Rapid divergence of local populations with different color forms in the dung beetle *Phelotrupes auratus* revealed by population genomics analyses

Yoshifumi Araki

Contents

Chapter 1. General introduction

1.1. Geographic variation in adaptive traits and population divergence.....4

1.2. Geographic color polymorphisms and hybrid zones.....4

1.3. Estimation of demographic history and genetic data resources.....6

1.4. Geographic color variation in *Phelotrupes auratus*.....8

Chapter 2. Population genetic structure underlying the geographic variation in beetle structural colour with multiple transition zones

2.1. Introduction.....11

2.2. Materials and Methods.....14

2.3. Results.....19

2.4. Discussions.....24

2.5. Conclusions.....29

Chapter 3. Whole-genome resequencing reveals recent divergence of geographic populations of the dung beetle *Phelotrupes auratus* with color variation

3.1. Introduction.....53

3.2. Materials and Methods.....56

3.3. Results.....60

3.4. Discussions.....63

3.5. Conclusions.....66

Chapter 4. General discussion

4.1. Geographic color variation and hybrid zones of *P. auratus*.....76

4.2. Demographic history of *P. auratus* in the Kinki District.....77

4.3. Perspective.....78

4.4. Conclusions.....80

Acknowledgements.....81

References.....82

Chapter 1. General Introduction

1.1. Geographic variation in adaptive traits and population divergence

Ecological speciation is the concept that divergent selection takes place due to changes or differences within environments, which drive reproductive isolation and population divergence, resulting in speciation (Nosil et al., 2009; Schluter, 2001). The first step of speciation is the introduction of a species into a different environment (or a major change in an existing environment), leading to development of a new population. Adaptation to novel conditions leads to the evolution of adaptive traits, and mutations are accumulated in the genes underlying those traits. However, gene flow tends to homogenize the gene pool and reduce differences among populations (Giles & Goudet, 1997). Therefore, restricted gene flow or strong natural selection for adaptive traits are necessary for population divergence (Bolnick & Fitzpatrick, 2007). When genetically diverged populations make secondary contact, the fitness of hybrid individuals can be reduced due to sterility or inviability (Shuker et al., 2005). In addition, although hybrid individuals with intermediate phenotypes may be more adapted to an intermediate environment, they can have reduced fitness in the environments of their parents (Campbell, 2004). Therefore, to better understand the early stages of population divergence, I should determine the underlying mechanisms that maintain population divergence as well as the historical backgrounds of diverged populations in species with geographic variation in adaptive traits.

1.2. Geographic color polymorphisms and hybrid zones

Among the many adaptive traits that characterize an animal, coloration has adaptive

significances (Endler & Mappes, 2017). Animal coloration often acts as a visual signal for various functions, such as predator avoidance by warning coloration or cryptic coloration as well as sexual signals for mate attraction (Protas & Patel, 2008). *Heliconius* butterflies are a textbook example of the functions of animal coloration; their wing patterns vary geographically (Hoyal Cuthill & Charleston, 2015; Mallet & Joron, 1999) and act not only as warning signals but also as signals for mate choice (Merrill et al., 2014). In particular, mimetic and cryptic coloration are involved in predator avoidance and strongly reflect the biotic and abiotic environments; thus different color morphs often occur in different geographic areas (Endler, 1978). Geographic variation in animal coloration is evidence of adaptive evolution to different environments and has been the focus of many studies (Baldassarre et al., 2014; Curran et al., 2020; Hoekstra et al., 2004; Mullen & Hoekstra, 2008; Scordato et al., 2017).

When secondary contact occurs between sister populations that have diverged in a certain adaptive trait (e.g., body color), the traits of the hybrid individuals may be intermediate, reducing their fitness (Arias et al., 2016). Therefore, geographic color variation is often strongly associated with population genetic structure, and is associated with the formation of hybrid zones (Endler, 1977). Because the maintenance of geographic color variation through hybrid zones relies on the balance of the strength of divergent selection for local adaptation and the homogenizing effect of gene flow, the cline theory was proposed to evaluate the strength needed to maintain hybrid zones (Barton & Gale, 1993; Barton & Hewitt, 1985; Gay et al., 2008; Szymura & Barton, 1986). Based on that theory, it is possible to assess the balance between selection and gene flow from genetic or phenotypic clines in hybrid zones where individuals with intermediate phenotypes and genotypes occur due to interbreeding. Assuming no regional differences

in the strength of geographic barriers, it may be possible to use adjacent local populations with multiple hybrid zones and to compare genotypic and phenotypic clines to understand the strength of natural selection on coloration. Cline analysis would provide important insights into the mechanisms of maintaining geographic variation in organisms.

1.3. Estimation of demographic history and genetic data resources

Understanding the mechanisms of biological diversification requires not only the correct understanding of the nature of the current genetic and phenotypic divergence, but also estimations of the demographic histories of diverging populations to clarify divergence times, changes in population size, and the amount of gene flow among populations from a time–space perspective. The combined interpretation of data of past geographical and evolutionary events through molecular genetic data has revealed various environmental factors that promote population divergence (Bemmels et al., 2022; Hewitt, 2000, 2004). In recent years, the impacts of human activities on the evolution of organisms have become clear. During the early to mid-Holocene, humans began to practice agriculture and pastoralism, rapidly increasing their population size and impacts on the environment (Boivin et al., 2016). Environmental changes due to human activities have affected the evolution of adaptive traits of wild organisms (Sullivan et al., 2017) and have led to the decline and extinction of wild populations (Ceballos et al., 2015). Thus, it is important to consider the possibility that anthropogenic changes in the demography of wild organisms have been influenced by not only climatic and geographic events but also human activities.

In most previous studies of the demographic histories of wild organisms, analyses have been conducted on the order of tens of thousands to millions of years (Aoki et al., 2019; Derkarabetian et al., 2016; Hirase et al., 2021; B. Wang et al., 2019), while few

have focused on more recent population divergence events such as during the period after the end of the last glacial period (Zhang et al., 2021). One reason for this is that researchers have focused mainly on species-level divergence and less on population-level divergence. Another reason is that the resolution of available genome sequence data has been limited, which makes it difficult to reveal the diversification process within a species. For example, mitochondrial genomes have long been commonly used in genetic analyses (Avise et al., 1987). However, studies that use mitogenome data underestimate the effective population size due to haploidy and do not necessarily accurately estimate phylogenetic relationships, particularly among closely related populations, because the mitogenome is only matrilineally inherited and lacks information of paternal lines (Ballard & Whitlock, 2004; Lin & Danforth, 2004). Microsatellite loci are heritable nuclear genome markers with high mutation rates, and their sequence variation may reflect recent population divergence. However, the reliability of microsatellite data depends on the number of available loci, making the use of such data costly due to the need to develop species-specific primers (Zane et al., 2002). In addition, the high variation in mutation rates among loci (Schlotterer et al., 1998) makes them difficult to use in the estimation of divergence times. Although the problems of the above markers were overcome by genome-wide SNPs obtained by some methods, such as restriction-site associated DNA sequence (RAD-seq), it is important to note that it is difficult to detect minor mutations that have occurred very recently when the number of SNPs sampled from the entire genome is limited. Recently, however, the remarkable development of high-throughput sequencing technology has enabled the acquisition of large amounts of fine-scale sequence variation data from the whole genome, which provides powerful clues to the process of population divergence (Ellegren, 2014). In addition, the development of

methods for analyzing large amounts of genomic data has made it possible to reveal fine-scaled demographic histories (Excoffier et al., 2013; Li & Durbin, 2011; Schiffels & Wang, 2020; Terhorst et al., 2017; Zhou et al., 2020). The estimation of demographic histories using genome-wide SNPs obtained by whole-genome resequencing will be a powerful tool for revealing recent population divergences and will play an important role in elucidating the mechanisms that drive population divergence.

1.4. Geographic color variation in *Phelotrupes auratus*

This study focused on geographic color variation in the geotrupid dung beetle, *Phelotrupes auratus* (Motschulsky). *P. auratus* is a diurnal beetle in mountain areas that feeds on the dung of large wild mammals (Tsukamoto et al., 2014). Sika deer are considered the most important wild animal supplying food for *P. auratus* because their distribution range overlaps that of *P. auratus* (Toda & Akei, 2003). *P. auratus* exhibits notable variation in its structural metallic body coloration. Their elytral coloration is provided by a simple multilayer reflector structure, and the coloration becomes closer to reddish as the thickness of the cuticular layers increases (Akamine, Ishikawa, et al., 2011). Their elytral coloration is categorized into red, green, and indigo based on the peak wavelength on the elytral reflectance spectrum (Akamine et al., 2008). In the Kinki District of Honshu, the main island of Japan, these three color forms occur separately in different geographic areas; the green form occurs in the southern part of Lake Biwa, the indigo form is in the Kii Peninsula, and the red form is in the other areas (Akamine, Maekawa, et al., 2011; Mizuno, 1964). Although the adaptive significance of the coloration is unclear, Müllerian mimicry for predation avoidance has been suggested, because a sympatric congener, *Phelotrupes laevistriatus*, shows a similar geographic

coloration pattern to *P. auratus* (T. Watanabe et al., 2002). Geographic variation in the color forms may have resulted from geographic or genetic isolation mechanisms or local adaptation in different populations. Although a previous population genetics study based on the mitochondrial cytochrome c oxidase I (COI) gene sequence revealed phylogenetic relationships among populations across the Japanese Archipelago, haplotype groups are not clearly associated with color forms (Akamine, Maekawa, et al., 2011). Mitochondrial data reflect only limited aspects of the demographic history and the process of geographic division of the color forms; thus, a more detailed study using genome-wide sequence data is necessary to reveal the population structure underlying the geographic color variation. In addition, genomic analyses are needed to study the genetic basis of coloration. Although the genetic basis of structural coloration is poorly understood in other organisms, coloration is thought to be controlled by multiple loci (Brien et al., 2018; Thayer et al., 2020; L. Zhang et al., 2017), and this may also be the case in *P. auratus* (Akamine et al., 2008). Genome-wide detection of mutations that are significantly correlated with coloration may provide insight into the genetic basis of the control of structural color.

In this thesis, the divergence processes of local populations of *P. auratus* in the Kinki District, Honshu, Japan, which exhibit remarkable body color variation, is discussed. First, the population genetic structure among local populations was estimated to assess genetic divergence among color forms, and natural selection for coloration was assessed by cline analysis using RAD-seq data. In addition, the genetic basis for color regulation was investigated via GWAS (Chapter 2). Second, the whole genome sequence of *P. auratus* was assembled and whole-genome resequencing data were obtained, which provide more detailed data on the demographic histories of the populations, to estimate

coalescence times and gene flow between local populations (Chapter 3). Finally, all of the results were integrated to discuss the evolutionary processes and causes of intraspecific diversity in *P. auratus* (Chapter 4).

Chapter 2. Population genetic structure underlying the geographic variation in beetle structural color with multiple transition zones

2.1. Introduction

To understand the processes of adaptive population divergence that may result in speciation, it is important to study geographic variation in major adaptive traits (Endler, 1977). In general, geographic phenotypic changes are associated with hybrid zones of once-diverged populations or changes in optimal phenotypes among geographic regions (Endler, 1977). Transition zones of adaptive traits, including hybrid zones, have played a major role in understanding how phenotypic divergence among geographic populations has been formed and maintained by the balance between gene flow and selection, and cline analyses provide a powerful framework to analyze various aspects of hybrid zones (Barton & Hewitt, 1985). In cline analyses, we can infer the nature and strength of barriers to gene flow between populations by fitting cline models to genomic and quantified trait data (Barton & Gale, 1993; Barton & Hewitt, 1985; Gay et al., 2008; Szymura & Barton, 1986). Recently, various methods to analyze genomic data have been developed and used to understand the genetic population structure and gene flow involved in hybrid zones (Gompert et al., 2017).

Animal coloration is a typical adaptive trait that has various functions, and it is readily accessible for studies on adaptation (Endler & Mappes, 2017). Animal coloration sometimes differs among geographic regions of a species range that is more or less continuous and shows clinal changes at the boundary regions (Roulin, 2004). Thus, cline analyses are applicable to quantified color values showing geographic variation. In recent analyses of geographic color variation, analyses of genome-wide markers have played an

increasingly important role in characterizing the genetic population structure and phenotypic and genetic clines, which reflect the gene flow, selection, and demographic history underlying the color variation (Baldassarre et al., 2014; Brelsford et al., 2011; Curran et al., 2020; Hoekstra et al., 2004; Mullen & Hoekstra, 2008; Scordato et al., 2017). Furthermore, the strength of selection on color can be studied by comparing the clines for color-associated loci and neutral genetic markers (Mullen & Hoekstra, 2008). In addition, genome-based approaches can be a powerful tool to investigate the genetic basis of coloration traits (Knief et al., 2019; San-Jose & Roulin, 2017; Scordato et al., 2017).

In this study, I focused on intraspecific variation in the structural color of the geotrupid dung beetle, *Phelotrupes auratus* (Motschulsky), which exhibits metallic body colors perceived as red, green, and indigo by human eyes and as different wavelength ranges of reflectance spectra (Akamine, Ishikawa, et al., 2011). This species has a wide distribution range in Japan, and the distribution pattern of different colors is not related to latitude; hence, thermoregulation is unlikely a factor in color variation (Mizuno, 1964). Although the adaptive significance of the coloration is still unclear, Müllerian mimicry has been suggested, because a sympatric congener showed parallel color variation with *P. auratus* (T. Watanabe et al., 2002). In the Kinki District, Honshu, Japan, all of these color forms occur parapatrically in separate areas (Akamine et al., 2008; Mizuno, 1964; T. Watanabe et al., 2002), which implies the presence of an underlying population genetic structure associated with the distribution of the color form. The geographic partitioning of the color forms may stem from geographic or genetic isolation mechanisms or local adaptation in different populations. However, a previous population genetics study based on the mitochondrial cytochrome c oxidase I (COI) gene sequence revealed that different haplotype groups are not clearly associated with different color forms (Akamine,

Maekawa, et al., 2011). However, mitochondrial data reflect only limited aspects of demographic history and the process of geographic division of the color forms, thus a more detailed study using genome-wide sequence data is necessary to reveal the population structure underlying the geographic color variation. A genomic approach will also allow us to study the genetic basis of color. The coloration of *P. auratus* is considered a polygenic quantitative trait (Akamine et al., 2008) and may be controlled by multiple loci (Brien et al., 2018; Thayer et al., 2020; L. Zhang et al., 2017).

In this study, wI used restriction-site associated DNA (RAD) markers (Baird et al., 2008) and mitochondrial COI sequences to investigate the association between genetic composition and body color across a geographic range of *P. auratus*. I measured the reflectance spectrum to define elytral color form and clarified the distributions of different forms and their boundary (transition) zones. After confirming that the color variation was not related to abiotic environmental (climatic) variables, I analyzed the genetic population structure underlying the geographic division of the color forms to reveal the concordance/discordance between the genetic and color transitions. Then, I performed cline analyses to clarify whether the geographic division of the color forms was maintained primarily by selection on the color per se, or by intrinsic or extrinsic barriers to gene flow. In addition, I performed a genome-wide association study (GWAS) to identify single-nucleotide polymorphisms (SNPs) associated with color and examined how changes in the allelic frequencies of these SNPs were related to the transition of color forms. My study revealed heterogeneity among transition zones of the color forms, which may have resulted from interactions between varying levels of barriers to gene flow and selection for different colors possibly controlled by multiple genetic loci.

2.2. Materials and Methods

2.2.1. Study species and sampling

P. auratus is a dung beetle found on Japanese islands and the East Asian mainland (China, Korea, and Russian Far East). The beetles are active flyers and search for fresh dung of wild animals from April to October. During summer, a female buries a dung ball into the soil and lays a single egg into the dung. Larvae grow feeding on the provisioned dung and emerge as adult in the next year (Tateno et al., 2020). In Japan, the body color of *P. auratus* is usually red, while green populations occur sporadically. Indigo populations are restricted to two areas, the southern Kinki District of Honshu (the present study area) and Yaku Island in the southwestern-most part of the species range in Japan (Tsukamoto et al., 2014).

Between 2015 and 2019, I collected 179 *P. auratus* adults from 24 of 26 sites around Lake Biwa in the Kinki District, Honshu, Japan (P1–P24; Figure 2.1) using traps baited with horse dung. I did not find any evidence of the occurrence of *P. auratus* at sites P25 and P26 (Figure 2.1), suggesting discontinuous distribution in the north of Lake Biwa. Collected beetles were fixed with 99% ethanol and stored at –30°C until DNA extraction. After DNA extraction, all specimens were dried and used to measure elytral coloration.

2.2.2. Quantification of elytral color

I quantified the elytral color of individuals by measuring the reflection spectra of the surface of the center of the right elytra of dry specimens in a small dark room using a spectrometer (USB2000+UV-VIS-ES; Ocean Optics, USA). The value of λmax(α), the peak wavelength of the visible light region (α peak), was determined by averaging three measurements for each individual. Based on the frequency distribution pattern of λmax(α)

between 450 nm and 680 nm with three modes corresponding to red, green, and indigo color by human eyes as observed by Akamine *et al.* (2008) and ourselves, I defined the color form of each population as "red" if the median of λmax(α) of the population was 600 nm or more, "green" if it was 525–600 nm, and "indigo" if it was ≤525 nm.

2.2.3. Relationships between elytral color and climatic variables

Prior to conducting molecular genetic analyses, I first explored the possibility that the elytral color was associated with climatic factors. I obtained annual rainfall, mean temperature, and sunlight hours for each sampling site from the website of the Geospatial Information Authority of Japan (https://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-G02.html; Table S2.1). I performed multiple regression analyses for the effects of these climatic variables and latitude on the median λmax(α) scores at each site.

2.2.4. DNA extraction and sequencing

Total genomic DNA was extracted from the leg muscle using the DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany) following the manufacturer's protocol. I performed RAD sequencing for the 179 individuals following the library construction protocol of Etter, Bassham, Hohenlohe, Johnson, & Cresko (2011). The genomic DNA samples were digested with the restriction enzyme *Pst*I and barcoded with a unique five-nucleotide sequence in a P1 adapter. After ligation of the P1 adaptor at Kyoto University, the samples were sent to Hokkaido System Science Co. Ltd. (Sapporo, Japan) for library construction and sequencing in three runs on an Illumina HiSeq 2500 sequencer (Illumina, USA) with 101-base, single-end reads.

I also obtained 745 bp mitochondrial COI gene sequence data for the 179

individuals. PCR amplification was performed using primers C1-J-2195 (5'-TTG ATT TTT TGG TCA TCC AGA AGT-3') and TL2-N-3014 (5'-TCC AAT GCA CTA ATC TGC CAT ATT A-3') (Simon et al., 1994). The PCR program was as follows: initial denaturation at 94°C for 5 min; 25 cycles at 94°C for 30 s, 56°C for 30 s, and 72°C for 30 s; and a final extension at 72°C for 7 min. The PCR products were sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA), and sequencing was performed using the 3130xl Genetic Analyzer (Applied Biosystems).

2.2.5. Population structure analyses

I obtained a total of 532 million RAD sequences from 179 individuals (~3 million reads per individual). The sequence reads were processed by Stacks version 2.41 (Rochette et al., 2019) using a Phred score of 10 for read filtering in the *process_radtag* function, a minimum coverage depth of 2 to create a stack, a maximum of two different nucleotides to merge two different stacks (-m 2 and -M 2 in the *ustacks* function), and a maximum of one mismatch was allowed between sample loci (-n 1 in the *cstacks* function). An SNP data matrix (the mean coverage per SNP per individual was 10.14) was constructed for the RAD loci shared by 23 of 24 populations and identified in >50% of individuals within a population (-p 23 and -r 0.50 in the *populations* function).

