Title Page

Title

Phylogenomics reveals a gene flow history of reverse colonization

Authors

Aoki D[1,*], Senzaki M[2], Ando H[3], Odaya Y[4], Heim W[5,6], Kitazawa M[3,7], Wulf T[8], Trense D[9], Bastardot M[10], Fukuda A[11], Takahashi M[12,13], Kondo N[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

OCRID

Daisuke Aoki: 0000-0002-5780-8487

Masayuki Senzaki: 0000-0002-3843-5793

Haruko Ando: 0000-0001-7827-3883

Yoshiya Odaya: 0009-0004-1057-4005

Wieland Heim: 0000-0002-3262-2491

Munehiro Kitazawa: 0000-0003-0167-4464

Daronja Trense: 0000-0003-3598-8699

Masao Takahashi: 0000-0001-8385-6098

Natsuko Kondo: 0000-0002-0883-0506

Abstract

Islands have long been viewed as biodiversity sinks based on a microevolutionary assumption that island populations are small and genetically depauperate, having difficulties in re-colonizing continents. This conventional view is, however, currently challenged by “reverse colonization,” a macroevolutionary phenomenon where a continental lineage originates from island ancestors. This contradiction, therefore, raises an important question about how and what demographic histories of island populations heighten genetic variation and thereby allow successful reverse colonization. Here we address this question by

focusing on a rail species that likely experienced reverse colonization from Japan to the Eurasian continent, despite the expectation that island rails are “evolutionary dead-ends.” Our integrative phylogenomics supported the continental population originating from an island lineage, confirming reverse colonization. We unveiled a complex history of both intraspecific gene flow and introgression associated with reverse colonization. Particularly, we suggested that introgression may have played a pivotal role in restoring genetic variation post-reverse colonization, potentially facilitating continental population expansion. In theory, introgression assists adaptive radiation and diversification by rapidly introducing novel genetic variation. Thus, our findings shed light on how islands contribute to global biodiversity, thereby bridging a gap between micro- and macroevolution.

Main Text

Introduction

Island biological diversity contributes disproportionately to global biodiversity [1]. Islands, comprising 3.5% of the terrestrial land, contribute nearly 20% of global species richness and contribute more than 60% of extinct terrestrial species [2]. It has been assumed that the small and isolated landmass and the unique environment of the island cause small population size with low genetic variation [3] and specialization of genomes [e.g. 4], likely leading to high risks of extinction [5] and the "evolutionary dead-end" [6]. Based on these microevolutionary assumptions, islands have been viewed as sinks where lineages end diversification [7].

Recently, however, "reverse colonization," through which the island becomes the source of continental colonization and radiation, has become increasingly evident [2,7]. It has been reported from a variety of taxa, including birds, reptiles, bats, insects, and plants [7,8] across various kinds of islands, including Hawaii [9] and Japan [10]. Reverse colonization can be further followed by mainland diversification. Examples include the radiation of passerine species, representing almost 60% of the modern bird species richness [11], and Anolis lizard radiation in North America [12]. Previously, inference of reverse colonization has been done phenomenologically by detecting a macroevolutionary pattern in a phylogenetic tree where a continental lineage is derived from an island lineage [7]. However, the prevalence of reverse colonization suggests that there are overlooked microevolutionary processes by which insular organisms succeed in colonization and persist on the continents by overcoming aforementioned population genetics obstacles [7]. Therefore, reverse colonization needs to be assessed from microevolutionary processes to understand how islands potentially become a source of biodiversity.

Demographic history is one promising microevolutionary aspect that can help us understand how an island population overcomes its potentially low genetic variation. High genetic variation is important

for persistence and adaptation to novel environments, such as more competitive environments [13,14], like continents in comparison with islands. Given the several indications that insular populations may not necessarily be genetically depauperate [e.g. 8], there should be important demographic conditions under which reverse colonization is successful with retaining high genetic variation. For instance, island populations can maintain high genetic variation by avoiding severe population size decline due to unexpectedly large founder population size [15]. Alternatively, insular populations might have gained genetic variation through gene flow from other populations or other species [16]. These past demographic events linked with genetic variation are crucial to reflect its microevolutionary lesson in macroevolution because they critically determine whether speciation and diversification can initiate after population establishment [17].

Phylogeography can cover both macro- and microevolutionary arenas [18], and its application with high-throughput sequencing (HTS) techniques and a species distribution model (SDM) will provide both evidence for reverse colonization and its associated demography. HTS allows resolving phylogenetic trees while simultaneously considering gene flow and genetic drift [19,20]. HTS allows further reconstruction of demography beyond the phylogenetic tree, including population genetics structure [21] and historical changes in population size [22]. SDMs allow modeling range limits and shapes at particular periods, and therefore, their integration allows interpretation of the demographic parameters in spatially explicit ways [17], which helps associate demographic processes with reverse colonization.

The objective of this study is therefore to provide evidence of reverse colonization at the population level and reveal its historical demography in relation to genetic variation by applying phylogeography. To achieve this goal, we selected the Swinhoe's Rail (*Coturnicops exquisitus*), the world's smallest Asian rail species, as the ideal model case, because candidate phylogeographic scenarios are easily narrowed down, given that it is only the Eurasian descendent of the New World rail species group ("*Laterallus* group") [23]. Upon colonization from North America around the Pliocene to the Early Pleistocene [23], we expect that colonization by the Swinhoe's Rail occurred either 1) along the Japanese

Archipelago through the Kamchatka Peninsula, followed by continental colonization (i.e., reverse colonization scenario, a dotted line in Fig. 1a), or 2) along continental North Asia to East Asia, followed by colonization of the Japanese Archipelago (i.e., island colonization scenario, a hatched line in Fig. 1a). A reverse colonization scenario is likely in this species because a warm and wet environment that was presumably suitable for rails extended more along the islands than continental North Asia around this period [24]. This is contrasted by many rail species being the representatives of "end-of-colonization" on islands for extensive specialization and small population size [23,25], making this study further challenging to reinterpret islands as sources of biodiversity.

In this study, we first ask whether a Japanese population is a basal and a continental population nests within Japanese populations in a phylogenetic tree, supporting the reverse colonization scenario, by reconstructing phylogenetic relationships among populations through application of an HTS technique as well as standard mitochondrial analyses. Then, by integrating HTS and SDM, we compare genetic variations of each population and examine underlying historical demography, such as the presence, levels, and timings of intraspecific gene flow, introgression, and changes in population size. Finally, we evaluate a key demographic process associated with genetic variation of the continental population. We predict that the population under reverse colonization has retained genetic variation by maintaining an unexpectedly large population size or through gene flow from other populations or species.

Materials and methods

Details about the methods and datasets are described in the Supplementary Methods and scripts in the data depository [26].

Field sampling

The Swinhoe's Rail is poorly studied, and almost all of the presently recognized breeding areas were re-discovered recently, including the Amur region [27] and Lake Baikal [28] in Russia and some regions of

Japan, namely Aomori [29] and Hokkaido [30] (Fig. 1a). Despite its cryptic lifestyle inhabiting inaccessible wetlands, we developed our original sampling strategies [27,30] 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 (Fig. 1a, Table S1). We also captured wintering or migratory birds, including the Kanto region of Honshu Island (n = 37) and 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 [31], for 52 samples (Table S1). The latter is a type of genotyping by sequencing with conducting polymerase chain reactions at inter-simple sequence repeats. In exchange for smaller genome coverage, it is effective for low-quality DNA samples, best adapted to some of our dried feather or swab samples. For Sanger sequencing, we used the standard PCR and BigDye terminator protocols to sequence the *Cytb* gene. For MIG-seq, we used eight forward and eight reverse primers originally described by Suyama & Matsuki [31] to prepare a library mixture containing all the samples. The final mixture was applied to MiSeq Reagent Kit v3 150 cycles (Illumina) to obtain 150 bp paired-end reads. All the nucleotide sequences and raw reads obtained were deposited to DDBJ (Table S1).

The raw FASTQ reads for each sample were trimmed and filtered using Trimmomatic v. 0.39 [32]. Both paired and unpaired trimmed reads were aligned against a reference genome of a closely related species in the *Laterallus* group, the Black Rail *Laterallus jamaicensis coturniculus* (Table S1), using the

BWA-MEM algorithm of Burrows-Wheeler Aligner [33]. After the removal of individuals based on the levels of missingness, 49 samples were retained.

