PNAS NEXUS

PNAS Nexus, 2026, **5**, pgag037

https://doi.org/10.1093/pnasnexus/pgag037
Advance access publication 17 March 2026
Research Report

Discovering and protecting cryptic biodiversity: A case study of a previously undescribed, vulnerable bird species in Japan

Takema Saitoh [a,*1], Daria Shipilina [b,1,2], Canwei Xia [b,c], Lijun Zhang [b,c], Shin-Ichi Seki [d], Urban Olsson [e,f] and Per Alström [b,g,*]

[a]Yamashina Institute for Ornithology, 115 Konoyama, Abiko, Chiba 270-1145, Japan
[b]Animal Ecology, Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18 D, Uppsala SE-752 36, Sweden
[c]Ministry of Education Key Laboratory for Biodiversity and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing 100875, China
[d]Kansai Research Center, Forestry and Forest Products Research Institute, Nagaikyutaro 68, Kyoto 612-0855, Japan
[e]Systematics and Biodiversity, Department of Zoology, Gothenburg University, Box 463, Göteborg SE-405 30, Sweden
[f]Gothenburg Global Biodiversity Centre, Box 461, Gothenburg SE-405 30, Sweden
[g]State Key Laboratory of Animal Biodiversity Conservation and Integrated Pest Management, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China
*To whom correspondence should be addressed: Email: saitoh@yamashina.or.jp (T.S.); Email: per.alstrom@ebc.uu.se (P.A.)
[1]T.S. and D.S. contributed equally to this work.
[2]Present address: Division of Systematic and Evolutionary Botany, Department of Botany and Biodiversity Research, University of Vienna, Rennweg 14, A-1030 Vienna, Austria
Edited By Zhonghe Zhou

Abstract

Despite the escalating biodiversity crisis, many species remain unknown to science and may even disappear unnoticed. This is particularly true for many island populations. We illustrate the problem of detecting overlooked species and its consequences by exploring a rare and geographically restricted migratory songbird. We find that this consists of two—hence even rarer—species: the Japanese endemic Ijima's Leaf Warbler *Phylloscopus ijimae* from the Izu Islands and the Tokara Leaf Warbler from the Tokara Islands. We describe the latter as a new cryptic species, ie one that is morphologically highly similar to, but genetically distinct from, a known species. The genetic divergence is revealed by analyses of nuclear genome-wide and mitochondrial DNA and supported by differences in vocalizations, while the morphological differences are minimal. We evaluate key conservation genomic indicators, showing that both species show low levels of genetic diversity and signs of a decrease of effective population size. Our genome-wide analysis revealed short runs of homozygosity and a low estimated deleterious load, suggesting limited recent inbreeding and possible purging of harmful alleles—indicators of genetic recovery after past demographic fluctuations. Ijima's Leaf Warbler is already classified as Vulnerable as well as a "Natural Monument" in Japan, and we propose that the Tokara Leaf Warbler should retain this status, with continued focused monitoring. Our study not only highlights the importance of integrating genomics with taxonomy for uncovering cryptic avian diversity but also provides a critical foundation for future conservation efforts.

Significance Statement

Island species are particularly vulnerable to extinction, yet many remain undiscovered. Our study demonstrates how genomics can uncover hidden biodiversity and provide critical metrics of population health. Our findings highlight the urgent global need to formally recognize and evaluate the conservation status of cryptic species, ensuring that overlooked lineages receive protection in a rapidly changing world. This study adds a whole-genome conservation assessment in passerine birds to the growing body of work applying genomics to conservation.

Introduction

In the current global biodiversity crisis, the identification of overlooked species, such as cryptic taxa, is crucial for conservation, as their unrecognized status may leave them unprotected. Estimates of the number of species on Earth range from 2 million to 3 trillion, yet only a fraction has been described (1). A major contributor to this gap is the presence of cryptic taxa whose prevalence may be immense; for example, each morphology-based insect species may contain on average 3.1 cryptic species (2).

Modern genomic approaches are particularly powerful for detecting hidden diversity. They complement traditional taxonomic methods by identifying cryptic divergence that may not be evident

Competing Interest: The authors declare no competing interests.
Received: October 29, 2025. **Accepted:** January 26, 2026

morphologically. In addition, genomics provide essential conservation indicators—such as genetic diversity, inbreeding coefficients, and genetic load—that inform management and policy decisions (3, 4). These metrics are especially critical for small, isolated island populations, where geographic isolation fosters speciation but limited habitats heighten vulnerability to climate change, catastrophic events, and human activity. Yet, despite their potential, comprehensive genomic assessments of threatened species remain relatively rare.

Island birds, in particular, provide an ideal system for applying integrative taxonomic and conservation genomic approaches. Because birds are widely used as "barometers for planetary health" (5), accurate delimitation of species and assessment of their conservation status have broad implications for biodiversity protection. Avian diversity is better documented than that of any other major group of organisms, with more than 11,000 species currently recognized (6, 7). Yet, integrative taxonomy—where genomics complements phenotypic, ecological, and distributional data—has contributed to a steady increase in the number of recognized bird species over recent decades (e.g., (8), 8,590 species). The true total could be double the current figure (9), highlighting the need for further studies, particularly in taxa known to harbor cryptic diversity and in vulnerable, overlooked regions.

The avian family Phylloscopidae (leaf warblers) offers a striking example of hidden and under-described biodiversity, as demonstrated by the ~50% increase since the 1980s (6,; review in 10). These small, insectivorous, Old-World birds are known for their cryptic appearances but usually more easily distinguishable songs (11). Within this family, the species in the West Pacific archipelagos are particularly poorly studied, and it seems likely that in-depth investigations may increase the number of recognized species. This is supported by the recent descriptions of three new leaf warblers from Indonesia based on analyses of genomics, vocalizations, and morphology (12, 13). Yet species inhabiting West Pacific islands are likely vulnerable, based on the observation that more than 90% of bird extinctions in the past 400 years have occurred on islands (14). This highlights the need for evaluations of the conservation status of species in these archipelagos, especially based on modern genomic methods.

Ijima's Leaf Warbler *Phylloscopus ijimae* breeds in two Japanese archipelagos, separated by a distribution gap of ~1,000 km: the Izu Islands, southeast of Honshu, and the Tokara Islands, which are part of the Nansei Shoto or Ryukyu Islands, southwest of Kyushu (Fig. 1A). The latter population was only discovered in 1988 on the island of Nakanoshima (15), and the species has been observed in the breeding season on four other small nearby islands (16, 17). The species is listed as Vulnerable (VU; (18)) and as a "Natural Monument" (19). Since Higuchi and Kawaji (15), the Izu and Tokara Island populations of *P. ijimae* have been considered undifferentiated. However, based on an integrative study of genome-wide single-nucleotide polymorphisms (SNPs) and introns, mitochondrial DNA (mtDNA), morphology, and vocalizations, we conclude that these populations are strongly divergent—except in morphology—and show no signs of present or recent interbreeding. We describe the Tokara Islands population as a species new to science and evaluate its conservation status based on our genomic data.

