Dual expansion routes likely underlie the present-day population structure in a *Parnassius* butterfly across the Japanese Archipelago

Hideyuki Tamura[1], Tomoaki Noda[1], Mikiko Hayashi[1], Yuko Fujii[1], Noriko Iwata[1], Yuko Yokota[1], Masanori Murata[1], Chisato Tatematsu[1], Hideshi Naka[1#], Akio Tera, Katsumi Ono, Kakeru Yokoi[1$], Takanori Kato[1], Tomoko Okamoto[1], Koji Tsuchida[1*]

1: Laboratory of Insect Ecology, Faculty of Applied Biological Sciences, Gifu University, Yanagido 1-1, Gifu City, Gifu 501-1193, Japan.

[#]Present Address: The United Graduate School of Agricultural Sciences, Tottori University, Koyama-Minami 4-101, Tottori City, Tottori 680-8553, Japan

[$]Present Address: Insect Genome Research Unit, The National Agriculture and Research Organization, Owashi 1-2, Tsukuba City, Ibaraki 305-8634, Japan

*: to whom correspondence should be addressed.

Abstract

The Japanese Archipelago, comprising a series of isolated yet interconnected islands, had been geographically separated from the Eurasian continent. The linear topography presents a unique biogeographic context for dispersing organisms from the continent. In this study, we utilized mitochondrial DNA (mtDNA) and single nucleotide polymorphism (SNP) variation to elucidate the dispersal history of the Japanese clouded butterfly *Parnassius glacialis* across the Japanese Archipelago, including North Island (Hokkaido), Main Island (Honshu) and Shikoku Island. Our analysis of mtDNA (COI, COII) at 1192 base pairs (bps) revealed 49 haplotypes and identified three distinct haplotype groups in the network. These groups correspond geographically to East Japan, West Japan, and Chugoku-Shikoku. The Chugoku-Shikoku group is the most ancient lineage. Interestingly, the Chugoku-Shikoku lineage showed a closer network connection to the East Japan lineage than the geographically proximate West Japan lineage. Divergence time estimates suggest that the Chugoku-Shikoku lineage diverged from the continental *P. glacialis* approximately 3.05 million years ago (MYA). Subsequently, from the Chugoku-Shikoku lineage, the East Japan and West Japan lineages diverged around 1.05 MYA, with the subsequent divergence between the East and West Japan lineages occurring at approximately 0.62 MYA. Based on the 3067 SNP genotypes, population structure analysis revealed five distinct genetic structures within the Japanese Archipelago, indicating geographical differentiation. From the analyses by mtDNA and SNP variations, four primary genetic barriers were identified: between Hokkaido and Honshu, between East and Central Japan, within the Kansai region, and within the Chugoku region. The former three lines corresponded to the Blakiston Line, the Itoigawa-Shizuoka Tectonic Line, and Lake Biwa, respectively. These findings

suggest that *P. glacialis* diverged from the continental *P. glacialis* and expanded its range across the Japanese Archipelago via the North and South routes, establishing its current distribution.

Keywords: phylogeography, *Parnassius*, Japanese Archipelago, mtDNA, SNP

1. Introduction

Evolutionary biological factors, including distances between individuals and populations, stochastic events, and local population adaptations (e.g., Wright, 1943; Nosil and Crespi, 2004; Nosil et al., 2005), reflect relatively recent events. On the other hand, geological factors encompassing the formation of historical landforms reflect considerably older events than the former (e.g., Otofuji et al., 1985; Jolivet et al., 1994; Martin, 2011). The distribution of organisms is influenced by these two main factors, resulting in their current distribution patterns. Indeed, organisms carry the history of these influences.

The Proto-Japanese archipelago islands were located on the eastern edge of the Eurasian continent until 20 million years ago (MYA), constituting the easternmost part of the Eurasian continent. They originated from an accretionary wedge created by the subduction of the oceanic plate beneath the continental plate (Nakajima, 2018). According to the double-door hypothesis (Otofuji et al., 1985; Jolivet et al., 1994; Martin, 2011), two half-arcs in the south-west and the north-east were later differentiated and migrated from the continent, forming the Sea of Japan with them. These half-arcs underwent opposing rotations and moved to their approximate positions, separated by a valley known as Fossa Magna. This valley was later filled through

alluvial and volcanic activity at approximately 5 MYA, connecting the two half-arcs by land and roughly forming the present-day outline of the Japanese Archipelago.

As the Japanese Archipelago is linearly connected from subarctic to subtropical zones, species expansions into the Archipelago can be divided into two main routes (Fig. 1). The first is from the West, which can be further divided into two sub-routes: via the Korean Peninsula and the Ryukyu Archipelago. A land bridge connected eastern China and western Japan continuously from the Late Miocene to the end of the Pliocene (10–1.7 MYA; million years ago), allowing continuous immigration of continental species into the Archipelago. Kitamura and Kimoto (2004) estimated the presence of a land bridge between Kyushu and the continent at least 3.1 and 1.7 MYA, followed by subsequent submergence.

In contrast, the North route is also divided into two sub-routes: one via Sakhalin and the other via the Kuril Islands. The northern island of Hokkaido was repeatedly connected to the continent during the Pleistocene, including the last glacial periods (0.01−0.07 MYA, Fujimaki, 1994; Goto, 1994), with a land bridge between Sakhalin and Hokkaido facilitating the immigration of many faunal species from the continent. This land bridge is supported by evidence that the same phylogenetic insects are distributed in both Hokkaido and Sakhalin (e.g., Sota and Hayashi, 2007; Hayashi and Sota, 2014). Both the West and North routes are connected to the Eurasian continent, and it is thought that part of the lineages differentiated on the Eurasian continent spread and settled in the Japanese Archipelago. Occasionally, these lineages are dispersed from the Japanese Archipelago back to the continent via these routes (Tojo et al., 2017).

Species widely distributed across the Japanese Archipelago are excellent

subjects for elucidating the processes of their spread and exploring the mechanisms underlying their local adaptation. Among these species, butterflies of the genus *Parnassius* have attracted significant interest from researchers investigating their dispersal from the continent and their migration routes. This genus originated on the Tibetan Plateau and was initially adapted to alpine environments. It is believed that *P. glacialis* differentiated as it adapted to lowland climates after experiencing a population bottleneck in the lowlands of the continent (Si et al., 2020; Su et al., 2020; Tao et al., 2020; Zhao et al., 2022, 2023; Tian et al., 2023).

Parnassius glacialis is a butterfly species distributed across Hokkaido (Northern Island), Honshu (Main Island), and Shikoku Island in the Japanese Archipelago. It was initially described as *P. glacialis* but has recently been referred as a separate species, *P. citrinarius* without any supporting evidence (see supplementary materials for the species name used in this study). Yagi et al. (2001) reported that populations of both *P. citrinarius* (= *glacialis*) and *P. stubbendorfii* in Japan may have been isolated from the continent on the Japanese islands during the early Pleistocene glacial period (ca. 1.7-2.0 Ma). They noted that the genetic distance between the Japanese *P. citrinarius* (= *glacialis*) and continental *P. glacialis* is substantial enough to classify them as separate species compared to the genetic distances observed between some other *Parnassius* species. Recently, Nagata (2024) also found that *P. citrinarius* (= *glacialis*) shows differentiation between eastern and western Japan within the Archipelago. However, the relatively small numbers of nucleotide sequences used in the studies have limited the ability to further clarify the detailed degree of differentiation and the expansion processes into the Japanese Archipelago.

This study aimed to analyze the DNA variations of *P. glacialis* in Japan to

estimate the dispersal process from the continent and the degree of variation within the Japanese Archipelago. For this purpose, the COI and COII regions in the mitochondrial DNA (mtDNA) were analyzed. These mtDNA analyses revealed that the Japanese lineage can be divided into three main lineages. The relationship between these lineages and the continental *P. glacialis* was also examined, and the divergence dates were estimated using whole mtDNA genome sequencing. Furthermore, the recently developed GRAS-Di method (Enoki and Takeuchi, 2018; Hosoya et al., 2019) was employed to detect SNP loci in the species and to explore detailed population structure based on these loci.

2. Materials and Methods

2.1. Samples

We collected the butterflies with an insect net and taken back to the laboratory. Adult butterflies were collected at one to nine sites per prefecture, and the prefectural units were treated as a single population in this study. Their wings were excised and dried, while the remaining parts were stored in 100% ethanol (EtOH) at 4°C until DNA extraction. All samples were collected from areas where this species is not banned from collecting. From the samples, we selected 160 samples for 22 populations for SNP analyses and 185 samples for 24 populations for mtDNA analyses. Samples for mtDNA were collected in Fukui and Tochigi prefectures in addition to these sites for SNP analyses. DNA was extracted from each sample as described below.