We mainly adopted probabilistic analysis frameworks to account for uncertainties about the true genotype, by calculating genotype likelihoods using the software ANGSD [34]. Genotypes were hard-called using the '-doGeno' function of ANGSD for software not accepting genotype likelihoods as inputs (RAxML-ng and TreeMix). Genotype likelihoods were repeatedly calculated using the 'SAMtools' model for different subsets of the dataset. They included the entire dataset without population categorization (DATASET 1), the entire dataset with five populations (Russia, Kushiro, Tomakomai, Aomori, and Kanto) (DATASET 2), and the dataset with three genetic breeding populations, namely Kushiro, Tomakomai, and continental (Russia + Aomori), which was delineated based on our genetic structure analyses (principal component analysis (PCA) and ADMIXTURE at K = 3 showing clear genetic distinctiveness; Fig. 1b–c) (DATASET 3). Linkage disequilibrium among variant sites was calculated and pruned using ngsLD [35] for analyses that do not recommend linked sites. Genotype likelihoods were similarly calculated for outgroup sequences downloaded through the Sequence Read Archive, which included *Laterallus jamaicensis* (reference sample), *Atlantisia rogersi*, and *Zapornia atra* (Table S1). For demographic analyses, we used unfolded site frequency spectra (SFS) and two-dimensional SFS (2d-SFS) using 167,120 sites (including both polymorphic and monomorphic sites) for each population or population pair defined by the DATASET 3, calculated by the '-doSAF' option of ANGSD and the 'realSFS' function.

Genetic structure and genetic diversity indices

We examined global genetic structure using the DATASET 1 by PCA using PCAngsd [36] and ADMIXTURE analysis using NGSadmix [21] with K ranging from 1 to 6. Using the DATASET 3, we calculated two genetic diversity indices, Watterson's theta θ_W [37] and nucleotide diversity θ_π [38], and two neutrality test statistics, Tajima's *D* [39] and Fu's *Fs* [40], per reference scaffold and population by using the 'do_stat' option of the 'thetaStat' function with the SFS as inputs. We also calculated pairwise

weighted and unweighted Fst on unlinked SFS for each population pair using the 'fst' function implemented in 'realSFS'. Values below zero for the neutrality test statistics indicate population bottlenecks followed by demographic expansions. Diversity indices and neutrality statistics were compared among populations, using linear mixed models. To set populations as covariates, we constructed two independent models with either Tomakomai or Kushiro set as the base population of a dummy variable.

Inference of phylogenetic relationship

Phylogenetic relationships reconstructed using different datasets and methods differ from each other because they may reflect different gene genealogy of complex population histories, such as intraspecific gene flow and introgression [41]. Therefore, our phylogenetic reconstructions consist of four analyses, a mitochondrial phylogenetic tree, a maximum likelihood (ML) tree using SNP data (ML SNP tree), a population-level tree, and demographic modeling. Then, a population history that can reasonably explain all the results was comprehensively interpreted.

Mitochondrial phylogenetic tree: The time-calibrated Bayesian inference method was applied to an alignment of 772 bps of partial *Cytb* by using BEAST v.2.7.3 [42]. Samples included 12 haplotypes from 54 Swinhoe's rails and three outgroups (two yellow rails and one black rail). The standard molecular clock rate of 1.05%/lineage/Myr [43] was applied to estimate divergence times.

ML SNP tree: RaxML-ng [44] with the 'GTGTR4+G+ASC_LEWIS' model was applied to the unlinked SNPs acquired using the DATASET 2. A black rail was set as an outgroup and standard non-parametric bootstrapping with 100 replicates was simultaneously conducted.

Population-level tree: TreeMix [20] was used to infer a population-level tree that simultaneously considered gene flow and genetic drift. A model with a migration edge = 0 was initially estimated, which was used to compare model performance of models with migration edges ranging from one to five. Ten replicates of runs for each number of migration edges were conducted. Changes in log-likelihoods, total

variance in population relatedness, and residual fits were evaluated to determine the best number of migration edges [20,45]. Consistency was also checked among different replicates at the best number of migration edges. Among the replicates with the best number of edges, a run with the largest model likelihood was selected as the best-fit model.

Demographic modeling: A population history was modeled using fastsimcoal2 v. 2.7.0.9 [46], which can estimate gene flow and population size changes while simultaneously inferring the population branching history although outgroup sequences from the database cannot be included in the models. Using the unlinked 2d-SFS, six independent models were tested. In each model, the three populations (Tomakomai, Kushiro, and the continental) were assigned to POP1 to POP3 differently giving three different branching histories, with and without recent population size changes at t_{EXP} (Fig. 3a). In each model, we considered only bidirectional gene flow between adjacent populations (between Kushiro and Tomakomai, and Tomakomai and the continental) at t_0. We assumed a mutation rate of 1.91e-09 [47]. We applied 'bounded reference' to constrain the first split time (t_{DIV2}) between 100,000 and 1,000,000, whose upper bound was set according to the divergence time estimated for the two statistically supported mitochondrial clades (see results). The best model was selected out of the six models tested based on the Akaike Information Criterion (AIC).

Inference of demographic trajectories

Stairway Plot 2 (SWP2) can estimate demographic trajectories through time, independently of user-defined models [22]. Using the one-dimensional unfolded SFS, we separately estimated demographic changes for each of the three populations. The same mutation rate was used as the fastsimcoal2 analyses and analyses were run with default settings. In the SWP2 analysis, the timing that populations started to show different demography from one another could be interpreted as population divergence, which can further support population history.

We compared the effective population size (N_e) estimates with the contemporary population size to understand the spatial extent of the populations by converting the census population size (N_c) [48] to the effective population size of a plain in Hokkaido (including Tomakomai) by using a mean ratio of N_c:N_e = 1:0.44 for birds [49].

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, the Yellow Rail distributed in North America, and we projected the model to the present and past climatic environment conditions of North America and East Asia. The model projection was intended to estimate the breeding range of both the Swinhoe's Rail, the Yellow Rail, and their common ancestor. This benefits from utilizing a large dataset from eBird (https://ebird.org/) from North American countries. Meanwhile, this is strongly based on an assumption of niche conservatism; we assume that these two sister species retain the environment niche of their common ancestor.

The SDM was applied to the eBird dataset of the Yellow Rail using MaxEnt [50]. We generated a "zero-filled" dataset following the instruction of an R package auk [51], and we used presence and pseudo-absence points of this species per observation event and site as occurrence and background inputs, respectively. We used an R package pastclim v. 1.2 [52] to extract time slices of 22 variables [53]. We selected one variable per group of variables correlated >0.7, resulting in seven covariates for modeling. We tested 12 models with different parameter settings by using the function 'ENMevaluate' implemented in an R package ENMeval v. 2.0.4 [54], and the best model was selected by the lowest delta AIC for small sample size (AICc), which was then projected to seven different time slices of environmental variables, including three interglacial (0 (present), 0.424, and 0.790 million years ago (Mya)) and three glacial (0.029 (Last Glacial Maximum, LGM), 0.478, and 0.761 Mya) periods. The two oldest periods represented the initial population split inferred by the mitochondrial phylogenetic tree and the SWP2

analyses (split1 in Fig. 3), and the two intermediate periods roughly corresponded to the oldest estimate for the population expansion of the continental population (split2 in Fig. 3).

Results

Sequencing and reference mapping

MIG-seq resulted in paired-end reads of 4,834,210 overall and 92,966 per-sample average. Reads after quality filtering were mapped on the reference genome, giving 3,724,169 paired-end reads successfully mapped.

Phylogenetic relationship and gene flow

The multiple phylogenetic trees based on different methods/datasets exhibited different populations as the basal population of the Swinhoe's Rail, suggesting complex histories of intra- and interspecific gene flow. The Swinhoe's Rail diverged from the Yellow Rail around the Late Pliocene to the Early Pleistocene (median = 2.92 Mya), and two clades diverged within species around the Middle Pleistocene (the Chibanian epoch) (median = 0.5 Mya) but showing no apparent genetic structure between them, inferred by mitochondrial phylogenetic analyses (Fig. 2a, Fig. S1). Meanwhile, the Tomakomai population was inferred as basal in both the population-level tree by TreeMix (Fig. 2c) and demographic modeling by fastsimcoal2 (Fig. 3b), and the divergence of the Tomakomai population (split1) corresponded to the divergence time of the mitochondrial clades (0.981 Mya; Fig. 3b), suggesting that divergence inferred by the mitochondrial analyses possibly correspond to this split.

Multiple gene flow events were suggested by different analyses. Of the two mitochondrial clades, the Tomakomai samples were contained exclusively in one clade while the others were included in both clades (Fig. 2a, Fig. S1), suggesting post-divergence intraspecific gene flow from the ancestor of the Tomakomai population to the common ancestor of the other two populations. This is contradicted by

contemporary migrants being biased from both the Kushiro and continental populations to the Tomakomai population suggested by the best fastsimcoal2 model (AIC = 14019.1, Akaike weight = 0.515; Fig. 3a–b, Table S2) and ADMIXTURE analysis (red clusters in Tomakomai samples at K = 3; Fig. 1c). Together, both historical but post-divergence gene flow from Tomakomai to the other populations (or their common ancestors) and contemporary gene flow in the other direction were inferred, given that mitochondrial DNA reflects older events than autosomal SNPs [55]. Complex intraspecific gene flow was also supported by Fst which showed considerable differences among pairs (Table S3).

