**Supplementary Materials**

**Phylogenomics Reveals that Mitochondrial Introgression with Nuclear Introgression Characterizes Skua Species Proposed to be of Hybrid Origin**

Else K. Mikkelsen & Jason T. Weir

Table of Contents

[Supplementary Methods and Results 4](#_Toc115133498)

[Sequencing and Alignment 4](#_Toc115133499)

[**Table S1.** Samples sequenced for this study. Samples were obtained from the Royal Ontario Museum in Toronto, Canada. Sex was identified based on the heterozygosity and relative coverage of the Z chromosome. 5](#_Toc115133500)

[**Table S2.** Sequencing statistics for samples sequenced in this study. %GC measures the proportion of G and C nucleotides in the trimmed sequencing data. Heterozygosity is estimated by Rohan (Renaud et al. 2019) on autosomes. Divergence is measured as Kimura 2-parameter distances, averaged across autosomal 5-kb windows. 6](#_Toc115133501)

[**Table S3.** Sequence mapping statistics for samples sequenced in this study. 7](#_Toc115133502)

[**Sequence heterozygosity** 8](#_Toc115133503)

[**Figure S1. Heterozygosity across chromosome Z.** Heterozygosity was estimated by Rohan as the proportion of sites which are heterozygous within 1 Mb windows in each sample. The nine female samples show low estimates of heterozygosity across most of chrZ (which is in hemizygous state in females), but with a peak of very high heterozygosity from 0-10 Mb. This region is likely affected by paralogous W chromosome sequences mapping to the Z chromosome, and so was excluded from downstream analyses. 9](#_Toc115133504)

[**Figure S2. Heterozygosity across the genome.** Heterozygosity is estimated by Rohan as the proportion of sites which are heterozygous within 1 Mb windows in each sample. The autosomes are displayed in order of descending length, and the Z chromosome is plotted last. Across the genome, the three jaeger species (S. longicaudus, S. pomarinus, and S. parasiticus) have moderate to relatively high heterozygosity, while the four “Catharacta” skuas (S. skua, S. antarcticus, S. maccormicki, and S. chilensis) have markedly lower heterozygosity. 10](#_Toc115133505)

[Genotype calling pipeline 10](#_Toc115133506)

[Species Phylogeny 10](#_Toc115133507)

[Astral-III phylogeny 10](#_Toc115133508)

[**Figure S3.** Densitree visualization of the maximum likelihood gene trees used in the Astral-III species tree reconstruction. One percent of the gene trees were randomly subsampled and plotted as overlapping trees to show variation in topology and branch lengths between loci. For visualization purposes, gene trees were midpoint rooted using the phangorn package v2.8.1 (Schliep 2011) in R v4.1.2 (R Core Team 2018), scaled to a root age of 36.6, and converted to an ultrametric tree using chronos within the ape package v5.6.2 (Paradis and Schliep 2019). Incomplete lineage sorting and gene tree discordance is prevalent at the shallowest nodes. 11](#_Toc115133509)

[**Figure S4.** Astral-III phylogenies estimated using58,747autosomal loci (left) or754 Z chromosome loci (right) in a multispecies coalescent framework. Branch lengths are measured in coalescent units and reflect the amount of gene tree concordance at a node. Dashed lines indicate branches whose lengths are not estimated, as terminal branch lengths can only be estimated by Astral for taxa with more than one sample. Nodes are labelled with local posterior probabilities based on quartet frequencies, followed by normalized quartet scores, which reflect the proportion of gene trees that are concordant with the species topology at that node. 12](#_Toc115133510)

[**Table S4**. Internal branch lengths inferred by Astral-III and the expected and observed frequencies of discordant gene tree topologies under incomplete lineage sorting based on these estimated internal branch lengths. Branch lengths are measured in coalescent units at the internal branch in between the divergence of P1 from P2 and the divergence of (P1,P2) from P3. Observed proportions are the normalized quartet score of the discordant topology, which indicates the proportion of gene trees that support a topology in which P3 is sister to either P1 or P2. Under a scenario of incomplete lineage sorting in the absence of introgression, the frequencies of the two discordant topologies are expected to be equal, but introgression or ancestral population structure can result in an imbalance between the two frequencies. Note that since Astral-III branch length estimates are based on the observed frequencies of discordant gene trees, the expected and observed frequencies are not independent. “Catharacta” refers to the clade of S. skua, S. antarcticus, S. chilensis, and S. maccormicki. 12](#_Toc115133511)

[Starbeast3 Phylogeny 13](#_Toc115133512)

[**Figure S5.** Time-calibrated species tree inferred by Starbeast3. Node bars indicate the 95% HPD for the age of each node. All nodes have full posterior probability except for the shallowest two nodes (labelled). Circles next to the species images represent the relative effective population sizes estimated for each species, with the area of the circle proportional to effective population size. The phylogeny is magnified on the right to show greater detail at the shallowest nodes. 14](#_Toc115133513)

[mtDNA Assembly 14](#_Toc115133514)

[Mitochondrial phylogeny 15](#_Toc115133515)

[**Figure S6.** Maximum likelihood phylogeny of the whole mitogenomes of the Stercorariidae. The phylogeny is rooted with three species of Alcidae, and branch support values are based on 1000 bootstrap replicates. One copy of the duplicated region was removed prior to analysis, as the two copies showed evidence for concerted evolution and thus do not represent independently evolving sequence. 16](#_Toc115133516)

[**Figure S7.** Time-calibrated mitochondrial phylogeny produced by BEAST. Node bars show the 95% HPD of node ages. All nodes have full posterior probability, except for two nodes (labelled). The phylogeny is magnified and reflected on the right side to show detail at the shallower nodes. 17](#_Toc115133517)

[**Figure S8.** Histogram ofrelativedivergence between S. pomarinus and the four “Catharacta” skuas (S. chilensis, S. skua, S. antarcticus, and S. maccormicki) within 16.7-kb windows of the nuclear genome. Since mitochondrial sequences diverge at a much higher rate than nuclear sequences in birds, all genetic distances are standardized by dividing each Kimura two-parameter distance estimate by the genetic distance between S. pomarinus and S. parasiticus within the same window. The distributions are plotted as 30 separate, but highly overlapping distributions for each combination of the three S. pomarinus and ten “Catharacta” skuas. The point estimates for the relative mitochondrial divergence between S. pomarinus and the southern hemisphere skuas (S. maccormicki, S. antarcticus, and S. chilensis) are drawn as black lines, and the relative mitochondrial divergences between the S. pomarinus and S. skua samples are drawn as red lines. Only 0.44% of the nuclear windows fall equal to or lower than the average of the red lines, while 21.1% of windows are equal to or lower than the average of the black lines. 18](#_Toc115133518)

[Tests of Introgression 18](#_Toc115133519)

[**Phylonet** 18](#_Toc115133520)

[**Figure S9.** Maximum pseudolikelihood network returned by Phylonet. Introgression events are shown as red arrows. Internal branch lengths are estimated in coalescent units, but terminal branch lengths were not estimated, and are drawn as dotted lines of arbitrary length. 19](#_Toc115133521)

[**ABBA BABA tests** 19](#_Toc115133522)

[**Figure S10.** Visualization of estimates for D and fb for each pair of non-sister taxa. Estimates for D are shown above the diagonal, colour coded from low (blue) to high (red), with the colour saturation corresponding to the p-value (legend in bottom right). Estimates for fb are shown below the diagonal, ranging from low (white) to high (red). 20](#_Toc115133523)

[**Table S5.** D (ABBA-BABA test) and D3 statistics calculated for all trios. Raw p-values for each ABBA-BABA test are provided without correcting for multiple testing. D3 was averaged across 5-kb windows of the genome, and the proportion of windows for which D3 was negative is provided. A significantly positive value of D and a negative value of D3 are consistent with introgression between P2 and P3. Statistics are provided for all trios that are consistent with the estimated species tree, as well as the two alternate topologies at the node containing S. antarcticus, S. maccormicki, and S. chilensis where gene tree discordance is very high. 20](#_Toc115133524)

[**D3** 22](#_Toc115133525)

[**Table S6.** Genetic distances between the samples. Distances were calculated using the Kimura two-parameter correction for multiple substitutions, and are presented as percentages. Mitochondrial sequence divergence is given above the diagonal, and autosomal nuclear divergence is given below the diagonal. Autosomal heterozygosity estimated by Rohan is given on the diagonal. Cells are shaded according to their value from low (blue) to high (red), with separate scales for the mitochondrial and nuclear estimates. Mitogenome distances are calculated while including only one copy of the duplicated control region. 23](#_Toc115133526)

[**Figure S11.** Genetic distances across the genomes of S. parasiticus, S. pomarinus, S. skua, and S. chilensis are inconsistent with a hybrid speciation scenario. S. parasiticus shows similar levels of genetic divergence to all three taxa across the genome (green, blue, yellow; overlapping lines), with no indication of lower divergence with S. pomarinus (green). Similarly, S. pomarinus shows nearly identical genetic distances to both S. skua and S. chilensis (pink, orange; overlapping lines) that are higher than the distance between S. skua and S. chilensis (brown). Genetic distance was calculated in 5-kb windows as Kimura two-parameter distances, and lines were smoothed using a rolling mean of 200 windows. 24](#_Toc115133527)

[**Figure S12.** Genetic divergence across the genomes of the “Catharacta” skuas (S. skua, S. antarcticus, S. maccormicki and S. chilensis). Genetic distances are low and are highly similar between all pairs of species. Genetic distance was calculated in 5-kb windows as Kimura two-parameter distances, and lines were smoothed using a rolling mean of 200 windows. 25](#_Toc115133528)

[**Figure S13.** No evidence for introgression between S. pomarinus and the small jaegers. The genetic distance measures used to calculate D3 are shown plotted across the genome for all comparisons involving the small jaegers (S. parasiticus and S. longicaudus). The genetic distance between S. pomarinus and the small jaegers is highly similar to, and not lower than, the genetic distance between the “Catharacta” skuas and the small jaegers across the genome. The genetic distance between S. parasiticus and S. longicaudus is also plotted for reference. Note that lines for 29 different combinations of samples are plotted, with close overlap between sample pairs. 26](#_Toc115133529)

[**fDM and fb** 26](#_Toc115133530)

[MC1R Gene Tree 27](#_Toc115133531)

[**Figure S14**. Maximum likelihood phylogeny of the MC1R gene with the samples sequenced for this study, along with the 55 haplotypes sequenced by (Janssen and Mundy 2017). Nodes are labelled with bootstrap support at nodes with more than 50% support. The pale morph S. pomarinus haplotypes are placed as sister to a clade containing the “Catharacta” skuas, while the dark morph S. pomarinus haplotypes are nested within the “Catharacta” skuas, consistent with a history of possible introgression of the dark morph haplotype. Note that the placement of the two small jaegers (S. parasiticus and S. longicaudus) is discordant with the estimated species phylogeny, but similar discordant topologies are observed at 12% of 5-kb gene trees at this node (Table S4). 29](#_Toc115133532)

[Test for Mitonuclear Coadaptation 29](#_Toc115133533)

[**Table S7**. Mitonuclear-interacting genes examined to test for co-introgression with mitochondria. The positions of the genes identified within the pseudochromosome-scale Stercorarius parasiticus assembly are provided, with introns included but with 5’ and 3’ untranslated regions excluded. Chromosome names correspond to homologous chromosomes in the Alca torda genome, which was used to scaffold the S. parasiticus genome. 30](#_Toc115133534)

[**Table S8**. Z chromosome homologs of W chromosome-encoded genes examined to test for co-introgression with mitochondria and chrW. The positions of the genes identified within the scaffolded Stercorarius parasiticus assembly are provided, with introns included and with 5’ and 3’ untranslated regions excluded. 35](#_Toc115133535)

[Supplementary Note 1: Mitochondrial genome assemblies 36](#_Toc115133536)

[Supplementary Note 2: Mitochondrial Substitutions 39](#_Toc115133537)

[References 45](#_Toc115133538)

# Supplementary Methods and Results

## Sequencing and Alignment

We used FastQC ﻿ (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to assess sequencing quality, and then trimmed reads using fastp v0.23.2 (Chen et al. 2018) with the settings “--detect\_adapter\_for\_pe --trim\_poly\_g --cut\_front --cut\_front\_window\_size 1 --cut\_front\_mean\_quality 20 --cut\_right --cut\_right\_window\_size 4 --cut\_right\_mean\_quality 20 --length\_required 40 --correction”. This removes adapter sequences, removes poly-G tracts which are common artifacts in Novaseq sequences, trims 5’ bases below a quality of 20, trims the 3’ end where it falls below a quality of 20 in a 4 bp sliding window, and keeps the resulting reads that are at least 40 bp in length.

We then generated datasets mapped to two different reference genomes: *Stercorarius parasiticus* (ASM1339691v1, GCA\_013396915), a member of the ingroup Stercorariidae, and *Alca torda* (bAlcTor1, GCA\_008658365.1), a member of the outgroup Alcidae. Mapping to the outgroup resulted in slightly lower mapping quality, but we used the outgroup reference genome to reduce the impact of reference bias on analyses such as the ABBA BABA test that may be particularly sensitive to the effects of reference bias associated with variation between samples in their phylogenetic distance to the reference genome: in particular, using *S. parasiticus* as the reference genome could bias genotype calls against genotypes in which the sample differs from *S. parasiticus.* If this bias affects genotype calls in the outgroup, it may reduce the relative frequency of genotype patterns in which *S. parasiticus* carries a derived allele, leading to false signals in the ABBA BABA test. We used the *Alca torda* reference genome for the ABBA BABA test and related statistics (Fbranch, *f*dM) and for the comparison of genetic distances between the mitogenome and nuclear genome (as this relied on comparing estimates of genetic divergence between taxa with varying distances to *S. parasiticus*). We used *S. parasiticus* as the reference genome for all other analyses to take advantage of higher mapping quality in the stercorariids. As the *S. parasiticus* genome was not assembled into chromosomes, we scaffolded it into pseudochromosomes using the *Alca torda* genome as a reference with RagTag v2.2.0 (Alonge et al. 2021).

We mapped reads to the genomes using bowtie2 v2.4.4 (Langmead and Salzberg 2012) with the following settings: “--very-sensitive --no-unal --mp 5,1”. We used samtools v1.14 (Li et al. 2009) to sort the bam file and mark optical duplicates, and assessed results with qualimap v2.2.1 (Okonechnikov et al. 2016). Sample information is given in Table S1, sequencing statistics are provided in Table S2, and mapping statistics are provided in Table S3.