Results and discussion

Geographically separated island populations of *P. ijimae* are cryptic in morphology but differ in song

We confirmed that the Izu and Tokara Island populations are extremely similar morphologically: our comparison of museum specimens detected no plumage differences between them, and they were essentially undifferentiated in a principal component analysis (PCA) of morphometrics (Fig. S1, Table S1), although there were significant differences in tarsus length and total head length (Table S2). See Section S3.1 for further details.

In contrast, differences in song between the populations from Nakanoshima in the Tokara Islands (hereafter Tokara) and Miyakejima in the Izu Islands (hereafter Izu) were more pronounced (Fig. 1D–P, Section S3.2). In both populations, we classified the song into two main types (types I and II; Figs. 1 and S2), which were often alternated between. Type I, which was by far the commonest, differed significantly between Tokara and Izu in all measured variables (Table S3), and in sonograms only a fraction of the songs shared extremely similar syllables between the two islands (Fig. S2). A PCA generated three principal components (PCs) with eigenvalues >1.0, explaining in total 81.0% of the variance (Table S4). In plots of the three first PCs, the songs from Tokara and Izu formed largely separate clusters along PC1, with considerably more overlap along the other axes (Figs. 1H and S3). In a discriminant function analysis (DFA), 100% of the recordings from Tokara and 97% of the ones from Izu were correctly classified (Table S5). Although our sample of type II songs was small, there were significant differences between Tokara and Izu in a few variables (Table S6), and both the PCA and DFA identified differences between the two populations, with 90% of the recordings from Tokara and 92% of the ones from Izu correctly classified in the DFA (Tables S7 and S8, Fig. S4). In contrast to song, the contact calls were not found to be significantly different between Tokara and Izu, although the sample sizes were small (Tables S9 and S10, Figs. S5 and S6). See Section S3.2 for further details of vocalizations.

The poor morphological differentiation but stronger divergence in song between the Tokara and Izu populations is in agreement with several species of leaf warblers, while a few leaf warblers show the opposite pattern (Table S11).

Phylogenetic and population genomic analyses demonstrate a deep split between island populations of *P. ijimae*

The populations on Tokara and Izu are genetically deeply diverged, as concordantly shown by phylogenetic analyses of whole-genome autosomal SNPs (Figs. 2A and S8), genome-wide introns (Fig. 2B), and mtDNA (Figs. 2C, S7, S9, and S10), as well as population structure analyses (PCA, Fig. 3A; ADMIXTURE; Fig. 3B) and population summary statistics (F_{ST}, d_{XY}, Fig. 3C; Table S17; window-based scans Fig. S11).

The genome-wide SNP phylogeny (Fig. 2A) of the entire Eastern Crowned Warbler "*Phylloscopus coronatus* clade", including both *P. ijimae* populations, showed a clear divergence between the Tokara and Izu populations. However, the relative divergence between Lemon-throated Leaf Warbler *Phylloscopus cebuensis*, Philippine Leaf Warbler *Phylloscopus olivaceus* and the *P. ijimae* populations depended on the sampling scheme. When all samples were included, as in this figure ($n = 8$ Tokara, $n = 5$ Izu, $n = 2$ *P. olivaceus*, $n = 1$ *P. cebuensis*), the *P. ijimae* divergence was considerably deeper than when two samples per species ($n = 1$ *P. cebuensis*) were analyzed (Fig. S8). This might be due to the deeper intrapopulation coalescence captured by the larger samples of the two *P. ijimae* populations compared to the two other species. The phylogeny based on genome-wide introns (Fig. 2B) also revealed a deep divergence between the Tokara and Izu populations, comparable to or greater than that between multiple other leaf warbler

