

## Mitochondrial DNA variation of *Drosophila obscura* (Diptera: Drosophilidae) across Europe

PAVLE ERIĆ<sup>1,5</sup> , MARINA STAMENKOVIĆ-RADAK<sup>2,5</sup> , MILAN DRAGIĆEVIĆ<sup>1</sup> , MAARIA KANKARE<sup>3,5</sup> , MEGAN A. WALLACE<sup>4,5,\*</sup> , MARIJA SAVIĆ VESELINOVIĆ<sup>2,5</sup>  and MIHAILO JELIĆ<sup>2,5</sup> 

<sup>1</sup> Institute for Biological Research “Siniša Stanković”, National Institute of Republic of Serbia, University of Belgrade, Bulevar despota Stefana 142, 11000 Belgrade, Serbia; e-mails: pavle.eric@ibiss.bg.ac.rs, mdragicevic@ibiss.bg.ac.rs

<sup>2</sup> University of Belgrade, Faculty of Biology, Studentski trg 16, 11000 Belgrade, Serbia; e-mails: marina@bio.bg.ac.rs, marijas@bio.bg.ac.rs, mihailoj@bio.bg.ac.rs

<sup>3</sup> Department of Biological and Environmental Science, University of Jyväskylä, Survantie 9, 40014 Jyväskylä, Finland; e-mail: maaria.kankare@jyu.fi

<sup>4</sup> Institute of Evolutionary Biology, University of Edinburgh, Charlotte Auerbach Road, Kings Buildings, Edinburgh EH9 3FL, UK

<sup>5</sup> The European Drosophila Population Genomics Consortium (DrosEU)

**Key words.** *Cyt b*, genetic variation, population expansion, phylogeography

**Abstract.** *Drosophila obscura* is a common fruit fly that inhabits the temperate forests of Europe. While it is abundant in the north compared to other *Drosophila*, its density decreases southwards, where it is gradually replaced by other *Drosophila* species. This study describes variation in the mitochondrial *Cyt b* gene of *D. obscura* from several European populations. We observed a large number of haplotypes, together with the structuring of genetic variation. Genetic variation is higher in the west where O1 and related divergent haplotypes dominate. In the east, the O2 haplotype is most frequent, together with haplotypes that recently arose from it. In the central part of the species range, both O1 and O2 are equally present, along with many others. These data reveal signs of population expansions that probably happened earlier in the west, and more recently in the east. Though our conclusions are based on only one genetic marker, limiting the power of the analysis, the results imply either postglacial expansion from two unique sources or, more likely, eastwards stepping-stone expansion. This study adds important information on genetic variation and phylogeography to the obscure biology of *D. obscura*, a species that has the potential to become an interesting model in evolutionary biology and conservation genetics.

### INTRODUCTION

*Drosophila obscura*, the nominate species of the obscura group, is a Palearctic fruit fly whose distribution extends from Southern Europe to the central Fennoscandia, and from Western Europe to central Asia (Lakovaara & Saura, 1971; Brehm & Krimbas, 1991). In Europe, its relative abundance decreases from north to south (Brehm & Krimbas, 1991). In southern Finland, it is the most abundant *Drosophila* species (Lakovaara & Saura, 1971). Along with *D. subobscura*, *D. obscura* is the one of the two most common *Drosophila* species in Great Britain, although *D. subobscura* outnumbers it during most months (Shorrocks, 1975; Begon, 1978). In the Central Balkans, among obscura group species, it is much less abundant than *D. subobscura* (Stanić et al., 2002; Pavković-Lučić et al., 2012), but is readily collected at higher altitudes (long-term observa-

tion of the authors). Among obscura group species it forms a monophyletic triad with the two most related species: *D. ambigua* and *D. trisits* (Bachmann & Sperlich, 1993; Gao et al., 2007). Although these three species live in sympatry, *D. obscura* is more common (Stamenković-Radak et al., 2003) especially at higher altitudes compared to the other two, which are rarely collected there (long-term observation of the authors).

Species from the obscura group, especially American *D. pseudoobscura* and European *D. subobscura*, have long served as models in evolutionary biology (Anderson et al., 1975; Krimbas, 1993; Powell, 1997; Schaeffer et al., 2003; Balanyá et al., 2006; Savić Veselinović et al., 2019). The genetic variation of natural populations of *D. subobscura* has been well described using phenotypic, chromosomal, and molecular markers (Krimbas, 1993; Pascual

\* Current address: College of Life and Environmental Sciences, University of Exeter, Penryn Campus, Cornwall TR10 9FE, UK; e-mail: M.A.Wallace@exeter.ac.uk

et al., 2001; Fragata et al., 2010; Savić Veselinović et al., 2019). Particularly interesting is the pattern of mitochondrial (mtDNA) variation found in natural populations of *D. subobscura*, which has proven to be an excellent model for studying the selective forces that maintain sympatric mtDNA variation (Jelić et al., 2015; Savić Veselinović et al., 2019; Kurbalija Novičić et al., 2020).

In contrast to *D. subobscura*, limited data is available on genetic variation in natural populations of its sympatric counterpart *D. obscura*. Chromosomal inversion polymorphism has been studied in 13 isofemale strains (IFSs) collected across the European continent (Brehm & Krimbas, 1991). The sample size was limited, but the number of detected inversions would suggest that chromosomal variation in *D. obscura* is similar to that in *D. subobscura*. Genetic variation has also been assessed by enzyme loci in multiple populations from Fennoscandia (Lakovaara & Saura, 1971) where a large numbers of polymorphic loci were observed. However, so far, mtDNA variation has not been assessed in this species. There is also a lack of *D. obscura* population studies that cover wide geographic areas. This species is easily collected in the wild, easily bred in the laboratory, and is a promising model for studying historical and adaptive processes that have shaped the genetic variation of natural populations of the Palearctic.

In this paper, we describe variation in an 893 bp long sequence of the mitochondrial Cytochrome b (*Cyt b*) gene from several European populations of *D. obscura*. We record significant genetic differentiation among different regions of Europe, as well as different levels of within-population variation. We discuss our findings in light of colonisation from glacial refugia, admixture, and more recent processes that could have influenced the observed pattern of variation.

**MATERIALS AND METHODS**

Samples of *D. obscura* were collected from four European countries: Serbia, Finland, Germany, and Scotland, UK (referred to as four populations in the further text), covering a wide geographic range (Fig. 1). Samples from Serbia cover several distinct localities. Table 1 contains information on the collection sites, date of collection, and the number of specimens.

DNA was extracted using a method that enriches genomic DNA with mtDNA (Martinez et al., 1992). For Serbian samples, extraction was done from the F<sub>1</sub> progeny of the females collected in the wild. Extraction was conducted separately for the progeny of different females. For the rest of the samples, DNA was extracted from wild-caught individuals that had been kept in ethanol, and the final step of alkaline lysis was excluded to obtain enough DNA. An 893 bp fragment corresponding to the mitochondrial *Cyt b* gene was PCR amplified and sequenced with the following primers: *Cyt b*-F 5'-TTAT GGTT GATT ATTA CGAA-3' and *Cyt b*-R 5'- CAAA ACAT ATGC TTAT TCAA-3' (Gao et al., 2007). The PCR cycling conditions consisted of an initial denaturation step at 94°C for 3 min, 35 cycles: at 94°C for 50 s, 51.5°C for 1 min, and 72°C for 1 min; with a final extension at 72°C for 3 min. Amplified products were purified using the QIAquick PCR Purification kit (QIAGEN, Hilden, Germany), and sequencing reactions were performed using both primers (Macrogen inc. Amsterdam, The Netherlands). The obtained sequences were aligned in BioEdit v.7.2.5 (Hall, 2011), using the ClustalW algorithm with



**Fig. 1.** Map of sampling locations of *Drosophila obscura* throughout Europe. A magnified map of Serbia is given bottom right since more sampling locations are present in close proximity. Base map was taken from Google Maps, 2021 [https://www.google.com/maps/@51.0375553,24.7087047,2977433m/data=!3m1!1e3!5m1!1e4], June 8, 2021.

default parameters (Thompson et al., 1994). The final analysis included 185 sequences.

A Median Joining network (Bandelt et al., 1999) was calculated and plotted in R v4.1.0 (R Development Core Team, 2018), using the pegas package (Paradis, 2010). Epsilon was set to 0, to put the fewest possible median haplotypes in the network. A *Cyt b* sequence of *D. ambigua* (IFS R42, collected at Mt. Rtanj in 2014) was used as an outgroup. The network was plotted to show only single alternative mutational steps between median haplotypes.

