Title Page

Title

Phylogenomics reveals an island as a genetic reservoir of a continental population

Authors

AOKI Daisuke[1,*], SENZAKI Masayuki[2], ANDO Haruko[3], ODAYA Yoshiya[4], HEIM Wieland[5,6], KITAZAWA Munehiro[3,7], WULF Tom[8], TRENSE Daronja[9], BASTARDOT Mark[10], FUKUDA Atsunori[11], TAKAHASHI Masao[12,13], KONDO Natsuko[3]

[1.] Department of Wildlife Biology, Forestry and Forest Products Research Institute, Tsukuba, Japan 305-8687

[2.] Faculty of Environmental Earth Science, Hokkaido University, Sapporo, Japan, 060-0810

[3.] Biodiversity Division, National Institute for Environmental Studies, Tsukuba, Japan, 305-8506

[4.] Abiko City Museum of Birds, Abiko, Japan, 270-1145

[5.] Swiss Ornithological Institute, Sempach, Switzerland, 6204

[6.] Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland, 8006

[7.] Graduate School of Agriculture, Hokkaido University, Sapporo, Japan, 060-0809

[8.] Department of Agriculture, Ecotrophology, and Landscape Development, Anhalt University of Applied Science, Bernburg, Germany, 06406

[9.] Department of Biology, Institute of Integrated Natural Sciences, Koblenz-Landau University, Koblenz, Germany, 56070

[10.] Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland, 1015

[11.] The Wild Bird Society of Japan, Ibaraki Branch, Mito, Japan, 310-0055

[12.] Iwate Prefectural Museum, Morioka, Iwate, Japan, 020-0102

[13.] Oosekka-land, Hachinohe, Aomori, Japan, 031-0823

* Corresponding author: Daisuke AOKI, aokid@ffpri.affrc.go.jp, +81-29-829-8258

Current address

HEIM Wieland: Institute of Biology and Environmental Sciences, University of Oldenburg, Oldenburg, Germany, 26129

ODAYA Yoshiya: Chiba Natural History Museum and Institute, Chiba, Japan, 260-0852

OCRID

AOKI Daisuke: 0000-0002-5780-8487

SENZAKI Masayuki: 0000-0002-3843-5793

ANDO Haruko: 0000-0001-7827-3883

ODAYA Yoshiya: 0009-0004-1057-4005

HEIM Wieland: 0000-0002-3262-2491

KITAZAWA Munehiro: 0000-0003-0167-4464

TRENSE Daronja: 0000-0003-3598-8699

TAKAHASHI Masao: 0000-0001-8385-6098

KONDO Natsuko: 0000-0002-0883-0506

Data availability statement

Scripts and data for analysis are available on Dryad. Raw sequence reads are deposited to Sequence Read Archives (see Table S1). Settings and procedures for these processes are fully described in the main and supplementary texts and the original scripts.

Author Contributions

DA and MS conceived the ideas and designed the study with contributions by WH; MS, YO, WH, MK, and AF led the field work with substantial support by DA and MT; HA, NK, and DA led the laboratory procedures; DA conducted bioinformatics and analyzed the data with supports by HA and NK; DA coordinated the study, prepared the draft, led the writing and revised the manuscript with assistance from MS. All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Funding

This study was partly conducted with the support of a grant-in-aid for Scientific Research (C) to DA (no. 22K20670) and MS (no. 23H02243) from the Japanese Society for the Promotion of Science (JSPS).

Conflict of interest statement

The authors declare no conflicts of interest.

Acknowledgements

Some tissue and blood samples were kindly provided by the Yamashina Institute for Ornithology. We thank many members of our field teams; Toshio Sadakuni for sampling in Kushiro, Akio Miya and Jun-ichi Ebina for sampling at Hotokenuma, Aomori; Sergei M. Smirenski and Yury Anisimov for sampling in Russia; Shigeo Ozawa, Ryosuke Abe, Koichi Shirakawa, Daisuke Akoshima, Ayumu Hamachi, Hiroko Shirakawa, Mikiya Oikawa, and Shiomi Hakataya for sampling in the Kanto region. We also thank Nobuyoshi Nakajima for his support on sequencing runs on MiSeq. Our field surveys were conducted with permission from the landowners of each site, including Tomatoh. DNA samples were collected with permission from the Ministry of the Environment Government of Japan (permission # includes 21-31-0172-74, 19-129, 1810254, 2012217-16). We also acknowledge the provision of rings by the Bird Ringing Centre of Russia. Transportation of DNA samples was conducted under a Material Transfer Agreement between the National Institute for Environmental Studies and the Amur Bird Project based on the Access and Benefit-Sharing (ABS) agreement in the Nagoya Protocol. Bioinformatics was mostly performed on the supercomputer of AFFRIT, MAFF, Japan.

Abstract

Islands as biodiversity sinks due to their small population size and low genetic variation are recently challenged by a phenomenon where a continental lineage originates from islands. This raises an important question about through what demographic processes insular genetic variation is introduced to the continent and its evolutionary consequences on the continent. We address this question by focusing on an Asian rail species, the Swinhoe's Rail *Coturnicops exquisitus*, where frequent trans-oceanic dispersal in a continent-island system is expected. We reconstructed a spatiotemporally explicit demographic history of this species in the continent-island system, using integrative phylogenomics. Multiple lines of evidence supported asymmetric gene flow from an island population to the continental population that once diverged around the Middle Pleistocene. This has possibly impacted the continental genetic variation because ancient genetic variation may have been preserved in the large island population. Re-establishment of the island by the continental population was also detected, suggesting that the insular genetic variation has been maintained also by receiving a part of continental genetic variation. These results indicated that the maintenance of insular genetic variation through dynamic continent-island metapopulation may have allowed the island to act as a genetic reservoir for continental genetic variation.

Teaser text

What process allows island populations that are small and at the “evolutionary dead-end” to become the source of continental colonization? Using a cutting-edge integrative phylogenomics approach, we disentangled the demographic history of the little-studied Swinhoe's Rail in East Asia, where it has survived by maintaining a dynamic metapopulation across a continent-island

system. Particularly, gene flow from the island impacted the contemporary genetic variation of the continental population, possibly facilitating its demographic expansion. We suggested that maintenance of metapopulation and asymmetric gene flow from islands are key for islands to introduce their genetic variation to the continent.

Keywords

Coturnicops exquisitus, genetic variation, gene flow, island biogeography, metapopulation, reverse colonization

Main Text

Introduction

Island biological diversity contributes disproportionately to global biodiversity (Whittaker and Fernández-Palacios 2007). Islands, comprising 3.5% of the terrestrial land, contribute nearly 20% of global species richness and contribute more than 60% of extinct terrestrial species (Whittaker et al. 2017). Meanwhile, it has been assumed that the small and isolated landmass of the island causes a small population size with low genetic variation (Cerca et al. 2023), likely leading to high risks of extinction (Frankham 2005) and the "evolutionary dead-end" (Vamosi et al. 2014). Based on these microevolutionary assumptions, islands have been viewed as sinks where lineages become left behind by the global diversification processes (Bellemain and Ricklefs 2008).

Recently, a phenomenon through which the island becomes the source of continental colonization has been increasingly evident (Whittaker et al. 2017). Macroevolutionary studies have disentangled many cases of continental species or lineages that nest within island species radiation in phylogenetic trees, which is referred to as "reverse colonization" (Bellemain and Ricklefs 2008). A previous review reported that more than a third of the studied phylogenies of a various range of taxa in many continent-island systems exhibited the reverse colonization patterns (Bellemain and Ricklefs 2008), including birds and reptiles from West Indies to North or South America (Bellemain and Ricklefs 2008; Patton et al. 2021) and *Scaptomyza* flies from Hawaii to Asia (O'Grady and Desalle 2008). Mainland expansion and diversification can also be triggered by reverse colonization. For example, the radiation of passerines, now comprising

almost 60% of the modern global bird species, began with out-of-New Guinea colonization (Jønsson and Holt 2015).

To understand why reverse colonization is more common than previously thought, a population genetics (i.e., microevolutionary) study needs to investigate a demographic process whereby insular genetic variation is introduced to the continent. Genetic variation of a population at one place can be translocated to another via migration in the form of either population establishment or gene flow (Yamaguchi 2022), thus, some particular cases of these processes may be key to allowing insular organisms to back-colonize continents. To our knowledge, there has been only one study that modeled the population genetic process of continental establishment by insular organisms (Patiño et al. 2015). This study, however, used bryophytes, whose genetic variation may not be sensitive to demographic changes due to their abilities of asexual reproduction (Ellegren and Galtier 2016), leaving why many sexually reproductive animals exhibit reverse colonization uncovered. Furthermore, gene flow in the continent-island system has received much attention in the context of divergence-with-gene-flow, i.e., how island populations can diverge in the presence of gene flow (Warren et al. 2015). Their major focus has been genomic variation that escaped from the homogenizing effect of gene flow (Feder et al. 2012). As a result, studies investigating historical “movements” of genetic variation from islands to continents and its demographic consequence in the continent are lacking.

Phylogeography provides a suitable framework to study the directionality and demographic consequences of population establishment and gene flow in the continent-island system because it allows us to infer a historical process that is responsible for the distribution of present genetic variation (Lomolino et al. 2017). Phylogeography is especially powerful in reconstructing a spatiotemporally explicit history when it integrates demographic reconstruction

and species distribution modeling (SDM). Phylogenetic trees, which phylogeography is primarily based on, provide patterns that emerged as a result of demographic processes, while the processes themselves remain unmodeled (Harvey et al. 2019). At this point, demographic reconstruction using high-throughput sequencing (HTS) directly models demographic processes that resulted in the present genetic variation (Nadachowska-Brzyska et al. 2022). SDMs can estimate the timing and extent of historical range shifts with spatiotemporal information (Harvey et al. 2019). Therefore, these methods can directly estimate or strengthen the inference of the direction, intensity, and timing of population establishment or gene flow.

Here, we apply integrative phylogeography to investigate the directionality of population establishment and gene flow in a continent-island system and its demographic consequences in the continent. To achieve this goal, we selected the Swinhoe's Rail (*Coturnicops exquisitus*), the smallest Asian rail species, as the ideal model case. Demographic processes strongly determine the present genetic variation of birds compared with many other taxonomic groups (Ellegren and Galtier 2016), while they represent many cases of reverse colonization (Bellemain and Ricklefs 2008). Therefore, we expect an important demographic process underlying the phylogeography of birds in the continent-island system. Rails are the representatives that underwent repeated trans-oceanic dispersal events, inferred by ancestral area reconstruction on a phylogenetic tree (Garcia-R and Matzke 2021). A study has revealed that the Swinhoe's Rail originated in Asia through transcontinental colonization around the Pliocene to the Early Pleistocene from an ancestor in North America (Garcia-R and Matzke 2021). Therefore, we can expect that the Swinhoe's Rail is prone to frequent colonization events across the sea, which is suitable for our study purpose.