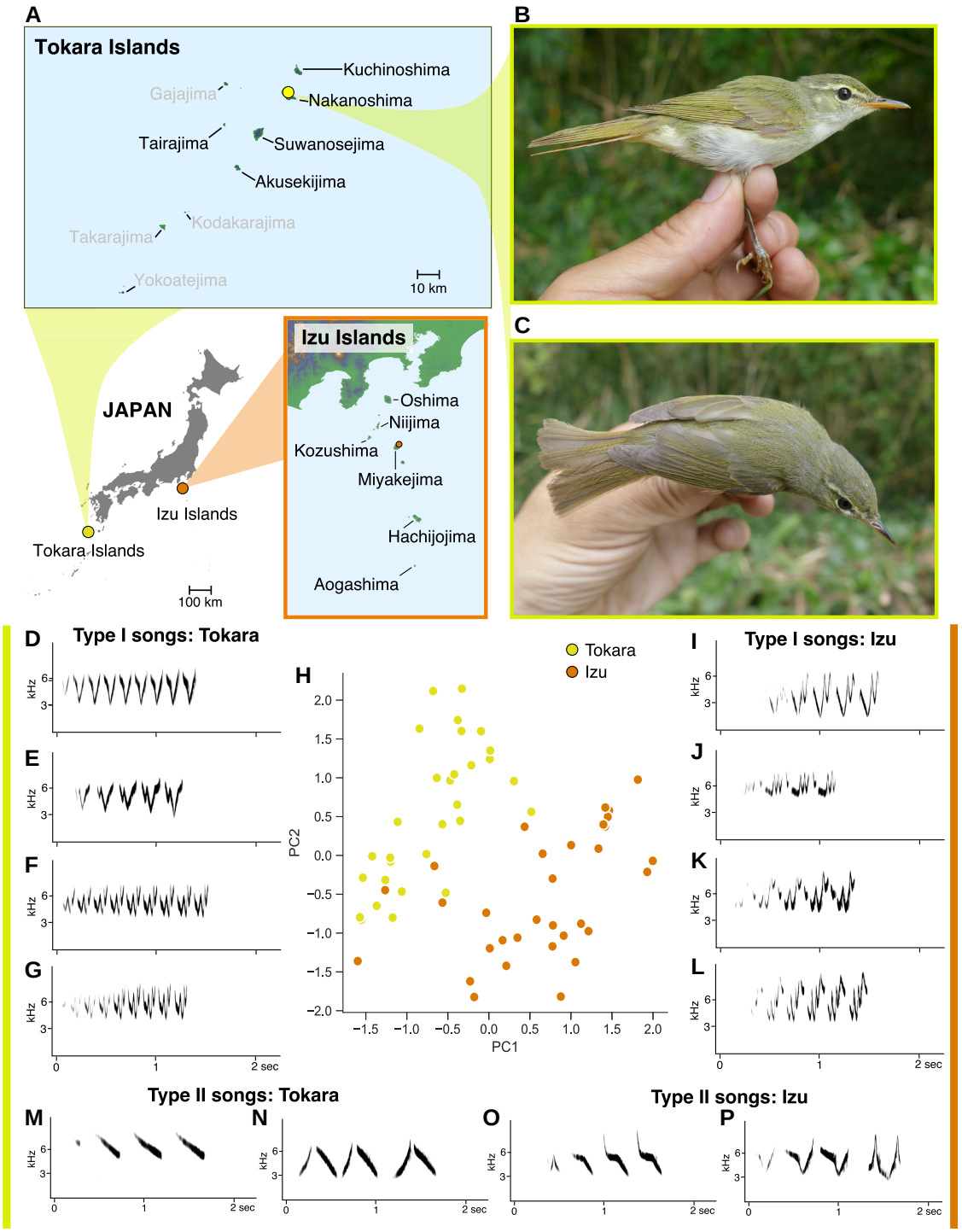

Fig. 1. Cryptic populations of *P. ijimae* are geographically separated by more than 1,000 km and have distinct songs. A) Global distribution of *P. ijimae*. Insets provide detailed maps of the Tokara Islands and northern Izu Islands (map source: Japan Aerospace Exploration Agency (2021). ALOS World 3D-30m DEM, V3.2, January 2021. Distributed by OpenTopography, https://doi.org/10.5069/G94M92HB. Accessed 2025 March 12). Island names in gray indicate locations where *P. ijimae* was not observed, while those in black indicate presence. B, C) Adult male, holotype of new species, Tokara Leaf Warbler (Yamashina Institute for Ornithology number YIO-76774), Nakanoshima, Tokara Islands, 2017 June 10 (photo: Per Alström; for additional photographs, see Table S18). Examples of single song strophes of type I songs from Tokara (D–G) and Izu (I–L), with a plot of the two PCs from a PCA based on 12 variables (H). Examples of single strophes of type II songs from Tokara (M and N) and Izu (O and P) (recordings: D: ML647192043; E: ML647191975; F: ML647192039; G: ML 647191951; M: 647192011; N: 647191965; all by P.A.; I: by T. Kabaya; J: ML647192103, by T.S.; K: XC749104; L: XC749102; O: XC749102), K, L, O by Geoff Carey; P: ML647356514, by Haruo Kuroda.

species. Notably, these include several sympatric pairs (indicated by asterisks), such as Sichuan Leaf Warbler *P. forresti* and Gansu Leaf Warbler *P. kansuensis*, which meet in a narrow contact zone in north central China with no evidence of interbreeding (20).

In the mitochondrial cytochrome b (cytb) tree of the entire leaf warbler family (Fig. 2C), the divergence between the Tokara and Izu populations (indicated by a dotted red line) was deeper than between more than 30 widely recognized species of leaf warblers,

