

Sex-biased Migration and Demographic History of the Big European Firefly *Lampyris noctiluca*

Ana Catalán ^{1,*} Daniel Gygyax ¹ Ulrika Candolin ² Sergio Tusso ³ Pablo Duchén ⁴ Sebastian Höhna ^{5,6}

¹Division of Evolutionary Biology, Ludwig-Maximilians-Universität München, Planegg-Martinsried 82152, Germany

²Organismal and Evolutionary Biology Research Programme, University of Helsinki, Helsinki 00014, Finland

³Division of Genetics, Ludwig-Maximilians-Universität München, Planegg-Martinsried 82152, Germany

⁴Faculty of Biology, Institute for Organismal and Molecular Evolutionary Biology, Johannes Gutenberg University of Mainz, 55128 Mainz, Germany

⁵Department of Earth and Environmental Sciences, Paleontology & Geobiology, Ludwig-Maximilians-Universität München, Munich 80333, Germany

⁶GeoBio-Center, Ludwig-Maximilians-Universität München, Munich 80333, Germany

*Corresponding author: E-mails: ana.catalan@gmail.com, catalan@bio.lmu.de.

Associate editor: Amanda Larracuente

Abstract

Differential dispersion between the sexes can impact the colonization process and demographic history of a species. Here, we explored the demographic history of the big European firefly, *Lampyris noctiluca*, which exhibits female neoteny. Distribution of *L. noctiluca* extends throughout Europe, but nothing is known about its colonization process. To investigate its demographic history, we produced the first *Lampyris* genome (653 Mb), including an IsoSeq annotation and the identification of the X chromosome. We collected 115 individuals from six populations of *L. noctiluca* (Finland to Italy) and generated whole-genome re-sequencing data for each individual. We inferred several population expansions and bottlenecks throughout the Pleistocene that correlate with glaciation events. Surprisingly, we uncovered strong population structure and low gene flow. We reject a stepwise, south to north, colonization history scenario and instead uncovered a complex demographic history with a putative eastern European origin. Analyzing the evolutionary history of the mitochondrial genome as well as X-linked and autosomal loci, we found evidence of a maternal colonialization of Germany, putatively from a farther western European population, followed by a male-only migration from south of the Alps (Italy). Overall, investigating the demographic history and colonization patterns of a species should form part of an integrative approach of biodiversity research. Our results provide evidence of sex-biased migration which is important to consider for demographic, biogeographic and species delimitation studies.

Keywords: population genetics, StairwayPlot 2, PoMo, ABC, *de novo* genome assembly, glow-worm, Lampyridae, female neoteny

Introduction

Fireflies are a diverse group of bioluminescent beetles with a wide distribution around the globe (Lewis et al. 2020). In Central Europe, around 38 firefly species have been reported (www.gbif.org), for which genomic resources and information about their demographic history is completely lacking. From all 38 European species, *Lampyris noctiluca* (the big European firefly or common glow-worm) has an especially wide distribution across Europe, inhabiting different types of ecological niches (De Cock 2009; Novak 2018). Its wide distribution makes this species particularly interesting to understand past migration patterns in Europe and present-day insect connectivity across highly fragmented landscapes. *L. noctiluca* adult females are neotenic, retaining a larval-like morphology, a characteristic that is present in several firefly species (Bocakova et al. 2007; South et al. 2011). Short after eclosion adult females start signaling by producing bioluminescent signals to attract males and die after laying eggs (Novak 2018; Van den Broeck et al. 2021). Short-distance migration can take place in both sexes during the larval stage (1 to 2

years), but in adulthood only winged males disperse (Lehtonen et al. 2021). Current distribution of *L. noctiluca* is marked by the last ice age, making northern latitudes and high-altitude locations only available for colonization after the retreat of the glaciers. Thus, range expansion in neotenic insects such as *L. noctiluca* is limited by the movement of the larvae (Lehtonen et al. 2021), is feasible for adult males and impossible for adult females. Male-biased migration can produce particular nucleotide diversity patterns, such as nuclear-mitochondria tree discordance (Hamilton et al. 2005) and can reduce nucleotide diversity levels when compared with species where gene flow is enabled by both sexes (Peart et al. 2020). Currently, we have almost no information about the impact of sex-specific neoteny on population genetic patterns (Eberle et al. 2019) or about the putative consequences that sex-specific neoteny can have in adaptive processes.

The current genetic diversity observed in a species includes the product of its past demographic and selection events. As an example, the last ice age sets boundaries on migration patterns which, together with ecological factors, have defined

Received: April 10, 2024. Revised: January 16, 2025. Accepted: February 10, 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

today's species distribution and genetic diversity (Hewitt 2000). In fruit flies and humans, the various forces shaping genetic diversity have been studied in detail, resulting in well-founded hypothesis for population changes through time, gene flow and adaptation (Stephan and Li 2007; Gutenkunst et al. 2009; Pool et al. 2012). This scenario is different for species with no obvious health or economic importance to humans, as is the case for many insects, but see (Catalán et al. 2022). Despite the importance of insects for their diverse roles in ecosystem function (Losey and Vaughan 2006), their high diversity in species number and diverse natural histories, has challenged in some way the generation of detailed knowledge of their ecology and evolutionary histories. Nevertheless, this knowledge is highly relevant to the maintenance and productivity of diverse ecosystems, particularly in the face of the current biodiversity crisis (Tihelka et al. 2021; Wagner et al. 2021). Astonishingly, most insect taxa lack the necessary genomic resources needed to generate comprehensive demographic and phylogeographic hypotheses, which are crucial for conservation purposes.

In this study, we sampled a total of 115 individuals from six populations of *L. noctiluca*, ranging from Finland to Italy and generated population-level whole-genome re-sequencing data. The goal of this study was to generate the first demographic hypothesis, including population size changes through time and migration events for *L. noctiluca* with a particular focus on the impact of female neoteny. Additionally, we assembled the first genome for *L. noctiluca*, contributing in this way to the genomic resources for fireflies and insect research. We investigated several population genetic statistics and generated hypothesis for *L. noctiluca*'s colonization patterns by taking

advantage of methods such as StairwayPlot2 (Liu and Fu 2015), PoMo (Polymorphism aware phylogenetic Model) (Borges et al. 2022) and Approximate Bayesian Computation (ABC) demographic modeling. We uncovered a complex demographic history, putatively with multiple migration routes from different glacial refugia and identified genetic footprints left by sex-biased migration. Our study contributes to generating a more complete vision of migration patterns in Europe and marks the first step into generating an integrative understanding of the forces shaping genetic diversity in fireflies.

Results

Genome Assembly

We generated long Nanopore and short Illumina reads to assemble the genome of *L. noctiluca* (*LaNoc*) (Fig. 1). From the four explored assembly strategies (MaSuRCA, Flye, Canu and Shasta) (Zimin et al. 2013; Kolmogorov et al. 2019; Shafin et al. 2020), the hybrid approach with MaSuRCA gave the best results (supplementary table S1, Supplementary Material online). *L. noctiluca*'s genome lies at the lowest quantile of known genome sizes of fireflies (Lampyridae) (Liu et al. 2017; Lower et al. 2017) (Fig. 1a) with 652,498,014 bp (Table 1). We identified the X chromosome using a female-to-male difference in genome coverage strategy, successfully detecting contigs with half coverage values in males, which is the heterogametic sex in fireflies (X0 or XY) (Wasserman and Ehrman 1986; Dias et al. 2007). The X chromosome is 19 Mbp in length, comprising ~2.9% of the genome (Fig. 1c). We identified a complete mitochondrial

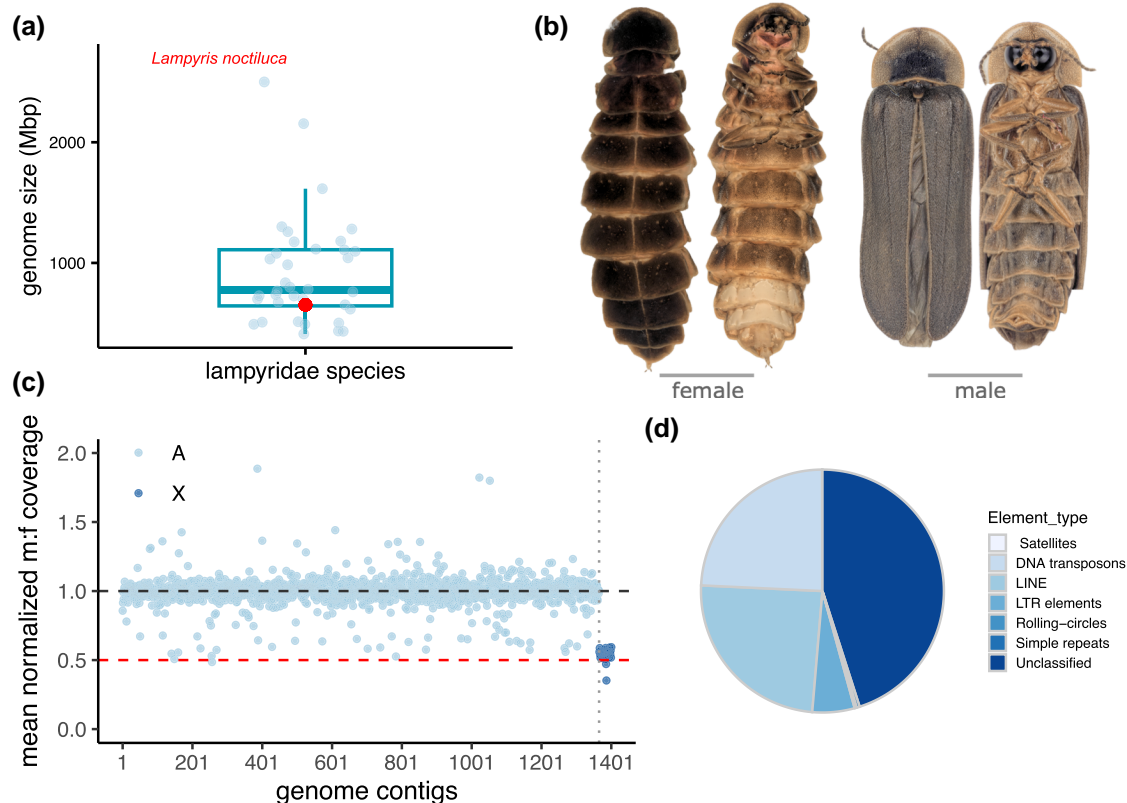


Fig. 1. Genome sequencing of *Lampyris noctiluca*. a) Boxplot showing genome size distribution of available fireflies including genome size of *Lampyris noctiluca* highlighted in red. b) Dorsal and ventral view of female (left) and male (right) images of *L. noctiluca*. c) Male-to-female ratio of normalized genome coverage across all assembled contigs. Dark blue dots, around m:f = 0.5 ratio, correspond to X chromosome linked contigs. d) Pie chart showing percentages of repetitive element types in the genome. Repeats constitute 41.76% of the genome.

