doi:10.1111/evo.14120

Integrative genomic phylogeography reveals signs of mitonuclear incompatibility in a natural hybrid goby population

Shotaro Hirase,[1,2,3] Ayumi Tezuka,[4] Atsushi J. Nagano,[4] Mana Sato,[2] Sho Hosoya,[2] Kiyoshi Kikuchi,[2] and Wataru Iwasaki[1,5,6,7,8,9]

[1]*Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113-0032, Japan*

[2]*Fisheries Laboratory, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Hamamatsu, Shizuoka 431-0214, Japan*

[3]*E-mail: cashirase@g.ecc.u-tokyo.ac.jp*

[4]*Faculty of Agriculture, Ryukoku University, Otsu, Shiga 520–2194, Japan*

[5]*Atmosphere and Ocean Research Institute, The University of Tokyo, Kashiwa, Chiba 277–8564, Japan*

[6]*Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba 277–8561, Japan*

[7]*Institute for Quantitative Biosciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-0032, Japan*

[8]*Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, Bunkyo-ku, Tokyo 113-0032, Japan*

[9]*E-mail: iwasaki@bs.s.u-tokyo.ac.jp*

Received December 18, 2019
Accepted September 30, 2020

Hybridization between divergent lineages generates new allelic combinations. One mechanism that can hinder the formation of hybrid populations is mitonuclear incompatibility, that is, dysfunctional interactions between proteins encoded in the nuclear and mitochondrial genomes (mitogenomes) of diverged lineages. Theoretically, selective pressure due to mitonuclear incompatibility can affect genotypes in a hybrid population in which nuclear genomes and mitogenomes from divergent lineages admix. To directly and thoroughly observe this key process, we *de novo* sequenced the 747-Mb genome of the coastal goby, *Chaenogobius annularis*, and investigated its integrative genomic phylogeographics using RNA-sequencing, RAD-sequencing, genome resequencing, whole mitogenome sequencing, amplicon sequencing, and small RNA-sequencing. *Chaenogobius annularis* populations have been geographically separated into Pacific Ocean (PO) and Sea of Japan (SJ) lineages by past isolation events around the Japanese archipelago. Despite the divergence history and potential mitonuclear incompatibility between these lineages, the mitogenomes of the PO and SJ lineages have coexisted for generations in a hybrid population on the Sanriku Coast. Our analyses revealed accumulation of nonsynonymous substitutions in the PO-lineage mitogenomes, including two convergent substitutions, as well as signals of mitochondrial lineage-specific selection on mitochondria-related nuclear genes. Finally, our data implied that a microRNA gene was involved in resolving mitonuclear incompatibility. Our integrative genomic phylogeographic approach revealed that mitonuclear incompatibility can affect genome evolution in a natural hybrid population.

KEY WORDS: **Dobzhansky-Muller incompatibility, hybridization, microRNA, mitonuclear interactions, oxidative phosphorylation, secondary contact.**

Hybridization between divergent lineages generates new allelic combinations, which often lead to inferior hybrid fitness or hybrid breakdown through Dobzhansky-Muller incompatibility (Dobzhansky 1937; Muller 1940). Mitonuclear incompatibility is one of the mechanisms that can hinder the formation of hybrid populations via dysfunctional interactions between proteins encoded in the nuclear and mitochondrial genomes (mitogenomes) (Burton and Barreto 2012; Sloan et al. 2017; Hill 2019). Mitonuclear interactions are essential in eukaryotic metabolic processes, as exemplified by the frequent coevolution of genes across both genomes (Barreto et al. 2018). Hybridization upon secondary contact of divergent populations can lead to dysfunctional mitonuclear interactions, which are typified by those among the oxidative phosphorylation (OXPHOS) proteins encoded in both genomes (Burton and Barreto 2012; Hill et al. 2019). Previous studies based on cytoplasmic hybrid cell lines and crossing experiments have demonstrated that mitonuclear incompatibility can substantially decrease hybrid fitness (Wallace 1999; Ellison and Burton 2006; Baris et al. 2017; Hill 2017; Hill 2019; Hill et al. 2019). However, contrary to the aforementioned observations, there have also been reports that mitogenome introgression occurs with little or no detectable movement of nuclear genes between divergent lineages (Glémet et al. 1998; Wilson and Bernatchez 1998; Doiron et al. 2002; Zieliński et al. 2013; Pons et al. 2014; Good et al. 2015). Therefore, whether mitonuclear incompatibility plays a direct and general role in driving evolutionary processes remains unclear and requires further investigation (Burton and Barreto 2012; Hill 2017; Sloan et al. 2017; Hill et al. 2019). Importantly, Baris et al. (2017) found two nuclear genetic clusters associated with two types of mitogenomes in a hybrid population of two fish lineages, and suggested the existence of natural selection upon mitonuclear interactions. Hybrid populations with multiple divergent mitogenomes could serve as a natural genetic system for identifying signs of mitonuclear incompatibility (Rand 2017), and exhaustive whole-genome investigations of such hybrid populations would effectively find evidence of mitonuclear incompatibility in natural environments (Burton et al. 2013; Rand 2017; Healty and Burton 2020).

In a previous study on *Chaenogobius annularis* populations, we identified an ideal system for pursuing this question (Hirase and Ikeda 2015). *Chaenogobius annularis* is a goby species distributed throughout the rocky coastlines of Japan and Korea (Akihito et al. 2002). Its two geographic lineages, the Pacific Ocean (PO) and Sea of Japan (SJ) lineages, diverged 3.5 million years ago due to isolation events in the Sea of Japan, and have now formed a hybrid population on the Sanriku Coast through secondary contact (Hirase et al. 2012; Hirase et al. 2016; Hirase and Ikeda 2014, 2015). Here, to directly identify signals of mitonuclear incompatibility in this natural hybrid population, we *de novo* sequenced the genome of *C. annularis* and investigated its integrative genomic phylogeography through RNA-sequencing, RAD-sequencing, genome resequencing, whole mitogenome sequencing, amplicon sequencing, and small RNA-sequencing. We found mitogenomic and likely associated nuclear genomic changes in the hybrid population, suggesting that mitonuclear incompatibility can actually affect genome evolution in natural hybrid populations.

Materials and Methods

SAMPLE COLLECTION AND DNA EXTRACTION

From 2006 to 2017, we collected *C. annularis* individuals using a hand net from 16 sampling locations along the Japanese coastline (Fig. 1; Table S1). Some of the collected specimens were previously subjected to mitochondrial DNA sequencing and nuclear microsatellite markers genotyping (Hirase et al. 2012; Hirase and Ikeda 2014; Hirase and Ikeda 2015). DNA was extracted from clipped fins by the phenol/chloroform method (Asahida et al. 1996) or by using Gentra Puregene Tissue Kits (QIAGEN, Hilden, Germany), and these samples were stored at −30 °C prior to analysis.