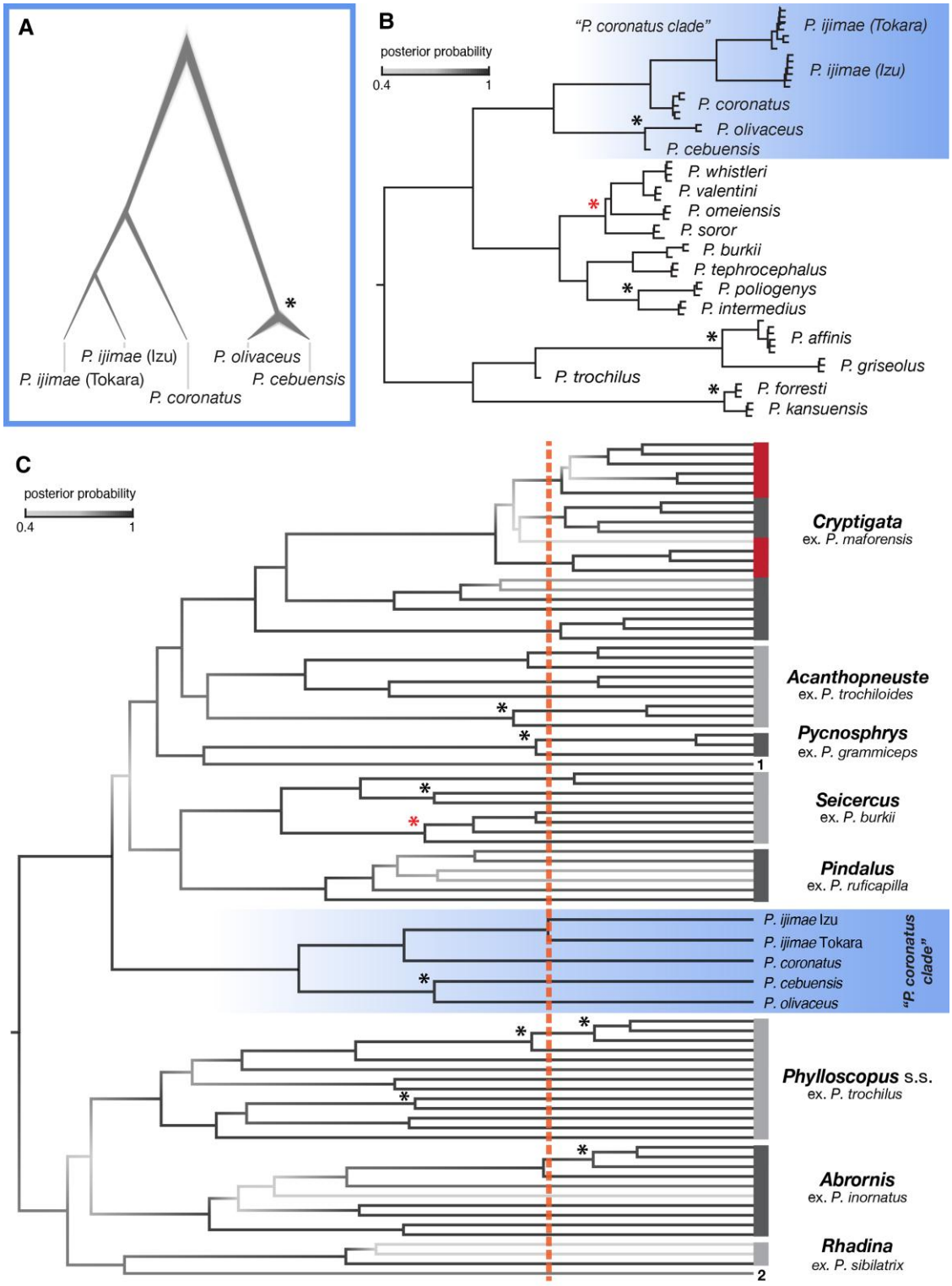

Fig. 2. Phylogenetic analyses suggest a deep divergence between the Tokara and Izu populations, on par with those between multiple closely related species of leaf warblers. A) Phylogenetic relationships and relative divergence within the "*P. coronatus* clade," including the two populations of *P. ijimae*, based on 15,830 genome-wide SNPs analyzed under the multispecies coalescent (SNAPPER) and including all samples. B) Phylogenetic relationships and relative divergence within the "*P. coronatus* clade" and some other *Phylloscopus* clades in which up to five species are sympatric, based on ~7,100 introns analyzed under the multispecies coalescent (ATPW). C) Phylogenetic relationships and relative divergence among all species in the family Phylloscopidae based on the mitochondrial cytb gene (1,041 base pairs), inferred in BEAST2. Posterior probabilities are indicated as shades of black, the West Pacific archipelago species are indicated by a red bar, and the vertical dotted line indicates the separation between the two *P. ijimae* populations: (1) *Phylloscopus emeiensis* and (2) *Phylloscopus neglectus*; all species names are shown in Fig. S7. In all trees, sympatric species with divergence times comparable to or younger than the two *P. ijimae* populations are indicated by asterisks; the red asterisk at the *Seicercus* clade indicates that three of the species are sympatric on the same mountain but mainly or entirely elevationally segregated. The names in C are subgeneric names, following Alström et al. (10), with one example of a species in each subgenus.

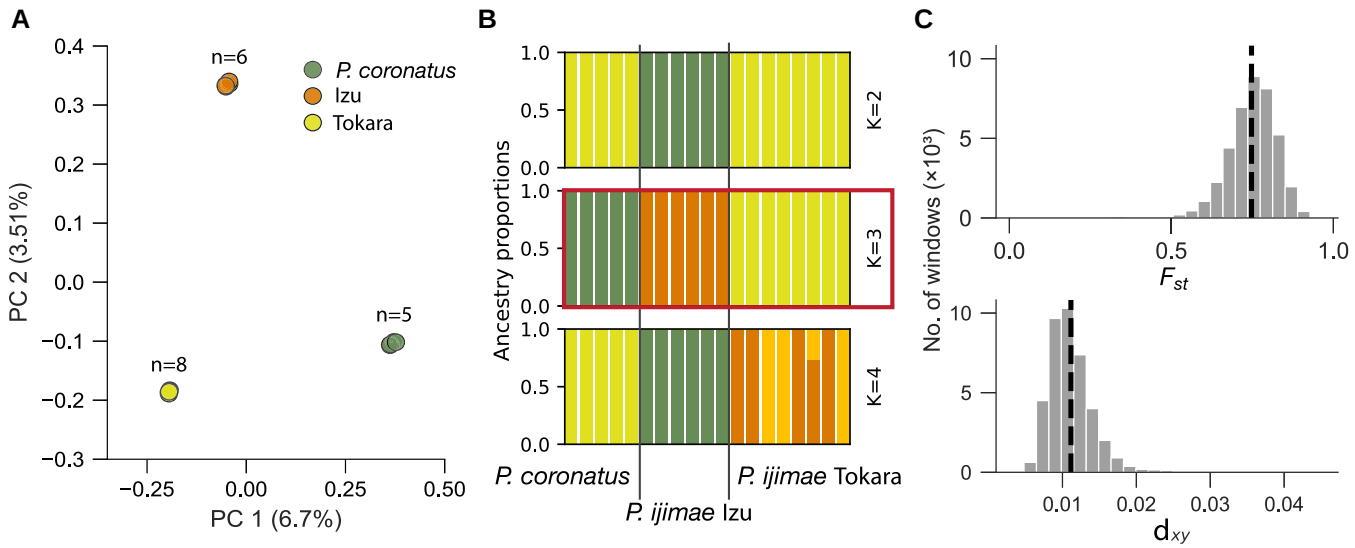

Fig. 3. Whole-genome analyses reveal distinct population structure and high divergence between *P. ijimae* populations. PCA (A) and ADMIXTURE (B) analysis of the two *P. ijimae* populations and their sister species *P. coronatus*. Color assignment on B is arbitrary. $K = 3$ had the lowest cross-validation error (highlighted by dark red rectangle, Fig. S12). C) Distributions of genomic differentiation (F_{ST}) and sequence divergence (d_{XY}) between Tokara and Izu populations. Dashed lines indicate mean values.

including 10 of the 11 recently radiated West Pacific island species (highlighted by red bars), and comparable to or deeper than between some pairs or groups of sympatric species (indicated by asterisks). A cytb tree including previously published sequences and sequences from birds on migration (Fig. S9) and a tree based on whole mitogenomes (Fig. S10) confirmed the presence of two deeply diverged lineages. Uncorrected p distances for cytb between the Tokara and Izu populations are 5.7–6.3% (vs. 0.00–0.68% within Nakanoshima and 0.00–0.39% within Izu). A rough estimate based on a cytb molecular clock rate of ~2.1%/million years (21) suggests a divergence time of ~2.8–3.2 million years ago.

The population genetic approaches suggested a complete lack of recent or contemporary gene flow between the two island populations: both the PCA (Fig. 3A) and ancestry analysis (ADMIXTURE; Fig. 3B) showed strong differentiation between the two *P. ijimae* populations. Genetic differentiation, as measured by fixation index F_{ST}, between the Tokara and Izu populations (mean 0.75 ± 0.08 SD; Fig. 3C, top, Fig. S11) was considerably higher than that observed between three marginally sympatric species pairs of leaf warblers, such as Greenish Warbler *Phylloscopus trochiloides* and Two-barred Warbler *Phylloscopus plumbeitarsus* in the famous "ring species complex" (Table S12). The observed F_{ST} values are high enough that the signal likely remains robust even when accounting for differences in F_{ST} estimation across studies, which may arise from variations in dataset filtering stringency, calculation methods, and normalization approaches (more details in Section S3.3). However, we note that the low genetic diversity and modest sample sizes in both *P. ijimae* populations may have also contributed to an elevated F_{ST}.

The absolute divergence metric d_{XY} between the Tokara and Izu populations (mean 0.01 ± 0.003 SD; Fig. 3C, bottom) was higher than at least some pairwise comparisons between closely related species; eg it was around three times higher than between the locally sympatric *P. trochiloides* and *P. plumbeitarsus* (0.0038221 on autosomes and 0.0032989 on the Z chromosome (22), and more than twice as high as that between the Collared Flycatcher *Ficedula albicollis* and the European Pied Flycatcher *Ficedula hypoleuca* (23), which are sympatric and completely reproductively isolated (24).

