Population genetic structure and evolution of Batesian mimicry in Papilio polytes from the Ryukyu Islands, Japan, analyzed by genotyping-by-sequencing

Yukuto Sato[1], Kaori Tsurui-Sato[1], Mitsuho Katoh[1], Ryosuke Kimura[1], Haruki Tatsuta[1], and Kazuki Tsuji[1]

[1]University of the Ryukyus

August 7, 2020

Abstract

Batesian mimicry is a striking example of Darwinian evolution, in which a mimetic species resembles toxic or unpalatable model species, thereby, receiving protection from predators. In some species exhibiting Batesian mimicry, non-mimetic individuals coexist in the same population despite the benefits of mimicry, and the relative abundance of mimetic individuals in the population varies among localities. In a previous study, we found that the mimic ratios (MRs), which varied among the Islands, were likely to be explained by the model abundance of each habitat, rather than isolation-by-distance or phylogenetic constraint in the swallowtail butterfly, Papilio polytes, on the Ryukyu Islands in Japan. Because the previous study, however, was based on the mitochondrial DNA (mtDNA) that may sometimes mislead conclusion stemming from the inherent genetic properties, this possibility was reexamined based on hundreds of nuclear single-nucleotide polymorphisms (SNPs) of 95 P. polytes individuals from five Islands of the Ryukyus. We found that the population genetic and phylogenetic structures of P. polytes largely corresponded to the geographic arrangement of the habitat Islands, and the genetic distances among island populations significantly correlated with the geographic distances, which was not evident by the mtDNA-based analysis. The revised SNP-based genetic distances with a partial correlation approach revealed that the MRs of P. polytes were strongly correlated with the model abundance of each island, implying that negative frequency-dependent selection shaped and maintained the mimetic polymorphism. Taken together, we suggest that predation pressure, not neutral factors, is major driving force to determine the relative abundance of Batesian mimicry in P. polytes from the Ryukyus.

Introduction

Batesian mimicry is an often mentioned and striking example of Darwinian evolution (Bates 1862; Cott 1940; Rettenmeyer 1970; Edmunds 1974; Ruxton et al. 2004, Kunte 2009). In this mimicry system, prey species resemble harmful or unpalatable model species, receiving protection from predators through visual imitation of toxic or otherwise defensive models. In some butterfly species, however, Batesian mimicry is found only in a subset of individuals of the population (Wallace 1865; Wickler 1968; Mallet & Joron 1999, Kunte 2009). Such polymorphism in Batesian mimicry has long been a matter of debate (Mallet & Joron 1999; Kunte 2009) as it is counterintuitive to the benefits of mimicry for avoiding predation (Mallet & Joron 1999; Kunte 2009). Functional variation of the autosomal gene *doublesex* (*dsx*) is related to the polymorphism of Batesian mimicry in butterflies of the genus *Papilio* (Clarke & Sheppard 1972; Kunte et al. 2014; Nishikawa et al. 2015; Komata et al. 2016; Zhang et al. 2017; Iijima et al. 2018; Iijima et al. 2019; Palmer & Kronforst 2020). However, determining the ecological factors and evolutionary mechanisms that shape and maintain mimetic polymorphism are long-standing unresolved challenges.

In a previous study, we proposed that negative frequency-dependent selection (NFDS) explains the polymorphic Batesian mimicry of the swallowtail butterfly *Papilio polytes* L. from the Ryukyu Islands in Japan

(Tsurui-Sato et al. 2019). The NFDS model (Barrett 1976; Turner 1978; Kunte 2009) predicts that the number of mimics in the population depends on the abundance of the model species in the habitat, because the defensive benefit of mimicry increases when the toxic or unpalatable model is more abundant. But if the mimics become overabundant relative to the models, the advantage of the mimicry decreases (Barrett 1976). Accordingly, the mimic ratio (MR; proportion of mimics in the population) is expected to approach an equilibrium with equal fitness between mimetic and non-mimetic types (Kunte 2009), determined by an abundance of model species as an ecological factor.

To evaluate this NFDS model for the maintenance of Batesian mimicry polymorphism, *P. polytes* from the Ryukyu Islands provides a good model system (Tsurui-Sato et al. 2019). This species is a common swallowtail butterfly in Southeast Asia and the southern part of East Asia, and exhibits polymorphic Batesian mimicry in females (Clarke & Sheppard 1972; Ford 1975; Kunte et al. 2014; Katoh et al. 2017). In the Ryukyu Islands, some females mimic unpalatable *Pachliopta aristolochiae* as a defense against avian predators (Uesugi 2000; Katoh et al. 2017), while others resemble males (Figure 1a). Interestingly, the MRs vary among the Ryukyu Island populations, and tend to be high where more of the model *P. aristolochiae* are present (Uesugi 2000; Tsurui-Sato et al. 2019). In particular, after the model species immigrated to Miyako Island (MYK in Figure 1b) and became established, the MR of this island increased rapidly from 1975 to 1989 (Uesugi 2000), implying local adaptation through NFDS (Tsurui-Sato et al. 2019).

As an alternative hypothesis, polymorphic Batesian mimicry may be explained by neutral evolutionary processes, such as isolation-by-distance and phylogenetic constraint, in addition to other hypotheses including sexual selection and ecological-physiological trade-offs (Burns 1966; Vane-Wright 1984; Cook et al. 1994; Ohsaki 2005; Katoh et al. 2020). In the simple neutral process, the mimetic and non-mimetic phenotypes have similar fitness independent of their frequencies. Accordingly, the MRs in local populations change through genetic drift and migration (Wright 1943; Ackermann & Cheverud 2004). Given this scenario, geographically closer populations (*i.e.*, neighboring island populations), which are expected to be genetically closer to each other, should exhibit similar MRs irrespective of the quantity of mimetic models on each island.

Our previous study (Tsurui-Sato et al. 2019) examined the possible association between the MR of *P. polytes* and the abundance of the mimetic model in the Ryukyu Islands, with respect to the NFDS hypothesis. The alternative, isolation-by-distance view of the MR variation among islands was tested by genetic analyses based on mitochondrial DNA (mtDNA). Strong correlation of the MRs of *P. polytes* with mimetic model abundance was found in five islands of the Ryukyu, rather than with geographic or genetic distances, incorporating newly added fieldwork data (Tsurui-Sato et al. 2019). In particular, the northern island populations were thought to be relatively recent immigrants having experienced population expansion, while their MRs correlated with the local abundance of model species that was probably caused by rapid evolution through NFDS (Tsurui-Sato et al. 2019). These results, however, may have suffered from incomplete lineage sorting or other biases because they are primarily based on a single locus. Hence, to evaluate and confirm the insights gathered from our former study, we address the population genetic characteristics and mimetic evolution of *P. polytes* of the same Ryukyu Islands, based on a much larger set of single nucleotide polymorphisms (SNPs) of nuclear genomes that enables more averaged and comprehensive evaluation of genetic diversity of the populations.