Bayesian inference of phylogeny implemented in Beast v2.6.2 (Bouckaert et al., 2019) was used to infer tree topology and then to construct a phylogenetic tree. Sequences of closely related species *D. ambigua* and *D. tristis* (NCBI accession number: EF216284.1) were used as outgroups to root the tree. We selected the best-fit substitution model for these data using likelihood ratio tests and Akaike information criterion (Akaike, 1973), implemented in jModelTest v2.1.10 (Posada, 2008). The best-fit model was a Tamura-Nei model of nucleotide substitution with a significant proportion of invariable sites (I) and gamma (G) distributed among-site rate heterogeneity (TnR+I+G). The Markov chain

**Table 1.** Sampling details. No. males/females – number of sampled males/females.

Country / Population	Locality	Code	Coordinates of Sampling sites	year	No. males / females
Serbia	Mt. Goč	SG	43°33'29.10"N	2015	24 females
			20°45'17.40"E		
			43°32'54.00"N		
			20°47'11.30"E		
Serbia	Mt. Tara	ST	43°56'58.09"N	2016	34 females
			19°21'27.18"E		
			43°55'04.37"N		
			19°25'13.19"E		
Serbia	Mt. Stara	SS	43°22'27.34"N	2016	20 females
			22°37'02.28"E		
			43°20'12.76"N		
Serbia	Mt. Maljen	MS	22°41'42.99"E	2017	2 females
			44°06'18.31"N		
Finland	Akaa	FA	19°58'56.55"E	2018	30 males/6 females
			61°08'58.69"N		
Germany	Fohrberg	GF	23°31'26.33"E	2015	17 males/4 females
			48°13'04.43"N		
Scotland, UK	Edinburgh	SC	07°49'09.10"E	2017–2019	35 males/13 females
			55°55'23.13"N		
			03°10'24.02"W		

Monte Carlo (MCMC) search was started from random trees and was run for  $5 \times 10^8$  generations, and the current tree was saved to file every 1000 generations. This generated an output of  $5 \times 10^5$  trees. After discarding  $10^8$  states (20% of samples) as “burn-in”, the remaining samples were summarized into a 50% majority-rule consensus tree, with clade posterior probabilities to approximate the posterior distribution of the phylogeny, calculated using TreeAnnotator v2.1.2 (BEAST package). The final tree was visualized and edited with FigTree v1.3.1 (<http://tree.bio.ed.ac.uk/software/figtree>). Mixing and convergence among chains, as well as stabilization of likelihood and parameter values, were assessed using the effective sampling size criterion (ESS values  $>200$  were considered acceptable) in Tracer v1.7 (Rambaut et al., 2018).

We then calculated nucleotide and haplotype diversity. Tajima's D (Tajima, 1989), Fu's  $F_s$  (Fu, 1997), Fu and Li's D-F (Fu & Li, 1993), and Fay and Wu's H (Fay & Wu, 2000) tests were used to test departure from mutation-drift equilibrium. We used several tests due to their differing statistical approaches, and power to infer demographic history (Ramos-Onsins & Rozas, 2002). This was important since populations differed in the number of analysed sequences. For Fu and Li's and Fay and Wu's tests, a *Cyt b* sequence from the most closely related species *D. ambigua* (IFS R42, collected at Mt. Rtanj in 2014) served as an outgroup. A McDonald-Kreitman test (McDonald & Kreitman, 1991) was implemented to compare the ratio of non-synonymous to synonymous change within and between species for *D. obscura* and *D. ambigua*. Changes in population size were examined by calculating the observed and expected pairwise differences (mismatch distribution) (Rogers & Harpending, 1992). The above parameters and tests were conducted using DNASP v.6.0 (Librado & Rozas, 2009). An extended Bayesian skyline plot (EBSP) (Heled & Drummond, 2008) was applied to additionally infer demographic history using BEAST2 (Bouckaert et al., 2019), with an appropriate substitution rate for *D. melanogaster* mtDNA (Haag-Liautard et al., 2008), and an assumption of 4 generations per year (Begon, 1978) which corresponds to the clock rate of 0.248. The length of the Markov chains was set to  $10^9$  for the EBSP, logging the parameters every 3000 iterations. Burn-in was set to discard 25% of the samples. Tracer (Rambaut et al., 2018) was used to assess the convergence of the chains.

An analysis of molecular variance (AMOVA) was implemented in order to partition variation between and within populations that represent the four geographical regions. An AMOVA was also conducted separately for the Serbian population that is comprised of four localities (with the exclusion of MS locality where only two individuals were available). Pairwise  $F_{ST}$  indices were calculated using Arlequin v.3.5.1.2 (Excoffier & Lischer, 2010). The Mantel test (Mantel, 1967) in Arlequin v.3.5.1.2 was used to test for a significant isolation-by-distance correlation. The Mantel test compares  $F_{ST}$  genetic distances with log e-transformed Euclidean spatial distances in kilometers. Another form of mantel test was employed, using two matrices of log e coordinates, one for longitude and one for latitude, to be compared to the  $F_{ST}$  matrix.

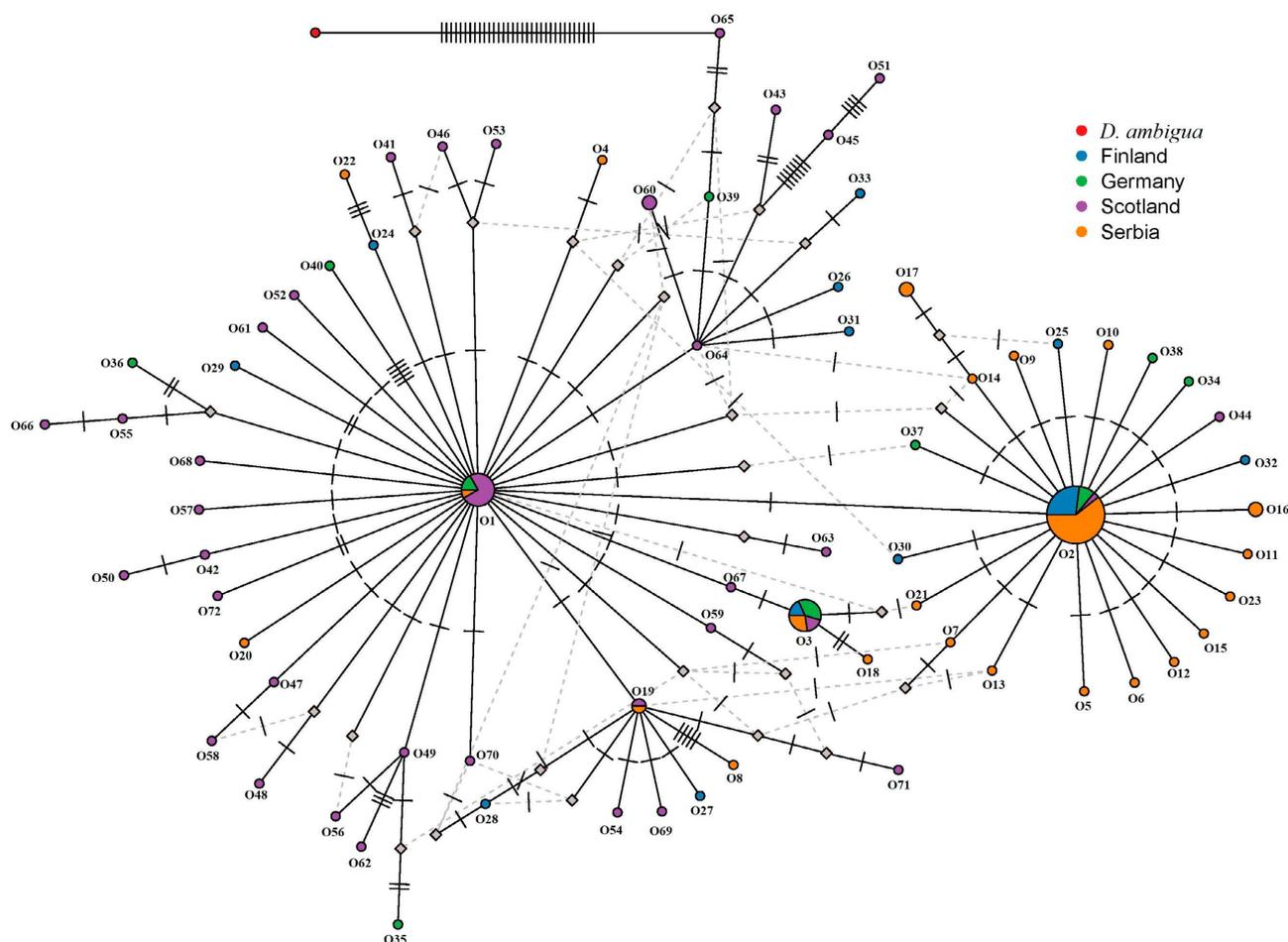
All samples were tested for the presence of *Wolbachia*. A PCR assay using 16S rDNA *Wolbachia*-specific primers (O'Neill et al., 1992) was used with a slight modification to the PCR conditions (García-Martínez et al., 1998). *Drosophila tristis* with known infection status served as a positive control (Erić et al., 2019). To exclude the possible presence of other maternally-transmitted bacteria (Hurst & Jiggins, 2005) we conducted microbiome sequencing. Two samples were made, each using 10 females originating from 10 different randomly chosen IFSs from Serbia. One sample included the O1 haplotype and 9 others that share its specific substitution on position 828. The other included the O2 hap-