Tokara Leaf Warbler: a species new to science

As shown above, the Tokara population is well differentiated genetically from the Izu population, indicating a long-standing lack of gene flow between them, with divergence comparable to or exceeding multiple distinct species of leaf warblers. This is further supported by differences in morphometrics and, especially, vocalizations. Higuchi and Kawaji (15) reported that males on Nakanoshima (Tokara Islands) responded to playback of songs from Izu (with no further details on the number of birds tested, etc.), which might suggest that the differences in songs would be insufficient to act as a reproductive isolating barrier in case the two populations were to come into contact. Although this might be true, speculating about the outcome of a potential—in our opinion unlikely—future secondary contact is futile (see Section S3.2 for further comments). Cytb sequences from birds sampled during autumn migration in early September at Kagoshima, southern Kyushu, are in the same clade as samples from Izu (Fig. S9), suggesting that the Izu Islands population migrates southwest along the coast of the main Japanese islands in autumn and then likely follows the Nansei Shoto island chain south toward its winter quarters in the Philippines. Accordingly, provided the same route is taken in spring, these migrants have the opportunity to stop and breed on the Tokara Islands, but in spite of this, there is no evidence of gene flow between Tokara and Izu, probably because of very strong innate philopatry (as is known from multiple birds; (25)). It also seems possible that the Tokara Islands were colonized from the Izu Islands (after prior colonization of the latter islands from a most recent common ancestor between *P. ijimae* and *P. coronatus* on the main Japanese islands), as suggested by the lower *Ne*, π, and more simple song on Tokara than on Izu. An alternative hypothesis is that the most recent common ancestor to the two populations was more widely distributed across Japan, but then disappeared from the larger islands—perhaps due to competition with the closely

related *P. coronatus*, which occurs on all of the large Japanese islands (26)—leaving geographically widely separated populations on the Izu and Tokara Islands. The deep divergence between these two populations may seem to favor this latter hypothesis.

Based on the above, we conclude that the two Japanese island populations fulfill the requirements for being treated as separate species (under the "biological" (27), "phylogenetic" (28), as well as "unified" (29, 30) species concepts). We describe the Tokara population as a species new to science (Section S4; Fig. 1B, C), which we in the following refer to as the Tokara Leaf Warbler (and the Izu population as Ijima's Leaf Warbler). Further information on, eg morphological variation, habitat, and breeding of the Tokara Leaf Warbler is presented in Section S4.

Conservation of Tokara Leaf Warbler and Ijima's Leaf Warbler: from monitoring to population genomics

By recognizing the Tokara Leaf Warbler as a new species, distinct from the already Vulnerable (VU, (18)), Ijima's Leaf Warbler, we highlight the need for an independent evaluation of its conservation status and provide genomic and ecological data to inform that process.

Although singing males have been previously observed on five islands in the Tokara archipelago (Fig. 1A), breeding has only been confirmed on Nakanoshima, and a recent study failed to find the species on one of these islands, Tairajima (17). Additionally, threats from invasive species, involving predation by introduced weasels *Mustela itatsi* and environmental modification by domestic goats, have been identified (17).

These findings underscore the importance of genome-wide data to understand demographic history and evaluate the evolutionary resilience of the Tokara Leaf Warbler.

Our analysis of nucleotide diversity indicated lower diversity in the Tokara Leaf Warbler (mean $\pi = 0.0023 \pm 0.001$ SD) and Ijima's Leaf Warbler (mean $\pi = 0.0034 \pm 0.003$ SD) compared with other *Phylloscopus* species (eg Common Chiffchaff *Phylloscopus collybita*, (31)), with the Tokara population showing significantly reduced diversity (Fig. 4A). Interestingly, the nucleotide diversity observed in Tokara Leaf Warbler was comparable to levels found in endangered bird species prior to significant population declines (eg Seychelles Paradise-Flycatcher *Terpsiphone corvina* (32)), although they were still substantially higher than ones observed in some critically endangered species (33, 34). Genome-wide nucleotide diversity was uniformly low across the genomes of both species, and we found no clear evidence of further localized reductions that would indicate recent selective sweeps or regions of low recombination (Fig. S11). This, together with a largely homogeneous landscape of F_{ST} (Fig. S11), supports a scenario in which demographic processes and historical allopatry—rather than strong divergent selection—have primarily shaped the observed genomic divergence.

The genetic diversity observed in both populations suggests a history of population bottlenecks. Comparisons of Tajima's *D* with their sister species, *P. coronatus*, which occurs throughout the main Japanese islands (used as a baseline; Fig. 4B), suggest divergent demographic histories: *P. coronatus* and Ijima's Leaf Warbler showed values close to neutral equilibrium. Tajima's *D* values for the latter are slightly negative, yet insufficiently different from zero to confirm a population growth or effect of selection. The Tokara Leaf Warbler displayed higher values, indicative of a past bottleneck. Site frequency spectrum analysis revealed that Ijima's Leaf Warbler has accumulated more rare alleles than the Tokara Leaf Warbler, potentially due to a faster recovery in the recent past (Fig. S13). Furthermore, analysis of recent changes in effective population size using stairway plots (Fig. S14) confirmed consistent historical population size decline in both species, with an earlier onset of decline in the Izu Leaf Warbler.

A moderate proportion of the genome was in a homozygous state, as shown by runs of homozygosity (ROH) analysis (Fig. 4C). Short and intermediate ROH classes (<100 kb and <1 Mb, respectively) were consistently more frequent in Tokara Leaf Warbler than in Ijima's Leaf Warbler, reflecting earlier inbreeding in the former. Despite reduced genetic diversity and evidence of bottlenecks, both populations showed signs of recovery and purging, as estimates of genetic load revealed a low proportion of high-impact (potentially deleterious) variants in both realized (~0.01%) and masked (0.03%) load (Fig. 4D and E).

Based on recent population size estimates (Fig. S14) and the conservation genetics statistics obtained here, and following the IUCN Red List criteria (35), it seems appropriate to place the Tokara Leaf Warbler in the same threat category, ie vulnerable (VU), as the one where *P. ijimae* (sensu lato) is placed (18). However, this could change quickly, as a small population, geographically restricted species is highly vulnerable to, eg epidemics, habitat change as a result of volcanic eruptions or other causes, and invasive species, not only on the breeding grounds but also on its wintering grounds and along its migration routes.

Our study shows that genomics can expose hidden biodiversity and guide urgent action to recognize and protect cryptic species in a rapidly changing world.

Materials and methods

Morphology

Five external measurements were taken in the field on 35 adult males from the Tokara Islands and 15 adult males from the Izu Islands, Abiko, Japan (Table S13). Mean values were compared using Mac Statistical Analysis ver. 2.0 (Esumi, Tokyo, Japan), and PCA was run using Mac EXCEL Multivariate Analysis ver. 1.0b (Esumi).

Vocalizations

Sound recordings were obtained from public databases and from our own collections. Songs from 33 individuals from the Izu Islands (Miyakejima, one from Hachijojima) and 29 individuals from the Tokara Islands (Nakanoshima) were selected for analysis (Table S18). Sonograms were created in Raven Pro 1.6.5 (Cornell Lab of Ornithology, USA), and eight variables were measured and four additional ones counted. The data were analyzed in SPSS 21.0 (IBM Corp., USA) using PCA and DFA to determine whether the sounds from different geographical areas could be distinguished. See Section S2.3 for details.