### **Table S1.** Samples sequenced for this study. Samples were obtained from the Royal Ontario Museum in Toronto, Canada. Sex was identified based on the heterozygosity and relative coverage of the Z chromosome.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Taxon** | **Common Name** | **Sample ID** | **NCBI BioSample ID** | **Collection Date** | **Locality** | **Sex** | **Sample Type** |
| ***Stercorarius longicaudus*** | Long-tailed Jaeger/Skua | MKP 990 (ROM Birds 156112) | SAMN31701265 | July 8, 1990 | Nunavut, 69.115°N, 105.088°W | Female | Muscle |
| ***Stercorarius parasiticus*** | Parasitic Jaeger (Arctic Skua) | B-20730 | SAMN31701266 | NA | Louisiana, USA | Female | Muscle |
| ***Stercorarius parasiticus*** | Parasitic Jaeger (Arctic Skua) | Genbank SAMN12253778 (B10K-DU-001-20) | SAMN12253778 | February 2, 1989 | Panama, 9.24°N, 82.31°W | Male | Muscle |
| ***Stercorarius pomarinus*** | Pomarine Jaeger/Skua | MKP 1559 (ROM Birds 157062) | SAMN31701267 | June 15, 1992 | Alaska, 71.290°N, 156.792 °W | Male | Muscle |
| ***Stercorarius pomarinus*** | Pomarine Jaeger/Skua (Dark Morph) | ROMO513777 (POJA4) | SAMN31701268 | November 13, 2011 | Lake Ontario, Canada | Female | Muscle |
| ***Stercorarius pomarinus*** | Pomarine Jaeger/Skua | 1B-2659 (ROM Birds 159823) | SAMN31701269 | November 27, 1996 | Lake Ontario, Canada, 43.197037°N, 79.457344°W | Female | Muscle |
| ***Stercorarius chilensis*** | Chilean Skua | MKP 2451 (ROM Birds 158363) | SAMN31701270 | February 14, 1995 | Chile, -52.941°N, 70.796°W | Female | Muscle |
| ***Stercorarius maccormicki*** | South Polar Skua | E23 (Antarctic Skua 10) | SAMN31701271 | pre-1997 | Cape Royds, Ross Island, Antarctica | Male | Blood |
| ***Stercorarius maccormicki*** | South Polar Skua | E67 (Antarctic Skua 7) | SAMN31701272 | pre-1997 | Cape Royds, Ross Island, Antarctica | Male | Blood |
| ***Stercorarius maccormicki*** | South Polar Skua | E68 (Antarctic Skua 8) | SAMN31701273 | pre-1997 | Cape Royds, Ross Island, Antarctica | Female | Blood |
| ***Stercorarius antarcticus lonnbergi*** | Subantarctic Skua (Brown Skua) | C54 (Chatham Island Skua 2) | SAMN31701274 | 1990-1991 | Chatham Islands | Male | Blood |
| ***Stercorarius antarcticus lonnbergi*** | Subantarctic Skua (Brown Skua) | C55 (Chatham Island Skua 3) | SAMN31701275 | 1990-1991 | Chatham Islands | Female | Blood |
| ***Stercorarius antarcticus lonnbergi*** | Subantarctic Skua (Brown Skua) | C72 (Chatham Island Skua 5) | SAMN31701276 | 1990-1991 | Chatham Islands | Female | Blood |
| ***Stercorarius skua skua*** | Great Skua | MKP 1592 (ROM birds 157280) | SAMN31701277 | July 17, 1992 | Iceland, 63.907°N, 16.705°W | Female | Muscle |
| ***Stercorarius skua skua*** | Great Skua | MKP 1593 (ROM birds 157281) | SAMN31701278 | July 17, 1992 | Iceland, 63.907°N, 16.705°W | Male | Muscle |

### **Table S2.** Sequencing statistics for samples sequenced in this study. %GC measures the proportion of G and C nucleotides in the trimmed sequencing data. Heterozygosity is estimated by Rohan (Renaud et al. 2019) on autosomes. Divergence is measured as Kimura 2-parameter distances, averaged across autosomal 5-kb windows.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Taxon** | **Sample ID** | **Shotgun Library type** | **Platform** | **Raw sequencing Data (Gigabases)** | **%GC** | **Heterozygosity** | **Divergence from *Alca torda*** |
| *Stercorarius longicaudus* | MKP990 (ROM Birds 156112) | PCR-free | Illumina HiSeq X | 29.25 | 46 | 0.71% | 4.55% |
| *Stercorarius parasiticus* | B-20730 | PCR-free | Illumina HiSeq X | 30.42 | 48 | 0.26% | 4.57% |
| *Stercorarius parasiticus* | GenBank SAMN12253778 (B10K-DU-001-20) | NA (GenBank) | Illumina HiSeq 2000 | 25.14 | 42 | 0.21% | 4.53% |
| *Stercorarius pomarinus* | MKP1559 (ROM Birds 157062) | PCR-free | Illumina HiSeq X | 27.33 | 48 | 0.42% | 4.65% |
| *Stercorarius pomarinus* | ROMO513777 (POJA4) | PCR-free | Illumina NovaSeq S4 | 19.56 | 43 | 0.31% | 4.57% |
| *Stercorarius pomarinus* | 1B-2659 (ROM Birds 159823) | PCR-free | Illumina NovaSeq S4 | 19.5 | 45 | 0.31% | 4.57% |
| *Stercorarius chilensis* | MKP2451 (ROM Birds 158363) | PCR-free | Illumina HiSeq X | 19.11 | 47 | 0.15% | 4.59% |
| *Stercorarius maccormicki* | E23 (Antarctic Skua 10) | PCR-free | Illumina NovaSeq S4 | 17.52 | 45 | 0.13% | 4.58% |
| *Stercorarius maccormicki* | E67 (Antarctic Skua 7) | PCR-free | Illumina HiSeq X | 21.84 | 43 | 0.12% | 4.57% |
| *Stercorarius maccormicki* | E68 (Antarctic Skua 8) | PCR-free | Illumina NovaSeq S4 | 18.54 | 43 | 0.13% | 4.58% |
| *Stercorarius antarcticus lonnbergi* | C54 (Chatham Island Skua 2) | PCR-free | Illumina NovaSeq S4 | 22.08 | 45 | 0.11% | 4.58% |
| *Stercorarius antarcticus lonnbergi* | C55 (Chatham Island Skua 3) | PCR amplified | Illumina HiSeq X | 25.26 | 46 | 0.11% | 4.56% |
| *Stercorarius antarcticus lonnbergi* | C72 (Chatham Island Skua 5) | PCR-free | Illumina NovaSeq S4 | 20.07 | 43 | 0.11% | 4.58% |
| *Stercorarius skua skua* | MKP1592 (ROM birds 157280) | PCR-free | Illumina HiSeq X | 25.23 | 47 | 0.13% | 4.57% |
| *Stercorarius skua skua* | MKP1593 (ROM birds 157281) | PCR-free | Illumina NovaSeq S4 | 14.97 | 46 | 0.12% | 4.59% |

### **Table S3.** Sequence mapping statistics for samples sequenced in this study.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Taxon** | **Sample ID** | **Average Depth mapped to *Alca torda*** | **Average Depth mapped to *S. parasiticus*** | **Relative depth of Z compared to autosomes** |
| *Stercorarius longicaudus* | MKP990 (ROM Birds 156112) | 14.8 | 16.4 | 0.54 |
| *Stercorarius parasiticus* | B-20730 (BROM573-07) | 13.7 | 15.2 | 0.54 |
| *Stercorarius parasiticus* | GenBank SAMN12253778 (B10K-DU-001-20) | 15.1 | 15.7 | 0.99 |
| *Stercorarius pomarinus* | MKP1559 (ROM Birds 157062) | 11.4 | 12.4 | 0.88 |
| *Stercorarius pomarinus* | ROMO513777 (POJA4) | 13.6 | 14.5 | 0.56 |
| *Stercorarius pomarinus* | 1B-2659 (ROM Birds 159823) | 13.3 | 14.3 | 0.55 |
| *Stercorarius chilensis* | MKP2451 (ROM Birds 158363) | 8.8 | 9.7 | 0.53 |
| *Stercorarius maccormicki* | E23 (Antarctic Skua 10) | 12.1 | 12.9 | 0.98 |
| *Stercorarius maccormicki* | E67 (Antarctic Skua 7) | 13.7 | 15.0 | 1.00 |
| *Stercorarius maccormicki* | E68 (Antarctic Skua 8) | 12.8 | 13.7 | 0.56 |
| *Stercorarius antarcticus lonnbergi* | C54 (Chatham Island Skua 2) | 14.4 | 15.8 | 0.96 |
| *Stercorarius antarcticus lonnbergi* | C55 (Chatham Island Skua 3) | 9.3 | 10.4 | 0.54 |
| *Stercorarius antarcticus lonnbergi* | C72 (Chatham Island Skua 5) | 13.8 | 14.8 | 0.56 |
| *Stercorarius skua skua* | MKP1592 (ROM birds 157280) | 11.6 | 12.7 | 0.54 |
| *Stercorarius skua skua* | MKP1593 (ROM birds 157281) | 9.9 | 10.6 | 0.95 |

### **Sequence heterozygosity**

We used Rohan (Renaud et al. 2019) to estimate sample heterozygosity in 1 Mb windows across the genome. This allowed for comparisons of heterozygosity between samples (Figure S1, S2), and for identification of the sex chromosomes and the sex of each sample. Differences between the sex chromosomes of male and female samples might confound sensitive tests for introgression if not accounted for. This could be particularly important if sequences from homologous regions of the W chromosome are mapped to the Z chromosome, inflating the differences between male and female samples. We identified sex based on relative heterozygosity and sequencing depth of the Z chromosome, which is present as two copies in males and one copy in females, and thus shows lower relative sequencing depth and estimated heterozygosity in females (Figure S1, Table S3). We further used the relative heterozygosity of females across chromosome Z to identify a region where females have highly elevated heterozygosity, suggesting that paralogous W-chromosome sequences are mapping to that region of the Z chromosome (Figure S1). This region was excluded from downstream analyses.

Chart

Description automatically generated with low confidence

### **Figure S1. Heterozygosity across chromosome Z.** Heterozygosity was estimated by Rohan as the proportion of sites which are heterozygous within 1 Mb windows in each sample. The nine female samples show low estimates of heterozygosity across most of chrZ (which is in hemizygous state in females), but with a peak of very high heterozygosity from 0-10 Mb. This region is likely affected by paralogous W chromosome sequences mapping to the Z chromosome, and so was excluded from downstream analyses.

A screenshot of a computer

Description automatically generated with medium confidence

### **Figure S2. Heterozygosity across the genome.** Heterozygosity is estimated by Rohan as the proportion of sites which are heterozygous within 1 Mb windows in each sample. The autosomes are displayed in order of descending length, and the Z chromosome is plotted last. Across the genome, the three jaeger species (S. longicaudus, S. pomarinus, and S. parasiticus) have moderate to relatively high heterozygosity, while the four “Catharacta” skuas (S. skua, S. antarcticus, S. maccormicki, and S. chilensis) have markedly lower heterozygosity.

## Genotype calling pipeline

We called genotypes of each sample using bcftools v1.14 (Li 2011) with probabilistic realignment disabled, and then merged the resulting VCF files with bcftools merge. To create a dataset of single nucleotide polymorphism markers (SNPs), we filtered to keep biallelic, autosomal SNPs that were not within 10 bp of an indel, with a minimum mapping quality and genotype quality of 25, with average depth lower than 17. We filtered individual genotypes for a minimum of 6× coverage and minimum genotype quality of 20.

## Species Phylogeny

### Astral-III phylogeny

To construct the Astral-III phylogeny, we built maximum likelihood gene trees from 5-kb sequence blocks throughout the genome. First, we produced reference-based genome sequences for each sample mapped to *S. parasiticus* using angsd -dofasta 3 (Korneliussen et al. 2014). This method selects the base with the highest effective sequencing depth for each sample at each position in the genome, while filling gaps with N’s. We filtered sequences for base call and mapping quality with the settings “-minMapQ 20 -minQ 20 -remove\_bads 1 -only\_proper\_pairs 1 -uniqueOnly 1 -doCounts 1 -setMinDepth 3”. We then divided the genome into blocks of 5000 bp using the script extract\_blocks.rb (https://github.com/mmatschiner/tutorials), resulting in 223,915 blocks. This block length was chosen to reduce the effects of recombination within blocks, while retaining enough variants per block to estimate the phylogeny. Since autosomes and sex chromosomes may have different rates of lineage sorting and introgression, we analyzed autosomal loci and Z chromosome loci separately. We discarded loci within the first 30 Mb of the Z chromosome to avoid the influence of paralogous W chromosome sequences that map to the pseudoautosomal region of the Z chromosome in female samples (Figure S1). We then thinned the dataset by discarding every second block to reduce the effects of linkage, such that each block was spaced at least 5 kb apart, and we then filtered to retain blocks with no more than 6% missing data in any stercorariid sample. This resulted in a dataset of 58,747 blocks on autosomes and 754 blocks on the Z chromosome. We then constructed maximum likelihood phylogenies for each sequence block using IQ-TREE v2.1.3 (Minh et al. 2020), while selecting a model of sequence evolution for each block with ModelFinder (Kalyaanamoorthy et al. 2017). Then, we summarized these gene trees to produce a species tree with Astral-III v5.7.3 (Zhang et al. 2018) (Figure S4). We calculated normalized quartet scores for each node within Astral-III, and we visualized the distribution of gene trees using Densitree (Bouckaert and Heled 2014) (Figure S3). Internal branch lengths and the frequencies of discordant topologies are provided in Table S4.

**A picture containing text, screenshot

Description automatically generated**

### **Figure S3.** Densitree visualization of the maximum likelihood gene trees used in the Astral-III species tree reconstruction. One percent of the gene trees were randomly subsampled and plotted as overlapping trees to show variation in topology and branch lengths between loci. For visualization purposes, gene trees were midpoint rooted using the phangorn package v2.8.1 (Schliep 2011) in R v4.1.2 (R Core Team 2018), scaled to a root age of 36.6, and converted to an ultrametric tree using chronos within the ape package v5.6.2 (Paradis and Schliep 2019). Incomplete lineage sorting and gene tree discordance is prevalent at the shallowest nodes.

**Graphical user interface, application

Description automatically generated**

### **Figure S4.** Astral-III phylogenies estimated using58,747autosomal loci (left) or754 Z chromosome loci (right) in a multispecies coalescent framework. Branch lengths are measured in coalescent units and reflect the amount of gene tree concordance at a node. Dashed lines indicate branches whose lengths are not estimated, as terminal branch lengths can only be estimated by Astral for taxa with more than one sample. Nodes are labelled with local posterior probabilities based on quartet frequencies, followed by normalized quartet scores, which reflect the proportion of gene trees that are concordant with the species topology at that node.

### **Table S4**. Internal branch lengths inferred by Astral-III and the expected and observed frequencies of discordant gene tree topologies under incomplete lineage sorting based on these estimated internal branch lengths. Branch lengths are measured in coalescent units at the internal branch in between the divergence of P1 from P2 and the divergence of (P1,P2) from P3. Observed proportions are the normalized quartet score of the discordant topology, which indicates the proportion of gene trees that support a topology in which P3 is sister to either P1 or P2. Under a scenario of incomplete lineage sorting in the absence of introgression, the frequencies of the two discordant topologies are expected to be equal, but introgression or ancestral population structure can result in an imbalance between the two frequencies. Note that since Astral-III branch length estimates are based on the observed frequencies of discordant gene trees, the expected and observed frequencies are not independent. “Catharacta” refers to the clade of S. skua, S. antarcticus, S. chilensis, and S. maccormicki.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| P1 | P2 | P3 | Branch Length | Expected frequencies | Observed (P1,P3) | Observed (P2, P3) |
| *“Catharacta”* | *S. pomarinus* | *S. longicaudus* & *S. parasiticus* | 7.7884 | 0.0001 | 0.0001 | 0.0001 |
| *S. parasiticus* | *S. longicaudus* | *“Catharacta”* & *S. pomarinus* | 1.7236 | 0.059 | 0.050 | 0.069 |
| *S. skua* | *S. maccormicki, S. antarcticus, & S. chilensis* | *S. pomarinus* | 0.6955 | 0.17 | 0.17 | 0.16 |
| *S. chilensis* | *S. antarcticus & S. maccormicki* | *S. skua* | 0.0506 | 0.32 | 0.29 | 0.34 |
| *S. antarcticus* | *S. maccormicki* | *S. chilensis* | 0.0303 | 0.32 | 0.35 | 0.29 |

### Starbeast3 Phylogeny

We built a species phylogeny using StarBeast3 v1.0.4 (Douglas et al. 2022) using a dataset of 107 5-kb phased haplotype blocks (Figure S5). To construct the blocks, we filtered genotype calls for a minimum mapping quality of 20, maximum average sequencing depth of 20×, and minimum individual genotype depth of 5×. We phased genotypes into haplotypes using read-aware phasing in shapeit v2.904 (Delaneau et al. 2013), with a minimum read quality and base quality of 20 in the extractPIRs step. We generated haplotype sequences from the resulting VCF file using bcftools consensus v1.14 (Li 2011). Since shapeit imputes missing data, we reverted imputed genotypes back to missing data using bedtools maskfasta v2.30.0 (Quinlan and Hall 2010). We divided the genome into 5 kb blocks. As the computational requirements of Starbeast3 preclude analyzing the entire dataset, we filtered to remove all blocks in which any stercorariid had more than 7% missing data, resulting in a dataset of 107 autosomal blocks spread throughout the genome, each a minimum of 10 kb apart.