Table 1 General genome assembly statistics

Statistic	Value
Genome size (bp)	652,498,014
No. of contigs	1,401
No. of IsoSeq transcripts	31,648
Largest contigs	4,636,015
Chr X length (bp)	19,305,488
Mitochondria (bp)	18,937
N50 (bp)	655,439
L50	292
CG %	35.17
Repeat content	41.76%
Genome heterozygosity	0.91%
BUSCO Insecta %	94.60

genome covering 18,937 bp. Based on PacBio IsoSeq RNA sequencing, we annotated ~31 K transcripts and delimited genomic regions into transcripts, exons, introns and intergenic. Repetitive elements comprised 41.76% of the genome, most of which are unclassified repeats (18.16%), followed by retro-elements (12.41%) and DNA transposons (10%) (Fig. 1d).

Population Structure

We sampled a total of six *L. noctiluca* populations from Finland, Germany-Kiel, Germany-Jena, Germany-Glonn, Switzerland, Italy and one outgroup specimen (*Lampyrus zenkeri*) collected in Greece, sequencing a total of 115 individuals (Fig. 2a). We identified 61,441,736 autosomal and 841,085 X chromosome single-nucleotide polymorphisms (SNPs) across all individuals. Genetic clustering approaches, such as PCA and STRUCTURE, showed well defined populations that clustered by country (Fig. 2b and c, supplementary fig. S1, Supplementary Material online). The three German populations are very closely related to each other, mostly forming one single genetic cluster, although the ADMIXTURE analysis at $k=4$ and $k=5$ revealed some level of population substructure (supplementary fig. S2, Supplementary Material online). The X chromosome population structure analysis mostly follows the patterns observed in the autosomes with the exception of $k=6$, where the Jena and Kiel German population do not segregate further into different genetic clusters as observed in the autosomes (supplementary fig. S3, Supplementary Material online).

Population differentiation (F_{st}) (Fig. 2d) and population divergence (D_{xy}) (supplementary fig. S4, Supplementary Material online) were calculated between populations, with both parameters showing strong population differentiation (Fig. 2d). F_{st} for the X chromosome show deeper population structure (supplementary fig. S5, Supplementary Material online), a result that can be explained by the smaller population size of the X chromosome. Overall, these results show strong population structure, which could be the result of limited migration, as mobilization only occurs at the larval stage and by flying adult males.

Past Population Size Changes

Changes in effective population size over time show an intrinsic characteristic of populations and provides information about past events. We used a Bayesian variant (implemented in the software RevBayes (Höhna et al. 2016) of the StairwayPlot approach (Liu and Fu 2015) to uncover single population demographic histories. All populations showed pronounced changes in population size over the last 100k

years (Fig. 3a). The impacts of these changes through time were also detected by Tajima's D calculated at autosomal and X-linked loci (Tajima 1989) (supplementary fig. S6, Supplementary Material online). Over the last 100k years, the effective population sizes varied from a maximum of 6.3 million individuals (Italy) to a minimum of 7k individuals (Switzerland and Finland). At least two deep bottleneck events were recovered for each population, however, the timing of these bottlenecks varied between populations (Fig. 3a). For the German populations, the timing of the bottlenecks overlapped, a pattern that could be driven by common environmental stressors leading to a bottleneck or through common ancestry (Stroeven et al. 2016; Seguinot et al. 2018). Towards the present, most populations, except for the Kiel population, showed a decline in effective population size. The observed decline in the last 10,000 years can be explained by the colonization of new habitats after the retreat of the last ice age resulting in a population range expansion (Excoffier et al. 2009). Alternatively, the recent population decline could also be explained by strong selective pressures.

Inference of the Time of Most Recent Common Ancestor

If the population divergence time is comparably recent and the population size comparably large, then the Most Recent Common Ancestor (MRCA) of a sample of individuals from a population is likely to predate the population divergence time. Furthermore, in such a scenario the time of the MRCA of samples from one population is like to be identical to samples from a sister population, due to deep coalescence events. Such a shared MRCA among populations leads to an excess of shared ancestral polymorphisms and incomplete lineage sorting. A shared MRCA is expected, for example, in scenarios of recent colonialization (more recent than, say, twice the effective population size * the generation time). Conversely, the MRCA will be distinct between populations if a population went through a bottleneck (forcing the MRCA to be younger) or older colonialization events. Traditional expectations for range expansions result in younger MRCA for the newly colonized regions as the assumption is that only few individuals participate in the range expansion (Excoffier et al. 2009).

To explore such a scenario, we inferred the MRCA (coalescent time) for all sampled individuals of a population, repeating the inference for each contig and each population. The inferred MRCAs for each population (except for the Swiss population) were ~800 kya for the autosomes and ~350 kya for the X chromosome (supplementary fig. S7, Supplementary Material online). Such an old time of the MRCA is caused by the large effective population size (see above). The time of the MRCA for each homologous contig was highly correlated across populations, but across contigs the time varied significantly (Fig. 3b), which provides evidence that a common ancestor was shared among populations. The younger time of the MRCA of the Swiss population, which is also reflected in its lower π values (Fig. 3c, supplementary fig. S8, Supplementary Material online), might be explained by a more recent founder event, a stronger bottleneck or smaller effective population size. The relatively old and shared common ancestor is also reflected in the high number of shared ancestral polymorphisms between the populations (supplementary fig. S9, Supplementary Material online). Overall, the shared time of the MRCA provides evidence for a recent colonialization (younger than 800 kya) and against a rapid range expansion through few individuals (Fig. 3b).

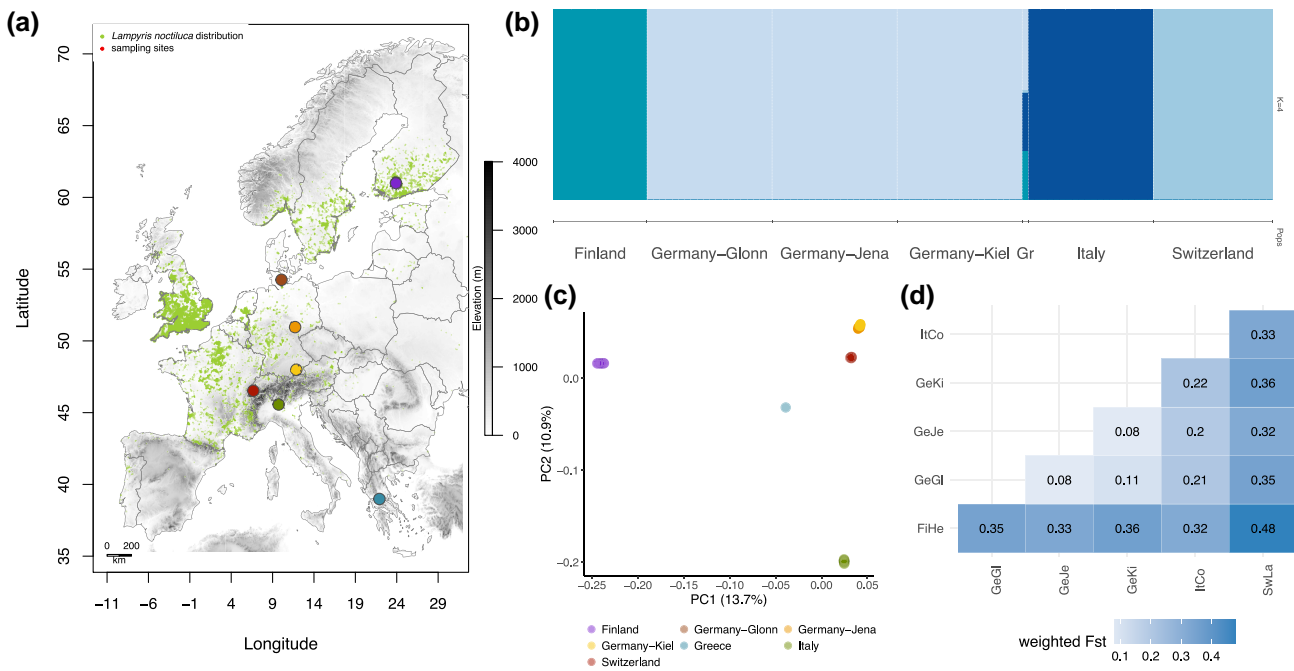


Fig. 2. Population structure in *L. noctiluca*. a) Map of Europe marking the locations of sample collection in color dots. Small green dots represent observations as reported by gbif. b) Genetic clusters assessed with STRUCTURE ($k = 4$). c) Principal component analysis (PCA). d) Pairwise population F_{st} statistic for population differentiation. FIHe: Finland, GeKi: Germany-Kiel, GeJe: Germany-Jena, GeGl: Germany-Glonn, ItCo: Italy, SwLa: Switzerland.

Population Colonization Hypothesis

We first explored a stepwise colonization process from south to north (Greece to Finland) as has been observed in humans, the fruit fly and other firefly species (Stephan and Li 2007; Excoffier et al. 2013; Catalán et al. 2022). We did not observe a declining pattern of nucleotide diversity (π) at the autosomes or X chromosome (supplementary fig. S8, Supplementary Material online), with increasing latitude (Fig. 3c) and an isolation by distance scenario was rejected by a Mantel test (P -value = 0.137, supplementary fig. S10, Supplementary Material online). The population's tree topology, inferred by a PoMo approach (De Maio et al. 2015; Borges et al. 2022), does not support a south to north colonization pattern (Fig. 4a). Instead, the PoMo tree showed a topology, where Greece (diverged < 80,000 ya) and Finland (diverged ~38,000 ya) show the deepest divergence times. Falling into the ice age period, the Swiss and the Italian populations, show divergence times of ~24,000 and ~21,000 ya, respectively. The three German populations diverged 8,000 to 10,000 yr ago, hinting a post ice age colonization. The inferred divergence times and tree topology open the question of the colonization process of these populations. In the case of the Italian population, it is plausible that it survived in a glacial refugia, as northern Italy was not covered by ice during this period (Patton et al. 2017). On the other hand, Finland and Switzerland were covered by glaciers at the estimated divergence times (Becker et al. 2016; Stroeven et al. 2016). The presence of unsampled or “ghost” populations could explain the deep divergence times retrieved for Finland and Switzerland.