2.2. Sequencing of mtDNA COI and COII

DNA was extracted from two or three legs per specimen using a NucleoSpin® Tissue kit (MACHEREY-NAGEL Düren, Germany). The isolated DNA was

138 resuspended in Tris-HCl buffer (pH 7.5) and stored at 4 °C. Polymerase chain reaction
139 (PCR) was conducted to amplify two gene regions: the contiguous mitochondrial
140 cytochrome oxidase I (COI) region of 721 base pairs (bps) and cytochrome oxidase II
141 (COII) region of 470 bps. Primers to amplify the COI region were JNfwd (5'-
142 GCAGGAACTGGATGAACAG-3') and Krev (5'-GAGTATCGTCGAGGTATCC-3 ')
143 (Dechaine et al., 2004). Primers used to amplify the COII region were PierreIII (5'-
144 AGAGAGTTTCACCTTTAATAGAACA-3'; Nazari et al., 2007) and Eva (5'-
145 GAGAACATTACTTGCTTTCAGTCATCT-3'; Bogdanowicz et al., 1993). Both
146 fragments were amplified using TaKaRa Ex Taq Hot Start Version (TaKaRa, Tokyo,
147 Japan). PCR conditions for the COI region were 94°C for 5 min, followed by 30 cycles
148 of 94°C for 30 s, 50°C for 30 s, and 72°C for 1 min. For the COII region, 95°C for 5
149 min, followed by 35 cycles of 94°C for 1 min, 50°C for 1 min, 72°C for 1 min, and
150 finally 72°C for 10 min. Sequencing reactions were performed using BigDye terminator
151 v3.1 (Applied Biosystems, Carlsbad, USA). All samples were sequenced using an ABI
152 PRISM 3100-Avant Genetic Analyzer (Applied Biosystems, Carlsbad, CA, USA).
153 Sequencing was conducted at the Life Science Experimental Center and Genome
154 Research Department, Gifu University. PCR fragments were sequenced in both
155 directions to ensure accuracy. The sequences were aligned the sequences using MEGA
156 6.06 (Tamura et al., 2013).
157 We estimated population demography using a Bayesian skyline plot by BEAST
158 v.2.7.6 (Drummond et al., 2006; Drummond & Rambaut, 2007). Nucleotide diversity,
159 haplotype diversity, Fu's F, and Tajima's D were estimated by Arlequine 3.5.2.2
160 (Excoffier and Lischer 2010).
161 *2.3. Whole Genome Sequencing of mtDNA for Three Lineages*

Our COI and COII analyses revealed that *P. glacialis* across the Japanese Archipelago is broadly divided into the East Japan, West, and Chugoku-Shikoku lineages (see Results). Therefore, three representative individuals were selected from these three lineages (the East Japan from Hokkaido, West Japan from Shiga, and Chugoku-Shikoku from Kochi), and DNA was extracted from the thoraxes of these specimens. Extraction was carried out using the previously mentioned kit, following the manufacturer's protocol. After DNA extraction, the following analyses were conducted at a commercial sequencing facility (Bioengineering Lab Co., Ltd., Kanagawa, Japan). Libraries were prepared according to the manual using the MGIEasy FS DNA Library Prep Set (MGI Tech). Following the provided manual, Cyclized DNA was prepared from the library and the MGIEasy Circularisation Kit (MGI Tech). DNBs were prepared according to the manual using the DNBSEQ-G400RS High-throughput Sequencing Kit (MGI Tech). Sequencing analysis of the produced DNBs was performed using the primer supplied with the High-throughput Pair-End Sequencing Primer Kit (App-D) (MGI Tech) and DNBSEQ-G400 (MGI Tech) platform at 2x150 bp. After removing adapter sequences using Cutadapt v. 4.0 (Martin, 2011), Sickle v. 1.33 (Joshi and Fass, 2011) was used to remove bases with a quality score of less than 20 and paired reads with less than 75 bases. SPAdes v. 3.15.5 (Bankevich et al. 2012) was used to assemble high-quality reads with the following parameters: a lower limit of read coverage (--cov-cutoff = 10) and mismatch correction (–careful). Mitochondrial sequences were extracted from the assembled sequences using Metaxa v. 2.2（Bengtsson-Palme et al., 2015）.

Adapter trimming and low-quality read removal were performed with Trimmomatic v.0.39 (Bolger et al., 2014). Sequence assembly was performed using

GetOrganelle v.1.7.7.1 (Jin et al., 2020). Alignment of these sequences together with the continental *Parnassius* sequences yielded a dataset of 15,185 bp, including indels (accession number = submitting).

2.4. SNP Detection

SNPs were detected from the extracted DNA using the GRAS-Di method. Enoki and Takeuchi (2018) presented the protocol at a conference, and Hosoya et al. (2019) subsequently published the detailed procedure. This protocol constructs libraries using two sequential PCR steps like MIG-seq (Suyama and Matsuki 2015). The first PCR primers consist of 10 bases of Illumina Nextera adaptor 3′-end sequences plus 3-base random oligomers (13 bases). The final PCR product is purified using columns or magnetic beads without size selection and then applied for sequencing on an Illumina platform. GRAS-Di has the advantages of simplicity in library construction and the ability to detect many SNPs. Libraries were prepared using the 2nd-step tailed PCR method; primers for 1st and 2nd PCR were followed by Hosoya et al. (2019). Cyclized DNA was prepared according to the manual using the produced library and the MGIEasy Universal Library Conversion Kit (App-A, MGI Tech). DNA Nanoballs (DNB) were prepared using the DNBSEQ DNB Rapid Make Reagent Kit (MGI Tech) and High-Throughput Pair-End Sequencing Primer Kit (App-D, MGI Tech), following the provided manuals. Sequencing was performed using DNBSEQ-T7RS High-throughput Sequencing Kit (MGI Tech) with the primer supplied with the High-Throughput Pair-End Sequencing Primer Kit (App-D) and DNBSEQ-T7 (MGI Tech) platform at a 2x150 bp condition. To remove primer sequences, the first three bases of each read were removed using Cutadapt v. 4.0 (Martin, 2011). Paired reads with a quality score of less than 30 and less than 75 bases were filtered out using Sickle v.1.33

(Joshi and Fass, 2011). Sequences after 75 nucleotides were retained to standardize read length for data analysis. The raw fastq sequence data of each sample were deposited into NCBI Sequence Read Archive (Supplemental Table 2).

Subsequently, denovo_map.pl of Stacks v. 2.62 (Catchen et al., 2013) was executed with the following parameters: minimum number of reads required to form a stack (m = 5), maximum number of nucleotide differences allowed between stacks for merging (M = 2), maximum number of nucleotide differences allowed when building the catalog (n = 2,), and assembly using paired-end reads (--paired). The SNPs information was obtained by running the program with the following settings: n = 2 (maximum number of nucleotide differences allowed when merging stacks) and assembly using paired end reads. These analyses were conducted at the commercial sequencing facility.

2.5. Phylogeographic Analyses

For COI and COII data, phylogenetic trees were constructed using the neighbor-joining (NJ) (Saitou & Nei, 1987), maximum likelihood (ML), and Bayesian inference (BI) methods. NJ trees were generated using MEGA X (Kumar et al. 2018) with the *p*-distance method. ML trees were constructed with RAxML (Stamatakis, 2014), and BI trees were built with MrBayes v.3.1.2 (Ronquist & Huelsenbeck, 2003). Bootstrap testing was conducted with 1,000 trials for the NJ and ML trees. ML analysis was performed using each model for each gene codon and assessed using Akaike information criterion (AIC) scores in Modeltest2 v. 2.3 (Nylander, 2004), and the selected model was HKY + I + G4. Similarly, BI was performed using each model for each gene codon and assessed using AIC scores in Kakusan4 (Tanabe, 2011), and the selected model was HKY + I + G. We aligned the sequences of 185 adults. As

234 outgroups, we used the continental *Parnassius* sequences: *Parunassius glacialis* in
235 China (NC 065029_1), *P. stubbendorfii* (OP 709281_1), *P. orleans* (NC 072335_1), and
236 *P. epaphus* (NC 026864_1). Haplotype networks were obtained with TCS v. 1.21
237 (Clement et al., 2000).
238 For whole mtDNA sequence data, phylogenetic trees were constructed using the
239 neighbor-joining (NJ) (Saitou & Nei, 1987), maximum likelihood (ML), and Bayesian
240 inference (BI) methods. NJ trees were generated using MEGA X (Kumar et al. 2018)
241 with the *p*-distance method. ML trees were constructed with RAxML (Stamatakis,
242 2014), and BI trees were built with MrBayes v.3.1.2 (Ronquist & Huelsenbeck, 2003).
243 Bootstrap testing was conducted with 1,000 trials for the NJ and ML trees. ML analysis
244 was performed using each model for each gene codon and assessed using Akaike
245 information criterion (AIC) scores in Modeltest2 v. 2.3 (Nylander, 2004), and the
246 selected model was TIM2 + G4. Similarly, BI was performed using each model for each
247 gene codon and assessed using AIC scores in Kakusan4 (Tanabe, 2011). The selected
248 model was GTR + G. We aligned the sequences of three adults of the three
249 representative lineages (East Japan, West Japan, and Chugoku-Shikoku). As outgroups,
250 we used the continental *Parnassius* sequences; *Parunassius glacialis* in China (NC
251 065029_1), *P. stubbendorfii* (OP 709281_1), *P. orleans* (NC 072335_1), and *P. epaphus*
252 (NC 026864_1), *P. acdestis* (NC 072544_1), *P. hide* (NC 072287_1), *P. cephalus* (NC
253 026457_1), *P. simo* (NC 072286_1).
254 We estimated divergence times using BEAST v. 2.7.6 (Drummond et al., 2006;
255 Drummond & Rambaut, 2007) with 10,000,000 generations of Markov-chain Monte
256 Carlo iterations, sampled every 1,000 generations under a relaxed clock model with an
257 uncorrelated lognormal distribution. Here, we used the HKY + G model and three