We calibrated the age of the root of the tree based on previous estimates for the divergence time between Alcidae and Stercorariidae. Smith & Clarke (2015) estimated the age of divergence to be 34.7 Ma, with a narrow confidence interval (34.2–35.5 Ma). This is similar to estimates from a previous study (36.6 Ma) based on independent calibrations (Weir and Mursleen 2013), and is consistent with the age of the oldest known Alcid fossil (34.2–36.0 Myr (Chandler and Parmley 2003; Smith 2015)). We used a log-normal prior distribution with a mean of 34.7 Myr and standard deviation of 0.008, which corresponds to a 95% density of 34.2–35.2 Myr.

We used a Birth Death model and strict clock for the species tree, with a log normal prior distribution for clock rate with mean of 0.002 and standard deviation of 2 (corresponding to 95% density of 5x10-6–0.01). For each locus we used an HKY site model with one gamma category (i.e., without gamma variation) and with the proportion of invariant sites estimated. We estimated the proportion of invariant sites rather than incorporating gamma variation in order to reduce computational cost; preliminary analyses with four gamma categories per locus reached the same species topology but ran too slowly to achieve a sufficient effective sample size for key parameters. We ran the analyses replicated 16 times for 100,000,000 generations each, and combined runs using LogCombiner v2.6.7 after discarding the first 20% of each run as burn-in.

A picture containing graphical user interface

Description automatically generated

### **Figure S5.** Time-calibrated species tree inferred by Starbeast3. Node bars indicate the 95% HPD for the age of each node. All nodes have full posterior probability except for the shallowest two nodes (labelled). Circles next to the species images represent the relative effective population sizes estimated for each species, with the area of the circle proportional to effective population size. The phylogeny is magnified on the right to show greater detail at the shallowest nodes.

## mtDNA Assembly

We assembled complete mtDNA genomes for each individual using NOVOPlasty v4.2.1 (Dierckxsens et al. 2017) from the raw sequencing reads after removing adapter sequences with Trimmomatic v0.38 (Bolger et al. 2014). We used the ND1 sequence of *Stercorarius maccormicki* (GenBank NC\_026125.1) (Han et al. 2016) as a reference seed sequence to initiate the assembly. For samples that initially failed to produce a complete assembly, we repeated the assembly process using a seed sequence from within the assembly gap. We aligned the resulting mitogenomes of all samples with the previously-published *S. maccormicki* mtDNA genome (GenBank NC\_026125.1) in Mesquite (Maddison and Maddison 2018) using MAFFT (Katoh and Standley 2013). For samples that produced multiple contigs, the sequences of other samples were used as references to assemble the contigs into complete mitogenomes. This resulted in a continuous, circularized mtDNA genome sequence for all samples.

To ensure that the mitogenome sequences did not contain numts (nuclear copies of mitochondrial sequences), we mapped the raw sequencing reads to the mitogenomes using bwa-mem v0.7.17 (Li and Durbin 2009), and quantified the frequency of each base at each position using Angsd -doCounts (Korneliussen et al. 2014). Since mitochondrial genomes are present at a much higher copy-number than the nuclear genome in bird muscle tissue, the mitochondrial sequence is expected to have much higher sequencing depth than the numt sequence (e.g., >1000× vs <20× depth). We identified two cases in which variants from a numt were incorporated into the mitogenome assembly, and manually corrected the sequence using the mapped reads with the higher-depth variants. For blood samples in which the mitogenome was not sequenced to greatly higher depth than the nuclear genome, we examined the region surrounding the numts that were characterized in the muscle samples, identified a third case of incorporation of a numt, and corrected it using the raw reads. To identify the locations of protein-coding genes, rRNA, and tRNA sequences, we annotated the mitogenome of *S. maccormicki* using the Mitos Webserver (Bernt et al. 2013). Details of the mitogenomes, including the description of a duplicated control region, are provided in Supplementary Note 1.

Since contamination or Illumina index hopping between samples could introduce a pattern of shared alleles potentially mimicking a signal of introgression, we also examined the allele counts in order to look for look for cases where a sample carried variants that match a different sample’s mitogenome sequence. Since mitochondria are usually present at very high copy number in bird cells, mitochondrial DNA is usually sequenced to a much higher depth than nuclear DNA in the shotgun sequencing protocol followed here. As a result, if no contaminating mitochondrial sequences can be detected, then nuclear contamination is likely to be negligible. No mitochondrial contamination was detected.

### Mitochondrial phylogeny

Details for the maximum likelihood phylogeny and time-calibrated Bayesian phylogeny of the mitogenome are provided in the main text. The maximum likelihood phylogeny is shown in Figure S6, and the time-calibrated Bayesian phylogeny is shown in Figure S7; the two analyses produced identical topologies.

***Graphical user interface, text

Description automatically generated***

### **Figure S6.** Maximum likelihood phylogeny of the whole mitogenomes of the Stercorariidae. The phylogeny is rooted with three species of Alcidae, and branch support values are based on 1000 bootstrap replicates. One copy of the duplicated region was removed prior to analysis, as the two copies showed evidence for concerted evolution and thus do not represent independently evolving sequence.

***A picture containing diagram

Description automatically generated***

### **Figure S7.** Time-calibrated mitochondrial phylogeny produced by BEAST. Node bars show the 95% HPD of node ages. All nodes have full posterior probability, except for two nodes (labelled). The phylogeny is magnified and reflected on the right side to show detail at the shallower nodes.

Since the maximum likelihood mitochondrial phylogeny placed *S. pomarinus* sister to *S. skua* (Figure S6, S7), we investigated whether the mitochondrial divergence between *S. pomarinus* and *S. skua* was shallower than expected based on the nuclear genome. In order to compare the mitochondrial and nuclear estimates of genetic divergence, which are driven by different mutation rates, we standardized the Kimura two-parameter genetic distances between each sample by dividing them by the genetic distance between *S. pomarinus* and *S. parasiticus*, which spans the deepest node of the Stercorariidae phylogeny and is not expected to be substantially affected by introgression. We repeated this process for the genetic distances between each *S. pomarinus* sample and each member of the “*Catharacta*” skua clade (*S. antarcticus, S. chilensis, S. skua, S. maccormicki*) for a dataset of 16.7-kb blocks of the genome. The 16.7-kb block size was chosen to match the single-copy length of the mitogenome. One copy of the mitochondrial duplication was excluded to avoid pseudoreplication in the calculation of genetic distance between mitogenomes. To avoid reference bias associated with differences in genetic distance to the reference genome, we used the dataset mapped to the genome of the outgroup *Alca torda*. Histograms of the standardized genetic distances are shown in Figure S8.

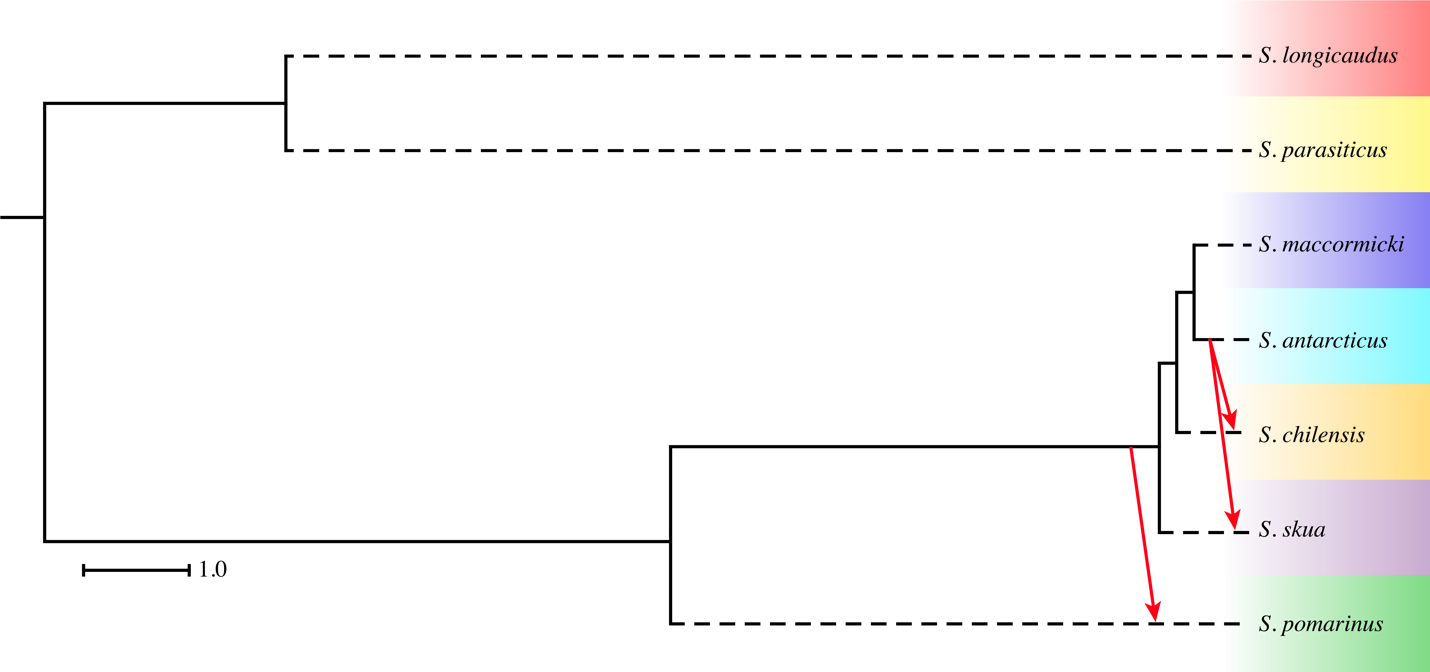
****

### **Figure S8.** Histogram ofrelativedivergence between S. pomarinus and the four “Catharacta” skuas (S. chilensis, S. skua, S. antarcticus, and S. maccormicki) within 16.7-kb windows of the nuclear genome. Since mitochondrial sequences diverge at a much higher rate than nuclear sequences in birds, all genetic distances are standardized by dividing each Kimura two-parameter distance estimate by the genetic distance between S. pomarinus and S. parasiticus within the same window. The distributions are plotted as 30 separate, but highly overlapping distributions for each combination of the three S. pomarinus and ten “Catharacta” skuas. The point estimates for the relative mitochondrial divergence between S. pomarinus and the southern hemisphere skuas (S. maccormicki, S. antarcticus, and S. chilensis) are drawn as black lines, and the relative mitochondrial divergences between the S. pomarinus and S. skua samples are drawn as red lines. Only 0.44% of the nuclear windows fall equal to or lower than the average of the red lines, while 21.1% of windows are equal to or lower than the average of the black lines.

## Tests of Introgression

### **Phylonet**

We built a phylogenetic network in a multispecies coalescent framework with Phylonet v3.8.2 (Than et al. 2008; Wen et al. 2018). As input, we used a dataset of 9899 phased autosomal 5-kb blocks. These blocks were constructed as described for the StarBeast3 analysis, but with a final filter for a maximum of 1000 bp missing data in any Stercorariid, rather than a maximum of 350 bp. Methods for the Phylonet analysis are provided in the main text. The best network inferred three introgression events, shown in Figure S9.



### **Figure S9.** Maximum pseudolikelihood network returned by Phylonet. Introgression events are shown as red arrows. Internal branch lengths are estimated in coalescent units, but terminal branch lengths were not estimated, and are drawn as dotted lines of arbitrary length.

### **ABBA BABA tests**

We used ABBA BABA test *D* statistics (Green et al. 2010) to test for an excess of shared derived alleles between non-sister lineages. We ran ABBA BABA tests using Dsuite (Malinsky et al. 2021), selecting the dataset mapped to the outgroup *Alca torda* to reduce the effects of reference bias. We used the SNP dataset further filtered for a maximum proportion of missing data of 17.6%. We restricted the analysis to the autosomes in order to remove any biases associated with W chromosome sequences that map to the Z chromosome in females (Figure S1). We used the alcids *Alca torda* and *Uria lomvia* as the outgroup in the P4 position for trios that included *S. longicaudus* and/or *S. parasiticus*, and used *S. parasiticus* and *S. longicaudus* as the outgroup for trios that do not involve those species. We visualized the results using the script plot\_d.rb (<https://github.com/mmatschiner/tutorials>) (Figure S10). Results are provided in Table S5. Statistical significance was assessed using a block jackknife approach within Dsuite.

To assess the impact of reference genome choice on the ABBA BABA test, we repeated the analysis using the dataset mapped to *Stercorarius parasiticus*. Results were qualitatively identical for the three introgression events that were inferred when using *Alca torda* as the reference genome, but with an additional statistically significant excess of shared derived alleles between *S. longicaudus* and the “*Catharacta*” + *S. pomarinus* clade (*D* = 0.03–0.05); this excess is reduced (*D=*0*.*018) and no longer statistically significant when using *Alca torda* as the reference genome. The signal of excess shared derived alleles in *S. longicaudus* is likely due to the impact of reference bias as the three species in the trio are different phylogenetic distances to the reference genome. We hypothesize that reference bias leads to a slight but statistically significant bias against calling sites in the outgroup Alcidae that differ from the *S. parasiticus* reference genome, leading to a bias against “BABA” pattern sites that are derived in *S. parasiticus,* and consequently a statistically significant excess of “ABBA” sites over “BABA” sites, mimicking a signature of introgression.

A picture containing square

Description automatically generated

### **Figure S10.** Visualization of estimates for D and fb for each pair of non-sister taxa. Estimates for D are shown above the diagonal, colour coded from low (blue) to high (red), with the colour saturation corresponding to the p-value (legend in bottom right). Estimates for fb are shown below the diagonal, ranging from low (white) to high (red).