In this study, we reconstruct a phylogeographic history of the Swinhoe's Rail and ask whether insular genetic variation is introduced to the continent via either population establishment and/or gene flow. We determine the directionality of population establishment by examining 1) the basality of an insular population and nestedness of a continental population within it in phylogenetic trees, 2) the demographic stability of populations, and 3) the presence and absence of suitable habitats at the time of population divergence. We assume that a derived population arises with a strong bottleneck from a stable ancestor at the time of a population split (e.g., Yeung et al. 2011). We determine the directionality of gene flow by 1) inspecting phylogenetic signals of gene flow, 2) directly estimating gene flow by demographic modeling, and 3) predicting historical range shifts of populations that gave rise to migrations by the SDM. Finally, we discuss the impact of these events on the population size and genetic variation in the continent.

Materials and methods

Field sampling

The Swinhoe's Rail is a poorly studied species. While breeding evidence is lacking for its previously known breeding ranges on the continent (Fig. 1a, dark gray areas) (Heim et al. 2019), the presently recognized breeding areas were re-discovered recently, including the Amur region (Heim et al. 2019) and Lake Baikal (Anisimova et al. 2019) in Russia and some regions of Japan, namely Aomori (Miya et al. 2005) and Hokkaido (Senzaki et al. 2020) (Fig. 1a). Despite its cryptic lifestyle inhabiting inaccessible wetlands, we developed our original sampling strategies (Heim et al. 2019; Senzaki et al. 2020) and collected feather, buccal swab, or blood samples successfully covering the entire extent of the known breeding area, Amur (n = 6) and Baikal (n =

1) for continental Russia, Kushiro (n = 12) and Tomakomai (n = 8) for Hokkaido Island, and Aomori (n = 4) for Honshu Island during their breeding season (Fig. 1a, Table S1). We also captured wintering birds in the Kanto region of Honshu Island (n = 37) and used a genetic sample of a migratory bird in Ishikari of Hokkaido Island (n = 1, see Supplementary Methods), to infer the migratory connectivity of this species and to speculate any historical range shifts. We additionally retrieved mitochondrial data from the NCBI nucleotide database for phylogenetic analyses (Table S1).

Laboratory procedures & Bioinformatics

The whole genomic DNA extracted from each tissue was used in the two experiments, sanger sequencing of the partial mitochondrial Cytochrome *b* (*Cytb*) region for 49 samples and one of the HTS techniques, MIG-seq (Suyama and Matsuki 2015), for 52 samples (Table S1). The latter is a type of genotyping by sequencing with polymerase chain reactions, which is effective for low-quality DNA samples in exchange for smaller genome coverage, best adapted to some of our dried feather or swab samples. MIG-seq resulted in paired-end reads of 4,834,210 overall and 92,966 per-sample average, which were trimmed and aligned against a reference genome of a closely related species, *Laterallus jamaicensis coturniculus* (WGS Accession ID, JAKCOX01, Table S1). After the removal of one breeding individual from Amur and two wintering individuals from Kanto exceeding the missing proportion >30% in the pre-runs of ANGSD (Fig. S1-S4, Supplementary Methods), the final dataset contained 7,396,956 including both paired and unpaired reads for 49 samples.

We mainly adopted probabilistic analysis frameworks to account for uncertainties about the true genotype, by calculating genotype likelihoods (GL) using the software ANGSD

(Korneliussen et al. 2014). We first calculated GLs separately for five regional populations by setting the '-minInd' to be half of the total number of samples for each population. Then, for different downstream analyses, GLs were repeatedly calculated using the 'SAMtools' model on the intersecting sites of different data subsets and populations using different filtering parameters, which resulted in different numbers of sites used (Table 1, S2). We specified different values for the minimum depth ('-setMinDepth'), depending on how downstream analyses use GLs, including 1) 4 for those directly using GLs (PCA, ADMIXTURE, distAngsd, RAxML, the D-statistic), 2) 7 for those using allele frequency estimated by GLs (TreeMix, diversity index calculation, Stairway Plot 2, Fastsimcoal2), and 3) 10 for the SNAPP analysis that uses called SNPs. These values were determined by the pre-analyses (Fig. S1-S4, Supplementary Methods). For those using only variant sites, we set 1e-06 for '-SNP-pval' and 0.05 for '-minMaf', and for SNAPP specifically, genotypes were called using the '-doGeno' function of ANGSD with 0.9 for '-postCutoff. Populations were defined by the five sampling regions (Fig. 1a) or two genetic breeding groups, namely island (Kushiro + Tomakomai) and continental (Russia + Aomori) (Table 1). The latter was determined based on our genetic structure analyses (Fig. 1b–c, Fig. S6-S10). Linkage disequilibrium among variant sites was calculated and pruned using ngsLD (Fox et al. 2019) for analyses that do not account for linkage disequilibrium (PCA, ADMIXTURE, and SNAPP). For demographic analyses, we used folded site frequency spectra (SFS) and two-dimensional folded SFS (2d-SFS), calculated by the '-doSAF' option of ANGSD and the 'realSFS' function. Overall analytical flow and parameter settings for ANGSD and its related analyses are described in Supplementary Methods and Fig. S5.

Genetic structure

We examined global genetic structure by PCA using PCAngsd (Meisner and Albrechtsen 2018), ADMIXTURE analysis using NGSadmix (Skotte et al. 2013) with K ranging from 1 to 6, and NeighborNet (Bryant and Moulton 2004) using pairwise genetic distance calculated by distAngsd (Zhao et al. 2022) (see Supplementary Methods for settings and details of these analyses). We calculated the D-statistic for each trio of four breeding populations using the 'doabbababa2' program of ANGSD (Soraggi et al. 2018) to confirm the clusters found in these analyses. We included two species of the American *Laterallus* group (*Laterallus jamaicensis,* Accession SRX14572910 and *Atlantisia rogersi*, Accession SRX6608087) as outgroups in the distAngsd analysis as well as supplementary analyses of PCA and ADMIXTURE (Fig. S6-S8). After we found that these species form an evolutionary unique cluster to the Swinhoe's Rail, we included only *Laterallus jamaicensis* as an outgroup in phylogenetic reconstruction.

Inference of phylogenetic relationship

Phylogenetic trees reconstructed using different datasets and methods may reflect gene flow and incomplete lineage sorting (ILS) differently (Twyford and Ennos 2012). Therefore, we reconstructed a phylogenetic history by comprehensively interpreting the results of four analyses, a mitochondrial phylogenetic tree (mtDNA tree), a maximum likelihood tree (RAxML tree), a graph-based population tree using allele frequency data (TreeMix), and a coalescent-based population tree using SNP data (SNAPP tree).

mtDNA tree: The time-calibrated Bayesian inference method was applied to an alignment of 772 bps of partial *Cytb* by using an uncorrelated lognormal relaxed clock model and the coalescent exponential population prior in BEAST v.2.7.3 (Bouckaert et al. 2019). Samples

included 12 haplotypes from 54 Swinhoe's rails and three outgroups (two *C. noveboracensis* and one *L. jamaicensis*). The standard molecular clock rate of 1.05%/lineage/Myr (Weir and Schluter 2008) was applied to estimate divergence times. A Markov chain Monte Carlo (MCMC) analysis was run for 2e+07 generations, and the tracelog and treelog files and trees were sampled every 1,000 generations. Convergence was checked on the log file by following Aoki et al. (2018). A maximum clade credibility tree was drawn with mean heights after 10% of the sampled trees were discarded as burn-in. We confirmed the inferred phylogenetic tree by reconstructing a haplotype network. For other parameter settings and the haplotype network, see Supplementary Methods.

RAxML tree: A maximum likelihood tree was estimated on a dataset consisting of both variant and invariant sites, using RAxML-ng (Kozlov et al. 2019) with the 'GTGTR4+G' model. We used GLs of ANGSD output as its input file (CATG format) by normalizing GLs so that all the ten possible genotype states add up to 1 for each site, and turning '--prob-msa' on. Standard non-parametric bootstrapping with 100 replicates was simultaneously conducted.

TreeMix: TreeMix (Pickrell and Pritchard 2012) was used to infer a population-level tree that simultaneously considered gene flow and genetic drift. We used the glactools software (Renaud 2017) to convert a VCF output from an ANGSD analysis to get an ACF file that contains estimated allele frequencies based on GLs for all sites. We set '--noprivate' in the 'acf2treemix' program to obtain a TreeMix input file for variant sites by removing sites that are polymorphic only to a certain population. Each migration edge ranging from zero to five was tested for ten replicates with a block size of 28 for jackknife sampling. We determined the block size as the half-life of an LD decay curve estimated by ngsLD (Supplementary Methods, Fig. S14). We used an R package 'OptM' v. 0.1.6 to determine the best number of migration edges scoring the

highest Δm score (Fitak 2021), at which a run with the highest model likelihood was selected as the best-fit model. Model likelihoods, total variance in population relatedness, and residual fits were also evaluated to determine model performance (Pickrell and Pritchard 2012; Fitak 2021), and model consistency was also checked among different replicates at the best number of migration edges.

SNAPP tree: SNAPP infers a population tree based on a coalescent model using unlinked SNPs without missing data (Bryant et al. 2012). Therefore, we selected three or four samples that scored the lowest missing proportion from each breeding population (three from Aomori due to its small sample size) and obtained unlinked hard-called SNPs by ANGSD. We adopted an approach that infers the tree topology while simultaneously estimating divergence times (Stange et al. 2018). We used two different log-normal distributions (lognormal(5.11, 0.09) and lognormal(8.95, 0.0225) in million years ago, Mya) as constraints for the calibration node (a split between the Swinhoe's Rail and *Laterallus jamaicensis*), which corresponded to its divergence time estimated by our clock-calibrated mtDNA tree (Fig. 2a) and fossil-calibrated mtDNA tree by Garcia-R et al. (2020), respectively. We performed 500,000 MCMC iterations to get posterior trees.

Demographic inferences

We reconstructed the historical demography of the continental and island populations based on the results of summary statistics, demographic trajectory inferences (Stairway Plot 2, SWP2), and parameter estimations in demographic models (fastsimcoal2, FSC2). These methods are complementary to each other because SWP2 is a nonparametric model that can flexibly infer complex effective population size (N_e) changes over time yet without modeling gene flow, while

FSC2 has an advantage of jointly modeling gene flow and N_e changes but in a simplified user-defined model (Nadachowska-Brzyska et al. 2022). Because the contemporary N_e may differ from the model estimates due to recent demographic changes that may not be detected by some of the models (Nadachowska-Brzyska et al. 2022), we estimated the contemporary N_e of the island population by converting estimated census population size (N_c) in a plain of Hokkaido in early nineteenth century (Kitazawa et al. 2022), using a mean ratio of N_c:N_e = 1:0.44 for birds (Frankham 1995).