DE NOVO GENOME SEQUENCING AND ASSEMBLY

For the *de novo* genome sequencing, DNA was extracted from one SJ lineage individual sampled at Aomori in 2015 (Fig. 1). Library construction and sequencing using the Illumina HiSeq 2500 and PacBio RS II sequencing platforms were carried out by Macrogen, Korea using standard protocols (Table S2). The insert sizes for mate-pair sequencing were 3000, 5000, and 8000 bp. Sixteen cells were used for PacBio RS II sequencing. The Illumina paired-end and mate-pair reads were filtered using the "*platanus_trim*" and "*platanus_internal_trim*" tools within the Platanus software package version 1.2.4 (Kajitani et al. 2014), respectively, with default parameters. Genomic contigs were constructed using filtered paired-end reads within Platanus using default settings. The generated contigs and the PacBio long reads were assembled using the DBG2OLC pipeline (Ye et al. 2016) with the following parameters: LD10, MinLen 200, KmerCovTh 2, MinOverlap 10, AdaptiveTh 0.001, and RemoveChimera 1. SSPACE (Boetzer et al. 2010), which were used to link the output scaffolds and form superscaffolds using the filtered mate-pair reads. Assembly completeness was evaluated using BUSCO version 3 (Simão et al. 2015) with default parameters.

RNA-SEQUENCING

Six individuals from the PO lineage and six individuals from the SJlineage were subjected to RNA-sequencing (Table S3). Brains, livers, gills, and muscle tissues were immediately stored in RNAlater (QIAGEN, Hilden, Germany) after dissection and were kept at −30 °C. We extracted total RNA using RNeasy

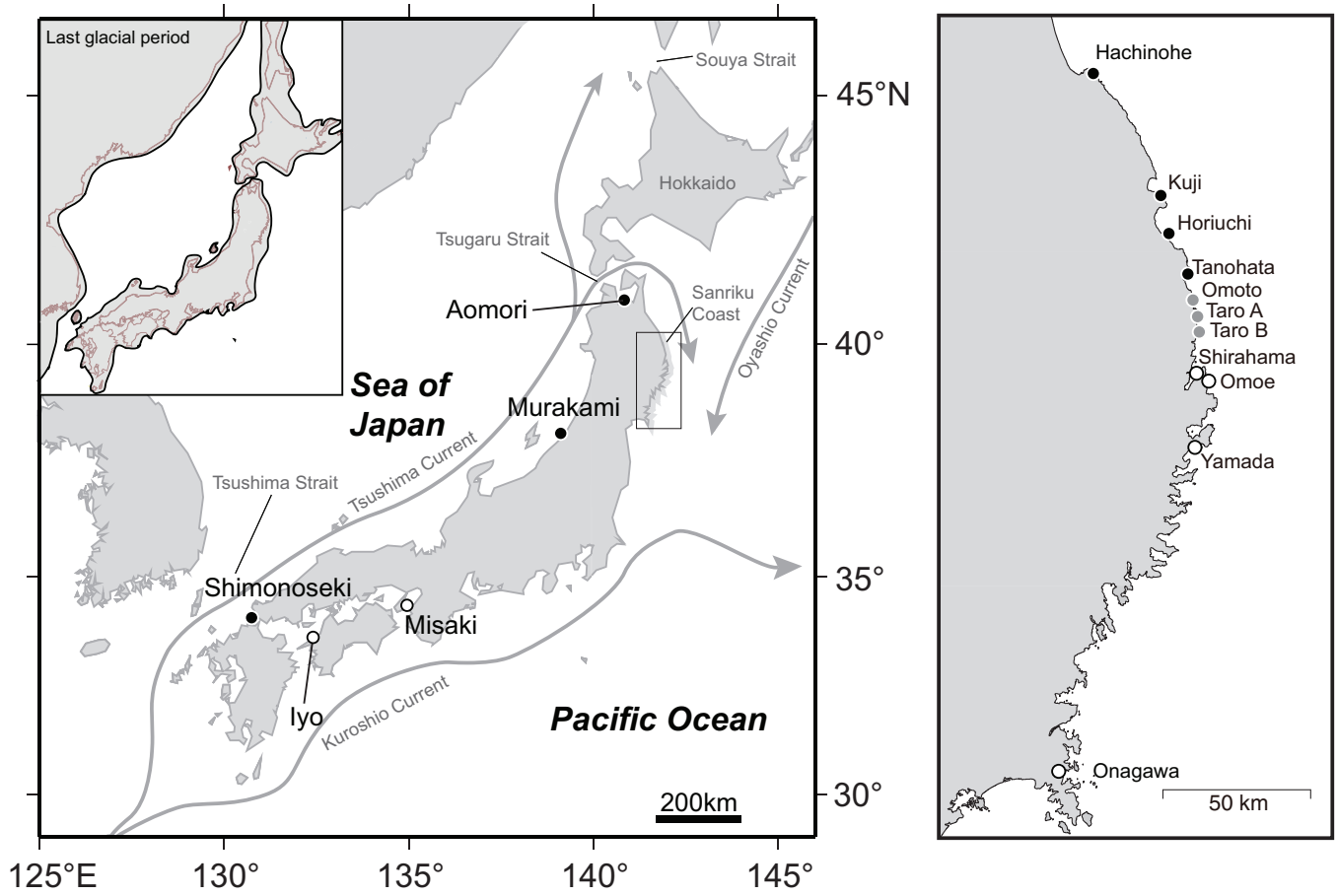

Figure 1. ***Chaenogobius annularis*** **sampling locations. The closed and open circles indicate populations that exclusively comprise individuals with Sea of Japan and Pacific Ocean lineage mitogenomes, respectively. In the enlarged Sanriku Coast map on the right, the gray circles indicate populations where both mitogenomes coexist. Coastlines during the last glacial period are shown in the top-left inset map and were taken from Crusius et al. (1999).**

Mini Kit (QIAGEN, Hilden, Germany). Library preparation was conducted following the methodology of Nagano et al. (2015). Then, 100-bp sequences from both sides of the DNA fragments were sequenced using Illumina HiSeq 2000. To remove the adapter and other Illumina-specific sequences and to cut low-quality regions, the RNA-sequencing reads were filtered using Trimmomatic version 0.32 (*-phred33 ILLUMINACLIP TruSeq3-SE.fa:2:30:10 LEADING:19 TRAILING:19 SLIDINGWINDOW:30:20 AVGQUAL:20 MINLEN:51*) (Bolger et al. 2014).

GENOME ANNOTATION AND DIFFERENTIAL EXPRESSION ANALYSIS

For gene prediction, we used *ab initio* and transcript-based methods with the unmasked genome sequence to avoid missing coding genes. The *ab initio* gene prediction was performed using AUGUSTUS version 2.5.5 (Stanke and Waack 2003), whose gene model parameters were trained using core eukaryotic genes in the *C. annularis* genome predicted by CEGMA version 2.5 (Parra et al. 2007). PASA version 2.0.2 pipeline (Haas et al. 2003) was used for transcript-based gene predictions, where the transcript sequences were constructed by feeding all RNA-sequencing data through to genome-guides assembly using the Trinity (release 17 July 2014) (Grabherr et al. 2011) and TopHat version 2.1.1 (Trapnell et al. 2009) software packages. Genes predicted by AUGUSTUS and that were supported by PASA version 2.0.2 pipeline were used in this study. Gene functions were annotated using TBLASTN searches (*E*-value cutoff: 10^{-10}) against stickleback, fugu, Atlantic cod, and medaka protein sequences. Genes likely to be related to mitonuclear incompatibility and OXPHOS genes were subsequently identified by BLASTP searches against the NCBI *nr* database followed by manual checking.