### **Table S5.** D (ABBA-BABA test) and D3 statistics calculated for all trios. Raw p-values for each ABBA-BABA test are provided without correcting for multiple testing. D3 was averaged across 5-kb windows of the genome, and the proportion of windows for which D3 was negative is provided. A significantly positive value of D and a negative value of D3 are consistent with introgression between P2 and P3. Statistics are provided for all trios that are consistent with the estimated species tree, as well as the two alternate topologies at the node containing S. antarcticus, S. maccormicki, and S. chilensis where gene tree discordance is very high.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **P1** | **P2** | **P3** | **P4** | ***D*** | ***p*-value** | ***f*4-ratio** | ***D*3** | ***D*3 Proportion** |
| *S. maccormicki* | *S. antarcticus* | *S. skua* | *S. longicaudus, S. parasiticus* | 0.23 | **0** | 0.15 | -0.054 | 0.44 |
| *S. chilensis* | *S. antarcticus* | *S. skua* | *S. longicaudus, S. parasiticus* | 0.22 | **0** | 0.15 | -0.064 | 0.42 |
| *S. chilensis* | *S. antarcticus* | *S. maccormicki* | *S. longicaudus, S. parasiticus* | 0.17 | **0** | 0.20 | -0.050 | 0.45 |
| *S. maccormicki* | *S. antarcticus* | *S. chilensis* | *S. longicaudus, S. parasiticus* | 0.21 | **0** | 0.13 | -0.047 | 0.44 |
| *S. maccormicki* | *S. chilensis* | *S. antarcticus* | *S. longicaudus, S. parasiticus* | 0.046 | **3.80 x10-6** | 0.047 | 2.01x10-3 | 0.50 |
| *S. maccormicki* | *S. skua* | *S. pomarinus* | *S. longicaudus, S. parasiticus* | 0.094 | **0** | 0.12 | -6.34x10-3 | 0.49 |
| *S. antarcticus* | *S. skua* | *S. pomarinus* | *S. longicaudus, S. parasiticus* | 0.094 | **2.74 x10-16** | 0.11 | -4.69x10-3 | 0.49 |
| *S. chilensis* | *S. skua* | *S. pomarinus* | *S. longicaudus, S. parasiticus* | 0.083 | **2.13 x10-10** | 0.11 | -6.67x10-3 | 0.49 |
| *S. skua* | *S. pomarinus* | *S. parasiticus* | Alcidae | 0.066 | **2.99 x10-7** | 1.39 x10-3 | -9.73x10-4 | 0.49 |
| *S. chilensis* | *S. pomarinus* | *S. longicaudus* | Alcidae | 0.063 | **1.16 x10-6** | 1.71 x10-3 | 2.04 x10-4 | 0.50 |
| *S. chilensis* | *S. pomarinus* | *S. parasiticus* | Alcidae | 0.062 | **2.83 x10-6** | 1.37 x10-3 | 5.37 x10-4 | 0.50 |
| *S. maccormicki* | *S. pomarinus* | *S. parasiticus* | Alcidae | 0.062 | **6.70 x10-6** | 1.32 x10-3 | -1.84 x10-3 | 0.48 |
| *S. maccormicki* | *S. pomarinus* | *S. longicaudus* | Alcidae | 0.061 | **3.53 x10-6** | 1.59 x10-3 | -1.86 x10-3 | 0.48 |
| *S. antarcticus* | *S. pomarinus* | *S. parasiticus* | Alcidae | 0.060 | **1.07 x10-5** | 1.29 x10-3 | -2.04 x10-3 | 0.48 |
| *S. skua* | *S. pomarinus* | *S. longicaudus* | Alcidae | 0.059 | **4.93 x10-5** | 1.53 x10-3 | -1.14 x10-3 | 0.49 |
| *S. antarcticus* | *S. pomarinus* | *S. longicaudus* | Alcidae | 0.058 | **2.45 x10-7** | 1.53 x10-3 | -2.06 x10-3 | 0.48 |
| *S. parasiticus* | *S. longicaudus* | *S. pomarinus* | Alcidae | 0.018 | 0.013 | 2.91 x10-3 | 6.38 x10-4 | 0.51 |
| *S. parasiticus* | *S. longicaudus* | *S. skua* | Alcidae | 0.018 | 0.014 | 2.41 x10-3 | 1.65 x10-4 | 0.50 |
| *S. parasiticus* | *S. longicaudus* | *S. maccormicki* | Alcidae | 0.018 | 0.014 | 2.47 x10-3 | 7.45 x10-6 | 0.50 |
| *S. parasiticus* | *S. longicaudus* | *S. antarcticus* | Alcidae | 0.018 | 0.010 | 2.37 x10-3 | -4.22 x10-6 | 0.50 |
| *S. antarcticus* | *S. chilensis* | *S. pomarinus* | *S. longicaudus, S. parasiticus* | 0.00076 | 0.478 | 7.26 x10-4 | 1.95 x10-3 | 0.50 |
| *S. maccormicki* | *S. antarcticus* | *S. pomarinus* | *S. longicaudus, S. parasiticus* | 0.015 | 0.122 | 0.014 | -1.67 x10-3 | 0.50 |
| *S. maccormicki* | *S. chilensis* | *S. pomarinus* | *S. longicaudus, S. parasiticus* | 0.016 | 0.066 | 0.017 | 4.01 x10-4 | 0.50 |
| *S. maccormicki* | *S. chilensis* | *S. skua* | *S. longicaudus, S. parasiticus* | 0.011 | 0.244 | 7.62 x10-3 | 0.011 | 0.51 |
| *S. antarcticus* | *S. chilensis* | *S. longicaudus* | Alcidae | 0.0020 | 0.468 | 2.57 x10-5 | -2.27 x10-3 | 0.48 |
| *S. chilensis* | *S. antarcticus* | *S. parasiticus* | Alcidae | 0.0040 | 0.424 | 4.13 x10-5 | 2.58 x10-3 | 0.53 |
| *S. chilensis* | *S. maccormicki* | *S. longicaudus* | Alcidae | 0.0017 | 0.472 | 2.30 x10-5 | 2.05 x10-3 | 0.52 |
| *S. chilensis* | *S. maccormicki* | *S. parasiticus* | Alcidae | 0.0089 | 0.348 | 1.01 x10-4 | 2.36 x10-3 | 0.52 |
| *S. chilensis* | *S. skua* | *S. longicaudus* | Alcidae | 0.0041 | 0.432 | 6.13 x10-5 | 1.32 x10-3 | 0.51 |
| *S. chilensis* | *S. skua* | *S. parasiticus* | Alcidae | 0.0 | 0.500 | 0 | 1.48 x10-3 | 0.51 |
| *S. maccormicki* | *S. antarcticus* | *S. longicaudus* | Alcidae | 0.0060 | 0.380 | 7.23 x10-5 | 1.52 x10-4 | 0.50 |
| *S. maccormicki* | *S. antarcticus* | *S. parasiticus* | Alcidae | 0.0035 | 0.433 | 3.45 x10-5 | 1.46 x10-4 | 0.50 |
| *S. maccormicki* | *S. skua* | *S. longicaudus* | Alcidae | 0.0046 | 0.399 | 6.51 x10-5 | -7.74 x10-4 | 0.49 |
| *S. parasiticus* | *S. longicaudus* | *S. chilensis* | Alcidae | 0.014 | 0.038 | 1.91 x10-3 | 2.91 x10-4 | 0.50 |
| *S. skua* | *S. antarcticus* | *S. longicaudus* | Alcidae | 0.0015 | 0.472 | 1.93 x10-5 | 9.54 x10-4 | 0.51 |
| *S. skua* | *S. antarcticus* | *S. parasiticus* | Alcidae | 0.0077 | 0.348 | 8.06 x10-5 | 1.11 x10-3 | 0.51 |
| *S. skua* | *S. maccormicki* | *S. parasiticus* | Alcidae | 0.0027 | 0.437 | 3.13E-05 | 9.34E-04 | 0.51 |

### **D3**

The *D*3 statistic (Hahn and Hibbins 2019) tests for introgression between non-sister taxa by testing for a reduction in the average genetic distance between the hybridizing taxa. Under a scenario of pure incomplete lineage sorting in the absence of introgression or ancestral population structure, the expected value of *D*3 is zero. We calculated *D*3 for all trios using a dataset of genetic distance estimates for the 223,915 5000-bp sequence blocks generated for the Astral-III phylogeny (prior to filtering for missing data). We calculated the Kimura two-parameter distance between each sample within each block using Emboss distmat v6.6.0 (Rice et al. 2000). For each species trio, we filtered the blocks further such that no sample within the trio had more than 6% missing data.We calculated *D*3in *R* v4.1.2(R Core Team 2018) using the formula (Hahn and Hibbins 2019), where *d* is the genetic distance between two populations in a trio of three populations — P1, P2, and P3 — in which P1 and P2 are sisters and P3 is a population that may have introgressed with either P1 or P2.Results for all trios are provided in Table S5. The average Kimura two-parameter genetic distances between all samples are provided in Table S6, and a genome scan of the distances are shown in Figure S11, Figure S12, and Figure S13.



Note that since *D*3 detects a reduction in the genetic distance between two species after hybridization, it will have little power to detect hybridization when sequence diversity within the species is comparable to sequence divergence between species, as is likely the case in *S. pomarinus* (in which sequence divergence between *S. pomarinus* individuals is similar to sequence divergence between *S. pomarinus* and “*Catharacta*” skuas across most of the genome (Table S6)). To illustrate this, we calculated a theoretical maximum value of *D*3 for the trios that test for introgression between *S. pomarinus* in P3 and “*Catharacta*” skuas in P1 and P2. We placed different *S. pomarinus* individuals in P2 and P3 with “*Catharacta*” skuas in P1, to mimic a situation in which the P2 population received 100% introgression from *S. pomarinus*. Despite representing 100% introgression, the magnitude of *D*3 was low (*D*3=-0.01), highlighting the low power of the test for this trio, and indicating that the low magnitude of the observed values (*D*3=-0.0047 to -0.0067) are not necessarily inconsistent with the strong ABBA BABA test result. In contrast, the theoretical maximum value of *D*3 for a trio testing for introgression between *S. parasiticus* and *S. pomarinus* (calculated by placing different *S. parasiticus* individuals in P2 and P3 while *S. pomarinus* in in P1) is -0.77, and so the low magnitude of the observed values (-0.0011 to 0.00053) are inconsistent with substantial introgression from *S. parasiticus* or *S. longicaudus* into *S. pomarinus*, rejecting the previously-proposed hybrid origin of *S. pomarinus* involving one of the small jaegers.

### **Table S6.** Genetic distances between the samples. Distances were calculated using the Kimura two-parameter correction for multiple substitutions, and are presented as percentages. Mitochondrial sequence divergence is given above the diagonal, and autosomal nuclear divergence is given below the diagonal. Autosomal heterozygosity estimated by Rohan is given on the diagonal. Cells are shaded according to their value from low (blue) to high (red), with separate scales for the mitochondrial and nuclear estimates. Mitogenome distances are calculated while including only one copy of the duplicated control region.



**Chart, line chart

Description automatically generated**

### **Figure S11.** Genetic distances across the genomes of S. parasiticus, S. pomarinus, S. skua, and S. chilensis are inconsistent with a hybrid speciation scenario. S. parasiticus shows similar levels of genetic divergence to all three taxa across the genome (green, blue, yellow; overlapping lines), with no indication of lower divergence with S. pomarinus (green). Similarly, S. pomarinus shows nearly identical genetic distances to both S. skua and S. chilensis (pink, orange; overlapping lines) that are higher than the distance between S. skua and S. chilensis (brown). Genetic distance was calculated in 5-kb windows as Kimura two-parameter distances, and lines were smoothed using a rolling mean of 200 windows.

Chart, line chart

Description automatically generated

### **Figure S12.** Genetic divergence across the genomes of the “Catharacta” skuas (S. skua, S. antarcticus, S. maccormicki and S. chilensis). Genetic distances are low and are highly similar between all pairs of species. Genetic distance was calculated in 5-kb windows as Kimura two-parameter distances, and lines were smoothed using a rolling mean of 200 windows.

Chart

Description automatically generated with medium confidence

### **Figure S13.** No evidence for introgression between S. pomarinus and the small jaegers. The genetic distance measures used to calculate D3 are shown plotted across the genome for all comparisons involving the small jaegers (S. parasiticus and S. longicaudus). The genetic distance between S. pomarinus and the small jaegers is highly similar to, and not lower than, the genetic distance between the “Catharacta” skuas and the small jaegers across the genome. The genetic distance between S. parasiticus and S. longicaudus is also plotted for reference. Note that lines for 29 different combinations of samples are plotted, but are closely overlapping.

### **fDM and fb**

We estimated the proportion of introgression using the *f*-branch statistic *fb*, and *fDM.* Each statistic estimates the proportion of the genome that may have introgressed: while *fb* is designed to estimate a single genome-wide proportion and incorporates results from multiple trios, *fDM* can be applied at a sliding window scale within a single trio. *fb* builds upon the *f*4 admixture ratio statistic to account for the non-independence between *f*4 admixture ratio tests involving lineages that share branches of a phylogenetic tree. This method can detect gene flow involving internal or external branches, when the level of gene flow is greater than approximately 1%. To calculate *f*b,we used Dsuite (Malinsky et al. 2021), which first calculates *D* and *f*4 statistics for all trios using the autosomal SNP dataset, and then estimates *f*b for each branch of the species tree. *fDM* is based on the *f*d statistic, which is a more accurate estimate of the proportion of introgression than *D* (Martin et al. 2015)*.*

## MC1R Gene Tree

Variation in the protein MC1R has been demonstrated to form the basis of the dark and light morph plumage colour polymorphism in *S. pomarinus* and *S. parasiticus* (Janssen and Mundy 2017). It has previously been observed that the dark morph haplotype of *S. pomarinus* shares a higher similarity to the “*Catharacta*” skua haplotypes than does the light morph *S. pomarinus* haplotype, and this led to speculation that the light morph haplotype might have introgressed from one of the small jaegers (*S. parasiticus* or *S. longicaudus)* into *S. pomarinus* (Janssen and Mundy 2017). However, the lack of phylogenetic signal within the short MC1R sequence precluded confident rooting of the gene tree and makes the most likely scenario of introgression or ancestral polymorphism difficult to evaluate.

We located the MC1R sequence in the *Stercorarius parasiticus* assembly using BLAST v2.12.0+ (Camacho et al. 2009) with the MC1R sequence of *Charadrius vociferus* (XP\_009888634.1). This produced a single top hit (95.9% identity, 100% coverage: chr7: 45887798-45888739). We extracted the region within 10 kb upstream and downstream of MC1R with bcftools view v1.14 (Danecek et al. 2021), filtering for a minimum mapping quality of 20. We used shapeit v2.904 (Delaneau et al. 2013) to phase haplotypes using read-aware phasing with minimum read quality and base quality of 20. We generated haplotype sequences from the phased genotypes using bcftools consensus v1.14, and then converted imputed genotypes back to missing data using bedtools maskfasta v2.30.0 (Quinlan and Hall 2010). To determine whether our *S. parasiticus* and *S. pomarinus* samples carried light or dark morph haplotypes, we compared their genotypes to the previously-published coding differences between dark and light morph haplotypes in these species (Janssen and Mundy 2017). Since we found that our dark-morph *S. pomarinus* sample was heterozygous for both the light and dark alleles, we manually examined the aligned sequencing reads in IGV (Robinson et al. 2011) to verify phasing and to check for evidence of incorrectly called heterozygous sites. We identified and corrected four sites that were called as homozygous but in which the sample had reads from both dark and light alleles (for example, a site called as homozygous for the light morph allele, but in which 4/14 of its reads actually carried the dark morph allele).

We constructed a maximum likelihood phylogeny of the MC1R gene and flanking noncoding sequence using IQtree2 v2.1.3 (Minh et al. 2020) with 1000 bootstrap replicates. The pale-morph *S. pomarinus* haplotype was placed as sister to a clade containing the “*Catharacta*” skuas (*S. skua, maccormicki, chilensis,* and *antarcticus*)and the dark *S. pomarinus* haplotypewith 94% bootstrap support. This pattern is inconsistent with the previously proposed scenario of introgression of the pale haplotype from a small jaeger to *S. pomarinus.* While recombination within the flanking or coding sequence of MC1R could potentially obscure the phylogenetic history of the causal SNPs responsible for the plumage coloration, we observed only a single derived SNP exclusively shared between the pale *S. pomarinus* haplotype and the small jaegers within MC1R and flanking regions, while ten derived SNPs are exclusively shared between *S. pomarinus* and all “*Catharacta*” skuas, rejecting an introgressed origin for the pale morph *S. pomarinus* MC1R haplotype from a small jaeger.

The maximum likelihood phylogeny placed the dark *S. pomarinus* allele nested within the “*Catharacta*” skuas with 100% bootstrap support. This pattern could be produced via incomplete lineage sorting of ancestral dark and light alleles, or via introgression of the dark allele from the “*Catharacta*”skuas into *S. pomarinus*. For incomplete lineage sorting to produce the maximum likelihood topology, at least three ancestral haplotypes would need to be retained within the “*Catharacta*” lineage (in addition to two in *S. pomarinus*). Further, we note that the dark *S. pomarinus* allele differs from *S. antarcticus* at only a single basepair in the 1775 bp alignment (at which the dark *S. pomarinus* haplotype shares a derived allele with the light haplotype – a possible instance of gene conversion). For the topology to have been produced by incomplete lineage sorting, these ancestral haplotypes would need to have been retained since the common ancestor of “*Catharacta*” and *S. pomarinus* with no mutations in the *S. antarcticus* haplotype and only the single mutation in the dark *S. pomarinus* haplotype. Overall, given genomic evidence for gene flow between *S. pomarinus* and *S. skua*, we consider introgression from “*Catharacta*” to *S. pomarinus* to be the most likely explanation for the origin of the dark morph within *S. pomarinus*.