calibration time points according to Si et al. (2020). The calibrations for the molecular dating were based on two butterfly fossils: *Praepapilio colorado* Durden (Papilionidae) from the Green River Shale of Colorado (USA) of mid-Eocene early Lutetian age (41–48 MYA) and *Thaites ruminiana* Scudder (subfamily Parnassiinae) from Aix-en-Provence (southern France) of late Oligocene Chattian age (23–28 MYA) (Durden and Rose 1978, Jong 2017, Condamine et al. 2018); thus, the crown group divergence of the Parnassiinae was constrained between 23 MYA and 48 MYA. The earliest divergence time of seven *Parnassius* subgenera other than *P. epaphus,* exclusive of subgenus *Parnassius* was set to be 16–37 Ma.

2.6. SNP Data Analysis

We used STRUCTURE v. 2.0 (Pritchard et al., 2000) to delineate the number of genetically identified clusters (*K*) and assign individuals to clusters without prior information on their origin population. We calculated an *ad hoc* criterion of *ΔK* (Evanno et al., 2005) for determining the optimal *K* value using STRUCTURE Harvester (Earl & Holdt, 2012). The assignment index for each individual in each population was calculated using the CLUMPAK server (Kopelman et al., 2015). Genetic variation associated with various hierarchical levels of populations and their fixation index (*F*) values were quantified using analysis of molecular variance (AMOVA; Excoffier et al., 1992) in GenoDive ver. 3.0. (Meirmans, 2020). We conducted principal component analysis (PCA) using GenoDive v. 3.0.

We used Netview R v.1.0 operated by R v.4.2.3 (Neuditschko et al. 2012; Steinig et al. 2016) to reveal fine-scale population stratification independent of *a priori* ancestry information. We generated a population network based on a shared allele distance matrix (1-identity by state (IBS)) generated with PLINK v.1.9. The network was

visualized by R v.4.2.3. We also used TreeMix v.1.13 (Pickrell and Pritchard, 2012) to investigate patterns of population splits and cross-population gene flow.

3. Results

3.1. Phylogeography

We detected and aligned COI (721 bps) and COII (470 bps) for each individual (accession number = in submitting). Although, if the samples in Okayama were generally included in the Chugoku Region were put into West Japan (Haplotype 06), three distinct clusters were identified in the Haplotype network by TCS (Fig. 2), corresponding to the West Japan, East Japan, Chugoku-Shikoku respectively. Demography statistics were calculated for the three strains, and the Chugoku-Shikoku strain showed a trend towards higher nucleotide diversity than the other strains (Table S1). Fu's *F* values were significantly negative in the West Japan and East Japan lineages, while Tajima's *D* was not significantly negative, suggesting a weak tendency towards population expansion. On the other hand, neither value was negative for the Chugoku-Shikoku lineage. Skyline plots showed recent population increases in West Japan and the Chugoku -Shikoku lineage (Fig. S1) but not in the East Japan lineage.

The three phylogenetic trees could be combined into one consensus tree (Fig. 3). There, the posterior probability and bootstrap value for the branching of the Chugoku-Shikoku lineage were highest, from which the East Japan and West Japan lineages diverged. A clear difference between the three phylogenetic trees was the position of the continental *P. glacialis*. The position of the continental *P. glacialis* was out of Japanese lineages in NJ (Fig. S2). However, the ML and BI trees contained that position within the Japanese lineages (Fig. S3 and 3).

306 The continental *P. glacialis* was included within *P. glacialis* in Japan, so the
307 phylogenetic tree was rebuilt using the whole mtDNA sequences of the continental
308 *Parnassius* species and those of three lineages of *P. glacialis* in Japan (accession
309 number = in submitting) by the three phylogenetic methods and Beast. The consensus
310 tree (Fig. S4) showed the continental *P. glacialis* located outside *P. glacialis* in Japan,
311 which was derived from the continental *P. glacialis*, which was supported with high
312 probability (1/100/100: BI/ML/NJ). The oldest lineage of *P. glacialis* in Japan was the
313 Chugoku-Shikoku lineage (Fig. 4), which was supported with a high probability
314 (1/100/100). *P. glacialis* in Japan diverged from the continental *P. glacialis* at 3.05
315 MYA ; after that, the remaining the East Japan and West Japan lineages diverged at 1.05
316 MYA, showing again that the Chugoku-Shikoku lineages were most ancestral ones
317 across the Japanese Archipelago. The East Japan and West Japan lineages diverged at
318 0.62 MYA.

319 *3.2. Population structure depicted by SNPs*

320 We detected 3067 SNP loci using a method of GRAS-Di (Enoki and Takeuchi,
321 2018; Hosoya et al., 2019) and analyzed with STRUCTURE based on the information
322 of those loci for 160 individuals. The results showed that ΔK (Evanno et al., 2005) for
323 determining the optimal K value using STRUCTURE Harvester (Earl & Holdt, 2012)
324 was highest at $K = 5$, which was the most appropriate cluster number (Fig. 5). Four clear
325 genetic clusters were found at $K = 5$: Hokkaido and Tohoku (dark blue), Koshinetsu
326 (blue), Chubu (white) and Shikoku (orange)—in the case of $K = 6$, a new cluster
327 appeared in Shizuoka, indicating that the populations across the Japanese Archipelago
328 were divided into eastern and western Japan, with Shizuoka as the boundary.

329 The results analyzed by NetviewR were consistent with those of STRUCTURE

(Fig. 6). Characteristically, Hokkaido was isolated with the Blakiston Line between Hokkaido and Honshu (main Island), and the Kansai was divided into two groups. One group of the Kansai was connected to the Chugoku, while the other was connected to the Chubu. Individuals from the Kansai connected to the Chugoku were collected from the western shore side of Lake Biwa to Hyogo prefecture. In contrast, those from the Kansai connected to the Chubu were collected from the eastern shore side of Lake Biwa to Toyama prefecture, with Lake Biwa acting as one geographical barrier. Another clear barrier was found between the northern regions (Tohoku, Koshinetsu, and Chubu) and the central regions (Chubu and Kansai).

The individuals from the Chubu connected to those of the Kansai, including those collected in northern Gifu and Toyama prefectures. The other individuals of the Chubu were connected to those of the Koshinetsu and Tohoku and included individuals from southern Gifu and Shizuoka. The individuals from Sado Island in the Sea of Japan were not connected to the individuals from Niigata, where it is administratively included, but rather to those from the Tohoku. These results were generally consistent with the Treemix results (Fig. S5).

The PCA results were consistent with the STRUCTURE and NetviewR results (Fig. S6). Hokkaido and Shizuoka prefectures were located separately, and the three populations in the Shikoku region also formed a single cluster.

4. Discussion

We revealed that *P. glacialis* in Japan was divided into three distinct genetic lineages in mtDNA of COI + COII analyses: East Japan, spanning from Hokkaido to the Koshinetsu region; West Japan, spanning Chubu to the Chugoku region; and the

354 Chugoku-Shikoku region. The results of the SNP analyses generally supported these
355 genetic structures. Of particular interest is that the Chugoku-Shikoku lineage was
356 estimated to have the oldest branching time. Furthermore, the Haplotype network
357 showed that the East Japan lineage was connected to the Chugoku-Shikoku lineage
358 rather than the geographically closer West Japan lineage (Fig. 2).

359 In the phylogenetic trees with the continental *Parnassius* species as the
360 outgroup, three clusters were generally identified in the trees from the three methods
361 (Fig. 3). However, phylogenetic relationships other than the Chugoku-Shikoku lineage
362 were unclear in the ML method (Fig. S3). Despite this, all phylogenetic trees suggested
363 that the Chugoku-Shikoku lineage diverged first. The phylogenetic tree based on 1192
364 bp of mtDNA sometimes showed the continental *P. glacialis* included within the
365 phylogenetic tree of *P. glacialis* in Japan, but this issue was resolved using the whole
366 mtDNA sequence (Fig. S4). Our branching time estimates indicate that *P. glacialis* in
367 Japan diverged from the continental *P. glacialis* approximately 3.05 MYA, with the two
368 remaining lineages diverging at 1.05MYA from the Chugoku-Shikoku lineage,
369 confirming that the Chugoku-Shikoku lineage is the most ancestral across the Japanese
370 Archipelago. Furthermore, the haplotype network demonstrated that the East Japan
371 lineage is linked to the Chugoku-Shikoku lineage rather than to the geographically
372 closer West Japan lineage (Fig. 2). These results suggest that the East Japan lineage may
373 have been derived from the Chugoku-Shikoku lineage, followed by the divergence of
374 the West Japan lineage from the East Japan lineage.