Assessment of Sex-biased Migration Due to Female Neoteny

We investigated sex-biased migration by searching for mitochondrial to nuclear phylogenetic incongruity. The maternal (mitochondrial) relationships between populations revealed

one fundamental incongruence with the nuclear tree: the Swiss population being sister to the German populations (Fig. 4b), whereas in the nuclear tree the Italian population is sister to the German populations (Fig. 4a). We further investigated sex-biased migration by estimating PoMo population trees from X-linked contigs. The X-linked contigs corroborated the mitochondrial history, where the Swiss population is sister to the German populations in 57% of the contigs and the Italian population was sister to the German populations in 43% (Fig. 4c, supplementary fig. S11, Supplementary Material online). In contrast, the autosomal contigs showed a strong Italian-German sister relationship (89% of contigs). The uncovered nuclear-mito-X incongruency hints at a scenario of sex-biased migration.

To explore putative biases of introgression events on the X chromosome from Switzerland and Italy into the German populations we applied an ABBA-BABA model testing two trios: Trio 1: P1 = Italy, P2 = Germany, P3 = Switzerland, O = Greece; and trio 2: P1 = Germany, P2 = Italy, P3 = Switzerland, O = Greece. For the two tested trios we did not detect either an excess of ABBA nor of BABA, showing that when the whole chromosome is evaluated, there is no signal of introgression between the populations from Germany and Switzerland or Germany and Italy. This result is in line with our ADMIXTURE and F_{st} analysis. Outliers of the admixture fraction (f_{dM}) were identified and highlighted in red (Fig. 5). Outliers represent windows in the genome with an introgression signal between P2-P3 ($f_{dM} > 0$) or P1-P3 ($f_{dM} < 0$). From the observed outlier distribution in the two tested trios, there are no outlier enrichment along the X chromosome contigs. Both trios show outliers that support migration between the German populations and Italian population and between the German populations and the Swiss population at specific genomic regions. In the case of a single introgression event, e.g. introgression of a big chromosomal inversion (Rocha et al. 2023), a tight clustering of f_{dM} outliers can be expected. Additionally, we did window-wise PoMo

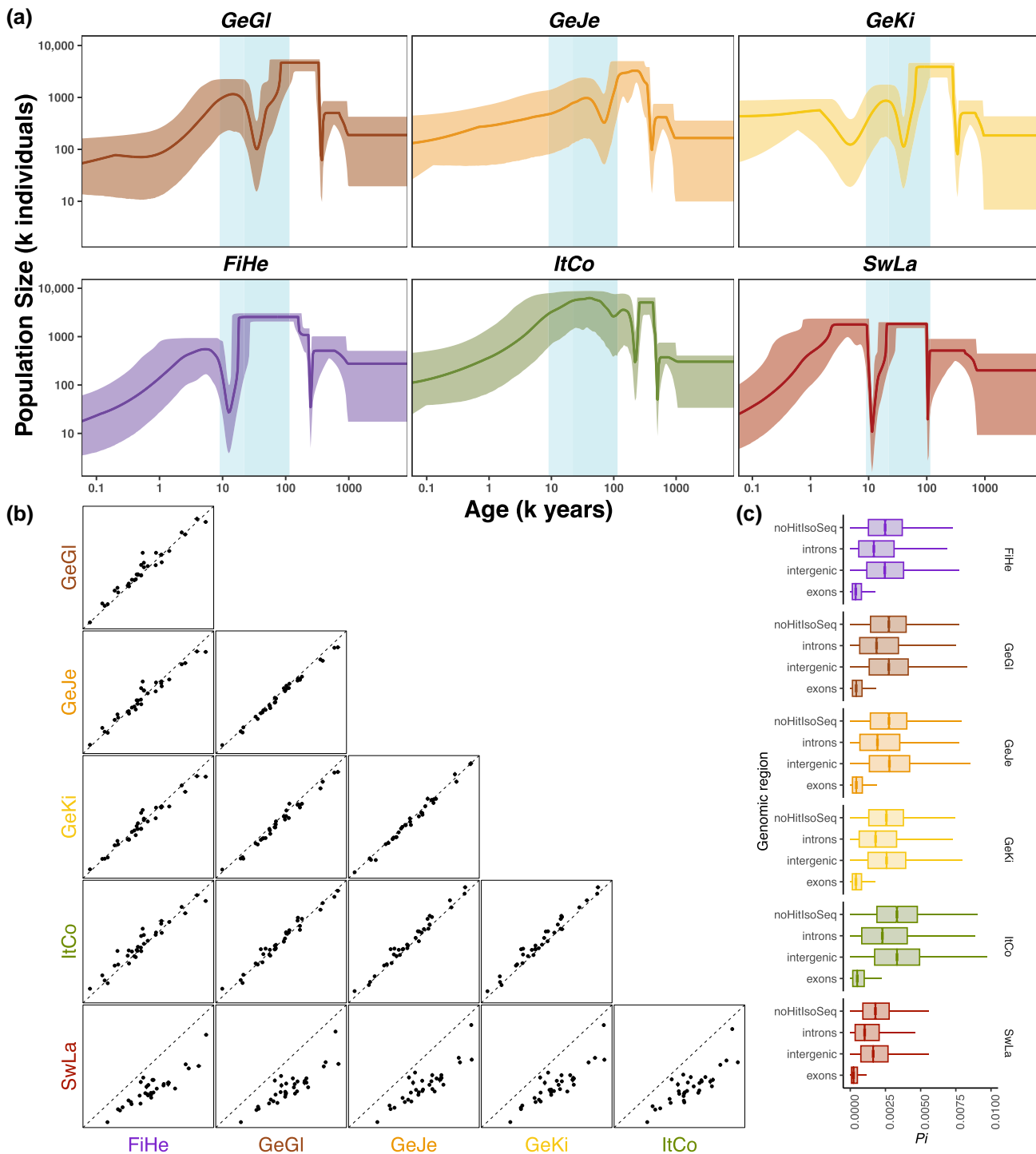


Fig. 3. Population size changes past in time. a) StairwayPlot estimation of changes in past population size. Lighter color represents 95% credible intervals. Light blue vertical rectangles mark the start and the end of the last ice age (9,000—115,000 years ago). b) Estimation of coalescent ages (most recent common ancestor, MRCA) of all sampled individuals of a population for 35 autosomal sequences. FiHe: Finland, GeKi: Germany-Kiel, GeJe: Germany-Jena, GeGI: Germany-Glonn, ItCo: Italy, SwLa: Switzerland. c) Pairwise nucleotide diversity estimates of π for the autosomes at introns, exons and intergenic regions.

analyses using windows containing 200 and 1,000 SNPs to further explore the chance of linked stretches of the X chromosome being introgressed from the Swiss or Italian population into the German populations. We did not observe continuous windows of the X chromosome sharing the same posterior probability for either the Swiss or Italian population being sister to the German populations, supporting no linkage across windows (supplementary fig. S12, Supplementary Material online).

Testing for Population Gene Flow

Our STRUCTURE/ADMIXTURE analyses showed little evidence of gene flow (Fig. 2b, supplementary fig. S1 and S2 and S3, Supplementary Material online), which is contrary to the expectation of one single widespread European species with recent, post-glacial expansion. On the other hand, strong population (sub-)structure and low gene flow matches the

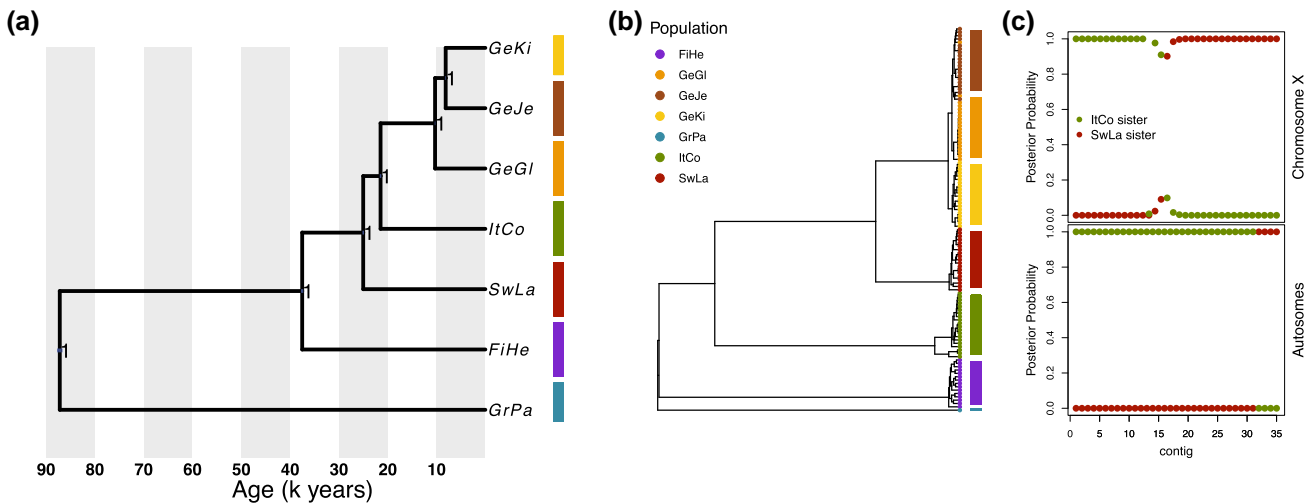


Fig. 4. PoMo phylogenetic analysis and nuclear-mitochondria-X tree discordance. a) Genetic relationship across populations estimated by a PoMo analysis. b) Maximum likelihood phylogenetic tree of the mitochondria. c) PoMo posterior probability of either Italy or Switzerland being sister to Germany. FIHe: Finland, GeKi: Germany-Kiel, GeJe: Germany-Jena, GeGI: Germany-Glonn, ItCo: Italy, SwLa: Switzerland, GrPa: Greece-Paleokastro.