Materials and Methods

Sample collection and DNA extraction

Sample *P. polytes* were collected as described in our previous study (Tsurui-Sato et al., 2019) from eight islands of the Ryukyu Archipelago (the Ryukyus), southern Japan (Kikai, Amami, Okinawa, Aguni, Miyako, Tarama, Ishigaki, and Taketomi Islands). Among these samples, we analyzed 95 females from the following five islands, where more than ten individuals of mimetic and non-mimetic females were obtained: 11, 46, 12, 14, and 12 individuals of Kikai (abbreviated as KIK), Okinawa (OKI), Miyako (MYK), Ishigaki (ISG), Taketomi (TKT), respectively (Figure 1). These are basically identical to the samples analyzed for mtDNA in the study of Tsurui-Sato et al. (2019), apart from an additional four samples from MYK and two samples from ISG that were included in this study for nuclear SNP analysis. Butterflies were caught using

2

a bag-net between 2014-2016, examined to record the sex and mimic type, and stored at -30 °C until DNA extraction. Thus, the same individuals were neither collected nor analyzed twice. DNA was extracted from the entire middle and hind legs of the frozen samples using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) after hand shearing, manual homogenization, and enzymatic digestion of the samples by proteinase K, as described by Tsurui-Sato et al. (2019). The eluted DNA was quantified using a Nanodrop 2000C spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and stored at -30 °C.

PCR amplification and genotyping-by-sequencing for nuclear SNP analysis

To obtain genotyping data for random SNPs from the nuclear genome using a PCR-based method, we applied multiplexed inter-simple sequence repeat (ISSR) genotyping by sequencing (MIG-seq) analysis (Suyama & Matsuki 2015). Since frozen leg samples of *P. polytes* butterflies are relatively small in size and weight, the total genomic DNA obtained from these samples was relatively low and the DNA quality was variable due to diverse field and sample storage conditions. Therefore, it was difficult to apply PCR-free SNP analysis methods, such as restriction site-associated DNA sequencing (RAD-seq) that require relatively large amounts of high-quality DNA, to our *P. polytes* samples. On the other hand, PCR-based methods, such as MIG-seq, are capable of analyzing smaller amounts of DNA, although the number of consensus SNPs identified among individuals may be limited compared to that with RAD-seq or whole genome resequencing methods (Davey et al. 2011; Suyama & Matsuki 2015).

To amplify and sequence the ISSR regions from genomic DNA using the MIG-seq method, two rounds of PCR were performed as described by Suyama & Matsuki (2015). The reaction mixture of the 1st round PCR included 1.0 µL of sample DNA, 0.2 µM of each MIG-seq set-1 primer (Suyama & Matsuki 2015), 0.035 µL of Multiplex PCR Enzyme Mix, and 3.5 µL of 2 × Multiplex PCR Buffer from the Multiplex PCR Assay Kit Ver. 2 (Takara Bio, Shiga, Japan) for each sample. The thermal cycling profile of the 1st round PCR was as follows: 94 °C for 1 min followed by 27 cycles at 94 °C for 30 sec, 48 °C for 1 min, and 72 °C for 1 min, with a final extension at 72 °C for 10 min.

The 1st-round PCR product was diluted 50-fold in RNase-free water (Thermo Fisher Scientific/Invitrogen, Carlsbad, USA) and subjected to 2nd-round indexing PCR to add the dual-index tags (A5 and A7 series) and flow cell binding site sequences of the Illumina DNA sequencer platforms (Illumina, San Diego, CA, USA). The PCR mixture contained 2.5 µL of diluted 1st-round PCR product, 0.2 µM of each forward and reverse indexing primer, 0.24 µL of PrimeSTAR GXL DNA Polymerase (Takara Bio), 0.96 µL of dNTP mixture, and 2.4 µL of 5 × PrimeSTAR GXL Buffer (Takara Bio). The thermal cycling profile was as follows: 94 °C for 1 min followed by 15 cycles at 98 °C for 10 sec, 54 °C for 15 sec, and 68 °C for 30 sec. The amplified MIG-seq DNA libraries were equivalently pooled among samples, extracted from 1.0 % L03 agarose gel (Takara; sizes ranging from 400 to 800 bp) using a MinElute gel extraction kit (Qiagen), purified using AMPure XP magnetic beads (Beckman Coulter, High Wycombe, UK) using a standard purification protocol, quantified by a Qubit 3.0 fluorometer (Thermo Fisher Scientific) and Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA, USA), and sequenced in a single lane of the HiSeq X platform (Illumina) to generate 2 × 151-bp paired-end sequences.

Quality-based sequence data filtering and **de novo***SNP calling*

Raw sequencing data of MIG-seq amplicons from the genomic DNA of*P. polytes* (DRA accession number: DRA010473) underwent quality filtering using relevant software/scripts and *de novo* SNP calling using the software package Stacks version 1.45 (Catchen et al. 2013). The first 17 nucleotides of the 5'-ends of each read were uniformly deleted to remove MIG-seq forward primer sequences using Cutadapt version 1.14 (Martin 2011). The adapter primer-derived sequences that sometimes remained or were repeated in the opposite 3'-ends of each read were carefully removed by Cutadapt, twice, using a [?] 8-bp length accordance and a [?] 10% base mismatch. Finally, the low-quality (Phred score < 10) 3'-tails of each read were trimmed by Cutadapt and the shorter sequences (< 120 bp in length) were excluded using a custom Perl script.

The quality-filtered forward reads (read R1) were analyzed for *de novo* SNP calling. First, all forward reads from each sample were clustered into putative loci within an individual based on sequence similarity

using the program Ustacks that is provided in the software package Stacks. Threshold values of minimum coverage depth of respective loci and maximum number of mismatched bases among alleles were set to 4 and 3 (2.2-2.5% per locus), respectively, based on preliminary analysis using eight individuals from all five islands (Supplementary Figure S1a and S1b). Next, consensus loci among individuals were identified using the program Cstacks, with a threshold value of 3 for the maximum number of mismatch bases among alleles from other individuals (2.2-2.5% per locus), as determined based on preliminary analysis of the same individuals as above (Supplementary Figure S1c). Finally, SNP calling was performed using the programs Sstacks and Populations to generate an SNP matrix table with less than 50% missing values for each locus across all individuals. All individuals were treated provisionally as one population (pre-definition of subpopulations was not applied) to avoid biased output of missing values affected by a given subpopulation structure. One SNP site was randomly selected from each locus for outputting SNP genotypes using the option 'write-random-snp' for the Populations.