Interspecific gene flow (i.e., introgression) from the *Laterallus* group lineage [cf. 23] to the continental population was influential, supported by a TreeMix model with one migration edge that best explained the intraspecific genetic covariance (log-likelihood = 148.557, migration weight = 0.080; Fig. 2c, S2-S4, Supplementary Discussion). Introgression was most likely from the Yellow Rail, given that the analysis inferred introgression from an unsampled *Laterallus* group species that split from the ancestor of the *Laterallus* group simultaneously with the Swinhoe's Rail (see Supplementary Discussion), and this reasonably explained that introgression caused the continental population to be basal in the ML SNP tree without considering gene flow (Fig. 2b, Fig. S5). Introgression occurred exclusively to the entire continental population, supported by genetic homogeneity in the PCA (PC1, 99% variance explained; Fig. 1b), ADMIXTURE analysis (K = 2–5; Fig. 1c), and the ML SNP tree (Fig. 2b).

Overall, the Tomakomai population would most likely be the basal population of the Swinhoe's Rail, and the common ancestor of the Kushiro and continental population diverged from it around the Middle Pleistocene (Fig. 6a–b). The continental population was its most derived population, which received both gene flow from the insular population and introgression from the *Laterallus* group lineage on the continent (Fig. 6c–d), supporting the reverse colonization scenario. Aomori of Honshu Island was re-colonized by the continental population.

Historical demography of reverse colonization and contemporary genetic variation

The reverse colonization scenario inferred by the tree topology was well corroborated by the demographic history. The common ancestor of the Kushiro and continental populations (ANC1) occurred through a founder event, supported by the split1 associated with a drastic population decline around the early Middle Pleistocene (0.9–0.8 Mya, SWP2), while leaving the Tomakomai population demographically stable (Fig. 3b–c). This caused a one- (SWP2) to two-fold (fastsimcoal2) decline in the effective population size of the ANC1 in comparison with the Tomakomai or the common ancestor of the entire populations (ANC2), resulting in N_e = 2,400 (fastsimcoal2) to 151,000–159,000 (SWP2). Historical stability of the Tomakomai population was also supported by the highest neutrality test values around zero (Fig. 4c–d), suggesting that the Tomakomai population was likely the source population from which a founder event occurred. Because the large and stable effective population size of the Tomakomai population was contradicted by its small contemporary effective population size (roughly two orders of magnitude smaller than estimates by SWP2 or fastsimcoal2; Table 1), it indicated that the Tomakomai population could have experienced drastic declines recently ~20,000 years ago which cannot be detected by our demographic reconstruction [56].

Then, another founder event from the ANC1 and its subsequent population expansion to the continent gave rise to the continental population, supported by the split2 associated with an increased population size of the continental population towards the late Middle Pleistocene (0.3–0.6 Mya, SWP2) to the Late Pleistocene (0.0538 Mya, fastsimcoal2), while leaving the Kushiro population demographically stable (Fig. 3b–c). Again, the demographic stability of the Kushiro population at the split2 suggests that the continental population was derived from a founder event. A genetic bottleneck specific to the continental population was supported by the highest level of drift parameter in the TreeMix analysis (drift parameter = 0.107; Fig. 2c) in a model considering introgression, while it was lowest among populations without introgression (drift parameter = 0.100; Fig. S5). This indicated that the additional founder event

causing genetic bottleneck was possibly associated specifically with continental colonization and that introgression was responsible for counteracting genetic drift.

The high genetic variation of the continental population in comparison with the two Japanese populations (Fig. 4a–b, Table S4) contradicted the repeated population bottlenecks inferred for the continental population. Finally, the population expansion on the continent caused an increase in effective population size by 1.5× (SWP2) to 20× (fastsimcoal2), resulting in their significantly negative neutrality test values (Fig. 4c–d).

SDMs

The best-fit model (response curve 'LQ' + regularization multiplier 1.0, AICc = 2446.725, model weight = 0.54, AUC = 92.5%) based on the present Yellow Rail occurrence data were projected to the environment layers of both North America and East Asia during different glacial and interglacial periods. We evaluated our best-fit model to be well reflecting the present distribution of both the Yellow Rail and the Swinhoe's Rail, and our assumptions of niche conservatism are thought to be met (see Discussion and Supplementary Results, Fig. S6).

During both glacial and interglacial periods, Alaska was suitable for this species, which was possibly the potential source of Asian colonization. The Japanese Archipelago was more likely to have been colonized by rails in Alaska than continental East Asia around the onset of the Middle Pleistocene (~0.8 Mya). This was supported by the abundant and stable species' suitable habitat of the Kuril and the Japanese Archipelagoes being more exposed to Alaska (Fig. 5e, f) than the continental East Asian (e.g., current breeding ranges, such as Amur) that was rather isolated by potentially unsuitable habitats in North Asia that do not resemble the present North American environment (colored in black in Fig. 5) during both the interglacial and glacial periods. Meanwhile, secondary contacts and introgression between North American and Asian continental populations could have been facilitated towards the Late Pleistocene, supported by suitable regions gradually expanded in North Asia along the coastline of the Sea of Okhotsk

(arrows in Fig. 5), towards Amur. At the same time, suitable habitats became increasingly less available along the Japanese Archipelago except for the LGM. Therefore, SDMs supported our phylogenomic scenarios that the Japanese Archipelago was more likely to be the source area for the Swinhoe's Rail before the Middle Pleistocene, and introgression was likely to have occurred on the continent from its sister species on either North America or the Asian part of Beringia (e.g. Kamchatka) towards the Late Pleistocene.

Discussion

Despite the importance of reverse colonization for tackling the traditional assumption of islands as sinks of biodiversity [2], how potentially genetically depauperate island populations [57] can successfully colonize the continent remained unexamined [7]. To answer this question, using an integrative phylogenomics approach, we examined reverse colonization not only through phylogenetic patterns but also from demographic perspectives, including population size changes, intraspecific gene flow, and introgression. We disentangled a complex population history of the Swinhoe's Rail; an American ancestry colonized the Japanese Archipelago first, and its island descendent back-colonized the Eurasian continent, representing reverse colonization. We found that the back-colonized continental population represented high contemporary genetic variation even after severe population bottlenecks. Demographic analysis suggested that introgression from an American sister lineage and intraspecific gene flow might explain its unexpectedly high genetic variation.

Demography affecting genetic variation under reverse colonization

It has long been assumed that the small effective population size of island populations [5,57] leads to low genetic variation and hence a low chance of reverse colonization [8]. Our results showed that the ancestral insular population retained a large effective population size, corroborating the recent arguments that the insular population is not necessarily genetically depauperate [8]. This is inferred by comparison with

previous pairwise sequentially Markovian coalescent (PSMC) analyses in birds that could be roughly comparable to our SWP2 analyses; our mean N_e (Table 1) was similar to that for large island populations, New Guinea (mean N_e = 325,000) [58], and continental populations (mean N_e = 362,456) [59]. Rather, an important decline in population size was associated with colonization along the Japanese Archipelago (Fig. 2c, Fig. 3b–c, Fig. 6b), which resulted in as a small population size as oceanic island populations (mean N_e = 94,944) [59]. This may have strongly lowered genetic variation because colonization along island chains usually causes severe genetic bottlenecks and facilitates non-adaptive genomic evolution [15]. Given that reverse colonization has been frequently reported along archipelagic chains [e.g. 8,60], reverse colonization processes may be responsible for reducing genetic variability more than the island population size itself.

Meanwhile, as high contemporary genetic variation (especially pairwise nucleotide diversity) of the continental population as the stable Tomakomai population contradicted many other cases experiencing population bottlenecks towards the late Middle to the Late Pleistocene [61,e.g. 62,63]. Introgression might have introduced a profound effect on restoring genetic variation of the continental population as has been supported by our TreeMix analysis (Fig. 2c, Fig. S5) to a larger extent than intra-specific gene flow and the other candidate process, mutation. Our results showed that genetic variation was not heightened in the Kushiro population despite a signature of historical intraspecific gene flow similar to the continental population in our mitochondrial analyses. This suggested that intraspecific gene flow had a minor effect on heightening genetic variation. The latter was supported by the genetic homogeneity of the continental population despite its widespread distribution (Fig. 1a–c), indicating that genetic variation accumulated more rapidly through introgression than through genetic mutation that should have been preceded by spatial and population expansion [61].