We repeated the analysis by adding the additional 55 Stercorariid coding sequences from (Janssen and Mundy 2017) to the alignment. This produced a concordant topology placing the dark *S. pomarinus* haplotypes nested within the “*Catharacta”* skuas, and placing the pale *S. pomarinus* haplotypessister to that clade (Figure S14).

***Graphical user interface

Description automatically generated***

### **Figure S14**. Maximum likelihood phylogeny of the MC1R gene with the samples sequenced for this study, along with the 55 haplotypes sequenced by (Janssen and Mundy 2017). Nodes are labelled with bootstrap support at nodes with more than 50% support. The pale morph S. pomarinus haplotypes are placed as sister to a clade containing the “Catharacta” skuas, while the dark morph S. pomarinus haplotypes are nested within the “Catharacta” skuas, consistent with a history of possible introgression of the dark morph haplotype. Note that the placement of the two small jaegers (S. parasiticus and S. longicaudus) is discordant with the estimated species phylogeny, but similar discordant topologies are observed at 12% of 5-kb gene trees at this node (Table S4).

## Test for Mitonuclear Coadaptation

Since we found strong evidence for mitochondrial introgression from *S. pomarinus* into *S. skua*, we tested whether the nuclear genes that interact with the mitochondrially-encoded proteins have also introgressed. Similarly, since mitochondrial introgression implies that W chromosome introgression would also have occurred due to their shared maternal inheritance, we tested whether the Z chromosome homologs of W chromosome genes have co-introgressed along with the W. Reference protein sequences of each gene were obtained from the NCBI Gene database ortholog collections, which are produced in the NCBI Eukaryotic Genome Annotation pipeline using local synteny and protein sequence information. One reference bird protein per gene was selected, with preference for sequences from the Charadriiformes (*Calidris pugnax* or *Charadrius vociferus*)when available. We then located each gene in the *Stercorarius parasiticus* assembly using tBLASTn v2.12.0+ (Camacho et al. 2009). When multiple hits were detected (as in the case of paralogs), we selected the hit with higher protein sequence identity. If the best hit had less than 70% sequence identity with the reference protein along all exons, we considered it to be of uncertain orthology and excluded it from further analysis. We located 207 genes (Table S7, S8), but were unable to locate 15 genes (NDUFA13, NDUFB7, NDUFB11, UQCC3, COX5B, COX6B1, COX7A1, COX8A, OXA1L, PET100, SCO2, SARS2, MRPL4, MRPL11, HARS2) which may be missing from the *Stercorarius parasiticus* assembly or lack sufficient similarity to the reference protein. We then extracted these sequences from all samples using angsd -dofasta 3, and aligned the sequences using MAFFT v7.490 (Katoh and Standley 2013). We used IQtree v2.1.3 (Minh et al. 2020) to build maximum likelihood trees for each gene, and then identified genes which showed gene tree topologies consistent with co-introgression with the mitochondrion: topologies in which the two *S. skua* samples form a clade with at least one *S. pomarinus* sample to the exclusion of other species (i.e., *S. skua* nested within *S. pomarinus* or sister to *S. pomarinus*). To generate a null expectation for the proportion of genes that would show this topology in the absence of mitonuclear co-introgression, we assessed the proportion of gene trees with the putative mitonuclear co-introgression topology in the genome-wide dataset generated for Astral-III (56,005 gene trees). We identified gene trees with the mitochondrion-like topology using a custom script in R using the ape v5.5 and phangorn v2.8.1 packages, and the getDescendants() function (<http://blog.phytools.org/2012/01/function-to-get-descendant-node-numbers.html>) to identify trees that contained a clade with the two *S. skua* and at least one *S. pomarinus* sample to the exclusion of any other taxa. Results for the background proportion of gene trees with the mitonuclear topology are similar whether using the filtered dataset of 56,005 gene trees or the larger unfiltered dataset of 223,887 gene trees (6.5% vs 6.2%).

To determine whether any of the mitonuclear interacting genes that show a mitochondrial-like topology could have been driven to introgress due to coadaptation between mitochondrial and nuclear proteins, we compared the protein sequences encoded by the southern hemisphere skuas, *S. skua*, and *S. pomarinus.* If a gene were selected to introgress from *S. pomarinus* into *S. skua* due to mitonuclear coadaptation, we would expect to observe an amino acid substitution that differs between the southern hemisphere skuas and *S. skua/pomarinus*; however, none of the candidate genes contained this pattern.

### **Table S7**. Mitonuclear-interacting genes examined to test for co-introgression with mitochondria. The positions of the genes identified within the pseudochromosome-scale Stercorarius parasiticus assembly are provided, with introns included but with 5’ and 3’ untranslated regions excluded. Chromosome names correspond to homologous chromosomes in the Alca torda genome, which was used to scaffold the S. parasiticus genome.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Functional Category** | **Protein** | **Chromosome** | **Start** | **Stop** |
| Complex I Assembly | ACAD9 | chr11 | 22825514 | 22849435 |
| Complex I Assembly | ECSIT | VWZL01002810.1 | 58 | 899 |
| Complex I Assembly | FOXRED1 | chr18 | 496248 | 497754 |
| Complex I Assembly | NDUFAF8 | chr11 | 35723135 | 35723251 |
| Complex I Assembly | NUBPL | chr5 | 37535605 | 37620921 |
| Complex I Assembly | TIMMDC1 | chr1 | 100711012 | 100716333 |
| Complex I Assembly factor | NDUFAF1 | chr5 | 25656696 | 25667135 |
| Complex I Assembly factor | NDUFAF2 | chrZ | 21480725 | 21550552 |
| Complex I Assembly factor | NDUFAF3 | chr11 | 5944464 | 5945438 |
| Complex I Assembly factor | NDUFAF4 | chr3 | 81266277 | 81270192 |
| Complex I Assembly factor | NDUFAF5 | chr3 | 20745264 | 20751711 |
| Complex I Assembly factor | NDUFAF6 | chr2 | 137628460 | 137645237 |
| Complex I Assembly factor | NDUFAF7 | chr3 | 37597356 | 37603051 |
| Complex I subunit | NDUFA1 | chr7 | 10519095 | 10521402 |
| Complex I subunit | NDUFA10 | chr9 | 38765964 | 38807025 |
| Complex I subunit | NDUFA11 | chr21 | 1300634 | 1301642 |
| Complex I subunit | NDUFA12 | chr1 | 49789293 | 49800560 |
| Complex I subunit | NDUFA2 | chr6 | 30685512 | 30686734 |
| Complex I subunit | NDUFA3 | chr1 | 186086557 | 186086598 |
| Complex I subunit | NDUFA4 | chr2 | 29301172 | 29302917 |
| Complex I subunit | NDUFA4L2 | chr24 | 1867693 | 1867752 |
| Complex I subunit | NDUFA5 | chr1 | 24376180 | 24379480 |
| Complex I subunit | NDUFA6 | chr1 | 54316427 | 54318641 |
| Complex I subunit | NDUFA7 | chr21 | 1135743 | 1136522 |
| Complex I subunit | NDUFA8 | chr15 | 4066124 | 4068186 |
| Complex I subunit | NDUFA9 | chr1 | 77515286 | 77531165 |
| Complex I subunit | NDUFAB1 | chr8 | 7966716 | 7968478 |
| Complex I subunit | NDUFB1 | chr5 | 5790292 | 5791289 |
| Complex I subunit | NDUFB10 | chr8 | 6958479 | 6960235 |
| Complex I subunit | NDUFB2 | chr1 | 62728120 | 62728182 |
| Complex I subunit | NDUFB3 | chr9 | 29816709 | 29819522 |
| Complex I subunit | NDUFB4 | chr1 | 87324872 | 87326748 |
| Complex I subunit | NDUFB5 | chr6 | 2136677 | 2141904 |
| Complex I subunit | NDUFB6 | chr8 | 17852811 | 17855943 |
| Complex I subunit | NDUFB8 | chr10 | 19205477 | 19208810 |
| Complex I subunit | NDUFB9 | chr2 | 150542857 | 150546041 |
| Complex I subunit | NDUFC1 | chr4 | 13025067 | 13026075 |
| Complex I subunit | NDUFC2 | chr1 | 208670847 | 208671998 |
| Complex I subunit | NDUFS1 | chr9 | 31159234 | 31175488 |
| Complex I subunit | NDUFS2 | VWZL01000567.1 | 832 | 5105 |
| Complex I subunit | NDUFS3 | chr5 | 23559379 | 23563704 |
| Complex I subunit | NDUFS4 | chrZ | 18485872 | 18524541 |
| Complex I subunit | NDUFS5 | chr17 | 366038 | 367069 |
| Complex I subunit | NDUFS6 | chr2 | 93974075 | 93982691 |
| Complex I subunit | NDUFS7 | chr21 | 4015167 | 4018281 |
| Complex I subunit | NDUFS8 | chr5 | 68863080 | 68863223 |
| Complex I subunit | NDUFV1 | chr5 | 68990393 | 68992105 |
| Complex I subunit | NDUFV2 | chr2 | 107630354 | 107643000 |
| Complex I subunit | NDUFV3 | chr1 | 119075874 | 119083803 |
| Complex III Assembly | BCS1L | chr9 | 17693634 | 17698874 |
| Complex III Assembly | LYRM7 | chrZ | 48848575 | 48851782 |
| Complex III Assembly | PTCD2 | chrZ | 28033084 | 28046672 |
| Complex III Assembly | TTC19 | chr14 | 7855851 | 7859330 |
| Complex III Assembly | UQCC1 | chr13 | 17750856 | 17795891 |
| Complex III Assembly | UQCC2 | chr20 | 3440595 | 3444146 |
| Complex III subunit | CYC1 | chr1 | 202713181 | 202713384 |
| Complex III subunit | UQCR10 | chr13 | 19422380 | 19424109 |
| Complex III subunit | UQCR11 | chr21 | 498074 | 499506 |
| Complex III subunit | UQCRB | chr2 | 138096670 | 138101616 |
| Complex III subunit | UQCRC1 | chr11 | 2810688 | 2819164 |
| Complex III subunit | UQCRC2 | chr8 | 8867768 | 8878247 |
| Complex III subunit | UQCRFS1 | chr7 | 33325704 | 33326318 |
| Complex III subunit | UQCRH | chr12 | 24707879 | 24708052 |
| Complex III subunit | UQCRQ | chr6 | 42420895 | 42422065 |
| Complex IV assembly | APOPT1 | chr5 | 12442345 | 12449231 |
| Complex IV assembly | COA1 | chr2 | 56340269 | 56366711 |
| Complex IV assembly | COA3 | chr19 | 4560654 | 4560988 |
| Complex IV assembly | COA4 | chr1 | 206084268 | 206084651 |
| Complex IV assembly | COA5 | chr1 | 143797271 | 143800531 |
| Complex IV assembly | COA6 | chr3 | 43321053 | 43323027 |
| Complex IV assembly | COA7 | chr12 | 28207716 | 28209626 |
| Complex IV assembly | COX10 | chr11 | 28195428 | 28370435 |
| Complex IV assembly | COX11 | chr11 | 33510905 | 33512604 |
| Complex IV assembly | COX14 | chr24 | 323931 | 324086 |
| Complex IV assembly | COX15 | chr10 | 24500304 | 24503698 |
| Complex IV assembly | COX16 | chr5 | 29162115 | 29212725 |
| Complex IV assembly | COX17 | chr1 | 87728082 | 87730044 |
| Complex IV assembly | COX18 | chr4 | 1334265 | 1339351 |
| Complex IV assembly | COX19 | chr8 | 1995060 | 1996615 |
| Complex IV assembly | COX20 | chr3 | 39012435 | 39015872 |
| Complex IV assembly | FASTKD2 | chr9 | 30927430 | 30939274 |
| Complex IV assembly | LRPPRC | chr3 | 28331606 | 28413120 |
| Complex IV assembly | PET117 | chr3 | 8541111 | 8541269 |
| Complex IV assembly | ﻿SCO1 | chr11 | 25653925 | 25659270 |
| Complex IV assembly | SURF1 | chr15 | 1486114 | 1491007 |
| Complex IV subunit | COX4I1 | chr7 | 44214392 | 44217320 |
| Complex IV subunit | COX5A | chr8 | 19630559 | 19633417 |
| Complex IV subunit | COX6A1 | chr13 | 22202875 | 22204189 |
| Complex IV subunit | COX6B1 | VMED01000056.1 | 84990 | 85845 |
| Complex IV subunit | COX6C | chr2 | 139711407 | 139713271 |
| Complex IV subunit | COX7A2 | chr3 | 90229665 | 90232188 |
| Complex IV subunit | COX7A2L | chr3 | 26846934 | 26855108 |
| Complex IV subunit | COX7B | chr7 | 14499909 | 14500649 |
| Complex IV subunit | COX7C | chrZ | 77860056 | 77861314 |
| Complex IV translation | TACO1 | VWZL01009353.1 | 7 | 207 |
| Complex V Assembly Factor | ATPAF1 | chr12 | 25046214 | 25056158 |
| Complex V Assembly Factor | ATPAF2 | chr8 | 5458436 | 5463004 |
| Complex V Assembly Factor | TMEM70 | chr2 | 128635276 | 128636906 |
| Complex V subunit | ATP5F1A | chrZ | 4249695 | 4256103 |
| Complex V subunit | ATP5F1B | VWZL01005027.1 | 103 | 1354 |
| Complex V subunit | ATP5F1C | chr1 | 5010189 | 5016088 |
| Complex V subunit | ATP5F1D | chr21 | 3723061 | 3724452 |
| Complex V subunit | ATP5F1E | chr13 | 875235 | 875360 |
| Complex V subunit | ATP5G2 | chr24 | 818767 | 820026 |
| Complex V subunit | ATP5MC1 | chr19 | 7014586 | 7015772 |
| Complex V subunit | ATP5MC3 | chr9 | 36234382 | 36304056 |
| Complex V subunit | ATP5ME | chrZ | 51210444 | 51210524 |
| Complex V subunit | ATP5MF | chr8 | 4686483 | 4688795 |
| Complex V subunit | ATP5MG | chr18 | 7953504 | 7953831 |
| Complex V subunit | ATP5PB | chr20 | 6382203 | 6384698 |
| Complex V subunit | ATP5PD | chr11 | 31612689 | 31615218 |
| Complex V subunit | ATP5PF | chr1 | 111391211 | 111392707 |
| Complex V subunit | ATP5PO | chr1 | 114752258 | 114755497 |
| Replication | POLG | chr8 | 34358018 | 34368856 |
| Replication | SSBP1 | chr1 | 63175574 | 63178652 |
| Replication | TWNK | chr10 | 25918380 | 25922566 |
| Ribosomal subunit | DAP3 | chr23 | 1305744 | 1323070 |
| Ribosomal subunit | IMP3 | chr3 | 117918326 | 117918562 |
| Ribosomal subunit | MRPL1 | chr4 | 2576322 | 2589079 |
| Ribosomal subunit | MRPL10 | chr19 | 6502756 | 6504277 |
| Ribosomal subunit | MRPL12 | chr11 | 34853142 | 34856579 |
| Ribosomal subunit | MRPL13 | chr2 | 148826263 | 148843236 |
| Ribosomal subunit | MRPL14 | chr3 | 34077198 | 34080282 |
| Ribosomal subunit | MRPL15 | chr2 | 120341787 | 120349200 |
| Ribosomal subunit | MRPL16 | VMED01000072.1 | 45980 | 47631 |
| Ribosomal subunit | MRPL17 | chrZ | 11528324 | 11529060 |
| Ribosomal subunit | MRPL18 | chr3 | 51388159 | 51390513 |
| Ribosomal subunit | MRPL19 | chr3 | 116725033 | 116728343 |
| Ribosomal subunit | MRPL2 | chr3 | 7465566 | 7467102 |
| Ribosomal subunit | MRPL20 | chr16 | 6837283 | 6841165 |
| Ribosomal subunit | MRPL21 | chr5 | 69901625 | 69910452 |
| Ribosomal subunit | MRPL22 | chr6 | 48885668 | 48890856 |
| Ribosomal subunit | MRPL23 | chr5 | 52313137 | 52325111 |
| Ribosomal subunit | MRPL24 | VWZL01004084.1 | 75 | 224 |
| Ribosomal subunit | MRPL27 | chr11 | 31929168 | 31930781 |
| Ribosomal subunit | MRPL28 | chr8 | 16194534 | 16197295 |
| Ribosomal subunit | MRPL3 | chr2 | 46059493 | 46089324 |
| Ribosomal subunit | MRPL30 | chr1 | 144019771 | 144025072 |
| Ribosomal subunit | MRPL32 | chr2 | 55952335 | 55954465 |
| Ribosomal subunit | MRPL33 | chr3 | 30962735 | 30965085 |
| Ribosomal subunit | MRPL34 | chr21 | 4380645 | 4385269 |
| Ribosomal subunit | MRPL35 | chr4 | 47419035 | 47420386 |
| Ribosomal subunit | MRPL36 | chr2 | 93973184 | 93973294 |
| Ribosomal subunit | MRPL37 | chr12 | 29036434 | 29040401 |
| Ribosomal subunit | MRPL38 | chr11 | 30933505 | 30937495 |
| Ribosomal subunit | MRPL39 | chr1 | 111334748 | 111342771 |
| Ribosomal subunit | MRPL40 | chr13 | 33686285 | 33688139 |
| Ribosomal subunit | MRPL41 | chr15 | 11281095 | 11281499 |
| Ribosomal subunit | MRPL42 | chr1 | 49299644 | 49304702 |
| Ribosomal subunit | MRPL9 | VWZL01002321.1 | 7489 | 8960 |
| Ribosomal subunit | MRPS10 | chr3 | 25697419 | 25701183 |
| Ribosomal subunit | MRPS11 | chr8 | 34648940 | 34652014 |
| Ribosomal subunit | MRPS12 | VMED01000060.1 | 34666 | 35001 |
| Ribosomal subunit | MRPS14 | chr12 | 44445 | 47282 |
| Ribosomal subunit | MRPS15 | chr17 | 2179046 | 2184737 |
| Ribosomal subunit | MRPS16 | chr16 | 3255 | 3454 |
| Ribosomal subunit | MRPS17 | chr14 | 9973144 | 9974862 |
| Ribosomal subunit | MRPS18A | chr3 | 34550877 | 34568495 |
| Ribosomal subunit | MRPS18B | VWZL01000873.1 | 424 | 624 |
| Ribosomal subunit | MRPS18C | chr4 | 18292504 | 18293613 |
| Ribosomal subunit | MRPS2 | chr15 | 2682584 | 2685708 |
| Ribosomal subunit | MRPS21 | chr23 | 1804933 | 1805097 |
| Ribosomal subunit | MRPS22 | chr6 | 23036547 | 23042169 |
| Ribosomal subunit | MRPS23 | chr14 | 4858156 | 4860037 |
| Ribosomal subunit | MRPS24 | chr22 | 355305 | 356336 |
| Ribosomal subunit | MRPS25 | chr11 | 16885514 | 16886745 |
| Ribosomal subunit | MRPS26 | chr4 | 43992960 | 43993082 |
| Ribosomal subunit | MRPS27 | chrZ | 28051076 | 28098509 |
| Ribosomal subunit | MRPS30 | chrZ | 16567271 | 16571788 |
| Ribosomal subunit | MRPS31 | chr1 | 185553484 | 185572259 |
| Ribosomal subunit | MRPS33 | chr1 | 62833234 | 62833449 |
| Ribosomal subunit | MRPS34 | chr8 | 18208739 | 18210343 |
| Ribosomal subunit | MRPS35 | chr1 | 79406127 | 79427116 |
| Ribosomal subunit | MRPS36 | chrZ | 24550980 | 24555741 |
| Ribosomal subunit | MRPS5 | chr3 | 16050500 | 16107320 |
| Ribosomal subunit | MRPS6 | chr1 | 114852063 | 114863858 |
| Ribosomal subunit | MRPS7 | chr11 | 31496918 | 31499711 |
| Ribosomal subunit | MRPS9 | chr1 | 146774188 | 146801510 |
| Transcription | MTERF1 | chr1 | 59121390 | 59122580 |
| Transcription | POLRMT | chr21 | 3286904 | 3305710 |
| Transcription | TFAM | chr10 | 3094203 | 3103572 |
| Transcription | TFB1M | chr3 | 56957158 | 56983116 |
| Transcription | TFB2M | chr3 | 38166523 | 38177246 |
| tRNA processing | ELAC2 | chr11 | 26072304 | 26085423 |
| tRNA processing | PRORP | chr5 | 39288918 | 39339451 |
| tRNA processing | TRMT10C | chr1 | 93323430 | 93324674 |
| tRNA processing | HSD17B10 | VWZL01000864.1 | 205 | 609 |
| tRNA synthetase | AARS2 | chr3 | 33831813 | 33849830 |
| tRNA synthetase | CARS2 | chr1 | 150831071 | 150871129 |
| tRNA synthetase | DARS2 | chr12 | 2544024 | 2557454 |
| tRNA synthetase | EARS2 | chr8 | 8022100 | 8028203 |
| tRNA synthetase | FARS2 | chr2 | 71396942 | 71588983 |
| tRNA synthetase | FARSA | VWZL01004773.1 | 5360 | 7330 |
| tRNA synthetase | GARS1 | chr2 | 4733655 | 4761276 |
| tRNA synthetase | HARS2 | chr6 | 30710179 | 30734138 |
| tRNA synthetase | IARS2 | chr3 | 21194241 | 21223995 |
| tRNA synthetase | KARS1 | chr7 | 47190406 | 47199030 |
| tRNA synthetase | LARS2 | chr2 | 47196926 | 47282423 |
| tRNA synthetase | MARS2 | chr7 | 687056 | 694756 |
| tRNA synthetase | NARS1 | chrZ | 538256 | 549328 |
| tRNA synthetase | NARS2 | chr1 | 208851776 | 208877330 |
| tRNA synthetase | PARS2 | chr12 | 29143656 | 29145074 |
| tRNA synthetase | RARS2 | chr3 | 85307112 | 85345547 |
| tRNA synthetase | TARS2 | chr8 | 26446194 | 26458557 |
| tRNA synthetase | VARS2 | VWZL01008462.1 | 3 | 241 |
| tRNA synthetase | WARS2 | chr1 | 86058191 | 86080347 |
| tRNA synthetase | YARS2 | chr1 | 64871577 | 64877465 |