For differential expression analysis, reads from brain tissues were aligned to the genome sequence with TopHat (*–min-intron-length 40–max-intron-length 10,000*). Differentially expressed genes were retrieved using the "Cuffmerge" and "Cuffdiff" pipelines within Cufflinks version 2.2.1 (*–min-intron-length 40–max-intron-length 10,000*) (Trapnell et al. 2012).

RAD-SEQUENCING

Double-digest RAD libraries (Peterson et al. 2012) were prepared according to Peterson's protocol with slight modifications (Sakaguchi et al. 2015), and were sequenced using Illumina HiSeq 2500 to generate single-end 51-bp reads. Genomic DNA was digested using EcoRI and BglII (New England Biolabs, Beverly, MA, USA). To remove Illumina adapters and low-quality regions, the reads were filtered using Trimmomatic (*-phred33 ILLUMINACLIP TruSeq3-SE.fa:2:30:10 LEADING:19 TRAILING:19 SLIDINGWINDOW:30:20 AVGQUAL:20 MINLEN:51*). The filtered RAD-sequencing reads were aligned to the *C. annularis* genome sequence using the BWA-MEM algorithm in BWA version 0.7.8 (Li and Durbin 2009) with default settings. To avoid the effects of potential PCR duplicates, if multiple reads were aligned to the same position, all reads except for those of the highest mapping quality were removed using SAMtools version 0.1.18 (*rmdup -s*) (Li et al. 2009). SNPs were identified using "mpileup (*-Q 20*)" and "bcftools" within SAMtools, and were filtered using VCFtools version 0.1.12 (*–minQ 20–remove-indels–maf 0.05–max-alleles 2–min-meanDP 15–minDP 10*) (Danecek et al. 2011). SNP loci that were not present in 20% of the individuals in more than 10 populations were filtered out using the "pop_missing_filter.sh" tool within the dDocent version 2.2.7 software package (Puritz et al. 2014). Finally, to eliminate potential genotyping errors, SNP loci that showed divergence from the Hardy-Weinberg equilibrium (HWE; $P < 0.05$) in at least one population were removed using the "script filter_hwe_by_pop.pl" tool within the dDocent package. Due to the fact that this HWE filtering could have excluded SNP loci that are under strong selection, we confirmed that most of the excluded SNP loci showed excess heterozygosity, which implies genotyping errors (Chen et al. 2017), and that excluding this filtering did not affect our conclusions.

WHOLE GENOME RESEQUENCING

Genome resequencing was performed for six individuals of the SJ lineage and six individuals of the PO lineage (Table S4). Genome resequencing data were generated in Illumina HiSeq X Ten using the 250-bp paired-end protocol. All library construction and sequencing were carried out at BGI, China. The paired-end reads were filtered by Trimmomatic to remove the adapter and other Illumina-specific sequences, as well as low-quality regions (*-phred33 ILLUMINACLIP TruSeq3-SE.fa:2:30:10 LEADING:19 TRAILING:19 SLIDINGWINDOW:30:20 AVGQUAL:20 MINLEN:51*). The filtered paired-end reads were aligned to the *C. annularis* genome sequence using the BWA-MEM algorithm in BWA with default settings, followed by local realignment around indels using GATK version 3.7 (McKenna et al. 2010) with default settings. Then, SNPs were identified using "mpileup (*-Q 30*)" and "bcftools" within SAMtools, and were filtered using the "vcfutils.pl varFilter" within SAMtools and VCFtools (*–max-missing 1–minQ 20*) with default settings. We identified SNP sites that were fixed between the SJ and PO lineages within OXPHOS genes by calculating F_{ST} values in VCFtools ($F_{ST} = 1$). Genome-wide differentiation between the two lineages was investigated by calculating the F_{ST} values for nonoverlapping 15-kb sliding windows using VCFtools.

MITOGENOME SEQUENCING

To determine the mitogenome sequences of many *C. annularis* individuals, we performed parallel mitogenomic sequencing with the long PCR technique. *Chaenogobius annularis* mitogenomes were amplified using two sets of complementary primers (30 bp each) that were designed for the *C. annularis* tRNA-Lys and tRNA-Thr genes. The primer sequences were CAMT-Lys-L (5′-AGCATCAGCCTTTTAAGCTGAAAGTTGGTG-3′) and CAMT-Lys-H (5′-CACCAACTTTCAGCTTAAAAGGCTGATGCT-3′) for tRNA-Lys, and were tRNA-Thr-L (5′-GCCGGAGGTTAAAATCCTCCCTATTGC-3′) and tRNA-Thr-H (5′-GCAATAGGGAGGATTTTAACCTCCGGC-3′) for tRNA-Thr. Long PCR and library preparations were performed as described by Hirase et al. (2016), except for the use of dual indexed adapters and the NEBNext Ultra II DNA library prep Kit for Illumina in this study (New England BioLabs, Beverly, MA, USA). The dual indexed libraries were sequenced using Illumina MiSeq at Bioengineering Lab. Co., Ltd., Japan.

Mitogenomic reads were mapped to the reference sequences of the corresponding lineages using GS Reference Mapper version 3.7 (Roche, Basel, Switzerland) with the default settings. Nucleotides with quality scores of less than 30 were categorized as "N." In the following analyses, only mitogenomes that were longer than 16,000 bases and contained <3% of N bases were used. For each individual, two contigs from two long PCRs were concatenated and the coordinates were adjusted to place the $tRNA^{Phe}$ gene at the first position. The mitogenomic sequences were subjected to multiple alignment containing published mitogenomic sequences from several *C. annularis* populations (Hirase et al. 2016) and a mitogenomic sequence of *Chaenogobius gulosus* (NC_027193; a sister species of *C. annularis*) using the FFT-NS-2 algorithm within MAFFT version 7.058b (Katoh and Standley 2013). We annotated the mitogenome sequences using the method outlined in Hirase et al. (2016). Individuals with the same mitogenome sequences were grouped using Phylogears2 (Tanabe 2008).

Long PCR products were also used to sequence partial *ND4* genes with the BigDye Terminator Cycle Sequencing Kit version 3.1 (Thermo Fisher Scientific, Waltham, MA, USA) and an ND4-ago primer (5′-ATGGGACAGATCGCCTC-3′). Cycle sequencing reactions were purified by the ethanol/EDTA/sodium acetate method outlined in the BigDye Terminator version 3.1

Cycle Sequencing Kit manual and were analyzed in an ABI 3130 capillary sequencer (Thermo Fisher Scientific, Waltham, MA, USA).