375 Our results show that (1) the Chugoku-Shikoku lineage is the oldest lineage, and
376 (2) the East Japan lineage is linked to the Chugoku-Shikoku lineage rather than to the
377 geographically closer West Japan lineage. Two routes have been postulated for the

expansion of biota distribution in the Japanese Archipelago: A North route and a West route (Fig. 1). Let us now consider two scenarios in which the distribution is expanded from one of these routes. For example, if we assume that the East Japan lineage expanded its distribution from the North route and then derived the West Japan lineage and the Chugoku-Shikoku lineage, the Chugoku-Shikoku lineage would be the most recent lineage. This would be inconsistent with our data, which indicates that the Chugoku-Shikoku lineage was the oldest.

Let us also consider another scenario in which the Chugoku-Shikoku lineage expanded its distribution from the West route, from which the West Japan and East Japan lineage were subsequently derived. In this scenario, the most ancestral Chugoku-Shikoku lineage would be closely related to the West Japan lineage, which would also be inconsistent with our data. In other words, assuming a single route would not resolve the discrepancy with our data. The most appropriate route to resolve these discrepancies is a scenario in which two routes have been followed. Namely, the oldest Chugoku-Shikoku lineage expanded into the Japanese Archipelago via the West route. In contrast, the East Japan lineage, which is phylogenetically most closely linked to it, expanded its distribution into the Japanese Archipelago via the North route. The scenario in which the West Japan lineage diverged from the East Japan lineage in the Japanese Archipelago is the scenario that can explain our data without contradiction. In other words, this species is thought to have differentiated from the continental *P. glacialis* in mainland China, expanded its distribution in the Japanese Archipelago via two routes, and formed a contact zone in the central mainland (Honshu).

Organisms that have differentiated traits and phylogeny from east to west within the Japanese Archipelago have been known, and it has been suggested that they may

have expanded their distribution in the Japanese Archipelago via two routes, one to the North and the other to the West (Suzuki et al., 1996, 2002; Yoshimura et al., 2001; Schoville et al., 2013; Hayashi and Sota, 2014). However, neither of these phenomena has been confirmed by phylogeography. The present data are the first results supporting the two-route hypothesis.

SNP analyses with the GRAS-Di detected 3067 loci. STRUCTURE analysis indicated $K = 5$ was the most appropriate population number. Four clusters were identified at $K = 5$ (Fig. 5): Hokkaido/Tohoku (dark blue), Koshinetsu (blue), Chubu (white), and Shikoku (orange). These populations are connected from the north to the south of the Japanese Archipelago, and these boundaries may act as barriers to gene flow.

Fig. 7 compares the boundaries of the clusters as identified from mtDNA (A) and SNP (B). The mtDNA results observed two clear barriers (a and b). One barrier (b) corresponds well to the Itoigawa-Shizuoka Tectonic Line, which marks the western end of the Fossa Magna. The other barrier divides the area, including Shikoku region and Hiroshima prefectures, from the area further east. The former barrier is known to divide organisms east and west of this line in other species, and geohistorical influences have been repeatedly noted (Suzuki et al., 1996, 2002; Yoshikawa et al., 2001). Conversely, the latter barrier corresponds to the boundary between Hiroshima and Okayama Prefectures, but it has not been previously identified as a biological boundary.

In contrast, five genetic clusters with four barriers (i to iv in Fig. 7B) were identified in the nuclear genes of SNPs: (1) Hokkaido, (2) Tohoku, (3) East Chubu including Koshinetsu and South Chubu, (4) Hokuriku including Toyama, north Gifu and East of Shiga, (5) West Japan including West Kansai, Chugoku, and Shikoku. A closer

examination reveals that barrier (i) runs through the central part of Kansai and is associated with the boundary at Lake Biwa. The areas between barriers (i) and (ii) include Toyama, northern Gifu, and the eastern shore of Lake Biwa in Shiga. Southern Gifu Prefecture, however, shows greater connectivity with the Koshinetsu region between barriers (ii) and (iii). Historically, the northern and southern administrative regions of Gifu Prefecture differed, and natural conditions remain distinct even today. This phenomenon is also observed in another butterfly, *Luehdorfia japonica*, where mtDNA haplotypes from the Chubu to the Kansai region are relatively complex and intricately intermingled in a relatively narrow range (Suzuki et al., 2023). Thus, this distribution may reflect gene flows through relatively recent human activity rather than geohistorical influences.

The combined map of barriers in Fig. 7C shows an area where the three mtDNA and SNP lineages are contacting. The area enclosed by the dotted square indicates this contact zone. In allopatric and parapatric speciation, the post-branching contact of divergent lineages is termed secondary contact (Coyne and Orr, 2004). Here, we used this term to denote the situation where different lineages of the same species are in contact. Comparing the distribution regions of mtDNA (A) and SNP (B) genotypes, SNP (B) is more finely divided. Upon closer examination, we can recognize areas of discrepancy between the two markers. For instance, the position of line (a) in mtDNA differs from that of line (i) in SNP (B). Additionally, line (b) of mtDNA is situated between lines (ii) and (iii) of SNPs. This indicates that individuals entering these two regions belong to the same group for SNPs, but to different groups for mtDNA. This indicates that SNPs are spreading across mtDNA boundaries, suggesting that the spreading nature of the gene is faster for SNPs than for mtDNA.

The triangles in Fig 7C illustrate that from the red line of the mtDNA lineage boundary, there is a nuclear gene expansion towards the blue line of the SNP boundary. This indicates that the gene flow from the boundary (a) to the east opposes the gene flow direction from the boundary (b) to the west, which corresponds to the boundary of Lake Biwa. In other words, nuclear gene flow directions appear to conflict within the contact zone. The biological significance of the Lake Biwa boundary is currently unknown but will be an area of future investigation.

A feature of this result is the difference in the degree of introgression between nuclear DNA and mtDNA: the triangles in Fig. 7C indicate regions of greater introgression of nuclear DNA than mtDNA. A factor that may contribute to this difference in the introgression of nuclear DNA and mtDNA is activity differences between males and females. The behavior of organisms can be divided into two types: male philopatry, where males tend to stay in their territories, and female philopatry, where females tend to remain in their territories (Greenwood, 1980). In the case of butterflies, including this species, most exhibit female philopatry, and the sex ratio of actively flying adults is generally skewed toward males. In this case, nuclear DNA would be expected to spread more easily.

SNP analysis indicates that the Chugoku-Shikoku lineage appears to have been extending eastward. In contrast, the East Japan lineage seems to have been expanding westward across the Sea of Japan in the past. This suggests that the two lineages currently come into contact around Lake Biwa. If this scenario holds, Haldane's law (Coyne and Orr 2004) may be observed, where heterozygous individuals, particularly females in Lepidopteran insects, are more disadvantaged when hybridizing. However, data from amateur entomologists indicate no significant mortality increases were

observed when mating individuals from distant collecting sites (Ono and Tera, personal communication). On the other hand, anecdotal evidence suggests that *P. glacialis* in Japan mates with the continental *P. glacialis* to produce viable offspring. If confirmed, this implies that reproductive isolation between the two lineages between China and Japan is not well operating and warrants careful consideration in future studies.

The present study demonstrates that *P. glacialis* colonized Japan via North and West routes to the Japanese Archipelago. While previous studies (e.g., Hayashi and Sota, 2014) inferred this dual-route entry, it has been genetically confirmed in this study. Despite undergoing long-distance migration across the Sea of Japan via these two routes, there appears to be no apparent fertility deficit in the contact zone. This study provides insights into the species' distant history of this species. Future research on adaptive genes will be crucial for understanding their recent and future evolutionary trajectories.

5. Conclusion

The results of this study indicate that *Parnassius glacialis* in Japan differentiated from the continental *P. glacialis* and subsequently migrated to the Japanese Archipelago via two expansion routes: a West route and a North route. The most ancestral lineage entered the Japanese Archipelago via the West route, while its derived lineages spread via the North route. The lineage derived from the North route and the lineage from the West route are thought to be currently in contact in the central part of the Japanese Archipelago. Within the contact zone of these two lineages, differences in the distribution of nuclear genes and mtDNA were observed, suggesting that female philopatry is one of the contributing factors.

Acknowledgments

We thank T. Takeda, N. Nagahama, T. Takeuchi, M. Kawada, H. Matsuno, A. Yokokura, T. Kudo, J. Ogasawara for their sample collection and providing samples.

References

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology 19, 455–477. https://doi.org/10.1089/cmb.2012.0021

Bogdanowicz, S.M., Wallner, W.E., Bell, J., Odell, T.M., Harrison, R.G., 1993. Asian gypsy moths (Lepidoptera: Lymantriidae) in North America: evidence from molecular data. Annals of the Entomological Society of America 86, 710–715. http://doi.org/10.1006/mpev.1999.0744

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology 22, 3124–3140. http://doi.org/10.1111/mec.12354

Clement, M., Posada, D., Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. Molecular Ecology 9, 1657–1659.