### **Table S8**. Z chromosome homologs of W chromosome-encoded genes examined to test for co-introgression with mitochondria and chrW. The positions of the genes identified within the scaffolded Stercorarius parasiticus assembly are provided, with introns included and with 5’ and 3’ untranslated regions excluded.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Protein** | ***Ficedula* locus** | **Chromosome** | **Start** | **Stop** |
| SMAD7 | ENSFALG00000008945 | chrZ | 3331904 | 3352777 |
| CTIF | ENSFALG00000008943 | chrZ | 3503065 | 3381118 |
| SMAD2 | ENSFALG00000008934 | chrZ | 3778952 | 3818086 |
| C18orf25 | ENSFALG00000008886 | chrZ | 4212186 | 4183301 |
| ATP5F1A | ENSFALG00000008875 | chrZ | 4249695 | 4256103 |
| UBAP2 | ENSFALG00000009935 | chrZ | 9505462 | 9436256 |
| DCAF12 | ENSFALG00000009894 | chrZ | 9575111 | 9551247 |
| UBAP1 | ENSFALG00000009887 | chrZ | 9602559 | 9612827 |
| FAM219A | ENSFALG00000009864 | chrZ | 9724795 | 9718670 |
| VCP | ENSFALG00000009780 | chrZ | 10664346 | 10647968 |
| GOLPH3 | ENSFALG00000002160 | chrZ | 11952193 | 11941430 |
| ZFR | ENSFALG00000002149 | chrZ | 12084272 | 12039942 |
| SUB1 | ENSFALG00000002143 | chrZ | 12096704 | 12104906 |
| NIPBL | ENSFALG00000002058 | chrZ | 13677084 | 13773871 |
| PRKAA1 | ENSFALG00000002013 | chrZ | 15272366 | 15249830 |
| RPL37 | ENSFALG00000002006 | chrZ | 15299631 | 15297090 |
| ZNF131 | ENSFALG00000011128 | chrZ | 16002095 | 16014688 |
| SNX18 | ENSFALG00000010790 | chrZ | 18850117 | 18869666 |
| MIER3 | ENSFALG00000010987 | chrZ | 19924775 | 19908350 |
| ZSWIM6 | ENSFALG00000011056 | chrZ | 21643956 | 21754190 |
| KIF2A | ENSFALG00000011073 | chrZ | 21998103 | 22032207 |
| SREK1 | ENSFALG00000009835 | chrZ | 23470675 | 23503113 |
| MRPS36 | ENSFALG00000009866 | chrZ | 24551831 | 24555741 |
| CERT1 | ENSFALG00000010073 | chrZ | 26663820 | 26729381 |
| TNPO1 | ENSFALG00000010292 | chrZ | 27873902 | 27823671 |
| MAP1B | ENSFALG00000010312 | chrZ | 28123305 | 28111049 |
| RFX3 | ENSFALG00000010515 | chrZ | 29612402 | 29552087 |
| CDC37L1 | ENSFALG00000010536 | chrZ | 30154552 | 30160752 |
| ZFAND5 | ENSFALG00000010733 | chrZ | 38232660 | 38221718 |
| ENSFALG00000003294.1 | ENSFALG00000003294 | chrZ | 40769470 | 40685751 |
| UBQLN1 | ENSFALG00000006359 | chrZ | 42881021 | 42867961 |
| HNRNPK | ENSFALG00000012478 | chrZ | 42986586 | 42972219 |
| ENSFALG00000012406.1 | ENSFALG00000012406 | chrZ | 47146675 | 47170734 |
| NFIL3 | ENSFALG00000014613 | chrZ | 48356066 | 48354693 |
| HINT1 | ENSFALG00000004845 | chrZ | 48862027 | 48863064 |
| RNF38 | ENSFALG00000000978 | chrZ | 49631168 | 49725121 |
| KCMF1 | ENSFALG00000005137 | chrZ | 51566024 | 51552717 |
| ZNF462 | ENSFALG00000002876 | chrZ | 53301831 | 53352567 |
| CHD1 | ENSFALG00000010499 | chrZ | 55703455 | 55663997 |
| FEM1C | ENSFALG00000002587 | chrZ | 62099736 | 62084858 |
| CKMT2 | ENSFALG00000012745 | chrZ | 75282705 | 75298535 |
| RASA1 | ENSFALG00000010471 | chrZ | 78052910 | 78092863 |
| ARRDC3 | ENSFALG00000009068 | chrZ | 79763458 | 79751787 |
| ENSFALG00000014649.1 | ENSFALG00000014649 | chrZ | 82795627 | 82796418 |

# Supplementary Note 1: Mitochondrial genome assemblies

Complete mtDNA genomes were assembled for each individual from the whole genome shotgun sequences using NOVOPlasty. Assemblies range from 18,951 bp to 19,070 bp (Table S9) and include the standard vertebrate set of 13 protein-coding genes, 2 rRNA genes, and 22 tRNA genes. The ND3 gene contains the untranslated nucleotide insertion that is present in the ND3 sequence of many birds (Mindell et al. 1998).

We found that a large region of the mitogenome has been duplicated in the Stercorariid mitogenomes. The duplication is approximately 2318 bp in length: it begins at position 581 of cytochrome b, extends through tRNAThr, tRNAPro, ND6, tRNAGlu, and terminates prior to the microsatellite repeat region. A duplication in this region has been noted in many other avian taxa (reviewed in (Mackiewicz et al. 2019)), including a member of Charadriiformes (Verkuil et al. 2010). The duplication is not present in two previously-assembled Stercorariid mitogenomes (*Stercorarius maccormicki:* KM401546.1, *S. parasiticus*: MN356186.1), likely owing to the difficulty of assembling duplicated regions with short-read data.

The duplication shows signs of concerted evolution between the two copies, as has been noted previously in other birds (eg. Sammler, Bleidorn and Tiedemann, 2011; Zhou *et al.*, 2014; Mackiewicz *et al.*, 2019): the two assembled copies of the duplication are nearly identical between paralogs within a single mitogenome, but differ between orthologs of different samples. The exception is a short region of approximately 180-190 bp that is divergent between the two copies of the duplication, flanked by regions that are identical between duplicates. The duplication also appears to be present in two Alcids (*Alca torda*: CM018102.1and *Uria aalge:* MN356418.1), and the breakpoints of the duplication appear to be shared between the Stercorariids and *Alca torda*,suggesting that this duplication could be ancestral to the Stercorariidae and Alcidae clade (*Uria aalge* appears to have subsequently lost much of the duplicated control region).

We confirmed the presence of the repeat by removing one copy of the duplication from the assembly and mapping raw sequencing reads to the single-copy assembly. The average sequencing depth outside of the duplication in *S. pomarinus* MKP1559 was approximately 28,800×, while sequencing depth was approximately 56,500× inside of the duplicated region. This strongly suggests that the region is present as two copies, and that both duplicates are present on the mitochondrion (as opposed to a numt in the nuclear genome, which would not lead to greatly elevated sequencing depth as the nuclear genome was sequenced to less than 20× coverage).

NOVOPlasty detected low-frequency mitochondrial variants in all samples. Variants can result from sequencing errors, NUMTs (nuclear copies of mitochondrial sequences which may map to the mitochondrial sequence), heteroplasmy (via somatic mutation or rare biparental inheritance), or contamination. Contamination could potentially mimic a signal of introgression, and would likely manifest as low-frequency mitochondrial variants that match the haplotype of another sample. A shared numt was detected in all Stercorariids at position 5849-6488 (relative to the sequence of *S. maccormicki*), encompassing a portion of ND4 and ND5, and all of tRNA-H, tRNA-S1, and tRNA-L1. The full sequence of the second numt was reconstructed with flanking regions that do not align to the mitogenome (indicating nuclear sequence); the sequence of this numt is given in Figure S15. Each sample also possessed additional unique low-frequency variants which may be sequencing errors or instances of heteroplasmy. No mtDNA contamination was detected. Since mtDNA is usually present at much higher copy number than nuclear DNA in bird tissue (Table S9) and would therefore be most likely to be sequenced in the event of contamination, this indicates that contamination is unlikely to affect this study.