To determine genetic differences among populations, I performed population structure analyses for the RAD SNPs using the *snmf* function of the LEA package in R (Frichot & François, 2015). Three iterations were performed for $K = 1$ to 5, and the best value for K was estimated from the score of the cross-entropy criterion. I also performed principal component analyses (PCAs) based on the SNPs using the pcaMethods package

in R (Stacklies et al., 2007) to assess differences in the genetic composition of individuals among the geographic populations. For individuals of putative hybrid populations (P8 and P23), I calculated heterozygosity with regard to SNPs showing different allele frequencies between the two parental populations adjacent to the hybrid populations (P7 and P9 for P8; P22 and P24 for P23) using the introgress package in R (Gompert & Buerkle, 2010). A hybrid index (Buerkle, 2005) was calculated simultaneously. High and low values of heterozygosity indicate a recent hybridization and a long-lasting hybridization, respectively (Capblancq et al., 2020; Fitzpatrick, 2012).

The COI sequences were aligned using MEGA X (Kumar et al., 2018). I constructed a median-joining network (Bandelt, Forster, & Röhl, 1999) of COI haplotypes using PopART version 1.7 (Leigh & Bryant, 2015). Population clustering was performed using SAMOVA 2.0 (Dupanloup et al., 2002) to reveal the population structure based on the COI sequences. I compared the results when the number of groups varied from two to four, and selected the best result with the highest F_{CT} score, representing the fixation index among the groups.

2.2.6. Estimation of phenotypic and geographic clines between populations

I performed analyses of the hybrid zone using the HZAR package in R (Derryberry et al., 2014) to understand the differences between the genetic and phenotypic (color) clines at the three transition zones of the color morphs (i.e., between the west/red and the south/green, the south/green and the south/indigo, and the south/green and the east/red). I used the Q-value of the most frequent ancestry at the best K obtained in population structure analyses and the normalized individual $\lambda\max(\alpha)$ scores for this estimation. Because all transition zones included P9, I used the cumulative distance between adjacent

populations starting from P9 for the geographic distance axis, assuming that migration occurred only between adjacent populations. I compared 16 cline shape models, including a null model (no clinal change), in a Metropolis-Hastings MCMC run of 100,000 generations following and initial burn-in of 10,000 generations, and the model with the lowest corrected AIC (AICc) was chosen as the best fit. The differences in the estimated center and width between genetic and color clines at each transition zone were evaluated using the 95% confidence intervals of these parameters.

2.2.7. Estimation of the demographic history

To reveal the demographic history of the groups classified by the genetic clusters and color forms, I performed a neutrality test with Tajima's *D* (Tajima, 1989a, 1989b) for SNP data for each population group using the pegas package in R (Paradis, 2010). I tested if the mean *D* value significantly differed from zero assuming a beta distribution as implemented in the *tajima.test* function in pegas. I also obtained *R2* statistics (Ramos-Onsins & Rozas, 2002) for neutrality tests against population growth from the data matrix of SNPs using the *R2.test* function in pegas. The data matrix was divided by each population group and processed using the vcfR package in R (Knaus & Grünwald, 2017). In addition, I also estimated Tajima's *D* and Fu's F_S (Fu, 1997) from the mitochondrial COI sequence data using Arlequin 3.5.2.2 (Excoffier & Lischer, 2010). Negative values of Tajima's *D* suggest population bottleneck followed by population expansion (Ramos-Onsins & Rozas, 2002; Tajima, 1989a). Negative values of Fu's F_S and positive *R2* values suggest population expansion (Ramos-Onsins & Rozas, 2002).

2.2.8. GWAS

I performed a GWAS to identify SNPs in the RAD sequences that were significantly associated with the peak wavelength, λmax(α), using the rrBLUP package in R (Endelman, 2011). The analyses were performed using the λmax(α) scores of individuals and SNPs with a minor allele frequency > 0.05, and including PC1 scores resulting from the PCA as cofactors to consider the population structure. I adjusted the significance level of the results according to false discovery rate (FDR; < 0.05). The allelic frequency distribution for significant SNPs was obtained by population groups. Finally, I calculated the linkage disequilibrium (LD) between SNPs for each genetic cluster and evaluated the strength of the association between SNPs using the genetics package in R (Warnes, Gorjanc, Leisch, & Man, 2019).

To explore whether SNPs relating to color were associated with specific gene regions, I performed a BLASTX search (Gish & States, 1993; https://blast.ncbi.nlm.nih.gov/Blast.cgi) against the GenBank non-redundant protein sequence (nr) database using the sequences of RAD loci that contained these SNPs.

2.3. Results

2.3.1. Geographic distribution of color forms

I obtained the elytral reflection spectra data for all of the individuals collected from 24 sites and classified the populations into the three color forms based on the average peak wavelength, λmax(α) (Figure 2.2a, Table 2.1). The populations with the red elytral color were found west of Lake Biwa (populations P1–P7) and separately to the east of Lake Biwa (populations P20–P24). The green populations were found in the southeast, south of Lake Biwa (populations P8–P11, and P19), and the indigo populations were found to the south of the green populations (populations P13–P18) (Figure 2.1). Population P8

exhibited an exceptionally wide λmax(α) value range, although the average was slightly lower than 600 nm, and it was therefore not appropriate to be assigned to any color form. The coloration distribution (Figure 2.2b) showed many individuals with λmax(α) values around 600 nm, indicating a continuous transition between the red and green forms, whereas there were few individuals with λmax(α) values from 530–540 nm, indicating a discontinuous transition from the green to indigo forms.

Although the elytral color was apparently related to the latitude in the study area (latitude, $t_{1,22} = 9.287$, $P < 0.001$), climatic variables did not show significant relationships to the latitude (annual rainfall, $t_{1,22} = 1.188$, $P = 0.247$; annual mean temperature, $t_{1,22} = –1.430$, $P = 0.167$; annual sunlight hours, $t_{1,22} = –0.512$, $P = 0.614$). The multiple regression of climatic variables on elytral color did not show any significant relationships between climatic variables and elytral color (annual rainfall, $t_{1,20} = 0.569$, $P = 0.576$; annual mean temperature, $t_{1,20} = –0.762$, $P = 0.455$; annual sunlight hours, $t = –0.291$, $P = 0.774$).

2.3.2. Geographic population genetic structure

I obtained the RAD sequence data for 179 individuals from 24 areas and constructed a data matrix of 16,739 RAD loci containing 129,106 SNPs. First, I performed population structure analyses using the sNMF algorithm. Based on the value of the cross-entropy criterion, the optimal number of ancestral clusters was three (Figure S2.1), comprising the southern, western, and eastern clusters (mean Q-value = 0.603, 0.305, and 0.092, respectively) (Figure 2.3). There were no genetic differences among the green, indigo, and red forms that occurred to the southeast of Lake Biwa. Population clustering analyses divided the *P. auratus* populations in the study area into five groups: the west/red group

(populations P1–P7), the east/red group (population P24), the south/red group (populations P20–P22), the south/green group (populations P9–P11 and P19), and the south/indigo group (populations P13–P18). The P8 and P23 populations showed a mixture of two genetic clusters and were treated separately as hybrid populations (Figure 2.3).

The PCA showed that the first two principal component scores (PC1, PC2) clearly explained the variation in the population structure of *P. auratus* (proportion of variance explained [PVE]: PC1, 16.27%; PC2, 5.52%). PC3 (PVE = 1.14%), PC4 (PVE = 1.23%), and others were excluded from subsequent analyses because their PVE values were sufficiently small. The PC1 score revealed the distinction between the west/red population and the other populations, and the PC2 score indicated the distinction between the east/red and south/red populations. In the PC1-PC2 plot, three major clusters were recognized, corresponding to the west/red, the east/red (P24 individuals), a combined south/green, the south/indigo, and the south/red groups (Figure 2.4). The individuals from P8, a hybrid population, were distributed between the west/red and the south clusters, and those from another hybrid population at P23 were distributed between the east/red and south clusters. Individuals in P8 and P23 had intermediate hybrid index values with respect to adjacent red and green populations as expected but exhibited low levels of heterozygosity, indicating that these populations were not of recent origin (Figure 2.5).

I identified 110 haplotypes of the COI sequences among 179 individuals from 24 sites. The haplotype network showed that most haplotypes from the west/red populations were segregated from the other haplotypes (Figure 2.6). Consistent with the haplotype network, SAMOVA analyses divided populations of *P. auratus* into two geographic groups; the northwestern group containing P1–P3 and P5 (populations in the west/red

group), and the other contained all of the remaining populations (F_{CT} = 0.431, P < 0.001, Table S2.2).

2.3.3. Phenotypic and genetic clines

I used the Q-value of the south cluster at K = 3, obtained from the genetic population structure analyses, to analyze the structure of the genetic clines between populations of the different color forms (Figure 2.7, Table 2.2 and Table S2.3). In the west/red–south/green transition zone (Figure 2.7a), the 95% confidence intervals of cline center and width overlapped between genetic and color clines, indicating a concordance between the genetic and color clines (Table 2.2). On the other hand, the genetic cline between the south/green and south/indigo groups was flat (i.e., not clinal) despite the distinct color cline (Figure 2.7b, Table 2.2). Between the south/green and the east/red via the south/red, the steep genetic cline occurred only around P22, while the color cline was gentle and wide (Figure 2.7c, Table 2.2); both genetic and color clines differed in the center position (Table 2.2).

2.3.4. Demographic history

The Tajima's D scores estimated from the SNP data were significantly negative in each population group (Table 2.3), suggesting a recent population expansion following a bottleneck. The *R2* statistics also showed a significant departure from neutrality in each population group (Table 2.3), which implies population growth.

The Tajima's D scores estimated from the mitochondrial COI gene sequences were significantly negative in the south/green, south/indigo, and south/red groups, and were also negative but not significant in the west/red and east/red groups (Table 2.4). Fu's F_S

values were all significantly negative (Table 2.4). Thus, the mitochondrial data also implied that all of the populations experienced a recent population expansion.

2.3.5. SNPs associated with elytral coloration

Four SNPs (FDR < 0.05) were significantly associated with peak wavelengths, λmax(α) (Figure 2.8). SNP35526_34 contained an allele that was dominantly associated with a small λmax(α) (<525 nm), indicating indigo coloration; this allele occurred only in the south/indigo populations (Figure 2.8a). In the south/green–south/indigo transition zone, the allele frequency of this SNP showed a stepped cline concordant with the color cline but discordant with the unchanged overall genetic structure (Figure 2.7b, Table S2.3). Both alleles of SNP108387_29 showed an association with a large λmax(α) (>600 nm; red coloration) (Figure 2.8b); the allele frequency cline was concordant with the color cline but discordant with the genetic cline in both the west/red–south/green and east/red–south/green transition zones (Figure 2.7a, c, Table S2.3). SNP76613_62 and SNP202132_6 contained alleles associated with a large λmax(α) (>600 nm), indicating red coloration; these alleles occurred predominantly in the populations of the west/red and east/red groups (Figure 2.8c, d). In the west/red–south/green and east/red–south/green transition zones, the allele frequency clines of these SNPs were concordant with both the genetic and color clines, except for a discordance in the center of SNP202132_62 with the genetic cline center in the east/red–south/green transition zone (Figure 2.7a, c, Table S2.3). There was no significant LD between the four SNPs (Table S2.4).

The translated sequence of the RAD locus containing SNP35526_34 best matched a thioredoxin of the rhinoceros beetle *Oryctes borbonicus* (GenBank accession

no. KRT78961.1; query cover, 96%; identity, 70.97%; E = 3×10^{-6}). The SNPs appeared to result in synonymous substitution. I detected no matched coding sequences for RAD sequences containing the other SNPs.

2.4. Discussion

2.4.1. Population genetic structure underlying geographic color variation

Genetic population structure analyses using RAD sequence data divided the populations in my study area into three genetic clusters: west, east, and south (Figures 3, 4). However, population clustering using COI sequences resulted in two groups: the west cluster and all other groups (Table S2.2). These results imply deep divergence between the west cluster and the other populations, and shallow divergence between the east and south clusters. The genetic population structure results were somewhat inconsistent with the geographic distribution of the color forms, as I identified no appreciable changes in the genetic composition across the range of the green and indigo forms, despite there being a clear transition between these color forms (Figure 2.2). This pattern may indicate that while different colors are selected on different sides of the green/indigo boundary, gene flow has not been strongly restricted. In animal color transition zones, clines of color phenotypes under natural or sexual selection often show different patterns from those of neutral genetic markers, including instances in which the color cline is evident despite the absence of a clear cline for the neutral genetic markers (Baldassarre et al., 2014; Brelsford et al., 2011; Curran et al., 2020; Hoekstra et al., 2004; Scordato et al., 2017; Stankowski et al., 2017).

Meanwhile, I identified clinal changes in the genetic composition corresponding to the color clines in the two transition zones between the red and green forms, which

implies that the red/green transition zones were hybrid zones of previously diverged populations. The genetic and color clines were consistent with each other in the west transition zone but not in the east transition zone (Figure 2.7a, c, Table S2.3). It is difficult to judge whether color divergence in the west transition zone was due to recent contact, limited gene flow by endogenous and/or exogeneous factors, or selection for different colors solely based on the comparison between clines. However, individuals of the hybrid populations (P8, P23) exhibited low heterozygosity, which implies that the hybrid zones were not of recent origin (Fitzpatrick, 2012); the divergence in the west transition zone may have been maintained by both limited gene flow and selection for different colors. On the other hand, in the east transition zone, the genetic divergence was shallow, and hence the discordance between the genetic and color clines may have been driven by selection for different colors (Baldassarre et al., 2014).

2.4.2. Genetic background of structural color variation

The epicuticle of *P. auratus* elytra is formed by alternative stacking of two types of layers with electron-dense and electron-lucent materials, and the elytral coloration is a structural color based on multilayer reflectors, depending on the density of the layered structure (Akamine, Ishikawa, et al., 2011). The color variation among individuals within populations in terms of the peak wavelength $\lambda max(\alpha)$ indicate that coloration is under the control of quantitative trait loci affecting the layered structure, such as the thickness of the different layers, and ultimately, the peak wavelength (Akamine, Ishikawa, et al., 2011). Note that there seems to be little environmental variation in the coloration of *P. auratus* as I have not noticed temporal variation in color at the same localities and the present study clarified that color differences existed despite the similar climatic conditions in the

study area.

My GWAS implies that coloration (peak wavelength) is controlled by multiple, unlinked loci, including one locus with major effects on the indigo color and two loci with major effects on the red color (Figure 2.8). Thus, my study revealed a novel aspect of beetle structural color variation that has been explained by single-locus controls based on crossing experiments of color phenotype categorizations by the human eye (Favila et al., 2000; Paarmann et al., 2007). Although the genetic basis of structural coloration in the elytra of beetles is largely unknown, substantial information has been obtained for the structural color in wing scales of butterflies. Brien *et al.* (2018) found that iridescent structural color of *Heliconius erato* is a quantitative trait controlled by multiple genes. In some butterflies, the *optix* patterning gene controls coloration including iridescent structural color by controlling the thickness of lamina of wing scales (Thayer et al., 2020; L. Zhang et al., 2017).

The putative locus for the indigo color had a dominant allele that led to the expression of the indigo form, and this allele was not present in the red and green forms. Together with the discontinuous occurrence of the peak wavelength at ~525 nm (i.e., the boundary between the indigo and green colors; Figure 2.2b), it is likely that the locus for the indigo color controls the expression of a small $\lambda max(\alpha)$ independent from other loci that control the expression of the red/green colors (i.e., relatively large $\lambda max(\alpha)$ values). The profile of allelic frequency across the color transition zones showed that the indigo allele occurred only in the south part of my study area, and the dominance effect may have contributed to the maintenance of indigo color in the south area despite the presence of alleles for red/green colors at other loci. As in the textbook example of the industrial melanism controlled by the *carbonaria* allele in *Biston betularia* (Cook & Saccheri, 2013),

a dominant allele would enable a rapid response to selection while allowing the population to retain the recessive allele and undertake a reverse evolution when the direction of selection has altered.

In this study, I used RAD sequence data to detect SNPs indirectly linked to loci responsible for specific color phenotypes. Of the four SNPs associated with color, I found that the SNP for indigo/green was located in a thioredoxin coding region. Thioredoxin is involved in many biological processes, including melanin synthesis (Castro-Sowinski et al., 2007), and it may also be involved in controlling the cuticular structure that produces the iridescent colors of butterfly wings (L. Zhang et al., 2017). Therefore, allelic differences in this gene may affect structural coloration in beetles. However, because the SNPs detected in the putative thioredoxin gene sequence were synonymous, the variation responsible for this coloration may coexist in other parts of this gene region. Future studies should analyze whole-genome sequence data of *P. auratus* to thoroughly explore the sequence variation responsible for differences in coloration.

2.4.3. Processes and factors of geographic color variation

My results imply that the different color forms may have been maintained in different regions due to selection for different colors, although it is unclear what is the selective factor for body color. I cannot completely rule out the possibility that the color difference is selectively neutral and the apparently discontinuous distribution of color forms has resulted from the random fixation of particular color forms during bottleneck events, as suggested by the negative Tajima's *D* values (Tables 2.3, 2.4); if so, obscure genetic differences may be attributed to very recent population divergence. On the other hand,

the loss of genetic difference can result from frequent gene flow; in this case, however, it would be difficult for color differences to be maintained under frequent gene flow, particularly between the green and indigo forms.

Because structural coloration can have a function in thermoregulation (Biró et al., 2003; Kroiss et al., 2009), geographic variation in structural coloration may be related to local adaptation of thermoregulation (Bosi et al., 2008; Davis et al., 2008). In *P. auratus*, however, I found no significant relationships between elytral color and climatic variables. Thus, I could rule out local climatic adaptation as a selection factor (see also Mizuno, 1964). Beetle body color may be related to predation avoidance and mate recognition. However, assortative mating for the color form is unlikely as male and female beetles copulate around the dung to which they are attracted. The brilliant metallic color of *P. auratus* may be related to aposematism. In a previous study of color variation in *P. auratus*, the Müllerian mimicry hypothesis was proposed based on the similarity of geographic color variation patterns with a sympatric congener, *Phelotrupes laevistriatus* (T. Watanabe et al., 2002), although it is unknown whether these beetles are unpalatable to bird and mammal predators. Probable cases of Müllerian and Batesian mimicry have been reported in African scarabaeid dung beetles with bright metallic colors (Cambefort, 2014). The glossy or iridescent color of beetles may promote predator learning of unpalatable prey (Waldron et al., 2017) or may function as camouflage against brilliant backgrounds and be effective for avoiding predation (Kjernsmo et al., 2020). Clarification is needed about which type of selection acts on the different color forms in *P. auratus* to understand the reason for the geographic differences in the color forms.

2.5. Conclusions

I investigated the genetic population structure of three transition zones between three color forms of the geotrupid dung beetle *P. auratus* within an area, and found that color clines existed even when gene flow was not restricted as a whole. My results imply that the geographic divergence of coloration is maintained by geographically specific natural selection. I also found SNPs associated with different color forms and inferred that a transition between different combinations of colors was controlled by different genetic loci. Thus, my study demonstrates that geographic color variation with three color forms can be maintained by nonuniform interactions among barriers to gene flow, locally specific selection on different colors, and the effects of different color loci. Future studies are needed to identify the genes that control color and to estimate the strength of the selection that has acted on the corresponding genomic regions, as well as to reveal the selection factors on colorations.