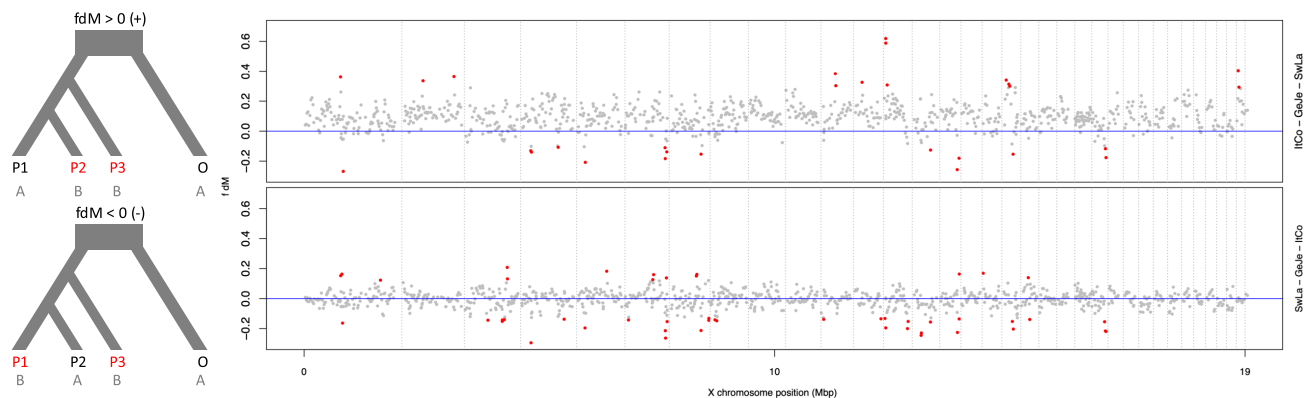


Fig. 5. Window-based ABBA-BABA analysis on the X chromosome. Two trios were tested for introgression ((P1,P2), P3), O): (1) Italy—Germany—Switzerland and (2) Switzerland—Germany—Italy. Values of $f_{dM} > 0$ indicate introgression between P2 and P3, $f_{dM} < 0$ indicates introgression between P1 and P3. Outliers were identified as $Q3 + (1.5 \times \text{Inter Quartile Range})$ and $Q1 - (1.5 \times \text{IQR})$, highlighted in red. Vertical dotted gray lines mark the lengths of the contigs comprising the X chromosome.

expectation of limited migration due to female neoteny. We computed f -branch statistics to assess the proportion of shared alleles across populations. We tested migration events assuming the most probable autosomal population tree, where the Italian population is sister to the German populations, and the most probable mitochondrial population tree, where the Swiss population is sister to the German populations. In the first scenario, we identified two putative migration events, the first between the ancestral population of Italy and Germany with the Finish population, sharing 8% of alleles, and the same ancestral population with the Swiss population, sharing 14% of shared alleles (Fig. 6a). In the second scenario, we identified multiple migration events. First, three migration events between the Finish population and ancestral population of the Jena and Kiel populations (<1% of shared alleles), the ancestral populations of all German populations (2% of shared alleles), and the ancestral population of the Swiss and German populations (7% of shared alleles). Second, a migration event between the Swiss population and ancestral population of the Jena and Kiel populations (1% of shared alleles).

Third, a migration event between the Swiss population and the Italian populations (1% to 2% of shared alleles). The f -branch statistics revealed different patterns of shared alleles in the autosomes and mitochondria.

ABC Analysis

We used an ABC framework to explore the contribution of a ghost population as a founder ancestral population and to further assess the population relationships between Germany—Italy—Switzerland. We tested the following models: *Model 1*: The Finish population as the most ancestral population from which the Swiss and Italian populations are derived independently. The Swiss population is sister to the German populations. *Model 2*: The Finish population as the most ancestral population from which the Italian population derives. The Swiss and German populations derive from the Italian population. The Italian population is sister to the German populations. *Model 3*: The Finish population as the most ancestral population from which the Swiss population derives. The Italian and German populations derive from the Swiss population. The Swiss population is sister to the German populations.

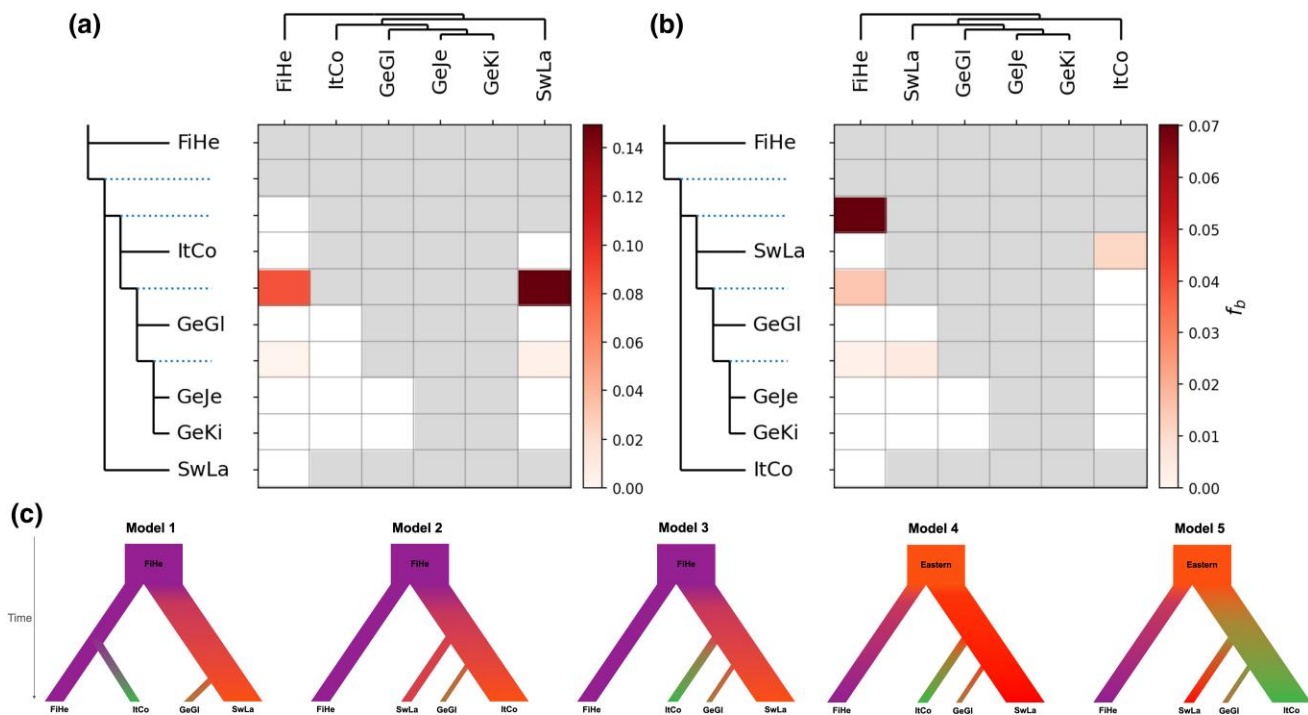


Fig. 6. f -branch statistics and ABC modelling. a) f -branch statistic investigating autosomal migration. b) f -branch statistic investigating mitochondrial migration. Gradient bar shows the fraction of shared alleles between tips and branches. c) Tested demographic scenarios presented. *Model 1*: Finland as the most ancestral population from which Switzerland and Italy derive independently. Swiss population is sister to the German populations. *Model 2*: Finnish population as the most ancestral population from which the Italian population derives. The Swiss and German populations derive from the Italian population. The Italian population is sister to the German population. *Model 3*: The Finnish population as the most ancestral population from which the Swiss population derives. The Italian and German populations derive from the Swiss population. The Swiss population is sister to the German populations. *Model 4*: A ghost population (e.g. a population from the East) represents the most ancestral population, from which the Finnish and Swiss populations derive independently. The Italian and German populations derive from the Swiss population; and the Swiss population is sister to the German populations. *Model 5*: Same as Model 4 but with the Italian population being sister to the German populations. FiHe: Finland, GeGI: Germany-Glenn, ItCo: Italy, SwLa: Switzerland.

Model 4: A ghost population (e.g. Eastern population) represents the most ancestral population, from which the Finnish and Swiss populations derive independently. The Italian and German populations derive from the Swiss population and the Swiss population is sister to the Germany populations. *Model 5*: A ghost population (e.g. Eastern population) represents the most ancestral population, from which the Finnish and Italian populations derive independently. The Swiss and German populations derive from the Italian population. The Italian population is sister to the German populations.

Using 12,572 SNPs retrieved from intergenic regions, we calculated the joint site frequency spectrum (JSFS), which showed that there is little to no fixed differences between pairs of populations (especially for populations pairs that exclude the Finnish population), and that most of the polymorphisms present in this dataset are either private or shared between each pair of populations (supplementary table S2 and S3, Supplementary Material online). Posterior probabilities for the five tested models were the following: *Model 1* (0.000), *Model 2* (0.000), *Model 3* (0.0224), *Model 4* (0.8858), and *Model 5* (0.0918). Parameter estimation on *Model 4* showed that the effective population size is around 1.448 million individuals and that the German populations split from the Swiss about 266k generations ago. Moreover, the Italian population split from a putative ghost population about 944k generations ago. On the other hand, the Swiss population split from the Italian population about 446k generations ago, and the Finnish population split more recently from the same ghost population about 614k generations ago (Table 2). ABC was not able to retrieve accurate estimates of population-specific N_e values.

We also used ABC to test ancestry contributions in the X chromosome from the Swiss and Italian populations into the German population and tested for the most probable sister group to the German populations (model tested: supplementary fig. S14, Supplementary Material online). The Swiss being sister to the German population (Model 4X, supplementary fig. S14, Supplementary Material online) showed a posterior probability of ~80%, whereas the Italian population and the German populations being sister populations (Model 5X, supplementary fig. S14, Supplementary Material online) showed a posterior probability of 20%.

Discussion

The big European firefly *L. noctiluca* is a charismatic insect species, widely distributed across Europe, where neotenic females signal their presence to males by producing light. Female neoteny limits migration and gene flow to one sex, as only winged males can disperse at the adult stage. We used a genomic approach to investigate population structure, past demographic events and gene flow in *L. noctiluca* with the main goal of understanding the impact of sex-biased migration into present nucleotide diversity patterns in insects.