**Table S9. Lengths of mitochondrial genomes generated in this study.** Note that the *S. maccormicki* sequence from GenBank is approximately 2.3 kb shorter than the samples in this study, as it contains only one copy of the duplicated control region and is presumably incomplete.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Species** | **Sample** | **mtDNA Length** | **Tissue type** | **Average Coverage** | **% of Sequencing reads** |
| *S. longicaudus* | MKP990 | 18984 | Muscle | 10038 | 0.6 |
| *S. parasiticus* | B-20730 | 18951 | Muscle | NA | 0.46 |
| *S. parasiticus* | SAMN12253778 | 18960 | Muscle | 12412 | 0.85 |
| *S. pomarinus* | MKP1559 | 19070 | Muscle | 32422 | 2 |
| *S. pomarinus* | ROMO513777 | 19031 | Muscle | 7741 | 0.66 |
| *S. pomarinus* | IB-2659 | 19002 | Muscle | 967 | 0.09 |
| *S. chilensis* | MKP2451 | 19025 | Muscle | 4076 | 0.42 |
| *S. maccormicki* | E23 (10) | 19019 | Blood | 100 | <0.01 |
| *S. maccormicki* | E67 (7) | 19021 | Blood | 65 | <0.01 |
| *S. maccormicki* | E68 (8) | 19012 | Blood | 88 | <0.01 |
| *S. maccormicki* | Genbank KM401546.1 | >16669 | NA | NA | NA |
| *S. antarcticus* | C54 (2) | 19004 | Blood | 70 | <0.01 |
| *S. antarcticus* | C55 (3) | 19022 | Blood | 84 | <0.01 |
| *S. antarcticus* | C72 (5) | 19022 | Blood | 83 | <0.01 |
| *S. skua* | MKP1592 | 19041 | Muscle | 11586 | 0.77 |
| *S. skua* | MKP1593 | 19034 | Muscle | 7499 | 0.95 |

|  |
| --- |
| >Stercorarius\_antarcticus\_NUMT  GCCTTTGTGTGTCGGCCGGAGGCTGGCACTTGGAGGGACAGGGTGTCGTGCATCCATAATGGATCAAAGCGGGAAATGAGTCTGCGAGTCAGAACCAATGAGATCCATGGCCAGGCCTGGCTGTGCCCATGGCGGTGGCTGGAGACCAGCATGRCCACAGCGTAGCTCARGTCAGACTKACACTGCATTTCACATCCCTGGTCCCCTTTTCCCTGATTGCAAGTACCAGATCCGGGTGAGCATCACCTGTGCTGGCCCATCTGGCAGCTGTGCTAAGAAGCTGGCCCCAAGAATCCAGAGAAACCTCCAGGTCTGTTTGCACCCTGCAGCGTCCACTTCCCATGAGTGTCAGGGTGATTGGAGTCCCCAGGGGTGACGTGGGGTCATTGGAATGGAATTGATAAAATAAGATCTTGTCCACTAATACAAACTACGAACACACGCACAGCCGCATCCTTCTCCTAACACGCAGCCTACAACCCCTACTACCACTCATAGCCACATGATGACTCCTAGCCAACTTAACAAACATGGCATTACCCCCAACCACAAACCTAATAGCAGAGCTAACAATTGTAATCGCACTATTCAACTCATCCTCCTTCAGAATCATCCTCACCGGAATCGCAACCCTACTGACTGCTTCCTACACCTTATTCATGCTGCTAATAACTCAAGGAGGGACACTCCCCACCCACATCACATCAATCCAAAACTCCAACACACGAGMACATCTCCTAATAACCCTCCACATCATCCCCCTACTCCTCCTAATCCTAAAACCAGAACTAATCTCAGGAATCCCCTCATGCAAGTATAGTTTTAACNCAAATATTAGACTGTGATTCTAAAAATAGAAGTTAAACTCTTCTTACCTGCTGAGGGGCAGTTCAATCAACAAGAACTGCTAATTCTTGTATCTGAGTTTAAGACCTCAGCCCCCTCACTTTTAAAGGATAATAGCAATCCATTGGTCTTAGGAACCACCCATCTTGGTGCAAATCCAAGTAAATGTAATGGAAGCCGCACTACTCCTTAATACCTCCATAATTTTAACACTCCAGCATGTCTCTGCTCTGAAGCAAAGCCTTTGCACAGTTGAGGGGGCCTTCAAAGCAGGTTTCTACTCTCAGCCCAAATGTGCCACAGCTGAGGTGCTGCAGCCCAGACCTGCCCCTGATGCCTTCAGCCTGGGGCTCAGAGAGGAGGAGCTGGAGCTGTGTGTGCAACCGCTTCCTTCCACYTGGAGGGAATAACAGCTGAGGGGTGCTGGGACAGCTCACATGCCTGGGCTGCTGTGATGGGGGGGTTCAGGATGATCAGGAGGCCAGAAAGGTCAGGAGGGTCTTGCCCTCAGTGTGAAGGAGCTGCTGAGATCCGCTCCTCTCTGGGATGAACGAGTTAGGTGAGCCCTTGTAGGTGAGTATCGGAGGAGAGGCTGATAGGGGTGACATTGKGGCGGGAGMCAAAGAACCTATTAAAGTCCCCTTGGCTGATG |

**Figure S15.** Nuclear mitochondrial sequence (numt) detected in *S. antarcticus*. This numt was shared by all Stercorariidae, and likely occurred in a common ancestor before their divergence.

# Supplementary Note 2: Mitochondrial Substitutions

To explore possible functional consequences of the mitochondrial introgression from *S. pomarinus* to *S. skua*, we assessed coding differences between the *S. pomarinus/skua* haplotypes and the southern hemisphere skua haplotypes — since *S. skua* appears to be very recently diverged from the southern hemisphere skuas based on nuclear data, it likely carried an ancestral haplotype that was similar to the modern southern hemisphere skuas.

The *S. pomarinus/S. skua* haplotypes differ from all southern hemisphere skuas with 13 amino acid substitutions (Table S10). We used the maximum likelihood topology to infer the most parsimonious Stercorariid ancestral state of each position, and found that 4 of these changes were likely derived in *S. pomarinus/S. skua* while 2 were identified as derived in the southern hemisphere taxa. Two sites contained a third allele in the small jaegers, and thus the ancestral state was classified as ambiguous.

The remaining five substitutions that differ between *S. skua/pomarinus* and the southern hemisphere species were also variable between *S. parasiticus* and *S. longicaudus*. Surprisingly, in all five cases, the southern hemisphere skuas carried the same allele as *S. longicaudus* while *S. skua/S. pomarinus* carried the same allele as *S. parasiticus*. While this pattern of convergence could be generated by chance (*p=*0.031), it is interesting to note that *S. longicaudus* breeds farther north in the high arctic than the other jaegers. It may be possible that selection for thermal adaptation in the high arctic and Antarctic has driven protein convergence between the southern hemisphere skuas and *S. longicaudus*.

**Table S10**. Derived substitutions in mitochondrially-encoded protein-coding and rRNA genes. The most parsimonious ancestral state was inferred based on the maximum likelihood topology. When *S. skua* and *S. pomarinus* differed from the southern hemisphere skuas, the alleles were compared to *S. parasiticus* and *S. longicaudus* to determine which allele is more likely to be the ancestral state. Cases were considered ambiguous when a single change of state could not be assigned unambiguously to either lineage, such as when *S. longicaudus* and *S. parasiticus* differed from each other or encoded a third amino acid. No nonsynonymous changes are observed amongst *S. skua, S. pomarinus, S. antarcticus, S. maccormicki,* or *S. chilensis* in ND3, ND4L, CO1, or CO2.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Gene** | **Derived in *S. skua*** | **Derived in *S. skua* and *S. pomarinus*** | **Derived in southern hemisphere skuas** | **Polymorphic in southern hemisphere skuas** | **Derived in *S. pomarinus*** | **Ambiguous, derived in (*pomarinus* + *skua*) or southern skuas** |
| ND1 | 1 | 0 | 0 | 0 | 0 | 1 |
| ND2 | 1 | 0 | 1 | 1 (antarcticus3,55) | 1 | 0 |
| ND4 | 2 | 0 | 1 | 1 (maccormicki026125) + 1 (maccormicki8,10) + 1 (maccormicki10) | 0 | 1 |
| ND5 | 0 | 1 | 0 | 0 | 1 | 4 |
| ND6 | 0 | 1 | 0 | 0 | 1 | 0 |
| Cytb | 0 | 0 | 0 | 1 (*chilensis*) + 2 (*chilensis*, antarcticus5,3, maccormicki026125) +1 (maccormicki8) | 0 | 0 |
| ATP6 | 0 | 1 | 0 | 1 (maccormicki026125) + 1 (maccormicki10) | 0 | 1 |
| ATP8 | 0 | 1 | 0 | 0 | 1 (MKP1559) | 0 |
| CO3 | 1 | 0 | 0 | 0 | 0 | 0 |
| Total amino acid | 5 | 4 | 2 | 10 | 4 | 7 |
| 12S rRNA | 0 | 1 | 3 | 0 | 0 | 1 |
| 16S rRNA | 2 | 4 | 3 | 3 (antarcticus2) + 1 (antarcticus3,55)+1 (maccormicki7,8,10)+1(maccormicki8) | 1 (MKP1559) | 1 |
| Total rRNA | 2 | 5 | 6 | 6 | 1 | 2 |

To test whether positive selection on derived alleles may have driven adaptive introgression of the mtDNA from *S. pomarinus* to *S. skua*, we tested for episodic diversifying selection on the four proteins that contained derived amino acid changes shared by *S. skua* and *S. pomarinus* (ND5, ND6, ATP6, and ATP8) using MEME (Murrell et al. 2012), which identifies sites that may have been subject to episodic positive selection on a given branch. Since this test is designed for between-lineage divergence data rather than within-population polymorphism data, we included one haplotype per population. The *S. skua/pomarinus*-specific substitutions in ND4 and ND5 were not predicted to be under positive selection, but the substitution in ATP6 was detected as likely evolving under positive selection in *S. pomarinus/skua* (p=0.044, EBF= 35202.78). This mutation (M1V) is a change in the start codon from Methionine and valine, a mutation that has also been observed in humans in ATP6, where it was sequenced in patients with suspected mitochondrial disorders but did not appear to alter ATP6 translation (Dubot et al. 2004). Experimental research will be required to determine whether the use of this alternative start codon has altered ATP6 transcription in *S. pomarinus* and *S. skua,* and whether the amino acid change has altered protein efficiency.

We further examined the potential impacts of the substitutions using PROVEAN (Conte et al. 2017), which predicts whether a protein substitution is likely to carry a functional impact based on its conservation in related taxa. Eight substitutions were predicted to be functionally non-neutral by PROVEAN:1 in *S. skua,* 2 shared by *S. skua* and *S. pomarinus,* 1 shared by all the southern hemisphere skuas, 1 shared by four southern hemisphere samples, 2 polymorphic in *S. maccormicki*, and 1 polymorphic in *S. pomarinus* (Table S11). These sites are flagged as functionally relevant due to their conservation in other bird lineages, and may represent deleterious mutations or lineage-specific positively selected mutations. The site predicted by MEME to be under positive selection in *S. skua/S. pomarinus* (M1V in ATP6) was also predicted to be non-neutral by PROVEAN (Table S11).

**Table S11**. PROVEAN scores and prediction of the deleteriousness of the observed mitochondrial substitutions, based on the level of conservation of the site in other taxa. Mutations are first tested using the *S. skua* allele as the reference allele (columns 2-4), and then tested using the *S. skua* allele as the mutant allele (columns 6-8). PROVEAN scores lower than -2.5 are likely to have functional impacts on the protein, and scores lower than -4.1 are highly likely to have functional impacts.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Gene** | **Mutation away from *S. skua*** | **PROVEAN score** | **Prediction** | **# sequences** | **Reverse mutation** | **PROVEAN score (reverse)** | **Prediction (reverse)** | **Derived taxon** |
| ND1 | F163L | -0.219 | Neutral | 6072 | L163F | 0.272 | Neutral | *S. skua* |
| ND1 | I182V | -0.561 | Neutral | 6072 | V182I | 0.564 | Neutral | Ambiguous |
| ND2 | T10I | -0.96 | Neutral | 11266 | I10T | 1.042 | Neutral | *S. pomarinus* |
| ND2 | G74S | 3.587 | Neutral | 11266 | S74G | -3.579 | **Non-neutral** | *S. skua* |
| ND2 | I195V | -0.199 | Neutral | 11266 | V195I | 0.204 | Neutral | Southern hemisphere skuas |
| ND2 | V328I | 0.641 | Neutral | 11266 | I328V | -0.645 | Neutral | *S. antarcticus* 3,5 |
| ND4 | M97V | 0.007 | Neutral | 6026 | V97M | -0.007 | Neutral | *S. skua* |
| ND4 | M101L | -0.112 | Neutral | 6026 | L101M | 0.112 | Neutral | Ambiguous |
| ND4 | T191A | 0.117 | Neutral | 6026 | A191T | -0.118 | Neutral | *S. skua* |
| ND4 | I314V | 0.031 | Neutral | 6026 | V314I | -0.031 | Neutral | Southern hemisphere skuas |
| ND4 | T337M | -4.313 | **Non-neutral** | 6026 | M337T | 4.312 | Neutral | *S. maccormicki 026125* |
| ND4 | I100T | -0.909 | Neutral | 6026 | T100I | 0.905 | Neutral | *S. maccormicki 10* |
| ND4 | T94I | -0.643 | Neutral | 6026 | I94T | 0.631 | Neutral | *S. maccormicki* 10,8 |
| ND5 | A3T | 1.351 | Neutral | 2139 | T3A | -1.271 | Neutral | Ambiguous |
| ND5 | T65I | -0.48 | Neutral | 2139 | I65T | 0.383 | Neutral | *S. skua & S. pomarinus* |
| ND5 | T277M | -1.153 | Neutral | 2139 | M277T | 1.149 | Neutral | Ambiguous |
| ND5 | L430M | -0.176 | Neutral | 2139 | M430L | 0.072 | Neutral | *S. pomarinus* |
| ND5 | T442I | -2.166 | Neutral | 2139 | I442T | 2.192 | Neutral | Ambiguous |
| ND5 | M540T | 0.046 | Neutral | 2139 | T540M | -0.069 | Neutral | Ambiguous |
| ND6 | I108V | -0.088 | Neutral | 1077 | V108I | 0.072 | Neutral | *S. skua & S. pomarinus* |
| ND6 | S116N | -0.341 | Neutral | 1077 | N116S | 0.34 | Neutral | *S. pomarinus* |
| ATP6 | V1M | 3.301 | Neutral | 5474 | M1V | -3.3 | **Non-neutral** | *S. skua & S. pomarinus* |
| ATP6 | S36N | 0.436 | Neutral | 5474 | N36S | -0.507 | Neutral | *S. maccormicki 026125* |
| ATP6 | T82M | -3.831 | **Non-neutral** | 5474 | M82T | 3.826 | Neutral | Ambiguous |
| ATP6 | V191I | 0.445 | Neutral | 5474 | I191V | -0.491 | Neutral | *S. maccormicki 10* |
| ATP8 | L36P | 4.983 | Neutral | 1127 | P36L | -6.045 | **Non-neutral** | *S. skua & S. pomarinus* |
| ATP8 | T53A | -2.905 | **Non-neutral** | 1127 | A53T | 2.891 | Neutral | *S. pomarinus MKP1559* |
| CO3 | S38N | 1.033 | Neutral | 4527 | N38S | -1.033 | Neutral | *S. skua* |
| cytb | Y110H | -3.088 | **Non-neutral** | 69462 | H110Y | 3.089 | Neutral | *S. chilensis,* *S. antarcticus 5,3, & S. maccormicki* 026125 |
| cytb | S226T | -0.801 | Neutral | 69462 | T226S | 0.593 | Neutral | *S. chilensis,* *S. antarcticus 5,3, & S. maccormicki* 026125 |
| cytb | A331T | -1.689 | Neutral | 69462 | T331A | 1.853 | Neutral | *S. chilensis* |
| cytb | G39S | -3.901 | **Non-neutral** | 69462 | S39G | 3.743 | Neutral | *S. maccormicki* 8 |

To determine whether any of the coding differences between *S. skua/pomarinus* and the southern hemisphere skuas likely cause structural impacts on the proteins, we modelled the protein structures of the mitochondrially-encoded proteins in *S. pomarinus* using Phyre2 (Kelley et al. 2015), and then assessed the impacts of each substitution using Missense3D (Ittisoponpisan et al. 2019). Most substitutions were predicted to have minimal impact on protein structure. Predicted impacts of substitutions in each protein are summarized below.

***NADH: Ubiquinone Oxidoreductase Core Subunit 1 (ND1)***

*S. skua* carries a derived substitution in ND1 (L163F) that is convergent with *S. longicaudus* and not shared with *S. pomarinus*. This substitution is not predicted to have a large effect on protein structure.