AMPLICON SEQUENCING

Divergent SNP sites within the OXPHOS genes between the PO and SJ lineages were genotyped for the Taro B population using Ion AmpliSeq technology (Thermo Fisher Scientific, Waltham, MA, USA) with slight modifications (Sato et al. 2019). Ion AmpliSeq primer pools were designed by the Ion AmpliSeq Designer (Thermo Fisher Scientific, Waltham, MA, USA). Multiplex PCR amplification was performed using 20-ng genomic DNA using Multiplex PCR Assay Kit version 2 (Takara, Shiga, Japan) and the designed primers. The PCR conditions comprised an initial denaturation at 94 °C for 1 min, followed by 30 cycles of denaturation at 98 °C for 10 s, and annealing and extension at 68 °C for 10 min. The PCR products were purified using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA) and DNA concentrations were measured using Qubit dsDNA BR Assay Kits (Thermo Fisher Scientific, Waltham, MA, USA). The purified PCR products were digested with the USER enzyme (New England Biolabs, Beverly, MA, USA) at 37°C for an hour and were purified using ×1.8 volume Agencourt AMpure XP (Beckman Coulter, Brea, CA, USA). Approximately 10-200 ng of the PCR product was then end-repaired and 3′-adenylated, ligated with Illumina adapters, and PCR-enriched with Illumina sequencing dual indexes using the NEBNext Ultra II DNA library prep Kit for Illumina (New England BioLabs, Beverly, MA, USA). The dual indexed library was sequenced by Illumina MiSeq at the Bioengineering Lab. Co., Ltd., Japan. The reads were aligned to the *C. annularis* genome using the BWA-MEM algorithm in BWA. Only alignments located 600-bp around the targeted SNP sites were considered. The SNPs were identified using SAMtools mpileup (*-Q 30*) and were filtered using VCFtools (*–minGQ 15*).

SMALL RNA-SEQUENCING

Four individuals sampled in the hybrid zone (two at Tanohata and two at Shirahama; Fig. 1) were used for small RNA-sequencing. Their brains were immediately stored in RNAlater (QIAGEN, Hilden, Germany) after dissection and were kept at −30 °C. All library construction and sequencing were carried out at Macrogen Inc., Korea. Briefly, Total RNA was extracted using Trizol Thermo Fisher Scientific, Waltham, MA, USA), and small RNA libraries were prepared using TruSeq Small RNA Library Prep Kits (Illumina, San Diego, CA, USA). The libraries were sequenced using Illumina HiSeq 2500 to generate single-end 51-bp reads. We extracted reads that were similar to the RAD-tag sequence, including the 20th, 21st, and 22nd SNP loci, using Magic-BLAST version 1.3.0 (Boratyn et al. 2018). Then, the extracted small RNA-sequencing reads were aligned to the *C. annularis* genome using BWA aln/samse with default parameters. MiRBoost (Tempel et al. 2015) equipped with the cross-species option was used to estimate whether the genomic regions to which the small RNA-sequencing reads were aligned encoded microRNA (miRNA) genes.

POPULATION GENETICS ANALYSES

The population structure in the hybrid zone was investigated based on the nuclear genome SNP data obtained by RAD-sequencing. The genetic diversity of each population (effective numbers of alleles and observed heterozygosity) was estimated using GenoDive version 2.0b27 (Meirmans et al. 2004). We performed clustering analysis using ADMIXTURE (Alexander et al. 2009) with K assumption ranging 1 to 10 with default parameters. To remove the effect of linkage disequilibrium, we removed SNP loci that showed high r^2 values ($r^2 > 0.5$) with other sites, which were calculated using PLINK version 1.07 (Purcell et al. 2007). Cross validation scores were used to determine appropriate K values (Alexander and Lange 2011). The level of contemporary gene flow was estimated using BayesAss3-SNPs (Wilson and Rannala 2003).

Biased introgression of mitogenomes was assessed using BGC version 1.0 (Gompert and Buerkle 2012) based on the mitogenomic haplotypes and 319 SNP sites that showed divergence between the two lineages ($F_{ST} = 1$). Missing genotype data were filled in based on the population allele frequency data using GenoDive. Genotype frequencies in the Iyo, Misaki, and Onagawa populations were used to represent the PO lineage, and those in the Shimonoseki, Murakami, and Aomori populations were used to represent the SJ lineage. All individuals sampled between Tanohata and Omoe were pooled into one admixed population. Mitogenomic genotypes were coded as diploid homozygotes, as in Trier et al. (2014). We ran three independent chains of 50,000 steps each and recorded samples from the posterior distributions every 20th step following a 20,000-step burn-in. When the 95% credibility intervals of the α and β cline parameters did not cross zero, we identified them as loci with biased introgression.

Genetic differentiation within the Taro B population and hybrid zone was assessed by k-means clustering using GenoDive, ADMIXTURE, and principal component analysis (PCA) using pcadapt version 4.0 (Luu et al. 2017). For the Taro B population, we used 47 individuals with available RAD and amplicon-sequencing data. Genes located within ±50,000 bp of the SNP loci (50,000 bp before and after) were identified as genes likely to be related to mitonuclear incompatibility. Functional enrichment analysis was carried out using DAVID version 6.8 (Huang et al. 2008) with human Ensembl Gene ID. Enriched terms with EASE score of 0.2 or less were retrieved. Subcellular protein localization prediction was performed by TargetP 1.1 Server with default settings (Emanuelsson et al. 2000). miRNA gene coordinates and

their secondary structures were predicted using miRNAFold (Tav et al. 2016) with default settings.

MITOGENOME ANALYSES

MEGA 5.05 (Tamura et al. 2011) with 95% cutoff partial deletion, Jukes-Cantor distance, and 1000 bootstrap replications was applied to the mitogenome alignment to reconstruct neighbor-joining (NJ) trees (Saitou and Nei 1987). It should be noted that NJ trees based on Jukes-Cantor distance are reportedly good at estimating phylogenetic relationships among closely related lineages (Nei and Kumar 2000; Hirase et al. 2016). The K_a/K_s ratios were calculated for 13 mitochondrial genes by pairwise comparisons using DNAsp version 5.10 (Librado and Rozas 2009). To avoid dividing by 0, one nonsynonymous mutation was added to nonsynonymous substitutions identified by DNAsp, as in Mishmar et al. (2003). Nucleotide sites under positive selection for each gene were extracted using TreeSAAP version 3.2 (Woolley et al. 2003) with the whole mitogenome tree. We only considered the most radical amino acid property changes scoring between 6 and 8 and with z-scores < 0.001. Mitochondrial protein domains were predicted using a hidden Markov model in TMHMM Server version 2.0 (Krogh et al. 2001) with default settings.

Results

DE NOVO GENOME SEQUENCING AND RNA-SEQUENCING OF *Chaenogobius annularis*

We *de novo* sequenced the genome of an SJ-lineage *C. annularis* individual sampled in Aomori, Japan (Fig. 1). A total of 164 Gb Illumina paired-end reads were used for contig assembly, 20 Gb PacBio reads were used for subsequent hybrid assembly, and 118 Gb Illumina mate-pair reads were used for final scaffolding (Table S5). The 747-Mb final assembly had an N50 scaffold size of 0.9 Mb and consisted of 1897 scaffolds (Table S5). We applied BUSCO to evaluate genomic completeness by searching for a set of highly conserved *Actinopterygii* orthologs, and found 4098 genes (89.4%).

For genome annotation and comparative transcriptomics, we conducted RNA-sequencing of the brain, liver, muscle, and/or gill tissues of six PO-lineage and six SJ-lineage individuals (Table S3). Combined with the *ab initio* method, 31,692 protein-coding genes were predicted in the genome. Among these, 25,010 protein-coding genes were successfully annotated through comparisons with protein-coding genes of other fish species. For comparative purposes, brain tissues were targeted because the samples were readily available, and because brain tissue contains many mitochondria, and is particularly vulnerable to mitochondrial dysfunction (Calabrese et al. 2001). Comparative transcriptomics of the brains of the six PO-lineage and six SJ-lineage individuals revealed significantly different expression levels for 4.3% of the annotated genes (false discovery rate [FDR] < 0.05), where 1.8% and 2.5% were overexpressed in the PO and SJ lineages, respectively.