Population genetic and molecular phylogenetic analyses

To examine the phylogeographic and evolutionary history of *P. polytes* from the Ryukyu Islands in relation to their Batesian mimicry, we performed population genetic and molecular phylogenetic analyses. Among the generated SNP data matrix, we focused on loci with less than 20% missing values across all 95 individuals. We removed the individuals having more than 30% missing values across those loci, which was probably due to the relatively low quality of initial genomic DNA. Then, the SNP loci with less than 20% of re-calculated missing values across the remaining individuals were used for the following genetic analysis.

Chi-squared tests were performed to remove the SNP loci indicating significant deviation from Hardy-Weinberg equilibrium (HWE; $p < 0.05$) in any one of the five island populations, using Arlequin version 3.5.2.2 (Excoffier & Lischer 2010). On the basis of the remaining multi-locus SNP genotypes, the population genetic structure was estimated using the program Structure, version 2.3.4 (Pritchard et al. 2000). The implemented admixture model was applied to cluster individuals into genetically similar subpopulations considering each SNP locus as genetically independent. The length of burn-in period and number of Markov Chain Monte Carlo (MCMC) iterations after burn-in were set to 50,000. Population genetic parameters including observed and expected heterozygosity, pairwise F_{ST} among island populations (Slatkin 1995) standardized by the estimated divergence times, and average gene diversity (or average heterozygosity; Nei 1987) over SNP sites were calculated using Arlequin.

SNP-based phylogenetic analysis was performed with both concatenation- and coalescent mode-based methods. First, multi-locus SNP genotype data was concatenated in a random order of alleles for each locus using a custom Perl script, since the MIG-seq loci were considered to be independent and thus the gametic phases were unknown. The obtained multi-FASTA formatted data was analyzed to estimate the molecular phylogenetic network tree based on the NeighborNet algorithm (Bryant & Moulton 2004) and Kimura's two-parameter model of nucleotide substitution (Kimura 1980) using the SplitsTree 5 version 5.1.7 (Huson & Bryant 2006). Coalescent model-based phylogenetic analysis was performed using the software package SNAPP (SNP and AFLP Package for Phylogenetic analysis; Bryant et al. 2012) implemented in BEAST 2 version 2.6.2 (Bouckaert et al. 2014; Bouckaert et al. 2019). The analysis was performed with default parameters and settings including the Jukes-Cantor model (Jukes & Cantor 1969) of nucleotide substitution and 10,000,000 MCMC iterations.

Based on the same concatenated SNP site data, a molecular phylogenetic tree was estimated by the neighbor-joining (NJ) method (Saito & Nei 1987) based on Kimura's two-parameter model of nucleotide substitution (Kimura 1980) using MEGA 7 version 7.0.14 (Kumar et al. 2016). The average number of pairwise differences and Nei's net number of nucleotide differences D_A (Nei & Li 1979) of SNP sites among the five island populations was also estimated. Based on the genetic distance D_A of the SNP sites, the population-level tree among the islands was inferred by the NJ method. Genetic differentiation (Φ_{ST}) of the SNP sites among the islands was also calculated using D_A divided by the average number of pairwise differences among islands. Coordination of two-dimensional nonmetric multidimensional scaling (nMDS) of the SNP profile was estimated using PAST version 4.03 (Hammer et al. 2001) based on the Kimura's two-parameter model

4

of nucleotide substitution (Kimura 1980), to calculate similarity scores among the samples.

Mantel and partial Mantel tests regarding the evolution of Batesian mimicry

Possible associations among genetic distance D_A, geographic distance, and MR of the five island populations of *P. polytes* were assessed by the Mantel test (Mantel 1967) and partial Mantel test (Smouse et al. 1986). The D_A s of the SNP sites among the five island populations, estimated in the previous section, were used as average genetic distances between islands. Geographical distances among the islands were measured by Tsurui-Sato et al. (2019) and the same data was used in the present analysis. These were distances in kilometers between the centers of sampling regions (*i.e.*, city hall or downtown area) of each island (Fig. 1a). However, the location of the Nakijin village office was used for OKI because the butterflies were collected in the Nakijin region, in the northern part of OKI. The MR and the advantage index (AI) of Batesian mimicry of *P. polytes* used in this study were calculated by Tsurui-Sato et al. (2019) based on the methods of Uesugi (2000) and Sekimura et al. (2014). The Mantel test was performed using Arlequin and the partial Mantel test was performed using a custom R script developed on the R platform version 3.4.2 (R Core Team 2017), based on 10,000 permutation tests according to the Method 1 by Legendre (2000).

Results

General results of genotyping-by-sequencing

We obtained a total of 115,550,624 pairs of raw DNA sequences with an average of 1,216,322 ± 51,480 (mean ± S.E.) per sample. Following the base call quality- and sequence length-based filtering of the forward reads (Phred score [?] 10 and [?] 120 bp; see Materials and Methods), a total of 94,897,949 reads (998,926 +- 44,136 per sample) remained. Based on these quality-filtered forward reads, on average, 4,864 +- 164 SNP loci were obtained for each sample with a mean coverage depth of 48 +- 0.24. Among the 95 individuals, 1,258, 799, 517, 233, and 82 consensus SNP loci were identified with less than 50, 40, 30, 20, and 10 percentage missing values for each locus, respectively. Focusing on the 233 loci with < 20% missing rates, two individuals (M9 of MYK and T121 of TKT) showed > 30% missing values per sample and were excluded from the analysis. Based on the re-calculated missing rates, a total of 259 SNP loci with < 20% missing values for each locus from 93 samples were used in the population genetic and phylogenetic analyses.

Population genetic structure and characteristics of **P. polytes** *in the Ryukyu Islands*

The SNP-based population genetic analyses indicated that two almost distinct *P. polytes* populations exist in the five Ryukyu islands (Figure 1). The analysis was performed using 238 SNP site data from the 93 individuals after excluding 21 sites that showed significant deviation from the HWE at $p < 0.05$. The results of the population assignment test showed that one population (indicated by gray in Figures 1b and 1c) corresponded to the northern islands (KIK and OKI), and another (indicated by black in Figures 1b and 1c) to the southern islands (MYK, ISG, and TKT).