Introgression potentially results not only in microevolutionary consequences but also in a macroevolutionary outcome, diversification [64]. As our results showed, introgression can introduce genetic variation without disrupting within-species (population) divergence more likely than intraspecific

gene flow [65]. Such a population has the potential to undergo adaptive evolution in a novel environment [64,66] and repeated niche expansion, potentially resulting in diversification [e.g. 3,16]. Our case was unique in that both the back-colonized population and the source of introgression were produced through transcontinental colonization (Fig. 6a, c). A chance of introgression through reverse colonization is probably much higher in a conceivable situation where a source continental species persists on the continent where its insular descendent back-colonizes after completion of speciation. Therefore, introgression is one hopeful candidate to answer why diversification can be often led by reverse colonization. In this study, we could not quantify the degree of a gain in genetic variations through introgression by specifying introgressed genomic regions, because of the low genomic resolution of our dataset and not including genomic data of its possible source, the Yellow Rail. Nevertheless, it may underestimate but not overestimate the level of introgression, supported by a remnant of a positive residual between the continental population and the *Laterallus* group even after adding a migration edge between them [Fig. S3, 20]. Future investigations need to include these samples to evaluate the relative contribution of introgression to successful reverse colonization.

Reverse colonization in phylogeographic contexts

The island ancestry of the Swinhoe's Rail was possibly linked with the higher availability of their suitable habitats along the Japanese Archipelago than the continent. This scenario may also be adequate to the period of species divergence (the Late Pliocene to the Early Pleistocene; Fig. 2a) because the climate of the Early Pleistocene resembles the early Middle Pleistocene with low amplitudes of glacial cycles more than later Pleistocene periods with much severer cycles [67] (see Supplementary Discussion). The island ancestry might have occupied not only a part of Hokkaido but also Honshu Island (Fig. 6a). This is supported by 1) the two orders of magnitude larger population size for the ancestral and Tomakomai populations than its contemporary size, 2) suitable areas in Honshu during glacial and older interglacial periods but not in the present in our SDMs, and 3) our Kanto wintering samples showing an affinity towards Japanese breeding birds (Fig. 1b–c, 2b), since wintering regions of short-distance migratory birds

may reflect the past breeding locality [68]. Because islands are characterized by more stable climates relative to mainland environments [69], islands may have high potential to serve as refugia from which organisms eventually expand to the continent, as has been inferred in several taxonomic groups [bryophytes, 8, bats, 10]. Therefore, reverse colonization could be a biogeographic event that is associated likely with climatic fluctuations.

Limitations in demographic estimates

Despite several limitations in our demographic inference, these effects should be minor in our conclusions. First, effective population size estimates by different approaches reflect different historical events, making important differences in the level of population size changes. For example, the population size change for the split1 differed between the two analyses by more than one order of magnitude (Fig. 3, Table 1), which blurs the strength of the population bottleneck and hence possibly compromises the importance of introgression. Simplified demographic models of the fastsimcoal2 analysis may have concealed important demographic parameters, or the effect of gene flow cannot be taken into account by SWP2 resulting in an overestimation of the population size [56]. However, the largest estimate of the population size after bottleneck by SWP2 (N_e = 151,000–159,000) is still small in comparison with small oceanic island populations (mean N_e = 94,944) [59], not precluding the importance of introgression in reverse colonization.

Second, an assumption of niche conservatism for the SDM potentially causes an overestimation of the breeding range of the Swinhoe's Rail, because island colonization may lead to specialization, narrowing its niche width [70]. However, this only confines us to draw the breeding limit of the Swinhoe's Rail itself, but not the total range occupied by the two sister species. Hence, this does not affect our inference of their secondary contact. Rather, we suggest a possibility that introgression may have introduced genetic variants preferred in the continent, resulting in the continental range expansion.

Conclusion

Demographic processes after population establishment are fundamental to bridging a gap between micro- and macroevolution since they determine population genetics structure that is significantly linked with subsequent speciation and diversification [17]. This study reconstructed a dynamic demographic history of reverse colonization by a species belonging to a group that repeatedly underwent island colonization; the Japanese ancestral population that was established from Alaska colonized the Eurasian continent with severe population bottlenecks, and introgression from an American lineage has possibly assisted colonization by providing genetic variability. Islands have unique adaptive landscapes different from the continent, providing ecological opportunities for island diversification, but its consequence on the continental biodiversity has not been discussed. We argue that reverse colonization creates an important opportunity for an insular species to hybridize with its continental relatives, and it may generate high genetic variation and novel allelic combinations that detect unoccupied adaptive peaks on the continent [71], leading to another ecological opportunity for diversification on the continent. Interpreting islands as “evolutionary turning points” will facilitate our understanding of evolutionary mechanisms underlying global biodiversity.

Tables

Table 1 Comparison of effective population size estimates for the common ancestor and present Tomakomai populations derived by different methods

	Stairway Plot 2	Fastsimcoal2	N_c-based
N_e (common ancestor)	607,000–818,000	56,200	-
N_e (Tomakomai)	867,000	1,010,000	4,840

Figures

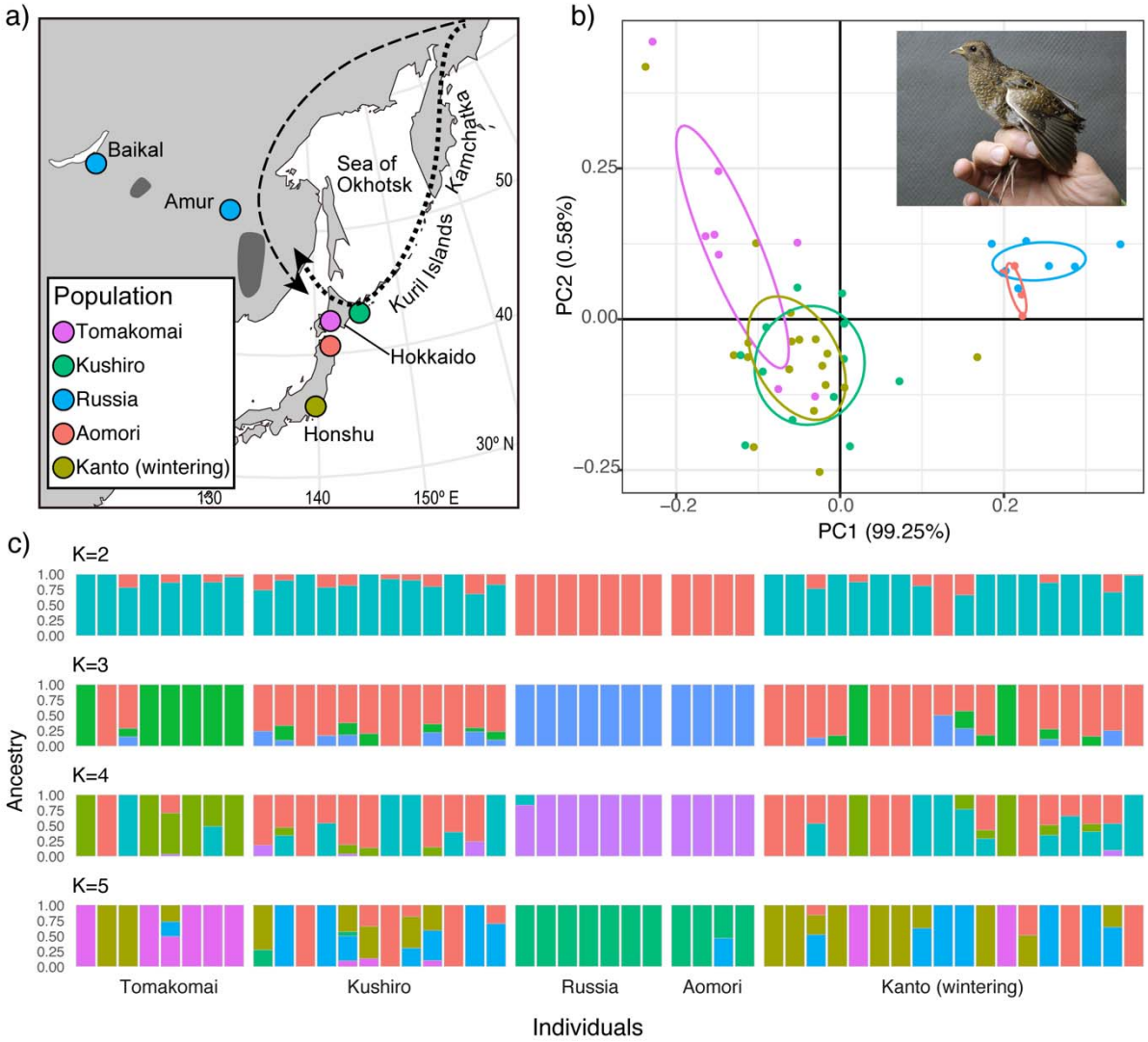

Figure 1 A map of sampling localities of Swinhoe's rails (a), and their population genetic structure inferred by (b) a principal component analysis (PCA) and (c) an ADMIXTURE analysis conducted using genotype likelihoods of 3,016 SNPs. (a, b) Dotted and hatched arrows indicate the reverse colonization and island colonization scenarios, respectively. 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 [27]. (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.