Sample collection and sequencing

DNA samples (pectoral muscle or blood) were collected and stored in 95% ethanol in the lab at −20 °C (Table S14). Total genomic DNA was extracted using the DNeasy Blood & Tissue Kit (QIAGEN) following the manufacturer's instructions. DNA libraries with an average insert size of ~350 bp were prepared and sequenced on the Illumina NovaSeq 6000 (paired-end, 150 bp reads) at SciLife Lab (Stockholm, Sweden). A few cytb sequences were also obtained by Sanger sequencing, as described in Olsson et al. (36).

Variant calling and filtering

Variant calling for the main dataset (dataset #1) was conducted for 14 individuals representing two island populations (Table S14). Raw reads were quality filtered with Trimmomatic (37) and aligned to

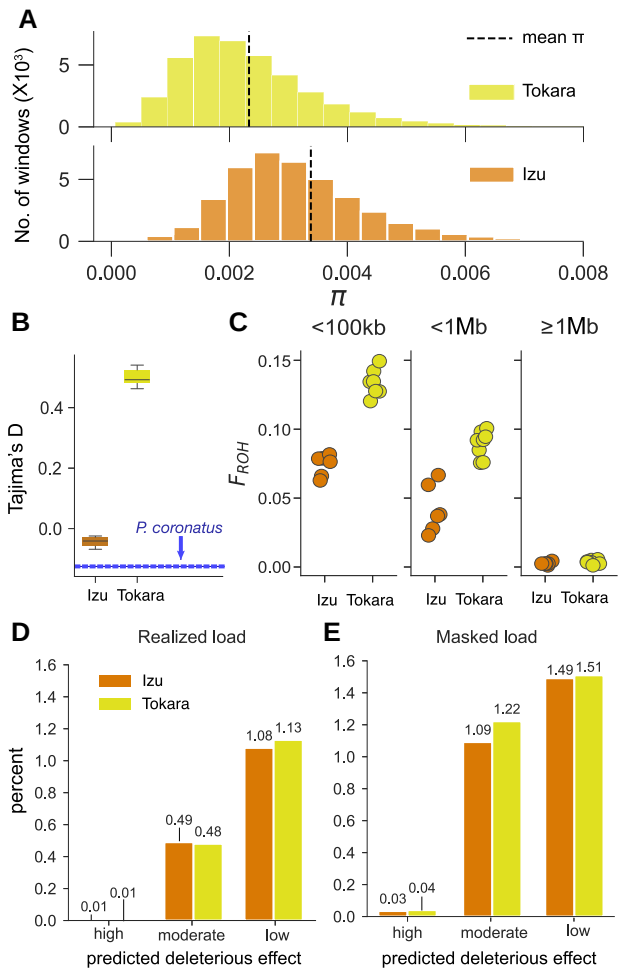

Fig. 4. Genomic diversity, inbreeding, and deleterious mutation load in two island populations of *P. ijimae*. A) Distribution of nucleotide diversity (π) across 10 kb windows; vertical dashed lines indicate the population means. B) Tajima's *D* values for the two island population and a baseline value for the sister species *P. coronatus* ($n = 5$; indicated by a blue horizontal dashed line), which occurs on the main Japanese islands. Since Tajima's *D* is sensitive to sample size, we generated all possible combinations of five individuals and summarized variation in Tajima's *D* estimates in boxplots. C) Inbreeding coefficients based on ROH (F_{ROH}) across three ROH length classes. D) Realized genetic load, estimated as the proportion of homozygous variants in three effect categories. E) Masked (heterozygous) load.

the *Phylloscopus trochilus* reference genome (38) using BWA (39); mapping statistics are provided in Table S15. For each individual in the main dataset (dataset #1), variants were called using the HaplotypeCaller tool from GATK to produce gVCF (genomic VCF) files. Details on SNP calling and filtering are provided in Section S2.4.1. The resulting file covered 864,824,543 bp, of which 54,196,166 were high-quality variant calls.

For further examination of individuals from sister and outgroup species, we utilized a rapid SNP calling approach within the enhanced Sarek Nextflow pipeline (40) (dataset #2). All samples passed intermediate mapping quality control (Table S16). The obtained set of markers (414,770 SNPs) was filtered according to the current best practices and described below in Section S2.4.1.

Phylogenetic analysis

We used previously published cytb sequences for all 71 *Phylloscopus* species (10) and added an individual from Tokara (one from Izu was already available). Details on alignment and BEAST2 specification can be found in Section S2.4.4. The same tree as in Fig. 2C, with all species names shown, is presented in Fig. S7. Additionally, we analyzed a dataset including all our cytb sequences and all those available on NCBI (in total 35 individuals) and two outgroup species (*P. coronatus* and *P. olivaceus*), with settings and methods as above (Fig. S9). Finally, whole mitochondrial consensus sequences were extracted from earlier obtained BAM files with *samtools consensus* (minimum depth 7, including indels, coordinates CM046213.1:1–16942). Consensus genomes were aligned in MAFFT (FFT-NS-2), and phylogenies were inferred with BEAST2 (Fig. S10 and Section S2.4.4).

For whole-genome phylogenetic reconstruction, we utilized dataset #2 and employed SNAPPER, a Bayesian method implemented in BEAST2 for inferring species trees and demographic history from unlinked binary markers using diffusion models of allele-frequency dynamics (41) (see extended methods section in Section S2.4.2). We used BEAUti to prepare the XML input file, assuming a Yule speciation model and incorporating a gamma distribution to model rate heterogeneity.

We used the BirdScanner pipeline (42) (github.com/Naturhistoriska/birdscanner, accessed 2025 June 3) to extract sequence homologs of intronic loci from the FASTA sequences compiled for each individual sample. The FASTA files of the resulting parsed-out matrices were processed with ATPW (Align-and-Trees-Parallel-Workflow; (43)), which estimates a species tree using the weighted ASTRAL algorithm implemented in ASTER (44).

Population genetic summary statistics

Accurate estimation of species divergence and diversity relies on high-quality information on both SNP and invariant sites. For this in-depth population genetic analysis, we used dataset #1 (*P. ijimae*–specific variant and invariant sites) and calculated population summary statistics using a high-accuracy approach implemented in pixy (45) software. We used this high-quality dataset to quantify the number of fixed polymorphisms between two populations of *P. ijimae*.