Summary statistics: We calculated Watterson's theta θ_W (Watterson 1975) and Tajima's *D* (Tajima 1983) per reference scaffold and population. Values of Tajima's *D* around zero indicate higher demographic stability, and values below zero (especially <-1) indicate a strong ancient bottleneck (Gattepaille et al. 2013). Furthermore, a large variance of Tajima's *D* among scaffolds indicates a moderate bottleneck, while a small variance indicates a population expansion (Gattepaille et al. 2013). Per-scaffold estimates of these indices were compared between populations, using linear mixed models. We used log-transformed scaffold size and population identity as the explanatory variables, where the island population was set as the base of a dummy variable to be compared with the continental population (see Supplementary Methods).

SWP2: We separately estimated demographic changes for the continental and island populations based on their one-dimensional folded SFSs, using Stairway Plot 2 v. 2.1.1 (Liu and Fu 2020). We assumed a mutation rate of 1.91e-09 (Nam et al. 2010) and a generation time of 1 year, and analyses were run with default settings. The timing that populations started to show different demography from one another could be interpreted as population divergence. We clipped the trajectories of the last 1,000 years off from the plot due to large uncertainties.

FSC2: We tested five models that differ from each other in terms of the presence and timing of migration matrices (Fig. 4a). We estimated t_{MIG} at which both the continental and island populations experienced demographic changes and migration matrices may have switched, by fixing their divergence time at 0.5 Mya that was collectively agreed by multiple approaches (Fig. 2a, 2d, 4c). Between the two populations, we modeled no migration throughout ("no migration"), migration only after ("recent migration") or before ("ancient migration") t_{MIG}, and migration throughout at the same rates ("constant migration") or different rates ("different migration") across t_{MIG}. We conducted 100 different replicates of 1,000,000 coalescent simulations in each of 40–100 optimization ECM cycles for parameter estimates, using fastsimcoal2 v. 2.7.0.9 (Excoffier et al. 2021). The same mutation rate as the SWP2 analyses was used. A simulation that scored the smallest difference between 'MaxEstLhood' and 'MaxObsLhood' was selected from each model, and models with the ΔAIC (delta Akaike Information Criterion, the difference between the lowest AIC and an AIC in focus) <2 were selected as the best-fit models. The ranges of parameter values for simulation were described in Table S5.

SDMs

Because the known breeding range of the Swinhoe's Rail is restricted to only a handful of localities, we modeled that of its sister species, *C. noveboracensis* distributed in North America, and projected it to the present and past climatic environment conditions of East Asia to estimate the distribution of historical suitable habitats of the Swinhoe's Rail. This benefits from utilizing a large dataset from eBird (https://ebird.org/) from North American countries, while this is based on an assumption of niche conservatism that the two sister species retain the environment niche of their common ancestor. We justify this assumption because they show strong similarities in morphology and breeding ecology and have been treated as a single species until recently

(Senzaki et al. 2020; Taylor and Sharpe 2020). Fossils of their ancestral species in North America occurring 5.5 Mya also resembled the present Yellow Rail (Bickart and Chandler 1990), suggesting that the ecology of this species group has been probably evolutionary stable.

The SDM was applied to the eBird dataset of *C. noveboracensis* using MaxEnt (Phillips et al. 2006). We used only occurrences of this species collected between June 1st to August 24th to model its breeding distribution. We generated a "zero-filled" dataset following the instruction of an R package auk (Strimas-Mackey et al. 2018), and we used the presence (1,345 records) and pseudo-absence points (10,000 records randomly sampled out of 7,663,704 total records) of this species per observation event and site as occurrence and background inputs, respectively. We used an R package 'pastclim' v. 1.2.3 (Leonardi et al. 2023) to extract time slices of 22 variables (Krapp et al. 2021) and selected one variable per group of variables correlated >0.7, resulting in seven covariates for modeling (see Supplementary Methods). We tested 12 models with different parameter settings by using the function 'ENMevaluate' implemented in an R package ENMeval v. 2.0.4 (Kass et al. 2021), and the best model was selected by the lowest ΔAIC for a small sample size (AICc). The best model was projected to present and six past time slices of environmental variables, consisting three Late Pleistocene periods (0.019 (Last Glacial Maximum, LGM), 0.05, 0.122 Mya) and three late Middle Pleistocene periods (0.434, 0.461, and 0.491 Mya), which corresponded to glacial-interglacial ranges of t_{MIG} estimated in the FSC2 (Fig. 4b-c) and population divergence time (Fig. 2a, 2d, 4d), respectively. Details about SDM are described in the Supplementary Methods.

Results

Phylogenetic relationship and gene flow

The multiple phylogenetic trees based on different methods collectively inferred a history of population divergence between the continent and islands and post-divergence gene flow from islands to the continent. The population divergence was supported by our mtDNA tree (Fig. 2a), TreeMix (Fig. 2c), SNAPP tree (Fig. 2d), and the D-statistic (Fig. S10) that exhibited bifurcating topologies between one clade consisting of Tomakomai and Kushiro samples (insular clade) and the other clade consisting of Russia and Aomori samples (continental clade). This was also supported by ADMIXTURE at K = 2 which scored the highest ΔK value (Fig. 1c, Fig. S7-S8). Post-divergence gene flow was suggested from the insular to continental clades by our mtDNA tree, given that Russian samples were found equally in both clades, while samples collected from Tomakomai and Kushiro were found almost exclusively from the insular clade (with one exception from Kushiro) (Fig. 2a, Fig. S11). Although our RAxML tree (Fig. 2b) did not show the bifurcating topology, it can be also explained by the asymmetrical gene flow from islands to the continent, given that Russian (and two Aomori, see below) samples nested within a clade of Tomakomai and Kushiro samples. The divergence time of the two clades was estimated at around 0.5 Mya, during the Middle Pleistocene (the Chibanian epoch), which was supported by different calibration methods, data, and algorithms, including the mtDNA tree (Fig. 2a), the SNAPP tree (Fig. 2d), and SWP2 (Fig. 4c).

Gene flow was also inferred by the best-fit TreeMix model with migration edge = 1 scoring the highest Δm value (Fig. S15). Meanwhile, the gene flow was directed from the common ancestor of this species to Russia, differently from the gene flow inferred by the other

trees. Furthermore, in the different replicates of migration edge = 1, models yet with lower log-likelihoods showed gene flow between different populations (Fig. S18). A model already scored a high variance explained >0.998 without any migration edge (Fig. S15, S17) where the Russian population was indicated as the basal while it did not well corroborate tree topologies from the other analyses (Fig. S16). These results may be explained by the possibility that ancient genetic variation was introduced to the Russian population by gene flow, while the source remains unknown from this analysis (see Supplementary Discussion).

Aomori samples showed a high affinity to and a low level of genetic differentiation from Russian samples in all the phylogenetic trees (Fig. 2), PCA, ADMIXTURE (Fig. 1b-c), NeighborNet (Fig. S9), and the D-statistic (Fig. S10), suggesting that Aomori was recolonized by the continental population. This was less likely an artifact due to its small sample size, because it was supported even in analyses using absolute metrics (e.g., genetic distance for NeighborNet), and they were not assigned to a meaningful genetic cluster that was differentiated from Russian samples in the analyses of PCA and ADMIXTURE even when Kanto wintering samples were removed (Fig. 1c, S6-S8). Two Aomori samples, however, clustered with Tomakomai samples in the RAxML tree (Fig. 2c).

Kanto wintering samples were identified as closely related to Hokkaido (especially Kushiro) samples by PCA and ADMIXTURE (Fig. 1b-c). Meanwhile, three of four mtDNA haplotypes from Kanto was found only from Russia or continental clades (Fig. 2a, S11). This may indicate a possibility that there is an unsampled population between Kushiro and continental Russia, thereby continental mitochondrial haplotypes are introgressed while retaining insular genetic variation in nuclear DNA, such as Sakhalin Island.

Historical demography and contemporary genetic variation

The high genetic variation was shown for the continental population in comparison with the island population (Fig. 3a, Table S3), and the FSC2 suggested that gene flow from the island to the continent has been important for shaping contemporary genetic variation. The best-fit models in the FSC2 with ΔAIC <2.0 included the "different migration" model (ΔAIC = 0.0, Akaike model weight = 0.586) and the "constant migration" model (ΔAIC = 0.692, Akaike model weight = 0.414), which supported the presence of gene flow both before and after t_{MIG} = 0.011–0.018 Mya, which corresponded to the LGM (~0.02 Mya) (Fig. 4a-c, Table S4). These two best-fit models agreed on asymmetrical gene flow from the island to the continent, which was 2.1–2.4 times higher than the opposite direction from the LGM to the present. In contrast, the direction and strength of gene flow before the LGM remained inconclusive, due to significant differences between the two models. The "different migration" model illustrated a large inflow of migrants to the insular population, whereas the "constant migration" model estimated constant asymmetrical gene flow from the island to the continent since its population split.

Our demographic analyses suggested that the island population remained stable, or, if any, experienced a modest level of population expansion for a long period since the population split until the LGM, supported by Tajima's *D* values close to zero (Fig. 3b, Table S4), and the 1.5–7.7-fold (best FSC2 models) or 2.2-fold (SWP2) increases in N_e (Fig. 4b–c, Table S3). Meanwhile, discordant results were obtained for the demographic history of the continental population by different methods, obscuring the presence of a population bottleneck for this population. A population expansion after a strong ancient population bottleneck was suggested by the significantly low value of Tajima's *D* close to -1 (median value = -0.919) with a smaller inter-scaffold variance than the island population (Fig. 3b) and the best FSC2 models (Fig. 4b-c).

In contrast, the SWP2 suggested a population expansion soon after the population split around 0.5 Mya without any signature of a population bottleneck.

A significant post-LGM population bottleneck was inferred for the island population by the best FSC2 models, decreasing by 98–99% of its original size (Fig. 4b-c). Our SWP2 results did not contradict it despite its large uncertainties towards the present (Fig. 4d). This was also supported by the small contemporary effective population size of Ishikari Plain in Hokkaido (including Tomakomai) (N_e = 4,840) calculated from the census population size estimates around the early nineteenth century, which well corroborated the results of the FSC2.

SDMs

The best-fit model (response curve 'LQ' + regularization multiplier 1.0, AICc = 2446.725, Akaike model weight = 0.54, AUC = 92.5%) based on the present breeding occurrence data reflected the present distribution of *C. noveboracensis* (Supplementary Results, Fig. S19). The projection of this model to East Asia also well predicted the current distribution of the Swinhoe's Rail (Fig. 5a, arrows), and our assumptions of niche conservatism are thought to be met (see Discussion). Although the wide area of the continental range did not resemble any climate in North America inferred by the negative values in the multivariate environmental similarity surface (MESS) analyses (gray-shaded grids in Fig. 5, see Supplementary Methods), boundaries of these regions were at the edge of present distribution of the Swinhoe's Rail (Fig. 5, arrows), suggesting that a large part of them may correspond to uninhabitable areas for this species. Therefore, we mainly interpreted areas not covered by regions with negative values of the MESS as potential distribution ranges of the Swinhoe's Rail.