Condamine, F.L., Rolland, J., Höhna, S., Sperling, F.A.H., Sanmartín, I., 2018. Testing

522 the role of the Red Queen and Court Jester as drivers of the macroevolution of
523 Apollo butterflies. Systematic Biology 67, 940–964.
524 http://doi.org/10.1098/rsbl.2014.1049
525 Coyne, J.A., Orr, H.A., 2004. Speciation. Sinauer, Sunderland, UK.
526 De Jong, R., 2017. Fossil butterflies, calibration points and the molecular clock
527 (Lepidoptera: Papilionoidea). Zootaxa 4270, 1–63.
528 Drummond, A., Ho, S., Phillips, M., Rambaut, A., 2006. Relaxed phylogenetics and
529 dating with confidence. PLoS Biology 4, e88.
530 http://doi:10.1371/journal.pbio.0040088
531 Drummond, A., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by
532 sampling trees. BMC Evolutionary Biology 7, 214. http:// doi:10.1186/1471-2148-
533 7-214
534 Durden, C.J., Rose, H., 1978. Butterflies from the middle Eocene: the earliest
535 occurrence of fossil Papilionoidea (Lepidoptera). The earliest occurrence of fossil
536 Papilionoidea (Lepidoptera). Texas Memorial Museum, The University of Texas
537 at Austin.
538 Earl, D., vonHoldt, B., 2012. STRUCTURE HARVESTER: a website and program for
539 visualizing STRUCTURE output and implementing the Evanno method.
540 Conservation Genetics Resources 4, 359–361. http://dx.doi.org/10.1007/s12686-
541 011-9548-7
542 Enoki, H., Takeuchi, Y., 2018. New genotyping technology, GRAS-Di, using next
543 generation sequencer. Plant and Animal genome conference XXVI.
544 Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of
545 individuals using the software STRUCTURE: a simulation study. Molecular

546 Ecology 14, 2611–2620. http://doi.org/10.1111/j.1365-294X.2005.02553.x
547 Excoffier, L., Smouse, P.E., Quattro, J.M., 1992. Analysis of molecular variance
548 inferred from metric distances among DNA haplotypes: application to human
549 mitochondrial DNA restriction data. Genetics 131, 479–491.
550 Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to
551 perform population genetics analyses under Linux and Windows. Molecular
552 Ecology Resources 10, 564–567. http://dx.doi.org/10.1111/j.1755-
553 0998.2010.02847.x
554 Fujimaki, Y., 1994. Across the strait: fauna of mixture. In: Ishigaki, K., Fukuda, M.
555 (Eds.), Formation history of nature of Hokkaido. Hokkaido University Press,
556 Sapporo, pp. 167-179.
557 Goto, A., 1994. Fishes in rivers and lakes: origin and adaptive strategy. In: Ishigaki, K.,
558 Fukuda, M. (Eds.), Formation history of nature of Hokkaido. Hokkaido University
559 Press, Sapporo, pp. 150-166.
560 Greenwood, P.J., 1980. Mating systems, philopatry, and dispersal in birds and
561 mammals. Animal Behaviour 28, 1140–1162.
562 Hayashi, M., Sota, T., 2014. Quaternary donaciine beetles (Coleoptera, Chrysomelidae)
563 in Japan: Colonization and divergence patterns inferred from fossil and molecular
564 data. Quaternary International 341, 255–266.
565 https://doi.org/10.1016/j.quaint.2013.08.022
566 Hosoya, S., Hirase, S., Kikuchi, K., Nanjo, K., Nakamura, Y., Kohno, H., Sano, M.,
567 2019. Random PCR-based genotyping by sequencing technology GRAS-Di
568 (genotyping by random amplicon sequencing, direct) reveals genetic structure of
569 mangrove fishes. Molecular Ecology Resources 19, 1153–1163.

https://doi.org/10.1111/1755-0998.13025

Jolivet, L., Tamaki, K., Fournier, M., 1994. Japan Sea, opening history and mechanism: A synthesis. Journal of Geophysical Research: Solid Earth 99, 22237–22259. https://doi.org/10.1029/93JB03463

Joshi, N.A., Fass, J.N., 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle

Kitamura, A., Kimoto, K., 2004. Reconstruction of the Akihisa Southern Channel of the Japan Sea at 3.9-1.0 Ma. Quaternary Research 43, 417–434 (in Japanese with English summary). http://doi.org/10.4116/jaqua.43.417

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., Mayrose, I., 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources 15, 1179–1191. https://doi.org/10.1111/1755-0998.12387

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Molecular Biology and Evolution 35, 1547–1549. https://doi.org/10.1093/molbev/msy096

Martin, A.K., 2011. Double saloon door tectonics in the Japan Sea, Fossa Magna, and the Japanese Island Arc. Tectonophysics 498, 45–65. https://doi.org/10.1016/j.tecto.2010.11.016

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12. https://doi.org/10.14806/ej.17.1.200

Meirmans, P.G., 2020. genodive version 3.0: Easy-to-use software for the analysis of

genetic data of diploids and polyploids. Molecular Ecology Resources 20, 1126–1131. https://doi.org/10.1111/1755-0998.13145

Nagata, N., 2024. Phylogeography of *Parnassius citrinarius* based on mitochondrial DNA reveals large differences in genetic structure between the Eastern and Western Japan. Zoological Science 41. https://doi.org/10.2108/zs230072

Nakajima, K., 2018. Tectonics of sedimentary basins in and around Japan since the opening of the Sea of Japan. Journal of the Geological Society of Japan 124, 693–722 (in Japanese with English summary). http://doi.org/10.5575/geosoc.2018.0049

Nazari, V., Zakharov, E.V., Sperling, F.A.H., 2007. Phylogeny, historical biogeography, and taxonomic ranking of Parnassiinae (Lepidoptera, Papilionidae) based on morphology and seven genes. Molecular Phylogenetics and Evolution 42, 131–156. https://doi.org/10.1016/j.ympev.2006.06.022

Neuditschko, M., Khatkar, M.S., Raadsma, H.W., 2012. NetView: A high-definition network-visualization approach to detect fine-scale population structures from genome-wide patterns of variation. PloS ONE 7, e48375. http://doi.org/10.1371/journal.pone.0048375

Nosil, P., Crespi, B.J., 2004. Does gene flow constrain adaptive divergence or vice versa? a test using ecomorphology and sexual isolation in *Timema cristinae* walking-sticks. Evolution 58, 102–112. http://doi.org10.1111/j.0014-3820.2004.tb01577.x

Nosil, P., Vines, T.H., Funk, D.J., 2005. Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution 59, 705–719. http://doi.org/10.1111/j.0014-3820.2005.tb01747.x

618 Nylander, J.A.A., 2004. MrModeltest v2. Program distributed by the author.
619 Otofuji, Y.-I., Matsuda, T., Nohda, S., 1985. Opening mode of the Japan Sea inferred
620 from the palaeomagnetism of the Japan Arc. Nature 317, 603–604.
621 http://doi.org/10.1038/317603a0
622 Pickrell, J., Pritchard, J., 2012. Inference of population splits and mixtures from
623 genome-wide allele frequency data. PLoS Genetics, e1002967.
624 http://doi.org/10.1038/npre.2012.6956.1
625 Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using
626 multilocus genotype data. Genetics 155, 945–959.
627 Ronquist, F., Huelsenbeck, J.P., 2003. MrMayes 3: Bayesian phylogenetic inference
628 uder mixed models. Bioinformatics 19, 1572–1574.
629 http://doi.org/10.1093/bioinformatics/btg180
630 Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for
631 reconstructing phylogenetic trees. Molecular Biology and Evolution 4, 406–425.
632 http://doi.org/10.1093/oxfordjournals.molbev.a040454
633 Schoville, S.D., Uchifune, T., Machida, R., 2013. Colliding fragment islands transport
634 independent lineages of endemic rock-crawlers (Grylloblattodea: Grylloblattidae)
635 in the Japanese Archipelago. Molecular Phylogenetics and Evolution 66, 915–927.
636 https://doi.org/10.1016/j.ympev.2012.11.022
637 Si, C., Chen, K., Tao, R., Su, C., Ma, J., Li, L., Hao, J., Yang, Q., 2020. Genetic
638 differentiation and divergence time of Chinese *Parnassius* (Lepidoptera:
639 Papilionidae) species based on nuclear internal transcribed spacer (Its) sequence
640 data. Journal of Entomological Science 55, 520–546.
641 http://doi.org/10.18474/0749-8004-55.4.520

