of COI haplotypes was reconstructed in TCS v1.21 (Clement *et al.*, 2000). Standard genetic diversity indices were calculated in Arlequin v 3.1 (Excoffier *et al.*, 2005). These included total number of haplotypes, haplotype diversity (H; the probability that two randomly chosen haplotypes in the sample are different (Nei, 1987)) and nucleotide diversity (π_n; the probability that two randomly chosen homologous nucleotide sites are different (Tajima, 1983; Nei, 1987)). Global and among population Φ_{ST} values were calculated as an estimate of the genetic differentiation. Exact tests of pairwise population differentiation (Raymond & Rousset, 1995; Goudet *et al.*, 1996) were conducted with a Markov Chain of 100 000 steps.

A mismatch distribution analysis was performed in Arlequin to test whether the distribution of pairwise differences fits a model of sudden demographic expansion (Slatkin & Hudson, 1991; Rogers & Harpending, 1992). Deviations from neutrality were examined by performing Fu's F_s test (Fu, 1997). This test uses the infinite-site model and is sensitive to departure from population equilibrium due to recent mutations. A large negative value generally indicates population demographic expansion. In total, 5000 samples were simulated in the analysis.

Microsatellite analyses

Allelic variability and pairwise population F_{ST} values were calculated in Arlequin v 3.1 (Excoffier *et al.*, 2005) with significance for the latter values tested based on 100 000 permutations. Exact tests of Hardy–Weinberg equilibrium, both globally and within each population, were conducted with default parameters in Arlequin (1 million steps in the Markov chain and 100 000 dememorization steps). A significant heterozygote deficit compared with the expected values suggested the presence of null alleles in eight loci. Hence, we also analysed the data set in the software FreeDNA (Chapuis & Estoup, 2007), which detects and corrects for the presence of null alleles in each population. The program implements the expectation maximization (EM) algorithm of Dempster *et al.*, 1977 to estimate frequencies of null alleles and adjusts genotype frequencies according to the ENA correction method (Chapuis & Estoup, 2007). Pairwise F_{ST} values were estimated with the method of Weir, 1996 using the ENA-corrected values. A Mantel test (Mantel, 1967) was conducted in Genepop v4.0 using default parameters (Rousset, 2008) to test for geographic structuring among populations due to isolation-by-distance (Rousset, 1997). The analysis was repeated excluding Morocco from the analysis.

STRUCTURE v2.3 (Pritchard *et al.*, 2000) was used to identify distinct genetic populations and assign individuals to these populations. STRUCTURE implements a model-based clustering method within a Bayesian framework to infer population structure. We analysed the data set with K (the prior on the maximum number of populations) values ranging from 1 to 10 using the admixture model (where each individual is allowed to have a fraction of its genotype from more than one population). The LOCPRIOR model, which assists clustering by making use of information on the origin of samples, was turned on. The analysis was run for 1 000 000 MCMC (Markov chain Monte Carlo) steps after a burnin of 100 000 steps. The analysis was rerun after discarding the locus 'Polalb 6', which failed to amplify in the Moroccan samples. The 10-locus data set was also analysed using the dominant markers model

(Falush *et al.*, 2007) by turning on the 'RECESSIVEALLELES' option in the program and setting the recessive value to an allele not observed at that locus (as suggested by the authors of STRUCTURE). This model, although designed primarily to deal with dominant markers such as AFLPs, allows more accurate results with codominant markers when null alleles are present in the data set (Pritchard *et al.*, 2000; Falush *et al.*, 2007). Individuals that did not amplify a particular locus were considered to have missing data, rather than being treated as null homozygotes.

Wolbachia-related assays

Wolbachia specific primers wsp81F (5'-TGG TCC AAT AAG TGA TGA AGA AAC3-') and wsp691R (5'-AAA AAT TAA ACG CTA CTC CA-3'; (Zhou *et al.*, 1998) amplify a gene (*wsp*) encoding a surface protein of *Wolbachia*. The primer pair has been extensively used to detect the presence of the bacterium in DNA extracts. PCR products visualized as bands in a standard 1% agarose gel indicate the presence of the bacterium in the DNA extract, thus providing a rapid method of detecting *Wolbachia* infestation in a large number of samples. We used this assay to ascertain infestation in our samples. The forward and reverse primers were concatenated with the universal primers T7Promoter and T3, respectively, to facilitate sequencing (see Wahlberg & Wheat, 2008 for more details). PCR protocols were as for COI with the exception that the annealing temperature was 55 °C and 8 μL product was checked on the gel. All PCR sets included positive and negative controls. Because the *Wolbachia* assays were conducted 3 years after COI was sequenced, samples that did not produce a visible band on the gel were tested with the LCO–HCO primer pair for the integrity of the DNA extracts.

Four *Wolbachia* genes, in addition to *wsp*, were sequenced from 20 samples (one to three from each population; Appendix 1) to ascertain strain diversity in *P. c-album*. These genes (ftsZ, gatB, groEL and gltA) were chosen from published studies demonstrating their variability within *Wolbachia* (Casiraghi *et al.*, 2005; Baldo *et al.*, 2006; Ros *et al.*, 2009). Appendix 2 lists the primers used along with their respective annealing temperatures.

Nuclear gene sequencing

We also sequenced a relatively fast-evolving nuclear gene, ribosomal protein subunit 5 (RPS5), from these 20 *P. c-album* samples in order to compare variability in COI to that in a nuclear gene region. PCR protocols were the same as for *wsp*. A Fisher's exact test was performed in R v 2.11.1 (R Development Core Team, 2010) to test for a difference in polymorphism between COI and RPS5, after correcting for differences in effective population size (*sensu* Hudson *et al.*, 1987).

Results

Mitochondrial sequences

A 36-bp (between 643 and 681 bp) region contained missing data for several sequences and was hence deleted from the analysis to avoid possible ambiguities in haplotype assignment. The final data set thus consisted of 1414 bp. Seventeen unique haplotypes were found among the 84 sequenced specimens of *P. c-album*. Haplotype diversity was 40.25% (SD ± 6.9%), and nucleotide diversity was 0.039% (SD ± 0.34%). Respective values for each population are presented in Table 1. Pairwise Φ_{ST} are shown in Table 2. In the exact tests of population differentiation, Spain was significantly differentiated from UK and Sweden. None of the other pairwise comparisons was significant.

The haplotype network of *P. c-album* was star-shaped (Fig. 1) and contained four missing haplotypes. Of the 84 individuals, 65 (77.3%) carried the same haplotype (*Widespread*), which formed the centre of the network and was found in all nine populations. Of the remaining 16 haplotypes, 13 were represented by a single individual. Two haplotypes (*UK1* and *UK2)* were represented by two samples each from UK, whereas one haplotype (*SW1*) was represented by two individuals from Sweden.