Figures and Tables

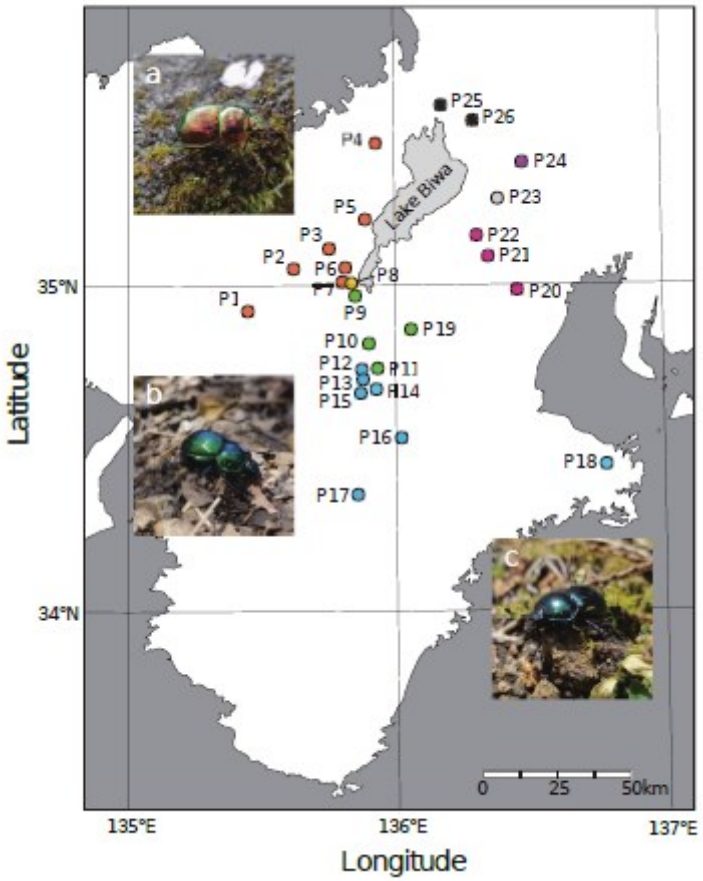

Figure 2.1. Study sites (P1–P26) in the Kinki District, Honshu, Japan. Circle colors indicate the population groups: orange, west/red; green, south/green; blue, south/indigo; pink, south/red; magenta, east/red; yellow, hybrid population between the west/red and the south/green; gray, hybrid population between the south/red and the east/red; black, *P. auratus* not found. Photographs of *P. auratus*: (a) red form at P6; (b) green form at P9; and (c) indigo form at P16. Photo credit: Y. Araki.

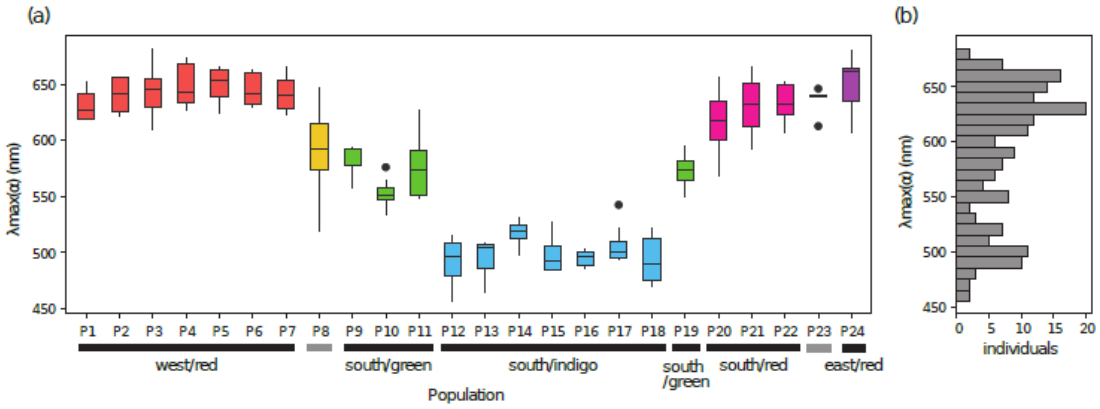

Figure 2.2. (a) Box plot of the color score (λmax(α)) of *P. auratus* at each site. Horizontal bar, median; rectangular box, interquartile range; vertical bar, range; dots, outliers. (b) Frequency distribution of the color score for all specimens.

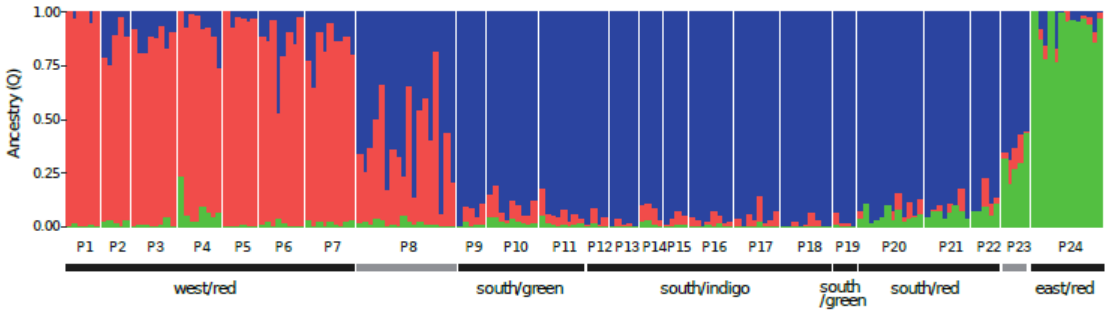

Figure 2.3. The ancestry scores for individuals of each site resulting from the population structure analyses at the optimal number of clusters $K = 3$: blue, the south cluster; orange, the west cluster; green, the east cluster.

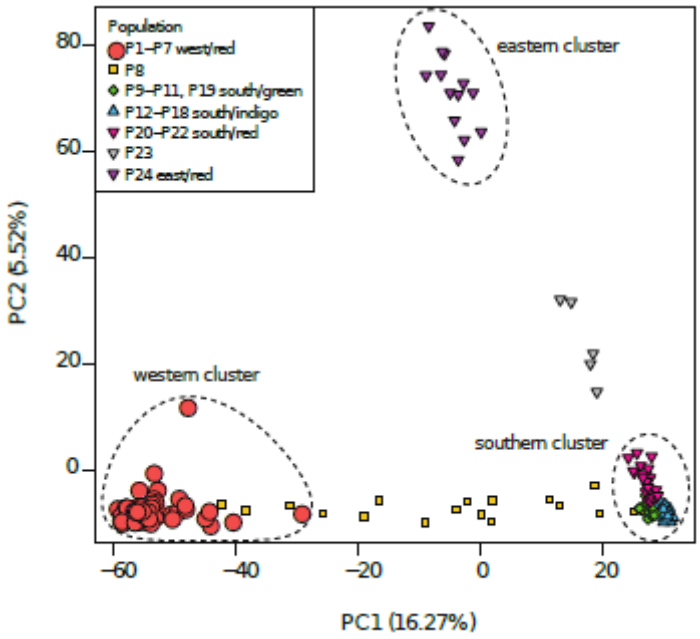

Figure 2.4. Plot of the PC1 and PC2 scores resulting from principal component analyses of SNPs. The proportions of variance explained by PC1 and PC2 were 16.27% and 5.52%, respectively. Colors of symbols indicate the different population groups as in Figure 2.1.

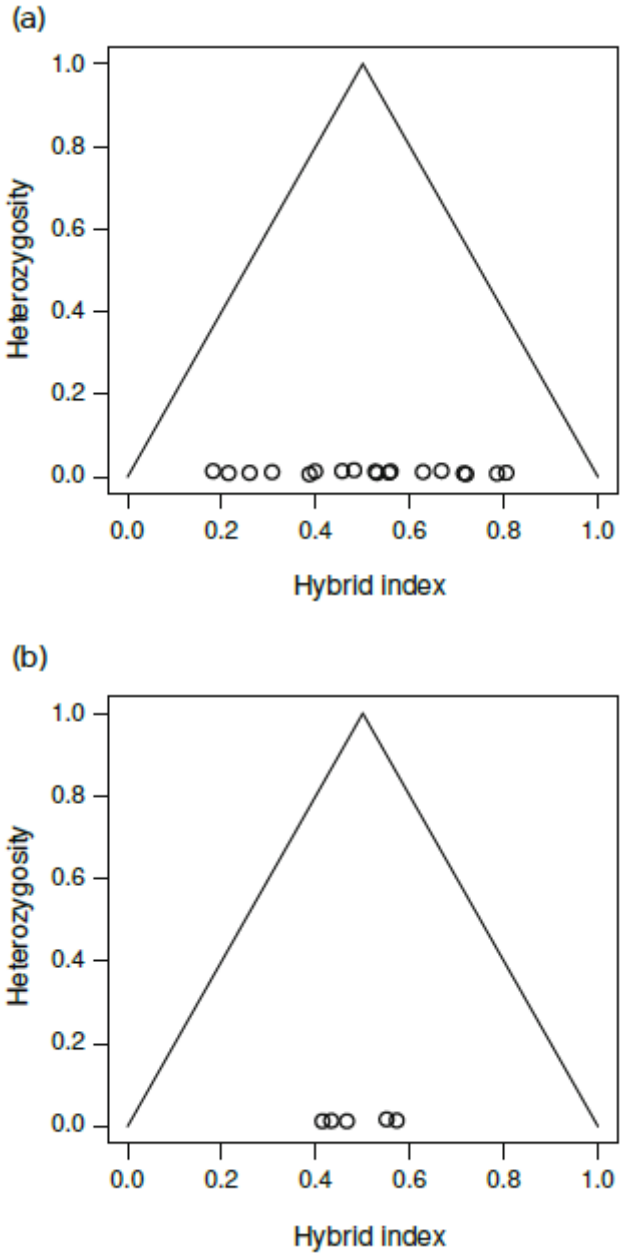

Figure 2.5. Triangle plot of heterozygosity (the fraction of markers heterozygous for alleles from parental populations) versus the hybrid index for individuals in populations (a) P8 (parental populations: P7 and P9) and (b) P23 (parental populations: P22 and P24).

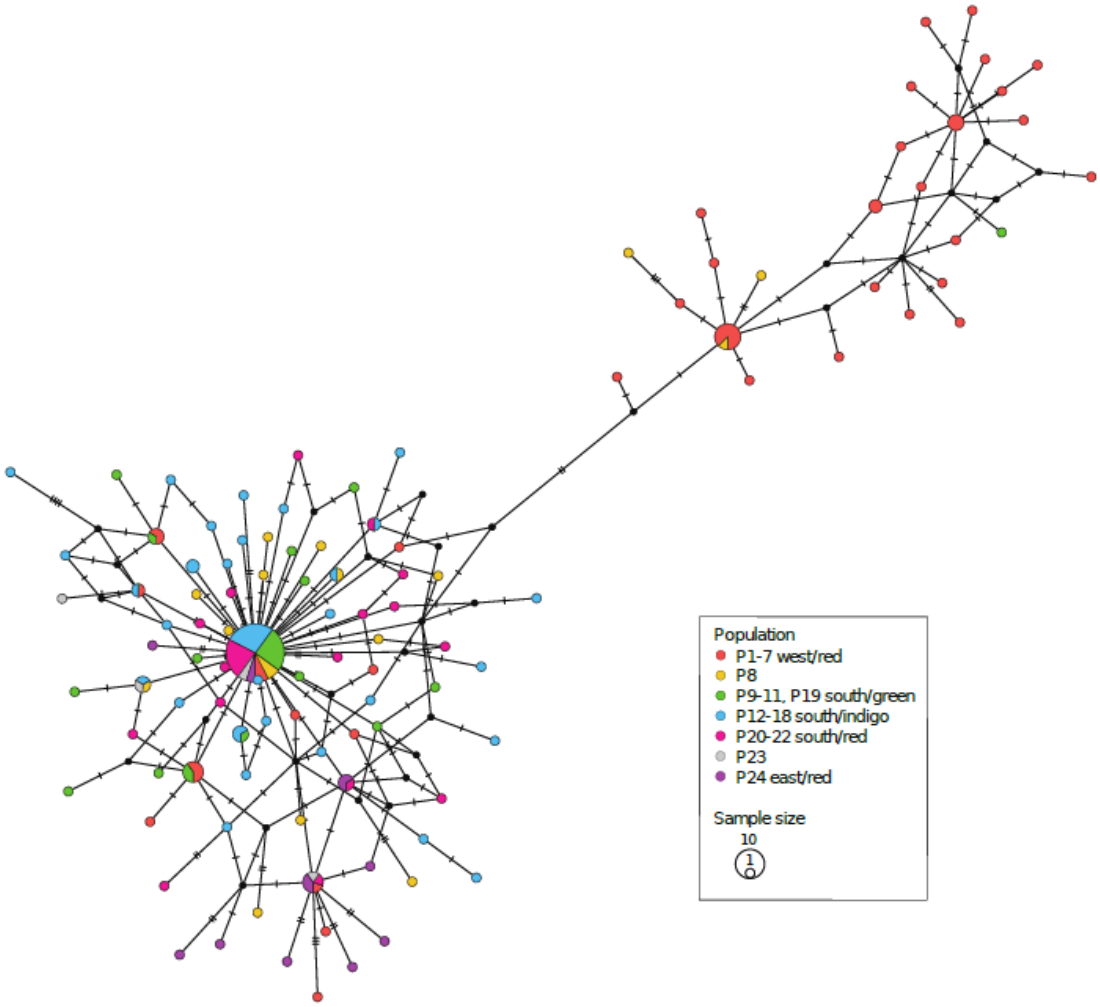

Figure 2.6. Haplotype network of mitochondrial COI gene sequences consisting of 110 haplotypes. Black dots indicate intermediate haplotypes that did not appear in the samples. The circle size reflects the sample size of the haplotype. Colors of symbols indicate the different population groups as in Figure 2.1.

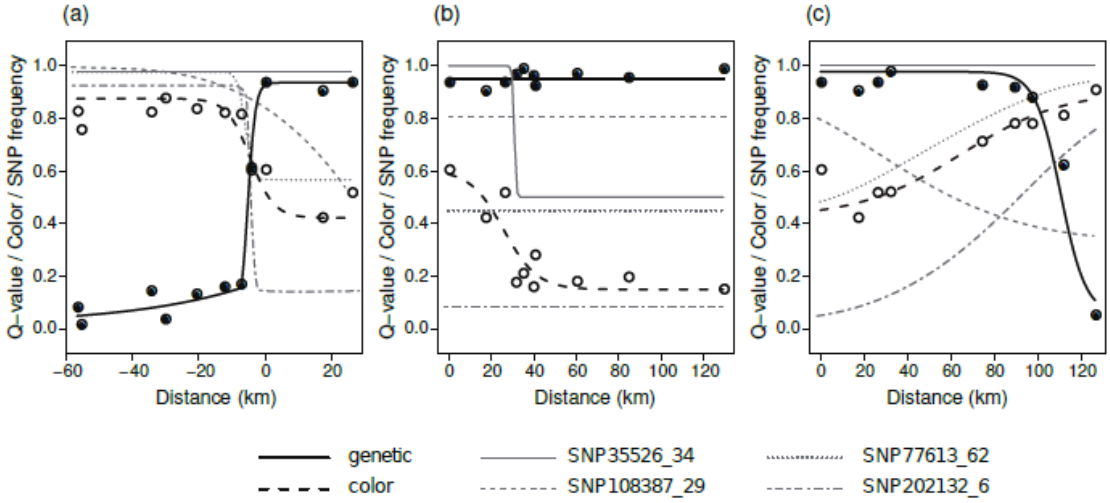

Figure 2.7. Geographic clines for population genetic assignment value (black circles and solid bold lines), elytral color (open circles and dashed bold lines), and allele frequency of the four color-related SNPs (SNP35526_34, solid lines; SNP108387_29, dashed lines; SNP76613_62, dotted lines; SNP202132_6, dot-dashed lines). The population genetic assignment value is the ancestry values of the south cluster (Q-value) obtained in the population structure analyses. Color is represented as the normalized scores of λmax(α). (a) The west/red to the south/green group (P1–P11); (b) the south/green (P9–P11) to the south/indigo groups (P9–P18); (c) the south/green to the east/red groups, via the populations of the south/red group (P9–P11, P19, P20–P22).

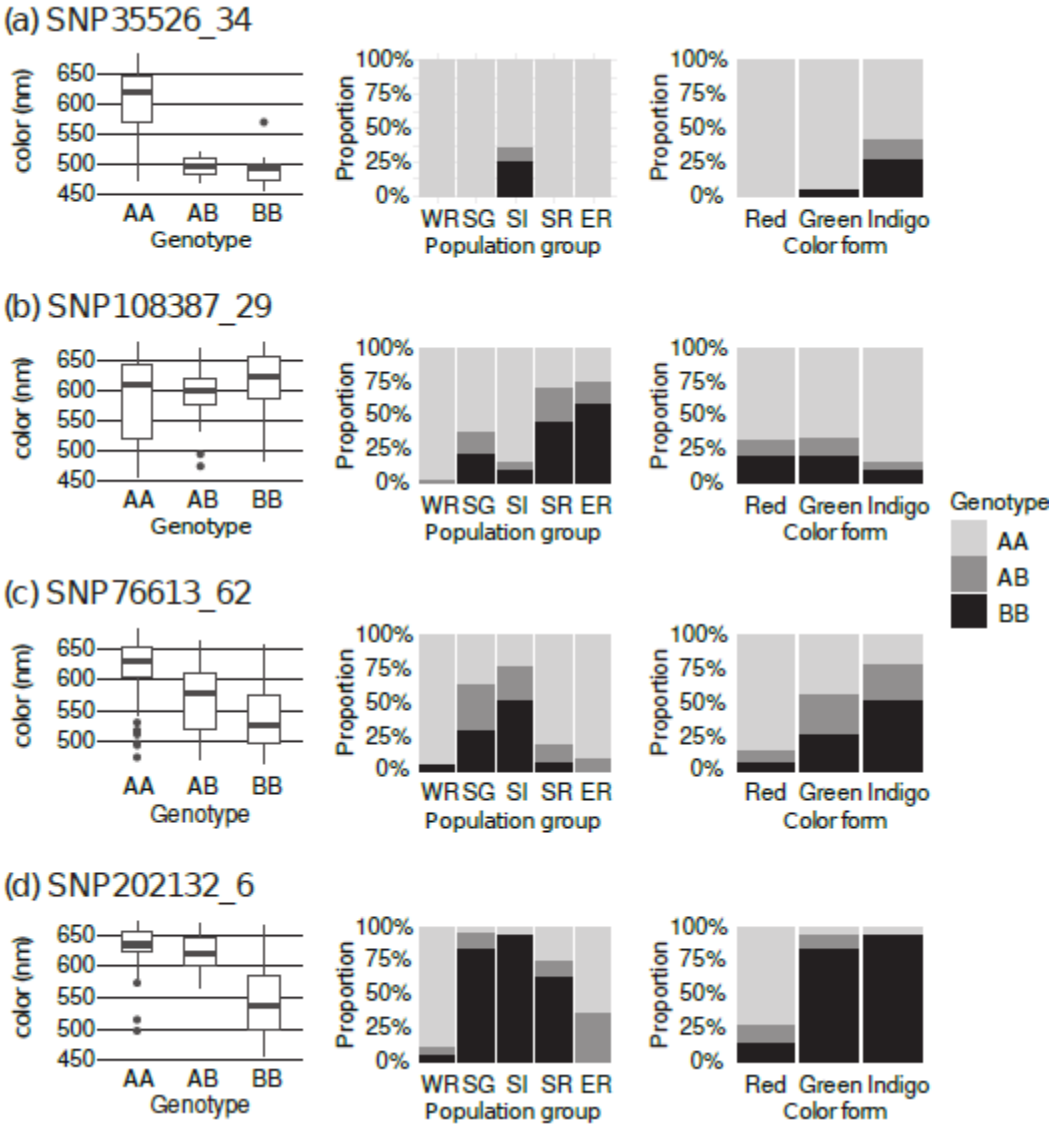

Figure 2.8. Box plot of color (λmax(α)) and individuals with each genotype (left) and the proportion of each genotype in each group (right) for the four SNPs associated with the color score. (a) SNP35526_34 (FDR = 0.007): (b) SNP108387_29 (FDR = 0.032): (c) SNP76613_62 (FDR = 0.050): (d) SNP202132_6 (FDR = 0.050). In the boxplot; horizontal bar, median; rectangular box, interquartile range; vertical bar, range; dots, outliers. Population groups in stacked bar plot; WR, the west/red group; SG, the south/green group; SI, the south/indigo group; SR, the south/red group; ER, the east/red group.

Table 2.1. The summary statistics of the color variation in *P. auratus*: population number, population names, cumulative distance starting from P9 (Mt. Otowa), the number of samples, the ranges, and the medians of the reflectance wavelengths.