To investigate genetic variation between closely related species, we used data set #2 (SNP set with outgroup species) and performed PCA using PLINK v1.90 b 4.9 (46). For this analysis, we additionally filtered out singleton SNP sites and SNPs that were in linkage (ie SNPs with r^2 values <0.1 using sliding windows of 50 SNPs with a step size of 10 SNPs). PLINK preferentially keeps variants with higher nonmajor allele frequency. Genetic structure was also assessed with the likelihood method ADMIXTURE (47). Each number of population clusters (K) from 1 to 6 was run independently, after which cross-validation errors calculated using the cross-entropy criterion were compared and visualized in Fig. S12.

Conservation genomics statistics

ROH, defined as continuous homozygous segments indicative of autozygosity, were identified using BCFtools/RoH (48) with a 30-generation threshold (-G30) and allele-frequency information from the VCF (–AF-tag AF). We filtered variants to retain only those with Phred quality scores >30. ROH analyses focused on segments ≥1 Mb and ≥100 SNPs to capture signatures of recent inbreeding.

To estimate deleterious mutation load, we annotated SNP variants generated above using SnpEff (49). We created a custom SnpEff database using annotation for the *P. trochilus* reference

genome (38), previously used for mapping. For each population we extracted SNPs that were homozygous (these variants are fixed and represent realized load) and heterozygous (masked load). Next, we classified each SNP set by predicting its functional impact using the standard settings in SnpEff. SnpEff reports effect impact for every cite/SNP, including HIGH (such as loss of start codon, gain of stop codon, and exon loss), MODERATE (eg codon insertion and nonsynonymous variants), and LOW (eg synonymous variants, switch between alternative start codons, and variants in the intron) effect categories, which we summarized using custom Python scripts.

We also assessed population demographic history using Tajima's D (50), calculated in nonoverlapping 1 Mb windows across the genome with VCFtools (51), to detect deviations from neutrality associated with population contractions or expansions. Tajima's D compares nucleotide diversity (π) to Watterson's θ from the number of segregating sites to test whether the allele-frequency spectrum deviates from neutral equilibrium expectations.

We inferred historical changes in effective population size (N_e using Stairway Plot v2 (52)), which models population size trajectories from folded site frequency spectra (SFS). For each population, we used easySFS github.com/isaacovercast/easySFS, accessed 2024 May 3) to generate folded SFS from filtered VCF files, projecting to 12 and 16 chromosomes, respectively. Input SFS were generated from 848 Mb of callable genome and included both polymorphic and invariant sites. The Stairway plot was run using multiple random breakpoints (Tokara: 2, 5, 7, 10; Izu: 7, 15, 22, 28), with 200 bootstrap replicates per configuration. A mutation rate of 3.3×10^{-9} per site per generation and a generation time of 1 year were assumed, based on estimates from related passerines (53). Output files were plotted across a time range of 0.1–400,000 years with default N_e axis scaling.

Acknowledgments

The authors are most grateful to the Institute of Zoology, Chinese Academy of Sciences, Beijing; University of Washington Burke Museum, Seattle; Zoological Institute and Museum, University of Greifswald; and Field Museum of Natural History, Chicago, for DNA samples; and the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation, the Swedish Research Council, and the SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure (to P.A.). The authors are also most grateful to Yoshimitsu Shigeta, Yutaka Yamamoto, and Shigeki Asai for DNA samples; to Yasuko Iwami for preparing the specimens; and to Haruo Kuroda for providing sound recordings. They are grateful to Staffan Bensch, Frank Rheindt, and two anonymous reviewers for comments on the manuscript and to David J. Weber for archiving their media at the Macaulay Library, Cornell Lab of Ornithology.

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

This work was supported by the Swedish Research Council Vetenskapsrådet (grant no. 2019-04486), Jornvall Foundation, Carl Trygger Foundation (CTS 20-6), and Olle Engkvists Stiftelse, all to P.A.; Suntory Fund for Bird Conservation and Scientific Research (specially designated research promotion) from the Ministry of Education, Culture, Sports, Science and Technology in Japan to T.S.

Author Contributions

Takema Saitoh (Conceptualization, Resources, Data curation, Formal analysis, Funding acquisition, Validation, Investigation, Visualization, Writing—original draft, Project administration, Writing—review & editing), Daria Shipilina (Conceptualization, Resources, Data curation, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing—original draft, Writing—review & editing), Canwei Xia (Data curation, Formal analysis, Investigation, Visualization, Writing—original draft, Writing—review & editing), Lijun Zhang (Data curation, Investigation, Writing—original draft), Shin-Ichi Seki (Resources, Validation, Investigation, Writing—original draft, Writing—review & editing), Urban Olsson (Conceptualization, Data curation, Formal analysis, Investigation, Visualization, Writing—original draft, Writing—review & editing), and Per Alström (Conceptualization, Resources, Formal analysis, Supervision, Funding acquisition, Validation, Investigation, Methodology, Writing—original draft, Project administration, Writing—review & editing)

Data Availability

All data and code supporting this study will be publicly available upon acceptance. Raw sequencing reads are deposited in the European Nucleotide Archive (ENA) under project accession PRJEB105759. Analysis scripts and workflows are available on the GitHub repository https://github.com/DaSh-bash/Phylloscopus_ijimae. Photographs of live birds and bioacoustic recordings are archived in the Macaulay Library (Cornell Lab of Ornithology, Cornell University; Table S18). DNA samples and specimens were collected under Ministry of the Environment Government of Japan permits no. 10-0217 and no. 1703012 and Agency for Cultural Affairs Government of Japan permit no. 4-1364. The collection of samples in 2020 was approved by the Yamashina Institute for Ornithology's Animal Research Ethics Committee (formed in 2019), no. 2020-001 (all to T.S.).

References

1 Wiens JJ. 2023. How many species are there on earth? Progress and problems. *PLoS Biol*. 21:e3002388.
2 Li X, Wiens JJ. 2023. Estimating global biodiversity: the role of cryptic insect species. *Syst Biol*. 72:391–403.
3 Allendorf FW, Hohenlohe PA, Luikart G. 2010. Genomics and the future of conservation genetics. *Nat Rev Genet*. 11:697–709.
4 Supple MA, Shapiro B. 2018. Conservation of biodiversity in the genomics era. *Genome Biol*. 19:131.
5 BirdLife International. *State of the World's Birds 2022: insights and solutions for the biodiversity crisis*. BirdLife International, Cambridge, UK, 2022.
6 AviList Core Team. 2025. AviList: The global avian checklist, v2025. doi:10.2173/avilist.v2025
7 Rheindt FE, *et al.* 2025. AviList: a unified global bird checklist. *Biodivers Conserv*. 34:3359–3376.
8 Mayr E, Amadon D. 1951. A classification of recent birds. *Am Mus Novit*. 1496:1–42.

9 Barrowclough GF, Cracraft J, Klicka J, Zink RM. 2016. How many kinds of birds are there and why does it matter? *PLoS One*. 11: e0166307.