lotype and 9 others with its specific variant on position 828. At the time of the analysis, these lines had been kept in the laboratory for 5 generations. DNA was extracted from pooled individuals using a previously published protocol for microbiome sequencing in *Drosophila* (Kapun et al., 2020). The microbiome sequencing was performed by Fisabio (Valencia, Spain) and included Illumina 16S V3–V4 amplicon library preparation and MiSeq 300 bp paired-end sequencing. Primer sequences [forward: CCT ACG GGN GGC WGC AG, reverse: GAC TAC HVG GGT ATC TAA TCC (Klindworth et al., 2013)] were removed with BBDuk (<https://jgi.doe.gov/data-and-tools/bbtools/>) using kmer length 15 and allowing for 2 mismatches. The generated data was analysed in R v4.1.0 (R Development Core Team, 2018) using dada2 for error estimation, sequence denoising, merging, and chimera removal (Callahan et al., 2016) as described previously (Beribaka et al., 2021). In short, sequences were first trimmed to 240 bp for forward and 210 bp for reverse reads, and all sequences containing more than 2 or 4 expected errors (for forward and reverse reads respectively) were discarded. After denoising, sequence pairs were merged using a minimum overlap of 20 bases without mismatches, and all sequences shorter than 400 or longer than 428 were discarded. Taxonomy assignment up to genus level was performed using the Silva v132 (<https://www.arb-silva.de/documentation/release-132/>) database with the RDP Naive Bayesian Classifier algorithm (Wang et al., 2007) as implemented in dada2 with default parameters.

## RESULTS

Among the 185 obtained sequences, we detected 72 different haplotypes for the *Cyt b* gene (NCBI accession numbers: MZ337620.1–MZ337804.1). There were 73 segregating sites, among which two had three alternating nucleotides. Among all segregating sites, 59 were synonymous, and 16 non-synonymous substitutions (Table S1). Moreover, 43 changes were observed only once, while 32 were observed in multiple individuals. One specific substitution (G to A on position 828 which changes valine to isoleucine) is of particular interest due to its presence in 73 individuals. The variant with valine is particularly frequent in Serbia (0.875), followed by Finland (0.750), and Germany (0.524), whilst it is rare in Scotland, UK (0.083). The lowest diversity was recorded in Serbia ( $\pi = 0.0146$ ,  $hd = 0.546$ ), while the highest was in Scotland ( $\pi = 0.0387$ ,  $hd = 0.964$ , Table 1). The majority of haplotypes were singletons.

A Median Joining network of *D. obscura* *Cyt b* haplotypes is presented in Fig. 2. The O1 haplotype is the most common in Scotland, together with a large number of very divergent haplotypes that share its specific substitution on position 828. Also, Scottish haplotypes are more closely related to the outgroup species *D. ambigua* sequence (Fig. 2). The O2 haplotype is particularly frequent in Serbia, where most other haplotypes share its specific substitution on position 828 and differ in only one mutational step from O2. In addition to the O1 and O2 haplotypes, O3 is also observed in more than two sampled individuals. O3 shares a specific substitution at position 828 with the O1 haplotype. Except for O1, which was not observed in Finland, these three most frequent haplotypes are present in all analysed regions. Haplotype O2 is the most frequent overall, maybe due to the overrepresentation of samples from Serbia.



**Fig. 2.** Median Joining Network based on an 893bp fragment of the *Cyt b* gene in *Drosophila obscura*. The size of circles corresponds to the number of individuals with the observed sequence. Dashes on lines represent nucleotide substitutions. Dashed lines represent alternative one mutational step reticulations. Different geographical regions are presented with different colours indicated in the legend within the figure.

Apart from splitting outgroup species as different clades, the Bayesian tree (Fig. S1) shows one sequence from Scotland (O65) as a monophyletic clade (PP = 1), and a few shallow nodes splitting sister haplotypes (PP > 0.7). After that, the phylogenetic relationships could not be resolved confidently, as the posterior probability is very low for the rest of the nodes (PP < 50%).

The results of Tajima’s D, Fu’s Fs, Fu and Li’s D-F, and Fay and Wu’s H are presented in Table 2. Negative values were observed in all cases, the majority of which were statistically significant. The McDonald-Kreitman test (Table 3) showed that the ratio of non-silent to silent variation was greater within species than between species. A statistically

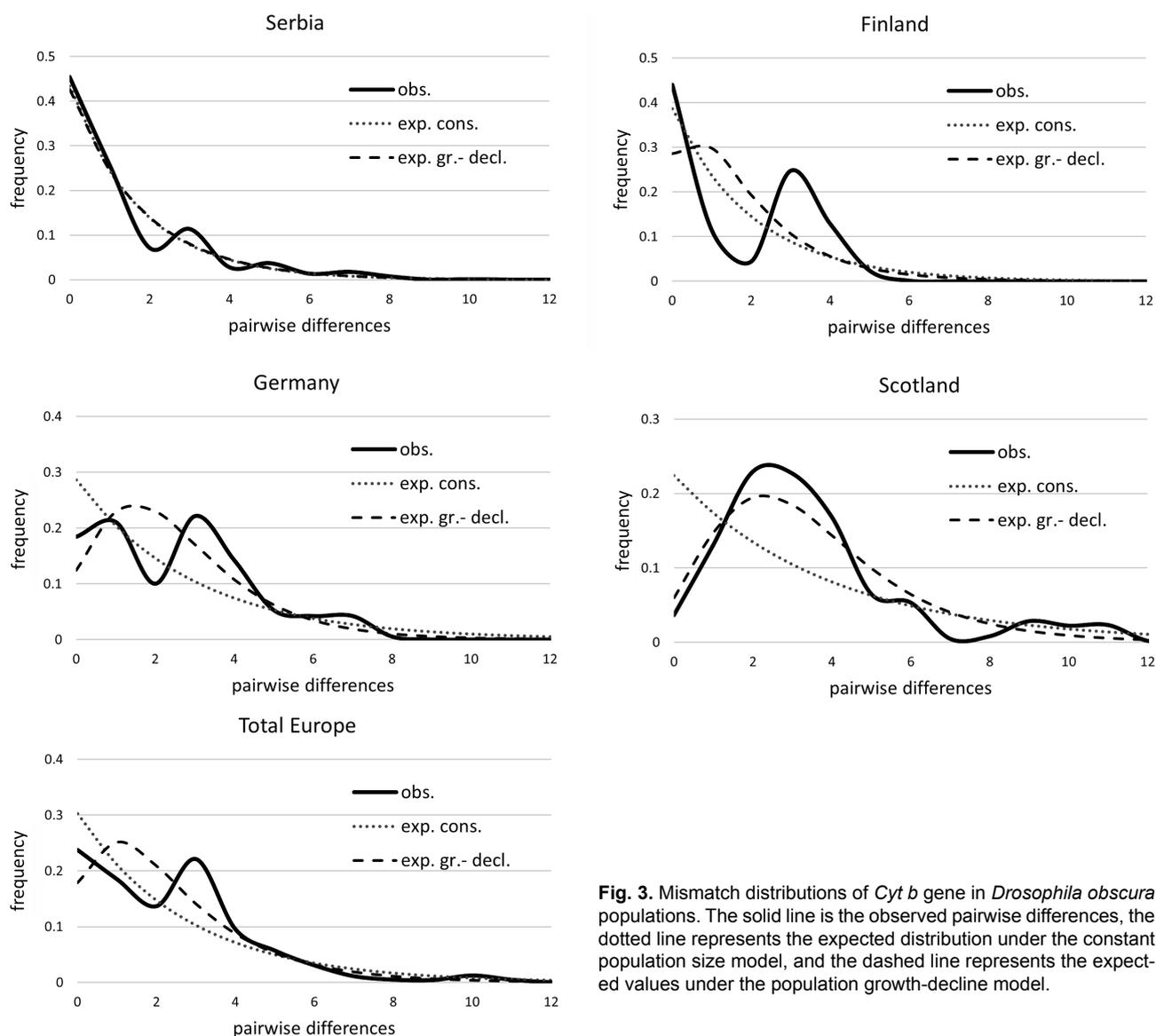
significant departure from neutrality was observed for the Serbian population (SR) and the German (GF) population.