Population	(Locality name)	Distance (km)	No. of samples	Range (nm)	Median (nm)
P1	(Mt. Myoken)	-55.579	6	618.817 – 652.493	626.990
P2	(Mt. Atago)	-34.519	5	621.844 – 655.702	642.059
P3	(Kibune)	-20.870	8	608.918 – 681.883	644.823
P4	(Mt. Bunagatake)	-56.702	8	627.106 – 672.674	642.777
P5	(Mt. Uchimi)	-30.260	6	624.311 – 665.416	653.971
P6	(Mt. Hiei)	-12.459	8	629.009 – 662.440	641.492
P7	(Mt. Daimonji)	-7.441	9	622.290 – 665.086	640.279
P8	(Mt. Nagara)	-4.397	18	518.666 – 646.275	592.189
P9	(Mt. Otowa)	0	5	558.131 – 594.174	592.7
P10	(Mt. Jubu)	17.119	9	533.983 – 576.041	551.640
P11	(Mt. Kasagi)	26.068	8	548.416 – 626.942	573.057
P12	(Mt. Tomyoji)	31.345	4	456.870 – 515.056	496.234
P13	(Gansenji Temple)	34.771	5	464.2143 – 508.294	503.972
P14	(Mt. Ittai)	40.350	4	498.415 – 531.606	519.711
P15	(Ishikiri touge)	39.541	4	484.450 – 526.972	492.371
P16	(Murou)	60.017	8	485.332 – 503.154	497.126
P17	(Mt. Yoshino)	84.356	8	493.666 – 542.536	500.752
P18	(Mt. Asamagatake)	129.240	9	470.227 – 521.573	490.383
P19	(Shigaraki)	31.991	4	550.028 – 595.477	573.699
P20	(Mt. Kiraramine)	74.156	12	568.845 – 656.476	616.957
P21	(Mt. Nihonkoba)	89.157	8	592.757 – 664.865	632.307
P22	(Mt. Takatori)	97.264	5	606.214 – 651.601	632.139
P23	(Mt. Ryozen)	111.680	5	612.636 – 645.947	639.270
P24	(Sekigahara)	126.531	13	607.342 – 680.342	661.116

Table 2.2. Estimated models and parameters of genetic and phenotypic (color) clines in the three transition zones between the south/green group and the others.

Cline[1]	Best fit model[2] scaling/tails	c (km)[3]	w (km)[4]	p_{min}[5]	p_{max}[5]	δ[6]	τ[6]
(a) Between west/red (P1-P7) – south/green (P9-P11)							
Q-value	fixed/left	-5.688 ± 0.025 (-10.739 ~ -0.132)	28.523 ± 0.142 (7.266 ~ 61.641)	0.015 (fixed)	0.936 (fixed)	46.927 ± 0.298 (left)	0.489 ± 0.003 (left)
Color	fixed/none	-0.70 ± 0.088 (-17.629 ~ 19.280)	55.855 ± 0.258 (8.727 ~ 97.620)	0.421 (fixed)	0.876 (fixed)	-	-
(b) Between south/green (P9-P11) – south/indigo (P12-P18)							
Q-value	null	-	-	0.952	0.952	-	-
Color	fixed/none	24.990 ± 0.120 (3.295 ~ 52.190)	71.270 ± 0.336 (9.616 ~ 126.910)	0.149 (fixed)	0.604 (fixed)	-	-
(c) Between south/green (P9-P11) – east/red (P24) via south/red (P20-P22)							
Q-value	fixed/none	110.46 ± 0.047 (100.620 ~ 119.630)	31.71 ± 0.181 (5.950 ~ 78.930)	0.252 (fixed)	0.968 (fixed)	-	-
Color	fixed/none	65.360 ± 0.165 (30.310 ~ 97.780)	83.500 ± 0.293 (17.930 ~ 127.630)	0.421 (fixed)	0.908 (fixed)	-	-

[1]Q-value, the mean of the ancestry scores of the south cluster in the population structure analysis; color, the mean of the normalized individual λmax(α) scores.

[2]The best fit model, represented by a pair of the "scaling" (none, p_{min}/p_{max} fix to 0/1; fixed, p_{min}/p_{max} fix to observed parameter; or free,

p_{min}/p_{max} is desired as free parameters) and "tails" (none, no exponential tails; right, one exponential tail on the right; left, one exponential tail on the left; mirrored, two exponential tails mirrored about the cline center; or both, two tails with independent parameters).

[3]c, the cumulative distance to the cline center from P9 (Mt. Otowa). Values are mean ± SE and 95% CI in parentheses.

[4]w, the width of the cline calculated as 1/maximum slope. Values are mean ± SE and 95% CI in parentheses.

[5]p_{min} and p_{max}, the minimum and the maximum frequency at the end of the cline, respectively.

[6]δ and τ, the exponential curve (tails) shape parameters. Values are mean ± SE.

Table 2.3. Neutrality tests by Tajima's *D* and *R2* for the five population groups using the SNP data.

Group	Sample size	number of SNPs	Tajima's *D*	*P* (*D*)	*R2*	*P* (*R2*)
west/red	50	129106	-3.667	< 0.001	0.016	< 0.001
south/green	26	129106	-4.005	< 0.001	0.033	< 0.001
south/indigo	42	129106	-3.745	< 0.001	0.019	< 0.001
south/red	25	129106	-4.025	< 0.001	0.036	< 0.001
east/red	13	129106	-4.443	< 0.001	0.082	0.003

Table 2.4. Neutrality tests by Tajima's *D* and Fu's F_S for the five population groups using the COI sequences.

Group	Sample size	π	Tajima's *D*	*P* (*D*)	Fu's F_S	*P* (F_S)
west/red	50	5.647	-0.934	0.181	-25.328	< 0.001
south/green	26	2.191	-2.376	0.001	-12.426	< 0.001
south/indigo	42	2.459	-2.339	0.002	-26.684	< 0.001
south/red	25	2.180	-2.208	0.001	-15.120	< 0.001
east/red	13	3.000	-1.390	0.082	-5.137	< 0.001

π, the sum of the pairwise differences divided by the number of pairs.

Supplemental information

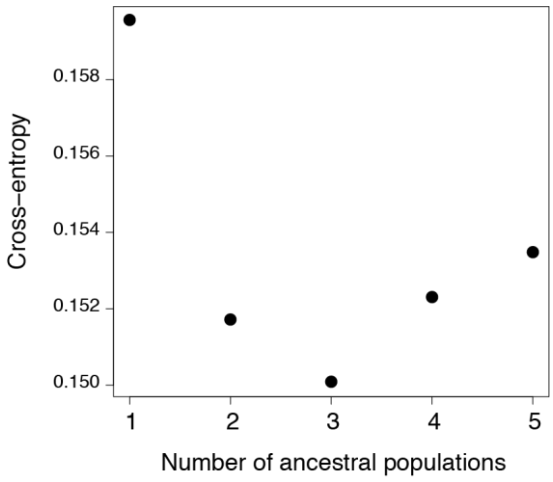

Figure S2.1. Values of the cross-entropy criterion for sNMF runs for the number of ancestral population (*K*) from 1 to 5.

Table S2.1. The locality and environmental data of each population.

Population	Locality name	Latitude (°N)	Longitude (°E)	Median of λmax(α) (nm)	Annual rainfall (mm)	Annual average temperature (°C)	Annual sunlight hours
P1	Mt. Myoken	34.9288	135.4674	626.99	1521.1	11.5	1738.6
P2	Mt. Atago	35.0600	135.6342	642.059	1509.0	10	1670.8
P3	Kibune	35.1222	135.7633	644.823	1974.0	12.6	1640.0
P4	Mt. Bunagatake	35.4495	135.9324	642.777	1389.2	12.3	1776.4
P5	Mt. Uchimi	35.2133	135.8942	653.971	2029.6	8.9	1551.8
P6	Mt. Hiei	35.0640	135.8224	641.492	1592.1	12.1	1730.0
P7	Mt. Daimonji	35.0197	135.8117	640.279	1518.6	13.3	1768.8
P8	Mt. Nagara	35.0159	135.8447	592.189	1519.1	13.6	17857
P9	Mt. Otowa	34.9768	135.8530	592.7	1481.8	12.3	1807.6
P10	Mt. Jubu	34.8296	135.9090	551.64	1459.9	11.1	1797.9
P11	Mt. Kasagi	34.7534	135.9410	573.057	1371.1	13.7	1799.0
P12	Mt. Tomyoji	34.7499	135.8835	496.234	1394.7	14.1	1806.9
P13	Gansenji Temple	34.7192	135.8879	503.972	1393.6	13.4	1795.1
P14	Mt. Ittai	34.6889	135.9365	519.711	1388.8	12.2	1775.3
P15	Ishikiri touge	34.6767	135.8798	492.371	1387.1	12.2	1781.7
P16	Murou	34.5391	136.0285	497.126	1513.4	11.9	1664.1
P17	Mt. Yoshino	34.3631	135.8704	500.752	1473.3	12.7	1631.7
P18	Mt. Asamagatake	34.4614	136.7764	490.383	2345.0	13	1962.1
P19	Shigaraki	34.9991	136.4488	573.699	2359.7	10.6	1663.8
P20	Mt. Kiraramine	35.1022	136.3424	616.957	2121.6	9.5	1619.8
P21	Mt. Nihonkoba	35.1666	136.3004	632.307	1815.4	11.6	1657.9
P22	Mt. Takatori	35.2804	136.3768	632.139	1962.9	8.5	1719.0
P23	Mt. Ryozen	35.3934	136.4644	639.27	1632.1	15.4	1979.5
P24	Sekigahara	35.2627	136.5234	661.116	2154.8	10.1	1736.4

Table S2.2. Fixation indices (F_{CT}) for the population groups computed by SAMOVA based on the COI sequences.

Population grouping	Number of groups	F_{CT}	*P*-value
(P1, P2, P3, P5), (others)	2	0.431	< 0.001
(P1, P2, P3), (P5), (others)	3	0.420	< 0.001
(P1, P3), (P2), (P5), (others)	4	0.413	< 0.001

Table S2.3. Estimated models and parameters showing top three AICc of genetic and phenotypic (color) clines in the three transition zones.

Cline[1]	Fitting model[2] scaling/tails	AICc	c (km)[3]	w (km)[4]	p_{min}[5]	p_{max}[5]	δ[6]	τ[6]
(a) Between west/red (P1-P7) – south/green (P9-P11)								
Q-value	fixed/left	9.585	-5.688 ± 0.025 (-10.739 ~ -0.132)	28.523 ± 0.142 (7.266 ~ 61.641)	0.015 (fixed)	0.936 (fixed)	46.927 ± 0.298 (left)	0.489 ± 0.003 (left)
	none/mirror	9.834	-4.480 ± 0.032 (-10.561 ~ 2.431)	42.756 ± 0.195 (13.594 ~ 76.907)	0.000	1.000	41.260 ± 0.316 (mirror)	0.020 ± 0.003 (mirror)
	free/none	10.143	-5.090 ± 0.022 (-9.564 ~ -0.010)	15.721 ± 0.122 (1.816 ~ 46.568)	0.097 ± 4.888e-04	0.893 ± 6.758e-04	-	-
Color	fixed/none	6.036	-0.704 ± 0.088 (-17.629 ~ 19.280)	55.855 ± 0.258 (8.727 ~ 97.620)	0.421 (fixed)	0.876 (fixed)	-	-
	free/none	9.304	8.755 ± 0.196 (-29.470 ~ 51.084)	56.788 ± 0.270 (5.653 ~ 98.030)	0.370 ± 0.002	0.816 ± 0.001	-	-
	fixed/mirror	9.593	-0.281 ± 0.104 (-18.092 ~ 22.755)	55.407 ± 0.257 (8.465 ~ 97.421)	0.421 (fixed)	0.876 (fixed)	50.232 ± 0.283 (mirror)	0.497 ± 0.003 (mirror)
SNP35526_34	null	6.755	-	-	0.977	0.977	-	-
	fixed/none	7.823	-4.808 ± 0.192 (-49.219 ~ 27.000)	52.247 ± 0.245 (4.273 ~ 88.360)	0.893 (fixed)	1.000	-	-

SNP108387_29	free/none	11.085	-2.569 ± 0.279 (-55.377 ~ 29.383)	45.265 ± 0.299 (1.964 ~ 87.634)	0.875 ± 0.003	0.972 ± 2.575e-04	-	-
	none/none	11.193	23.360 ± 0.044 (13.740 ~ 29.680)	61.190 ± 0.132 (37.580 ~ 86.690)	0.000	1.000	-	-
	fixed/none	11.758	10.670 ± 0.080 (-4.157 ~ 26.910)	59,970 ± 0.168 (28.547 ~ 88.250)	0.500	1.000	-	-
SNP76613_62	free/none	12.003	11.100 ± 0.120 (-13.640 ~ 28.676)	40.223 ± 0.214 (4.069 ~ 83.621)	0.451 ± 0.002	0.955 ± 3.319e-04	-	-
	free/none	13.000	0.305 ± 0.132 (-24.724 ~ 26.234)	41.528 ± 0.271 (2.235 ~ 86.771)	0.479 ± 0.002	0.944 ± 4.54e-04		
	fixed/none	13.120	-3.147 ± 0.070 (-16.760 ~ 11.710)	55.765 ± 0.181 (22.880 ~ 87.270)	0.500 (fixed)	1.000 (fixed)	-	-
	fixed/mirror	13.331	-3.150 ± 0.071 (-16.331 ~ 12.279)	52.295 ± 0.196 (18.683 ~ 86.934)	0.500 (fixed)	1.000 (fixed)	46.343 ± 0.256 (mirror)	0.514 ± 0.003 (mirror)
SNP202132_6	free/none	12.582	-3.417 ± 0.041 (-10.097 ~ 7.068)	19.002 ± 0.178 (1.281 ~ 66.323)	0.185 ± 0.001	0.912 ± 0.001	-	-
	fixed/left	13.286	-1.535 ± 0.046 (-9.467 ~ 8.917)	48.303 ± 0.199 (14.881 ~ 85.243)	0.083 (fixed)	1.000 (fixed)	36.546 ± 0.258 (left)	0.444 ± 0.003 (left)
	fixed/mirror	13.341	-1.616 ± 0.046 (-9.485 ~ 8.980)	47.506 ± 0.217 (13.520 ~ 84.675)	0.083 (fixed)	1.000 (fixed)	36.456 ± 0.258 (mirror)	0.434 ± 0.003 (mirror)
(b) Between south/green (P9-P11) – south/indigo (P12-P18)								
Q-value	null	3.201	-	-	0.952	0.952	-	-

	fixed/none	4.813	60.580 ± 0.348 (4.440 ~ 124.700)	70.500 ± 0.365 (4.705 ~ 127.500)	0.904 (fixed)	0.988 (fixed)	-	-
	fixed/mirror	9.189	60.557 ± 0.350 (4.204 ~ 124.833)	70.996 ± 0.364 (4.713 ~ 127.373)	0.904 (fixed)	0.988 (fixed)	64.197 ± 0.383 (mirror)	0.498 ± 0.003 (mirror)
Color	fixed/none	5.803	24.990 ± 0.120 (3.295 ~ 52.190)	71.270 ± 0.336 (9.616 ~ 126.910)	0.149 (fixed)	0.604 (fixed)	-	-
	null	9.406	-	-	0.344	0.344	-	-
	none/none	9.439	16.310 ± 0.081 (1.997 ~ 32.440))	116.120 ± 0.111 (88.589 ~ 129.49)	0.000	1.000	-	-
SNP35526_34	fixed/none	11.763	40.500 ± 0.147 (18.780 ~ 75.020)	65.950 ± 0.399 (2.310 ~ 127.720)	0.500 (fixed)	1.000 (fixed)	-	-
	free/none	12.347	54.811 ± 0.375 (4.089 ~ 126.641)	64.657 ± 0.413 (2.690 ~ 127.139)	0.668 ± 0.001	0.876 ± 7.976e-04	-	-
	fixed/right	13.856	40.175 ± 0.142 (16.656 ~ 72.814)	68.562 ± 0.423 (2.704 ~ 127.461)	0.500 (fixed)	1.000 (fixed)	58.465 ± 0.402 (right)	0.465 ± 0.003 (right)
SNP108387_29	null	11.410	-	-	0.806	0.806	-	-
	fixed/none	12.836	38.700 ± 0.177 (7.732 ~ 78.190)	107.500 ± 0.166 (67.465 ~ 129.120)	0.500 (fixed)	1.000 (fixed)	-	-
	free/none	13.192	44.961 ± 0.325 (1.915 ~ 121.154)	68.369 ± 0.363 (4.321 ~ 126.942)	0.623 ± 0.002	0.885 ± 0.001	-	-

SNP76613_62	null	20.996	-	-	0.449	0.449	-	-
	free/none	25.608	87.599 ± 0.364 (4.811 ~ 128.723)	63.392 ± 0.381 (2.686 ~ 126.924)	0.390 ± 0.001	0.639 ± 0.002	-	-
	fixed/left	25.612	106.543 ± 0.137 (77.981 ~ 128.413)	102.199 ± 0.217 (49.825 ~ 129.109)	0.000	1.000	8.762 ± 0.067 (left)	0.097 ± 0.001 (left)
SNP202132_6	null	8.428	-	-	0.083	0.083	-	-
	fixed/none	8.803	42.610 ± 0.250 (4.567 ~ 103.500)	84.160 ± 0.311 (14.198 ~ 127.900)	0.000 (fixed)	0.250 (fixed)	-	-
	none/none	12.456	4.170 ± 0.033 (0.136 ~ 12.860)	77.630 ± 0.210 (42.966 ~ 122.700)	0.000	1.000	-	-
(c) Between south/green (P9-P11, P19) – east/red (P24) via south/red (P20-P22)								
Q-value	fixed/none	8.189	110.460 ± 0.047 (100.620 ~ 119.630)	31.710 ± 0.181 (5.950 ~ 78.930)	0.051 (fixed)	0.978 (fixed)	-	-
	free/none	8.983	111.928 ± 0.056 (99.726 ~ 122.710)	20.927 ± 0.148 (1.720 ~ 58.155)	0.113 ± 0.001	0.911 ± 4.083e-04	-	-
	none/left	9.009	110.713 ± 0.060 (99.155 ~ 122.383)	52.848 ± 0.285 (16.606 ~ 111.913)	0.000	1.000	31.508 ± 0.275 (left)	0.285 ± 0.003 (left)
Color	fixed/none	5.004	65.360 ± 0.165 (30.310 ~ 97.780)	83.500 ± 0.293 (17.930 ~ 127.630)	0.421 (fixed)	0.908 (fixed)	-	-
	none/none	8.787	34.500 ± 0.095	120.500 ± 0.081	0.000	1.000	-	-

			(15.160 ~ 53.100)	(100.140 ~ 129.700)				
	fixed/left	9.295	66.301 ± 0.176 (29.955 ~ 98.899)	82.731 ± 0.300 (18.277 ~ 127.671)	0.421 (fixed)	0.908 (fixed)	64.672 ± 0.373 (left)	0.499 ± 0.003 (left)
SNP108387_29	fixed/none	7.962	24.950 ± 0.126 (2.685 ~ 52.940)	91.620 ± 0.270 (33.817 ~ 128.270)	0.333 (fixed)	1.000 (fixed)	-	-
	null	11.194	-	-	0.510	0.510	-	-
	none/none	12.358	72.630 ± 0.107 (52.150 ~ 92.830)	120.450 ± 0.082 (99.360 ~ 129.740)	0.000	1.000	-	-
SNP76613_62	fixed/none	8.709	56.320 ± 0.149 (28.530 ~ 84.650)	105.49 ± 0.174 (63.610 ~ 128.970)	0.375 (fixed)	1.000 (fixed)	-	-
	none/none	10.016	21.410 ± 0.096 (3.442 ~ 40.630)	117.120 ± 0.102 (91.926 ~ 129.51)	0.000	1.000	-	-
	fixed/mirror	12.643	52.952 ± 0.164 (20.174 ~ 83.815))	98.140 ± 0.232 (44.298 ~ 128.765)	0.375 (fixed)	1.000 (fixed)	53.967 ± 0.379 (mirror)	0.481 ± 0.003 (mirror)
SNP202132_6	none/none	8.454	90.770 ± 0.095 (71.250 ~ 110.400)	108.490 ± 0.152 (75.62 ~ 129.200)	0.000	1.000	-	-
	fixed/none	9.326	76.690 ± 0.120 (52.780 ~ 99.870)	102.940 ± 0.177 (64.490 ~ 128.660)	0.000 (fixed)	0.818 (fixed)	-	-
	fixed/left	11.586	80.479 ± 0.129 (54.602 ~ 105.429)	95.140 ± 0.232 (43.750 ~ 128.103)	0.000 (fixed)	0.818 (fixed)	55.620 ± 0.380 (left)	0.477 ± 0.003 (left)

[1]Q-value, the mean of the ancestry scores of the south cluster in the population structure analysis; color, the mean of the normalized individual λmax(α) scores.
[2]The best fit model, represented by a pair of the "scaling" (none, p_{min}/p_{max} fix to 0/1; fixed, p_{min}/p_{max} fix to observed parameter; or free, p_{min}/p_{max} is desired as free

parameters) and “tails” (none, no exponential tails; right, one exponential tail on the right; left, one exponential tail on the left; mirrored, two exponential tails mirrored about the cline center; or both, two tails with independent parameters).
[3]c, the cumulative distance to the cline center from P9 (Mt. Otowa). Values are mean ± SE and 95% CI in parentheses.
[4]w, the width of the cline calculated as 1/maximum slope. Values are mean ± SE and 95% CI in parentheses.
[5]p_{min} and p_{max}, the minimum and the maximum frequency at the end of the cline, respectively.
[6]δ and τ, the exponential curve (tails) shape parameters. Values are mean ± SE.