Our results inferred variable effective population sizes over the last 100k years ranging from a maximum of 6.3 million individuals to a minimum of 7k individuals most likely severely affected by the last ice age (Fig. 3a). Unsurprisingly, the large effective population sizes lead to old estimates of the common ancestor of individuals from the same population (~800 kya).

Table 2 ABC parameter estimates and their respective priors

Parameter	Prior	Mode	95% quantiles
$\log_{10}(N_e)$	unif(4.5,7.5)	6.161 ($N_e = 1,448,772$ ind.)	(4.43,7.01)
T_{GeCl}	unif($0.1 * 20,000/4N_e$, $10 * 20,000/4N_e$)	$0.046 * 4N_e$ generations ago	(0,1.42)
T_{SwLa}	unif(T_{GeCl} , $10 * 25,000/4N_e$)	$0.163 * 4N_e$ generations ago	(0,3.47)
T_{ItCo}	unif(T_{SwLa} , $10 * 40,000/4N_e$)	$0.077 * 4N_e$ generations ago	(0,2.20)
T_{FiHe}	unif($0.1 * 40,000/4N_e$, $10 * 40,000/4N_e$)	$0.106 * 4N_e$ generations ago	(0,2.23)

Here, N_e stands for effective population size, T corresponds to the split time of each population.

The common ancestor of individuals from the same population was in fact the same as between population (i.e. shared common ancestor), as shown by the strong correlation of estimates per contig when compared between contigs (Fig. 3b). Thus, populations of *L. noctiluca* share ancestral polymorphisms and the divergence between populations must have occurred after (i.e. more recent) than the age of the common ancestor. Despite the presence of ancestral polymorphisms, we inferred high levels of population structure with little gene flow using several approaches (STRUCTURE, ADMIXTURE, TreeMix and f -branch statistic). Our new PoMo divergence time estimation using full-genome data and all samples per population, inferred the main population-splitting events during the last ~40k years, during the last ice age as opposed to after it. Interestingly, we uncovered nuclear-mitochondrial-X incongruency, which provides insights into sex-biased colonization patterns and gene flow. Some of these insights reveal male-only migration and female-lead population range expansions (Fig. 4b).

High Levels of Population Structure and Species Status

We uncovered high levels of population structure where populations showed strong differentiation according to country. Only the three German populations formed a single genetic cluster, showing the lowest retrieved F_{ST} values (0.08 to 0.11) (Fig. 2). The German populations presented recent divergence times, which suggest that Germany was colonized by a single event. We suggest a south to north migration pattern within the German populations is plausible, reflected by the correlated pattern of shared alleles to latitude (ADMIXTURE, k4-k5) and by the binomial distribution of the permutation Mantel test (supplementary fig. S2 and S7, Supplementary Material online). The remaining populations formed defined genetic clusters with high F_{ST} values ranging 0.21 to 0.48, suggesting low levels of gene flow. Similarly high F_{ST} values have been retrieved for between species comparisons, such as between different species of *Heliconius* butterflies (Van Belleghem et al. 2018) or between closely related fox species (L. Rocha et al. 2023). This observation raises the question of the species status across populations of *L. noctiluca* and the effect that neoteny can have in the process of speciation.

Population History and the Addition of a Ghost Population

The estimated PoMo tree (Fig. 4a) placed the Finish population as the population with the deepest divergence, followed by the Swiss, Italian and then German populations. This topology rules out an isolation by distance, stepwise colonization pattern from south to north. The deep divergence of the Finish population and our ABC analysis suggests that the founder of

this population is only distantly related to the rest of the sampled populations.

According to *L. noctiluca*'s geographical distribution, Model 4 and Model 5, which tested for a ghost population putatively coming from Eastern Europe, could correspond to a founder or genetically contributing population to the Finish population (Fig. 6b). Interestingly, the addition of an Eastern population in our ABC analysis as a proxy for a glacial refugia population resulted in the Italian population showing an older divergence time than the Finish population. The divergence time of the Swiss population was about ~25 kya, which corresponds to a period when Switzerland was still covered by a glacier (Seguinot et al. 2018), surely hindering colonization at that time. The relatively deep divergence time of the Swiss population poses the hypothesis that an unsampled Western population might be the founder of the Swiss population. Sampling *L. noctiluca* populations from Western and Eastern European regions will further shed light on the population history and dynamics of this species and reveal common faunal biogeographical patterns (Schmitt and Varga 2012; Fonseca et al. 2023).

Nevertheless, the above presented divergence times should be treated cautiously. The mutation rate is not known for this species yet; therefore, the estimated divergence times can change accordingly. Additionally, the estimated divergence times also depend on the model of population size changes used and the force of selection. The present study constitutes the first to estimate divergence times using full genomes and multiple-population data and the first to propose a comprehensive demographic species for *L. noctiluca*.

Shared Alleles Across Populations

Using an f -branch statistics approach, we were able to detect putative migration within the Italian–German internal branch and two populations: Finland and Switzerland, reaching 8% and 14% of migration fraction, respectively. The detection of migration involving an internal branch can be interpreted as migration events happening in the past or between unsampled lineages (Malinsky et al. 2018; Suvorov et al. 2022). We consistently found migration between an internal branch and Finland, as further explored with an f -branch analysis based on the mitochondrial tree (Fig. 6a) and by a TreeMix analysis (supplementary fig. S13, Supplementary Material online), further suggesting a connection with an unsampled population or past migration with Finland. Our population structure and gene flow analysis suggest high population structure between the collected populations with the presence of putative ancestral gene flow. Currently, we do not have information about the generational displacement of larvae or adults, information that would contribute to the comprehension of the migration biology of *L. noctiluca*.

Sampling of additional populations will contribute to have more information on migration breadth of this species.

Sex-biased Migration

We uncovered a nuclear-mitochondrial-X chromosome tree discordance, where the sister population to the German populations differs according to the genomic region tested. The mitochondrial tree shows a scenario where Germany was initially colonized by the ancestral population of the Swiss population, possibly of Western European origin. In such a scenario, female larva set the limit for a range expansion. The autosomal signal on the other hand, suggests a scenario where the Italian population is closer related to the German populations, a signal which we hypothesize might be driven by male-biased migration between Italy to Germany. The X chromosome nicely captures both, the autosomal and mitochondrial signals when tested different methods (PoMo, ABC, and ABBA-BABA), further supporting a male-biased migration scenario. The divergence time of X-linked contigs supporting a Swiss-German split is older than that of X-linked contigs supporting the Italian-German split ([supplementary fig. S11, Supplementary Material](#) online), further supporting a putative German colonization by Western European population. The detection of nuclear-mitochondrial-X chromosome incongruence shows the putative effect that neoteny can have on colonization and migration patterns.

Conclusions

Female neoteny, which limits dispersal capability, can play an important role in a species' range expansion and migration. In this study, we studied the demographic history and sex-biased migration of the big European firefly *L. noctiluca*. We generated a high-quality genome assembly, which includes PacBio IsoSeq gene annotation, transposable elements characterization and the identification of the X chromosome. Population sampling of 115 individuals from Italy to Finland produced the first results on population structure, nucleotide diversity levels, and demographic inference of *L. noctiluca*. We found very strong population structure with very low migration between populations. We applied a novel approach for full-genome population data, PoMos, to estimate divergence times, population relationships and discordance among genomic regions. Our results demonstrate that *L. noctiluca*'s range expansion is followed by migration from males, the sex with higher dispersal capability (introgression). The migration signal was not detected by traditional approaches and sex-biased migration should be considered for future model and hypothesis development. Finally, the population colonization of *L. noctiluca* is more complex than expected, where we hypothesize that unsampled populations from western, eastern and southern Europe have a big potential of further explaining the colonization history of this species.

Methods

Sample Collection, DNA Extraction, and Sequencing

We sampled specimens in seven locations in Europe. Males were collected using a funnel light trap with a yellow led lamp (2.0 to 2.2 V). Light traps were put in the ground shortly after sunset and left on for two hours. Females were collected by walking along transects. Collected individuals were stored in 96% ethanol at -4°C . For each population 15 to 20

individuals were collected. A Greek *Lampyris zenkeri* individual served as an outgroup.

High molecular DNA was extracted from one single male collected in Lausanne ([supplementary table S5, Supplementary Material](#) online) using the kit MagAttract HMW DNA kit (Qiagen) following manufacturer's guidelines. DNA fragment sizes and integrity were checked with a 1% agarose gel and a Femto Pulse system (Agilent). Long DNA fragments were sequenced from a single male individual using Nanopore PromethION in one Flongle flow cell and run for 72 h. Nanopore sequencing was done by the SciLifeLab in Uppsala, Sweden.

Short molecular DNA was extracted using the Monarch Genomic DNA Purification kit (New England BioLabs). DNA quality and integrity were assessed with a Nanodrop and an Agilent 5400. Illumina 150 bp paired-end reads were generated for each sample, aiming at a $15\times$ coverage, with the exception of the genome's individual from which $60\times$ sequencing depth was generated. Library prep and sequencing was outsourced to Novogene, China.

De Novo Genome Assembly

Base calling for Nanopore reads was done with Guppy (4.0.11). A hybrid genome assembly was performed with MaSuRCA v4.0.5 ([Zimin et al. 2013](#)) using 15Gb of Nanopore reads and $60\times$ depth 150 bp paired-end Illumina reads from the same individual, followed by two rounds of haplotype purging with Purge_dups v1.2.5 ([Guan et al. 2020](#)). Sequences not belonging to the class Insecta were identified using Blobtools v1.1.1 ([Laetsch and Blaxter 2017](#)) and removed from the assembly. No genome polishing was performed for *L. noctiluca*, as further polishing lead to worse BUSCO scores. Genome size and genome heterozygosity levels were estimated with GenomeScope ([Vurture et al. 2017](#)). Genome statistics were calculated with Quast v5.0.2 ([Mikheenko et al. 2018](#)) and genome completeness was assessed with BUSCO v5.2.2 ([Simão et al. 2015](#)) using the dataset Insecta. Repeats were annotated using RepeatModeler v1.0.11 and the generated custom made repeat library was used for genome masking with RepeatMasker v4.1.2 ([Smit et al. 2015](#)).