COEXISTENCE OF TWO MITOGENOMES IN A STABLE HYBRID POPULATION IN A SECONDARY CONTACT ZONE

To reveal admixture patterns in the nuclear genomes of the two lineages upon secondary contact on the Sanriku Coast, RAD-sequencing was performed for 267 individuals from 16 locations around the Japanese archipelago (Fig. 1). The average numbers of quality-filtered reads and total bases per individual were 3.1 million (range: 1.1-8.9 million) and 0.16 Gb (range: 0.06-0.45 Gb), respectively. Mapping the RAD-sequencing reads to the genome detected 5300 SNP loci (minor allele frequency >0.05).

Genetic diversity peaked in the Taro B population on the Sanriku Coast (Fig. S1). Clustering analysis by ADMIXTURE assigned the PO- and SJ-lineage individuals to different clusters and detected their nuclear genomic admixture in Taro B and in neighboring populations (Fig. 2A). Notably, all individuals from Taro B had well-admixed nuclear genomes, where both genomes are mixed at a ratio of approximately 50% ($K = 2$; Hybrid indices of individuals between the PO and SJ lineages are shown in Fig. S2). The cross-validation error plateaued around $K = 5$ (Fig. S3), at which point the Taro B and neighboring populations tended to be assigned to clusters outside of the PO and SJ lineages (Fig. 2A), suggesting that the genomic compositions of these hybrid populations are distinct from both lineages. Limited immigration (approximately 1% per generation) among the Taro B population and the neighboring populations was estimated (Table S6). A triangle plot showed that all hybrids in the Taro B population had low interpopulation heterozygosity (Fig. 2B). Note that, although an F1 hybrid has a hybrid index of 0.5 and is heterozygous at all divergent genomic loci (i.e., interpopulation heterozygosity $= 1.0$), an individual from a stable hybrid population has a hybrid index of 0.5 and low heterozygosities due to fixation after drift and/or selection (Gompert et al. 2014). Collectively, these results indicate that the Taro B population is composed of late hybrid generations.

Next, we conducted mitogenomic sequencing of the *C. annularis* individuals and visualized its introgression around the secondary contact zone (Fig. 2C). We observed a steep cline in the ratios of the two mitogenomes around Taro B. Bayesian Genomic Cline (BGC) analysis showed a high β value of the mitogenomes ($\beta = 2.6$; top 5% of the β distributions), which demonstrated that the clines are significantly steeper than the neutral expectation. This observation indicates the existence of mitonuclear incompatibility and/or different migration behaviors

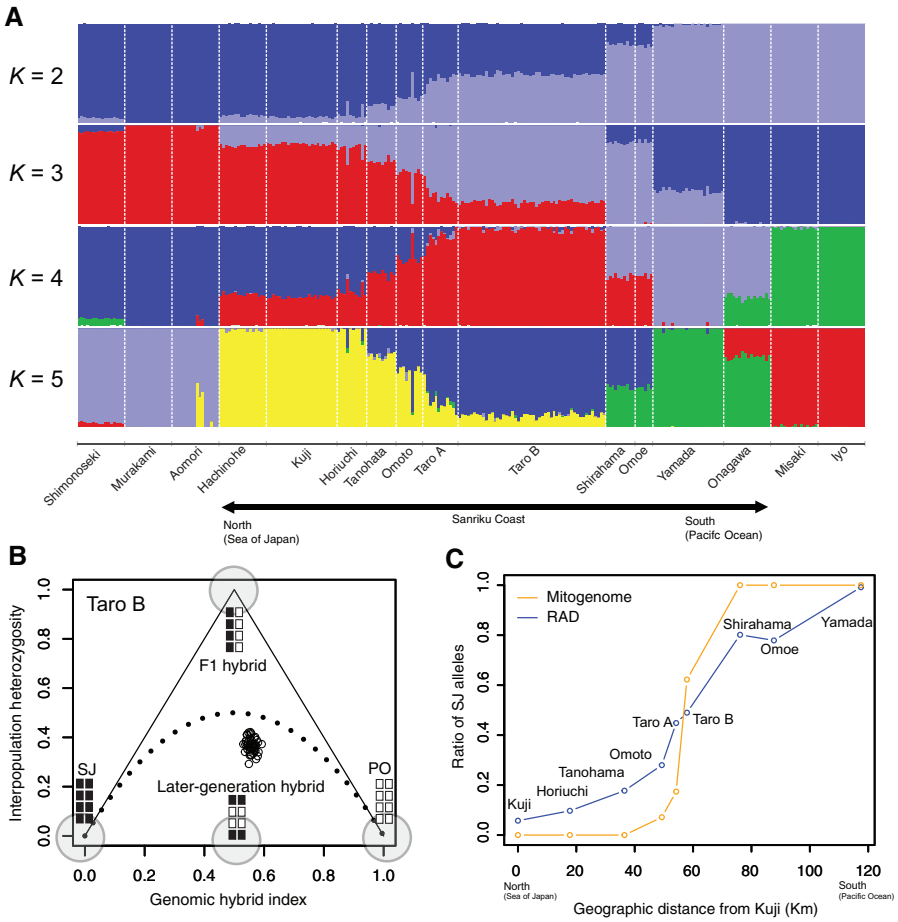

Figure 2. **Genetic characteristics of the hybrid populations. (A) Individual admixture proportions (*q* values) estimated by ADMIXTURE and SNP data obtained by RAD-sequencing (*K* = 2-5). (B) The relationships between the genomic hybrid indices and interpopulation heterozygosities of the Taro B individuals. The genomic hybrid indices are 0 and 1 for the SJ- and PO-lineage nuclear genomes, respectively. (C) Geographic clines of the ratios of SJ mitogenomes (orange) and the average *q* values at *K* = 2 (blue). The *x*-axis shows geographic distances from the Kuji population on the north side (Fig. 1).**

between males and females (Toews and Brelsford 2012); we could not assess if this was the case because a sex-identification method for *C. annularis* has not yet been established. Regardless of the cause, the two types of mitogenomes from the PO and SJ lineages still coexist in the Omoto, Taro A, and Taro B populations.

Next, we questioned whether the coexistence of the two mitogenomes affected nuclear genotypes in the late hybrid populations under limited immigration. For an in-depth analysis, we performed a PCA of the RAD locus data of 47 individuals from Taro B (Fig. 3A). The nuclear genomes of individuals that had PO (named PO-TaroB-major and PO-TaroB-minor; explained later) and SJ (SJ-TaroB) mitogenomes were not clearly separated from each other, indicating the existence of genetic exchange between the two groups of individuals within this population. It may be notable that two minor clusters, each formed of three individuals with PO mitogenomes, were identified (clusters X and Y in Fig. 3A). These clusters appeared even if we removed SNP loci that showed high r^2 values ($r^2 > 0.5$) with other sites to mitigate effects of linkage disequilibrium, and they were also identified by *k*-means clustering and ADMIXTURE analyses of the Taro B population data (Fig. S4). When RAD locus data of the neighboring populations were added to the analysis, these minor clusters did not become closer to those of the neighboring populations (Fig. 3B); thus, individuals in these two minor clusters are unlikely to be immigrants from neighboring populations but unique to the Taro B population. It is also notable that excess heterozygosity was detected in both of these minor clusters (cluster X: observed heterozygosity [H_O] = 0.286, expected heterozygosity [H_E] = 0.217; cluster Y: H_O = 0.277, H_E = 0.209) but not in the major cluster (H_O = 0.286, H_E = 0.286). Due to the fact that small-data-size resampling from the major cluster did not show such excess heterozygosity (Fig. S5), this observation was likely to be because of specific genetic selection in the minor clusters, rather than sampling artifacts.