Table S2.4. Linkage disequilibrium indices between the SNPs selected by the GWAS within the west, south, and east genetic clusters.

Genetic cluster Pair of SNP sites	*D*	*D'*	*r*	X^2	*P*-value[1]
West cluster					
SNP35526_34[†]/SNP108387_29	-	-	-	-	-
SNP35526_34[†]/SNP76613_62	-	-	-	-	-
SNP35526_34[†]/SNP202132_6	-	-	-	-	-
SNP108387_29/SNP76613_62	-4.335e-04	0.897	0.020	0.033	0.854
SNP108387_29/SNP202132_6	0.001	0.255	0.191	2.691	0.101
SNP76613_62/SNP202132_6	-7.837e-04	0.940	-0.028	0.060	0.807
South cluster					
SNP35526_34/SNP108387_29	0.004	0.045	0.031	0.042	0.837
SNP35526_34/SNP76613_62	-0.002	0.058	-0.015	0.021	0.885
SNP35526_34/SNP202132_6	-0.018	0.997	-0.154	2.222	0.136
SNP108387_29/SNP76613_62	-0.051	0.410	-0.227	6.203	0.013
SNP108387_29/SNP202132_6	-0.022	0.397	-0.134	1.971	0.160
SNP76613_62/SNP202132_6	0.040	0.427	0.260	7.423	0.006
East cluster					
SNP35526_34[†]/SNP108387_29	-	-	-	-	-
SNP35526_34[†]/SNP76613_62	-	-	-	-	-
SNP35526_34[†]/SNP202132_6	-	-	-	-	-
SNP108387_29/SNP76613_62	0.028	0.997	0.293	1.902	0.168
SNP108387_29/SNP202132_6	-7.000e-06	0.001	-9.020e-05	2.000e-07	1.000
SNP76613_62/SNP202132_6	-6.056e-02	0.999	-0.333	2.219	0.136

[1]All nonsignificant after the Bonferroni correction at $\alpha = 0.05$.
[†]No genotype data for this SNP site within the genetic cluster.

Chapter 3. Whole-genome resequencing reveals recent divergence of geographic populations of the dung beetle *Phelotrupes auratus* with color variation

3.1. Introduction

Knowledge of the history of population divergence is crucial to our understanding of organisms' diversification mechanisms. The evolutionary histories of wild organisms have been shaped largely by past climatic and geographic events. For example, environmental changes that occurred during the last glacial period strongly influenced species' distribution patterns. The combined interpretation of past climatic and geographic events and molecular population genetic data has provided a powerful means to elucidate the evolutionary history of organisms over relatively long time scales such as the Quaternary (Hewitt, 2000, 2004). In recent years, the impact of human activities on the evolution of organisms has gradually become clear. Human-induced environmental changes have affected the evolution of adaptive traits in many organisms (Sullivan et al., 2017), and have led to the decline and extinction of wild populations (Ceballos et al., 2015). During the early to mid-Holocene, humans began to practice agriculture and pastoralism, rapidly increasing their population size and impact on the environment (Boivin et al., 2016). It is thus important to consider the possibility that evolutionary changes in the populations of wild organisms have been influenced by not only climatic and geographic events but also human activities.

Genetic variation between local populations is determined by the balance of the homogenizing effect of gene flow and the diversifying effect of local adaptation (Endler, 1977). Geographic variation in adaptive traits often reflects genetic differences among populations due to restricted gene flow. Among many adaptive traits, coloration and

pigment patterns are inferred as inter- or intra-specific signals that have been used to define populations in studies of evolutionary processes (Curran et al., 2020; Hoyal Cuthill & Charleston, 2015; McLean & Stuart-Fox, 2014; Mullen et al., 2009; Stankowski et al., 2017). Recently, the development of high-throughput sequencing technology has enabled the acquisition of large amounts of fine-scale sequence variation data that provide powerful clues about the processes underlying population divergence through whole-genome resequencing (Ellegren, 2014) or reduced-representation approaches such as restriction site-associated DNA sequencing (RAD-seq) (Baird et al., 2008). Currently, great efforts are being made to estimate past demographic changes and evolutionary events that have occurred in various wildlife populations based on large amounts of genomic data (Excoffier et al., 2013; Li & Durbin, 2011; Schiffels & Wang, 2020; Terhorst et al., 2017; Zhou et al., 2020).

Here, I focused on the geotrupid dung beetle *Phelotrupes auratus* (Motschulsky). It exhibits notable variation in its structural metallic body coloration, which is provided by a simple multilayer reflector structure (Akamine, Ishikawa, et al., 2011) and is categorized as red, green, or indigo based on the peak wavelength on the elytral reflectance spectrum (Akamine et al., 2008; Chapter 2). *Phelotrupes auratus* is a diurnal beetle that inhabits montane forests; its food resource is the dung of large wild mammals, especially sika deer (Tsukamoto et al., 2014). In the Kinki District of Honshu, the main island of Japan, the three color forms occur separately in different geographic areas (Akamine, Maekawa, & Kon, 2011; Figure 1a); however, the mechanism by which this variation formed and has been maintained remains unclear. My previous population-structure analysis using RAD markers (Chapter 2) showed that *P. auratus* populations in the Kinki District diverged into five geographic groups defined by combinations of three

color forms and three genetic clusters—west/red (WR), east/red (ER), south/indigo (SI), south/green (SG), and south/red (SR)—and that geographic color variation was maintained by hybrid transition zones with steep genetic clines (i.e., barriers to gene flow) or transition zones without genetic differentiation. In the latter zone type, I inferred the action of strong selection on different body colors in different regions. However, as the function of body color in *P. auratus* is unclear, the inference that color differentiation is maintained solely by selection despite frequent gene flow was questionable. Obscure genetic differences between sister populations may have been caused by very recent population divergence, even if they are separated by barriers to gene flow. Therefore, the lack of detection of genetic differentiation in Chapter 2 may be attributable to the low resolution of the RAD-seq data and/or the methodology of my population clustering analysis. Analysis using a larger genomic dataset, such as that provided by whole-genome resequencing, may provide new insights into geographic color diversification in *P. auratus*.

In this study, to construct robust demographic history models and estimate the dynamics of effective population sizes, gene flow, and divergence times among populations, I constructed a *de novo* draft genome of *P. auratus* and analyzed genome-wide single-nucleotide polymorphism (SNP) data obtained by whole-genome resequencing of individuals sampled from five local populations. I found that the five local populations diverged recently—within several thousand years, under the influence of human activity—and that no substantial gene flow had occurred among the diverged populations. My findings reveal an unexpected influence of human activity on population differentiation through the divergence of a conspicuous trait in this insect species.

3.2. Materials and Methods

3.2.1. Sampling and individual coloration data

Between 2019 and 2020, *P. auratus* adults were collected using traps baited with horse dung at six locations around Lake Biwa in the Kinki District, Honshu, Japan (Figure 1a). To obtain five or more individuals from each of the five local populations of *P. auratus* identified in the region, one sampling site was selected for each population based on my previous data; SG was sampled at two locations to obtain sufficient samples. A total of 27 beetles were collected (Table 1) and fixed in 99% ethanol; one beetle collected from SI was immersed in RNAlater solution for DNA and RNA sampling, and then stored at −30°C until DNA/RNA extraction. As in Chapter 2, I quantified the elytral colors of individuals using a spectrometer (USB2000+UV-VIS-ES; Ocean Optics, Rochester, NY, USA) to measure the reflection spectra of the elytral surfaces of dry specimens in a dark room, and classified the color forms as indigo (< 525 nm), green (525–600 nm), or red (> 600 nm) (Chapter 2).

3.2.2. Genomic DNA extraction and sequencing

Total genomic DNA was extracted from the flight muscle and testes using a Genomic-tip 20/G kit (QIAGEN, Hilden, Germany), following the manufacturer's protocol. The quality of DNA samples was assessed with Qubit (Thermo Fischer Scientific), NanoPhotometer (Implen), and 2200 TapeStation (Agilent) instruments. For genome assembly, the genomic DNA of an SI individual (SI03) was sent to GeneBay, Inc. (Yokohama, Japan) for library construction and long-read sequencing using the ONT PromethION system (Oxford Nanopore Technologies). Short-read sequencing was performed with 150-bp paired-end reads in three runs on an Illumina HiSeq X Ten

sequencer (Illumina, San Diego, CA, USA) at Macrogen Japan (Tokyo, Japan). For genome assembly, a single sample (SI03) was sequenced using one lane of the HiSeq X Ten sequencer. For resequencing, 10 samples from the SI and ER groups and 15 samples from the WR, SG, and SR groups were separately sequenced using two lanes of the HiSeq X Ten sequencer. Library construction for SI and ER samples was performed at Macrogen Japan and that for WR, SG, and SR samples was performed at Kyoto University using the NEB Next Ultra II FS DNA Library Prep Kit for Illumina (E7805S; New England Biolabs Japan, Tokyo, Japan). Adapters on raw reads were trimmed using Porechop (v0.2.4, https://github.com/rrwick/Porechop) for ONT PromethION long reads and fastp (Chen et al., 2018) for Illumina HiSeq X short reads.

3.2.3. Transcriptomic RNA extraction and sequencing

For gene prediction on the reference *P. auratus* genome, I obtained mRNA sequence data from an adult SI specimen (not included in Table 1) fixed in RNAlater solution. Total RNA was extracted using a Monarch Total RNA Miniprep Kit (New England Biolabs Japan), following the manufacturer's protocol. Stranded paired-end 150-bp library construction and sequencing were conducted using GENEWIZ (South Plainfield, NJ, USA). Adapters on raw reads were trimmed using fastp.

3.2.4. Genome assembly for reference genome construction

Genome size was estimated with Illumina short reads using GenomeScope (Vurture et al., 2017, http://qb.cshl.edu/genomescope/) based on the result of the 31-mer count obtained with Jellyfish v2.3.0 (Marçais & Kingsford, 2011). I *de novo* assembled a draft genome of *P. auratus* (individual SI03) at the scaffold level using Illumina HiSeq X short reads

and ONT PromethION long reads by a hybrid assembly procedure using MaSuRCA 3.4.2 (Zimin et al., 2013). The constructed draft genome was polished with racon 1.4.13 (Vaser et al., 2017) and Pilon 1.23 (Walker et al., 2014), then curated using Purge Haplotigs 1.0.4 (Roach et al., 2018). Using BUSCO v5.2.5 (Simão et al., 2015), the quality of the curated draft genome was checked against the "endopterygota_odb10" database (updated on 10 September 2020, containing 2124 single-copy ortholog genes); masked tandem repeats were detected using RepeatModeler 2.0.1 (Flynn et al., 2020) and RepeatMasker 4.1.1 (Smit et al., 2015). Gene prediction was conducted with the BRAKER v2.1.4 pipeline (Barnett et al., 2011; Brůna et al., 2021; Hoff et al., 2016, 2019; Lomsadze et al., 2014; Stanke et al., 2006, 2008) using the draft genome and mapped RNA-seq data. RNA-seq mapping was performed with hisat2 v2.2.1 (Kim et al., 2019) and samtools.

3.2.5. Read mapping and variant calling

Short reads were mapped to the draft genome with bwa-mem 0.7.17 (Li & Durbin, 2009) and sorted with picard SortSam v2.25.7 (https://broadinstitute.github.io/picard/). The average read depth was 16.02, excluding SI03 (read depth, 154.11). SNP calling was conducted for each sample using bcftools 1.12 (Li et al., 2009); genotype likelihoods were computed using the mpileup command and variants were called using the call command. Based on the average read depth of each sample, generated by samtools 1.13 (Li et al., 2009), I collected SNPs with Phred quality scores > 10 and depths between two-thirds (or 10 if the average depth was < 15×) and twice the average depth, and removed indels. Called variants were merged into a single VCF file with bcftools merge, then filtered with maf > 0.05, max missing < 0.15, -m = 2, and -M = 2 to remove rare substitutions, missing abundant sites, and multiple substitution sites.–

3.2.6. Phylogenetic and genetic population structure analyses

A phylogenetic tree was constructed using all the obtained SNP data and the maximum likelihood method in IQ-TREE 2.1.2 (Minh et al., 2020) with 1,000 UF-bootstrap iterations (Hoang et al., 2018) and 1,000 bootstrap replicates for the SH-like approximate likelihood ratio test (SH-aLRT) (Guindon et al., 2010). The best-fitting substitution model was estimated using ModelFinder (Kalyaanamoorthy et al., 2017) with an ascertainment bias correction model (Lewis, 2001); as a result PMB+F+ASC+R4 was selected. In the optimal topology, five individuals of WR were assigned to the outgroup based on the results of Chapter 2, although I did not assume the monophyly of WR individuals prior to the phylogenetic analysis. I also performed population genetic structural analysis using the ADMIXTURE v1.3.0 software (Alexander et al., 2009) with the number of ancestral clusters, *K*, ranging from 2 to 5 and principal component analyses using the pcaMethods package in the R software (Stacklies et al., 2007) based on the SNPs to assess differences in the genetic compositions of individuals among the populations.

3.2.7. Demographic history

I used SMC++ v1.15.2 (Terhorst et al., 2017) with the number of spline knots set to 25 to estimate the demographic history of each population and population divergence times using the SNP data, excluding SNPs on the scaffolds < 100 kb in length. Previously published genome-wide mutation rates in dipteran, lepidopteran, and hymenopteran insects are similar, ranging from 2.8e-09 to 3.4e-09 per base per haploid genome per generation (Keightley et al., 2014, 2015; Liu et al., 2017; Yang et al., 2015). Because data on the mutation rate and generation time of *P. auratus* are lacking, I used the rate of

Heliconius melpomene (2.9×10^{-9} per generation; Keightley et al., 2015) as the mutation rate in the coalescent simulations and 2 years as the generation time (Tateno et al., 2020). Estimations were conducted 20 times with 20 iterations, and the median and 95% confidence intervals (CIs) of the estimated size history were calculated using the nonparametric bootstrap method. In addition, I estimated demographic parameters using fastsimcoal2.7 (Excoffier et al., 2013), based on the topology estimated by IQ-TREE2. The coalescent simulation requires constraints on the divergence time of one or more nodes of the tree. In the absence of prior information on the divergence times of *P. auratus* populations, I used the results of the SMC++ analysis to determine the upper constraints on divergence time and effective population size. Recent secondary contacts and past gene flow were designed for contiguous (or ancestral) populations in the current geographic distribution. To simplify the simulations, I assumed that each population had a constant size with no growth after population divergence; to disregard recombination rates by minimizing linkage disequilibrium among the SNPs used in the analysis (Excoffier et al., 2013), LD-based SNP pruning was conducted using the SNP data and PLINK 1.90 (Purcell et al., 2007) with the command --indep-pairwise 10000 10 0.2. The model was run 100 times with the following settings: number of coalescent simulations in each cycle, 100,000; expectation/conditional maximization (ECM) cycles, 40; minimum ECM cycles, 10; minimum observed site frequency spectrum (SFS) entry count, 10; and removal of monomorphic sites for SFS. The 95% CIs of the parameters were calculated using a non-parametric bootstrap method.

3.3 Results

3.3.1. Genome sequence of *P. auratus*

The genome size of *P. auratus* estimated by the GenomeScope analysis was 680,802,890 bp with 2% heterozygosity. The assembled draft genome contained 9,727 scaffolds with a total length of 867,702,560 bp (scaffold N50, 148,001 bp) and was thus longer than the estimated size, possibly due to insufficient removal of haplotigs. Gene prediction resulted in the detection of 30,643 transcripts in 28,887 protein-coding regions from the draft genome; 30,252 (98.72%) transcripts were annotated in the BLAST search. BUSCO score quality assessment revealed 1,810 (85.2%) complete single-copy, 226 (10.6%) duplicated, 30 (1.4%) fragmented, and 58 (2.8%) missing BUSCOs. Variant calling against the draft genome for all resequencing data (n = 27) resulted in the assembly of a dataset with 9,184,739 SNPs in 8,838 scaffolds (rate of missing SNP data, < 0.01%). I used this dataset in the subsequent analyses with and without SNP pruning for LD. After LD pruning, 247,662 SNPs remained in the dataset (rate of missing SNP data, 3.51%).

3.3.2. Phylogenetic and population genetic structural analyses

The phylogenetic analysis revealed that individuals from WR, ER, and the southern populations (SI, SG, and SR) formed different clades. Within the southern populations, one SG individual (SG05) was included in the clade of SI individuals, and one SG individual (SG06) was included in the clade of SR individuals. The population structural analysis (Figure 2) showed that the optimal number of ancestral groups (K) was 2 based on cross-validation error values (0.75 at K = 2; 0.88 at K = 3; 1.04 at K = 4; 1.23 at K = 5); individuals were divided into one group including the WR and ER populations and another including the SI, SG, and SR populations. However, individuals from WR and ER were clearly discriminated at K = 3 (Figure 2). The first two principal components showed genetic differences among the three groups: WR, ER, and the southern population

group (SI, SG, and SR) (proportions of variance explained: PC1, 10.90%; PC2, 10.05%; Figure 3). Although there was little genetic divergence among the southern populations, the genetic divergence between SR and SI+SG individuals was also shown by the PC1–PC2 plot. These population genetic structure results were similar to my previous RAD-seq results (Chapter 2).

3.3.3. Demographic history

I first estimated the demographic history of each population and the divergence times between pairs of populations based on the SNP data using SMC++. The temporal changes in the effective population size were consistently similar among all populations until approximately 50,000 years ago (Figure 4). The estimation of divergence times showed that WR and ER diverged from the lineage of the southern populations (SI, SG, and SR) approximately 27,000–47,000 years ago and a subsequent divergence among populations occurred approximately within the last 9,000 years (Table S1).

I performed a coalescent simulation with fastsimcoal2 to reconstruct the population divergence history with gene flow using a model with the upper limit of the effective population size set to 100,000 and that of the first divergence event from the most recent common ancestor (MRCA) set to 30,000 generations (60,000 years) ago, based on the results of the SMC++ analysis. Parameter estimation using the demographic model revealed the following demographic histories. WR and A1 (MRCA of ER, SI, SG, and SR) diverged 3,830 (95% CI, 3,410–4,220) years ago, ER and A2 (MRCA of SI, SG, and SR) diverged 2,086 (95% CI, 1,788–2,326) years ago, SR and A3 (MRCA of SI and SG) diverged 607 (95% CI, 490–716) years ago, and SI and SG diverged 195 (95% CI, 174–208) years ago (Figure 5). For the effective population size, the A1 population

experienced a severe bottleneck; the A2 and ER populations showed slight increases; and further population increases occurred in SI, SG, and SR (Figure 5, Table S2). Although I assumed continuous gene flow between populations, the estimated effective number of migrants (N_em) was much lower than unity (10^{-10} to 10^{-7}) for all cases (Figure 5, Table S3), suggesting a lack of substantial gene flow (Slatkin, 1987).