The individual-level plot of nMDS exhibited a consistent pattern, in which the northern and southern island populations clustered separately (Figure 1d). The northern individuals (colored in blue and pink in Figure 1d) were plotted in lower value spaces of coordinate 1 (horizontal axis of the plot), and the southern ones (colored in green, orange, and yellow in Figure 1d) in higher value spaces. The estimated population genetic indicators based on these SNP sites (not a whole genomic average) showed relatively lower average gene diversity over SNP sites in the northern island populations (0.035-0.048), while those of the southern islands were relatively higher (0.057-0.102) (Table 1). In particular, the southern ISG and TKT had larger numbers of polymorphic sites (23 and 29) relative to their smaller sample sizes.

Phylogenetic relationship and mimetic characters of **P. polytes** *in the Ryukyu Islands*

We again identified separate northern (KIK and OKI) and southern (MYK, ISG, and TKT) island groups by molecular phylogenetic analyses (Figure 2 and Supplementary Figure S2). A phylogenetic network based on concatenated SNP site data clearly indicated two large, northern and southern clades (Figure 2a), which contained subclades mostly corresponding to individual islands. The northern subclade N1 was comprised

of OKI samples, and N2 consisted of mainly KIK individuals and also included two OKI individuals. The southern subclades S1, S2, and S3 were mostly composed of MYK, TKT and ISG, and ISG samples, respectively, and were partly a mixture of other southern island individuals. The molecular phylogenetic tree (Supplementary Figure S2) showed a similar subclade structure largely corresponding to KIK, OKI, MYK, ISG, and ISG and TKT Island individuals, although the separation was somewhat obscure compared to that of the phylogenetic network (Figure 2a). These two phylogenetic analysis consistently indicated a clear separation of northern and southern populations, and the mimetic phenotypes (shown by white outlined letters in Figures 2a and Supplementary Figure S2) were independent from the phylogeny, not forming monophyletic groups.

A population-level phylogenetic tree (Figure 2b) based on the genetic distance D_A indicated a relationship among island populations, which was congruent with individual-level phylogenetic relationships (Figures 2a and Supplementary Figure S2). The KIK and OKI populations constituted the northern clade, and MYK, ISG, and TKT constituted the southern clade, with MYK in a relatively basal position. This relationship among the island populations appeared to be consistent with those indicated by the standardized pairwise genetic differentiation scores Φ_{ST} of the SNP sites (Table 2). The northern KIK and OKI, and southern TKT and ISG pairs showed low (closer) scores, respectively (Φ_{ST}, 0.125 and 0.014), and MYK showed intermediate scores relatively closer to other southern ones (Φ_{ST} 0.045-0.102). The island-level branching patterns were, however, not so clear in the Slatkin's pairwise F_{ST} (Table 2) and phylogenetic trees (Figures 2a and Supplementary Figure S2). Such ambiguities regarding the relationships among island populations or subclades were also shown by the coalescent model-based phylogenetic tree densities (Figure 3 and Supplementary Figure S3). The posterior density of simulated coalescent trees, however, also indicated the existence of large northern (KIK and OKI) and southern (MYK, ISG, and TKT) populations and the non-monophyletic origin of the mimetic phenotype of *P. polytes* (indicated by white outlined letters in Figures 3 and Supplementary Figure S3).

Genetic distances and other factors possibly associated with the MR of **P. polytes**

The genetic distance (D_A) and MR difference among the five island populations of *P. polytes* exhibited no significant association by the Mantel test based on the nuclear SNP data obtained in this study (Table 3). On the other hand, we confirmed the strong correlation between the MR and AI controlling for the nuclear SNP-based genetic distances by the partial Mantel test (Table 4), consistent with the results from the mtDNA data from our previous study (Tsurui-Sato et al. 2019). Two methods of calculating AI (Uesugi 2000; Tsurui-Sato et al. 2019) were used, and both provided significant results at $p < 0.01$ and $p < 0.05$, respectively (bold faces in Table 4). In addition, we found a significant correlation between the geographic and genetic distances of these nuclear SNPs among the five islands by the Mantel test at $p < 0.05$ (Table 3), whereas such a correlation was not detected based on mtDNA analysis (Tsurui-Sato et al. 2019).

Discussion

We investigated the population genetic and phylogenetic characteristics of *P. polytes* in five islands of the Ryukyus, Japan, based on information from hundreds of nuclear SNPs (Figures 1-3, Tables 1 and 2). Our results confirm that the MR of these butterfly populations can be explained by local model species abundance (indicated by AI or AI [Uesugi]; Table 4) rather than genetic distances or other factors such as geographic and environmental distances (Table 3). Interestingly, the current analysis based on nuclear SNPs detected a significant correlation between geographic and genetic distances (Table 3), while such a correlation was not detected by mtDNA analysis in our previous study (Tsurui-Sato et al. 2019). This implies that the results of mtDNA analysis of one locus may have been affected by incomplete lineage sorting or smaller effective population sizes, leading to less accurate results. Hundreds of nuclear SNPs appear to reflect the micro-evolutionary history of *P. polytes* in the Ryukyu Islands more accurately, although SNP analysis also has inherent problems such as ascertainment bias (Hartl & Clark 2007; Lachance & Tishkoff 2013). This is a methodological limitation as SNP analysis often depends on common polymorphic nucleotide sites that do not necessarily represent the whole genomic and meta-population average. In addition, the current analysis used 238 SNP sites, which is a relatively small number of genetic markers, due to our own technical limitations

6

associated with the amount and quality of DNA samples (see Materials and Methods). The estimated genetic distances based on nuclear SNPs, however, seem to be vastly improved because they correlate with the geographic distances as expected (Table 3). These nuclear SNP-based genetic distances, along with the geographic and environmental distances, do not explain the MR distribution across the Ryukyu Islands (Table 3). Accordingly, given that the MR and AI are strongly correlated after controlling for genetic and other distances (Table 4), we suggest that the polymorphic Batesian mimicry of *P. polytes* in the Ryukyu Islands (Figure 1) has been shaped and maintained through NFDS. This view, which has been proposed theoretically (Barrett 1976; Turner 1978; Kunte 2009) and evidenced by field survey and mtDNA analysis (Tsurui-Sato et al. 2019), was evaluated and confirmed by the present study by adding nuclear genomic support.