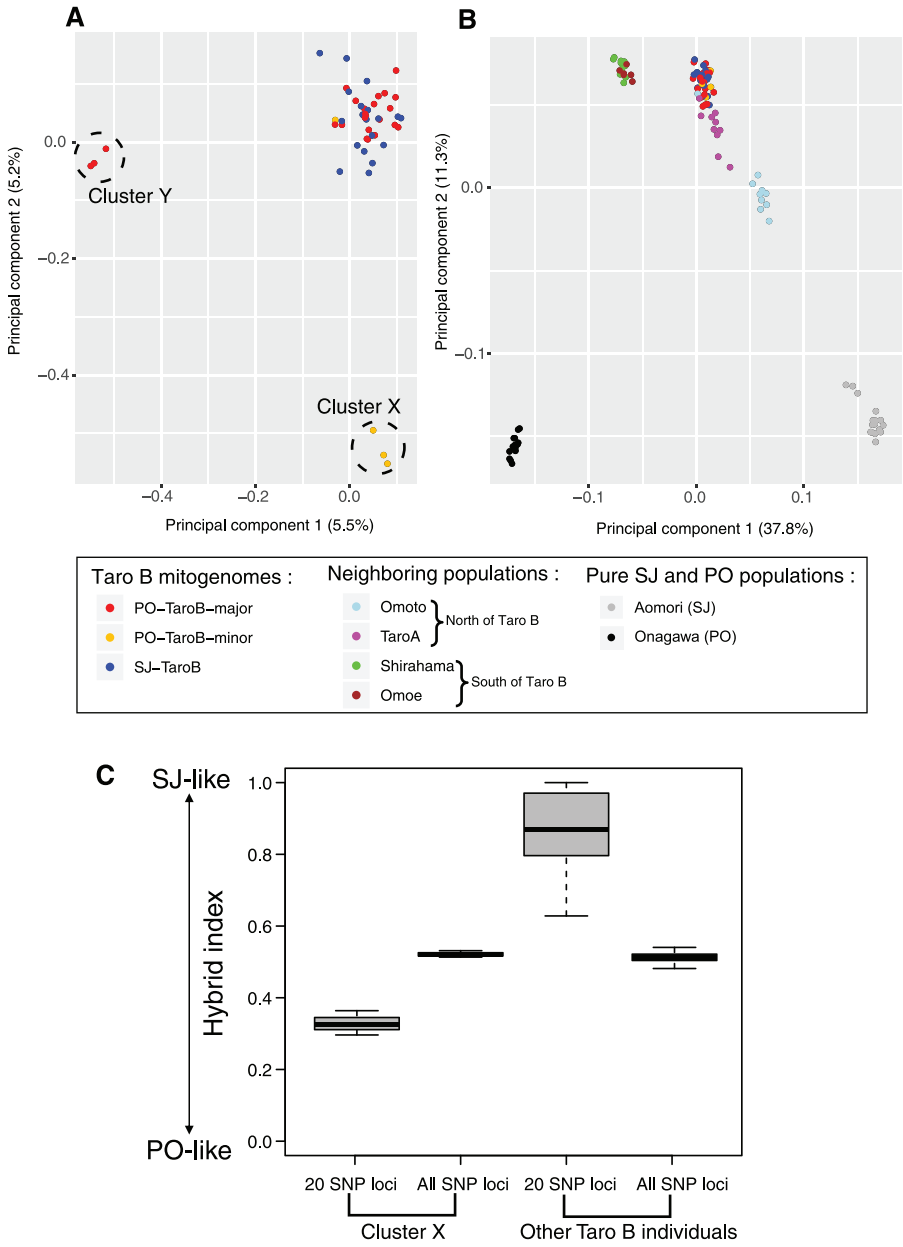

Figure 3. **Principal components analysis of the SNP data. (A) Results based on RAD-sequencing of the Taro B population. Red, orange, and blue indicate individuals with PO-TaroB-major, PO-TaroB-minor, and SJ-TaroB mitogenomes, respectively. (B) Results based on RAD-sequencing of the Taro B and neighboring populations. Light blue and pink indicate individuals from Omoto and Taro A, respectively, which are located to the north of Taro B. Green and purple indicate Shirahama, and Omoe, respectively, which are located to the south of Taro B (Fig. 1). Gray and black indicate individuals from Onagawa and Aomori, respectively (Fig. 1). (C) Hybrid indices of the nuclear genome SNP loci of the individuals in cluster X and others from the Taro B population. The hybrid indices are calculated for all SNP loci, and for the 20 SNP loci that contributed to principal component 2 in the PCA analysis (Fig. 3A) and were associated with the types of mitogenomes. The hybrid index becomes one for SJ-lineage nuclear genomes.**

OXPHOS genes on nuclear genomes are known to be important in mitonuclear incompatibility (Burton and Barreto 2012). Thus, we also performed amplicon sequencing of 15 OXPHOS genes that were successfully annotated in the *C. annularis* genome and had divergent single-nucleotide variants (SNVs) between the two lineages (Table S7) in the 47 individuals in the Taro B population. PCA using the OXPHOS gene data alone did not produce any clusters associated with the mitogenomic haplotypes (Fig. S6), suggesting that the 15 OXPHOS genes were not associated with selection in the Taro B population.

INCREASED NONSYNONYMOUS MUTATIONS IN HYBRID POPULATION MITOGENOMES

The observations above could be because (1) mitonuclear incompatibility does not yet exist between the PO and SJ lineages or (2) mitonuclear incompatibility has been secondarily resolved during the hybridization process. Through further investigation of the mitogenomic sequences within the Taro B population, we found that the mitogenomic haplotypes in the Taro B population are generally monophyletic in the *C. annularis* mitogenomic tree (Fig. 4), in accordance with the infrequent immigration scenario. Specifically, the PO-lineage mitogenomes were split to a monophyletic clade (PO-TaroB-major) and a single haplotype (PO-TaroB-minor), whereas the SJ lineage mitogenome formed a single monophyletic clade (SJ-TaroB). Most mitogenomes belonged to the PO-TaroB-major clade, except for four mitogenomes that had the PO-TaroB-minor haplotype. Compared to the closely-related PO lineage mitogenomes of the neighboring Yamada population, the PO-TaroB-major haplotypes contained seven nonsynonymous substitutions in total, in the *ND2* (codons 44 and 342), *ND4* (184, 185, and 270), *COI* (489), and *COIII* (230) genes, whereas the PO-TaroB-minor haplotype contained one nonsynonymous substation in the *Cytb* gene (28) (Fig. 5A). Pairwise comparisons of the mitogenomic sequences of the Taro B and neighboring populations were performed to calculate K_a/K_s ratios (number of nonsynonymous substitutions per nonsynonymous site over synonymous substitutions per synonymous site) (Fig. 5B). The K_a/K_s ratios of these five genes between the Taro B and neighboring populations revealed significant increase of nonsynonymous substitutions in the Taro B population (Wilcoxon rank sum test, $P < 0.01$; Fig. 5B).