Repeated shifts of suitable habitats were suggested from 0.5 Mya to the present. Generally, during each glacial period (Fig. 5-b, c, e, f), suitable habitats were present equally in both the continent and the Japanese Archipelago, yet isolated by the Japan Sea. This corroborated our genetic analyses that suggested a population split event between the continent and island around 0.5 Mya. Meanwhile, during interglacial periods (Fig. 5-a, d, g, including the present), suitable habitats became more abundant in the continent and continuous from island to continent, which also corroborated the genetic inference of migration from the island population to the continent in these periods. Limited suitable habitats in Japan also agreed small contemporary population size on the island. The inference of migration from the island to the continent should not be due to the assumption of niche conservatism. This is because the assumption may lead to an underestimation of the island's suitability, which detracts from the potential capacity of the island as refugia, since details of the prediction in Japan, such as low or no suitability in Tomakomai and Aomori (Fig. 5a), did not well reflect the present distribution.

Discussion

Despite the potential importance of islands as biodiversity sources (Whittaker et al. 2017), whether and how genetic variation of an island population is introduced to the continent remained unexamined (Bellemain and Ricklefs 2008). Using an integrative phylogeography in the Swinhoe’s Rail, we showed higher genetic diversity and a larger population size for the continental population than the island population, while the continental genetic variation was impacted by gene flow from the insular population. These results indicate that gene flow from an island to a continent can potentially play an important role in the microevolutionary consequences of a continental population. In contrast, we were unable to conclude the directionality of population establishment at the time of population split for this species; multiple

phylogenetic trees showed bifurcating topologies, at which colonization direction cannot be unambiguously resolved, and the continental population bottleneck remained ambiguous due to incongruence among demographic analyses.

Islands as potential reservoirs of genetic diversity for continental populations

We suggest that the islands have served as sources where genetic diversity has been long preserved, and its introgression might have been important for restoring genetic variation and facilitating the expansion of the continental population in this species. Before the LGM, the ancstral island population was estimated to have retained a large N_e that was comparable to that of avian continental populations (mean N_e = 362,456) (Leroy et al. 2021) and large tropical island populations, such as New Guinean birds (mean N_e = 325,000) (Pujolar et al. 2022), while the N_e of the ancestral continental population was estimated to be as small as that of oceanic island populations (mean N_e = 94,944) (Leroy et al. 2021) by our FSC2 results. In such a large population, adaptive variation can be more likely to be preserved than in a small population where weakly deleterious mutations accumulate (Leroy et al. 2021), and asymmetric gene flow can directly restore genetic variation and increase population size in such a situation (Fitzpatrick et al. 2016).

Discordance between population size estimates by the FSC2 and SWP2 may be also explained by the presence of strong gene flow to the continent. A demographic trajectory inferred by SWP2 is expected to show an increase in effective population size in the presence of gene flow, even though populations have not experienced any expansion (Nadachowska-Brzyska et al. 2022). Therefore, the genetic bottleneck inferred in the FSC2 analyses might have been concealed in SWP2 by gene flow to the continental population. Alternatively, a lack of

parameters for population-specific demographic change in our FSC2 modeling may have concealed complex fluctuating histories (long-term N_e, Nadachowska-Brzyska et al. 2022), which may have exaggerated the impact of a bottleneck. However, a long-term N_e reflects a harmonic mean of a demographic trajectory (Nadachowska-Brzyska et al. 2022), and hence, the effect of past bottlenecks on the continental genetic variation should not be undervalued.

The impact of the gene flow should not be discounted by the incongruent inferences of gene flow among different phylogenetic trees either. For example, tree topology is affected less likely by gene flow history when using a limited number of SNPs (like the SNAPP tree) than a more comprehensive dataset (i.e., the RAxML tree) (Zhang et al. 2021). The basal position of two Aomori samples in the RAxML tree (Fig. 2b) may be explained by this, i.e., recent but undetected gene flow by the other analyses, while it still remains possible that an artifact due to their relatively high missing data proportions have influenced their positions (Supplementary Results; Fig. S12-S13; Smith et al. 2020). Gene flow, and not ILS, is a likely explanation for the position of Russian samples in the mtDNA tree because the estimated divergence time of the two mitochondrial clades corresponded to the population split time estimated by our analyses on our MIG-seq data, whereas a genealogical split older than a population split is expected for ILS (Twyford and Ennos 2012).

Population persistence through a continent-island metapopulation

We were unable to distinguish the directionality of population establishment, possibly because this species may have survived multiple glacial cycles through a metapopulation in this continent-island system, which we suggest is likely for the following two reasons. First, gene flow from the island to continental populations could have occurred recurrently in each glacial

cycle, because our SDM analyses predicted similar latitudinal range shifts of suitable habitats even before the LGM (Fig. 5e-g). The best-fit FSC models also supported the presence of gene flow during this period, although they significantly differ in direction and intensity (Fig. 4b-c, Table S3). This inconsistency may be related to a limitation of the FSC2 model that cannot parameterize a large number of gene flow and demographic change events (Nadachowska-Brzyska et al. 2022). Therefore, different models may have reflected gene flow occurring in different periods, and gene flow from the island during the pre-LGM period may be possible as the "constant migration" model inferred. Second, Aomori of Honshu Island was inferred to have re-established recently in the wave of continental population expansion, suggesting a possibility that the island population has been also sustained by migrants from the continent. In fact, Honshu might have been occupied previously by the present Hokkaido ("island") population during a glacial period, supported by 1) a significantly larger historical N_e than its contemporary N_e, 2) continuous suitable habitats across Honshu and Hokkaido inferred by the SDM, and 3) a high affinity of Kanto samples towards Hokkaido breeding birds, given that wintering regions of short-distance migratory species may reflect the past breeding range (Zink and Gardner 2017). Although our assumption of niche conservatism in the SDM may bias our inference of the recurrence of gene flow, it can only lead to underestimation but not overestimation of the Japanese Archipelago as glacial refugia, given that the present suitable habitats on the Japanese islands were predicted at a lesser extent than the actual range (i.e., Tomakomai and Aomori).

Therefore, we propose a possible scenario where the Swinhoe's Rail has survived multiple glacial periods by maintaining a metapopulation in which the Japanese Archipelago played a pivotal role. Because the island population has preserved a large glacial refugium while retaining high (and possibly ancient) genetic variation (Fig. 6a), post-glacial gene flow to the

continent (Fig. 6b) may have facilitated the continental expansion (Fig. 6c), which then resulted in the re-establishment or gene flow to the island, allowing the persistence of the island population for the next cycle (Fig. 6d). This scenario contradicts many previous studies that have found small genetic variation and N_e of island populations where migrants occurred only from their continental counterparts. However, they were biased to taxa that have undergone strong phenotypic or genomic divergence on islands (e.g., Wang et al. 2016; Van Doren et al. 2017; Delmore et al. 2020; Recuerda et al. 2021), where post-divergence gene flow from islands may be prevented. Our scenario focuses on the system where organisms form a metapopulation in the archipelagic or land-bridge island system (Seki et al. 2007; Mapel et al. 2021; Tsujimoto et al. 2023), where populations may have been dynamically maintained. In fact, some Japanese endemic lineages have extended their ranges to the far eastern edge of the Eurasian continent (e.g. *Agropsar philippensis*, Craig and Feare 2020; *Carduelis sinica*, Saitoh et al. 2020; *Ficedula narcissina*, Clement et al. 2021), and similar continent-island population systems are recognized in other regions (Bellemain et al. 2008; Garcia-R et al. 2017), suggesting the universality of such a system.

A recent review emphasizes the importance of long-term population persistence through a dynamic metapopulation for speciation and adaptive radiation to progress (Harvey et al. 2019). Our view also emphasizes the evolutionary consequences of metapopulation in continental populations, differently from the traditional perspective that focuses on the persistence of an overall metapopulation by itself (Hanski 1998). At this point, a metapopulation structure across the continent-island system can be perceived as a transition state for insular populations to undergo reverse colonization, thereby promoting long-term persistence and adaptation to the continental environment. Future studies need to model complex trajectories of both population

size and migration rates over time to support the recurrence of gene flow (e.g., Wang et al. 2020), which will disentangle the evolutionary impacts of a continent-island metapopulation to adaptation and diversification on the continent, extending our microevolutionary idea to a macroevolutionary framework.

Conclusion

Historical demography is fundamental to understanding a determinant of population genetic structure that is significantly linked with subsequent speciation and diversification (Harvey et al. 2019). This study reconstructed the demographic dynamics of a rail species in a continent-island system. We found concrete evidence of asymmetrical gene flow from the Japanese Archipelago to the continental population, which was possibly associated with its restoration of genetic variability and population expansion. Furthermore, we suggested that this species may have persisted as a metapopulation across both the islands and the continent. These findings were probably unexpected from the conventional continent-island view, especially since the limited breeding range of this species was only recently discovered in Japan (Senzaki et al. 2020). To fully understand the processes that have given rise to the present biodiversity, evolutionary studies on the current and apparent species distribution are insufficient. Reconstructing a long history of population dynamics of organisms in a spatiotemporally explicit way will shed light on the previously unfilled gap between micro- and macroevolutionary concepts and facilitate our understanding of global biodiversity patterns.

Tables

Table 1 Dataset and the number of sites used in each analysis

Analyses	Linkage	Site type	Outgroup	Populations	# Sites	# Samples	Notes
PCAngsd, ngsAdmixture	unlinked	V	Non	4 breeding[1], 1 wintering[2]	1,709	49	-
NeighborNet (distAngsd)	linked	V/I	Laja/Atro	4 breeding	162,050	32	GL was calculated by bcftools
RAxML, ABBA-BABA	linked	V/I	Laja	4 breeding	162,050	31	-
TreeMix	linked	V	Laja	4 breeding	2,299	31	Population-specific private alleles were removed by glactools
SNAPP	unlinked	V	Laja	4 breeding	927	16	The top three/four samples with the lowest missing allele proportions in each population were selected
Diversity indices, Fst, Fastsimcoal2, Stairway Plot 2	linked	V/I	Non	2 genetic[3]	210,121	30	-

Abbreviation: V, variants; V/I, both variants and invariants; Laja, *Laterallus jamaicensis*; Atro, *Atlantisia rogersi*; [1]Tomakomai, Kushiro, Russia, Aomori; [2]Kanto; [3]Island, Continental.

Figure legends

Figure 1 A map of sampling localities of Swinhoe's rails (a), and their population genetic structure inferred by (b) a principal component analysis (PCA) and (c) an ADMIXTURE analysis conducted using genotype likelihoods of 1,709 SNPs. (a, b) The colors of dots and inertia ellipses represent different regional populations. Dark-grayed regions are previously known breeding areas where present breeding evidence is lacking. (c) Genetic assignment of individuals to each number of genetic clusters (K = 2 to 5) is shown. Colors do not correspond to the sampling localities. The inset photo by Marc Bastardot.