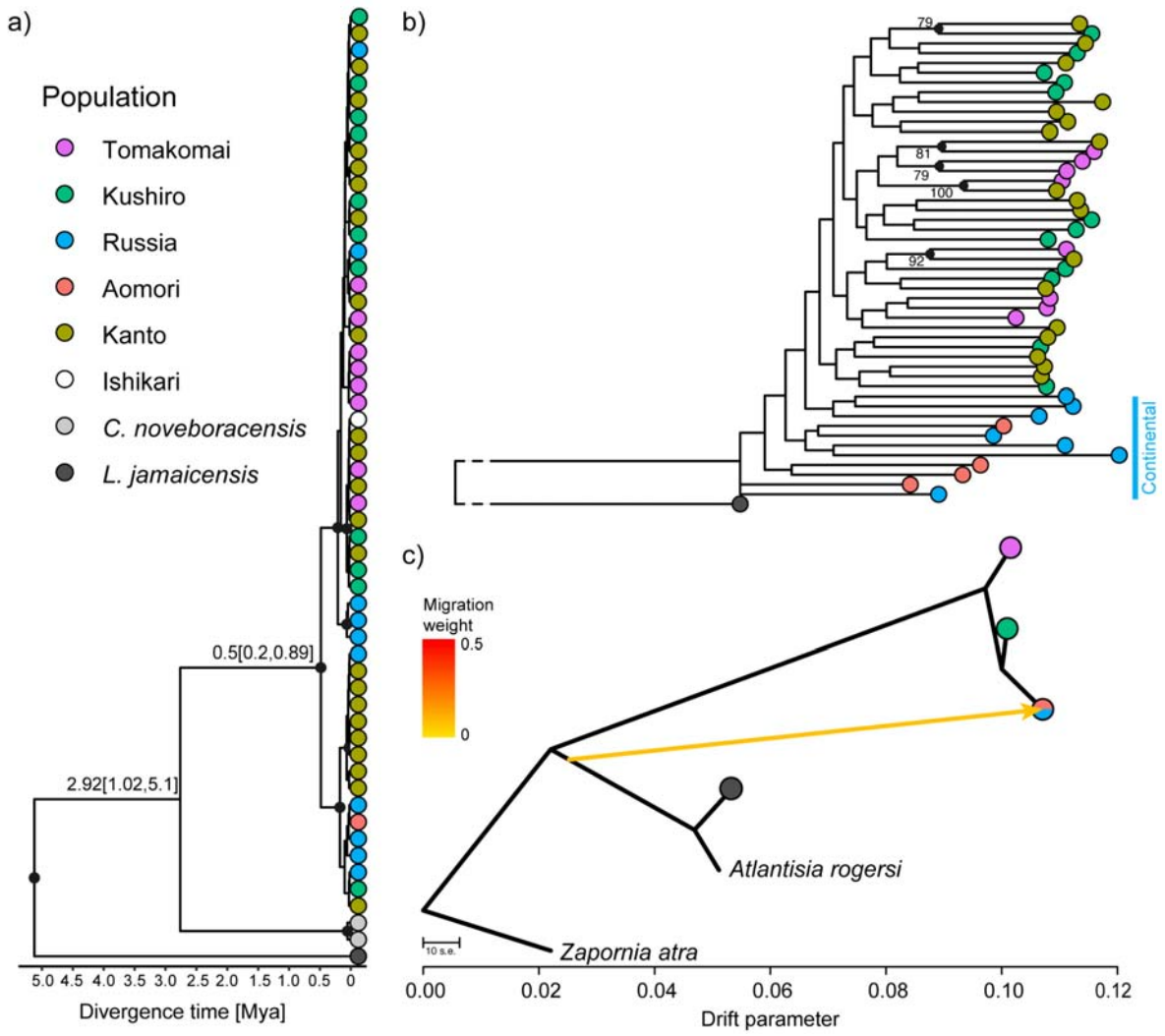

Figure 2 Phylogenetic trees of the Swinhoe's Rail reconstructed using different methods and genetic markers. (a) A mitochondrial phylogenetic tree using 772 bp of the *Cytb* gene. Posterior probabilities >0.9 were indicated by black dots on nodes. Numbers at the nodes indicate median divergence time estimates with 95% highest posterior density within brackets. (b) A maximum likelihood (ML) tree was reconstructed using 3,288 unlinked SNPs. Only nodes with bootstrap values >70 are annotated with black nodes. Branch lengths are shortened (hatched) for drawing. (c) A population-level tree reconstructed by TreeMix using 1,912 unlinked polymorphic loci. Branch lengths indicate drift parameters, the strength of drift specific to each branch. The yellow arrow indicates estimated gene flow. *C. noveboracensis* = *Coturnicops noveboracensis*, Yellow Rail; *L. jamaicensis* = *Laterallus jamaicensis*, Black Rail.

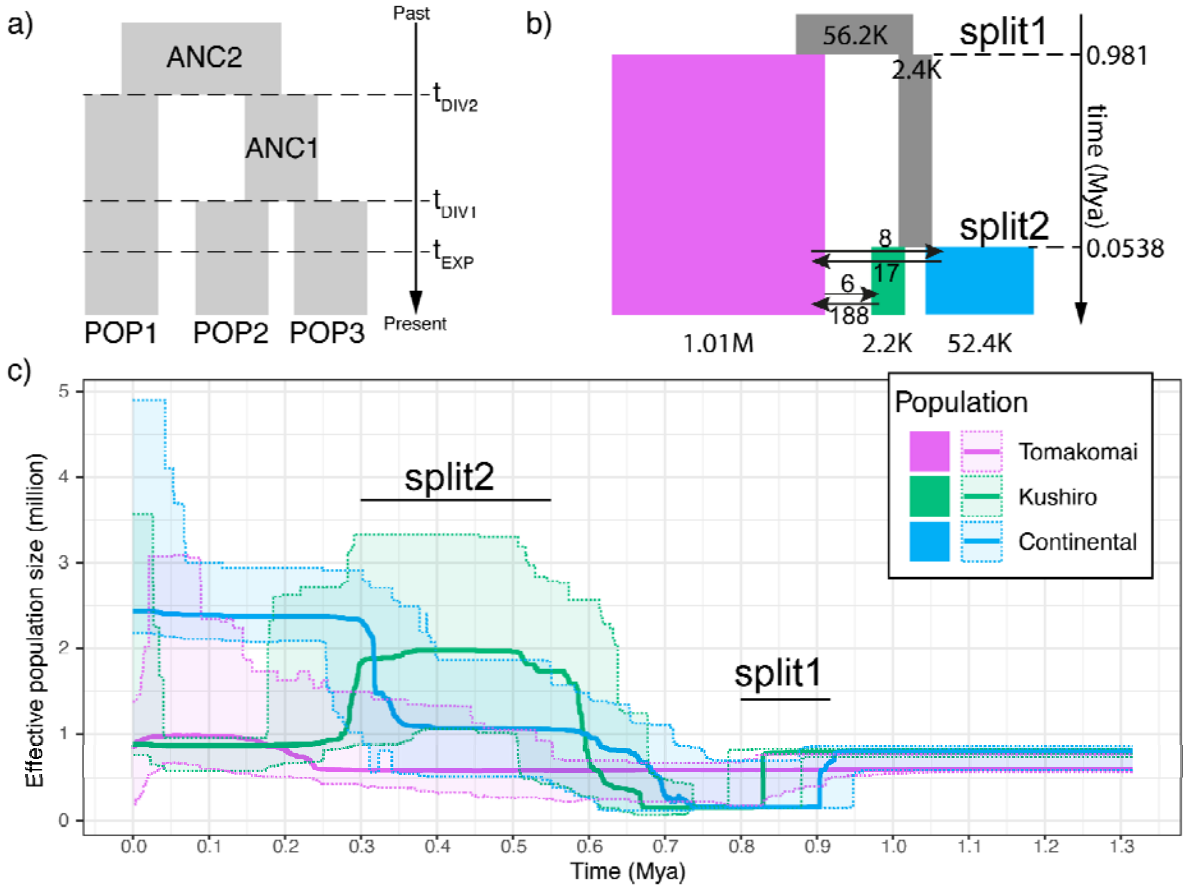

Figure 3 Demographic inference of the three genetic populations of the Swinhoe's Rail based on site frequency spectra, using (a, b) fastsimcoal2 and (c) Stairway Plot 2. (a) Generalized schematic representation of the six models tested. Parameters with upper cases represent effective population size at specific event times t_{EVENTS}. Effective population size at the time t_{EXP} and migration are not shown. (b) The best-fit model with parameter estimates. Population size is shown in either million (M) or kilo (K). Arrows with numbers between populations represent the number of migrants/year. (c) Changes in effective population size (in millions) over time (in million years) are shown for each genetic population. Bold solid lines and ribbons around them show median and 95% confidence intervals, respectively. Effective population size and time are scaled by mutation rate (1.91e-09) and generation time (1 year).