The most notable Taro B-specific nonsynonymous mitogenomic substitutions were those in the *ND4* gene of the PO-TaroB-major mitogenome. ND4 is a transmembrane proton-pumping protein in mitochondrial complex I, which uses energy liberated during electron transfer from NADH to O_2 in the OXPHOS process (Sánchez-Caballero et al. 2016) and is frequently under positive selection (Garvin et al. 2015). The proton gradient generated across the inner mitochondrial membrane is used by ATP synthase (complex V) to chemiosmotically drive ATP synthesis. Two of the three nonsynonymous substitutions (codons 184 and 270) changed the amino acids of the ND4 protein from T and V to A and I, respectively, and the latter are the same amino acids as those found in all SJ mitogenomes in the Aomori and neighboring hybrid populations (Fig. 5A). TreeSAAP analysis showed that the two nonsynonymous substitutions occurred in different branches independently, suggesting they were convergent substitutions (Fig. 4). Although statistical tests could not be performed because of the limited numbers of substitution sites, it should be noted that such convergent substitutions are consistent with the assumption that they were beneficial because they

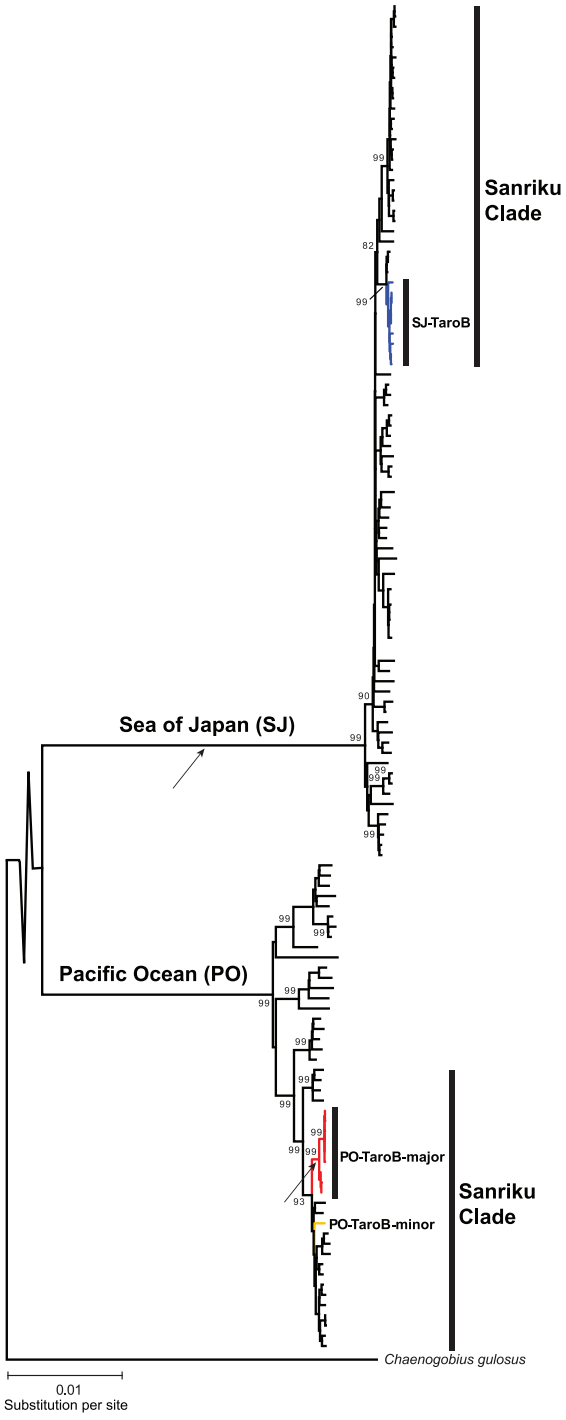

Figure 4. A phylogenetic tree of 132 *C. annularis* mitogenomic haplotypes (distinct sequences) with an outgroup (*C. gulosus*). A neighbor-joining tree based on Jukes-Cantor distance was reconstructed with 1000 bootstrap replications. Branch numbers indicate bootstrap values that strongly (≥90%) support geographical clades. The bootstrap value supporting the Sanriku clade within the SJ lineage (82%) is also shown. Branch lengths are proportional to the estimated numbers of nucleotide substitutions per site. PO-TaroB-major (nine haplotypes), PO-TaroB-minor (one), and SJ-TaroB (nine) clades are noted. Arrows show the branches in which two convergent nonsynonymous substitutions in the *ND4* gene occurred.

A

Gene	ND2		ND4			Cyt*b*	COI	COIII
Codon	44	342	184	185	270	28	489	230
PO (Yamada)	A	V	T	T	V	A	A	E
PO-TaroB-minor	A	V	T	T	V	V	A	E
PO-TaroB-major	A or T	M	A	A or T	I	A	G	Q or E
SJ (Aomori)	A	V	A	T	I	A	A	E