642 Sota, T., Hayashi, M., 2007. Comparative historical biogeography of *Plateumaris* leaf
643 beetles (Coleoptera: Chrysomelidae) in Japan: interplay between fossil and
644 molecular data. Journal of Biogeography 34, 977–993.
645 https://doi.org/10.1111/j.1365-2699.2006.01672.x
646 Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-
647 analysis of large phylogenies. Bioinformatics 30, 1312–1313.
648 http://doi.org/10.1093/bioinformatics/btu033
649 Steinig, E.J., Neuditschko, M., Khatkar, M.S., Raadsma, H.W., Zenger, K.R., 2016.
650 netview p: a network visualization tool to unravel complex population structure
651 using genome-wide SNPs. Molecular Ecology Resources 16, 216–227.
652 https://doi.org/10.1111/1755-0998.12442
653 Su, C., Xie, T., Wang, Y., Si, C., Li, L., Ma, J., Li, C., Sun, X., Hao, J., Yang, Q., 2020.
654 Miocene diversification and high-altitude adaptation of *Parnassius* butterflies
655 (Lepidoptera: Papilionidae) in Qinghai–Tibet Plateau revealed by large-scale
656 transcriptomic data. Insects 11, 754. http://doi.org/10.3390/insects11110754
657 Suyama, Y., Matsuki, Y., 2015. MIG-seq: an effective PCR-based method for genome-
658 wide single-nucleotide polymorphism genotyping using the next-generation
659 sequencing platform. Scientific Reports 5, 16963.
660 http://doi.org/10.1038/srep16963
661 Suzuki, H., Sato, Y., Fujiyama, S., Ohba, N., 1996. Allozymic differentiation between
662 two ecological types of flashing behavior in the Japanese firefly, *Luciola cruciata*.
663 Japanese Journal of Entomology 64, 682–691.
664 Suzuki, H., Sato, Y., Ohba, N., 2002. Gene diversity and geographic differentiation in
665 mitochondrial DNA of the Genji firefly, *Luciola cruciata* (Coleoptera:

666 Lampyridae). Molecular Phylogenetics and Evolution 22, 193–205.
667 http://dx.doi.org/10.1006/mpev.2001.1046
668 Suzuki, S., Suzuki, T., Shibuya, Y., Goto, M., Yokoyama, J., Kato, T., Ando, M.,
669 Okamoto, T., Tsuchida, K., 2023. Expansion processes of two emblematic
670 Luehdorfia butterflies across the Japanese Archipelago. Journal of Biogeography
671 50, 1710–1723. https://doi.org/10.1111/jbi.14682
672 Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6:
673 Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and
674 Evolution 30, 2725–2729. http://doi.org/10.1093/molbev/mst197
675 Tanabe, A.S., 2011. Kakusan4 and Aminosan: two programs for comparing
676 nonpartitioned, proportional and separate models for combined molecular
677 phylogenetic analyses of multilocus sequence data. Molecular Ecology Resources
678 11, 914–921. https://doi.org/10.1111/j.1755-0998.2011.03021.x
679 Tao, R., Xu, C., Wang, Y., Sun, X., Li, C., Ma, J., Hao, J., Yang, Q., 2020.
680 Spatiotemporal differentiation of alpine butterfly *Parnassius glacialis*
681 (Papilionidae: Parnassiinae) in China: Evidence from mitochondrial dna and
682 nuclear single nucleotide polymorphisms. Genes 11, 188.
683 https://doi.org/10.3390/genes11020188
684 Tian, X., Mo, S., Liang, D., Wang, H., Zhang, P., 2023. Amplicon capture
685 phylogenomics provides new insights into the phylogeny and evolution of alpine
686 *Parnassius* butterflies (Lepidoptera: Papilionidae). Systematic Entomology 48,
687 571–584. https://doi.org/10.1111/syen.12591
688 Tojo, K., Sekiné, K., Takenaka, M., Isaka, Y., Komaki, S., Suzuki, T., Schoville, S.D.,
689 2017. Species diversity of insects in Japan: Their origins and diversification

processes. Entomological Science 20, 357–381. http://doi.org/10.1111/ens.12261

Wright, S., 1943. Isolation by distance. Genetics 28, 114–138.

Yagi, T., Katoh, T., Chichvarkhin, A., Shinkawa, T., Omoto, K., 2001. Molecular phylogeny of butterflies *Parnassius glacialis* and *P. stubbendorfii* at various localities in East Asia. Genes & Genetic Systems 76, 229–234. http://doi.org/10.1266/ggs.76.229

Yoshikwa, T., Ide, K., Kubota, Y., Nakamura, Y., Takebe, H., Kusaoke, H., 2001. Intraspecific genetic variation and molecular phylogeny of *Luciola cruciata* (Coleoptera: Lampyridae) inferred from the mitochondrial ND5 gene sequences. Japanese Journal of Entomology 4, 117–127 (in Japanese with English summary). http://doi.org/10.20848/kontyu.4.4_117

Zhao, Y., He, B., Tao, R., Su, C., Ma, J., Hao, J., Yang, Q., 2022. Phylogeny and biogeographic history of Parnassius butterflies (Papilionidae: Parnassiinae) reveal their origin and deep diversification in West China. Insects 13, 406. http://doi.org/10.3390/insects13050406

Zhao, Y., Su, C., He, B., Nie, R., Wang, Y., Ma, J., Song, J., Yang, Q., Hao, J., 2023. Dispersal from the Qinghai-Tibet plateau by a high-altitude butterfly is associated with rapid expansion and reorganization of its genome. Nature Communications 14, 8190. http://doi.org/10.1038/s41467-023-44023-2

Author contributions: CrediT

H.N., A.T., K.O., T.O and K.T. conceived the study; H.T., T.N., M.H., Y.F., N.I., Y.Y., M.M., C.T., H.N., A.T., K.O., and K.T. collected the samples and extracted DNA. H.T., T.N. M.H., Y.F., N.I., Y.Y., M.M., and C.T. analysed genetic variation of mtDNA; K.Y,

T.K., T.O., and K.T. conducted data analyses; K.T. wrote the manuscript; all authors approved the final version of the manuscript.

Conflicts of interest

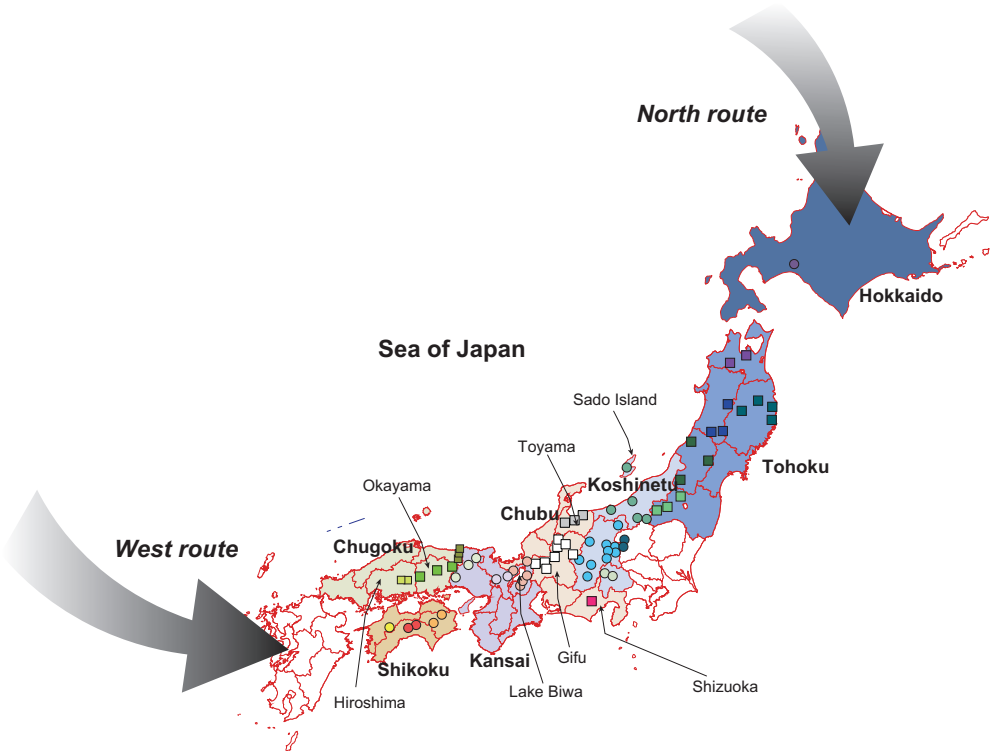
The authors declare that they have no competing interests.

Data accessibility

All genetic and related data used in the analyses will be deposited in Dryad.