Identification of the X Chromosome

To identify the putative contigs belonging to the X chromosomes we compared male-to-female (m:f) coverage ratio across contigs. *L. noctiluca* males are expected to be the heterogametic sex, thus we expect a m:f coverage ratio on the X to lie near 0.5. We used Illumina reads from 2 samples of each sex. Read quality control was done with FastQC v0.11.9 ([Andrews 2010](#)) and trimming of adaptors and tails was done with Cutadapt v3.4 ([Martin 2011](#)) using a threshold of Phred < 20 . The curated reads were mapped to the hard masked genome, Repeat Masker, v4.1.2 ([Smit et al. 2015](#)) with BWA v0.7.17 ([Li and Durbin 2009](#)). Duplicate reads were removed from the bam files with Picard (V 2.20.8). Coverage was calculated with Deeptools v3.5.0 ([Ramírez et al. 2016](#)), for 10Kb windows across the genome and normalized using RPKM (Reads Per Kilobase Million). Each 10 kb window coverage level was normalized by dividing it by the mean coverage value of the 5 five largest autosomal contigs. These five contigs were manually selected by choosing the five largest contigs with a male-to-female coverage ratio of 1 ± 0.1 . Contigs smaller than 30 kb were filtered out leaving only

contigs with at least 3 data points (i.e. three 10 kb windows). We then performed a non-parametric Wilcoxon Rank Sum Test to test for significant differences in contig coverage values between sexes and applied a Bonferroni multiple test correction. Male-to-female coverage ratios were calculated only from contigs with significant differences in coverage between sexes. Contigs with a m:f ratio $0.4 \leq x \leq 0.6$ were considered to belong to the X chromosome.

Identification of Coding Sequences Using PacBio IsoSeq Transcriptome Data

RNA was extracted from heads and thorax + abdomen of one female and one male using the Monarch Total RNA Miniprep Kit (New England, BioLabs). RNA was purified by ethanol precipitation and equal concentrations of head and thorax + abdomen RNA was pooled for sequencing, separately for each sex. Equal RNA concentration of each body type ensures equal probability of transcript sequencing. IsoSeq libraries and sequencing were done by Novogene where 69 and 91 sub-read bases (Gb) were produced for the male and female sample, respectively. Primer removal, multiplexing of raw reads and clustering was done with IsoSeqv3 v3.8.0 (<https://github.com/PacificBiosciences/IsoSeq>). In order to have a single non-redundant transcript set for the species, transcript collapsing was performed with Cupcake v29.0.0 (https://github.com/Magdoll/cDNA_Cupcake) and BUSCO in transcriptome mode was run to assess transcriptome completeness. Curated transcripts were mapped back to the genome with minimap2 v2.14 (Li 2018) and a gtf file was produced denoting intergenic, exons and intronic regions (Appendix 1).

Processing of Population Level Whole-Genome re-sequencing

Illumina short read sequences were trimmed with TrimmGalore! V0.6.6 (Krueger 2012) and FastQC v0.11.9 was used to filter out bases with a phred score < 20 and reads shorter than 20 bp (Andrews 2010). Reads were mapped back to the genome with BWA v0.7.17 (Li and Durbin 2009). Mapped files in BAM format were curated by removing PCR duplicates with Picard v2.20.8, and low-quality reads (Q20) were discarded using SAMtools v1.10 (Li et al. 2009).

GATK v4.1.9 (Auwera et al. 2013) was used to call SNPs and indels via local re-assembly of haplotypes with HaplotypeCaller. Joint genotyping of all sequenced samples was done with GenotypeGVCFs. VCFs statistics were drawn with bcftools stats and gatk VariantsToTable. Quality scores thresholds were applied for minimum and maximum read depth [20,1568], fisher strand [FS=10], strand bias [SOR=3], root mean square mapping quality [MQ=40] and nucleotide quality by depth [DP=2]. Only variants with a QUAL > 30 were kept, as well as only SNPs (indels were removed) and biallelic sites. A SNP missingness of 0.25 across all samples was allowed. Sites in the VCF file which overlapped with repetitive elements were excluded from the analysis. An additional set of VCF files which included monomorphic sites was generated, where the GATK tag `-select-type-to-include NO_VARIATION` was used.

Population Genetic and Structure Analysis

Population genetic analyses were done separately on the autosomes and the X chromosome. Nucleotide polymorphism diversity (depicted by π per site) and Tajima's D were

estimated with VCFtools v0.1.14 (Danecek et al. 2011), separately for exons, introns and intergenic regions on sliding windows of 10,000 base pairs.

Population structure was first explored via Principal Component Analysis (PCA) in Plink v1.09 (Chang et al. 2015), filtering out linked sites with $r^2 > 0.2$. ADMIXTURE (Alexander et al. 2009) and fineSTRUCTURE (Lawson et al. 2012) were run using the unlinked SNP set, for K1-K6. Population differentiation was calculated as F_{st} (Weir and Cockerham 1984) across all population pairs with VCFtools in windows of 10,000 bp. Genetic distances (D_{xy}) (Wakeley 2016) were calculated for every pair of populations using pixy (Korunes and Samuk 2021).

Population Size Estimation

We estimated population size trajectories using the StairwayPlot approach (Liu and Fu 2020) within a Bayesian statistical framework as implemented in RevBayes (Höhna et al. 2016). We used SNPs from the autosomes as data, comprising a total of 56 to 59 million SNPs per population. We assumed a total genome size of 411.5Mb for the computation of monomorphic sites. This number of monomorphic sites is lower than the total genome size as we estimated that 30% to 40% of the variable sites were filtered out, thus we reduced the corresponding genome size accordingly. The number of variable sites is informative about the actual effective population size but not about the population size changes over time. We explored the impact of data filtering and different genomic regions by performing the StairwayPlot analysis for all SNPs (main results), and only intergenic regions, exons or introns.

For each population we computed the site frequency spectrum in RevBayes directly from the VCF file. The complete site frequency spectrum had between 30 (for Finland) to 40 (Italy and Germany) categories. We used the folded site frequency spectrum because the sites were not polarized.

We assumed a mutation rate of $2.8E-9$ per site per year and a generation time of one year (Keightley et al. 2015). Our prior model for the population size trajectory assumed autocorrelated, log-normal distributed displacements with an expectation of one order of magnitude over the total timespan. We explored several additional prior models, including uncorrelated models where population sizes per interval are drawn from a log-normal prior distribution. We ran a Markov chain Monte Carlo simulation for 250,000 iterations with sampling every 10 iterations. We applied the same settings for all populations. We plotted the resulting population size trajectories using the R package *RevGadgets* (Tribble et al. 2022).

Coalescent Estimations of the Last Common Ancestor

We estimated the time of the MRCA of all sampled individuals within a population using a genealogy-based coalescent approach as implemented in RevBayes (Billenstein and Höhna 2024). We constructed multiple sequence alignments including invariant sites per contig for each population. We inferred the genealogy based on the alignments using the following model assumptions. We assumed a standard phylogenetic GTR + I nucleotide mutation model. We assumed a coalescent process prior on the genealogy and a “known” mutation rate of $2.8E-9$ per site per year. We ran four replicated Markov chain Monte Carlo analyses for 50,000 iterations each with 197 moves per iteration. We checked for convergence using

the R package *convenience* (Fabreti and Höhna 2022). From the posterior sample of genealogies, we extracted the root age to represent the time of the MRCA. We performed this genealogy-based coalescent analysis for 33 contigs from the autosomes and 34 contigs from the X chromosome.

Demographic Inference With PoMos

We estimated the population relationship and divergence times between populations using PoMos implemented in RevBayes (Borges et al. 2022). PoMo models use as data allele counts per population and therefore can handle multiple individuals per population efficiently. Changes in allele frequencies are modeled using a Moran process combined with a boundary mutation process. Instead of a 4-state nucleotide process we converted the data into binary states. We used a total of 61,441,738 SNPs from the autosomes and implemented an ascertainment bias correction for not using monomorphic sites. We assumed a “known” mutation rate of 2.8E-9 and an average effective population size of 100k. In further analyses we explored the impact of the a priori assumed average effective population size. We assumed a uniform prior distribution on both topology and divergence times. We ran two replicated Markov chain Monte Carlo analyses for 50,000 iterations each. We computed the *maximum a posteriori* topology and mean divergence times from the posterior samples. We checked for convergence using the R package *convenience* (Fabreti and Höhna 2022). We plotted the population tree using the R package *RevGadgets* (Tribble et al. 2022).

Demographic Inference With ABC

Data Collection

Data consists of single nucleotide polymorphisms (SNPs) obtained from 10 randomly chosen intergenic regions coming from four populations of *L. noctiluca*. The populations are Helsinki-Finland (FiHe), Lausanne-Switzerland (SwLa), Italy (ItCo), and Germany (GeGl). A total of 12,572 SNPs coming from the intergenic regions were kept for downstream analyses (supplementary table S5, Supplementary Material online). The “FiHe” population consisted of $n = 15$ individuals, the “SwLa” population with $n = 20$ individuals, the “ItCo” population with $n = 20$ individuals, and the “GeGl” with $n = 19$ individuals, yielding a total of 74 sampled individuals. Recombination rates for each of the intergenic regions were estimated using the software *LDhat* (Auton and McVean 2007).

Observed Summary Statistics

We calculated a total of 370 summary statistics, including: number of segregating sites S , Watterson’s θ_W (Watterson 1975), π , Tajima’s D (Tajima 1989), linkage disequilibrium Z_{ns} (Kelly 1997), the folded SFS, Weir-Cockerham’s F_{st} (Weir and Cockerham 1984), distance of Nei (Nei and Li 1979) and the Wakeley-Hey “W” summaries of the joint SFS. All the above-mentioned statistics are unaffected by the polarization (or lack thereof) of the observed SNPs.

Demographic Models

We tested five different demographic scenarios: scenario (1) FiHe and SwLa split from an ancestral Finish population, then ItCo splits directly from FiHe, and finally GeGl splits from SwLa; scenario (2) FiHe and ItCo split from an ancestral Finish population, then SwLa splits from ItCo, and finally GeGl splits from ItCo; scenario (3) same as scenario (2), but

GeGl splits from SwLa; scenario (4) Same as scenario (3), but both FiHe and ItCo split from a putative “Eastern” population; scenario (5) Same as scenario (4), but with GeGl and ItCo being sister populations (Fig. 6c). With these different demographic scenarios, we covered biologically plausible population histories of European *L. noctiluca*.