The mismatch distribution showed a negative slope, implying the leading edge of a unimodal curve. The curve did not deviate from that expected under a model of sudden expansion (SSD = 0.00037164, P ($SSD_{sim} \geq SSD_{obs}$) = 0.86450000). Fu's F_s value was significantly negative ($F_s = -22.73881$, P ($sim\_F_s \leq obs\_F_s$) < 0.0001).

Microsatellites

All 10 loci were variable, with allele numbers ranging from 13 to 35 (Table 3, Appendix 3). One locus (Polalb 6) did not amplify in the butterflies from Morocco. The Belgian population was represented by a single individual and was hence deleted from pairwise population comparisons. Eight loci had significantly lower heterozygosity

Table 1 Total number of mtDNA haplotypes as well as nucleotide and haplotype diversity values for population. Belgium, France and Finland were excluded from the analysis because they were represented by less than five individuals.

	No. of haplotypes (samples)	Haplotype diversity (H)	Nucleotide diversity (π_n)
Macedonia	3 (5)	0.7000 ± 0.2184	0.001132 ± 0.000933
Morocco	2 (9)	0.2222 ± 0.1662	0.000157 ± 0.000231
Russia	3 (14)	0.2747 ± 0.1484	0.000303 ± 0.000326
Spain	2 (19)	0.1053 ± 0.0920	0.000074 ± 0.000144
Sweden	6 (16)	0.6167 ± 0.1347	0.000607 ± 0.000506
UK	4 (14)	0.5824 ± 0.1372	0.000474 ± 0.000433

Table 2 Pairwise Φst values calculated from mtDNA haplotype frequencies in Arlequin. Φst values of population pairs that were significantly different in the exact tests of differentiation are in bold.

	Belgium	Finland	France	Macedonia	Morocco	Russia	Spain	Sweden
Belgium	–							
Finland	0.00	–						
France	0.00	0.00	–					
Macedonia	0.00	0.00	0.00	–				
Morocco	0.00	0.15	0.15	0.09	–			
Russia	0.00	0.07	0.07	0.10	0.00	0.00		
Spain	0.00	0.34	0.34	0.23	0.02	0.01	0.00	
Sweden	0.00	0.00	0.00	0.06	0.00	0.01	**0.03**	0.00
UK	0.00	0.03	0.03	0.09	0.02	0.04	**0.07**	0.04

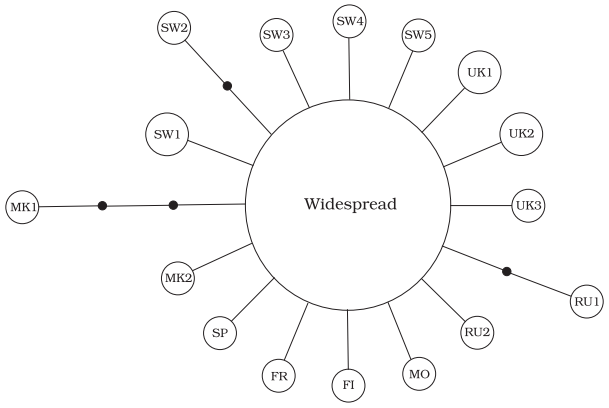

Fig. 1 Statistical parsimony network of the 17 *Polygonia c-album* cytochrome oxidase subunit I haplotypes. SW – Sweden; UK – United Kingdom; RU – Russia; SP – Spain; FI – Finland, MO – Morocco; FR – France; MK – Macedonia; BE – Belgium. The central haplotype 'Widespread' was found in 65 individuals, whereas SW1, UK1 and UK2 were represented by two individuals each. Remaining haplotypes were restricted to single individuals.

Table 3 Allelic variability, expected and observed heterozygosity values estimated from genotyping 85 *Polygonia c-album* individuals for 10 microsatellite loci. Observed heterozygosity values significantly lower than expected are in bold.

Locus	Allelic range	No. of alleles	Exp. *H*	Obs. *H*
Polalb 2	120–164	17	0.87549	**0.36111**
Polalb 5	269–309	35	0.94654	**0.425**
Polalb 6	69–139	26	0.92074	**0.54795**
Polalb 7	173–233	24	0.92471	**0.48649**
Polalb 8	117–145	13	0.84901	**0.28395**
Polalb 10	117–181	24	0.94457	**0.37681**
Polalb 11	205–257	14	0.80383	0.71765
Polalb 17	81–119	15	0.89946	**0.23729**
Polalb 20	175–213	15	0.89614	0.83529
Polalb 12	119–163	22	0.92556	**0.4878**

values compared with the expected Hardy–Weinberg heterozygosity values (all except Polalb 11 and Polalb 20; Table 3). This indicates the possible presence of null alleles in these loci, which was corroborated by FreeNA. However, F_{ST} estimates with and without correction using the ENA method showed similar patterns of differentiation. In both the cases, Morocco was differentiated from other populations with F_{ST} values >0.1 (Table 4), whereas all other pairwise F_{ST} values were <0.1. Morocco was significantly differentiated from all populations except Belgium and France, which were represented by one and three individuals, respectively. Russia was significantly different from Morocco, Sweden, Finland and Spain (Table 4).

The Mantel test indicated no significant isolation-by-distance (one tailed P value = 0.1), and this was true even when Moroccan samples were excluded from the analysis. In all STRUCTURE analyses, the log likelihood value increased as we increased the prior on the maximum number of populations (K) from one to two, indicating that there were at least two genetic clusters in the data set. However, we found no consistent pattern of increase or decrease in the log likelihood values as we increased K from 2 to 10. We first report results from the analysis on the 10-locus data set without imposing the dominant markers model (i.e. not 'correcting' for the presence of null alleles). A well-defined cluster comprised of the Moroccan individuals was recovered in all runs except with $K = 1$. Another cluster, of the Russian individuals, was recovered in runs with $K = 3$–6, whereas this cluster was less cohesive with higher K values. In keeping with the principle that the minimum number of populations that capture the structure should be chosen (Pritchard *et al.*, 2000, in the documentation to STRUCTURE), we depict results from the run with $K = 3$ (Fig. 2). We do note that we were unable to objectively choose the best value for K. The results with respect to the Moroccan and Russian clusters were mirrored in the analysis where the locus Polalb 6 was discarded, confirming that the missing data did not affect results. With the dominant markers model imposed (on the 10-locus data set), we obtained similar results, i.e. the Moroccan cluster was well resolved with all values of K, whereas the Russian cluster was well defined only with lower K values.