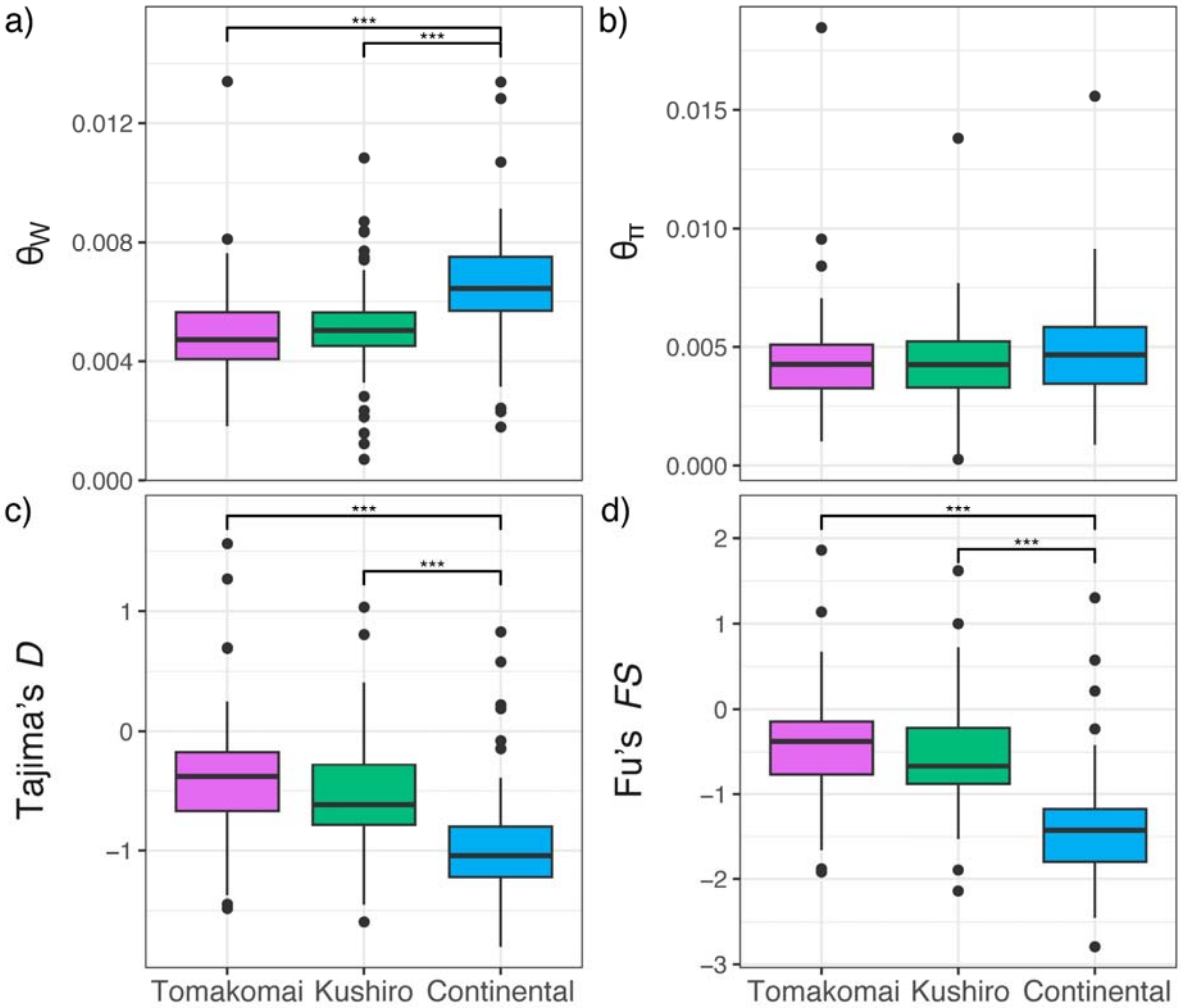

Figure 4 Comparison of scaffold-level statistics across three genetic populations, including two genetic diversity indices, (a) Watterson's theta and (b) nucleotide diversity, and two neutrality test statistics, (c) Tajima's *D* and (d) Fu's *FS*. Population pairs with significant differences in the values of each index were annotated with asterisks (*** $p < 0.001$, no signature for $p > 0.05$ after Bonferroni correction).

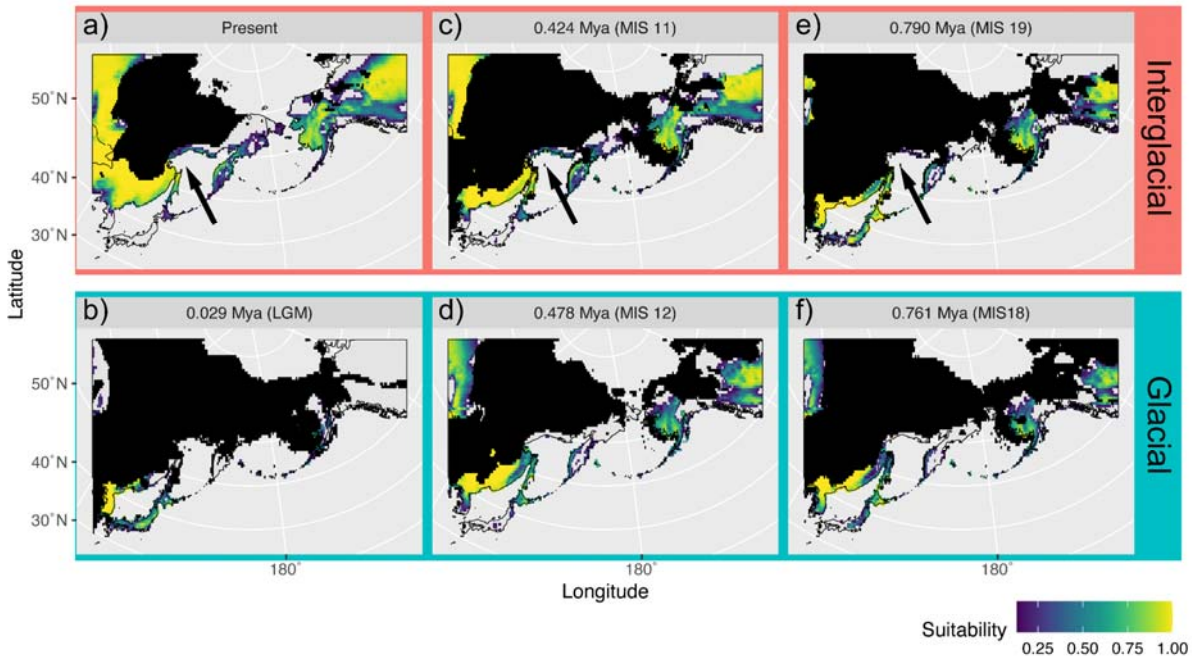

Figure 5 Species distribution models constructed using the present occurrence records of the Yellow Rail in North America were projected to Eurasia and North America in different time slices, including interglacial periods (a, c, e) as well as glacial periods (b, d, f). Color gradients from indigo to yellow represent environmental suitability for rails. Regions with black are where model extrapolation was restricted due to the absence of the climate in the modeling domain.