Figure 2 Phylogenetic trees of the Swinhoe's Rail reconstructed using different methods and genetic markers. (a) A mitochondrial phylogenetic tree (mtDNA tree) using 772 bp of the *Cytb* gene. Posterior probabilities >0.9 were indicated by black dots on nodes. Numbers at the nodes indicate median divergence time estimates with 95% highest posterior density within brackets (bold) and posterior probabilities (italic). (b) A maximum likelihood (ML) tree (RAxML tree) was reconstructed using 162,050 sites including both variant and invariant sites. Only nodes with bootstrap values >70 are annotated with black circles. (c) A population-level tree reconstructed by TreeMix using the allele frequency of 2,299 polymorphic loci, excluding population-specific private alleles. Branch lengths indicate drift parameters, the strength of drift specific to each branch. The yellow arrow indicates the estimated gene flow. (d) Posterior population-level tree reconstructed by SNAPP analyses using 927 SNPs with no missing data. Values at each node (white) indicate divergence time estimates for each node with their credible intervals within brackets. Two different divergence time estimates were obtained for different calibration constraints.

Figure 3 Comparison of scaffold-level statistics between the island and continental populations, including (a) Watterson's theta and (b) Tajima's *D*. Significant differences in the values of each index were annotated with asterisks (* $p < 0.05$, ** $p < 0.01$).

Figure 4 Demographic inference of the three genetic populations of the Swinhoe's Rail based on site frequency spectra, using (a, b, c) fastsimcoal2 (FSC2) and (d) Stairway Plot 2. Asterisks indicate fixed values in each model. Effective population size and time are scaled by mutation rate (1.91e-09) and generation time (1 year). (a) Generalized schematic representation of the five models tested. Parameters with upper cases represent the effective population size we modeled. Arrows with different colors represent different migration matrices within the model. (b, c) The two best-fit FSC2 models with estimated parameters. Numbers by or on branches and arrows represent population sizes and the number of migrants/year, respectively. The size of each box is not scaled to the corresponding population size. Divergence time is shown in million years. (d) Changes in effective population size (in millions) over time (in million years) are shown for the island and continental populations. Bold lines and thin ribbons around them show median and 95% confidence intervals, respectively. The gray range indicates the approximate period of population divergence.

Figure 5 Species distribution models constructed using the present occurrence records of *Coturnicops noveboracensis* in North America were projected to East Asia in different time slices (a-g). On the top right panel, a proxy for paleoclimatic temperature (δ^{18}O) (Lisiecki and Raymo 2005; Spratt and Lisiecki 2016) was plotted, and periods that corresponded to each time slice (a-g) were indicated on it. Higher δ^{18}O values indicate higher temperature. Color gradients from indigo to yellow represent environmental suitability for rails. Black arrows in the projection to the present climate layer (a) indicate the presently or previously known distribution areas of

this species. Regions with gray shades and colors are where model extrapolation was restricted due to the absence of the climate in the modeling domain and regions with ice sheets or deserts, respectively.

Figure 6 Schematic representation of our recurrent continent-island metapopulation scenario of the Swinhoe's Rail; the archipelago acted as a genetic reservoir during a glacial period (a), and gene flow to the continent (b) facilitates continental population expansion (c), and re-establishment of island population guarantees the island population to persist until the next glacial period (d). The time series is shown in anticlockwise. Arrows on maps indicate movement of individuals as colonization or expansion (solid) and gene flow (hatched). Filled circles represent stable (solid) and dynamic (hatched) distribution ranges of each population.

Figures

Figure 1

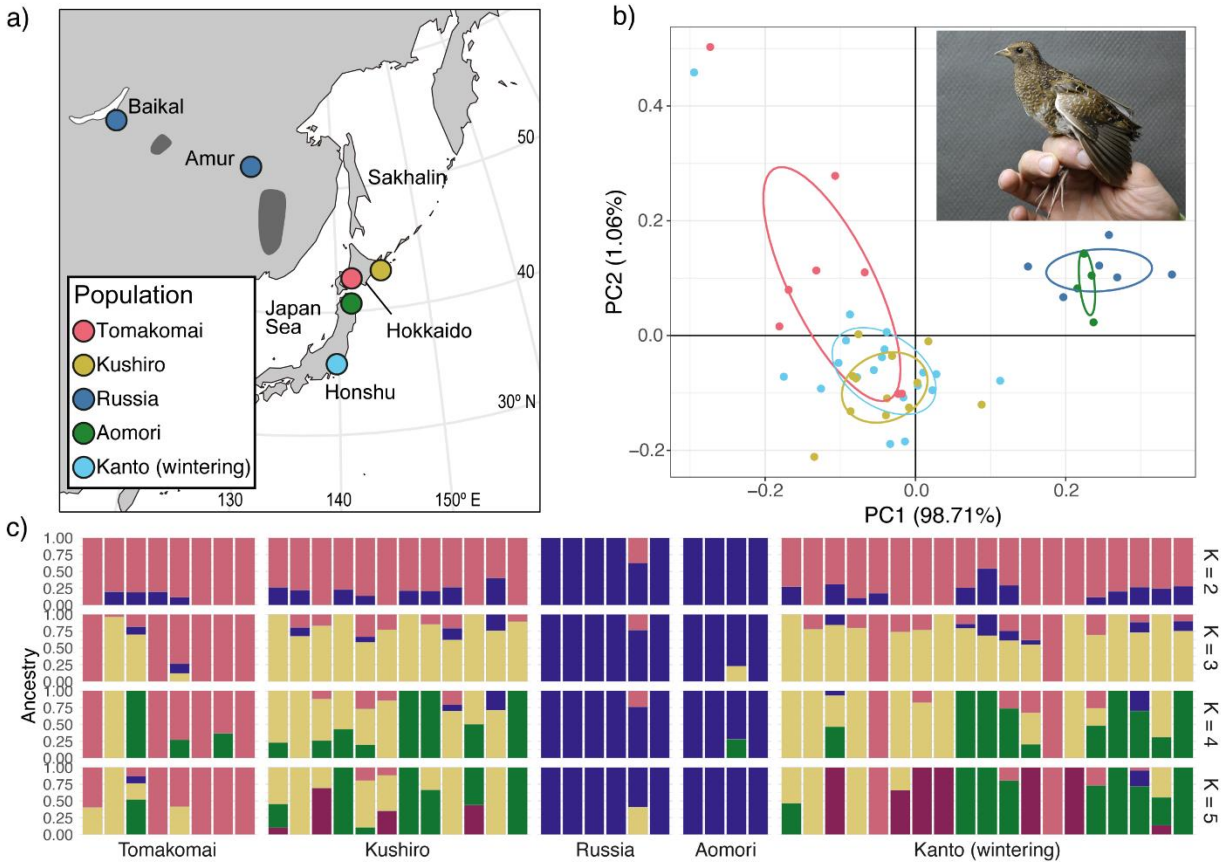


Figure 2

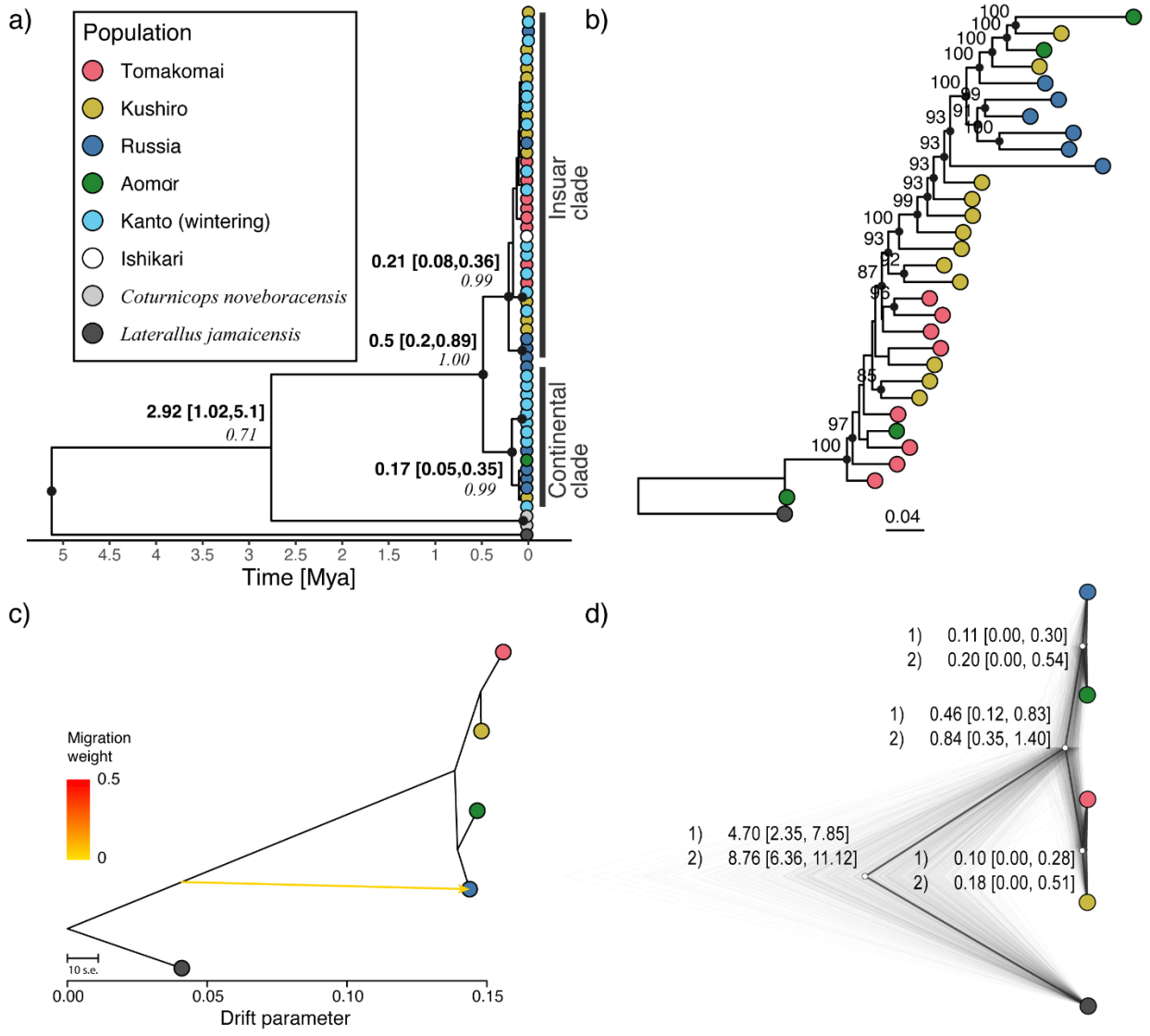

a)
Population
Tomakomai
Kushiro
Russia
Aomori
Kanto (wintering)
Ishikari
Coturnicops noveboracensis
Laterallus jamaicensis
0.21 [0.08,0.36]
0.99
0.5 [0.2,0.89]
1.00
2.92 [1.02,5.1]
0.71
0.17 [0.05,0.35]
0.99
Insular clade
Continental clade
5 4.5 4 3.5 3 2.5 2 1.5 1 0.5 0
Time [Mya]
b)
0.04
c)
Migration weight
0.5
0
10 s.e.
0.00 0.05 0.10 0.15
Drift parameter
d)
1) 0.11 [0.00, 0.30]
2) 0.20 [0.00, 0.54]
1) 0.46 [0.12, 0.83]
2) 0.84 [0.35, 1.40]
1) 4.70 [2.35, 7.85]
2) 8.76 [6.36, 11.12]
1) 0.10 [0.00, 0.28]
2) 0.18 [0.00, 0.51]