Table 4 Estimates of pairwise population differentiation based on microsatellite data measured as F_{ST} values. Numbers to the left of the diagonal are F_{ST} estimates without correcting for the presence of null alleles, calculated in Arlequin. F_{ST} values in bold are significantly greater than zero. Numbers to the right of the diagonal are F_{ST} values estimated after ENA correction for null alleles in FreeNA.

	UK	Russia	Morocco	Sweden	Finland	Spain	Macedonia	France
UK	N/A	0.02	0.16	0.00	0.04	0.01	−0.01	0.06
Russia	0.02	N/A	0.15	0.02	0.09	0.02	0.02	0.06
Morocco	**0.19**	**0.21**	N/A	0.14	0.20	0.14	0.15	0.20
Sweden	0.01	**0.05**	**0.19**	N/A	0.05	0.00	0.00	0.04
Finland	0.02	0.06	**0.18**	0	N/A	0.06	−0.04	0.07
Spain	0.01	**0.06**	**0.19**	0.00	0.01	N/A	−0.01	0.05
Macedonia	0.02	**0.07**	**0.19**	0.03	−0.03	0.01	N/A	0.01
France	0.05	0.02	0.08	0.01	−0.01	0.02	0.00	N/A

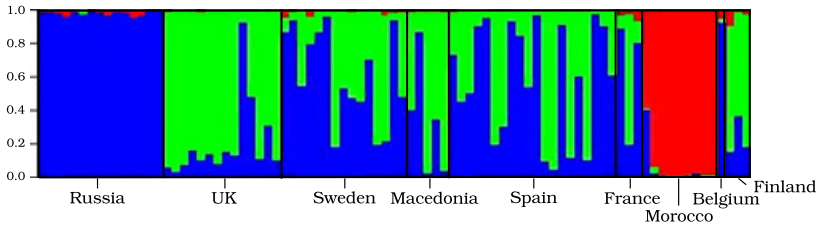

Fig. 2 Population structure inferred in the STRUCTURE analysis of the 10-locus data set with K = 3. Each individual is represented by a single vertical line with a maximum of three coloured segments, with lengths proportional to each of the three inferred clusters. Numbers on the horizontal axis correspond to the eight predefined populations. 1 corresponds to the Russian individuals and 7 to the Moroccan ones.

For all loci except Polalb 11, we were unable to amplify all individuals. We ran a F_{ST} population comparison in Arlequin (among Russian, UK, Swedish, Spanish and Moroccan populations) including those loci with most representation and least deficit of heterozygotes (Polalb 5, Polalb 7, Polalb 8, Polalb 11, Polalb 20 and Polalb 22). The Moroccan population was strongly and significantly different from the other populations. Similarly, the Russian population was significantly different from the rest (data not shown).

Wolbachia

Of the 85 samples, 82, representing all nine populations, produced a distinguishable band on the gel after PCR amplification with the *wsp* primers. Two samples subsequently failed to amplify LCO–HCO, indicating degradation of DNA in storage. The 20 *wsp* PCR products sequenced resulted in identical sequences. Similarly, the *ftsz, gatB, Wglt* and *groEL* sequences displayed no variation across individuals. All five sequences belonged to *Wolbachia* supergroup B (sequences submitted to GenBank, accession numbers JN093149–53).

Nuclear gene sequencing

The 20 RPS5 sequences (602 bp) had a total of 25 segregating sites, and 24 of these positions were heterozygous in one or more individuals. COI (1414 bp) had four segregating sites for the same set of individuals. The four times higher effective population size of autosomal genes compared with that of mitochondrial genes was taken into account by scaling the number of segregating sites in COI by a factor of four (*sensu* Hudson *et al.*, 1987). Nucleotide polymorphism in COI was significantly lower than that in RPS5 ($P < 0.001$).

Discussion

Rapid expansion of mtDNA

Analyses of the mitochondrial data set indicate a rapid expansion of the mitochondrial genome of *P. c-album* in the Palearctic. The haplotype network is a typical star-shaped network with a widespread central haplotype and geographically restricted 'satellite' haplotypes radiating from it (Fig. 1). This pattern is consistent with an exponentially growing population (Slatkin & Hudson, 1991). The widespread haplotype is ubiquitous across the species range, whereas no population is significantly differentiated genetically. The lack of differentiation over such a large geographic expanse suggests a recent and rapid expansion of the mitochondrial genome. This is further corroborated by the results from the mismatch distribution analysis and the significantly negative Fu's *F* value, which fit with a model of sudden demographic expansion.

Wolbachia and its effects

Wolbachia are known to manipulate the reproductive ecology of their hosts in order to gain a selective advantage. The most common strategy is cytoplasmic incompatibility, where infected males cannot produce viable offspring with uninfected females or females infected with a different strain (Werren, 1997). The net effect is that uninfected females are selected against (Jansen *et al.*, 2008). Because the bacterium is maternally inherited, this allows it to spread rapidly in the host population (Hurst *et al.*, 1993). They are also known to distort sex ratios in favour of females, which is also thought to increase prevalence rates in host populations (Hurst & Majerus, 1993). In some cases, *Wolbachia* provide direct selective benefits to the host; for example, in *Drosophila melanogaster, Wolbachia* infection decreases susceptibility to mortality by a range of RNA viral infections (Hedges *et al.*, 2008; Teixeira *et al.*, 2008). Therefore, *Wolbachia* have evolved multiple mechanisms that facilitate rapid spread across populations of a host species. Being maternally inherited, mtDNA hitchhikes along with the bacterium (Turelli *et al.*, 1992; Hurst *et al.*, 1993; Rasgon *et al.*, 2003; Hurst & Jiggins, 2005). This ultimately results in selective sweeps through which a single mtDNA haplotype dominates (Turelli & Hoffmann, 1991; Jiggins, 2003; Ballard & Whitlock, 2004; Hurst & Jiggins, 2005). Moreover, the effective population size of mtDNA at equilibrium is reduced because mutations in uninfected females are rapidly lost (Johnstone & Hurst, 1996).

Our *Wolbachia* assays show a near 100% prevalence of the bacterium in *c-album* and strongly indicate that all populations have the same strain. It must be noted that the PCR assays used to ascertain infection status are not error free (Jeyaprakash & Hoy, 2000) – it is thus possible that the individual that was not positive in the assay may still be infected. The infection rate we report in this species is one of the highest reported so far. We have been using *P. c-album* from various populations in laboratory experiments for over two decades and have never observed any marked bias in favour of females through the course of innumerable laboratory rearings. Neither has this been reported among wild populations, to our knowledge. We suspect that the high prevalence rate of the bacterium is maintained by the induction of cytoplasmic incompatibility.