Graphs of mismatch distribution are presented in Fig. 3. All graphs show some raggedness for the observed function. A unimodal distribution with a peak at 0 observed pairwise differences is found for sequences from Serbia. A unimodal distribution with a peak at 2–3 observed pairwise nucleotide differences is found in Scotland. A bimodal distribution was observed for sequences from Finland, Germany, and total Europe. Visual inspection shows that only samples from Scotland show a good fit to a population expansion scenario in contrast to constant population size.

**Table 2.** Parameters of genetic diversity and Tajima’s D, Fu’s Fs, Fu and Li’s D-F, and Fay and Wu’s H tests that measure departure from mutational-drift equilibrium.

Sample location	Diversity		Tajima		Fu		Fu & Li		Fay & Wu			
	$\pi$	hd	D	P	Fs	P	F	P	D	P	H	P
Serbia, SR	0.00146	0.546	<b>-2.451</b>	**	<b>-23.208</b>	***	<b>-3.846</b>	*	<b>-3.673</b>	*	<b>-13.839</b>	**
Finland, FA	0.00178	0.560	<b>-2.110</b>	*	<b>-6.241</b>	*	<b>-2.670</b>	*	-2.277	ns	<b>-9.305</b>	**
Germany, GF	0.00279	0.816	-1.661	ns	-2.429	ns	-1.860	ns	-1.514	ns	<b>-7.316</b>	*
Scotland UK, SC	0.00387	0.964	<b>-2.273</b>	**	<b>-42.140</b>	***	-2.322	ns	-1.684	ns	<b>-15.638</b>	*

$\pi$  – nucleotide diversity; hd – haplotype diversity; ns – nonsignificant; \* $P < 0.05$ ; \*\* $P < 0.02$ ; \*\*\* $P < 0.001$ , thresholds of  $P$  values are given based on the DNASP software; significant values are given in bold letters.



**Fig. 3.** Mismatch distributions of *Cyt b* gene in *Drosophila obscura* populations. The solid line is the observed pairwise differences, the dotted line represents the expected distribution under the constant population size model, and the dashed line represents the expected values under the population growth-decline model.

EBSP estimates of the time of demographic expansion are given in Fig. 4. The analysis with all sequences included shows that population expansion started around 3000 years ago (Fig. 4a). When analysing the Serbian population the graph gives signs of recent expansion that started around 1200 years ago but with notable uncertainty (Fig. 4b). The EBSP plot for samples from Scotland estimates population expansion to have started at around 8000 years ago, but the confidence interval gets wider as we go back in time (Fig. 4c). The German sample was very small and the EBSP function did not detect past population expansion,

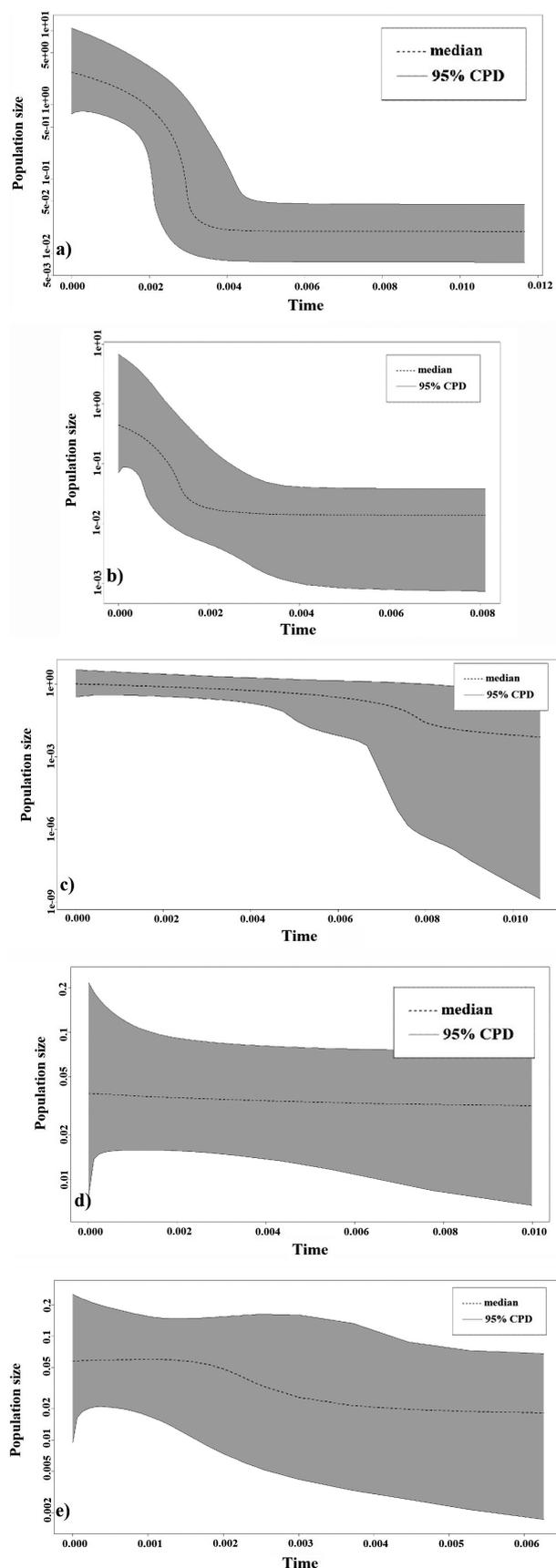
as the Markov chain did not converge even when we increased the number of iterations (Fig. 4d). The population from Finland, despite the small sample size, showed signs of a mild population expansion ranging from 3000 to 2000 years in the past, though with a wide confidence interval (Fig. 4e).

The results of the AMOVA show that 13.24% of the total variation is present among populations ( $F_{ST} = 0.1324, P < 0.001$ ). Pairwise  $F_{ST}$  values and the significance of their difference from zero are presented in Table 4. The most differentiated populations are Serbia and Scotland, while

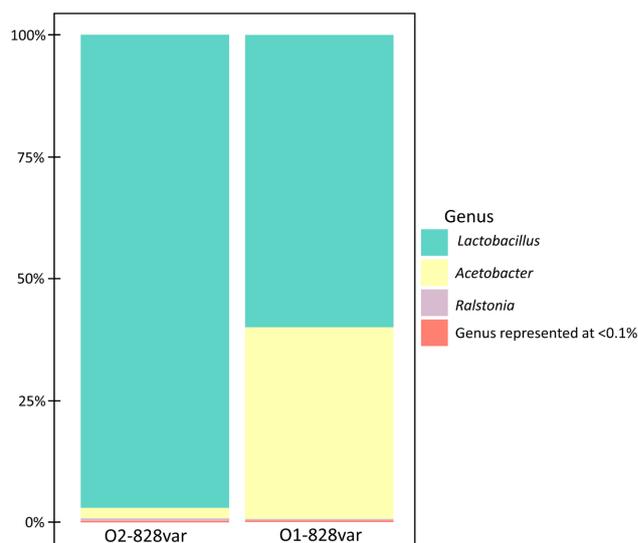
**Table 3.** Results of the McDonald-Kreitman test.