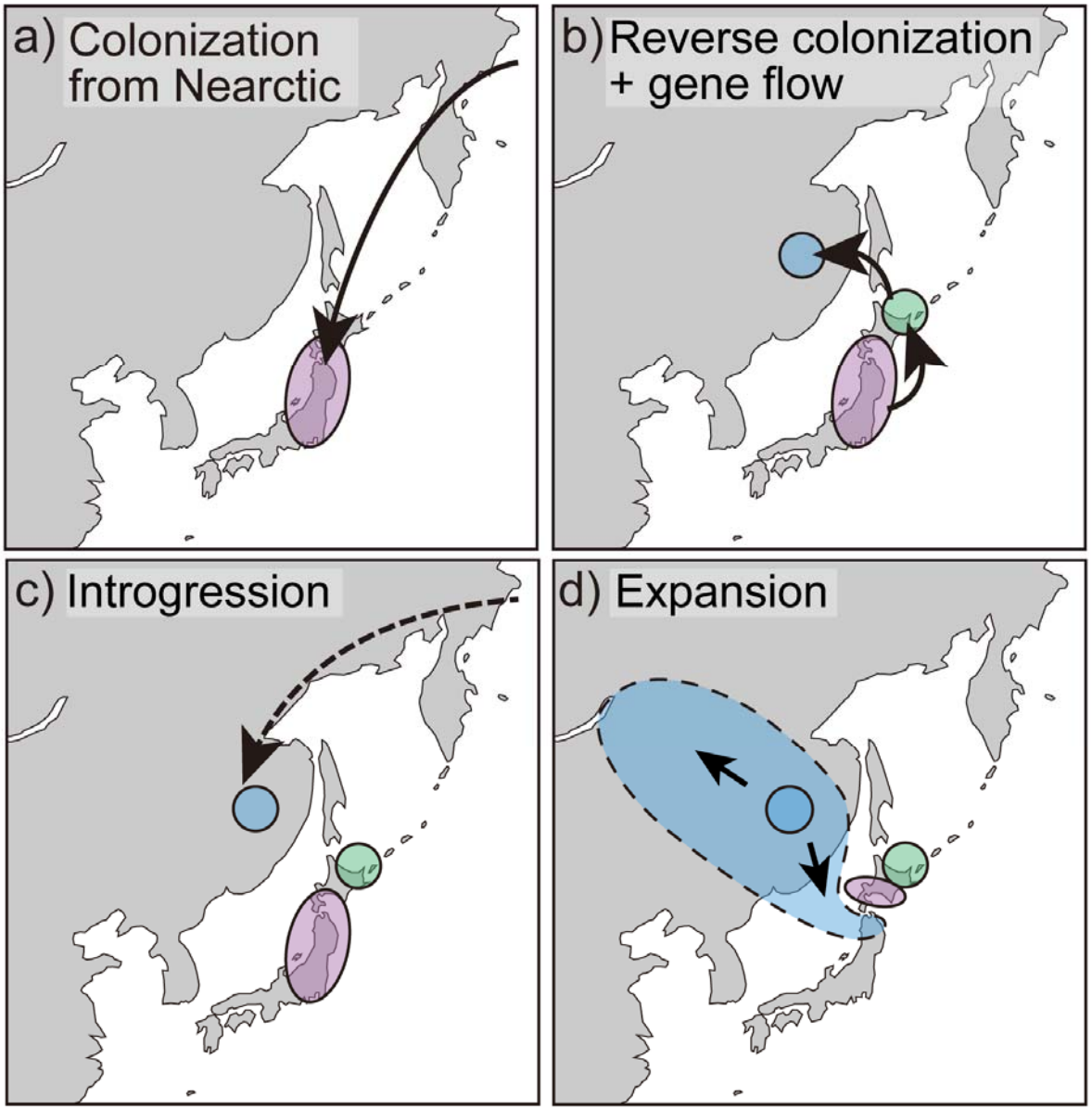

Figure 6 Schematic representation of our reverse colonization scenario of the Swinhoe's Rail. In this scenario, (a) the ancestral population of rails colonized the Japanese Archipelago from North America, then (b) the Asian continent was back-colonized by this insular population along the island chain, and (c) the continental population received introgression from its close relative on the continent. Finally, (d) the continental population expanded while the insular population shrank its population size.

Data Availability Statement

A preprint is available at [72]. The datasets and codes used in this study are available in the Dryad Digital Repository [26].

References

1. Whittaker RJ, Fernández-Palacios JM. 2007 *Island Biogeography: Ecology, Evolution, and Conservation*. Oxford: Oxford University Press.
2. Whittaker RJ, Fernández-Palacios JM, Matthews TJ, Borregaard MK, Triantis KA. 2017 Island biogeography: Taking the long view of nature's laboratories. *Science* **357**, eaam8326. (doi:10.1126/science.aam8326)
3. Cerca J *et al.* 2023 Evolutionary genomics of oceanic island radiations. *Trends in Ecology & Evolution* **38**, 631–642. (doi:10.1016/j.tree.2023.02.003)
4. Burga A, Wang W, Ben-David E, Wolf PC, Ramey AM, Verdugo C, Lyons K, Parker PG, Kruglyak L. 2017 A genetic signature of the evolution of loss of flight in the Galapagos cormorant. *Science* **356**, eaal3345. (doi:10.1126/science.aal3345)
5. Frankham R. 2005 Genetics and extinction. *Biological Conservation* **126**, 131–140. (doi:10.1016/j.biocon.2005.05.002)
6. Vamosi JC, Armbruster WS, Renner SS. 2014 Evolutionary ecology of specialization: insights from phylogenetic analysis. *Proceedings of the Royal Society B: Biological Sciences* **281**, 20142004. (doi:10.1098/rspb.2014.2004)
7. Bellemain E, Ricklefs R. 2008 Are islands the end of the colonization road? *Trends in Ecology & Evolution* **23**, 461–468. (doi:10.1016/j.tree.2008.05.001)
8. Patiño J, Carine M, Mardulyn P, Devos N, Mateo RG, González-Mancebo JM, Shaw AJ, Vanderpoorten A. 2015 Approximate bayesian computation reveals the crucial role of oceanic islands for the assembly of continental biodiversity. *Systematic Biology* **64**, 579–589. (doi:10.1093/sysbio/syv013)
9. O'Grady P, DeSalle R. 2008 Out of Hawaii: the origin and biogeography of the genus Scaptomyza (Diptera: Drosophilidae). *Biology Letters* **4**, 195–199. (doi:10.1098/rsbl.2007.0575)

10. Ikeda Y, Motokawa M. 2021 Phylogeography of the Japanese greater horseshoe bat *Rhinolophus nippon* (Mammalia: Chiroptera) in Northeast Asia: New insight into the monophyly of the Japanese populations. *Ecology and Evolution* **11**, 18181–18195. (doi:10.1002/ece3.8414)

11. Jønsson KA, Holt BG. 2015 Islands contribute disproportionately high amounts of evolutionary diversity in passerine birds. *Nat Commun* **6**, 8538. (doi:10.1038/ncomms9538)

12. Alföldi J *et al.* 2011 The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477**, 587–591. (doi:10.1038/nature10390)

13. Hart SP, Turcotte MM, Levine JM. 2019 Effects of rapid evolution on species coexistence. *Proceedings of the National Academy of Sciences* **116**, 2112–2117. (doi:10.1073/pnas.1816298116)

14. Vellend M, Geber MA. 2005 Connections between species diversity and genetic diversity. *Ecology Letters* **8**, 767–781. (doi:10.1111/j.1461-0248.2005.00775.x)

15. Sendell-Price AT, Ruegg KC, Robertson BC, Clegg SM. 2021 An island-hopping bird reveals how founder events shape genome-wide divergence. *Molecular Ecology* **30**, 2495–2510. (doi:10.1111/mec.15898)

16. Choi JY *et al.* 2021 Ancestral polymorphisms shape the adaptive radiation of *Metrosideros* across the Hawaiian Islands. *Proceedings of the National Academy of Sciences* **118**, e2023801118. (doi:10.1073/pnas.2023801118)

17. Harvey MG, Singhal S, Rabosky DL. 2019 Beyond reproductive isolation: demographic controls on the speciation process. *Annu. Rev. Ecol. Evol. Syst.* **50**, 75–95. (doi:10.1146/annurev-ecolsys-110218-024701)

18. Lomolino MV, Riddle BR, Whittaker RJ. 2017 *Biogeography: Biological Diversity across Space and Time*. 5th Edition. Maryland: Sinauer Associates.

19. Beichman AC, Huerta-Sanchez E, Lohmueller KE. 2018 Using genomic data to infer historic population dynamics of nonmodel organisms. *Annu. Rev. Ecol. Evol. Syst.* **49**, 433–456. (doi:10.1146/annurev-ecolsys-110617-062431)

20. Pickrell JK, Pritchard JK. 2012 Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genetics* **8**, e1002967. (doi:10.1371/journal.pgen.1002967)

21. Skotte L, Korneliussen TS, Albrechtsen A. 2013 Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**, 693–702. (doi:10.1534/genetics.113.154138)

22. Liu X, Fu Y-X. 2020 Stairway Plot 2: demographic history inference with folded SNP frequency

spectra. *Genome Biology* **21**, 280. (doi:10.1186/s13059-020-02196-9)

23. Garcia-R JC, Matzke NJ. 2021 Trait-dependent dispersal in rails (Aves: Rallidae): Historical biogeography of a cosmopolitan bird clade. *Molecular Phylogenetics and Evolution* **159**, 107106. (doi:10.1016/j.ympev.2021.107106)

24. Tindall JC, Haywood AM, Salzmann U, Dolan AM, Fletcher T. 2022 The warm winter paradox in the Pliocene northern high latitudes. *Climate of the Past* **18**, 1385–1405. (doi:10.5194/cp-18-1385-2022)

25. Garcia-R JC, Joseph L, Adcock G, Reid J, Trewick SA. 2017 Interisland gene flow among populations of the buff-banded rail (Aves: Rallidae) and its implications for insular endemism in Oceania. *Journal of Avian Biology* **48**, 679–690. (doi:10.1111/jav.01201)

26. Aoki D *et al.* 2023 Phylogenomics reveals a gene flow history of reverse colonization. Dryad Digital Repository. (http://dx.doi.org/10.5061/dryad.37pvmcvs1)

27. Heim W, Trense D, Heim A, Kamp J, Smirenski SM, Wink M, Wulf T. 2019 Discovery of a new breeding population of the Vulnerable Swinhoe's Rail *Coturnicops exquisitus* confirmed by genetic analysis. *Bird Conservation International* **29**, 454–462. (doi:10.1017/S0959270918000138)