ABC Simulations

We performed simulations with the program *msms* (Ewing and Hermisson 2010). For each of the five demographic models described above (Fig. 1) we simulated segregating sites for 74 individuals (15, 20, 20, and 19 individuals representing the “FiHe”, “SwLa”, “ItCo” and “GeGl” populations, respectively). From the simulated sites, we calculated all summary statistics described above. All priors are shown in Table 2. We repeated this whole simulation process 20,000 times.

Model Choice and Parameter Estimation

With all 20,000 simulations per model we calculated the posterior probabilities of each of the five demographic scenarios using the R package *abc* (Csilléry et al. 2012). Model choice was based on the following summary statistics per population: θ_W , π , Tajima’s D , Z_{ns} , W statistics, distance of Nei, and F_{st} . Parameter estimation on the best model was accomplished by using both the rejection (Tavare et al. 1997; Pritchard et al. 1999) and regression (Beaumont et al. 2002) algorithms using the same R package *abc*. To reduce dimensionality while keeping the maximum amount of information still available we used partial least squares (pls) in the context of ABC (Wegmann et al. 2009).

ABC Analysis for the X Chromosome

A similar ABC analysis was performed with X chromosome data. A total of 13,112 SNPs were recovered from 10 randomly chosen X-linked intergenic regions. Models 4 and 5, renamed as Models 4X and 5X (supplementary fig. S14, Supplementary Material online) were simulated with this new dataset. Model choice followed the same procedure as above.

Mitochondrial Genome Tree Inference

We constructed a multiple sequence alignment for the mitochondrial genome comprising of 115 sequences of 18,937 bp each. We estimated the mitochondrial relationship using a standard phylogenetic approach as implemented in RevBayes (Höhna et al. 2016). We assumed a GTR + GI mutation model with 4 rate categories (slow to faster evolving sites). We assumed a mitochondrial mutation rate of 1.34E-8 per site per year (Pons et al. 2010). We assumed a uniform prior distribution on the root age, the topology and node ages between lineages. We ran two replicated Markov chain Monte Carlo simulations for 50,000 iterations with 614.6 moves per iteration, sampling every 10th iteration. We checked for convergence using the R package *convenience* (Fabreti and Höhna 2022). We plotted the mitochondrial tree using the R package *RevGadgets* (Tribble et al. 2022).

Analysis of Gene Flow

We performed several types of analyses to infer gene flow and migration between the sampled *L. noctiluca* populations. Our primary analysis consisted of a PoMo (Borges et al. 2022) analysis, which we performed separately for 35 autosomal and 34

X-linked contigs. These contigs were chosen based on length. For each contig, we computed for each SNP the allele frequency per population and used the allele frequencies as data. We used an ascertainment bias correction to condition on only variable sites. We ran two replicated Markov chain Monte Carlo analyses for 50,000 iterations. We computed the posterior probabilities of the Italian and the Swiss population being sister to the German populations.

Additionally, we explored putative migration events between the populations using the *f*₄-branch statistic as implemented in Dsuite (Malinsky et al. 2021). As input data, we used the previously curated VCF file including monomorphic sites. To disentangle correlated *f*₄-ratio results and to assign evidence of gene flow across the phylogenetic tree we used the *f*-branch metric implemented by the same authors. The most probable phylogenetic tree was used for the analysis, where Italy is sister group to Germany. A Treemix (Pickrell and Pritchard 2012) analysis was done for autosomal unlinked SNPs using the following command: `treemix -i $TreeInput -m $i -o lanoc_M_.$i -root GrPa -bootstrap -k 1000`. Treemix was run to investigate up to six migration events (six edges) using the Greek individual to root the population tree and allowing for 1,000 replications.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

We would like to thank Gabriele Kumpfmüller for her magnificent technical assistance in the lab. Andreas Tiraboschi (Gruppo Ecologico Colognese) for sharing his knowledge with us on Italian fireflies. Klaus Reinhold for guiding and showing us where to find fireflies in Greece. To Shanaka Thisara for his fine imaging skills of *L. noctiluca* specimens. This study was funded by the Société Vaudoise des Sciences Naturelles (SVSN) and the Société Académique Vaudoise (SAV) to PD and by the DFG SPP-1991 to SH and AC.

Data Availability

The SRA submission of the short read sequences can be found in [supplementary table S6, Supplementary Material online](#). The *Lampyrus noctiluca* genomes will be realized with the article on NCBI PRJNA1238231.

References

- Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19(9):1655–1664. <https://doi.org/10.1101/gr.094052.109>.
- Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Auton A, McVean G. Recombination rate estimation in the presence of hotspots. *Genome Res.* 2007;17(8):1219–1227. <https://doi.org/10.1101/gr.6386707>.
- Auweru GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinforma.* 2013;43:11.10.1–11.10.33. <https://doi.org/10.1002/0471250953.bi1110s43>.
- Beaumont MA, Zhang W, Balding DJ. Approximate Bayesian computation in population genetics. *Genetics.* 2002;162(4):2015–2035. <https://doi.org/10.1093/genetics/162.4.2025>.
- Becker P, Seguinot J, Jouvett G, Funk M. Last glacial maximum precipitation pattern in the Alps inferred from glacier modelling. *Geogr Helv.* 2016;71(3):173–187. <https://doi.org/10.5194/gf-71-173-2016>.
- Billenstein RJ, Höhna S. Comparison of Bayesian coalescent skyline plot models for inferring demographic histories. *Mol Biol Evol.* 2024;41(5):1–14. <https://doi.org/10.1093/molbev/msae073>.
- Bocakova M, Bocak L, Hunt T, Teraväinen M, Vogler AP. Molecular phylogenetics of elateriformia (Coleoptera): evolution of bioluminescence and neoteny. *Cladistics.* 2007;23(5):477–496. <https://doi.org/10.1111/j.1096-0031.2007.00164.x>.
- Borges R, Boussau B, Höhna S, Pereira RJ, Kosiol C. Polymorphism-aware estimation of species trees and evolutionary forces from genomic sequences with RevBayes. *Methods Ecol Evol.* 2022;2022(11):2339–2346. <https://doi.org/10.1111/2041-210X.13980>.
- Catalán A, Höhna S, Lower SE, Duchon P. Inferring the demographic history of the North American firefly *Photinus pyralis*. *J Evol Biol.* 2022;35(11):1488–1499. <https://doi.org/10.1111/jeb.14094>.
- Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4(1):1–16. <https://doi.org/10.1186/s13742-015-0047-8>.
- Csilléry K, François O, Blum MGB. Abc: an R package for approximate Bayesian computation (ABC). *Methods Ecol Evol.* 2012;3(3):475–479. <https://doi.org/10.1111/j.2041-210X.2011.00179.x>.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27(15):2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.
- De Cock R. Biology and behaviour of European lampyrids. *Res Signpost.* 2009;2:161–200. https://www.researchgate.net/publication/285849945_Biology_and_behaviour_of_European_lampyrids.
- De Maio N, Schrempf D, Kosiol C. Pomo: an allele frequency-based approach for species tree estimation. *Syst Biol.* 2015;64(6):1018–1031. <https://doi.org/10.1093/sysbio/syv048>.
- Dias CM, Schneider MC, Rosa SP, Costa C, Cella DM. The first cytogenetic report of fireflies (Coleoptera, Lampyridae) from Brazilian fauna. *Acta Zool.* 2007;88(4):309–316. <https://doi.org/10.1111/j.1463-6395.2007.00283.x>.
- Eberle J, Bazzato E, Fabrizi S, Rossini M, Colomba M, Cillo D, Uliana M, Sparacio I, Sabatinelli G, Warnock RCM, et al. Sex-biased dispersal obscures species boundaries in integrative species delimitation approaches. *Syst Biol.* 2019;68(3):441–459. <https://doi.org/10.1093/sysbio/syy072>.
- Ewing G, Hermisson J. MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics.* 2010;26(16):2064–2065. <https://doi.org/10.1093/bioinformatics/btq322>.
- Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. Robust demographic inference from genomic and SNP. *PLoS Genet.* 2013;9(10):e1003905. <https://doi.org/10.1371/journal.pgen.1003905>.
- Excoffier L, Foll M, Petit RJ. Genetic consequences of range expansions. *Annu Rev Ecol Syst.* 2009;40(1):481–501. <https://doi.org/10.1146/annurev.ecolsys.39.110707.173414>.
- Fabreti LG, Höhna S. Convergence assessment for Bayesian phylogenetic analysis using MCMC simulation. *Methods Ecol Evol.* 2022;13(1):77–90. <https://doi.org/10.1111/2041-210X.13727>.
- Fonseca EM, Pelletier TA, Decker SK, Parsons DJ, Carstens BC. Pleistocene glaciations caused the latitudinal gradient of within-species genetic diversity. *Evol Lett.* 2023;7(5):331–338. <https://doi.org/10.1093/evlett/qrado30>.
- Guan D, Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics.* 2020;36(9):2896–2898. <https://doi.org/10.1093/bioinformatics/btaa025>.

- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 2009; 5(10):e1000695. <https://doi.org/10.1371/journal.pgen.1000695>.
- Hamilton G, Currat M, Ray N, Heckel G, Beaumont M, Excoffier L. Bayesian estimation of recent migration rates after a spatial expansion. *Genetics.* 2005;170(1):409–417. <https://doi.org/10.1534/genetics.104.034199>.
- Hewitt G. The genetic legacy of the quaternary ice ages. *Nature.* 2000;405(6789):907–913. <https://doi.org/10.1038/35016000>.
- Höhna S, Landis MJ, Heath TA, Boussau B, Lartillot N, Moore BR, Huelsenbeck JP, Ronquist F. RevBayes: bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst Biol.* 2016;65(4):726–736. <https://doi.org/10.1093/sysbio/syw021>.
- Keightley PD, Pinharanda A, Ness RW, Simpson F, Dasmahapatra KK, Mallet J, Davey JW, Jiggins CD. Estimation of the spontaneous mutation rate in *Heliconius melpomene*. *Mol Biol Evol.* 2015;32(1):239–243. <https://doi.org/10.1093/molbev/msu302>.
- Kelly JK. A test of neutrality based on interlocus associations. *Genetics.* 1997;146(3):1197–1206. <https://doi.org/10.1093/genetics/146.3.1197>.
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol.* 2019;37(5):540–546. <https://doi.org/10.1038/s41587-019-0072-8>.
- Korunes KL, Samuk K. Pixy: unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol Ecol Resour.* 2021;21(4):1359–1368. <https://doi.org/10.1111/1755-0998.13326>.
- Krueger F. 2012. Trim Galore! : A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. Available at https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
- Laetsch DR, Blaxter ML. BlobTools: interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. *F1000Res.* 2017;6:1287. <https://doi.org/10.12688/f1000research.12232.1>.
- Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of population structure using dense haplotype data. *PLoS Genet.* 2012;8(1):11–17. <https://doi.org/10.1371/journal.pgen.1002453>.
- Lehtonen TK, Babic NL, Piepponen T, Valkeinen O, Borshagovski AM, Kaitala A. High road mortality during female-biased larval dispersal in an iconic beetle. *Behav Ecol Sociobiol.* 2021;75(1):26. <https://doi.org/10.1007/s00265-020-02962-6>.
- Lewis SM, Wong CH, Owens ACS, Fallon C, Jepsen S, Thancharoen A, Wu C, De Cock R, Novák M, López-Palafox T, et al. A global perspective on firefly extinction threats. *Bioscience.* 2020;70(2):157–167. <https://doi.org/10.1093/biosci/biz157>.
- Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018;34(18):3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Liu GC, Dong ZW, He JW, Zhao RP, Wang W, Li XY. Genome size of 14 species of fireflies (Insecta, Coleoptera, Lampyridae). *Zool Res.* 2017;38:449–458. <https://doi.org/10.24272/j.issn.2095-8137.2017.078>.
- Liu X, Fu YX. Exploring population size changes using SNP frequency spectra. *Nat Genet.* 2015;47(5):555–559. <https://doi.org/10.1038/ng.3254>.
- Losey JE, Vaughan M. The economic value of ecological services provided by insects. *Bioscience.* 2006;56(4):311–323. [https://doi.org/10.1641/0006-3568\(2006\)56\[311:TEVOES\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2006)56[311:TEVOES]2.0.CO;2).
- Lower SS, Johnston JS, Stanger-Hall KF, Hjelmen CE, Hanrahan SJ, Korunes K, Hall D. Genome size in north American fireflies: substantial variation likely driven by neutral processes. *Genome Biol Evol.* 2017;9(6):1499–1512. <https://doi.org/10.1093/gbe/evx097>.
- Malinsky M, Matschiner M, Svardal H. Dsuite—fast D-statistics and related admixture evidence from VCF files. *Mol Ecol Resour.* 2021;21(2):584–595. <https://doi.org/10.1111/1755-0998.13265>.
- Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, Durbin R. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat Ecol Evol.* 2018;2(12):1940–1955. <https://doi.org/10.1038/s41559-018-0717-x>.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011. <https://journal.embnet.org/index.php/embnetjournal/article/view/200/479>.
- Mikheenko A, Pribelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics.* 2018;34(13):i142–i150. <https://doi.org/10.1093/bioinformatics/bty266>.
- Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A.* 1979;76(10):5269–5273. <https://doi.org/10.1073/pnas.76.10.5269>.
- Novak M. Redescription of immature stages of central European fireflies, Part 1: *Lampyrus noctiluca* (Linnaeus, 1758) larva, pupa and notes on its biology (Coleoptera: Lampyridae: Lampyrinae). *Zootaxa.* 2018;4378(4):516. <https://doi.org/10.11646/zootaxa.4378.4.4>.
- Patton H, Hubbard A, Andreassen K, Auric A, Whitehouse PL, Stroeven AP, Shackleton C, Winsborrow M, Heyman J, Hall AM. Deglaciation of the Eurasian ice sheet complex. *Quat Sci Rev.* 2017;169:148–172. <https://doi.org/10.1016/j.quascirev.2017.05.019>.
- Peart CR, Tusso S, Pophaly SD, Botero-Castro F, Wu CC, Auriolos-Gamboa D, Baird AB, Bickham JW, Forcada J, Galimberti F, et al. Determinants of genetic variation across eco-evolutionary scales in pinnipeds. *Nat Ecol Evol.* 2020;4(8):1095–1104. <https://doi.org/10.1038/s41559-020-1215-5>.
- Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 2012;8(11):e1002967. <https://doi.org/10.1371/journal.pgen.1002967>.
- Pons J, Ribera I, Bertranpetit J, Balke M. Nucleotide substitution rates for the full set of mitochondrial protein-coding genes in Coleoptera. *Mol Phylogenet Evol.* 2010;56(2):796–807. <https://doi.org/10.1016/j.ympev.2010.02.007>.
- Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM, Crepeau MW, Duchon P, Emerson JJ, Saelao P, Begun DJ, et al. Population Genomics of sub-saharan *Drosophila melanogaster*: African diversity and non-African admixture. *PLoS Genet.* 2012; 8(12):e1003080. <https://doi.org/10.1371/journal.pgen.1003080>.
- Pritchard JK, Seielstad MT, Feldman MW. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol Biol Evol.* 1999;16(12):1791–1798. <https://doi.org/10.1093/oxfordjournals.molbev.a026091>.
- Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 2016;44(W1):W160–W165. <https://doi.org/10.1093/nar/gkw257>.
- Rocha LJ, Silva P, Santos N, Nakamura M, Afonso S, Qinba A, Boratynski Z, Sudmant PH, Brito JC, Nielsen R, et al. North African fox genomes show signatures of repeated introgression and adaptation to life in deserts. *Nat Ecol Evol.* 2023;7(8):1267–1286. <https://doi.org/10.1038/s41559-023-02094-w>.
- Schmitt T, Varga Z. Extra-Mediterranean refugia: the rule and not the exception? *Front Zool.* 2012;9(1):1–12. <https://doi.org/10.1186/1742-9994-9-22>.
- Seguinot J, Ivy-Ochs S, Jouvét G, Huss M, Funk M, Preusser F. Modelling last glacial cycle ice dynamics in the Alps. *Cryosphere.* 2018;12(10):3265–3285. <https://doi.org/10.5194/tc-12-3265-2018>.
- Shafin K, Pesout T, Lorig-Roach R, Haukness M, Olsen HE, Bosworth C, Armstrong J, Tigyi K, Maurer N, Koren S, et al. Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. *Nat Biotechnol.* 2020;38(9):1044–1053. <https://doi.org/10.1038/s41587-020-0503-6>.

- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
- Smit A, Hubley R, Green P. RepeatMasker. 2015. <http://www.repeatmasker.org>.
- South A, Stanger-Hall K, Jeng ML, Lewis SM. Correlated evolution of female neoteny and flightlessness with male spermatophore production in fireflies (coleoptera: Lampyridae). *Evolution (N Y)*. 2011;65:1099–1113. <https://doi.org/10.1111/j.1558-5646.2010.01199.x>.
- Stephan W, Li H. The recent demographic and adaptive history of *Drosophila melanogaster*. *Heredity (Edinb)*. 2007;98(2):65–68. <https://doi.org/10.1038/sj.hdy.6800901>.
- Stroeve AP, Hättestrand C, Kleman J, Heyman J, Fabel D, Fredin O, Goodfellow BW, Harbor JM, Jansen JD, Olsen L, et al. Deglaciation of fennoscandia. *Quat Sci Rev*. 2016;147:91–121. <https://doi.org/10.1016/j.quascirev.2015.09.016>.
- Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, Price DK, Waddell P, Lang M, Courtier-Orgogozo V, et al. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. *Curr Biol*. 2022;32(1):111–123.e5. <https://doi.org/10.1016/j.cub.2021.10.052>.
- Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989;595(3):585–595. <https://doi.org/10.1093/genetics/123.3.585>.
- Tavare S, Balding DJ, Griffiths JRC, Donnelly P. Inferring coalescence TimesFrom DNA sequence. *Genetics*. 1997;145(2):505–518. <https://doi.org/10.1093/genetics/145.2.505>.
- Tihelka E, Cai C, Giacomelli M, Lozano-Fernandez J, Rota-Stabelli O, Huang D, Engel MS, Donoghue PCJ, Pisani D. The evolution of insect biodiversity. *Curr Biol*. 2021;31(19):R1299–R1311. <https://doi.org/10.1016/j.cub.2021.08.057>.
- Tribble CM, Freyman WA, Landis MJ, Lim JY, Barido-Sottani J, Kopperud BT, Hahn S, May MR. RevGadgets: an R package for visualizing Bayesian phylogenetic analyses from RevBayes. *Methods Ecol Evol*. 2022;13(2):314–323. <https://doi.org/10.1111/2041-210X.13750>.
- Van Belleghem SM, Baquero M, Papa R, Salazar C, McMillan WO, Counterman BA, Jiggins CD, Martin SH. Patterns of Z chromosome divergence among *Heliconius* species highlight the importance of historical demography. *Mol Ecol*. 2018;27(19):3852–3872. <https://doi.org/10.1111/mec.14560>.
- Van den Broeck M, De Cock R, Van Dongen S, Matthyssen E. Blinded by the light: artificial light lowers mate attraction success in female glow-worms (*lampyris noctiluca* L.). *Insects*. 2021;12(8):734. <https://doi.org/10.3390/insects12080734>.
- Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*. 2017;33(14):2202–2204. <https://doi.org/10.1093/bioinformatics/btx153>.
- Wagner DL, Grames EM, Forister ML, Berenbaum MR, Stopak D. Insect decline in the anthropocene: death by a thousand cuts. *Proc Natl Acad Sci U S A*. 2021;118(2):e2023989118. <https://doi.org/10.1073/pnas.2023989118>.
- Wakeley J. *Coalescent theory: an Introduction*. 1 ed. New York: Macmillan Learning; 2016.
- Wasserman M, Ehrman L. Firefly chromosomes, II. (*lampyridae*: coleoptera). *Florida Entomol*. 1986;69(4):755. <https://doi.org/10.2307/3495223>.
- Watterson GA. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol*. 1975;276(2):256–276. [https://doi.org/10.1016/0040-5809\(75\)90020-9](https://doi.org/10.1016/0040-5809(75)90020-9).
- Wegmann D, Leuenberger C, Excoffier L. Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics*. 2009;182(4):1207–1218. <https://doi.org/10.1534/genetics.109.102509>.
- Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution (N Y)*. 1984;38:1358–1370. <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x>.
- Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome assembler. *Bioinformatics*. 2013;29(21):2669–2677. <https://doi.org/10.1093/bioinformatics/btt476>.