724 **Legends of figures**

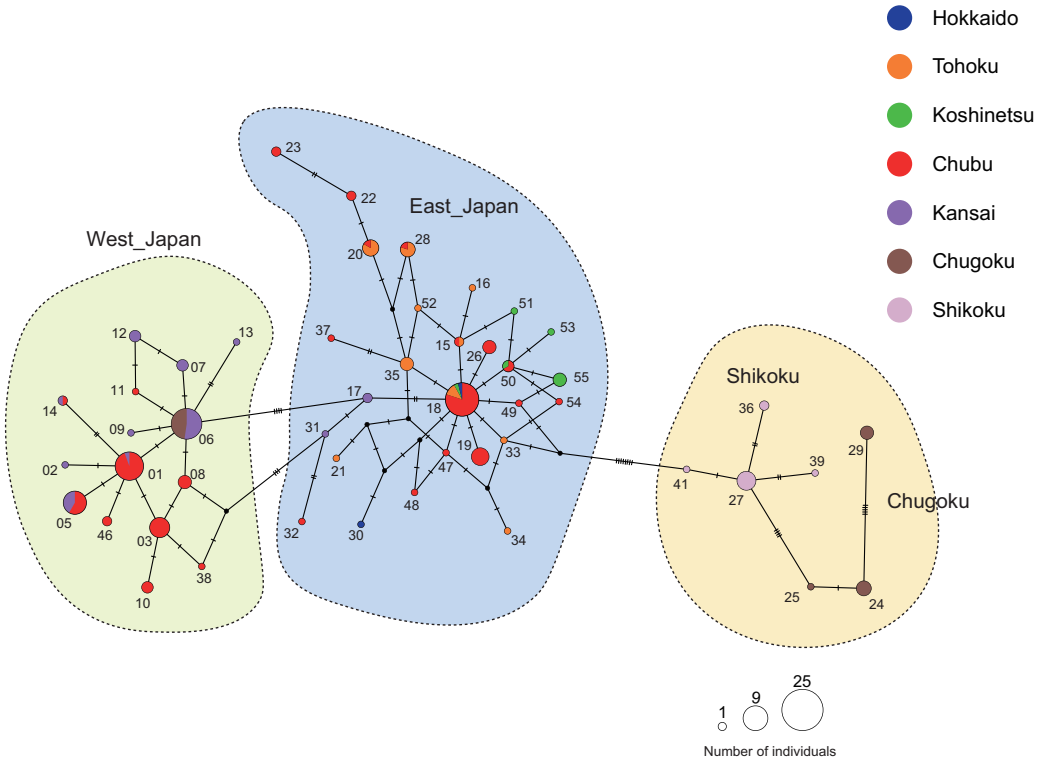
725 Fig 1. The outline of the present-day Japanese Archipelago and expected biological
726 expansion routes from the Eurasian continent, North and West routes. The regional
727 names are shown in bold, and prefecture names cited in the text were also shown.
728 Circles and squares indicate each collection site for SNP analyses in this study. Samples
729 for mtDNA were collected in Fukui and Tochigi prefectures in addition to these sites.
730
731 Fig. 2. Haplotype network depicted by TCS using the sequences of 49 haplotypes
732 detected in *Parnassius glacialis* in Japan. Each number was the haplotype number.
733
734 Fig. 3. A consensus tree of BI (Bayesian Inference), ML (Maximum Likelihood), and NJ
735 (Neighbor Joining) methods. The tree was the result of BI method. The values above
736 each branch showed the posterior probability of BI, the bootstrap value of ML, and that
737 of NJ, respectively. The values less than 0.6 of posterior probability and 60 of bootstrap
738 value were not shown in this figure. The bald letters indicated the haplotype numbers.
739
740 Fig. 4. Bayesian estimates of the divergence time obtained using BEAST ver. 2.7.6.
741 under HKY + G model, using whole mtDNA sequences of nine species and the three
742 lineages of *P. glacialis* in Japan. The values around each node indicate the estimated
743 divergence time. Horizontal bars indicate the 95% credible age interval at each node,
744 with a posterior probability 1. The representative sequences for the West Japan, East
745 Japan, and Chugoku-Shikoku lineages were obtained from the samples collected in
746 Shiga, Hokkaido, and Kochi prefectures.
747
748 Fig. 5. Population assignment using STRUCTURE from $K = 3$ to 9. Each population
749 (prefecture) and regional name (in bald letters) were shown in the bottom margin from
750 North (left) to South (right).
751
752 Fig. 6. The result of NetviewR using SNPs variations. Each line showed the connection
753 of each genotype.
754
755 Fig. 7. Population barriers estimated by (A) mtDNA, (B) SNP analyses. Our estimated
756 combined barriers were shown in (C). The dotted rectangle shows inferred contact zone
757 among the lineages.
758
759