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

29. Miya A, Sannohe S, Ebina J, Sekishita H. 2005 A survey report on Yellow Rail (*Coturnicops noveboracensis exquisitus*) observed during the breeding season at Hotokenuma Reclaimed Land in Misawa, Aomori Prefecture. *Japanese Journal of Ornithology* **54**, 104–107. (doi:10.3838/jjo.54.104)

30. Senzaki M, Kitazawa M, Sadakuni T, Takahashi M. 2020 Breeding evidence of the vulnerable Swinhoe's Rail (*Coturnicops exquisitus*) in Japan. *The Wilson Journal of Ornithology* **132**, 711–717. (doi:10.1676/19-45)

31. 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. *Sci Rep* **5**, 16963. (doi:10.1038/srep16963)

32. Bolger AM, Lohse M, Usadel B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. (doi:10.1093/bioinformatics/btu170)

33. Li H. 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*

34. Korneliussen TS, Albrechtsen A, Nielsen R. 2014 ANGSD: Analysis of Next Generation Sequencing

Data. *BMC Bioinformatics* **15**, 356. (doi:10.1186/s12859-014-0356-4)

35. Fox EA, Wright AE, Fumagalli M, Vieira FG. 2019 ngsLD: evaluating linkage disequilibrium using genotype likelihoods. *Bioinformatics* **35**, 3855–3856. (doi:10.1093/bioinformatics/btz200)

36. Meisner J, Albrechtsen A. 2018 Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* **210**, 719–731. (doi:10.1534/genetics.118.301336)

37. Watterson GA. 1975 On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**, 256–276. (doi:10.1016/0040-5809(75)90020-9)

38. Tajima F. 1983 Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437–460. (doi:10.1093/genetics/105.2.437)

39. Tajima F. 1989 The effect of change in population size on DNA polymorphism. *Genetics* **123**, 597–601. (doi:10.1093/genetics/123.3.597)

40. Fu Y-X. 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925. (doi:10.1093/genetics/147.2.915)

41. Twyford AD, Ennos RA. 2012 Next-generation hybridization and introgression. *Heredity* **108**, 179–189. (doi:10.1038/hdy.2011.68)

42. Bouckaert R *et al.* 2019 BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* **15**, e1006650. (doi:10.1371/journal.pcbi.1006650)

43. Weir JT, Schluter D. 2008 Calibrating the avian molecular clock. *Molecular Ecology* **17**, 2321–2328. (doi:10.1111/j.1365-294X.2008.03742.x)

44. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019 RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455. (doi:10.1093/bioinformatics/btz305)

45. Fitak RR. 2021 OptM: estimating the optimal number of migration edges on population trees using Treemix. *Biol Methods Protoc* **6**. (doi:10.1093/biomethods/bpab017)

46. Excoffier L, Marchi N, Marques DA, Matthey-Doret R, Gouy A, Sousa VC. 2021 fastsimcoal2: demographic inference under complex evolutionary scenarios. *Bioinformatics* **37**, 4882–4885. (doi:10.1093/bioinformatics/btab468)

47. Nam K *et al.* 2010 Molecular evolution of genes in avian genomes. *Genome Biology* **11**, R68. (doi:10.1186/gb-2010-11-6-r68)

48. Kitazawa M, Yamaura Y, Senzaki M, Hanioka M, Ohashi H, Oguro M, Matsui T, Nakamura F. 2022

Quantifying the impacts of 166 years of land cover change on lowland bird communities. *Proceedings of the Royal Society B: Biological Sciences* **289**, 20220338. (doi:10.1098/rspb.2022.0338)

49. Frankham R. 1995 Effective population size/adult population size ratios in wildlife: a review. *Genetics Research* **66**, 95–107. (doi:10.1017/S0016672300034455)

50. Phillips SJ, Anderson RP, Schapire RE. 2006 Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**, 231–259. (doi:10.1016/j.ecolmodel.2005.03.026)

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

52. Leonardi M, Hallett EY, Beyer R, Krapp M, Manica A. 2023 pastclim 1.2: an R package to easily access and use paleoclimatic reconstructions. *Ecography* **n/a**, e06481. (doi:10.1111/ecog.06481)

53. Krapp M, Beyer RM, Edmundson SL, Valdes PJ, Manica A. 2021 A statistics-based reconstruction of high-resolution global terrestrial climate for the last 800,000 years. *Sci Data* **8**, 228. (doi:10.1038/s41597-021-01009-3)

54. Kass JM, Muscarella R, Galante PJ, Bohl CL, Pinilla□Buitrago GE, Boria RA, Soley□Guardia M, Anderson RP. 2021 ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions. *Methods Ecol Evol* **12**, 1602–1608. (doi:10.1111/2041-210X.13628)

55. de Jong MJ *et al.* 2023 Range-wide whole-genome resequencing of the brown bear reveals drivers of intraspecies divergence. *Commun Biol* **6**, 1–16. (doi:10.1038/s42003-023-04514-w)

56. Nadachowska-Brzyska K, Konczal M, Babik W. 2022 Navigating the temporal continuum of effective population size. *Methods in Ecology and Evolution* **13**, 22–41. (doi:10.1111/2041-210X.13740)

57. Frankham R. 1997 Do island populations have less genetic variation than mainland populations? *Heredity* **78**, 311–327. (doi:10.1038/hdy.1997.46)

58. Pujolar JM *et al.* 2022 The formation of avian montane diversity across barriers and along elevational gradients. *Nat Commun* **13**, 268. (doi:10.1038/s41467-021-27858-5)

59. Leroy T *et al.* 2021 Island songbirds as windows into evolution in small populations. *Current Biology* **31**, 1303-1310.e4. (doi:10.1016/j.cub.2020.12.040)

60. Jønsson KA, Fabre P-H, Ricklefs RE, Fjeldsa J. 2011 Major global radiation of corvoid birds originated in the proto-Papuan archipelago. *Proceedings of the National Academy of Sciences* **108**,

2328–2333. (doi:10.1073/pnas.1018956108)

61. Excoffier L, Foll M, Petit RJ. 2009 Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics* **40**, 481–501. (doi:10.1146/annurev.ecolsys.39.110707.173414)

62. Boria RA, Brown SK, Matocq MD, Blois JL. 2021 Genome-wide genetic variation coupled with demographic and ecological niche modeling of the dusky-footed woodrat (*Neotoma fuscipes*) reveal patterns of deep divergence and widespread Holocene expansion across northern California. *Heredity* **126**, 521–536. (doi:10.1038/s41437-020-00393-7)

63. Saitoh T, Alström P, Nishiumi I, Shigeta Y, Williams D, Olsson U, Ueda K. 2010 Old divergences in a boreal bird supports long-term survival through the Ice Ages. *BMC Evol Biol* **10**, 35. (doi:10.1186/1471-2148-10-35)

64. Seehausen O. 2013 Conditions when hybridization might predispose populations for adaptive radiation. *Journal of Evolutionary Biology* **26**, 279–281. (doi:10.1111/jeb.12026)

65. Seehausen O. 2004 Hybridization and adaptive radiation. *Trends in Ecology & Evolution* **19**, 198–207. (doi:10.1016/j.tree.2004.01.003)

66. Taylor SA, Larson EL. 2019 Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nat Ecol Evol* **3**, 170–177. (doi:10.1038/s41559-018-0777-y)

67. Head MJ. 2021 Review of the Early–Middle Pleistocene boundary and Marine Isotope Stage 19. *Progress in Earth and Planetary Science* **8**, 50. (doi:10.1186/s40645-021-00439-2)

68. Zink RM, Gardner AS. 2017 Glaciation as a migratory switch. *Sci. Adv.* **3**, e1603133. (doi:10.1126/sciadv.1603133)

69. Weigelt P, Jetz W, Kreft H. 2013 Bioclimatic and physical characterization of the world's islands. *Proceedings of the National Academy of Sciences* **110**, 15307–15312. (doi:10.1073/pnas.1306309110)

70. Wilson EO. 1961 The nature of the taxon cycle in the Melanesian ant fauna. *The American Naturalist* **95**, 169–193.

71. Stroud JT, Losos JB. 2016 Ecological opportunity and adaptive radiation. *Annual Review of Ecology, Evolution, and Systematics* **47**, 507–532. (doi:10.1146/annurev-ecolsys-121415-032254)

72. Aoki et al. 2023 Phylogenomics reveals demography of reverse colonization in a rail species. *bioRxiv*. (doi: 10.1101/2023.09.27.559751)

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. DNA samples were collected with permission from the Ministry of the Environment Government of Japan. Our field surveys were conducted with permission from the landowners of each site, including Tomatoh. We 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. 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). Bioinformatics was mostly performed on the supercomputer of AFFRIT, MAFF, Japan.