Sample location	<i>N</i>	<i>F</i> , <i>P</i>	<i>G</i>	<i>P</i>	<i>G</i> ( <i>W</i> )	<i>P</i>	<i>G</i> ( <i>Y</i> )	<i>P</i>
SR, Serbia	8.826	ns	<b>5.756</b>	*	<b>5.411</b>	*	<b>3.851</b>	*
FA, Finland	1.824	ns	0.680	ns	0.133	ns	0.116	ns
GF, Germany	<b>10.714</b>	*	<b>5.956</b>	*	<b>5.457</b>	*	<b>3.879</b>	*
SC, Scotland UK	4.550	ns	2.581	ns	2.412	ns	1.335	ns

*N* – neutrality index; *F*, *P* – Fisher’s exact test *P*-value (two tailed); *G* – test; *G* (*W*), *G* (*Y*) – *G* test with Williams’ and Yates’ correction respectively; ns – nonsignificant; \**P* < 0.05; \*\**P* < 0.02; \*\*\**P* < 0.001, thresholds of *P* values are given based on the DNASP software; significant values are given in bold letters.



**Fig. 4.** Population expansion of *Drosophila obscura* estimated with EBSP, for the total sample (a), Serbian population (b), Scottish population (c), German population (d), and Finnish population (e). Confidence intervals are given as shaded EBSP. Time is given in million years.



**Fig. 5.** Presence of bacterial genera in *Drosophila obscura* samples, obtained by microbiome sequencing of 16S V3–V4 region. O1–828var contains individuals with the O1 variant at the 828<sup>th</sup> nucleotide and O2–828var contains individuals with the O2 variant at the 828<sup>th</sup> nucleotide.

the least differentiated are those from Serbia and Finland. Germany shows a similar level of differentiation to all three other populations. For the Serbian population, where four localities were available, the AMOVA shows a great similarity between the samples. Only 2.27% of the variation is observed among different geographic localities from Serbia ( $F_{ST} = 0.02269, P < 0.05$ ).

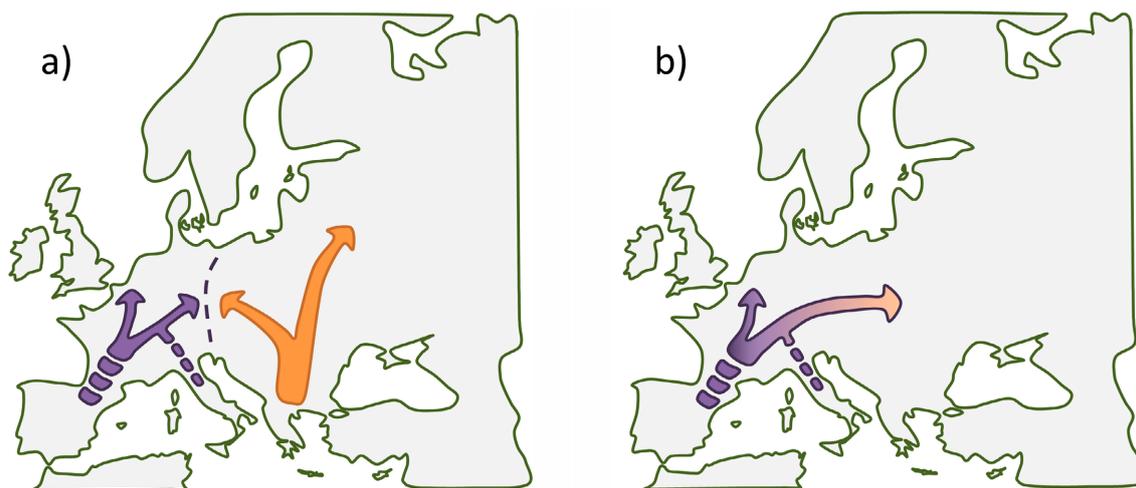
Significant isolation by distance is not observed in our data set. When geographical distances are used in the Mantel test, a positive correlation was observed ( $r = 0.304757; P = 0.330740$ ), which was not statistically significant. If latitude is considered, the correlation is negative, and also non-significant ( $r = -0.234862; P = 0.667790$ ). The highest correlation was observed for longitude ( $r = 0.735025; P = 0.082810$ ), which, though not statistically significant, shows signs of possible population differentiation along the East-West cline.

All samples were negative for the presence of *Wolbachia*. Microbiome sequencing of available IFSs shows a complete absence of maternally transmitted bacteria (not a single read was recorded for *Wolbachia*, *Spiroplasma*, *Microsporidia*, or *Rickettsia*) that could influence mtDNA variation (Fig. 5). *Lactobacillus* together with *Acetobacter* comprise more than 99% of the microbiota in both samples while *Ralstonia* comprises less than 0.5% and all the other genera represent less than 0.1%.

**Table 4.** Pairwise population differentiation calculated by  $F_{ST}$  values.

	Serbia	Finland	Germany
Finland	0.00174 ns		
Germany	<b>0.08485</b> ***	<b>0.02939</b> *	
Scotland	<b>0.2371</b> ***	<b>0.14119</b> ***	<b>0.05841</b> ***

ns – non significant; \* $P < 0.05$ ; \*\* $P < 0.02$ ; \*\*\* $P < 0.001$ ; significant values are given in bold letters.



**Fig. 6.** Maps of proposed postglacial expansion scenarios of *Drosophila obscura*. The non-continuous starting point of the arrows indicates uncertainty in the refugial source. (a) Colonisation from at least two different sources. Different colours indicate different refugia; (b) Expansion from western peninsulas with gradual colonization by eastwards stepping-stone expansion. Gradual colour change indicates a decrease in variation.

## DISCUSSION

In this study, we analyse nucleotide variation of the mitochondrial *Cyt b* gene in a widespread European *Drosophila* species, throughout its range. Our results show that *D. obscura* possesses a high level of mtDNA variation both within and between populations. The nucleotide diversity within populations observed for the *D. obscura* *Cyt b* gene (0.00146–0.00387) is generally higher compared to other *Drosophila* species: *D. melanogaster* (0.0009), *D. simulans* (0.0003), *D. yakuba* (0.0014) (Ballard & Kreitman, 1994) and *D. subobscura* (0.00087) (Erić et al., 2019). Sequences form a complex haplotype network with several star-shaped subnetworks, with a geographical structuring of genetic variation across Europe, mostly across the East-West axis. Interestingly, as O1 and O2 frequencies differ in the East and West, so do the frequencies of the other haplotypes that share their specific variants on the 828th nucleotide of the *Cyt b* gene. Variation is especially high in the western part of this species range. The haplotypes sharing the O1 variant on position 828 are more divergent from one another than the haplotypes sharing the O2 variant. Haplotypes sharing the O2 828 variant, in all but one case, differ from O2 by only one mutational step. Haplotypes O1 and O2 show similar frequencies in Central Europe. Although the number of individuals sampled from this area was restricted, and only one site from Germany was sampled, it is intuitive to conclude that the haplotype distribution in Central Europe is somewhere between eastern and western populations. The observed geographical structuring of mtDNA variation in *D. obscura* is especially interesting when compared to the sympatric *D. subobscura*, which shows geographic homogeneity in mtDNA variation across its entire range. During the past 40 years, studies of more than 30 populations of *D. subobscura* have shown that two main haplotypes are almost equally present in all populations and that there are less frequent population-specific haplotypes (Kurbalija Novičić et al., 2020).

Cytochrome b is highly conserved and the sequences are very similar, so posterior probabilities were low for all but a few nodes. Bayesian inference did not reconstruct the *Cyt b* phylogeny with high confidence, but we did draw a few conclusions from the analysis. The Bayesian tree shows that haplotypes that share a polymorphism at position 828 with the O1 haplotype, and that are frequent in the west, are indeed older and closer to *Drosophila ambigua*. The tree also shows that sequences that share the 828 O2 polymorphism are grouped together, but with a low posterior probability.

The observed excess of singletons and a large number of segregating sites are responsible for the negative values for the array of parameters that measure the departure of haplotype distribution from mutation-drift equilibrium. This result implies either population expansion or purifying selection. Positive values of neutrality indices from the McDonald-Kreitman test indicate an excess of non-silent polymorphism compared to divergence. This excess of amino acid polymorphism in the mtDNA, relative to divergence, is generally present in mice and humans (Nachman et al., 1994, 1996; Templeton, 1996), and also in *Drosophila* (Kaneko et al., 1993; Ballard & Kreitman, 1994; Rand et al., 1994), particularly for the *Cyt b* gene (Ballard & Kreitman, 1994; Erić et al., 2019). The observed pattern of non-silent polymorphism can be interpreted (Rand & Kann, 1996, 1998) in the light of the nearly neutral model (Ohta, 1992a, b) that predicts the accumulation of mildly deleterious alleles that persist for short time within a population and do not contribute to divergence (Nachman, 1998; Weinreich & Rand, 2000; Meiklejohn et al., 2007). Older, slightly deleterious haplotypes cannot become progenitors of new lineages due to natural selection (Grant, 2015). This could be the case for the haplotype pattern observed in Scotland. Recent increases in effective population size can also generate artefactual evidence of positive selection if substitutions are slightly deleterious and if there is no selection upon synonymous codon use (Eyre-Walker, 2002).