Figure 3

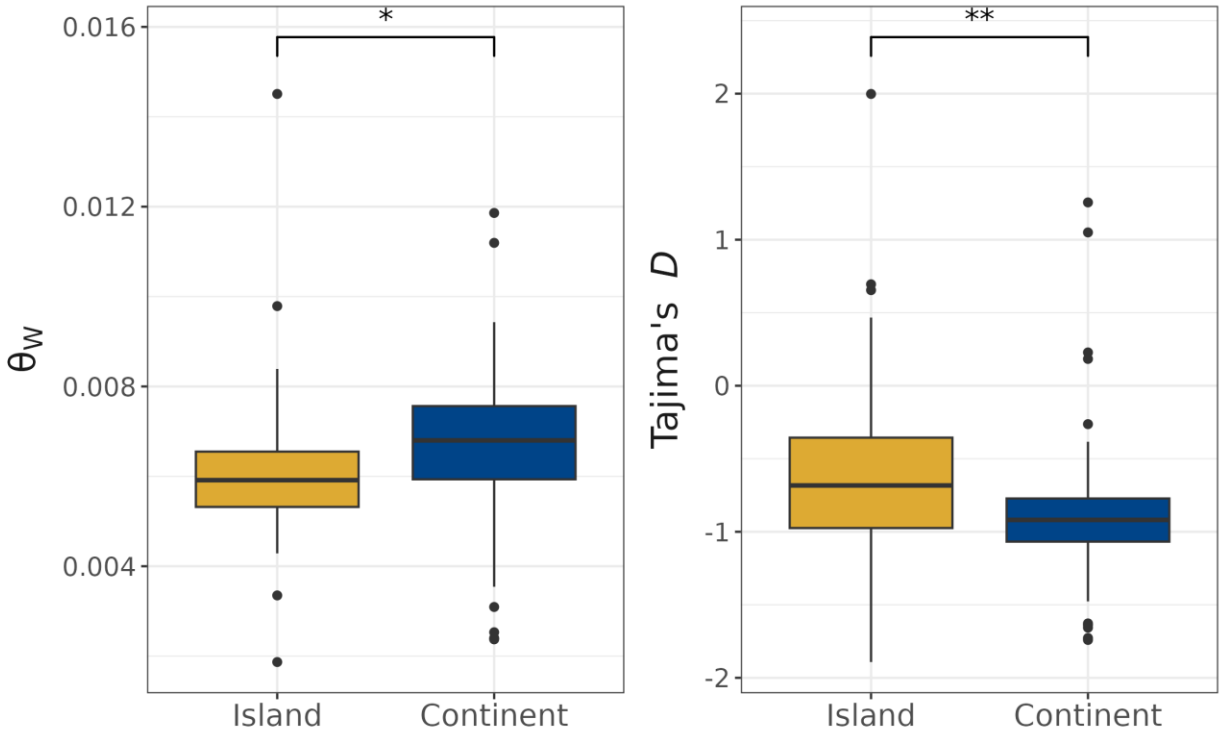

*
0.016
0.012
0.008
0.004
θW
Island
Continent
**
2
1
0
-1
-2
Tajima's D
Island
Continent

Figure 4

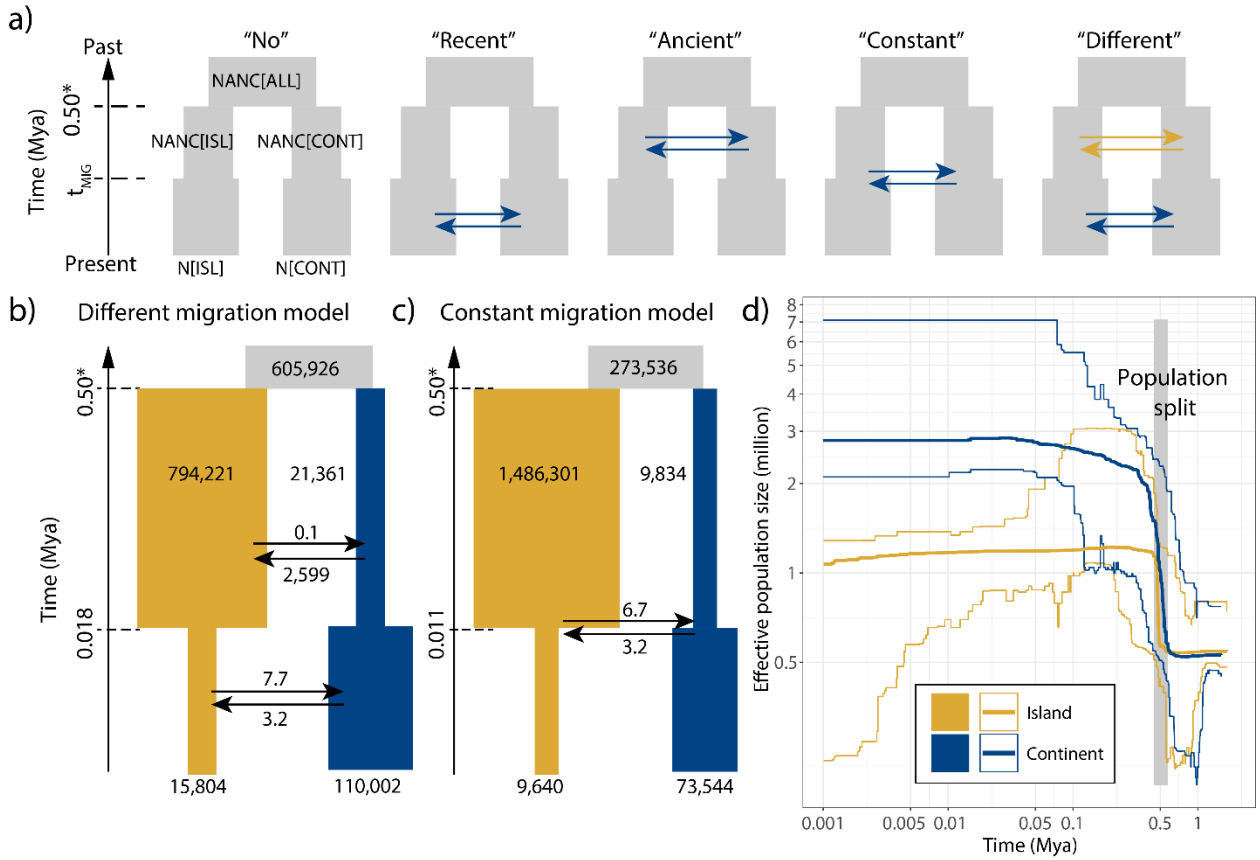

a)
Past
Present
Time (Mya)
0.50*
t_{MIG}
"No"
"Recent"
"Ancient"
"Constant"
"Different"
NANC[ALL]
NANC[ISL]
NANC[CONT]
N[ISL]
N[CONT]
b) Different migration model
605,926
794,221
21,361
0.1
2,599
7.7
3.2
15,804
110,002
c) Constant migration model
273,536
1,486,301
9,834
6.7
3.2
9,640
73,544
d)
Effective population size (million)
Population split
Island
Continent
Time (Mya)