Cytochrome oxidase subunit I is generally a fast-evolving gene, because of which it is commonly employed in phylogenetics and as a barcoding gene. RPS5 has a lower substitution rate in published nymphalid phylogenetic data sets (Wahlberg *et al.*, 2009a,b; Kodandaramaiah *et al.*, 2009; Peña & Wahlberg, 2008; Peña *et al.*, 2011). Wahlberg *et al.* (2009a) included 400 taxa across Nymphalidae, whereas Wahlberg *et al.* (2009a) studied the genus *Polygonia.* Hence, it is highly unlikely that RPS5 has an inherently higher mutation rate than does COI. We can also rule out balancing selection in RPS5 as most substitutions are synonymous – only three of the 23 lead to a change in amino acid. RPS5 has higher polymorphism in this species even under the assumption that both genes have the same mutation rate. This, coupled with the *Wolbachia* assays, leads us to conclude that indirect selection on the mitochondrial genome mediated by the spread of the bacterium has reduced mtDNA diversity drastically. The effects of *Wolbachia* infestation resulting in altered mtDNA structure have been shown in several groups including mosquitoes, butterflies, ants and beetles (Keller *et al.*, 2001; Shoemaker *et al.*, 2003; Shaikevich *et al.*, 2005; Gompert *et al.*, 2006; Narita *et al.*, 2006; Nice *et al.*, 2009; see Hurst & Jiggins, 2005 for more examples). Selective sweeps of *Wolbachia* leave traces very similar to that from a demographic expansion (Johnstone & Hurst, 1996; Hurst & Jiggins, 2005). That the mtDNA data correspond to a rapid expansion model further corroborates the *Wolbachia* scenario. The depleted variation in COI suggests that the symbiont invasion is quite recent. However, most RPS5 sequences possessed two or more heterozygotes; we were hence unable to reliably identify haplotypes and reconstruct a haplotype network.

Microsatellite null alleles

Historically, microsatellites have been difficult to develop for butterflies and other Lepidoptera (Meglecz & Solignac, 1998; Zhang, 2004). Null alleles have been reported in almost all papers describing the development of microsatellites in Lepidoptera (reviewed in Meglecz *et al.*, 2004). The presence of null alleles hence appears to be a feature of the lepidopteran genome, perhaps because of high mutation rates in the flanking regions of microsatellites. They have been considered problematic because they can bias F_{ST} estimates and population assignment (Oosterhout *et al.*, 2006; Chapuis & Estoup, 2007); thus, null allele-free loci are assumed to be preferable. However, it is difficult to assess the presence of null alleles before genotyping a substantially large number of individuals. Moreover, in the case of butterflies, loci without null alleles are relatively uncommon, making studies with only null allele-free loci infeasible unless data from the majority of loci are discarded. Fortunately, recent methods (Oosterhout *et al.*, 2006; Chapuis & Estoup, 2007) can correct for the bias introduced by null alleles, providing more realistic results. In this study, we found that the overall results were not strongly affected depending on whether or not these corrections were applied. A simulation study (Carlsson, 2008) showed that the presence of null alleles leads to a small overestimation of F_{ST} values (between 0.003 and 0.004) and slightly reduced the proportion of correctly assigned individuals in STRUCTURE. The author of the study goes on to state that the presence of null alleles is unlikely to have major impacts on conclusions regarding the presence or

absence of genetic differentiation. Our results are in agreement with the author's conclusions. F_{ST} estimates based on loci with null alleles are probably more robust to their presence than previously assumed.

Population structure vis-a-vis host-plant use

All microsatellite markers in our study had considerable variation, with a minimum of 14 alleles per locus (Table 3). This data set indicates strong differentiation of the Moroccan population from all other populations. The Russian population is also differentiated, although not as strongly as in the case of the Moroccan population. Geographic isolation due to the intervening marine barrier explains the differentiation of the Moroccan population. The Russian samples used in this study originated from the south-eastern part of the country and hence geographically relatively distant from the remaining populations. Surprisingly, there was no significant IBD, even when Morocco was excluded from the analysis, which indicates that the F_{ST} values were not correlated with distance. However, the F_{ST} analyses indicate that the Russian samples are significantly differentiated from several European populations, supporting the STRUCTURE results. Furthermore, Churkin (2003) states that the male genital morphology of the Siberian *P. c-album* is very different from the European butterflies. We hence conclude that the Russian samples used in this study comprise a distinct genetic cluster. Little is known about host-plant preference of this species in Russia, and future work on this population will be crucial in understanding the evolution of host-plant use in this species.

There was weak structuring among the remaining populations, suggesting considerable geneflow among them. Therefore, we find no support for the hypothesis that generalist populations (represented by Sweden and Belgium in this study) are genetically more proximate to each other in relation to specialist populations (represented by Spain and UK). We also do not find any evidence for latitudinal structuring based on voltinism. We reason that host-plant preferences have evolved more or less independently in different populations depending on local conditions. Thus, generalist and specialist traits appear to be evolutionarily plastic and little constrained by the phylogeographic history of populations, which is also of relevance in predicting the effects of climate change (Braschler & Hill, 2007).

Literature records indicate that the species feeds on *Ribes* spp in Morocco (Tennent, 1996). To our knowledge, *Ribes* is not one of the preferred species in Europe. It is possible that the Moroccan population has adapted to a different host-plant over time according to local conditions. We identity this population as a focal system to understand the evolution of host-plant preferences and to study early stages of host-plant mediated speciation.

Summary and conclusions

Population structure inferred from microsatellites suggested a strongly differentiated Moroccan population and to a lesser degree, another genetic cluster comprising the Russian individuals. However, mtDNA suggests a recent rapid expansion and little divergence among populations. We surmise that the best explanation is indirect selection on the mitochondrial genome by *Wolbachia*, resulting in rapid fixation of a single haplotype and reduced diversity overall. The data also suggest that specialist and generalist traits with respect to host-plant use are evolutionarily plastic and have evolved independently in different populations.

Acknowledgments

We acknowledge funding from the Swedish Research Council (Vetenskapsrådet) and the Strategic Research Programme EkoKlim at Stockholm University. UK was partly funded by the ERC grant EMARES during manuscript preparation. Samples were provided by Andrew B. Martynenko, Georg Nygren, Niklas Wahlberg, Slobodan Davkov, Anton Chichvarkhin, Michel Tarrier, Bengt Karlsson, A. G. Belik, Constantí Stefanescu, Jane Hill, Anssi Teräs and Reza Zahiri. We are also grateful to Karin Norén for help in the laboratory and Emily Hornett for providing information on *Wolbachia* primers.