This may be the case for the O2 haplotype, and other haplotypes characteristic for Eastern Europe, whose network reflects only recent expansion.

Mismatch distribution analysis shows unimodal distributions for the Serbian population, with low mismatch values. Unimodal distributions with high mismatch values are detected for Scotland. A unimodal distribution of mismatches, which reflects demographic expansion, moves to higher values as mutations accumulate over time in a population (Grant, 2015), which implies earlier population expansion in the western range of the species.

The analysis of infection status with maternally transmitted microorganisms is important since they share the same mode of inheritance with mtDNA. They can often confound the inference of evolutionary history obtained by mtDNA markers, as variation is altered by selection acting on these microorganisms (Hurst & Jiggins, 2005). Importantly, all our samples were negative for *Wolbachia*. Additionally, we conducted microbiome sequencing on the Serbian samples, and maternally transmitted bacteria were also excluded (*Spiroplasma*, *Microsporidia*, and *Rickettsia*). Although we cannot exclude their presence in non-tested samples, the main haplotypes that are present Europe-wide were covered within our sample. While *Lactobacillus* and *Acetobacter*, which were detected in our samples, are common bacteria also found in IFSs of *D. subobscura* and *D. melanogaster* kept in the laboratory, *D. obscura* shows less diversity of bacterial genera compared to these two species (Beribaka et al., 2021).

What historical processes could have shaped the observed pattern of mtDNA variation in *D. obscura*? The different compositions of haplotypes in the east and the west could imply postglacial colonisation from at least two different sources with an admixture in central Europe (Fig. 6a). The Balkan Peninsula would be the hypothetical source of the O2 haplotype, while the source of the O1 haplotype could be the Iberian or Apennine peninsula. The Central European population perfectly matches this scenario, as it possesses eastern and western lineages in almost equal frequency. In addition, graphs of mismatch distribution show two peaks that correspond to the peaks of two distinct lineages. In that respect, postglacial colonisation of this species would follow the hedgehog's scenario (Hewitt, 2004). Although our EBSP intervals for population expansion are very wide, the results imply that the expansion probably happened earlier in the West than the East. Then the question arises as to why expansion after the last glacial maximum happened at different times from different glacial refugia, and why there is such a difference in the level of variation between different refugial sources?

One might imagine another scenario, of postglacial expansion from western peninsulas, or western cryptic refugia, and then gradual colonisation of the European continent by eastwards stepping-stone range expansion (Fig. 6b). In the East, the O2 haplotype might have increased its frequency to the detriment of other haplotypes due to the bottleneck effect, and subsequently generated an array of young singleton haplotypes in its recent expansion. In this

scenario, postglacial colonisation of this species would be more similar to the brown bear's scenario (Hewitt, 2004). This hypothesis is supported by the earlier expansion of *D. obscura* in the west, approximately 8000 years ago, while the O2 haplotype network might reflect only recent expansion after its colonisation of the highlands of the Balkan Peninsula. These expansions may also include those that generally happen yearly after winter or summer contractions, which are also observed in *D. subobscura* (Castro et al., 2010; Christie et al., 2010; Erić et al., 2019). Signs of ancient historical expansions may also have been masked over by annual contractions and expansions. It is also important to understand that the time of the expansion generated by EBSP used in this study is just a rough estimation based on a constant number of generations per year (Begon, 1978) which could have varied through time and across different regions.

We should also stress a specific aspect of this species' biology, which could be responsible for the observed pattern of structure, and level, of genetic variation. In the Balkan Peninsula, *D. obscura* is found in great numbers only in higher altitudes, while moving to the north it is frequently found in the lowlands too. The fact that it is the most common *Drosophila* species in southern Fennoscandia (Lakovaara & Saura, 1971) implies that it is adapted to colder climates. So, it is expected for southern *D. obscura* to show different population size dynamics compared to northern Europe. On the other hand, ecological conditions could be similar since vegetation and climate in higher altitudes in the south are similar to those from northern Europe. Populations from the south are more isolated and prone to genetic drift. Although populations can reach high numbers of individuals in the summer, they are restricted to small geographical areas and less prone to gene flow in the optimal part of the season that could restore variation.

To more accurately decipher the population history that has caused the observed pattern of variation in *D. obscura* mtDNA more sampling is needed, especially from the Apennine and the Iberian Peninsula, where *D. obscura* is more likely to be found in higher altitudes. Further analysis should also be conducted with additional genetic markers, both mitochondrial and nuclear (Brito & Edwards, 2009). This species' distribution in higher altitudes and latitudes also makes it a potential model for studying evolutionary change due to global warming. Although *Drosophila* flies can migrate easily, woodland habitats suitable for *D. obscura* cannot be formed at the pace dictated by global warming. Additionally, in some regions, there is not much space left for the flies to move, either to higher altitudes or northwards. Being very abundant, and easy to collect (particularly in northern Europe), designate and maintain in the laboratory, together with its rich mtDNA polymorphism, makes this species a promising model for these sorts of evolutionary studies.

**ACKNOWLEDGEMENTS.** This research is financed by the Ministry of Education, Science and Technological Development of the Republic of Serbia (451-03-9/2021-14/200178 for MSR, MSV, and MJ; 451-03-9/2021-14/ 200007 to PE and MD). The research

is supported through European *Drosophila* Population Genomics Consortium (DrosEU, which is funded by a Special Topic 587 Networks (STN) grant from the European Society for Evolutionary Biology). MK is supported by the Academy of Finland (grant no. 322980). MW was supported by the UK Natural Environmental Research Council through the E3 doctoral training programme (NE/L002558/1). We are grateful to D.J. Obbard and F. Staubach for providing us with samples of *D. obscura*. We are also grateful to M. Stefanović and I. Budinski for their useful advice in statistical analysis. We express the debt of gratitude for the support in this study from deceased Prof. Marko Anđelković, who taught us how to run in the field of population genetics with a broad view over the horizon. The authors declare no conflicts of interest.