Figure 5

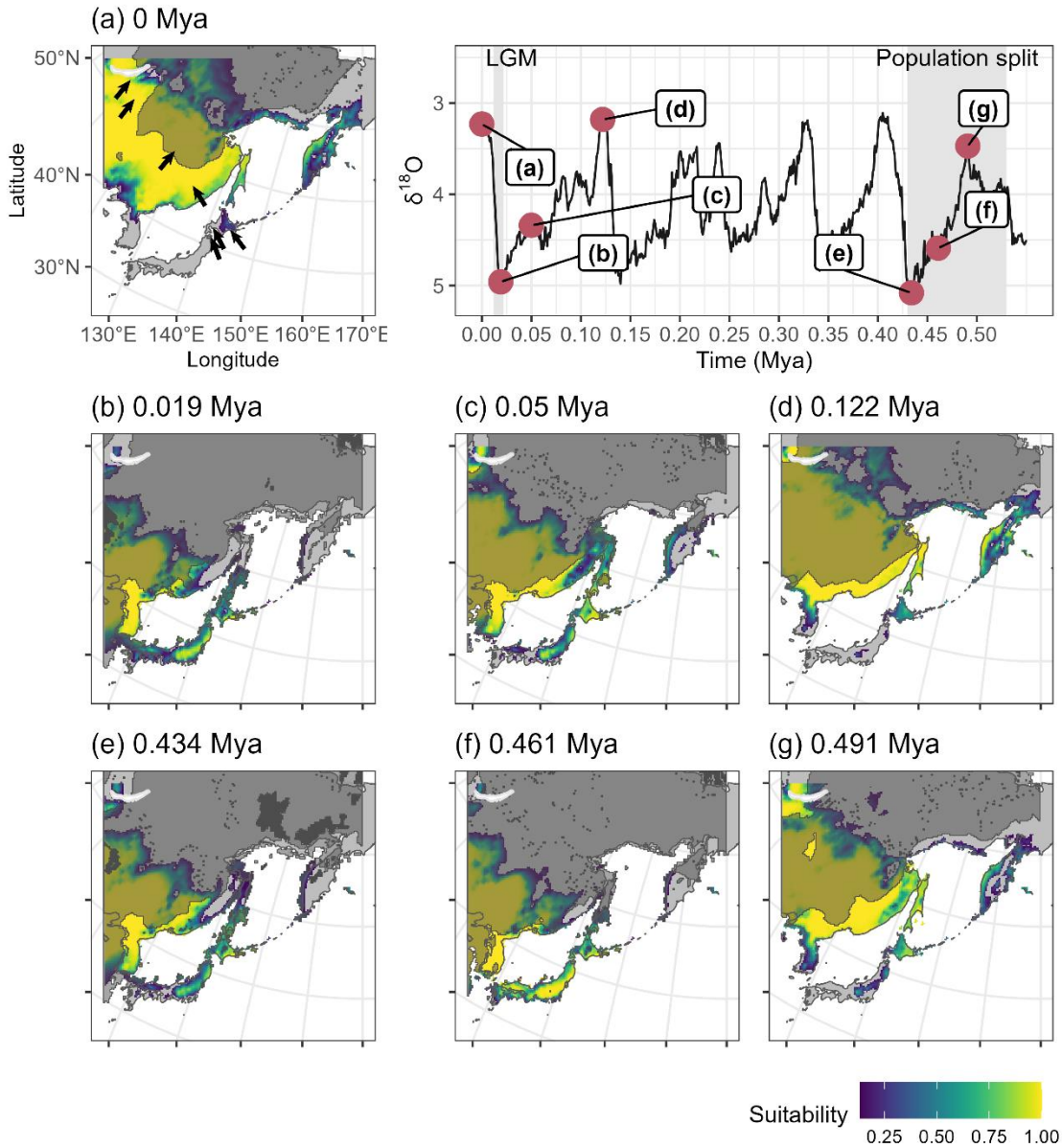
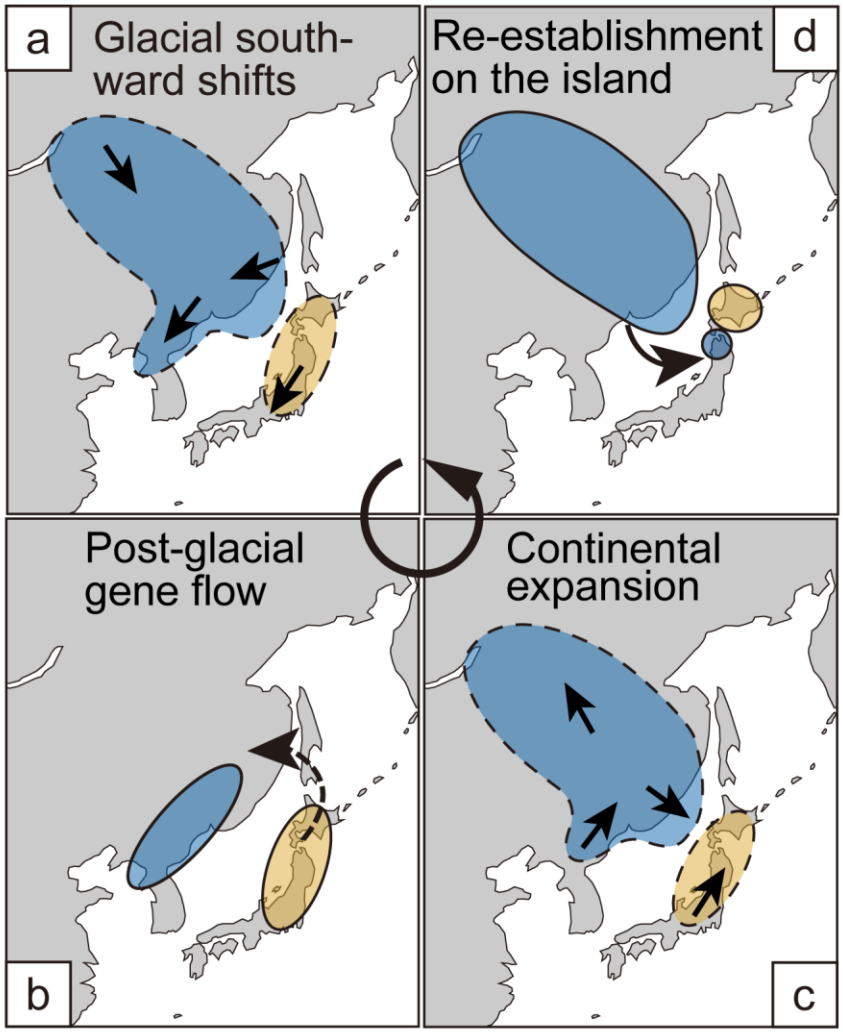

(a) 0 Mya
50°N
40°N
30°N
Latitude
130°E 140°E 150°E 160°E 170°E
Longitude
LGM
Population split
(a)
(b)
(c)
(d)
(e)
(f)
(g)
δ18O
3
4
5
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50
Time (Mya)
(b) 0.019 Mya
(c) 0.05 Mya
(d) 0.122 Mya
(e) 0.434 Mya
(f) 0.461 Mya
(g) 0.491 Mya
Suitability
0.25 0.50 0.75 1.00

Figure 6


a
Glacial south-
ward shifts
Re-establishment
on the island
d
Post-glacial
gene flow
Continental
expansion
b
c

References

Anisimova, V., Y. Anisimov, M. Bastardot, and W. Heim. 2019. First observation of Swinhoe's Rail *Coturnicops exquisitus* in the delta of Selenga river. Baikal Zoological Journal 2:112–113.

Aoki, D., G. Kinoshita, A. P. Kryukov, I. Nishiumi, S.-I. Lee, and H. Suzuki. 2018. Quaternary-related genetic differentiation and parallel population dynamics of the Eurasian Jay (*Garrulus glandarius*) in the circum-Japan Sea region. J. Ornithol. 159:1087–1097.

Bellemain, E., E. Bermingham, and R. E. Ricklefs. 2008. The dynamic evolutionary history of the bananaquit (*Coereba flaveola*) in the Caribbean revealed by a multigene analysis. BMC Evol. Biol. 8:240.

Bellemain, E., and R. E. Ricklefs. 2008. Are islands the end of the colonization road? Trends Ecol. Evol. 23:461–468.

Bickart, K. J., and R. M. Chandler. 1990. Recent advances in the study of Neogene fossil birds. Ornithol. Monogr. 44:1–161.

Bouckaert, R., T. G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina, J. Heled, G. Jones, D. Kühnert, N. De Maio, M. Matschiner, F. K. Mendes, N. F. Müller, H. A. Ogilvie, L. du Plessis, A. Popinga, A. Rambaut, D. Rasmussen, I. Siveroni, M. A. Suchard, C.-H. Wu, D. Xie, C. Zhang, T. Stadler, and A. J. Drummond. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 15:e1006650.

Bryant, D., R. Bouckaert, J. Felsenstein, N. A. Rosenberg, and A. RoyChoudhury. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29:1917–1932.

Bryant, D., and V. Moulton. 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol. Biol. Evol. 21:255–265.

Cerca, J., D. D. Cotoras, V. C. Bieker, R. De-Kayne, P. Vargas, M. Fernández-Mazuecos, J. López-Delgado, O. White, M. Stervander, A. J. Geneva, J. E. Guevara Andino, J. I. Meier, L. Roeble, B.

Brée, J. Patiño, J. M. Guayasamin, M. de L. Torres, H. Valdebenito, M. D. R. Castañeda, J. A. Chaves, P. J. Díaz, L. Valente, M. L. Knope, J. P. Price, L. H. Rieseberg, B. G. Baldwin, B. C. Emerson, G. Rivas-Torres, R. Gillespie, and M. D. Martin. 2023. Evolutionary genomics of oceanic island radiations. Trends Ecol. Evol. 38:631–642.

Clement, P., J. del Hoyo, D. Christie, N. Collar, G. M. Kirwan, and P. C. Rasmussen. 2021. Narcissus Flycatcher (*Ficedula narcissina*), version 1.1. *in* B. K. Keeney, ed. Birds of the World. Cornell Lab of Ornithology, NY.

Craig, A., and C. J. Feare. 2020. Chestnut-cheeked Starling (*Agropsar philippensis*) version 1.0. *in* J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie, and E. de Juana, eds. Birds of the World. Cornell Laboratory of Ornithology, NY.

Delmore, K., J. C. Illera, J. Pérez-Tris, G. Segelbacher, J. S. Lugo Ramos, G. Durieux, J. Ishigohoka, and M. Liedvogel. 2020. The evolutionary history and genomics of European blackcap migration. Elife 9:e54462.

Ellegren, H., and N. Galtier. 2016. Determinants of genetic diversity. Nat. Rev. Genet. 17:422–433.

Excoffier, L., N. Marchi, D. A. Marques, R. Matthey-Doret, A. Gouy, and V. C. Sousa. 2021. fastsimcoal2: demographic inference under complex evolutionary scenarios. Bioinformatics 37:4882–4885.

Feder, J. L., S. P. Egan, and P. Nosil. 2012. The genomics of speciation-with-gene-flow. Trends Genet. 28:342–350.

Fitak, R. R. 2021. OptM: estimating the optimal number of migration edges on population trees using Treemix. Biol Methods Protoc 6:bpab017.

Fitzpatrick, S. W., J. C. Gerberich, L. M. Angeloni, L. L. Bailey, E. D. Broder, J. Torres-Dowdall, C. A. Handelsman, A. López-Sepulcre, D. N. Reznick, C. K. Ghalambor, and W. Chris Funk. 2016. Gene flow from an adaptively divergent source causes rescue through genetic and demographic factors in two wild populations of Trinidadian guppies. Evol. Appl. 9:879–891.

Fox, E. A., A. E. Wright, M. Fumagalli, and F. G. Vieira. 2019. ngsLD: evaluating linkage disequilibrium

using genotype likelihoods. Bioinformatics 35:3855–3856.

Frankham, R. 1995. Effective population size/adult population size ratios in wildlife: a review. Genet. Res. 66:95–107.

Frankham, R. 2005. Genetics and extinction. Biol. Conserv. 126:131–140.

Garcia-R, J. C., L. Joseph, G. Adcock, J. Reid, and S. A. Trewick. 2017. Interisland gene flow among populations of the buff-banded rail (Aves: Rallidae) and its implications for insular endemism in Oceania. J. Avian Biol. 48:679–690.

Garcia-R, J. C., E. M. Lemmon, A. R. Lemmon, and N. French. 2020. Phylogenomic reconstruction sheds light on new relationships and timescale of rails (Aves: Rallidae) evolution. Diversity 12:70.

Garcia-R, J. C., and N. J. Matzke. 2021. Trait-dependent dispersal in rails (Aves: Rallidae): Historical biogeography of a cosmopolitan bird clade. Mol. Phylogenet. Evol. 159:107106.

Gattepaille, L. M., M. Jakobsson, and M. G. B. Blum. 2013. Inferring population size changes with sequence and SNP data: lessons from human bottlenecks. Heredity 110:409–419.

Hanski, I. 1998. Metapopulation dynamics. Nature 396:41–49.

Harvey, M. G., S. Singhal, and D. L. Rabosky. 2019. Beyond reproductive isolation: demographic controls on the speciation process. Annu. Rev. Ecol. Evol. Syst. 50:75–95.

Heim, W., D. Trense, A. Heim, and J. Kamp. 2019. Discovery of a new breeding population of the vulnerable Swinhoe's Rail *Coturnicops exquisitus* confirmed by genetic analysis. Bird Conserv. Int. 29:454–462.

Jønsson, K. A., and B. G. Holt. 2015. Islands contribute disproportionately high amounts of evolutionary diversity in passerine birds. Nat. Commun. 6:8538.

Kass, J. M., R. Muscarella, P. J. Galante, C. L. Bohl, G. E. Pinilla-Buitrago, R. A. Boria, M. Soley-Guardia, and R. P. Anderson. 2021. ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions. Methods Ecol. Evol. 12:1602–1608.