The current SNP analysis also clearly revealed the population genetic and phylogeographic characteristics of *P. polytes* in the Ryukyu Islands (Figures 1-3, Supplementary Figures S2 and S3). The results of population assignment (Figures 1b and 1c), nMDS (Figure 1d), and molecular phylogenetic analyses (Figures 2 and 3, Supplementary Figures S2 and S3) consistently segregated the two major northern and southern groups corresponding to the geographic distribution of habitat islands, not the mimetic types. Consequently, it appears that the mimetic phenotypes of female *P. polytes* are not determined by phylogenetic inertia, but exist as polymorphisms in each population, as previously indicated based on mtDNA analysis (Tsurui-Sato et al. 2019). Mitochondrial DNA analysis also identified the large northern and southern groups of *P. polytes* in the Ryukyus, but the MYK (Miyako Island) population was clustered with the northern group (Tsurui-Sato et al. 2019). The current analysis indicates the MYK population belongs to the southern group together with ISG and TKT (Figures 1 and 2, Table 2). This revised grouping is consistent with the known biogeographic gap, the "Kerama strait", dividing the fauna and flora of the Ryukyu archipelago (Figure 1; Ota 1998; Nakamura et al. 2009; Hirao et al. 2015). The clustering of MYK with the northern group by mtDNA analysis may have been affected by incomplete lineage sorting or other factors.

The resultant view of the genetic diversity of *P. polytes* in the Ryukyus is, however, highly comparable between the previous mtDNA and current nuclear SNP analyses. We found that the southern populations (ISG and TKT) were the source of mtDNA variation, and the northern populations (KIK, OKI, and MYK) were the derived, relatively recent immigrants (Tsurui-Sato et al. 2019) according to molecular phylogenetic, mismatch-distribution, and Tajima's D analyses (Tajima 1989a, 1989b). The current SNP analysis consistently indicated higher average gene diversity of SNP sites in the southern populations (0.065 and 0.102 in ISG and TKT, respectively; Table 1) and their relatively deep-branching pattern in phylogenetic trees (Figure 2, Supplementary Figure S2), implying higher genetic diversity and a longer history. It also indicates the relatively lower average gene diversity in the northern populations (0.048 and 0.035 in KIK and OKI, respectively; Table 1). In particular, the higher number of private alleles and the lower heterozygosity in the OKI population denote that this population harbors many low-frequency alleles generated by relatively recent mutations (Table 1), implying closer relationships among individuals and a shorter history. These observations support the view that the northern OKI population was established by relatively recent immigrants, and some of the OKI individuals further emigrated into the northernmost KIK Island analyzed in this study (Figures 1 and 2; Tsurui-Sato et al. 2019). The KIK population was estimated to have been established in the relatively recent past (31,000 to 45,000 years ago) after population expansion, based on the mtDNA variations (Tsurui-Sato et al. 2019), which does not conflict with the relatively young age of the KIK Island that is estimated to have arisen approximately 0.85 million years ago (Osozawa et al. 2012). A migration from the south in a northerly direction also seems probable because a low-level jet stream exists in this Ryukyu area during spring and summer (Japan Meteorological Agency 2016). Such a passive migration due to southern wind patterns has been reported in other flying insects such as the rice planthoppers *Sogatella furcifera* and *Nilaparvata lugens* (Kishimoto 1976; Seino et al. 1987). This phenomenon may have also affected the habitats of *P. polytes* .

In summary, the findings of Tsurui-Sato et al. (2019) and the present study suggest that the population genetic and phylogenetic characteristics of *P. polytes* in the Ryukyu Islands are highly compatible with the geographic pattern of their habitat islands, including possible migration in a northerly direction. The MR

of *P. polytes* populations in each island is, on the other hand, not explained by isolation-by-distance or other neutral factors (Table 3), as discussed above. This suggests that, in spite of past migration and gene flow that correlates with geographic distances (Figures 1 and 2, Table 3), the MR of *P. polytes* has been adjusted to the ecological conditions of each island, as determined directly by the abundance of mimetic model species (Table 4) and indirectly by predation pressure (Katoh et al. 2017). In general, the predation pressure on prey species (*P. polytes* in this case) is difficult to evidence in wild environments. In the case of butterflies, however, the strength of predation pressure can be estimated from a field survey of beak marks (*e.g.* , Ohsaki 1995). Our preliminary investigation for beak marks on mimetic females of *P. polytes* in the Ryukyu Islands suggested the predation by birds, the counts of which seem to be associated with the abundance of the model species (data not shown). We propose that the strong predation pressure on this butterfly (Katoh et al. 2017), rather than isolation-by-distance, phylogenetic constraint, or other environmental factors, has shaped and maintained the Batesian mimicry patterns of *P. polytes* observed in the Ryukyu Islands.

Further studies are needed to explore more comprehensively, the population genomic aspects of *P. polytes* in the Ryukyu Islands and molecular details and dynamics of the negative frequency-dependent evolution of the Batesian mimicry polymorphism across the islands. The PCR-based SNP analysis conducted herein has technical and theoretical limitations related to the number of available loci (Table 1), unexpected contamination from unknown parasitic/symbiotic organisms, and inherent analytical biases (Hartl & Clark 2007; Lachance & Tishkoff 2013), as partly discussed. Population-level whole genome resequencing should be performed to obtain more unbiased population genomic insights about *P. polytes* . Regarding the evolution of polymorphic Batesian mimicry, the responsible gene, *dsx,* in *Papilio* butterflies (Kunte et al. 2014; Nishikawa et al. 2015; Komata et al. 2016; Zhang et al. 2017; Iijima et al. 2018; Iijima et al. 2019; Palmer & Kronforst 2020) should be the focus of further investigation. Unveiling the molecular evolutionary dynamics of the *dsx* and allied genes across the Ryukyu Islands in association with the NFDS for mimetic types is necessary to understand the establishment of *P. polytes*mimicry interacting with ecological factors.

Acknowledgements

References

Ackermann, R. R., & Cheverud, J. M. (2004). Detecting genetic drift versus selection in human evolution.*Proceedings of the National Academy of Sciences of the United States of America, 101* (52), 17946–17951. doi: 10.1073/pnas.0405919102

Barrett, J. A. (1976). The maintenance of non-mimetic forms in a dimorphic Batesian mimic species. *Evolution, 30* (1), 82–85. doi: 10.1111/j.1558-5646.1976.tb00883.x

Burns, J. M. (1966). Preferential mating versus mimicry: disruptive selection and sex-limited dimorphism in *Papilio glaucus* .*Science, 153* (3735), 551–553. doi: 10.1126/science.153.3735.551.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C-H., Xie, D., . . . Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology, 10* (4), e1003537. doi: 10.1371/journal.pcbi.1003537

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., . . . Drummond, A. J. (2019) BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis.*PLoS computational biology, 15* (4), e1006650. doi: 10.1371/journal.pcbi.1006650

Bryant, D., & Moulton, V. (2004). Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution, 21* (2), 255–265. doi: 10.1093/molbev/msh018

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution, 29* (8), 1917–1932. doi: 10.1093/molbev/mss086

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology, 22* (11), 3124–3140. doi: 10.1111/mec.12354

Clarke, C. A., & Sheppard, P. M. (1972). The genetics of the mimetic butterfly *Papilio polytes* L. *Philosophical Transactions of the Royal Society B: Biological Sciences, 263* (855), 431–458. doi: 10.1098/rstb.1972.0006

Cook, S. E., Vernon, J. G., Bateson, M., & Guilford, T. (1994). Mate choice in the polymorphic African swallowtail butterfly, *Papilio dardanus* : male-like females may avoid sexual harassment. *Animal Behaviour, 47* (2), 389–397. doi: 10.1006/anbe.1994.1053

Cott, H. B. (1940). *Adaptive coloration in animals.* London: Methuen.