## REFERENCES

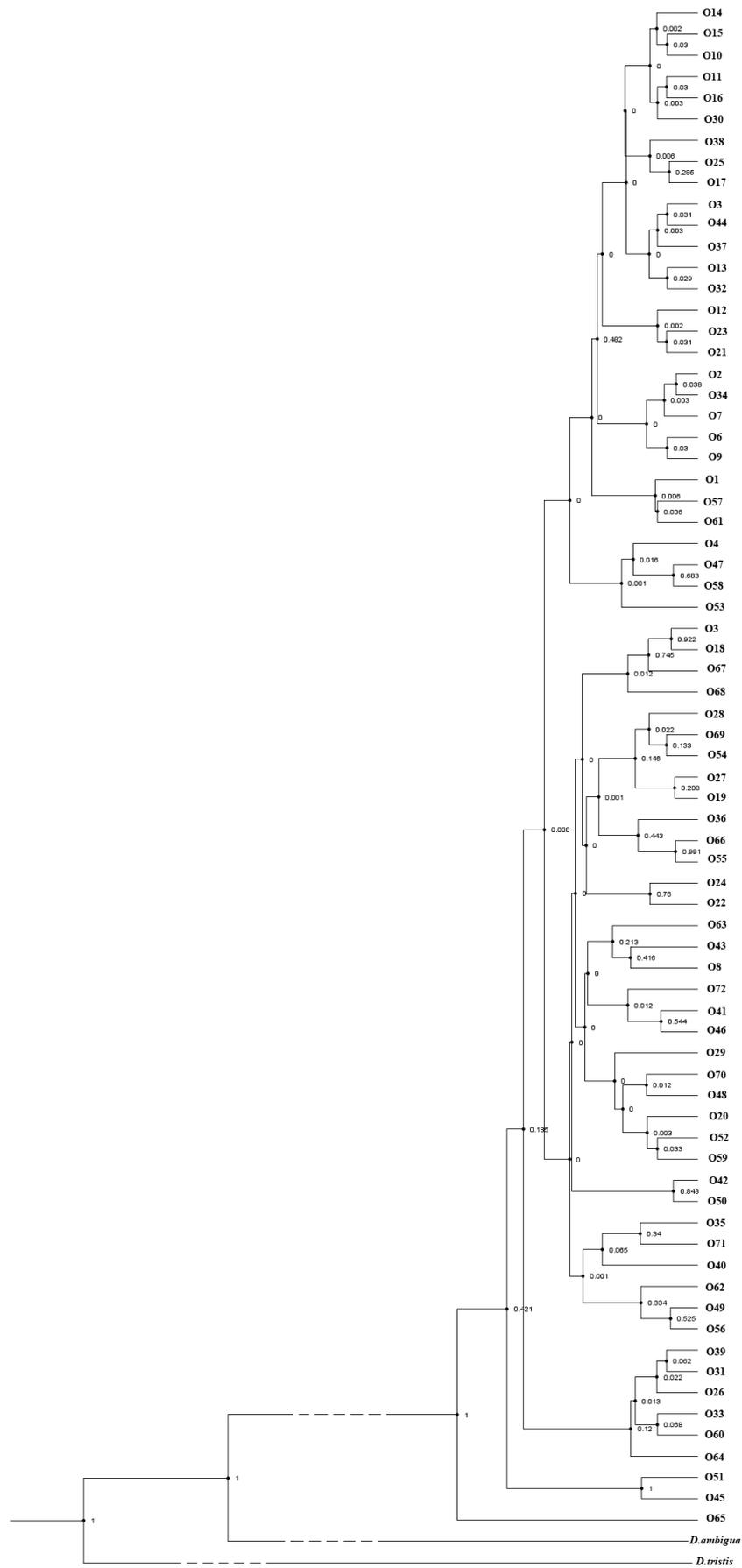
- AKAIKE H. 1973: Information theory and an extension of the maximum likelihood principle. In Petrov B.N. & Csaki F. (eds): *Proceedings of the 2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2–8, 1971*. Akademiai Kiado, Budapest, pp. 267–281.
- ANDERSON W., DOBZHANSKY T., PAVLOVSKY O., POWELL J. & YARDLEY D. 1975: Genetics of natural populations XLII. Three decades of genetic change in *Drosophila pseudoobscura*. — *Evolution* **29**: 24–36.
- BACHMANN L. & SPERLICH D. 1993: Gradual evolution of a specific satellite DNA family in *Drosophila ambigua*, *D. tristis*, and *D. obscura*. — *Mol. Biol. Evol.* **10**: 647–659.
- BALANYÁ J., OLLER J.M., HUEY R.B., GILCHRIST G.W. & SERRA L. 2006: Global genetic change tracks global climate warming in *Drosophila subobscura*. — *Science* **313**: 1773–1775.
- BALLARD J.W. & KREITMAN M. 1994: Unraveling selection in the mitochondrial genome of *Drosophila*. — *Genetics* **138**: 757–772.
- BANDELT H.J., FORSTER P. & ROHL A. 1999: Median-joining networks for inferring intraspecific phylogenies. — *Mol. Biol. Evol.* **16**: 37–48.
- BEGON M. 1978: Population densities in *Drosophila obscura* Fallén and *D. subobscura* Collin. — *Ecol. Entomol.* **3**: 1–12.
- BERIBAKA M., DIMKIĆ I., JELIĆ M., STANKOVIĆ S., PRŽULJ N., ANĐELKOVIĆ M. & STAMENKOVIĆ-RADAK M. 2021: Altered diversity of bacterial communities in two *Drosophila* species under laboratory conditions and lead exposure. — *Arch. Biol. Sci.* **73**: 17–29.
- BOUCKAERT R., VAUGHAN T.G., BARIDO-SOTTANI J., DUCHENE S., FOURMENT M., GAVRYUSHKINA A., HELED J., JONES G., KÜHNERT D., DE MAIO N. ET AL. 2019: BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. — *PLoS Comput. Biol.* **15**(4): e1006650, 28 pp.
- BREHM A. & KRIMBAS C.B. 1991: Inversion polymorphism in *Drosophila obscura*. — *J. Hered.* **82**: 110–117.
- BRITO P.H. & EDWARDS S.V. 2009: Multilocus phylogeography and phylogenetics using sequence-based markers. — *Genetica* **135**: 439–455.
- CALLAHAN B.J., McMURDIE P.J., ROSEN M.J., HAN A.W., JOHNSON A.J. & HOLMES S.P. 2016: DADA2: High-resolution sample inference from Illumina amplicon data. — *Nat. Methods* **13**: 581–583.
- CASTRO J.A., BARRIO E., GONZÁLEZ A., PICORNELL A., RAMON M.M. & MOYA A. 2010: Nucleotide diversity of a *ND5* fragment confirms that population expansion is the most suitable explanation for the mtDNA haplotype polymorphism of *Drosophila subobscura*. — *Genetica* **138**: 819–829.
- CHRISTIE J.S., PICORNELL A., MOYA A., RAMON M.M. & CASTRO J.A. 2010: Dynamics of the mtDNA haplotype variability in a *Drosophila subobscura* population over a two-year period. — *Open Evol. J.* **4**: 23–30.
- ERIĆ P., JELIĆ M., SAVIĆ-VESELINOVIĆ M., KENIG B., ANĐELKOVIĆ M. & STAMENKOVIĆ-RADAK M. 2019: Nucleotide diversity of *Cyt b* gene in *Drosophila subobscura* Collin. — *Genetika* **51**: 213–226.
- EXCOFFIER L. & LISCHER H.E.L. 2010: Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. — *Mol. Ecol. Resour.* **10**: 564–567.
- EYRE-WALKER A. 2002: Changing effective population size and the McDonald-Kreitman test. — *Genetics* **162**: 2017–2024.
- FAY J.C. & WU C.I. 2000: Hitchhiking under positive Darwinian selection. — *Genetics* **155**: 1405–1413.
- FRAGATA I., BALANYA J., REGO C., MATOS M., REZENDE E.L. & SANTOS M. 2010: Contrasting patterns of phenotypic variation linked to chromosomal inversions in native and colonizing populations of *Drosophila subobscura*. — *J. Evol. Biol.* **23**: 112–123.
- FU Y.X. 1997: Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. — *Genetics* **147**: 915–925.
- FU Y.X. & LI W.H. 1993: Statistical tests of neutrality of mutations. — *Genetics* **133**: 693–709.
- GAO J.J., WATABE H.A., AOTSUKA T., PANG J.F. & ZHANG Y.P. 2007: Molecular phylogeny of the *Drosophila obscura* species group, with emphasis on the Old World species. — *BMC Evol. Biol.* **7**: 87, 12 pp.
- GARCÍA-MARTÍNEZ J., CASTRO J.A., RAMÓN M., LATORRE A. & MOYA A. 1998: Mitochondrial DNA haplotype frequencies in natural and experimental populations of *Drosophila subobscura*. — *Genetics* **149**: 1377–1382.
- GRANT W.S. 2015: Problems and cautions with sequence mismatch analysis and Bayesian skyline plots to infer historical demography. — *J. Hered.* **106**: 333–346.
- HAAG-LIAUTARD C., COFFEY N., HOULE D., LYNCH M., CHARLESWORTH B. & KEIGHTLEY P.D. 2008: Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. — *PLoS Biol.* **6**(8): e204, 9 pp.
- HALL T. 2011: BioEdit: an important software for molecular biology. — *GERF Bull. Biosci.* **2**: 60–61.
- HELED J. & DRUMMOND A.J. 2008: Bayesian inference of population size history from multiple loci. — *BMC Evol. Biol.* **8**: 289, 15 pp.
- HEWITT G.M. 2004: Genetic consequences of climatic oscillations in the Quaternary. — *Philos. Trans. R. Soc. Lond. (B)* **359**: 183–195.
- HURST G.D. & JIGGINS F.M. 2005: Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. — *Proc. Biol. Sci.* **272**: 1525–1534.
- JELIĆ M., ARNQUIST G., NOVIČIĆ Z.K., KENIG B., TANASKOVIĆ M., ANĐELKOVIĆ M. & STAMENKOVIĆ-RADAK M. 2015: Sex-specific effects of sympatric mitonuclear variation on fitness in *Drosophila subobscura*. — *BMC Evol. Biol.* **15**: 135, 9 pp.
- KANEKO M., SATTÀ Y., MATSUURA E.T. & CHIGUSA S.I. 1993: Evolution of the mitochondrial ATPase 6 gene in *Drosophila*: unusually high level of polymorphism in *D. melanogaster*. — *Genet. Res.* **61**: 195–204.
- KAPUN M., BARRÓN M.G., STAUBACH F., OBBARD D.J., WIBERG R.A.W., VIEIRA J., GOUBERT C., ROTA-STABELLI O., KANKARE M., BOGAERTS-MÁRQUEZ M. ET AL. 2020: Genomic analysis of European *Drosophila melanogaster* populations reveals longitudinal structure, continent-wide selection, and previously unknown DNA viruses. — *Mol. Biol. Evol.* **37**: 2661–2678.