Kitazawa, M., Y. Yamaura, M. Senzaki, M. Hanioka, H. Ohashi, M. Oguro, T. Matsui, and F. Nakamura. 2022. Quantifying the impacts of 166 years of land cover change on lowland bird communities.

Proc. R. Soc. B. 289:20220338.

Korneliussen, T. S., A. Albrechtsen, and R. Nielsen. 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics 15:356.

Kozlov, A. M., D. Darriba, T. Flouri, B. Morel, and A. Stamatakis. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35:4453–4455.

Krapp, M., R. M. Beyer, S. L. Edmundson, P. J. Valdes, and A. Manica. 2021. A statistics-based reconstruction of high-resolution global terrestrial climate for the last 800,000 years. Sci Data 8:228.

Leonardi, M., E. Y. Hallett, R. Beyer, M. Krapp, and A. Manica. 2023. pastclim 1.2: an R package to easily access and use paleoclimatic reconstructions. Ecography , doi: 10.1111/ecog.06481.

Leroy, T., M. Rousselle, M.-K. Tilak, A. E. Caizergues, C. Scornavacca, M. Recuerda, J. Fuchs, J. C. Illera, D. H. De Swardt, G. Blanco, C. Thébaud, B. Milá, and B. Nabholz. 2021. Island songbirds as windows into evolution in small populations. Curr. Biol. 31:1303–1310.e4.

Lisiecki, L. E., and M. E. Raymo. 2005. A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records. Paleoceanography 20:PA1003.

Liu, X., and Y.-X. Fu. 2020. Stairway Plot 2: demographic history inference with folded SNP frequency spectra. Genome Biol. 21:280.

Lomolino, M. V., B. R. Riddle, and R. J. Whittaker. 2017. Biogeography: Biological Diversity across Space and Time. 5th Edition. Sinauer Associates, Maryland.

Mapel, X. M., E. F. Gyllenhaal, T. H. Modak, L. H. DeCicco, A. Naikatini, R. B. Utzurrum, J. O. Seamon, A. Cibois, J.-C. Thibault, M. D. Sorenson, R. G. Moyle, L. N. Barrow, and M. J. Andersen. 2021. Inter- and intra-archipelago dynamics of population structure and gene flow in a Polynesian bird. Mol. Phylogenet. Evol. 156:107034.

Meisner, J., and A. Albrechtsen. 2018. Inferring population structure and admixture proportions in low-depth NGS data. Genetics 210:719–731.

Miya, A., S. Sannohe, J. Ebina, and H. Sekishita. 2005. A survey report on Yellow Rail (*Coturnicops*

noveboracensis exquisitus) observed during the breeding season at Hotokenuma Reclaimed Land in Misawa, Aomori Prefecture. Jap. J. Ornithol. 54:104–107.

Nadachowska-Brzyska, K., M. Konczal, and W. Babik. 2022. Navigating the temporal continuum of effective population size. Methods Ecol. Evol. 13:22–41.

Nam, K., C. Mugal, B. Nabholz, H. Schielzeth, J. B. W. Wolf, N. Backström, A. Künstner, C. N. Balakrishnan, A. Heger, C. P. Ponting, D. F. Clayton, and H. Ellegren. 2010. Molecular evolution of genes in avian genomes. Genome Biol. 11:R68.

O'Grady, P., and R. Desalle. 2008. Out of Hawaii: the origin and biogeography of the genus *Scaptomyza* (Diptera: Drosophilidae). Biol. Lett. 4:195–199.

Patiño, J., M. Carine, P. Mardulyn, N. Devos, R. G. Mateo, J. M. González-Mancebo, A. J. Shaw, and A. Vanderpoorten. 2015. Approximate Bayesian computation reveals the crucial role of oceanic islands for the assembly of continental biodiversity. Syst. Biol. 64:579–589.

Patton, A. H., L. J. Harmon, M. Del Rosario Castañeda, H. K. Frank, C. M. Donihue, A. Herrel, and J. B. Losos. 2021. When adaptive radiations collide: Different evolutionary trajectories between and within island and mainland lizard clades. Proc. Natl. Acad. Sci. 118:e2024451118.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecol. Modell. 190:231–259.

Pickrell, J. K., and J. K. Pritchard. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8:e1002967.

Pujolar, J. M., M. P. K. Blom, A. H. Reeve, J. D. Kennedy, P. Z. Marki, T. S. Korneliussen, B. G. Freeman, K. Sam, E. Linck, T. Haryoko, B. Iova, B. Koane, G. Maiah, L. Paul, M. Irestedt, and K. A. Jønsson. 2022. The formation of avian montane diversity across barriers and along elevational gradients. Nat. Commun. 13:268.

Recuerda, M., J. C. Illera, G. Blanco, R. Zardoya, and B. Milá. 2021. Sequential colonization of oceanic archipelagos led to a species-level radiation in the common chaffinch complex (Aves: *Fringilla coelebs*). Mol. Phylogenet. Evol. 164:107291.

Renaud, G. 2017. glactools: a command-line toolset for the management of genotype likelihoods and allele counts. Bioinformatics 34:1398–1400.

Saitoh, T., K. Kawakami, Y. A. Red'kin, I. Nishiumi, C.-H. Kim, and A. P. Kryukov. 2020. Cryptic speciation of the Oriental Greenfinch *Chloris sinica* on oceanic islands. Zool. Sci. 37:280–294.

Seki, S.-I., H. Takano, K. Kawakami, N. Kotaka, A. Endo, and K. Takehara. 2007. Distribution and genetic structure of the Japanese wood pigeon (*Columba janthina*) endemic to the islands of East Asia. Conserv. Genet. 8:1109–1121.

Senzaki, M., M. Kitazawa, T. Sadakuni, and M. Takahashi. 2020. Breeding evidence of the vulnerable Swinhoe's Rail (*Coturnicops exquisitus*) in Japan. Wilson J. Ornithol. 132:711–717.

Skotte, L., T. S. Korneliussen, and A. Albrechtsen. 2013. Estimating individual admixture proportions from next generation sequencing data. Genetics 195:693–702.

Soraggi, S., C. Wiuf, and A. Albrechtsen. 2018. Powerful inference with the D-statistic on low-coverage whole-genome data. G3 8:551–566.

Spratt, R. M., and L. E. Lisiecki. 2016. A Late Pleistocene sea level stack. Clim. Past 12:1079–1092.

Stange, M., M. R. Sánchez-Villagra, W. Salzburger, and M. Matschiner. 2018. Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports Miocene closure of the Panamanian Isthmus. Syst. Biol. 67:681–699.

Strimas-Mackey, M., E. Miller, and W. M. Hochachka. 2018. auk: eBird Data Extraction and Processing with AWK. R package version 0.3.0.

Suyama, Y., and Y. Matsuki. 2015. MIG-seq: an effective PCR-based method for genome-wide single-nucleotide polymorphism genotyping using the next-generation sequencing platform. Sci. Rep. 5:16963.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460.

Taylor, B., and C. J. Sharpe. 2020. Swinhoe's Rail (*Coturnicops exquisitus*), version 1.0. *in* J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie, and E. de Juana, eds. Birds of the World. Cornell Lab of

Ornithology, Ithaca, NY.

Tsujimoto, D., H. Ando, H. Suzuki, K. Horikoshi, N. Nakajima, Y. Suyama, A. Matsuo, C. Fujii, and Y. Isagi. 2023. Has long-distance flight ability been maintained by pigeons in highly insular habitats? J. Biogeogr. 50:235–246.

Twyford, A. D., and R. A. Ennos. 2012. Next-generation hybridization and introgression. Heredity 108:179–189.

Vamosi, J. C., W. S. Armbruster, and S. S. Renner. 2014. Evolutionary ecology of specialization: insights from phylogenetic analysis. Proc. R. Soc. B 281:20142004.

Van Doren, B. M., L. Campagna, B. Helm, J. C. Illera, I. J. Lovette, and M. Liedvogel. 2017. Correlated patterns of genetic diversity and differentiation across an avian family. Mol. Ecol. 26:3982–3997.

Wang, K., I. Mathieson, J. O'Connell, and S. Schiffels. 2020. Tracking human population structure through time from whole genome sequences. PLoS Genet. 16:e1008552.

Wang, N., B. Liang, J. Wang, C.-F. Yeh, Y. Liu, Y. Liu, W. Liang, C.-T. Yao, and S.-H. Li. 2016. Incipient speciation with gene flow on a continental island: Species delimitation of the Hainan Hwamei (*Leucodioptron canorum owstoni*, Passeriformes, Aves). Mol. Phylogenet. Evol. 102:62–73.

Warren, B. H., D. Simberloff, R. E. Ricklefs, R. Aguilée, F. L. Condamine, D. Gravel, H. Morlon, N. Mouquet, J. Rosindell, J. Casquet, E. Conti, J. Cornuault, J. M. Fernández-Palacios, T. Hengl, S. J. Norder, K. F. Rijsdijk, I. Sanmartín, D. Strasberg, K. A. Triantis, L. M. Valente, R. J. Whittaker, R. G. Gillespie, B. C. Emerson, and C. Thébaud. 2015. Islands as model systems in ecology and evolution: prospects fifty years after MacArthur-Wilson. Ecol. Lett. 18:200–217.

Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. 7:256–276.

Weir, J. T., and D. Schluter. 2008. Calibrating the avian molecular clock. Mol. Ecol. 17:2321–2328.

Whittaker, R. J., and J. M. Fernández-Palacios. 2007. Island Biogeography: Ecology, Evolution, and Conservation. Oxford University Press, Oxford.

Whittaker, R. J., J. M. Fernández-Palacios, T. J. Matthews, M. K. Borregaard, and K. A. Triantis. 2017. Island biogeography: Taking the long view of nature's laboratories. Science 357:eaam8326.

Yamaguchi, R. 2022. Intermediate dispersal hypothesis of species diversity: New insights. Ecol. Res. 37:301–315.

Yeung, C. K. L., P.-W. Tsai, R. T. Chesser, R.-C. Lin, C.-T. Yao, X.-H. Tian, and S.-H. Li. 2011. Testing founder effect speciation: divergence population genetics of the spoonbills *Platalea regia* and *Pl. minor* (Threskiornithidae, Aves). Mol. Biol. Evol. 28:473–482.

Zhang, D., F. E. Rheindt, H. She, Y. Cheng, G. Song, C. Jia, Y. Qu, P. Alström, and F. Lei. 2021. Most genomic loci misrepresent the phylogeny of an avian radiation because of ancient gene flow. Syst. Biol. 70:961–975.

Zhao, L., R. Nielsen, and T. S. Korneliussen. 2022. distAngsd: Fast and accurate inference of genetic distances for next-generation sequencing data. Mol. Biol. Evol. 39:msac119.

Zink, R. M., and A. S. Gardner. 2017. Glaciation as a migratory switch. Sci Adv 3:e1603133.