References

Baldo, L., Dunning Hotopp, J.C., Jolley, K.A., Bordenstein, S.R., Biber, S.A., Choudhury, R.R. *et al.* 2006. Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Appl. Environ. Microbiol.* **72**: 7098–7110.

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

Bernays, E.A. 1998. Evolution of feeding behavior in insect herbivores. *Bioscience* **48**: 35–44.

Braschler, B. & Hill, J.K. 2007. Role of larval host plants in the climate-driven range expansion of the butterfly *Polygonia c-album*. *J. Anim. Ecol.* **76**: 415–423.

Carlsson, J. 2008. Effects of microsatellite null alleles on assignment testing. *J. Hered.* **99**: 616–623.

Casiraghi, M., Bordenstein, S.R., Baldo, L., Lo, N., Beninati, T., Wernegreen, J.J. *et al.* 2005. Phylogeny of Wolbachia pipientis based on gltA, groEL and ftsZ gene sequences: clustering of arthropod and nematode symbionts in the F supergroup, and evidence for further diversity in the Wolbachia tree. *Microbiology* **151**: 4015–4022.

Chapuis, M. & Estoup, A. 2007. Microsatellite null alleles and estimation of population differentiation. *Mol. Biol. Evol.* **24**: 621–631.

Churkin, S.V. 2003. Taxonomic notes on *Polygonia* Hubner, [1818] (Lepidoptera, Nymphalidae) with description of a new subspecies. *Helios (Moscow)* **4**: 132–147.

Clement, M., Posada, D. & Crandall, K.A. 2000. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* **9**: 1657–1659.

Dempster, A., Laird, N. & Rubin, D. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat Soc. Series B Stat. Methodol.* **39**: 38, 1.

Drès, M. & Mallet, J. 2002. Host races in plant-feeding insects and their importance in sympatric speciation. *Philos. Trans. R. Soc. B-Biol. Sci.* **357**: 471–492.

Ehrlich, P.R. & Raven, P.H. 1964. Butterflies and plants: a study in coevolution. *Evolution* **18**: 586–608.

Excoffier, L., Laval, G. & Schneider, S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **1**: 47–50.

Falush, D., Stephens, M. & Pritchard, J.K. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* **7**: 574–578.

Farrell, B.D. 1998. ''Inordinate Fondness'' explained: why are there so many beetles? *Science* **281**: 555–559.

Fu, Y.X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**: 915–925.

Gompert, Z., Nice, C.C., Fordyce, J.A., Forister, M.L. & Shapiro, A.M. 2006. Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly. *Mol. Ecol.* **15**: 1759–1768.

Gorbunov, P.Y. 2001. *The Butterflies of Russia: Classification, Genitalia, Keys for Identification (Lepidoptera: Hesperioidea and Papilionoidea)*. Thesis, Ekaterinburg.

Goudet, J., Raymond, M., de-Meeus, T. & Rousset, F. 1996. Testing Differentiation in diploid populations. *Genetics* **144**: 1933–1940.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**: 95–98.

Hebert, P.D.N., Cywinska, A., Ball, S.L. & deWaard, J.R. 2003. Biological identifications through DNA barcodes. *Proc. R. Soc. B* **270**: 313–322.

Hedges, L.M., Brownlie, J.C., O'Neill, S.L. & Johnson, K.N. 2008. Wolbachia and Virus Protection in Insects. *Science* **322**: 702.

Hilgenboecker, K., Hammerstein, P., Schlattmann, P., Telschow, A. & Werren, J.H. 2008. How many species are infected with *Wolbachia*; a statistical analysis of current data. *FEMS Microbiol. Lett.* **281**: 215–220.

Hudson, R.R., Kreitman, M. & Aguadé, M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**: 153–159.

Hurst, G.D.D. & Jiggins, F.M. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proc. R. Soc. Lond. B Biol. Sci.* **272**: 1525–1534.

Hurst, G.D.D. & Majerus, M.E.N. 1993. Why do maternally inherited microorganisms kill males? *Heredity* **71**: 81–95.

Hurst, G.D.D., Hurst, L.D. & Majerus, M.E.N. 1993. Problems and paradigms: altering sex ratios: the games microbes play. *Bioessays* **15**: 695–697.

Jansen, V.A.A., Turelli, M. & Godfray, H.C.J. 2008. Stochastic spread of Wolbachia. *Proc. R. Soc. Lond. B Biol. Sci.* **275**: 2769–2776.

Janz, N. 1998. Sex-linked inheritance of host-plant specialization in a polyphagous butterfly. *Proc. R. Soc. Lond. B* **265**: 1675–1678.

Janz, N. & Nylin, S. 2008. The oscillation hypothesis of host plant-range and speciation. In: *Specialization, Speciation, and Radiation: the Evolutionary Biology of Herbivorous Insects*. (K.J. Tilmon, ed.) pp. 203–215.University of California Press, Berkeley, California.

Janz, N., Nylin, S. & Wedell, N. 1994. Host plant utilization in the comma butterfly: sources of variation and evolutionary implications. *Oecologia* **99**: 132–140.

Janz, N., Nylin, S. & Nyblom, K. 2001. Evolutionary dynamics of host plant specialization: a case study of the tribe Nymphalini. *Evolution* **55**: 783–796.

Janz, N., Nylin, S. & Wahlberg, N. 2006. Diversity begets diversity: host expansions and the diversification of plant-feeding insects. *BMC Evol. Biol.* **6**: 4.

Jeyaprakash, A. & Hoy, M.A. 2000. Long PCR improves Wolbachia DNA amplification: wsp sequences found in 76% of sixty-three arthropod species. *Insect Mol. Biol.* **9**: 393–405.

Jiggins, F.M. 2003. Male-Killing Wolbachia and Mitochondrial DNA: selective Sweeps, Hybrid Introgression and Parasite Population Dynamics. *Genetics* **164**: 5–12.

Johnstone, R.A. & Hurst, G.D. 1996. Maternally inherited male-killing microorganisms may confound interpretation of mitochondrial DNA variability. *Biol. J. Linn. Soc.* **58**: 453–470.

Keller, L., Liautard, C., Reuter, M., Brown, W.D., Sundstrom, L. & Chapuisat, M. 2001. Sex ratio and Wolbachia infection in the ant Formica exsecta. *Heredity* **87**: 227–233.

Kodandaramaiah, U. & Wahlberg, N. 2007. Out-of-Africa origin and dispersal mediated diversification of the butterfly genus Junonia (Nymphalidae: Nymphalinae). *J. Evol. Biol.* **20**: 2181–2191.