3.4. Discussion

3.4.1. Demographic histories of the geographic populations

My demographic analysis revealed the history of divergence of the five geographic populations of *P. auratus*. The SMC++ results revealed long-term historical changes in the effective population sizes. The similarity in the patterns of demographic change among populations may reflect the low degree of genetic differentiation among populations until recently (Figure 4). Application of the split function in SMC++ revealed that WR diverged from the other populations first, approximately 49,000 years ago (Table S1). This result is consistent with the results of population structural analysis (Figure 2) and my previous finding that WR diverged deeply from the other populations (Chapter2). However, the divergence times estimated by SMC++ were much older than those estimated by fastsimcoal2. This difference may have resulted from methodological differences between the two programs. First, SMC++ completely ignores gene flow and estimates the demography of each population independently, and the estimated coalescent times are based solely on divergence in population size between populations (Terhorst et al., 2017); by contrast, fastsimcoal2 can consider gene flow between populations in demographic estimations (Excoffier et al., 2013). Second, SMC++ can use linkage information among SNPs in demographic estimations to consider linkage disequilibrium

(LD) and allow maximal use of SNP data (Terhorst et al., 2017), whereas fastsimcoal2 generally requires minimizing linkage disequilibrium in the SNP data by pruning linked SNPs to reduce the complexity of models and render the simulation feasible (Excoffier et al., 2013). These methodological differences can result in discrepancies in the estimation of demographic history. Regarding the demographic estimation, sequential Markovian coalescent (SMC)-based methods such as SMC++ are reliable for estimating relatively ancient demographic histories, whereas SFS-based methods such as fastsimcoal2 are reliable for estimating recent demographic histories (Patton et al., 2019). Accordingly, the divergence time estimation using SMC++, which is based on changes in estimated population sizes, may be inaccurate for recent population divergences. Further analyses are needed to understand the differences in divergence times estimated using different approaches.

The coalescent simulation conducted with fastsimcoal2 and the incorporation of gene flow in the model revealed that no substantial gene flow among the five populations had occurred since their divergence (Figure 5), suggesting that the geographic color variation of *P. auratus* is maintained mainly by barriers to gene flow. In Chapter 2, I inferred the occurrence of frequent gene flow among the three southern populations (SR, SG, and SI) and the presence of hybrid zones between WR and SG and between ER and SR (Chapter 2), and the genetic similarity among southern populations was also supported by the population structural analysis conducted in this study. However, the fastsimcoal2 analysis revealed that the divergence of the southern populations occurred within the past 600 years, implying that an insufficient number of unique genetic variations had accumulated in each population to detect divergence between local populations through population clustering analysis, despite the establishment of geographic isolation.

3.4.2 Factors restricting gene flow

Although *P. auratus* has dispersal potential by means of flight, they are confined to mountain forest areas and do not occur in wide open lands between lowland forests. The distribution of *P. auratus* among separate mountain areas in the Kinki District reflects this habitat selection pattern (Fig. 1a). Forest fragmentation can also lead to a decline in the supply of food resources for *P. auratus*, as forests are also the main habitat of large mammals in the Japanese islands. Therefore, forest fragmentation would have affected population divergence in this species.

The confinement of *P. auratus* to forest areas may also be related to the high risk of predation during flight outside of forests. The metallic coloration of beetles may be effective for predator avoidance when they fly among light and dark forest patches (Schultz, 1986). However, when beetles fly out into open land, they are easily tracked by birds. Thus, migration between patchy forests is likely to be difficult. The cost of subsocial behavior on parents with larval nests may also restrict the time allowed for dispersal by flight. Large dung beetles that construct deep nests and lay few, large eggs are expected to have low dispersal rates (Hanski & Cambefort, 2014). This trend is likely to occur in *P. auratus*, as females build nests at depths of 50–100 cm and lay only one large egg in each nest (Tateno et al., 2020); this relatively sedentary lifestyle may limit opportunities for dispersal and migration.

3.4.3 Factors influencing population divergence

My fastsimcoal2 analysis showed that *P. auratus* populations in the Kinki District diverged about 3,800 years ago and more recently (Figure 5). Although the Holocene

climate has been relatively stable and warm (Petit et al., 1999), minor local climatic changes (± 2°C) have occurred several times (Seppä et al., 2009; L. B. Wang et al., 2011). In Eurasia, marked climate cooling occurred 4,200–4,000 years ago, which may have caused the decline of several human civilizations, including the Jomon population in Japan (Yasuda et al., 2004). Japan also experienced climatic cooling 3,800–3,600 years ago (Kawahata et al., 2009; Yasuda et al., 2004). These cooling events approximately 4,200–3,600 years ago may have affected the population divergence of *P. auratus* approximately 3,800 years ago. The time of the second divergence event (2,000 years ago) coincides roughly with the beginning of the rapid increase in the human population in Japan due to the introduction of rice farming, which provided a stable food resource (Biraben, 2006; Schiffels & Durbin, 2014). Thus, population divergence 2,000 years ago and more recently may have been caused by habitat fragmentation due to declines in the populations of large mammals, such as deer, under the influence of human activities such as hunting and exploitation. The increased human population size may have reduced wild animal populations and fragmentated habitat (Y. Watanabe et al., 2019).

Declines in mammal populations due to hunting may have led to declines in the species richness and abundance of dung beetles (Andresen & Laurance, 2007). *Phelotrupes auratus* is thought to depend largely on sika deer dung, due to its overlapping distribution with sika deer (Toda & Akei, 2003). In addition, *P. auratus* likely has limited dispersal propensity due to its reproductive ecology. Thus, reduction and fragmentation of the sika deer population may have strongly affected the *P. auratus* population. Although *P. auratus* can also use dung produced by other large mammals as food resources, archaeological records for the Jomon period (1,200–2,400 years ago) and documents mentioning local products from the mid-Edo era (1730s) have shown that sika

deer and other large mammals such as wild boars had reduced their distribution areas by the 18th century (Tsujino et al., 2010). A genetic structural analysis also suggested that the sika deer population in Japan has experienced habitat fragmentation and bottlenecks due to human activity since the mid-19th century (Goodman et al., 2001). To test my hypothesis that human activity has impacted large mammal and dung beetle populations, it is necessary to study the demographic history of large mammals that provide dung food for *P. auratus* with sufficient temporal and spatial resolution.

3.4.4 Divergence of color morphs among local populations

Although I detected marked divergence of local populations with different coloration within short periods in this study, it remains uncertain whether different structural coloration has adaptive importance (Chapter 2). In a previous study, a Müllerian mimicry hypothesis was proposed based on the similarity of geographic color variation patterns between *P. auratus* and a sympatric congener, *Phelotrupes laevistriatus* (Watanabe et al., 2002). However, carcasses of the these beetles are often found in the feces of mammals such as raccoon dogs (Matsuyama et al., 2006) and chickens have been observed to feed on them without any problems (Kochi, 2001), implying that the *Phelotrupes* beetles are not unpalatable. As mentioned above, the metallic coloration of *P. auratus* may be effective for avoiding bird predation in forests, where they fly among light and dark patches (Schultz, 1986). However, color differences may not be important in this predator avoidance strategy. If differences in the structural coloration of *P. auratus* are not important, then color differences may have been fixed by chance. A reduction in the effective population size of *P. auratus* was occurring when the indigo and green forms diverged (Figure 5), which may indicate that this color divergence was driven by genetic

drift.

3.5. Conclusion

In this study, I conducted a detailed genomic analysis of genetic information from five local populations of *P. auratus*. My analyses revealed that geographic color variation in *P. auratus* is of very recent origin and is maintained by restricted gene flow. The recent divergence of these populations may have been influenced by increased human activity in the past few thousand years, providing an interesting example of the indirect impact of human activity on trait divergence in wild insect populations. My study implies the need for additional research conducted with a population genomics approach to reveal the historical relationships between the trophic levels, namely large mammals and dung beetles. Future studies of the demographic history of sika deer and other large mammals using genomic data will be essential to our understanding of the process of recent *P. auratus* population differentiation.

Figures and Tables

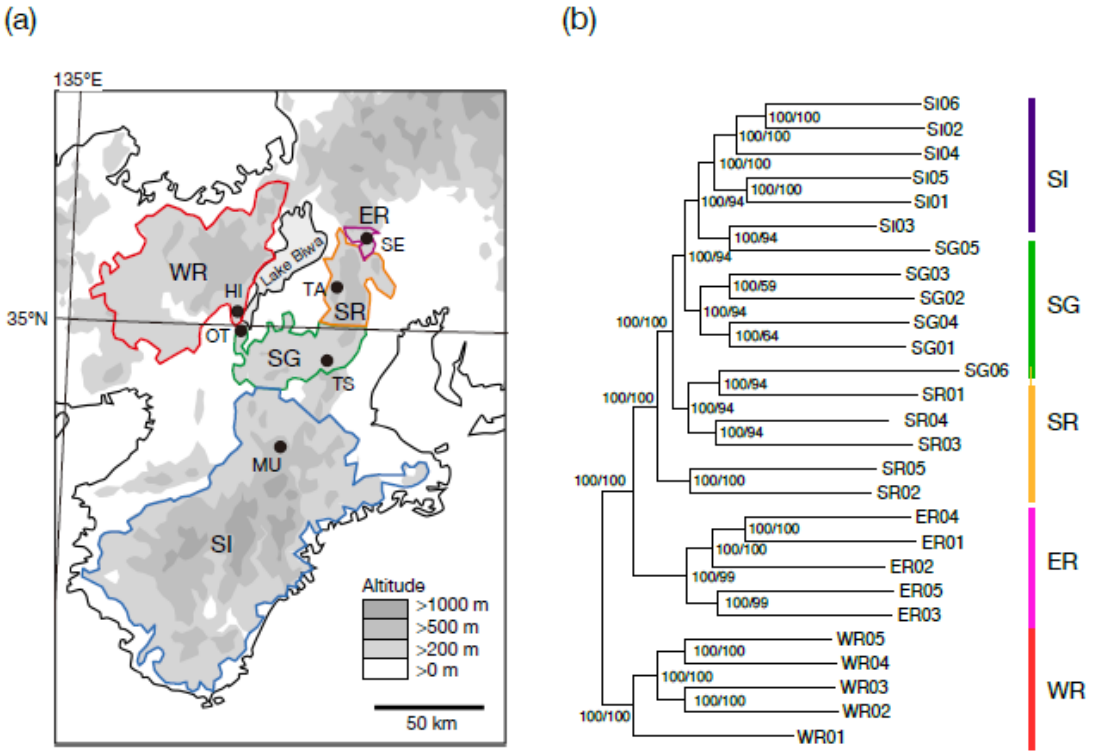

Figure 1.1. (a) Distribution areas and sampling sites of *P. auratus* in the Kinki District of Honshu, Japan. Distribution areas (encircled with different colors): red, west/red; purple, east/red; blue, south/indigo; green, south/green; orange, south/red. For sampling sites (closed circles), see Table 1. (b) Maximum likelihood tree of *P. auratus* individuals from the five local populations. Node support values indicated on branches are SH-aLRT and UF-bootstrap values (percentages).

Figure 3.2. Genetic population structure of *P. auratus* individuals at each site. Orange, ancestry 1; blue, ancestry 2; purple, ancestry 3. The optimal number of clusters (*K*) was found to be 2.

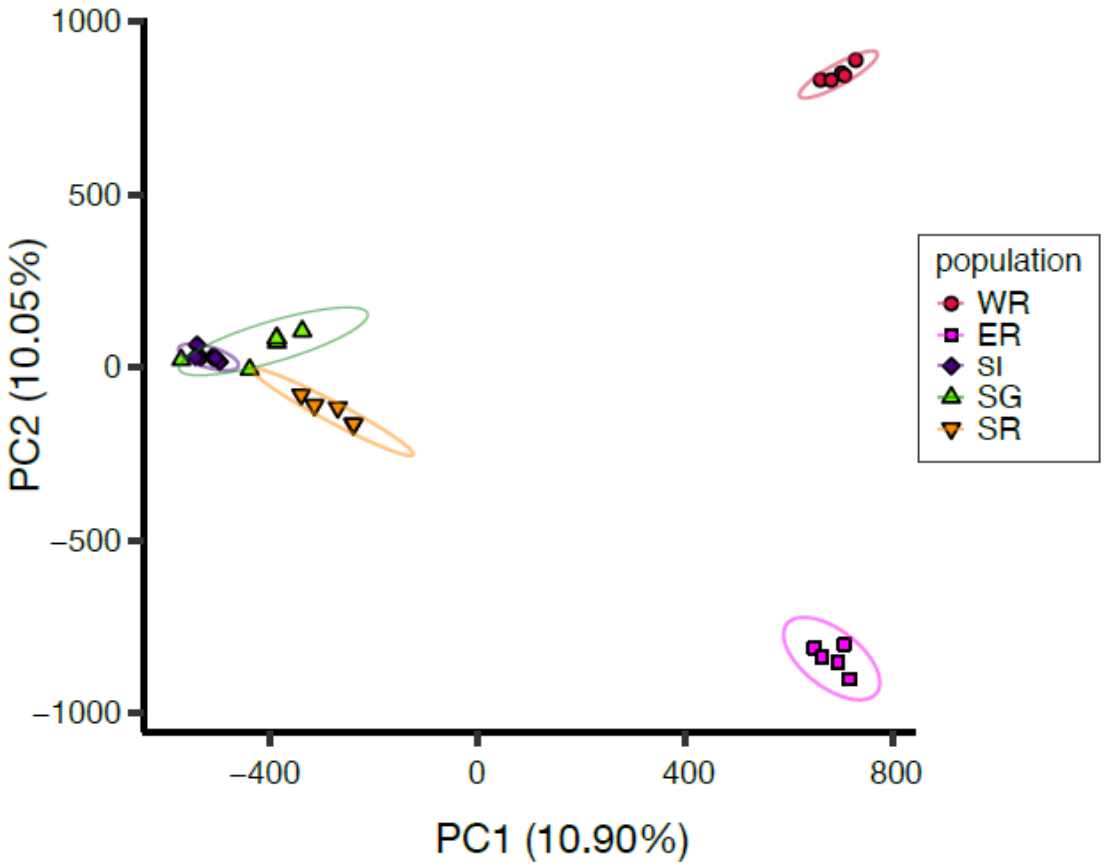

Figure 3.3. Principal components 1 (PC1) and 2 (PC2) scores of *P. auratus* individuals determined from principal component analyses of single-nucleotide polymorphisms. Local population: red circles, west/red; magenta squares, east/red; blue diamonds, south/indigo; green triangles, south/green; inverted pink triangles, south/red. Ellipses, 95% confidence ellipses of PC scores for each population.

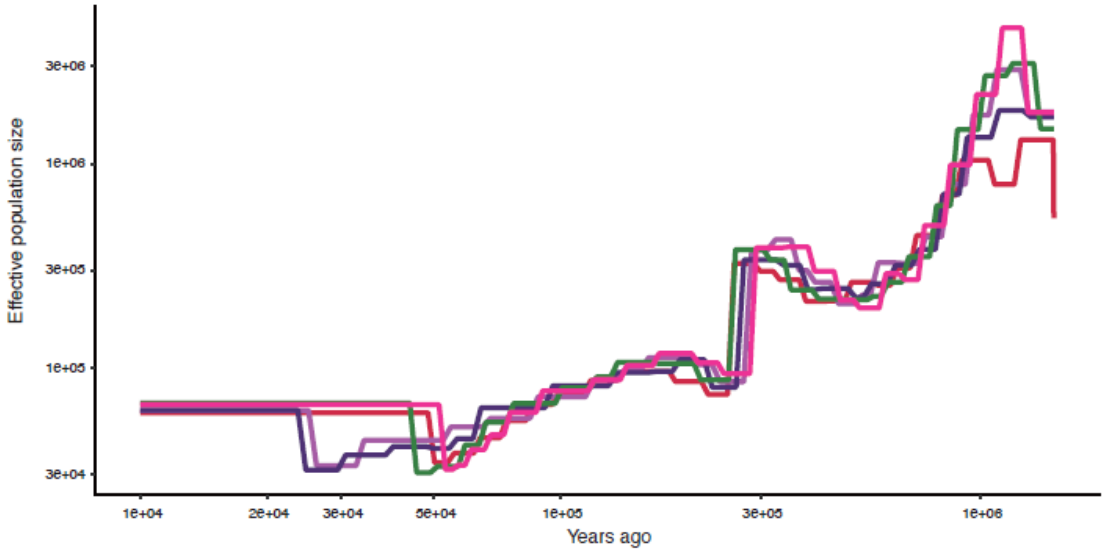

Figure 3.4. Historical changes in effective population sizes, inferred by SMC++. Line colors represent local populations: red, west/red; purple, east/red; blue, south/indigo; green, south/green; orange, south/red.

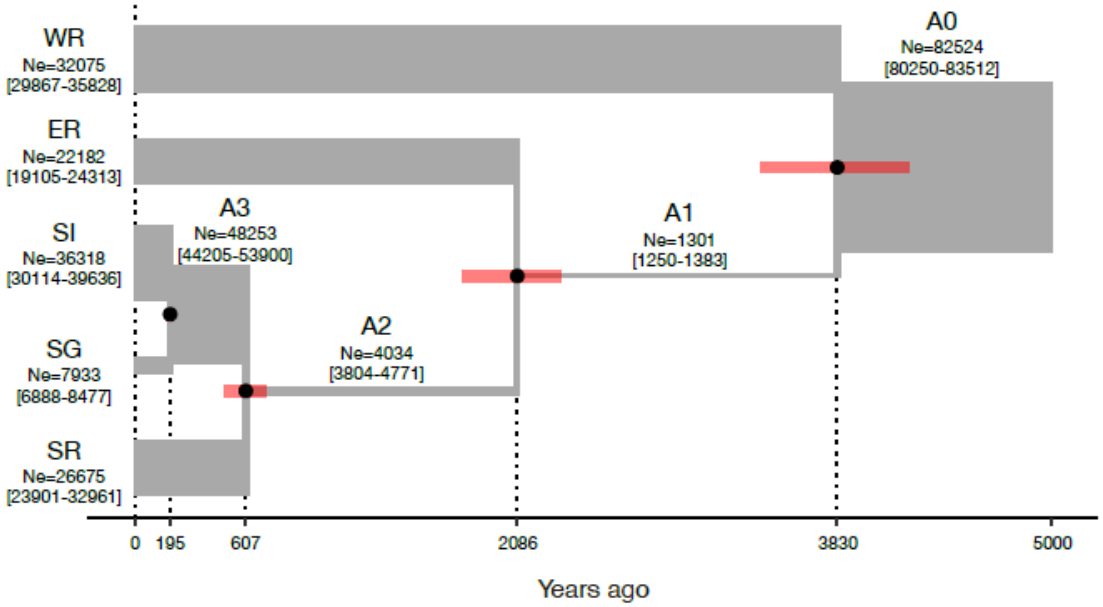

Figure 5. Demographic history of *P. auratus* populations inferred by coalescent simulation using fastsimcoal2 during the past 5,000 years. Horizontal bar widths indicate effective population sizes, N_e (median and 95% confidence intervals are indicated by numerals); black circles and red vertical bars indicate medians and 95% confidence intervals, respectively, of divergence times.

Table 3.1. *Phelotrupes auratus* individuals from which genomic DNA were extracted in this study.

Individual ID	Color: λmax(α) (nm)	Population group	Population; locality name; Latitude (°N), longitude (°E)
WR01	631.81	West/Red (WR)	HI; Mt. Hiei, Kyoto; 35.0640, 135.8224
WR02	632.14	”	”
WR03	620.05	”	”
WR04	660.79	”	”
WR05	642.28	”	”
ER01	665.09	East/Red (ER)	SE; Sekigahara, Gifu; 35.2627, 136.5234
ER02	677.06	”	”
ER03	664.75	”	”
ER04	645.28	”	”
ER05	673.21	”	”
SI01	483.45	South/Indigo (SI)	MU; Murou, Nara; 34.5391, 136.0285
SI02	494.14	”	”
SI03*	526.91	”	”
SI04	517.73	”	”
SI05	505.14	”	”
SI06	486.39	”	”
SG01	552.90	South/Green (SG)	OT; Mt. Otowa, Kyoto; 34.9768, 135.8530
SG02	579.57	”	”
SG03	638.50	”	”
SG04	589.58	”	”
SG05	596.04	”	TS; Tsuge, Mie; 34.8542, 136.2750
SG06	516.80	”	”
SR01	598.64	South/Red (SR)	TA; Mt. Takatori, Shiga; 35.2804, 136.3768
SR02	578.78	”	”
SR03	633.15	”	”
SR04	632.81	”	”
SR05	612.19	”	”

*, used for constructing the reference genome sequence.