Cox, M. P., Peterson, D. A., & Biggs, P. J. (2010). SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics, 11* , 485. doi: 10.1186/1471-2105-11-485.

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics, 12* (7), 499–510. doi: 10.1038/nrg3012.

Edmunds, M. (1974). *Defence in animals: a survey of anti-predator defences.* New York: Longman Inc.

Excoffier, L., & Lischer, H. E. (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. doi: 10.1111/j.1755-0998.2010.02847.x

Hammer, Ø., Harper, D. A. T., & Ryan, P. D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica, 4* (1), 1–9.

Hirao, T., Kubota, Y., & Murakami, M. (2015). Geographical patterns of butterfly species diversity in the subtropical Ryukyu Islands: the importance of a unidirectional filter between two source islands. *Journal of Biogeography, 42* (8), 1418–1430. doi: 10.1111/jbi.12501

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution, 23* (2), 254–267. doi: 10.1093/molbev/msj030

Iijima, T., Kajitani, R., Komata, S., Lin, C. P., Sota, T., Itoh, T., & Fujiwara, H. (2018). Parallel evolution of Batesian mimicry supergene in two *Papilio* butterflies, *P. polytes* and *P. memnon .Science Advances, 4* (4), eaao5416. doi: 10.1126/sciadv.aao5416

Iijima, T., Yoda, S., & Fujiwara, H. (2019). The mimetic wing pattern of *Papilio polytes* butterflies is regulated by a doublesex-orchestrated gene network. *Communications Biology, 2* , 257. doi: 10.1038/s42003-019-0510-7

Japan Meteorological Agency (2016). *Weather observation data search* (in Japanese). Retrieved from

http://www.data.jma.go.jp/obd/stats/etrn/index.php

Jukes, T. H., & Cantor, C. R. (1969) Evolution of protein molecules. In H. N. Munro (Ed.), *Mammalian protein metabolism* (pp. 21–132). New York: Academic Press.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution, 30* (4), 772–780. doi: 10.1093/molbev/mst010

Katoh, M., Tatsuta, H., & Tsuji, K. (2017). Rapid evolution of a Batesian mimicry trait in a butterfly responding to arrival of a new model. *Scientific Reports* , *7* , 6369.

Katoh, M., Tatsuta, H., & Tsuji, K. (2020). Mimicry genes reduce pre-adult survival rate in *Papilio polytes* : A possible new mechanism for maintaining female-limited polymorphism in Batesian mimicry. *Journal of Evolutionary Biology* , *in press* .

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences.*Journal of Molecular Evolution, 16* (2), 111–120. doi: 10.1007/BF01731581

Kishimoto, R. (1976). Synoptic weather conditions inducing long-distance immigration of planthoppers, *Sogatella furcifera* Horvath and*Nilaparvata lugens* Stål. *Ecological Entomology* ,*1* (2), 95–109. doi: 10.1111/j.1365-2311.1976.tb01210.x

Komata, S., Lin, C. P., Iijima, T., Fujiwara, H., & Sota, T. (2016). Identification of doublesex alleles associated with the female-limited Batesian mimicry polymorphism in *Papilio Memnon* . *Scientific Reports, 6* , 34782. doi: 10.1038/srep34782

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets.*Molecular Biology and Evolution, 33* (7), 1870–1874. doi: 10.1093/molbev/msw054

Kunte, K. (2009). Female-limited mimetic polymorphism: a review of theories and a critique of sexual selection as balancing selection.*Animal Behaviour, 78* (5), 1029–1036. doi: 10.1016/j.anbehav.2009.08.013

Kunte, K., Zhang W., Tenger-Trolander A., Palmer D.H., Martin A., Reed R.D., . . . & Kronforst, M. R. (2014). Doublesex is a mimicry supergene. *Nature, 507* (7491), 229–232. doi: 10.1038/nature13112

Legendre, P. (2000). Comparison of permutation methods for the partial correlation and partial Mantel tests. *Journal of Statistical Computation and Simulation* , 67(1), 37–73. doi: 10.1080/00949650008812035

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal, 17* (1), 10–12. doi: 10.14806/ej.17.1.200

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research, 27* (2), 209–220.

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* , 201–233. doi: 10.1146/annurev.ecolsys.30.1.201

Nakamura, K., Suwa, R., Denda, T. & Yokota, M. (2009) Geohistorical and current environmental influences on floristic differentiation in the Ryukyu Archipelago, Japan. *Journal of Biogeography, 36* (5), 919– 928. doi: 10.1111/j.1365-2699.2008.02057.x

Nei, M. (1987). *Molecular evolutionary genetics* . New York: Columbia University Press.

Nei, M., & Li., W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America 76* (10), 5269–5273. doi: 10.1073/pnas.76.10.5269

Nishikawa, H., Iijima, T., Kajitani, R., Yamaguchi, J., Ando, T., Suzuki, Y., . . . & Tabata, S. (2015). A genetic mechanism for female-limited Batesian mimicry in *Papilio* butterfly.*Nature Genetics, 47* (4), 405–409. doi: 10.1038/ng.3241

Ohsaki, N. (1995). Preferential predation of female butterflies and the evolution of Batesian mimicry. *Nature, 378* (6553), 173–175. doi: 10.1038/378173a0

Ohsaki, N. (2005). A common mechanism explaining the evolution of female-limited and both-sex Batesian mimicry in butterflies.*Journal of Animal Ecology, 74* (4), 728–734. doi: 10.1111/j.1365-2656.2005.00972.x

Ota, H. (1998). Geographic patterns of endemism and speciation in amphibians and reptiles of the Ryukyu archipelago, Japan, with special reference to their paleogeographical implications. *Researches on Population Ecology, 40* (2), 189–204. doi: 10.1007/BF02763404

Palmer, D. H., & Kronforst, M. R. (2020). A shared genetic basis of mimicry across swallowtail butterflies points to ancestral co-option of *doublesex* . *Nature Communications, 11* (1), 6. doi: 10.1038/s41467-019-13859-y

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155* (2), 945–959.