Kodandaramaiah, U., Peña, C., Braby, M.F., Grund, R., Müller, C.J., Nylin, S. *et al.* 2009. Phylogenetics of Coenonymphina (Nymphalidae: Satyrinae) and the problem of rooting rapid radiations. *Mol. Phylogenet. Evol.* **54**: 386–394.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**: 209–220.

Meglecz, E. & Solignac, M. 1998. Microsatellite loci for *Parnassius mnemosyne* (Lepidoptera). *Hereditas* **128**: 179–180.

Meglecz, E., Petenian, F., Danchin, E., D'Acier, A.C., Rasplus, J. & Faure, E. 2004. High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*. *Mol. Ecol.* **13**: 1693–1700.

Mitter, C., Farrell, B. & Wiegmann, B. 1988. The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? *Am. Nat.* **132**: 107–128.

Narita, S., Nomura, M., Kato, Y. & Fukatsu, T. 2006. Genetic structure of sibling butterfly species affected by Wolbachia infection sweep: evolutionary and biogeographical implications. *Mol. Ecol.* **15**: 1095–1108.

Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York, USA.

Nice, C.C., Gompert, Z., Forister, M.L. & Fordyce, J.A. 2009. An unseen foe in arthropod conservation efforts: the case of Wolbachia infections in the Karner blue butterfly. *Biol. Conserv.* **142**: 3137–3146.

Nygren, G.H., Nylin, S. & Stefanescu, C. 2006. Genetics of host plant use and life history in the comma butterfly across Europe: varying modes of inheritance as a potential reproductive barrier. *J. Evol. Biol.* **19**: 1882–1893.

Nylin, S. 1988. Host plant specialization and seasonality in a polyphagous butterfly, *Polygonia c-album* (Nymphalidae). *Oikos* **53**: 381–386.

Nylin, S. & Janz, N. 1993. Oviposition preference and larval performance in *Polygonia c-album* (Lepidoptera: Nymphalidae):

the choice between bad and worse. *Ecol. Entomol.* **18**: 394–398.

Nylin, S. & Wahlberg, N. 2008. Does plasticity drive speciation? Host-plant shifts and diversification in nymphaline butterflies (Lepidoptera: Nymphalidae) during the tertiary. *Biol. J. Linn. Soc.* **94**: 115–130.

Nylin, S., Nygren, G., Söderlind, L. & Stefanescu, C. 2009. Geographical variation in host plant utilization in the comma butterfly: the roles of time constraints and plant phenology. *Evol. Ecol.* **23**: 807–825.

Oosterhout, C.V., Weetman, D. & Hutchinson, W.F. 2006. Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. *Mol. Ecol. Notes* **6**: 255–256.

Peña, C. & Wahlberg, N. 2008. Prehistorical climate change increased diversification of a group of butterflies. *Biol. Lett.* **4**: 274–278.

Peña, C., Nylin, S. & Wahlberg, N. 2011. The radiation of Satyrini butterflies (Nymphalidae: Satyrinae): a challenge for phylogenetic methods. *Zool. J. Linn. Soc.* **161**: 64–87.

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

R Development Core Team 2010. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rasgon, J.L., Styer, L.M. & Scott, T.W. 2003. *Wolbachia*-induced mortality as a mechanism to modulate pathogen transmission by vector arthropods. *J. Med. Entomol.* **40**: 125–132.

Raymond, M. & Rousset, F. 1995. An exact test for population differentiation. *Evolution* **49**: 1280–1283.

Rogers, A.R. & Harpending, H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9**: 552–569.

Ros, V.I.D., Fleming, V.M., Feil, E.J. & Breeuwer, J.A.J. 2009. How diverse is the genus *Wolbachia*? Multiple-gene sequencing reveals a putatively new Wolbachia supergroup recovered from spider mites (Acari: Tetranychidae). *Appl. Env. Micr.* **75**: 1036–1043.

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

Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* **8**: 103–106.

Savela, M. 2010. http://www.nic.funet.fi/pub/sci/bio/life/insecta/lepidoptera/ditrysia/papilionoidea/nymphalidae/nymphalinae/nymphalis/index.html. Accessed August 2010.

Scriber, J.M. & Lederhouse, R.C. 1992. The thermal environment as a resource dictating geographic patterns of feeding specialization of insect herbivores. In: *Effects of Resource Distribution on Animal-plant Interactions* (M.R. Hunter, T. Ohgushi & P.W. Price, eds), pp. 429–466. Academic Press, San Diego, CA.

Shaikevich, E.V., Vinogradova, E.B., Platonov, A.E., Karan, L.S. & Zakharov, I.A. 2005. Polymorphism of mitochondrial DNA and infection with symbiotic cytoplasmic bacterium Wolbachia pipientis in mosquitoes of the Culex pipiens (Diptera, Culicidae) complex from Russia. *Russ. J. Genet.* **41**: 244–248.

Shoemaker, D.D., Keller, G. & Ross, K.G. 2003. Effects of Wolbachia on mtDNA variation in two fire ant species. *Mol. Ecol.* **12**: 1757–1771.

Slatkin, M. & Hudson, R.R. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**: 555–562.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**: 437–460.

Teixeira, L., Ferreira, Á. & Ashburner, M. 2008. The Bacterial Symbiont Wolbachia Induces Resistance to RNA Viral Infections in Drosophila melanogaster. *PLoS Biol.* **6**: e1000002.

Templeton, A.R., Crandall, K.A. & Sing, C.F. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132**: 619–633.

Tennent, W.J. 1996. *The Butterflies of Morocco, Algeria and Tunisia*. Gem, Wallingford, Oxfordshire, UK.

Thompson, J.N. 1988. Evolutionary genetics of oviposition preference in swallowtail butterflies. *Evolution* **42**: 1223–1234.

Tolman, T. & Lewington, R. 1997. *Butterflies of Britain and Europe*. HarperCollins Publishers, London.

Turelli, M. & Hoffmann, A.A. 1991. Rapid spread of an inherited incompatibility factor in California Drosophila. *Nature* **353**: 440–442.

Turelli, M., Hoffmann, A.A. & McKechnie, S.W. 1992. Dynamics of Cytoplasmic Incompatibility and mtDNA Variation in Natural Drosophila simulans Populations. *Genetics* **132**: 713–723.

Tuzov, V.K. 2000. *Guide to the Butterflies of Russia and Adjacent Territories. Volume 2*. Pensoft, Sofia, Bulgaria.

Vandewoestijne, S., Baguette, M., Brakefield, P.M. & Saccheri, I.J. 2004. Phylogeography of *Aglais urticae* (Lepidoptera) based on DNA sequences of the mitochondrial COI gene and control region. *Mol. Phylogenet. Evol.* **31**: 630–646.