Supplemental information

Table S3.1. Divergence time between populations of *P. auratus* estimated by SMC++.

Population pair	Divergence time Median (95% CI), years ago
WR – ER	48,901 (46,539-54,277)
WR – SI	41,873 (38,513-46,732)
WR - SG	33,656 (30,689-34,941)
WR – SR	40,927 (39,536-43,338)
ER – SI	35,282 (33,900-38,504)
ER – SG	33,499 (32,201-34,810)
ER – SR	27,446 (27,061-27,992)
SI – SG	6,753 (6,338-6,936)
SI – SR	8,950 (8,022-9,364)
SG - SR	5,742 (2,454-9,004)

Table S3.2. Medians and 95% CIs of estimated effective population sizes.

Population	median	(95% CI)
west/red	32,075	(29,867-35,828)
east/red	22,182	(19,105-24,313)
south/indigo	36,318	(30,114-39,636)
south/green	7,933	(6,888-8,477)
south/red	26,675	(23,901-32,961)
A0	82,524	(80,250-83,512)
A1	1,301	(1,250-1,383)
A2	4,034	(3,864-4,771)
A3	48,253	(44,205-53,900)

Table S3.3. Median and 95% CI of estimated effective number of migrants per generation (N_em).

Migration	From	To	Period	N_em (95% CI)
NeM123	SI	SG	Present–T1	2.00e-08 (1.71e-09–4.62e-07)
NeM132	SG	SI	Present–T1	1.55e-08 (1.53e-09–2.43e-07)
NeM134	SG	SR	Present–T1	6.30e-09 (7.33e-10–1.19e-07)
NeM143	SR	SG	Present–T1	4.52e-08 (7.81e-09–7.97e-07)
NeM224	A3	SR	T1-T2	1.10e-07 (5.54e-09–9.63e-07)
NeM242	SR	A3	T1-T2	4.69e-09 (6.25e-10–3.42e-08)

Chapter 4. General Discussion

4.1. Geographic color variation and hybrid zones of *P. auratus*

The correlation between color forms and population genetic structure or phylogenetic relationships in *P. auratus* was not clear in previous studies (Akamine, Maekawa, et al., 2011). My work using RAD-seq data divided *P. auratus* in the Kinki District into five local populations, as described above (WR, ER, SI, SG, and SR). Furthermore, SG and SR exhibited narrow and long-lasting hybrid zones in the boundary areas, with WR and ER belonging to a different genetic cluster. Therefore, gene flow was likely restricted by some barriers between the south cluster and other two clusters. The obscure population structure seen among SI, SG, and SR may be the result of either active gene flow or very recent population divergence. The results of the demographic analysis revealed that these three local populations were divided very recently, and there have been restricted gene flow among them. Although the factors preventing gene flow or color homogenization are still unclear, it is possible that some ecological characteristics of *P. auratus* prevent the migration of individuals between habitats.

The geographic color variation among local populations may be attributed to the effects of genetic drift or geographically specific natural selection. The demographic analysis based on the whole-genome SNP data showed that SI and SG were the most recently diverged local populations which showed decreases in effective population size, implying that the indigo and green forms are derived traits that the two south populations acquired independently due to the effect of genetic drift. On the other hand, it is difficult to evaluate the effects of natural selection because the function of coloration in *P. auratus* is still unclear. Therefore, it is necessary to identify the genes or genetic regions

responsible for color determination to understand how each color became fixed in each population. GWAS analysis using RAD-SNPs implied that the coloration of *P. auratus* was controlled by multiple loci. This analysis also found that a candidate gene encoding thioredoxin may be involved in the regulation of melanin production and structural coloration (Castro-Sowinski et al., 2007; L. Zhang et al., 2017). However, GWAS results must be interpreted with caution, as a hidden population genetic structure was revealed by whole-genome analysis in the five local populations considered in this work. It is challenging to detect the genetic regions responsible for coloration in these populations with existing GWAS methods, because mutations that are fixed within populations by chance could be false positives. Existing GWAS methods also have great difficulty identifying candidate genetic regions when traits are controlled by common variants with small effects or by rare variants with large effects (Korte & Farlow, 2013). To understand the genetic basis of the coloration of *P. auratus*, it is necessary to use different approaches such as quantitative trait locus (QTL) analysis.

4.2. Demographic history of *P. auratus* in the Kinki District

In various species, phylogenetic divergence occurred during the Last Glacial Maximum as a result of isolation into multiple refugia (Hewitt, 2004). However, the evolutionary events that have taken place since the beginning of the Holocene are not well understood. The Holocene was a period of relatively stable climate (Petit et al., 1999), and it is likely that human activities, along with geographical events, had a significant impact on the evolution of various organisms (Ceballos et al., 2015; Sullivan et al., 2017). However, minor climatic changes have occurred several times (Seppä et al., 2009; L. B. Wang et al., 2011). In particular, a cold climate between 4,200–3,600 years ago might have led to

a reduced food supply, causing the decline of several human civilizations (Kawahata et al., 2009; Yasuda et al., 2004). This timing roughly coincides with the first population divergence of *P. auratus* in the Kinki District, about 3,800 years ago. In the Kinki District, no major geographical events are known to have occurred in the last 4,000 years. Therefore, the subsequent population divergence that occurred in the past 2,000 years in this group may have been caused by increased human activity. The period around 2,000 years ago, when the second oldest population divergence occurred, coincides with the period when the population size of Japanese people began to increase dramatically (Biraben, 2006; Schiffels & Durbin, 2014). This explosive increase may have accelerated the pressure of anthropogenic activities on ecosystems, resulting in habitat fragmentation and the decline of large herbivores providing food resources.

Using a larger genomic dataset obtained by whole-genome resequencing, this study revealed that the divergence of differently colored populations occurred within the last 5,000 years and that there is currently no substantial gene flow among these populations. This restricted gene flow could be due to the dependence of *P. auratus* on the forest habitats. They depend on forest-living mammals for food (dung) supply, and migration between patchy forests across open lands may bear a high risk of predation. Thus, the forest fragmentation may have affected the population dynamics of *P. auratus* significantly.

4.3. Perspective

The cause of the restricted gene flow between sister populations remains unclear, but it may be due to limited migration between segregating habitats across open low lands. In

order to test this hypothesis, we need to accumulate more fundamental information on *P. auratus*, including detailed distribution areas and behavioral characteristics. On the other hand, it is important to identify the genetic regions responsible for controlling coloration and to clarify the evolutionary processes that have occurred in those regions to understand the reasons for the independently fixed colorations in different populations. The *P. auratus* populations studied for this thesis were not suitable for identifying responsible genetic regions in control of coloration via GWAS as different color forms occurred in genetically diverged local populations. Detection of such genes may be possible by GWAS only by using a single local population with multiple color polymorphisms, which would allow the effects of geographic genetic divergence to be ignored, or by quantitative trait locus (QTL) analysis using F2 hybrids or backcross individuals of different color forms.

This research implies that the population dynamics of *P. auratus* have been controlled directly by large herbivores and indirectly by human activities, which may have heavily impacted the herbivore populations. However, the dependency of *P. auratus* on large herbivores, in particular sika deer, is inferred only based on their overlapping distributions (Toda & Akei, 2003) and has not been tested either empirically through behavioral experiments or theoretically through genetic analysis. Therefore, it is important to understand how *P. auratus* uses animal dung in the wild, and to analyze the population histories of large mammals to evaluate the impact of their population dynamics on the population divergence of *P. auratus*. This would contribute to understanding of not only the mechanisms of population divergence in *P. auratus* but also how the increasing sika deer populations, which have become a problem in Japan, may affect local ecosystems.

4.4. Conclusions

In summary, this study demonstrates that the geographic color variation of *P. auratus* in the Kinki District has been driven by the rapid population divergence that occurred recently. The population clustering analysis revealed that they diverged into five isolated local populations by the combination of color form and genetic cluster, and the genetic differences between three population were obscure despite they showed different body colors (Chapter 2). The loss of genetic structure could be caused by frequent gene flow or very recent population differentiation. Demographic analysis showed that these populations diverged within the last 3,800 years and that gene flow between the sister populations was restricted (Chapter 3). This rapid population divergence was likely caused by the habitat fragmentation and the decline of large-mammal populations supplying food due to both climatic cooling and increasing human activity. Thus, the geographic color variation in *P. auratus* may have been acquired due to their isolation in patchy forests and may have been maintained due to the limitation of migration among segregated habitats. In this study, I provided an interesting example demonstrating that geographic population divergence in an insect with conspicuous differences in a trait (body color) may have occurred during a short period under the influence of human activity. In conclusion, the five local populations of *P. auratus* are in the earliest stages of speciation—namely, the process of acquiring novel genetic traits—and thus provide an important model for the study of organismal diversification.

Acknowledgements

I would like to thank my supervisor, Teiji Sota, with great sincerity for his helpful guidance and advice in my studies. I would also like to be grateful to Tomochika Fujisawa, Satoshi Yamamoto, and Shota Nomura for technical advises on genomic analysis, and Katsutoshi Watanabe, Craig Barnett, Masaki Hoso, Yuta Morii, and the past and present members of Animal Ecology Laboratory of Kyoto University for various supports, discussions, and comments. My studies were supported by JSPS KAKENHI (nos. 15H02637, 18H04010 to Teiji Sota).

References

Akamine, M., Ishikawa, K., Maekawa, K., & Kon, M. (2011). The physical mechanism of cuticular color in *Phelotrupes auratus* (Coleoptera, Geotrupidae). *Entomological Science*, *14*(3), 291–296. https://doi.org/10.1111/j.1479-8298.2011.00448.x

Akamine, M., Maekawa, K., & Kon, M. (2008). Geographic color variation of *Phelotrupes auratus* (Coleoptera, Geotrupidae) in the Kinki region, central Japan: A quantitative spectrophotometric analysis. *Entomological Science*, *11*(4), 401–407. https://doi.org/10.1111/j.1479-8298.2008.00286.x

Akamine, M., Maekawa, K., & Kon, M. (2011). Phylogeography of Japanese population of *Phelotrupes auratus* (Coleoptera, Geotrupidae) inferred from mitochondrial DNA sequences. *Zoological Science*, *28*(9), 652–658. https://doi.org/10.2108/zsj.28.652

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Andresen, E., & Laurance, S. G. (2007). Possible indirect effects of mammal hunting on dung beetle assemblages in Panama. *Biotropica*, *39*(1), 141–146. https://doi.org/10.1111/j.1744-7429.2006.00239.x

Aoki, K., Tamaki, I., Nakao, K., Ueno, S., Kamijo, T., Setoguchi, H., Murakami, N., Kato, M., & Tsumura, Y. (2019). Approximate Bayesian computation analysis of EST-associated microsatellites indicates that the broadleaved evergreen tree Castanopsis sieboldii survived the Last Glacial Maximum in multiple refugia in Japan. *Heredity*, *122*(3), 326–340. https://doi.org/10.1038/s41437-018-0123-9

Arias, M., le Poul, Y., Chouteau, M., Boisseau, R., Rosser, N., Théry, M., & Llaurens, V. (2016). Crossing fitness valleys: empirical estimation of a fitness landscape associated with polymorphic mimicry. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1829), 20160391. https://doi.org/10.1098/rspb.2016.0391

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A., & Saunders, N. C. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, *18*(1), 489–522. https://doi.org/10.1146/annurev.es.18.110187.002421

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, *3*(10), e3376. https://doi.org/10.1371/journal.pone.0003376

Baldassarre, D. T., White, T. A., Karubian, J., & Webster, M. S. (2014). Genomic and morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical introgression of a sexual signal. *Evolution*, *68*(9), 2644–2657. https://doi.org/10.1111/evo.12457

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*(4), 729–744. https://doi.org/10.1046/j.1365-294X.2003.02063.x

Bandelt, H. J., Forster, P., & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*(1), 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Barnett, D. W., Garrison, E. K., Quinlan, A. R., Stŕmberg, M. P., & Marth, G. T.

(2011). Bamtools: A C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics*, *27*(12), 1691–1692. https://doi.org/10.1093/bioinformatics/btr174

Barton, N. H., & Gale, K. S. (1993). Genetic analysis of hybrid zones. In *Hybrid zones and the evolutionary process* (pp. 13–45).

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, *16*(1), 113–148. https://doi.org/10.1146/annurev.es.16.110185.000553

Bemmels, J. B., Haddrath, O., Colbourne, R. M., Robertson, H. A., & Weir, J. T. (2022). Legacy of supervolcanic eruptions on population genetic structure of brown kiwi. *Current Biology*, *32*(15), 3389-3397.e8. https://doi.org/10.1016/j.cub.2022.05.064

Biraben, J.-N. (2006). The history of the human population from the first beginnings to the present day. In G. Caselli, J. Vallin, & G. Wunsch (Eds.), *Demography: Analysis and Synthesis* (Vol. 3, pp. 5–17). Elsevier. https://link.gale.com/apps/doc/CX2898300100/GVRL?u=kyotodai&sid=bookmark-GVRL&xid=a76b7ac7

Biró, L. P., Bálint, Z., Kertész, K., Vértesy, Z., Márk, G. I., Horváth, Z. E., Balázs, J., Méhn, D., Kiricsi, I., Lousse, V., & Vigneron, J. P. (2003). Role of photonic-crystal-type structures in the thermal regulation of a Lycaenid butterfly sister species pair. *Physical Review E*, *67*(2), 7. https://doi.org/10.1103/PhysRevE.67.021907

Boivin, N. L., Zeder, M. A., Fuller, D. Q., Crowther, A., Larson, G., Erlandson, J. M., Denhami, T., & Petraglia, M. D. (2016). Ecological consequences of human niche construction: Examining long-term anthropogenic shaping of global species

distributions. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(23), 6388–6396. https://doi.org/10.1073/pnas.1525200113

Bolnick, D. I., & Fitzpatrick, B. M. (2007). Sympatric speciation: models and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics*, *38*(1), 459–487. https://doi.org/10.1146/annurev.ecolsys.38.091206.095804

Bosi, S. G., Hayes, J., Large, M. C. J., & Poladian, L. (2008). Color, iridescence, and thermoregulation in Lepidoptera. *Applied Optics*, *47*(29), 5235. https://doi.org/10.1364/AO.47.005235

Brelsford, A., Milá, B., & Irwin, D. E. (2011). Hybrid origin of Audubon's warbler. *Molecular Ecology*, *20*(11), 2380–2389. https://doi.org/10.1111/j.1365-294X.2011.05055.x

Brien, M. N., Enciso-Romero, J., Parnell, A. J., Salazar, P. A., Morochz, C., Chalá, D., Bainbridge, H. E., Zinn, T., Curran, E. V., & Nadeau, N. J. (2018). Phenotypic variation in *Heliconius erato* crosses shows that iridescent structural colour is sex-linked and controlled by multiple genes. *Journal of the Royal Society Interface Focus*, *9*(1), 20180047. https://doi.org/10.1098/rsfs.2018.0047

Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M., & Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, *3*(1), 1–11. https://doi.org/10.1093/nargab/lqaa108

Buerkle, C. A. (2005). Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes*, *5*(3), 684–687. https://doi.org/10.1111/j.1471-8286.2005.01011.x

Cambefort, Y. (2014). Dung beetle in tropical savannas. In Illka Hanski & Y.

Cambefort (Eds.), *Dung Beetle Ecology* (pp. 156–178). Princeton University Press.

Campbell, D. R. (2004). Natural selection in Ipomopsis hybrid zones: implications for ecological speciation. *New Phytologist*, *161*(1), 83–90. https://doi.org/10.1046/j.1469-8137.2003.00919.x

Capblancq, T., Després, L., & Mavárez, J. (2020). Genetic, morphological and ecological variation across a sharp hybrid zone between two alpine butterfly species. *Evolutionary Applications*, *June 2019*, eva.12925. https://doi.org/10.1111/eva.12925

Castro-Sowinski, S., Matan, O., Bonafede, P., & Okon, Y. (2007). A thioredoxin of *Sinorhizobium meliloti* CE52G is required for melanin production and symbiotic nitrogen fixation. *Molecular Plant-Microbe Interactions®*, *20*(8), 986–993. https://doi.org/10.1094/MPMI-20-8-0986

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Science Advances*, *1*(5), 9–13. https://doi.org/10.1126/sciadv.1400253

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cook, L. M., & Saccheri, I. J. (2013). The peppered moth and industrial melanism : evolution of a natural selection case study. *Heredity*, *110*(3), 207–212. https://doi.org/10.1038/hdy.2012.92

Curran, E. V., Stankowski, S., Pardo-Diaz, C., Salazar, C., Linares, M., & Nadeau, N. J. (2020). Müllerian mimicry of a quantitative trait despite contrasting levels of

genomic divergence and selection. *Molecular Ecology*, *29*(11), 2016–2030. https://doi.org/10.1111/mec.15460

Davis, A. L. V., Brink, D. J., Scholtz, C. H., Prinsloo, L. C., & Deschodt, C. M. (2008). Functional implications of temperature-correlated colour polymorphism in an iridescent, scarabaeine dung beetle. *Ecological Entomology*, *33*(6), 771–779. https://doi.org/10.1111/j.1365-2311.2008.01033.x

Derkarabetian, S., Burns, M., Starrett, J., & Hedin, M. (2016). Population genomic evidence for multiple Pliocene refugia in a montane-restricted harvestman (Arachnida, Opiliones, Sclerobunus robustus) from the southwestern United States. *Molecular Ecology*, *25*(18), 4611–4631. https://doi.org/10.1111/mec.13789

Derryberry, E. P., Derryberry, G. E., Maley, J. M., & Brumfield, R. T. (2014). HZAR : hybrid zone analysis using an R software package. *Molecular Ecology Resources*, *14*(3), 652–663. https://doi.org/10.1111/1755-0998.12209

Dupanloup, I., Schneider, S., & Excoffier, L. (2002). A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, *11*(12), 2571–2581. https://doi.org/10.1046/j.1365-294X.2002.01650.x

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology and Evolution*, *29*(1), 51–63. https://doi.org/10.1016/j.tree.2013.09.008

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome*, *4*(3), 250–255. https://doi.org/10.3835/plantgenome2011.08.0024

Endler, J. A. (1977). *Geographic variation, speciation, and clines*. Princeton University Press.