R Core Team. (2017). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from *https://www.R-project.org*

Rettenmeyer, C. W. (1970). Insect mimicry. *Annual Review of Entomology, 15* (1), 43–74. doi: 10.1146/annurev.en.15.010170.000355

Ruxton, G.D., Sherratt, T.N., & Speed, M.P. (2004). *Avoiding attack.* Great Clarendon Street, Oxford: Oxford University Press.

Saitou, N., Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution, 4* (4), 406–425. doi: 10.1093/oxfordjournals.molbev.a040454.

Seino, H., Shiotsuki, Y., Oya, S., & Hirai, Y. (1987). Prediction of long distance migration of rice planthoppers to Northern Kyushu considering low-level jet stream. *Journal of Agricultural Meteorology, 43* (3), 203–208. doi: 10.2480/agrmet.43.203

Sekimura, T., Fujihashi, Y., & Takechi, Y. (2014). A model for population dynamics of the mimetic butterfly *Papilio polytes* in the Sakishima Islands, Japan. *Journal of Theoretical Biology, 361,* 133–140. doi: 10.1016/j.jtbi.2014.06.029

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics, 139* (1), 457–462.

Smouse, P. E., Long, J. C., & Sokal, R. R. (1986). Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology, 35* (4), 627–632. doi: 10.2307/2413122

Suyama, Y., & Matsuki, Y. (2015). MIG-seq: an effective PCR-based method for genome-wide single-nucleotide polymorphism genotyping using the next-generation sequencing platform. *Scientific Reports, 5* , 16963. doi: 10.1038/srep16963

Tajima, F. (1989a). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics, 123* (3), 585–595.

Tajima, F. (1989b). The effect of change in population size on population DNA polymorphism. *Genetics, 123* (3), 597–601.

Tsurui-Sato, K., Sato, Y., Kato, E., Katoh, M., Kimura, R., Tatsuta, H., & Tsuji, K. (2019). Evidence for frequency-dependent selection maintaining polymorphism in the Batesian mimic *Papilio polytes* in multiple islands in the Ryukyus, Japan. *Ecology and Evolution, 9* (10), 5991–6002. doi: 10.1002/ece3.5182

Turner, J. R. G. (1978). Why male butterflies are non-mimetic: natural selection, sexual selection, group selection, modification and sieving. *Biological Journal of the Linnean Society* , *10* (4), 385–432. doi: 10.1111/j.1095-8312.1978.tb00023.x

Uesugi, K. (2000). Avoiding attack in adult butterflies. In N. Ohsaki (Ed.), *The Natural history of butterflies* (pp. 106–123). Sapporo: Hokkaido University Press. (In Japanese.)

Vane-Wright, R. I. (1984). The role of pseudosexual selection in the evolution of butterfly colour patterns. In R. I. Vane-Wright & P. R. Ackery (Eds.), *The Biology of Butterflies* (pp. 251–253). New York: Academic Press.

Wallace, A. R. (1865) On the phenomena of variation and geographical distribution as illustrated by the Papilionidae of the Malayan region.*Transactions of the Linnean Society of London, 25* , 1–71.

Wickler, W. (1968) *Mimicry in plants and animals* . New York: World University Library, McGraw-Hill.

Wright, S. (1943). Isolation by distance. *Genetics, 28* (2), 114–138.

Zhang, W., Westerman, E., Nitzany, E., Palmer, S., & Kronforst, M. R. (2017). Tracing the origin and evolution of supergene mimicry in butterflies. *Nature Communications, 8* (1), 1269. doi: 10.1038/s41467-017-01370-1

Data Accessibility

Author Contributions

KT-S and YS designed the study. YS, KT-S, and MK performed the research. YS, KT-S, and HT analyzed data. YS and KT-S drafted the manuscript. RK, HT, and KT supervised the research. All authors contributed to writing the manuscript.

Figure legends

Figure 1. Mimicry system, sampling sites, and estimated population genetic structure of *Papilio polytes*. **(a)**Mimetic females of *P. polytes* resemble the model species*Pachliopta aristolochiae* , while non-mimetic females are similar in appearance to males. **(b)** Map of the five islands where sampling was performed in this study (denoted by thick outline): KIK, Kikai; OKI, Okinawa; MYK, Miyako; ISG, Ishigaki; TKT, Taketomi. Pie charts show the averaged population composition assigned to individuals of the island at K (number of populations) equaling 7 in the population assignment tests shown in panel (c). Color chart indicates corresponding population numbers 1-7. **(c)** Individual-level population compositions estimated by the population assignment tests at K equaling 2-7. Colors within the bar chart indicate respective populations 1-7 as denoted in panel (b). Mimetic individuals are denoted as white outlined letters in the sample names. **(d)** Coordinates of two-dimensional nMDS (nonmetric multidimensional scaling) of the SNP profiles. K1-K38, H1-H120, M3-M40, IS1-IS209, and T121-T153 are the samples from the Islands KIK, OKI, MYK, ISG, and TKT, respectively. Horizontal and vertical axes correspond to the estimated two-dimensional coordinates 1 and 2 where ranked differences in similarity scores among individuals were preserved. The normalized stress value of this plot was 0.320. Mimetic individuals are outlined in black.

Figure 2. An individual-level phylogenic network and a population-level phylogenetic tree of *P. polytes* from the five islands. **(a)** Phylogenetic network estimated from concatenated SNP site data. Colors indicate the island where the sample was collected: blue, KIK; pink, OKI; green, MYK; orange, ISG; yellow, TKT. Mimetic individuals are denoted as white outlined letters in the sample names. N1-2 and S1-3 indicate subclades within northern and southern clades, respectively. **(b)** The population-level tree among the five islands estimated by the neighbor-joining method (Saito & Nei 1987) based on the genetic distances D_A among islands calculated from concatenated SNP site data. The scale bar below the tree indicates the genetic distances in units of the net number of nucleotide differences of the SNP sites.

Figure 3. Probability density of possible coalescent trees among*P. polytes* individuals from the five islands. Colors indicate the island where the sample was collected: blue, KIK; pink, OKI; green, MYK; orange, ISG; yellow, TKT. Mimetic individuals are denoted as white outlined letters in the sample names. Samples were ordered according to those of the phylogenetic network shown in Figure 2. N1-2 and

S1-3 indicate subclades implied from the phylogenetic network and indicated in Figure 2. Green thick bar shows the clade structure implied from this coalescent-tree probability density.