At position 182, *S. pomarinus, S. skua,* and *S. parasiticus* encode Isoleucine, while the southern hemisphere taxa and *S. longicaudus* encode Valine. This substitution appears to have occurred multiple times throughout Charadriiformes: for example, within Laridae *Rissa tridactyla* (DQ385076.1) encodes Valine while *Sterna eurygnatha* (DQ385078.1) encodes Isoleucine. This change is not predicted to impact protein structure.

***NADH: Ubiquinone Oxidoreductase Core Subunit 2 (ND2)***

The ND2 sequence of *S. skua* differs from all other samples, including *S. pomarinus*, by a Serine to Glycine substitution at position 74, buried within the interior of the protein within the third transmembrane alpha helix. This is predicted to have an insignificant impact on protein structure, expanding the cavity volume by 0.4 Å3,and breaking a potential hydrogen bond with a Valine at position 70 (a distance of 3.07 Å).

The *S. skua* haplotype also differs from the southern hemisphere taxa by a substitution at position 195 that it shares with *S. pomarinus* and the two small jaegers (*S. parasiticus* and *longicaudus*) and thus is presumably derived in the southern hemisphere taxa with a substitution from isoleucine to valine (I195V). This mutation is in a hydrogen-bonded turn on the exterior of the protein, and is not predicted to alter protein structure.

***NADH: Ubiquinone Oxidoreductase Core Subunit 4 (ND4)***

*S. skua* carries a V97M substitution that is presumably convergent with *S. longicaudus* and *S. parasiticus,* and is not predicted to impact protein structure. Both *S. skua* and *S. pomarinus* differ from the southern hemisphere skuas at position 101, where *S. skua* and *S. pomarinus* encode Methionine, while the southern hemisphere taxa encode Leucine and the small jaegers encode Threonine. This substitution is not expected to affect protein structure.

*S. skua* also carries a unique A191T substitution, which is convergent with several other Charadriiformes (e.g., *Himantopus mexicanus* (DQ385132.1), *Charadrius vociferus* (DQ385133.1), *Chionis minor* (DQ385136.1)). This is a change from a hydrophobic to a neutral amino acid, and is expected to have minimal impact on protein structure, although it is predicted to introduce a potential hydrogen bond with an amino acid at position 166.

The southern hemisphere skuas carry a substitution which is unique to the southern hemisphere taxa and invariant in the other available Charadriiformes sequences, I314V, but this was not detected as evolving under episodic positive selection by MEME (*p*=0.10). This substitution is not predicted to impact protein structure.

The T337M substitution unique to the GenBank *S. maccormicki* 026125sequence is flagged as potentially evolving under positive selection by MEME (*p=*0.04), however since this allele is not fixed in the *S. maccormicki* population it is more likely to represent a slightly deleterious allele which has yet to be removed by selection, or a sequencing error. This is a change from a neutral to a hydrophobic amino acid and is predicted to disrupt a hydrogen bond with Asparagine at position 333.

***NADH: Ubiquinone Oxidoreductase Core Subunit 5 (ND5)***

*S. pomarinus* and *S. skua* differ from the southern hemisphere skuas at five positions in ND5. The ancestral state of only one of these positions could be identified (I65T, derived in *skua/pomarinus* and convergent with several other Charadriiformes such as *Sterna eurygnatha*). One of these positions (442) appears quite variable in the Charadriiformes, and the small jaegers carry a different allele (Valine) than the *pomarinus/skua* haplotype (Threonine) and the southern hemisphere skuas (Isoleucine). Interestingly, of the remaining three substitutions, the southern hemisphere skuas encode the same allele as *S. longicaudus* and *S. pomarinus/S. skua* encode the same allele as *S. parasiticus*.

***NADH: Ubiquinone Oxidoreductase Core Subunit 6 (ND6)***

*S. skua* and *S. pomarinus* differ from the other Stercorariids at a Valine to Isoleucine mutation at position 108. This allele is presumably derived, as the two small jaegers carry the Valine allele. This substitution is not predicted to alter protein structure.

***ATP Synthase Membrane Subunit 6 (ATP6)***

The ATP6 sequence is variable at position 82, where the southern hemisphere skuas and *S. longicaudus* have a Methionine, while *S. skua, S. pomarinus*, and *S. parasiticus* have a Threonine. The Threonine allele is presumably ancestral as it is also found in an outgroup Alcid (*Fratercula arctica*) and Larids (*Rissa tridactyla, Rynchops niger, Sterna eurygnatha*). The Methionine allele is predicted to disrupt a hydrogen bond with a methionine residue at position 78 (at a distance of 2.8 Å) and is a change from a neutral to a hydrophobic amino acid, but is not predicted to greatly alter the protein structure. This position does not appear to be greatly conserved in Charadriiformes, as a methionine allele is also observed in *Charadrius alexandrinus* (GenBank accession MK830739.1), *Pedionomus torquatus* (DQ385242.1), *Burhinus grallarius* (DQ385236.1), and *Nycticryphes semicollaris* (DQ385240.1); a Lysine allele is observed in several Scolopacidae (e.g. *Tringa erythropus,* GenBank AY894260.1; *Tringa solitaria*, GenBank AY894269.1; *Heteroscelus brevipes*, GenBank AY894259.1; *Catoptrophorus semipalmatus,* GenBank AY894258.1); an Alanine allele is observed in *Chionis minor* (DQ385238.1); and a serine allele is found in *Numenius arquata* (EU826408.1).

The sequence also differs at position 1, where *S. skua* and *S. pomarinus* use a valine start codon (GTG) instead of methionine (ATG) which is found in the other Stercorariids as well as all other Charadriiform ATP6 sequences currently on GenBank. The structure of the protein could not be modelled at this position.

***ATP Synthase Membrane Subunit 8 (ATP8)***

At position 36 in ATP8, *S. skua* and *S. pomarinus* contain Leucine, while all other Stercorariids encode Proline. This mutation is not predicted to affect the protein structure and this position does not appear to be highly conserved, as *Fratercula* also (presumably convergently) encodes Leucine, while *Rissa tridactyla* encodes Histidine. Notably, all Stercorariids appear to have a 2-amino acid insertion (SN) at position 40 relative to *Rissa* and *Fratercula*. *S. pomarinus* MKP1559 differs from *S. skua* and the other Stercorariids at position 53 with a T53A substitution, but the structure of the protein in this region could not be predicted.

***Cytochrome b (Cytb)***

While there are no fixed differences between *S. skua* and the southern hemisphere skuas in Cytb, four of the southern skuas (*S. chilensis, S. antarcticus* 5,3, and *S. maccormicki* 026125) have substitutions Y110H and S226T, *S. chilensis* carries substitution A331T, and *S. maccormicki* 8 carries S39G. Since these substitutions have not yet reached fixation in the southern hemisphere skuas, they are unlikely to have been driven by positive selection (as they would then be unlikely to be observed in polymorphic state) and are more likely segregating neutral or nearly neutral variants. These substitutions are not predicted to cause a severe change in protein structure, although Y110H is a change in an exposed amino acid from an uncharged neutral to a positively charged hydrophilic amino acid that can form a hydrogen bond with the serine a position 107 (a distance of 3.1 Å). The linked S226T substitution is not expected to carry a large effect. The A331T substitution of *S. chilensis* is a change from a hydrophobic amino acid to a neutral amino acid, and is predicted to introduce a hydrogen bond with amino acids at position 327 or 328. The functional impact of these changes is unclear.

***Cytochrome c Oxidase Subunit 3 (COIII)***

*S. skua* has a N38S substitution in COIII that is not shared with any other Stercorariid, which is a change of an exposed amino acid from hydrophilic to neutral in the S-bend between the first two alpha helices of the protein. This is predicted to expand the volume of the cavity by 21.0 Å3, but is not predicted to greatly affect protein structure. This substitution is also observed in *Rhynchops niger* (DQ385213.1).

# References

Alonge M., Lebeigle L., Kirsche M., Aganezov S., Wang X., Lippman Z.B., Schatz M.C., Soyk S. 2021. Automated assembly scaffolding elevates a new tomato system for high-throughput genome editing. bioRxiv.:1–17.

Bernt M., Donath A., Jühling F., Externbrink F., Florentz C., Fritzsch G., Pütz J., Middendorf M., Stadler P.F. 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69:313–319.

Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120.

Bouckaert R., Heled J. 2014. DensiTree 2: Seeing Trees Through the Forest. bioRxiv.:1–11.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T.L. 2009. BLAST+: Architecture and applications. BMC Bioinformatics. 10.

Chandler R.M., Parmley D. 2003. The earliest North American record of auk (Aves: Alcidae) from the Late Eocene of central Georgia. Oriole. 68:7–9.

Chen S., Zhou Y., Chen Y., Gu J. 2018. Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 34:i884–i890.

Conte G.L., Hodgins K.A., Yeaman S., Degner J.C., Aitken S.N., Rieseberg L.H., Whitlock M.C. 2017. Bioinformatically predicted deleterious mutations reveal complementation in the interior spruce hybrid complex. BMC Genomics. 18:1–12.

Danecek P., Bonfield J.K., Liddle J., Marshall J., Ohan V., Pollard M.O., Whitwham A., Keane T., McCarthy S.A., Davies R.M., Li H. 2021. Twelve years of SAMtools and BCFtools. Gigascience. 10:1–4.

Delaneau O., Howie B., Cox A.J., Zagury J.F., Marchini J. 2013. Haplotype estimation using sequencing reads. Am. J. Hum. Genet. 93:687–696.

Dierckxsens N., Mardulyn P., Smits G. 2017. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45.

Douglas J., Jiménez-Silva C.L., Bouckaert R. 2022. StarBeast3: Adaptive Parallelized Bayesian Inference under the Multispecies Coalescent. Syst. Biol. 71:901–916.

Dubot A., Godinot C., Dumur V., Sablonnière B., Stojkovic T., Cuisset J.M., Vojtiskova A., Pecina P., Jesina P., Houstek J. 2004. GUG is an efficient initiation codon to translate the human mitochondrial ATP6 gene. Biochem. Biophys. Res. Commun. 313:687–693.

Green R.E., Krause J., Briggs A.W., Maricic T., Stenzel U., Kircher M., Patterson N., Li H., Zhai W., Fritz M.H.Y., Hansen N.F., Durand E.Y., Malaspinas A.S., Jensen J.D., Marques-Bonet T., Alkan C., Prüfer K., Meyer M., Burbano H.A., Good J.M., Schultz R., Aximu-Petri A., Butthof A., Höber B., Höffner B., Siegemund M., Weihmann A., Nusbaum C., Lander E.S., Russ C., Novod N., Affourtit J., Egholm M., Verna C., Rudan P., Brajkovic D., Kucan Ž., Gušic I., Doronichev V.B., Golovanova L. V., Lalueza-Fox C., De La Rasilla M., Fortea J., Rosas A., Schmitz R.W., Johnson P.L.F., Eichler E.E., Falush D., Birney E., Mullikin J.C., Slatkin M., Nielsen R., Kelso J., Lachmann M., Reich D., Pääbo S. 2010. A draft sequence of the neandertal genome. Science (80-. ). 328:710–722.

Hahn M.W., Hibbins M.S. 2019. A Three-Sample Test for Introgression. Mol. Biol. Evol. 36:2878–2882.

Han Y.D., Baek Y.S., Kim J.H., Choi H.G., Kim S. 2016. Complete mitochondrial genome of the South Polar Skua Stercorarius maccormicki (Charadriiformes, Stercorariidae) in Antarctica. Mitochondrial DNA. Part A, DNA mapping, Seq. Anal. 27:1783–1784.

Ittisoponpisan S., Islam S.A., Khanna T., Alhuzimi E., David A., Sternberg M.J.E. 2019. Can Predicted Protein 3D Structures Provide Reliable Insights into whether Missense Variants Are Disease Associated? J. Mol. Biol. 431:2197–2212.

Janssen K., Mundy N.I. 2017. The genetic basis and enigmatic origin of melanic polymorphism in pomarine skuas (Stercorarius pomarinus). Proc. R. Soc. B Biol. Sci. 284.

Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., Von Haeseler A., Jermiin L.S. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods. 14:587–589.

Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

Kelley L.A., Mezulis S., Yates C.M., Wass M.N., Sternberg M.J. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10:845–858.

Korneliussen T.S., Albrechtsen A., Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics. 15:1–13.

Langmead B., Salzberg S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9:357–359.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 27:2987–2993.

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

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25:2078–2079.

Mackiewicz P., Urantówka A.D., Kroczak A., Mackiewicz D. 2019. Resolving Phylogenetic Relationships within Passeriformes Based on Mitochondrial Genes and Inferring the Evolution of Their Mitogenomes in Terms of Duplications. Genome Biol. Evol. 11:2824–2849.

Maddison W.P., Maddison D.R. 2018. Mesquite: a modular system for evolutionary analysis. Version 3.40. http://www.mesquiteproject.org.

Malinsky M., Matschiner M., Svardal H. 2021. Dsuite - fast D-statistics and related admixture evidence from VCF files. Mol. Ecol. Resour. 21:584– 595.

Martin S.H., Davey J.W., Jiggins C.D. 2015. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. Mol. Biol. Evol. 32:244–257.

Mindell D.P., Sorenson M.D., Dimcheff D.E. 1998. An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles [2]. Mol. Biol. Evol. 15:1568–1571.

Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., Von Haeseler A., Lanfear R., Teeling E. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. 37:1530–1534.

Murrell B., Wertheim J.O., Moola S., Weighill T., Scheffler K., Kosakovsky Pond S.L. 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 8:e1002764.

Okonechnikov K., Conesa A., García-Alcalde F. 2016. Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 32:292–294.

Paradis E., Schliep K. 2019. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 35:526–528.

Quinlan A.R., Hall I.M. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 26:841–842.

R Core Team. 2018. R: A language and environment for statistical computing. .

Renaud G., Hanghøj K., Korneliussen T.S., Willerslev E., Orlando L. 2019. Joint estimates of heterozygosity and runs of homozygosity for modern and ancient samples. Genetics. 212:587–614.

Rice P., Longden L., Bleasby A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 16:276–277.

Robinson J.T., Thorvaldsdóttir H., Winckler W., Guttman M., Lander E.S., Getz G., Mesirov J.P. 2011. Integrative Genomics Viewer. Nat. Biotechnol. 29:24–26.

Sammler S., Bleidorn C., Tiedemann R. 2011. Full mitochondrial genome sequences of two endemic Philippine hornbill species (Aves: Bucerotidae) provide evidence for pervasive mitochondrial DNA recombination. BMC Genomics. 12:1–10.

Schliep K.P. 2011. phangorn: Phylogenetic analysis in R. Bioinformatics. 27:592–593.

Smith N.A. 2015. Sixteen vetted fossil calibrations for divergence dating of Charadriiformes (Aves, Neognathae). Palaeontol. Electron. 18.1.4FC:1–18.

Smith N.A., Clarke J.A. 2015. Systematics and evolution of the Pan-Alcidae (Aves, Charadriiformes). J. Avian Biol. 46:125–140.

Than C., Ruths D., Nakhleh L. 2008. PhyloNet: A software package for analyzing and reconstructing reticulate evolutionary relationships. BMC Bioinformatics. 9.

Verkuil Y.I., Piersma T., Baker A.J. 2010. A novel mitochondrial gene order in shorebirds (Scolopacidae, Charadriiformes). Mol. Phylogenet. Evol. 57:411–416.

Weir J.T., Mursleen S. 2013. Diversity-dependent cladogenesis and trait evolution in the adaptive radiation of the auks (Aves: Alcidae). Evolution (N. Y). 67:403–416.

Wen D., Yu Y., Zhu J., Nakhleh L. 2018. Inferring phylogenetic networks using PhyloNet. Syst. Biol. 67:735–740.

Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics. 19:15–30.

Zhou X., Lin Q., Fang W., Chen X. 2014. The complete mitochondrial genomes of sixteen ardeid birds revealing the evolutionary process of the gene rearrangements. BMC Genomics. 15:1–9.