760

761 Fig. 1.

762

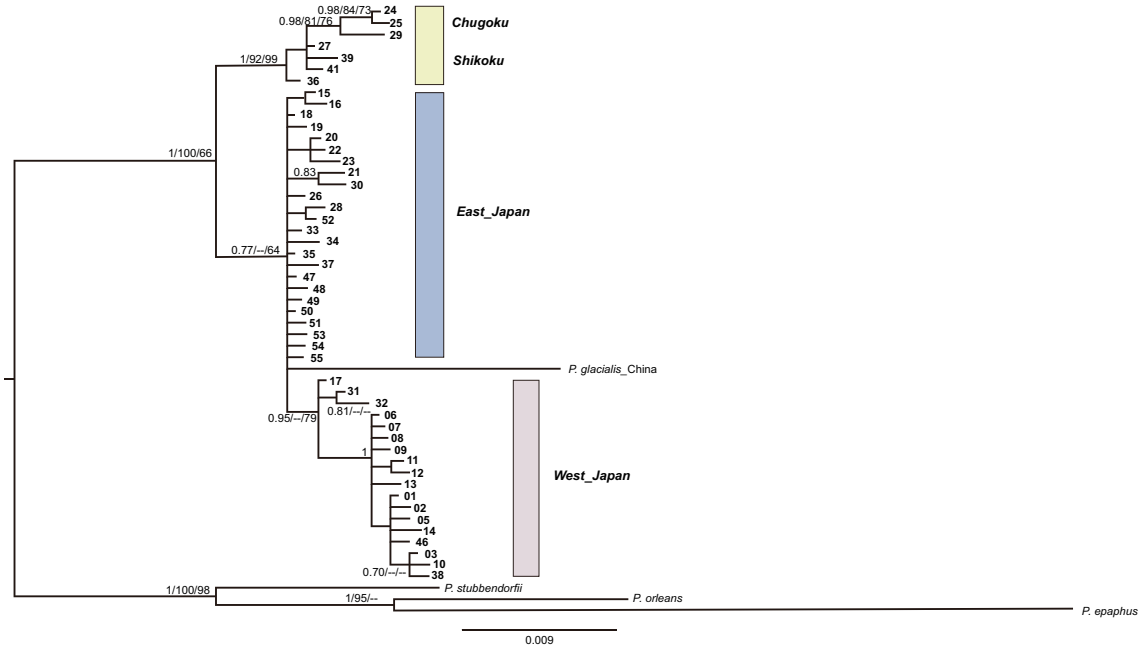


763

764 Fig. 2.

765

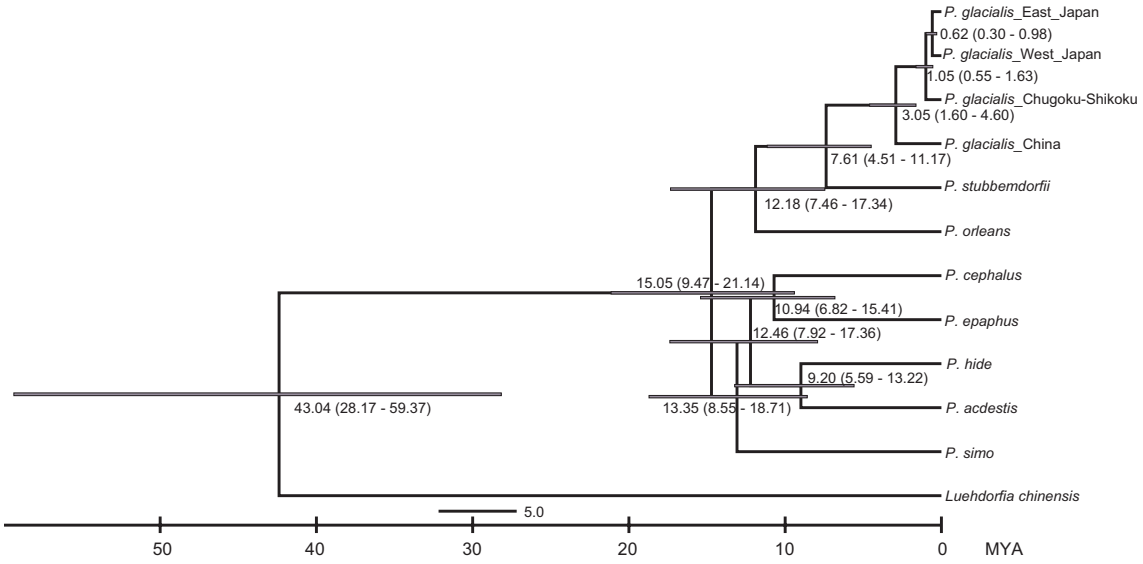
766



767

768 Fig. 3

769

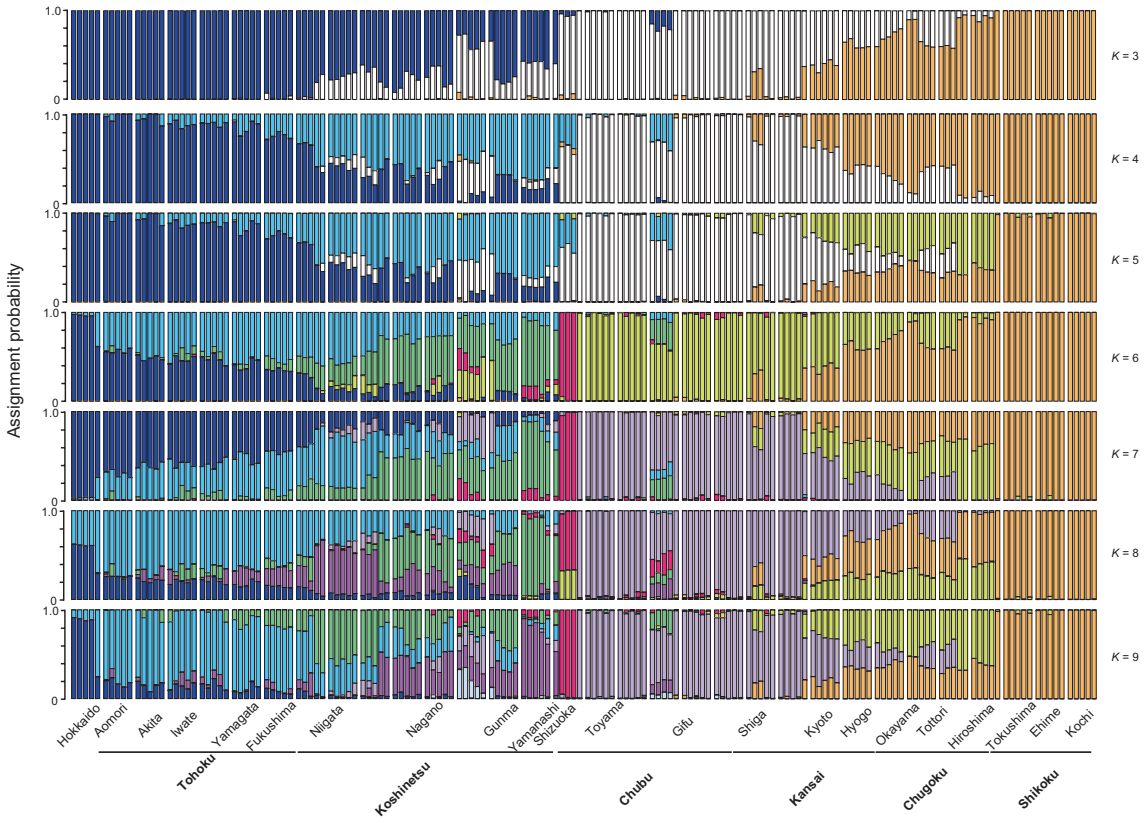


770

771 Fig. 4

772

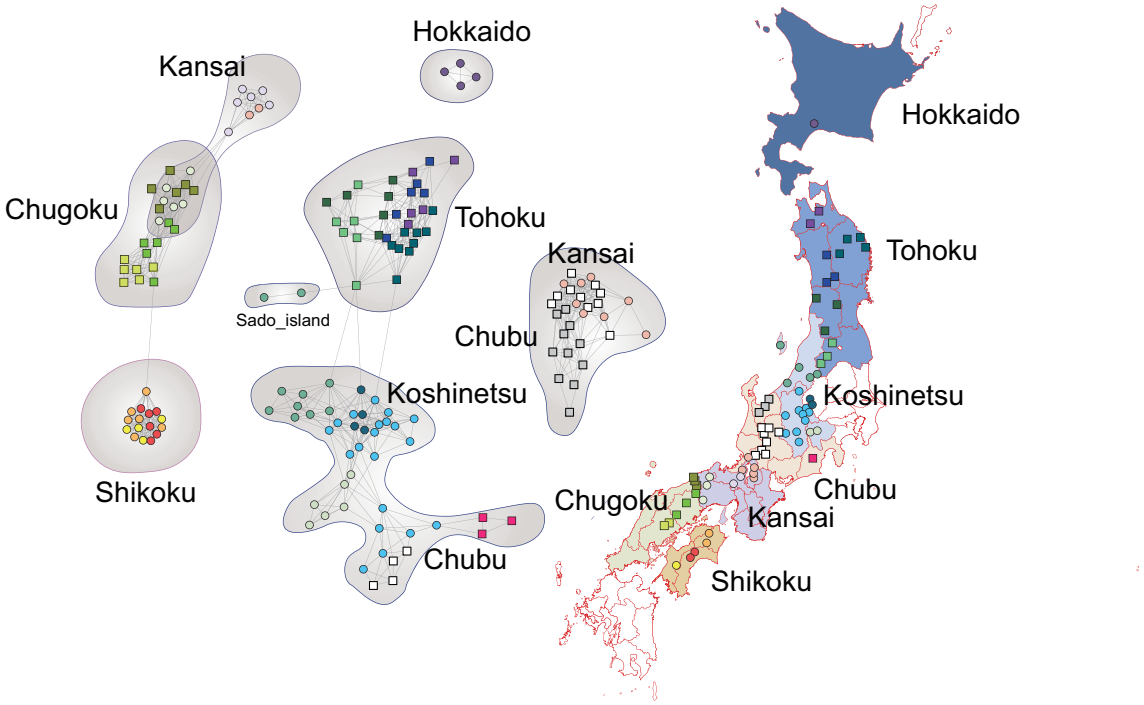
773



774

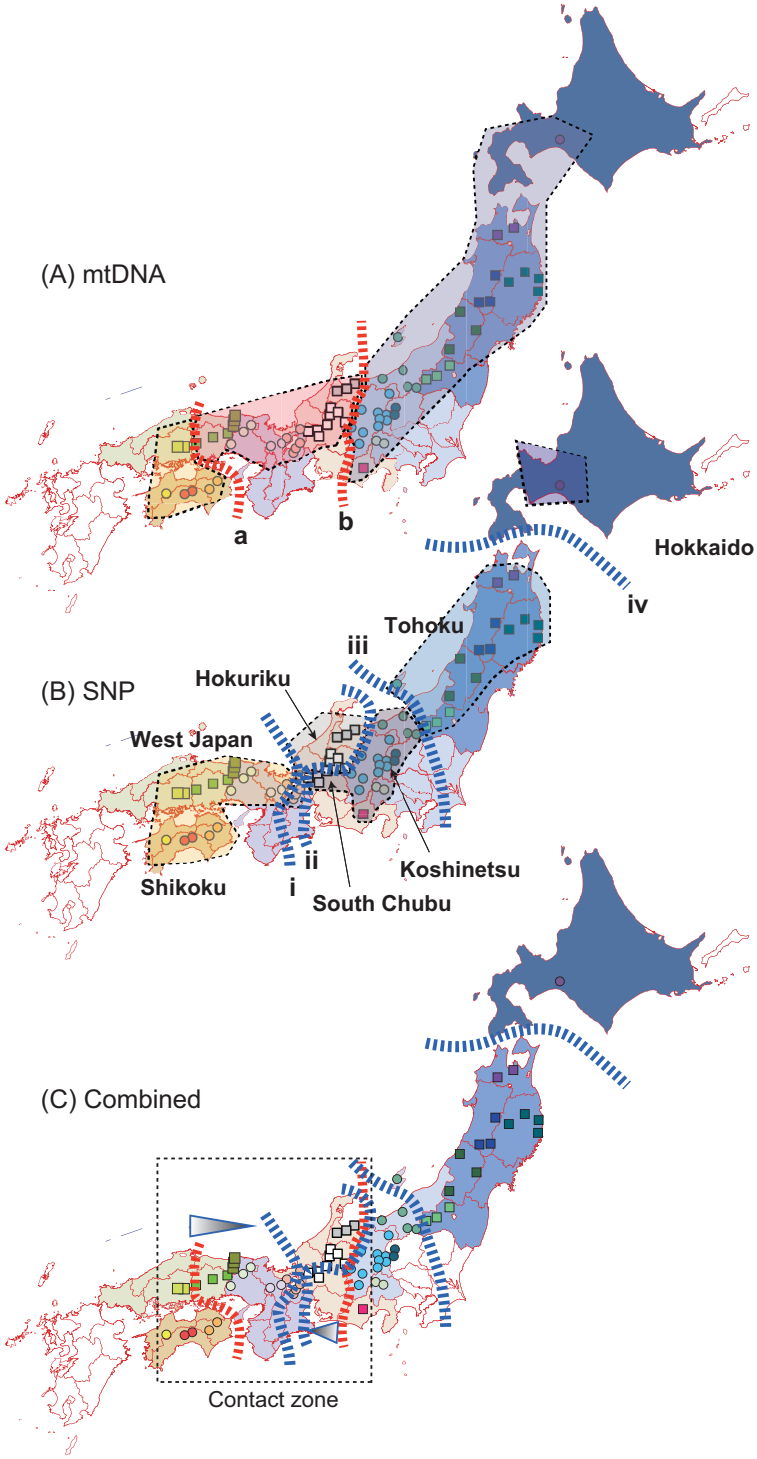
775 Fig. 5

776



777

778 Fig. 6



779

780 Fig. 7