Via, S. 2001. Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol. Evol.* **16**: 381–390.

Wahlberg, N. & Saccheri, I. 2007. The effects of Pleistocene glaciations on the phylogeography of Melitaea cinxia (Lepidoptera: Nymphalidae). *Eur. J. Entomol.* **104**: 675–684.

Wahlberg, N. & Wheat, C.W. 2008. Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera. *Syst. Biol.* **57**: 231–242.

Wahlberg, N., Weingartner, E., Warren, A. & Nylin, S. 2009a. Timing major conflict between mitochondrial and nuclear genes in species relationships of *Polygonia* butterflies (Nymphalidae: Nymphalini). *BMC Evol. Biol.* **9**: 92.

Wahlberg, N., Leneveu, J., Kodandaramaiah, U., Peña, C., Nylin, S., Freitas, A.V.L. *et al.* 2009b. Nymphalid butterflies diversify following near demise at the Cretaceous/Tertiary boundary. *Proc. R. Soc. Lond. B Biol. Sci.* **276**: 4295–4302.

Weingartner, E., Wahlberg, N. & Nylin, S. 2006. Dynamics of host plant use and species diversity: a phylogenetic investigation in *Polygonia* butterflies (Nymphalidae). *J. Evol. Biol.* **19**: 483–491.

Weir, B.S. 1996. *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA.

Werren, J.H. 1997. Biology of *Wolbachia*. *Annu. Rev. Entomol.* **42**: 587–609.

Werren, J.H. & Windsor, D.M. 2000. *Wolbachia* infection frequencies in insects: evidence of a global equilibrium? *Proc. R. Soc. B* **267**: 1277–1285.

Zhang, D. 2004. Lepidopteran microsatellite DNA: redundant but promising. *Trends Ecol. Evol.* **19**: 507–509.

Zhou, W., Rousset, F. & O'Neil, S. 1998. Phylogeny and PCR-based classification of *Wolbachia* strains using wsp gene sequences. *Proc. R. Soc. B* **265**: 509–515.

Appendix 1

List of samples used in this study along with their collection locality, GenBank accession numbers and their haplotypes as identified here

Voucher	Collection locality	Haplotype	GenBank (cytochrome oxidase subunit I)	GenBank (ribosomal protein subunit 5)
EW28-24	Belgium, Antwerpen	*Widespread*	JN093199	JN093173
EW39-4	Finland, Kaarina	*Widespread*	JN093216	JN093167
EW46-20	Finland, Turku	*FI*	JN093238	–
EW46-21	Finland, Turku	*Widespread*	JN093239	–
EW23-5	France, Aquitaine, Dordogne, Le Buga	*FR*	JN093178	JN093166
EW23-7	France, La Court	*Widespread*	JN093204	–
EW23-6	France, Languedoc-Roussillon, Aude, Roquefére	*Widespread*	JN093176	–
EW19-12	Macedonia	*Widespread*	JN093185	–
EW19-13	Macedonia	*Widespread*	JN093184	–
EW19-14	Macedonia	*MK1*	JN093183	–
EW19-15	Macedonia	*MK2*	JN093182	JN093163
EW19-16	Macedonia	*Widespread*	JN093181	JN093164
EW38-3	Morocco	*Widespread*	JN093213	JN093160
EW38-4	Morocco	*Widespread*	JN093209	JN093161
EW38-5	Morocco	*MO*	JN093210	–
EW38-6	Morocco	*Widespread*	JN093212	–
EW38-8	Morocco, Haut Atlas septentrional, Tizi-n-Oufraou	*Widespread*	JN093217	JN093162
EW38-7	Morocco, Moyen Atlas central, Col du Zad et env.	*Widespread*	JN093214	–
EW26-32	Morocco, Moyen Atlas central, Djebel Tarharhat	*Widespread*	JN093175	–
EW38-1	Morocco, Moyen Atlas central, Env. Azrou	*Widespread*	JN093211	–
EW38-2	Morocco, Moyen Atlas central, Env. Azrou	*Widespread*	JN093215	–
EW13-1	Russia, Buryatia Republic, Baruzinskiy Mts	*RU1*	JN093177	–
EW13-2	Russia, Buryatia Republic, Baruzinskiy Mts	*Widespread*	JN093207	–
EW13-5	Russia, Buryatia Republic, Mondy	*Widespread*	JN093195	–
EW13-6	Russia, Buryatia Republic, Mondy	*Widespread*	JN093194	–
EW13-14	Russia, Chita Region, Udocan Mts	*Widespread*	JN093198	JN093155
EW13-15	Russia, Chita Region, Udocan Mts	*Widespread*	JN093205	–
EW13-16	Russia, Chita Region, Udocan Mts	*Widespread*	JN093201	–
EW14-6	Russia, Chita Region, Udocan Mts	*Widespread*	JN093206	JN093172
EW37-2	Russia, Irkutsk region, Slyudyanka valley	*Widespread*	JN093218	–
EW26-2	Russia, Primorskiy Krai	*RU2*	JN093174	–
EW13-3	Russia, Primorskiy Krai, Vladivostok	*Widespread*	JN093197	–
EW13-4	Russia, Primorskiy Krai, Vladivostok	*Widespread*	JN093196	–
EW13-7	Russia, Primorskiy Krai, Vladivostok	*Widespread*	JN093200	JN093154
EW46-5	Spain, Catalonia	*Widespread*	JN093243	–
EW46-6	Spain, Catalonia	*Widespread*	JN093244	–
EW46-7	Spain, Catalonia	*Widespread*	JN093245	–
EW46-8	Spain, Catalonia	*Widespread*	JN093246	–
EW45-1	Spain, Catalonia, Can Liro	*SP*	JN093219	–
EW45-2	Spain, Catalonia, Can Liro	*Widespread*	JN093220	–
EW45-7	Spain, Catalonia, Can Liro	*Widespread*	JN093225	–
EW46-1	Spain, Catalonia, Can Liro	*Widespread*	JN093237	–
EW46-2	Spain, Catalonia, Can Liro	*Widespread*	JN093240	–
EW46-3	Spain, Catalonia, Can Liro	*Widespread*	JN093241	–
EW46-4	Spain, Catalonia, Can Liro	*Widespread*	JN093242	–
EW23-1	Spain, Catalonia, Girona Prov., El Cortalet	*Widespread*	JN093180	–
EW23-2	Spain, El Cortalet	*Widespread*	JN093179	JN093165
EW23-3	Spain, El Cortalet	*Widespread*	JN093208	–
EW23-4	Spain, El Cortalet	*Widespread*	JN093203	–
EW45-6	Spain, El Puig	*Widespread*	JN093224	–
EW45-3	Spain, Vallfornars	*Widespread*	JN093221	JN093168
EW45-4	Spain, Vallfornars	*Widespread*	JN093222	JN093169
EW45-5	Spain, Vallfornars	*Widespread*	JN093223	–