- KLINDWORTH A., PRUESSE E., SCHWEER T., PEPLIES J., QUAST C., HORN M. & GLOCKNER F.O. 2013: Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. — *Nucleic Acids Res.* **41**(1): e1, 11 pp.
- KRIMBAS C.B. 1993: *Drosophila subobscura: Biology, Genetics, and Inversion Polymorphism*. Dr. Kovac, Hamburg, 395 pp.
- KURBALIJA NOVIČIĆ Z., SAYADI A., JELIĆ M. & ARNQVIST G. 2020: Negative frequency dependent selection contributes to the maintenance of a global polymorphism in mitochondrial DNA. — *BMC Evol. Biol.* **20**(1): 20, 9 pp.
- LAKOVAARA S. & SAURA A. 1971: Genetic variation in natural populations of *Drosophila obscura*. — *Genetics* **69**: 377–384.
- LIBRADO P. & ROZAS J. 2009: DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. — *Bioinformatics* **25**: 1451–1452.
- MANTEL N. 1967: The detection of disease clustering and a generalized regression approach. — *Cancer Res.* **27**: 209–220.
- MARTINEZ D., MOYA A., LATORRE A. & FERERES A. 1992: Mitochondrial DNA variation in *Rhopalosiphum padi* (Homoptera: Aphididae) populations from four Spanish localities. — *Ann. Entomol. Soc. Am.* **85**: 241–246.
- MCDONALD J.H. & KREITMAN M. 1991: Adaptive protein evolution at the Adh locus in *Drosophila*. — *Nature* **351**: 652–654.
- MEIKLEJOHN C.D., MONTTOOTH K.L. & RAND D.M. 2007: Positive and negative selection on the mitochondrial genome. — *Trends Genet.* **23**: 259–263.
- NACHMAN M.W. 1998: Deleterious mutations in animal mitochondrial DNA. — *Genetica* **102–103**: 61–69.
- NACHMAN M.W., BOYER S.N. & AQUADRO C.F. 1994: Nonneutral evolution at the mitochondrial NADH dehydrogenase subunit 3 gene in mice. — *Proc. Natl. Acad. Sci. USA* **91**: 6364–6368.
- NACHMAN M.W., BROWN W.M., STONEKING M. & AQUADRO C.F. 1996: Nonneutral mitochondrial DNA variation in humans and chimpanzees. — *Genetics* **142**: 953–963.
- OHTA T. 1992a: The nearly neutral theory of molecular evolution. — *Annu. Rev. Ecol. Syst.* **23**: 263–286.
- OHTA T. 1992b: Theoretical study of near neutrality. II. Effect of subdivided population structure with local extinction and recolonization. — *Genetics* **130**: 917–923.
- O'NEILL S.L., GIORDANO R., COLBERT A.M., KARR T.L. & ROBERTSON H.M. 1992: 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. — *Proc. Natl. Acad. Sci. USA* **89**: 2699–2702.
- PARADIS E. 2010: pegas: an R package for population genetics with an integrated-modular approach. — *Bioinformatics* **26**: 419–420.
- PASCUAL M., AQUADRO C.F., SOTO V. & SERRA L. 2001: Microsatellite variation in colonizing and palearctic populations of *Drosophila subobscura*. — *Mol. Biol. Evol.* **18**: 731–740.
- PAVKOVIĆ-LUČIĆ S., SAVIĆ T., JELIĆ M., KENIG B., TANASKOVIĆ M., STAMENKOVIĆ-RADAK M. & ANDELKOVIĆ M. 2012: Note on the fauna of *Drosophila* (Diptera: Drosophilidae) and the first record of *Opomyza florum* (Diptera: Opomyzidae) from Mt. Goč, Serbia. — *Acta Entomol. Serb.* **17**: 45–51.
- POSADA D. 2008: jModelTest: phylogenetic model averaging. — *Mol. Biol. Evol.* **25**: 1253–1256.
- POWELL J.R. 1997: *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. Oxford University Press, New York, 776 pp.
- R DEVELOPMENT CORE TEAM 2018: *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. URL: <https://www.R-project.org/>
- RAMBAUT A., DRUMMOND A.J., XIE D., BAELE G. & SUCHARD M.A. 2018: Posterior summarization in bayesian phylogenetics using Tracer 1.7. — *Syst. Biol.* **67**: 901–904.
- RAMOS-ONSINS S.E. & ROZAS J. 2002: Statistical properties of new neutrality tests against population growth. — *Mol. Biol. Evol.* **19**: 2092–2100.
- RAND D.M. & KANN L.M. 1996: Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. — *Mol. Biol. Evol.* **13**: 735–748.
- RAND D.M. & KANN L.M. 1998: Mutation and selection at silent and replacement sites in the evolution of animal mitochondrial DNA. — *Genetica* **102**: 393–407.
- RAND D.M., DORFSMAN M. & KANN L.M. 1994: Neutral and non-neutral evolution of *Drosophila* mitochondrial DNA. — *Genetics* **138**: 741–756.
- ROGERS A.R. & HARPENDING H. 1992: Population growth makes waves in the distribution of pairwise genetic differences. — *Mol. Biol. Evol.* **9**: 552–569.
- SAVIĆ VESELINOVIĆ M., KURBALIJA NOVIČIĆ Z., KENIG B., JELIĆ M., PATENKOVIĆ A., TANASKOVIĆ M., PERTOLDI C., STAMENKOVIĆ-RADAK M. & ANDELKOVIĆ M. 2019: Local adaptation at fine spatial scale through chromosomal inversions and mito-nuclear epistasis: Findings in *Drosophila subobscura* (Diptera: Drosophilidae). — *Eur. J. Entomol.* **116**: 492–503.
- SCHAEFFER S.W., GOETTING-MINESKY M.P., KOVACEVIC M., PEOPLES J.R., GRAYBILL J.L., MILLER J.M., KIM K., NELSON J.G. & ANDERSON W.W. 2003: Evolutionary genomics of inversions in *Drosophila pseudoobscura*: evidence for epistasis. — *Proc. Natl. Acad. Sci. USA* **100**: 8319–8324.
- SHORROCKS B. 1975: The distribution and abundance of woodland species of British *Drosophila* (Diptera: Drosophilidae). — *J. Anim. Ecol.* **44**: 851–864.
- STAMENKOVIĆ-RADAK M., MILANOVIĆ M., SAVIĆ T. & ANDELKOVIĆ M. 2003: Adaptive significance of amylase polymorphism in *Drosophila* XIII. Old World obscura species subgroup divergence according to biochemical properties of  $\alpha$ -amylase. — *Genes Genet. Syst.* **78**: 23–28.
- STANIĆ S., KEKIĆ V. & PAVKOVIĆ-LUČIĆ S. 2002: A contribution to knowledge of Drosophilidae (Diptera) fauna in Kragujevac basin. — *Acta Entomol. Serb.* **1**: 151–154.
- TAJIMA F. 1989: Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. — *Genetics* **123**: 585–595.
- TEMPLETON A.R. 1996: Contingency tests of neutrality using intra/interspecific gene trees: The rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the hominoid primates. — *Genetics* **144**: 1263–1270.
- THOMPSON J.D., HIGGINS D.G. & GIBSON T.J. 1994: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. — *Nucleic Acids Res.* **22**: 4673–4680.
- WANG Q., GARRITY G.M., TIEDJE J.M. & COLE J.R. 2007: Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. — *Appl. Environ. Microbiol.* **73**: 5261–5267.
- WEINREICH D.M. & RAND D.M. 2000: Contrasting patterns of non-neutral evolution in proteins encoded in nuclear and mitochondrial genomes. — *Genetics* **156**: 385–399.

Received April 12, 2021; revised and accepted December 20, 2021

Published online February 9, 2022





**Fig. S1.** Bayesian phylogenetic tree with posterior probabilities given for each node. Long external dashed branches were shortened to make the figure more comprehensible.