10 Alström P, *et al.* 2018. Complete species-level phylogeny of the leaf warbler (Aves: Phylloscopidae) radiation. *Mol Phylogenet Evol*. 126:141–152.

11 Bairlein F, *et al.* Family sylviidae (warblers). In: del Hoyo J, Elliott A, Christie DA, editors. *Handbook of the birds of the world, vol. 12*. Lynx Edicions, Barcelona, 2006. p. 492–709.

12 Ng NSR, *et al.* 2018. A striking new species of leaf warbler from the Lesser Sundas as uncovered through morphology and genomics. *Sci Rep*. 8:15646.

13 Rheindt FE, *et al.* 2020. A lost world in Wallacea: description of a montane archipelagic avifauna. *Science*. 367:167–170.

14 Johnson TH, Stattersfield AJ. 1990. A global review of island endemic birds. *Ibis (Lond 1859)*. 132:167–180.

15 Higuchi H, Kawaji N. 1989. Ijima's willow warbler *Phylloscopus ijimae* of the Tokara Islands, a new breeding locality, in southwest Japan. *Bull Biogeogr Soc Jpn*. 44:11–15.

16 Seki S-I, Tokorozaki S, Mizoguchi F, Takagi S, Nakamura N, Crystal F. 2011. Avifauna of the Tokara Islands, northern Ryukyu Archipelago. *Bull FFPRI*. 10:183–229 (In Japanese).

17 Seki S-I. 2024. Breeding distribution of the isolated Iijima's leaf warbler population in the Tokara Islands and its ecological, morphological, and genetic characteristics. *J Yamashina Inst Ornithol*. 56:33–50 (in Japanese).

18 IUCN. 2023. The IUCN Red List of Threatened Species, version 2023-1. 2025, September 1. https://www.iucnredlist.org.

19 Agency for Cultural Affairs, Government of Japan. 2024. Database of cultural properties. 2025, September 1. https://kunishitei.bunka.go.jp/bsys/index.

20 Wu L, *et al.* 2023. Limited song-mixing without recent gene flow in a contact zone between two songbird species. *Mol Biol Evol*. 40:msad053. https://doi.org/10.1093/molbev/msad053

21 Weir JT, Schluter D. 2008. Calibrating the avian molecular clock. *Mol Ecol*. 17:2321–2328.

22 Irwin DE, Alcaide M, Delmore KE, Irwin JH, Owens GL. 2016. Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. *Mol Ecol*. 25:4488–4507.

23 Chase MA, Ellegren H, Mugal CF. 2021. Positive selection plays a major role in shaping signatures of differentiation across the genomic landscape of two independent Ficedula flycatcher species pairs. *Evolution*. 75:2179–2196.

24 Ålund M, *et al.* 2024. Tracking hybrid viability across life stages in a natural avian contact zone. *Evolution*. 78:267–283.

25 Paradis E, Baillie SR, Sutherland WJ, Gregory RD. 1998. Patterns of natal and breeding dispersal in birds. *J Anim Ecol*. 67:518–536.

26 *The Ornithological Society of Japan*. 2012. *Checklist of Japanese birds*. 7th revised ed. Ornithological Society of Japan, Sanda.

27 Mayr E. *Animal Species and evolution*. Belknap Press, Cambridge, MA, 1963.

28 Cracraft J. 1989. Speciation and its ontology: the empirical consequences of alternative species concepts. In: Otte D, Endler JA, editors. *Speciation and its consequences*. Sinauer, Sunderland, MA, pp. 28–59.

29 de Queiroz K. 2005. A unified concept of species and its consequences for the future of taxonomy. *Proc Calif Acad Sci*. 56: 196–215.

30 de Queiroz K. 2007. Species concepts and species delimitation. *Syst Biol*. 56:879–887.

31 Talla V, Kalsoom F, Shipilina D, Marova I, Backström N. 2017. Heterogeneous patterns of genetic diversity and differentiation in European and Siberian chiffchaff (*Phylloscopus collybita abietinus/P. tristis*). *G3 (Bethesda)*. 7:3983–3998.

32 Femerling G, *et al.* 2023. Genetic load and adaptive potential of a recovered avian species that narrowly avoided extinction. *Mol Biol Evol*. 40:msad256.

33 Guhlin JG, *et al.* 2025. Population genomics of yellow-eyed penguins uncovers subspecies divergence and candidate genes linked to respiratory distress syndrome. bioRxiv 683354. https://doi.org/10.1101/2025.10.20.683354, preprint: not peer reviewed.

34 Matthews TJ, *et al.* 2024. The global loss of avian functional and phylogenetic diversity from anthropogenic extinctions. *Science*. 386:55–60.

35 IUCN Standards and Petitions Committee. 2024. Guidelines for Using the IUCN Red List Categories and Criteria. Version 16. Prepared by the Standards and Petitions Committee. Downloadable from https://www.iucnredlist.org/documents/RedListGuidelines.pdf.

36 Olsson U, Alström P, Ericson PGP, Sundberg P. 2005. Non-monophyletic taxa and cryptic species: evidence from a molecular phylogeny of leaf-warblers (Phylloscopus, Aves). *Mol Phylogenet Evol*. 36:261–276.

37 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*. 30:2114–2120.

38 Lundberg M, *et al.* 2023. Extreme genomic differentiation driven by a chromosomal inversion in willow warblers. *Nat Commun*. 14:5162.

39 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 25:1754–1760.

40 Garcia M, *et al.* 2020. Sarek: a portable workflow for WGS analysis of germline and somatic variants. *F1000Res*. 9:63.

41 Stoltz M, *et al.* 2021. Bayesian inference of species trees using diffusion models. *Syst Biol*. 70:145–161.

42 Ericson PGP, *et al.* 2020. Parallel evolution of bower-building behavior in two groups of bowerbirds suggested by phylogenomics. *Syst Biol*. 69:820–829.

43 Nylander JAA. 2022. ATPW—Align and trees parallel workflow. https://github.com/nylander/Align-and-trees-parallel-workflow.

44 Zhang C, Mirarab S. 2022. ASTRAL-Pro: quartet-based species-tree inference despite paralogy. *Mol Biol Evol*. 39: msab316.

45 Korunes KL, Samuk K. 2021. Pixy: unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol Ecol Resour*. 21:1359–1368.

46 Chang CC, *et al.* 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 4:7.

47 Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 19: 1655–1664.

48 Narasimhan V, *et al.* 2016. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*. 32:1749–1751.

49 Cingolani P, *et al.* 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)*. 6:80–92.

50 Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 123:585–595.

51 Danecek P, *et al.* 2011. The variant call format and VCFtools. *Bioinformatics*. 27:2156–2158.

52 Liu X, Fu Y-X. 2020. Stairway plot 2: demographic history inference with folded site frequency spectra. *Genome Biol*. 21:280.

53 Zhang GJ, *et al.* 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*. 346: 1311–1320.