Appendix 1 (*Continued*)

EW46-10	Sweden, Uppland, Åkersberga	*Widespread*	JN093227	–
EW46-11	Sweden, Uppland, Åkersberga	*Widespread*	JN093228	–
EW46-12	Sweden, Uppland, Åkersberga	*Widespread*	JN093229	–
EW46-13	Sweden, Uppland, Åkersberga	*Widespread*	JN093230	–
EW46-14	Sweden, Uppland, Åkersberga	*Widespread*	JN093231	–
EW46-9	Sweden, Uppland, Åkersberga	*Widespread*	JN093247	–
EW46-15	Sweden, Uppland, Häggvik	*SW2*	JN093232	–
EW46-16	Sweden, Uppland, Häggvik	*Widespread*	JN093233	–
EW46-17	Sweden, Uppland, Häggvik	*SW1*	JN093234	–
EW46-18	Sweden, Uppland, Häggvik	*Widespread*	JN093235	–
EW46-19	Sweden, Uppland, Häggvik	*SW5*	JN093236	JN093170
EW17-5	Sweden, Uppland, Stockholm	*Widespread*	JN093187	JN093158
EW17-6	Sweden, Uppland, Stockholm	*SW4*	JN093186	–
EW17-7	Sweden, Uppland, Stockholm	*Widespread*	JN093202	JN093159
EW17-11	Sweden, Uppland, Vallentuna	*SW3*	JN093192	–
EW17-15	Sweden, Uppland, Vallentuna	*SW1*	JN093191	–
EW17-1	UK, Oxford	*Widespread*	JN093193	–
EW17-2	UK, Oxford	*UK2*	JN093190	–
EW17-3	UK, Oxford	*UK1*	JN093189	–
EW17-4	UK, Oxford	*Widespread*	JN093188	JN093157
EW45-8	UK, Yorkshire, Bishop Wood	*Widespread*	JN093226	JN093171
EW47-1	UK, Yorkshire, Bishop Wood	*UK1*	JN093248	JN093156
EW47-2	UK, Yorkshire, Bishop Wood	*Widespread*	JN093249	–
EW47-3	UK, Yorkshire, Bishop Wood	*UK3*	JN093250	–
EW47-4	UK, Yorkshire, Bishop Wood	*Widespread*	JN093251	–
EW47-5	UK, Yorkshire, Bishop Wood	*Widespread*	JN093252	–
EW47-6	UK, Yorkshire, Bishop Wood	*Widespread*	JN093253	–
EW47-7	UK, Yorkshire, Bishop Wood	*Widespread*	JN093254	–
EW47-8	UK, Yorkshire, Bishop Wood	*Widespread*	JN093255	–
EW47-9	UK, Yorkshire, Bishop Wood	*UK2*	JN093256	–

Appendix 2

PCR primers used to determine *Wolbachia* strain diversity among populations. 'T °C' refers to annealing temperatures used. 'bp' indicates the length of the sequence in *P. c-album*

Gene	Encoding	Forward Primer	Reverse Primer	T °C	bp	References
tsZ	Cell division protein	ftsZ_F1 (ATYATGGARCATATAAARGATAG)	ftsZ_R1 (TCRAGYAATGGATTRGATAT)	54	435	Baldo *et al.* (2006)
gatB	Glutamyl-tRNA(Gln) amidotransferase, subunit B	gatB_F1 (GAKTTAAAYCGYGCAGGBGTT)	gatB_R1 (TGGYAAYTCRGGYAAAGATGA)	54	369	Baldo *et al.* (2006)
groEL	HSP 60	groEL-Fd (CAACRGTRGSRRYAACTGCDGG)	GroEL-Rd (GATADCCRCGRTCAAAYTGC)	50	483	Ros *et al.* (2009)
gltA	Citrate synthase	WgltAF1 (TACGATCCAGGGTTTGTTTCTAC)	WgltARev2 (CATTTCATACCACTGGGCAA)	52	632	Casiraghi *et al.* (2005)

Appendix 3

Observed and expected heterozygosity values for each locus and population. Bold values indicate observed heterozygosity values that are significantly lesser than the expected values. Data for Finland, Belgium and France are not shown as they were represented by less than four individuals. *Polalb 6* was not amplifiable in the Moroccan samples

	UK		Russia		Morocco		Sweden		Spain		Macedonia	
Locus	Exp. *H*	Obs. *H*	Exp. *H*	Obs. *H*	Exp. *H*	Obs. *H*	Exp. *H*	Obs. *H*	Exp. *H*	Obs. *H*	Exp. *H*	Obs. *H*
Polalb 2	0.62	**0.00**	0.89	**0.53**	0.86	1.00	0.90	**0.31**	0.82	**0.18**	0.53	0.00
Polalb 5	0.94	**0.36**	0.94	**0.69**	0.54	0.67	0.93	**0.25**	0.93	**0.28**	0.96	0.75
Polalb 6	0.87	**0.67**	0.92	**0.43**	0.00	0.00	0.90	**0.50**	0.91	**0.53**	0.93	0.60
Polalb 7	0.88	**0.36**	0.89	**0.62**	0.62	**0.00**	0.90	**0.56**	0.95	**0.50**	0.93	**0.50**
Polalb 8	0.87	**0.21**	0.84	**0.57**	0.22	0.22	0.88	**0.29**	0.80	**0.05**	0.80	**0.40**
Polalb 10	0.89	**0.50**	0.94	**0.55**	0.78	**0.14**	0.93	**0.38**	0.92	**0.43**	0.80	0.00
Polalb 11	0.83	0.86	0.70	**0.40**	0.45	0.44	0.85	0.81	0.86	0.84	0.78	0.80
Polalb 17	0.91	**0.50**	0.89	**0.33**	0.75	**0.00**	0.91	**0.08**	0.88	**0.27**	0.89	**0.25**
Polalb 20	0.88	0.79	0.92	0.86	0.71	**0.56**	0.88	0.86	0.86	0.95	0.82	0.60
Polalb 22	0.88	**0.50**	0.90	**0.50**	0.75	**0.22**	0.89	**0.60**	0.92	**0.53**	0.89	**0.25**

Received 14 October 2010; revised 26 May 2011; accepted 30 May 2011