Island	# individuals	
KIK	11	
OKI	46	
MYK	11	
ISG	14	
TKT	11	
[a]Average (± s.d.) observed heterozygosity over polymorphic sites.	[a]Average (± s.d.) observed heterozygos	
[b]Average (± s.d.) expected heterozygosity over polymorphic sites.	[b]Average (± s.d.) expected heterozygos	
[c]Probability that two randomly chosen alleles at a homologous locus are different.	[c]Probability that two randomly chosen	

Table 1. Population genetic statistics of SNP loci/sites used in the analysis of the present study.

OKI	
MYK	
ISG	
TKT	

Note. Values on the left and right side of the slash denote pairwise F_{ST} (Slatkin 1995) and Φ_{ST} (Nei 1987) between each pa

Ταβλε 2. Παιρωισε $\Phi_{\Sigma T}$ ανδ $\Phi_{\Sigma T}$ οφ ΣΝΠ σιτες αμονγ τηε φιε ισλανδ ποπυλατιονς.

Table 3. Mantel tests for mimic ratio differences (MRD) among the five islands.

Independent variable	
AGD (ncSNPs)	
AGD (mtDNA)	
GD	
GD	
GD	
ED1	
ED2	

Note. Results of the present study are denoted by bold faces. Average genetic distances (AGD) based on nuclear SNP (ncSN

Table 4. Partial Mantel tests between mimic ratio (MR) and advantage index (AI) controlling for other factors.

Association	
MR-AI	
MR-AI [Uesugi]	

13

Note. Results of the present study are denoted by bold faces. Average genetic distances (AGD) based on nuclear SNP (ncS

Supplementary figure legends

Supplementary Figure S1. Examination of parameter settings of *de novo* SNP calling analysis using the Stacks software. These analyses were performed using preliminary sequence data of eight *P. polytes* individuals from all five islands with an average of 129,999 ± 5,245 (S.E.) of quality-filtered forward reads per sample. **(a)** Relationship between the values of minimum coverage depth and the obtained number of SNP loci and their grand mean depth across loci and individuals. Grand mean depth increased and obtained number of loci decreased when the threshold values of minimum coverage depth became higher. **(b)** Relationship between the maximum number of mismatches (in base pairs) between alleles of the same individual and the number of obtained SNP loci. When the threshold of mismatches allowed between alleles increased, the number of obtained loci decreased (over-merged), and vice versa. **(c)** Relationship between the maximum number of mismatches (in base pairs) among alleles from different individuals and the number of consensus SNP loci obtained. When the threshold of mismatches allowed among alleles increased, the number of consensus loci decreased (over-merged), and vice versa.

Supplementary Figure S2. Molecular phylogenic tree of *P. polytes* from the five islands. The neighbor-joining tree was estimated from the concatenated SNP site data. Numbers on the tree indicate support values for the node of the tree estimated from 1,000 bootstrap replications. Colors indicate the island where the sample was collected: blue, KIK; pink, OKI; green, MYK; orange, ISG; yellow, TKT. Mimetic individuals are denoted as white outlined letters for the sample name. N1-4 and S1-3 indicate subclades within northern and southern clades, respectively, implied from this phylogenetic tree. The scale bar indicates the estimated evolutionary distances in units of the number of nucleotide substitutions per site.

Supplementary Figure S3. Probability density of possible coalescent trees among *P. polytes* individuals from the five islands. Colors indicate the island where the sample was collected: blue, KIK; pink, OKI; green, MYK; orange, ISG; yellow, TKT. Mimetic individuals are denoted as white outlined letters for the sample name. Samples were ordered according to those of the neighbor-joining phylogenetic tree shown in Supplementary Figure S2. N1-4 and S1-3 indicate subclades implied from the phylogenetic tree and indicated in Supplementary Figure S2. Green thick bar shows the clade structure implied from this coalescent-tree probability density.
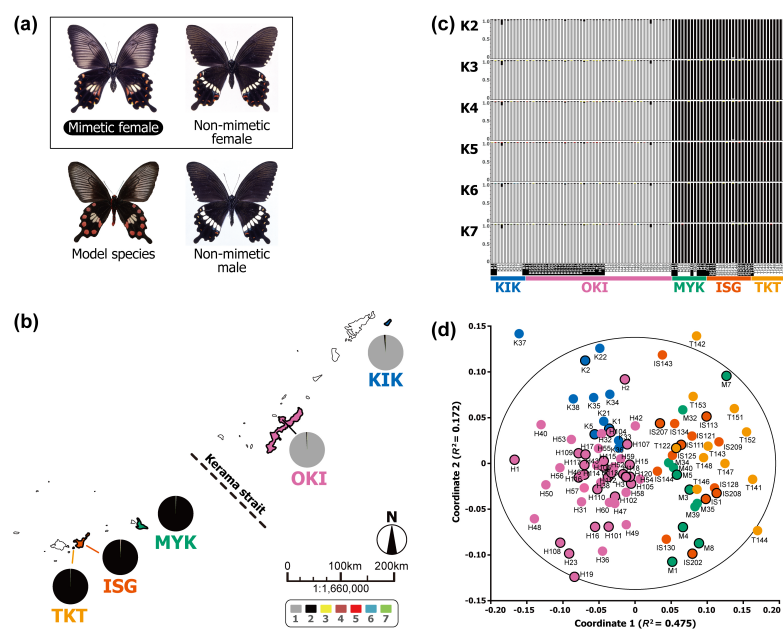
Figure 1

15

(a)

Northern clade

Southern clade

N1

H108 H48

H23 H16 H102 H57 H50

H1 H117 H116

H114 H38 H113 H46

H2 H115 H54 H113

H31 H43

H32 H113

H40 H109

H56

H17

H110

H120

H52

H3

H53 H107

H104 H112

H47

H42

H101

H60 H105 H89

H55 H106

H19

H36

K36 K33

K1 K5 K21

K35 K34

K2 K38

K22

K37

N2

S1

IS202 M1

M4 M35 M8

M5 M39 T152 T144

M40 T146 S2

M34 IS208

M3 IS1

IS128 T151

T143 T141

T147

T148

T153

IS209

IS121

IS134

IS144 IS111 T142

IS125 IS113 M7

M32

IS207 T142

IS130 IS143

S3

H58

H49

(b)

KIK

OKI

Northern clade

MYK

ISG

Southern clade

TKT

0.50

16