Endler, J. A. (1978). A predator's view of animal color patterns. In *Evolutionary Biology* (pp. 319–364). Springer US. https://doi.org/10.1007/978-1-4615-6956-5_5

Endler, J. A., & Mappes, J. (2017). The current and future state of animal coloration research. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1724), 20160352. https://doi.org/10.1098/rstb.2016.0352

Etter, P. D., Bassham, S., Hohenlohe, P. a, Johnson, E. a, & Cresko, W. a. (2011). Molecular methods for evolutionary genetics. In V. Orgogozo & M. V. Rockman (Eds.), *Molecular Methods for Evolutionary Genetics, Methods in Molecular Biology, vol. 772* (Vol. 772, Issue 6). Humana Press. https://doi.org/https://doi.org/10.1007/978-1-61779-228-1_9

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, *9*(10). https://doi.org/10.1371/journal.pgen.1003905

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Favila, M. E., Ruiz-Lizarraga, G., & Nolasco, J. (2000). Inheritance of a red cuticular color mutation in the scarab beetle *Canthon cyanellus cyanellus* LeConte (Coleoptera: Scarabaeidae). *The Coleopterists Bulletin*, *54*(4), 541–545. https://doi.org/10.1649/0010-065x(2000)054[0541:ioarcc]2.0.co;2

Fitzpatrick, B. M. (2012). Estimating ancestry and heterozygosity of hybrids using molecular markers. *BMC Evolutionary Biology*, *12*(1), 1–14. https://doi.org/10.1186/1471-2148-12-131

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(17), 9451–9457. https://doi.org/10.1073/pnas.1921046117

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929. https://doi.org/10.1111/2041-210X.12382

Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, *147*(2), 915–925. http://www.ncbi.nlm.nih.gov/pubmed/9335623

Gay, L., Crochet, P.-A., Bell, D. A., & Lenormand, T. (2008). Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. *Evolution*, *62*(11), 2789–2806. https://doi.org/10.1111/j.1558-5646.2008.00491.x

Giles, B. E., & Goudet, J. (1997). Genetic differentiation in *Silene dioica* metapopulations: Estimation of spatiotemporal effects in a successional plant species. *The American Naturalist*, *149*(3), 507–526. https://doi.org/10.1086/286002

Gompert, Z., & Buerkle, C. A. (2010). Introgress: A software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, *10*(2), 378–384. https://doi.org/10.1111/j.1755-0998.2009.02733.x

Gompert, Z., Mandeville, E. G., & Buerkle, C. A. (2017). Analysis of population genomic data from hybrid zones. *Annual Review of Ecology, Evolution, and Systematics*, *48*(1), 207–229. https://doi.org/10.1146/annurev-ecolsys-110316-022652

Goodman, S. J., Tamate, H. B., Wilson, R., Nagata, J., Tatsuzawa, S., Swanson, G. M., Pemberton, J. M., & McCullough, D. R. (2001). Bottlenecks, drift and differentiation: the population structure and demographic history of sika deer (*Cervus nippon*) in the Japanese archipelago. *Molecular Ecology*, *10*(6), 1357–1370. https://doi.org/10.1046/j.1365-294X.2001.01277.x

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Hanski, Ilkka, & Cambefort, Y. (2014). Spatial pricesses. In Ilkka Hanski & Y. Cambefort (Eds.), *Dung Beetle Ecology* (pp. 283–304). Princeton University Press.

Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*(6789), 907–913. https://doi.org/10.1038/35016000

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*(1442), 183–195. https://doi.org/10.1098/rstb.2003.1388

Hirase, S., Yamasaki, Y. Y., Sekino, M., Nishisako, M., Ikeda, M., Hara, M., Merilä, J., & Kikuchi, K. (2021). Genomic evidence for speciation with gene flow in broadcast spawning marine invertebrates. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msab194

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Hoekstra, H. E., Drumm, K. E., & Nachman, M. W. (2004). Ecological genetics of

adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. *Evolution*, *58*(6), 1329–1341. https://doi.org/10.1111/j.0014-3820.2004.tb01711.x

Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., & Stanke, M. (2016). BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*, *32*(5), 767–769. https://doi.org/10.1093/bioinformatics/btv661

Hoff, K. J., Lomsadze, A., Borodovsky, M., & Stanke, M. (2019). Whole-genome annotation with BRAKER. *Methods in Molecular Biology*, *1962*(0), 65–95. https://doi.org/10.1007/978-1-4939-9173-0_5

Hoyal Cuthill, J. F., & Charleston, M. (2015). Wing patterning genes and coevolution of Müllerian mimicry in Heliconius butterflies: Support from phylogeography, cophylogeny, and divergence times. *Evolution*, *69*(12), 3082–3096. https://doi.org/10.1111/evo.12812

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Kawahata, H., Yamamoto, H., Ohkushi, K., Yokoyama, Y., Kimoto, K., Ohshima, H., & Matsuzaki, H. (2009). Changes of environments and human activity at the Sannai-Maruyama ruins in Japan during the mid-Holocene Hypsithermal climatic interval. *Quaternary Science Reviews*, *28*(9–10), 964–974. https://doi.org/10.1016/j.quascirev.2008.12.009

Keightley, P. D., Ness, R. W., Halligan, D. L., & Haddrill, P. R. (2014). Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster*

full-sib family. *Genetics*, *196*(1), 313–320. https://doi.org/10.1534/genetics.113.158758

Keightley, P. D., Pinharanda, A., Ness, R. W., Simpson, F., Dasmahapatra, K. K., Mallet, J., Davey, J. W., & Jiggins, C. D. (2015). Estimation of the spontaneous mutation rate in *Heliconius melpomene*. *Molecular Biology and Evolution*, *32*(1), 239–243. https://doi.org/10.1093/molbev/msu302

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kjernsmo, K., Whitney, H. M., Scott-Samuel, N. E., Hall, J. R., Knowles, H., Talas, L., & Cuthill, I. C. (2020). Iridescence as camouflage. *Current Biology*, *30*(3), 551-555.e3. https://doi.org/10.1016/j.cub.2019.12.013

Knaus, B. J., & Grünwald, N. J. (2017). vcfr : a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44–53. https://doi.org/10.1111/1755-0998.12549

Knief, U., Bossu, C. M., Saino, N., Hansson, B., Poelstra, J., Vijay, N., Weissensteiner, M., & Wolf, J. B. W. (2019). Epistatic mutations under divergent selection govern phenotypic variation in the crow hybrid zone. *Nature Ecology and Evolution*, *3*(4), 570–576. https://doi.org/10.1038/s41559-019-0847-9

Kochi, A. (2001). The function of iridescence of the dung beetle *Geotrupes auratus*. *Konchu to Shizen*, *36*(10), 4–7. (In Japanese)

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, *9*(1), 29. https://doi.org/10.1186/1746-4811-9-29

Kroiss, J., Strohm, E., Vandenbem, C., & Vigneron, J.-P. (2009). An epicuticular

multilayer reflector generates the iridescent coloration in chrysidid wasps (Hymenoptera, Chrysididae). *Naturwissenschaften*, *96*(8), 983–986. https://doi.org/10.1007/s00114-009-0553-6

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*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Leigh, J. W., & Bryant, D. (2015). POPART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110–1116. https://doi.org/10.1111/2041-210X.12410

Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, *50*(6), 913–925. https://doi.org/10.1080/106351501753462876

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493–496. https://doi.org/10.1038/nature10231

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lin, C. P., & Danforth, B. N. (2004). How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets.

Molecular Phylogenetics and Evolution, *30*(3), 686–702. https://doi.org/10.1016/S1055-7903(03)00241-0

Liu, H., Jia, Y., Sun, X., Tian, D., Hurst, L. D., & Yang, S. (2017). Direct determination of the mutation rate in the bumblebee reveals evidence for weak recombination-associated mutation and an approximate rate constancy in insects. *Molecular Biology and Evolution*, *34*(1), 119–130. https://doi.org/10.1093/molbev/msw226

Lomsadze, A., Burns, P. D., & Borodovsky, M. (2014). Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Research*, *42*(15), 1–8. https://doi.org/10.1093/nar/gku557

Mallet, J., & Joron, M. (1999). Evolution of diversity in warning color and mimicry: polymorphisms, shifting balance, and speciation. *Annual Review of Ecology and Systematics*, *30*(1), 201–233. https://doi.org/10.1146/annurev.ecolsys.30.1.201

Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, *27*(6), 764–770. https://doi.org/10.1093/bioinformatics/btr011

Matsuyama, J., Hata, K., & Sone, K. (2006). Food habits of raccoon dogs, *Nyctereutes procyonoides* Gray, in Kagoshima Prefecture. *Research Bulletin of the Kagoshima University Forests*, *34*, 75–100. http://hdl.handle.net/10232/9105 (In Japanese)

McLean, C. A., & Stuart-Fox, D. (2014). Geographic variation in animal colour polymorphisms and its role in speciation. *Biological Reviews*, *89*(4), 860–873. https://doi.org/10.1111/brv.12083

Merrill, R. M., Chia, A., & Nadeau, N. J. (2014). Divergent warning patterns contribute to assortative mating between incipient Heliconius species. *Ecology and Evolution*, *4*(7), 911–917. https://doi.org/10.1002/ece3.996

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., Lanfear, R., & Teeling, E. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Mizuno, T. (1964). Geographic variation in the Japanese species of the genus Geotrupes. *Iden*, *18*(9), 24–27. (In Japanese)

Mullen, L. M., & Hoekstra, H. E. (2008). Natural selection along an environmental gradient: a classic cline in mouse pigmentation. *Evolution*, *62*(7), 1555–1570. https://doi.org/10.1111/j.1558-5646.2008.00425.x

Mullen, L. M., Vignieri, S. N., Gore, J. A., & Hoekstra, H. E. (2009). Adaptive basis of geographic variation: genetic, phenotypic and environmental differences among beach mouse populations. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1674), 3809–3818. https://doi.org/10.1098/rspb.2009.1146

Nosil, P., Harmon, L. J., & Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, *24*(3), 145–156. https://doi.org/10.1016/j.tree.2008.10.011

Paarmann, W., Rohe, W., Luchtrath, I., Assmann, T., & Mossakowski, D. (2007). Heredity of the elytral colour in adults of *Poecilus lepidus* Leske (Coleoptera, Carabidae). *Proceedings of the 13th European Carabidologists' Meeting, Blagoevgrad*, *75*, 183–194.

Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, *26*(3), 419–420. https://doi.org/10.1093/bioinformatics/btp696

Patton, A. H., Margres, M. J., Stahlke, A. R., Hendricks, S., Lewallen, K., Hamede, R. K., Ruiz-Aravena, M., Ryder, O., McCallum, H. I., Jones, M. E., Hohenlohe, P. A., & Storfer, A. (2019). Contemporary demographic reconstruction methods are robust to genome assembly quality: A case study in Tasmanian devils. *Molecular Biology and Evolution*, *36*(12), 2906–2921. https://doi.org/10.1093/molbev/msz191

Petit, J. R., Jouzel, J., Raynaud, D., Barkov, N. I., Barnola, J. M., Basile, I., Bender, M., Chappellaz, J., Davis, M., Delaygue, G., Delmotte, M., Kotiyakov, V. M., Legrand, M., Lipenkov, V. Y., Lorius, C., Pépin, L., Ritz, C., Saltzman, E., & Stievenard, M. (1999). Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature*, *399*(6735), 429–436. https://doi.org/10.1038/20859

Protas, M. E., & Patel, N. H. (2008). Evolution of coloration patterns. *Annual Review of Cell and Developmental Biology*, *24*(1), 425–446. https://doi.org/10.1146/annurev.cellbio.24.110707.175302

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Ramos-Onsins, S. E., & Rozas, J. (2002). Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, *19*(12), 2092–2100. https://doi.org/10.1093/oxfordjournals.molbev.a004034

Roach, M. J., Schmidt, S. A., & Borneman, A. R. (2018). Purge Haplotigs: allelic contig

reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, *19*(1), 460. https://doi.org/10.1186/s12859-018-2485-7

Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, *28*(21), 4737–4754. https://doi.org/10.1111/mec.15253

Roulin, A. (2004). The evolution, maintenance and adaptive function of genetic colour polymorphism in birds. *Biological Reviews*, *79*(4), 815–848. https://doi.org/10.1017/S1464793104006487

San-Jose, L. M., & Roulin, A. (2017). Genomics of coloration in natural animal populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1724). https://doi.org/10.1098/rstb.2016.0337

Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, *46*(8), 919–925. https://doi.org/10.1038/ng.3015

Schiffels, S., & Wang, K. (2020). MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent. In J. Y. Dutheil (Ed.), *Statistical Population Genomics* (Vol. 2090, pp. 147–166). Humana Press. https://doi.org/10.1007/978-1-0716-0199-0_7

Schlotterer, C., Ritter, R., Harr, B., & Brem, G. (1998). High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. *Molecular Biology and Evolution*, *15*(10), 1269–1274. https://doi.org/10.1093/oxfordjournals.molbev.a025855

Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology & Evolution*, *16*(7), 372–380. https://doi.org/10.1016/S0169-5347(01)02198-X

Schultz, T. D. (1986). Role of structural colors in predator avoidance by tiger beetles of the genus Cincindela (Coleoptera: Cincindelidae). *Bulletin of the Entomological Society of America*, *32*(3), 142–146. https://doi.org/10.1093/besa/32.3.142

Scordato, E. S. C., Wilkins, M. R., Semenov, G., Rubtsov, A. S., Kane, N. C., & Safran, R. J. (2017). Genomic variation across two barn swallow hybrid zones reveals traits associated with divergence in sympatry and allopatry. *Molecular Ecology*, *26*(20), 5676–5691. https://doi.org/10.1111/mec.14276

Seppä, H., Bjune, A. E., Telford, R. J., Birks, H. J. B., & Veski, S. (2009). Last nine-thousand years of temperature variability in Northern Europe. *Climate of the Past*, *5*(3), 523–535. https://doi.org/10.5194/cp-5-523-2009

Shuker, D. M., Underwood, K., King, T. M., & Butlin, R. K. (2005). Patterns of male sterility in a grasshopper hybrid zone imply accumulation of hybrid incompatibilities without selection. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1580), 2491–2497. https://doi.org/10.1098/rspb.2005.3242

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., & Flook, P. (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, *87*(6), 651–701. https://doi.org/10.1093/aesa/87.6.651

Slatkin, M. (1987). Gene flow and the geographic structure of natural populations.

Science, *236*(4803), 787–792. https://doi.org/10.1126/science.3576198

Smit, A. F. A., Hubley, R., & Green, P. (2015). *RepeatMasker Open-4.0. 2013-2015*. RepeatMasker Open-4.0. http://www.repeatmasker.org

Stacklies, W., Redestig, H., Scholz, M., Walther, D., & Selbig, J. (2007). pcaMethods a bioconductor package providing PCA methods for incomplete data. *Bioinformatics*, *23*(9), 1164–1167. https://doi.org/10.1093/bioinformatics/btm069

Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, *24*(5), 637–644. https://doi.org/10.1093/bioinformatics/btn013

Stanke, M., Schöffmann, O., Morgenstern, B., & Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics*, *7*, 1–11. https://doi.org/10.1186/1471-2105-7-62

Stankowski, S., Sobel, J. M., & Streisfeld, M. A. (2017). Geographic cline analysis as a tool for studying genome-wide variation: a case study of pollinator-mediated divergence in a monkeyflower. *Molecular Ecology*, *26*(1), 107–122. https://doi.org/10.1111/mec.13645

Sullivan, A. P., Bird, D. W., & Perry, G. H. (2017). Human behaviour as a long-term ecological driver of non-human evolution. *Nature Ecology and Evolution*, *1*(3), 1–11. https://doi.org/10.1038/s41559-016-0065

Szymura, J. M., & Barton, N. H. (1986). Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. *Evolution*, *40*(6), 1141–1159. https://doi.org/10.1111/j.1558-5646.1986.tb05740.x

Tajima, F. (1989a). Statistical method for testing the neutral mutation hypothesis by

DNA polymorphism. *Genetics*, *123*(3), 585–595. http://www.ncbi.nlm.nih.gov/pubmed/6374681

Tajima, F. (1989b). The effect of change in population size on DNA polymorphism. *Genetics*, *123*(3), 597–601. http://www.ncbi.nlm.nih.gov/pubmed/2599369

Tateno, H., Kitamori, Y., & Kondo, Y. (2020). Some ecological notes on *Phelotrupes auratus auratus* Motschulsky (Coleoptera, Geotrupidae). *Gekkan Mushi*, *596*(2), 11–23. (In Japanese)

Terhorst, J., Kamm, J. A., & Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics*, *49*(2), 303–309. https://doi.org/10.1038/ng.3748

Thayer, R. C., Allen, F. I., & Patel, N. H. (2020). Structural color in Junonia butterflies evolves by tuning scale lamina thickness. *ELife*, *9*, 1–21. https://doi.org/10.7554/eLife.52187

Toda, N., & Akei, K. (2003). Distribution of *Phelotrupes auratus* Moschulsky (Coleoptera, Geotrupidae) in Aichi Prefecture, Japan. *Science Report Toyohashi Museum National History*, *13*, 5–9. (In Japanese)

Tsujino, R., Ishimaru, E., & Yumoto, T. (2010). Distribution patterns of five mammals in the Jomon period, middle Edo period, and the present, in the Japanese Archipelago. *Mammal Study*, *35*(3), 179–189. https://doi.org/10.3106/041.035.0304

Tsukamoto, K., Inagaki, M., Kawahara, M., & Mori, M. (2014). *The earth-boring dung beetle, Phelotrupes auratus, in Japan*. Mushi-sha. (In Japanese)

Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, *27*(5), 737–746.

https://doi.org/10.1101/gr.214270.116

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*, *33*(14), 2202–2204. https://doi.org/10.1093/bioinformatics/btx153

Waldron, S. J., Endler, J. A., Valkonen, J. K., Honma, A., Dobler, S., & Mappes, J. (2017). Experimental evidence suggests that specular reflectance and glossy appearance help amplify warning signals. *Scientific Reports*, *7*(1), 257. https://doi.org/10.1038/s41598-017-00217-5

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*, *9*(11). https://doi.org/10.1371/journal.pone.0112963

Wang, B., Mojica, J. P., Perera, N., Lee, C., Lovell, J. T., Sharma, A., Adam, C., Lipzen, A., Barry, K., Rokhsar, D. S., Schmutz, J., & Mitchell-olds, T. (2019). Ancient polymorphisms contribute to genome-wide variation by long-term balancing selection and divergent sorting in *Boechera stricta*. *Genome Biology*, *20*(1), 1–15. https://doi.org/https://doi.org/10.1186/s13059-019-1729-9

Wang, L. B., Yang, Z. S., ZHang, R. P., Fan, D. J., Zhao, M. X., & Hu, B. Q. (2011). Sea surface temperature records of core ZY2 from the central mud area in the South Yellow Sea during last 6200 years and related effect of the Yellow Sea Warm Current. *Chinese Science Bulletin*, *56*(15), 1588–1595. https://doi.org/10.1007/s11434-011-4442-y

Warnes, G., Gorjank, G., Leisch, F., & Man, M. (2019). *Genetics: population genetics. R package version 1.3.8.1.2*. https://cran.r-project.org/web/packages/genetics/index.html

Watanabe, T., Tanigaki, T., Nishi, H., Ushimaru, A., & Takeuchi, T. (2002). A quantitative analysis of geographic color variation in two Geotrupes dung beetles. *Zoological Science*, *19*(3), 351–358. https://doi.org/10.2108/zsj.19.351

Watanabe, Y., Naka, I., Khor, S. S., Sawai, H., Hitomi, Y., Tokunaga, K., & Ohashi, J. (2019). Analysis of whole Y-chromosome sequences reveals the Japanese population history in the Jomon period. *Scientific Reports*, *9*(1), 2–9. https://doi.org/10.1038/s41598-019-44473-z

Yang, S., Wang, L., Huang, J., Zhang, X., Yuan, Y., Chen, J. Q., Hurst, L. D., & Tian, D. (2015). Parent-progeny sequencing indicates higher mutation rates in heterozygotes. *Nature*, *523*(7561), 463–467. https://doi.org/10.1038/nature14649

Yasuda, Y., Fujiki, T., Nasu, H., Kato, M., Morita, Y., Mori, Y., Kanehara, M., Toyama, S., Yano, A., Okuno, M., Jiejun, H., Ishihara, S., Kitagawa, H., Fukusawa, H., & Naruse, T. (2004). Environmental archaeology at the Chengtoushan site, Hunan Province, China, and implications for environmental change and the rise and fall of the Yangtze River civilization. *Quaternary International*, *123–125*, 149–158. https://doi.org/10.1016/j.quaint.2004.02.016

Zane, L., Bargelloni, L., & Patarnello, T. (2002). Strategies for microsatellite isolation: a review. *Molecular Ecology*, *11*(1), 1–16. https://doi.org/10.1046/j.0962-1083.2001.01418.x

Zhang, L., Mazo-Vargas, A., & Reed, R. D. (2017). Single master regulatory gene coordinates the evolution and development of butterfly color and iridescence.

Proceedings of the National Academy of Sciences of the United States of America, *114*(40), 10707–10712. https://doi.org/10.1073/pnas.1709058114

Zhang, X., Rayner, J. G., Blaxter, M., & Bailey, N. W. (2021). Rapid parallel adaptation despite gene flow in silent crickets. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-020-20263-4

Zhou, Y., Browning, S. R., & Browning, B. L. (2020). A fast and simple method for detecting identity-by-descent segments in large-scale data. *The American Journal of Human Genetics*, *106*(4), 426–437. https://doi.org/10.1016/j.ajhg.2020.02.010

Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics*, *29*(21), 2669–2677. https://doi.org/10.1093/bioinformatics/btt476