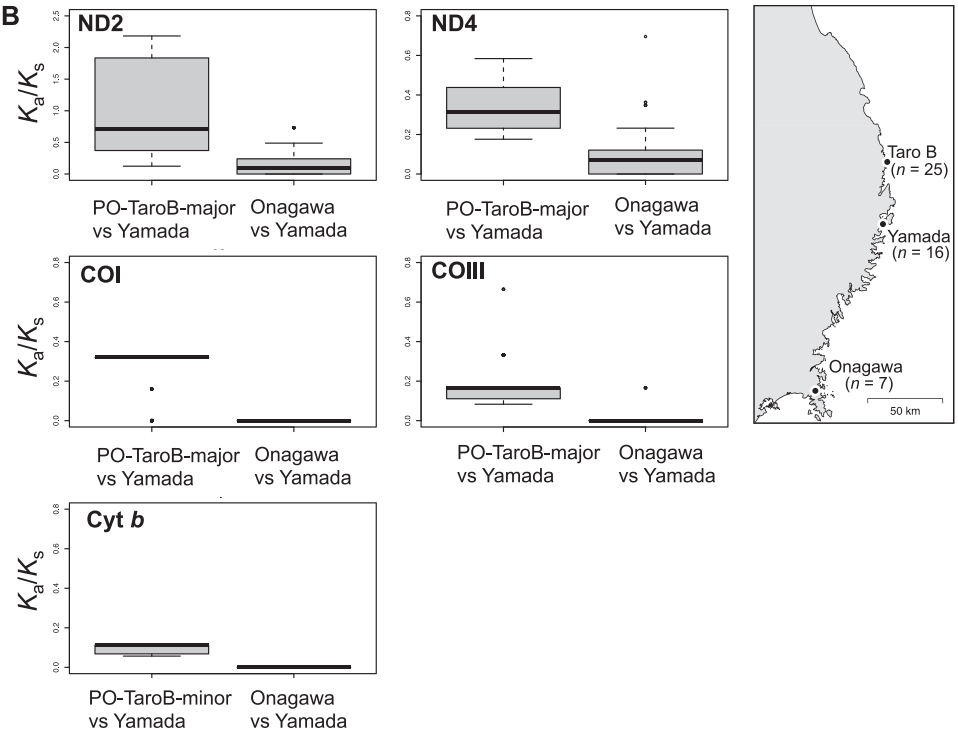

Figure 5. **Increased nonsynonymous mutations on the mitogenomes of Taro B individuals. (A) Amino acids coded for by the nonsynonymous substitution sites. In addition to those of the PO-TaroB-major and PO-TaroB-minor mitogenomes, the amino acids of the consensus mitogenomic sequences of the neighboring populations (Yamada and Aomori for the PO and SJ groups, respectively) are shown. (B) Plots of pairwise K_a/K_s values calculated for each mitogenomic gene. The values between the Taro B and Yamada populations are compared to those between the Yamada and Onagawa populations as background rates. The PO-TaroB-major haplotypes were used to analyze the *ND2*, *ND4*, *COI*, and *COIII* genes, and the PO-TaroB-minor haplotype was used to analyze the *Cytb* gene. The box limits and lines within them represent quartiles and medians, respectively. The whiskers represent the maximum and minimum values within the 1.5× interquartile ranges. For reference, the locations of the Taro B, Yamada, and Onagawa sites and the numbers of PO mitogenomes used for this analysis are shown in the right-hand-side map.**

ameliorated the mitonuclear incompatibility between the PO lineage mitogenomes and the SJ lineage nuclear genome. Topology predictions by TMHMM Server located these two amino acids (A and I) within the core transmembrane helices, not within the flexible loop regions. TreeSAAP analysis of selection on amino acid properties using phylogenetic trees (Woolley et al. 2003) showed that the nonsynonymous substitution at codon 184 (and another at codon 185) could lead to significant physicochemical changes that affect alpha-helix structures (AA-score: 6; $P < 0.001$). It may also be notable that previous studies have suggested that mutations in ND4 transmembrane regions change its ATP-synthesizing activity (Torres-Bacete et al. 2007) and can be under positive selection on proton pumping activity (Garvin et al. 2011).

Overall, the increase of nonsynonymous mutations suggests that mitonuclear incompatibility between the PO-lineage mitogenomes and SJ-lineage nuclear genome existed and led to positive selection pressure on the PO-lineage mitogenomes in the Taro B population. On the other hand, such evolution was not observed in the SJ-TaroB mitogenome, possibly due to the absence of mitonuclear incompatibility between the SJ-lineage mitogenome and PO-lineage nuclear genome and/or due to increases in the frequencies of SJ-lineage nuclear mitochondria-related genes.

ENRICHMENT OF MITOCHONDRIA-RELATED GENES IN GENOMIC REGIONS THAT ARE LIKELY TO BE UNDER SELECTION DUE TO MITONUCLEAR INCOMPATIBILITY

Although the 24 PO-TaroB-major mitogenome contained seven nonsynonymous substitutions in total, the PO-TaroB-minor mitogenome contained only one nonsynonymous mutation and was possessed by only four individuals. Thus, we assumed that these four individuals with the PO-TaroB-minor haplotype were still under the influence of mitonuclear incompatibility. Under this assumption, their nuclear genome genes, especially those with mitochondria-related functions, would be under negative selection.

Among the four individuals with the PO-TaroB-minor haplotype, three had nuclear genomes that belonged to one of the two minor clusters in the PCA (cluster X in Fig. 3A). The association between cluster X and the PO-TaroB-minor mitogenomic haplotype was statistically significant (Fisher exact test; $P < 0.01$). Because cluster X was separated from the others by principal component 2 in the PCA (Fig. 3A), 20 SNP loci that contributed significantly to this component were retrieved (FDR < 0.1; Table 1). The hybrid indices of these SNP loci of the individuals in cluster X were 0.32-0.37 (Fig. 3C; the hybrid index becomes 0 if all SNP loci have homozygous PO alleles). These were significantly smaller than the hybrid indices of all SNP loci, meaning that the individuals in cluster X had more PO-like nuclear genomic regions close to those 20 SNP loci than the genomic average (Wilcoxon rank sum test, $P < 0.01$). In contrast, the hybrid indices of the same 20 SNP loci in other Taro B individuals were significantly larger than those of all SNP loci (Fig. 3C; Wilcoxon rank sum test, $P < 0.01$). This means that the nuclear genomic regions that were divergent between the individuals in cluster X and other Taro B individuals were more PO-like in the former. It should be noted that this observation is in line with the assumption that individuals with the PO-TaroB-minor haplotype may be still under the influence of mitonuclear incompatibility against SJ-lineage nuclear genomes.

Fifty-nine genes were retrieved (Table S8 and a summary in Table 2) from the genomic regions around the 20 SNP loci. Surprisingly, no nuclear OXPHOS genes were included among them; however, functional enrichment analysis using DAVID version 6.8 yielded 36 significant terms, including "mitochondrion" and "mitochondrial inner membrane" (EASE score ≤ 0.2; Table S9). Mitochondria-related functions were allocated to 10 proteins: Phosphatidylinositol 4-kinase type 2 alpha (*PI4K2A*), FAST kinase domains 5 (*FASTKD5*), serine hydrolase-like protein 2 (*SERHL2*), essential mitochondrial MCU regulator (*SMDT1*), ATP/GTP binding protein 1 (*AGTPBP1*), EF-hand domain family member D1 (*EFHD1*), NLR family member X1 (*NLRX1*), solute carrier family 25 member 42 (*SLC25A42*), iron responsive element binding protein 2 (*IREB2*), and 28S ribosomal protein S27 (*MRPS27*).

GENOME RESEQUENCING AND RNA-SEQUENCING ANALYSES OF MITOCHONDRIA-RELATED NUCLEAR GENES LIKELY TO BE UNDER SELECTION DUE TO MITONUCLEAR INCOMPATIBILITY

If the abovementioned 10 mitochondria-related genes were under selection due to mitonuclear incompatibility, they should be associated with the PO-TaroB-minor mitogenome and should also need to be differentiated between the PO and SJ lineages. We thus conducted whole-genome resequencing of six PO-lineage (three from Onagawa and three from Misaki) and six SJ-lineage (three from Murakami and three from Aomori) individuals. Mapping the resequenced reads to the genome detected 8,181,853 SNVs, 805,134 insertions, and 945,311 deletions in the PO lineage and 3,157,276 SNVs, 504,753 insertions, and 656,569 deletions in the SJ lineage, on average (Table S4). The mean pairwise nucleotide diversities calculated for 15-kb windows were 0.0054 and 0.0030 in the PO and SJ lineages, respectively. The mean weighted (Weir and Cockerham's) F_{ST} value was 0.50 (Fig. S7), showing substantial genetic divergence between the PO and SJ lineages. Of the 10 genes, *FASTKD5*, *AGTPBP1*, and *NLRX1* had divergent nonsynonymous substitutions between the PO and SJ lineages (Tables 2 and S8). Because differences can also occur at the expression level, the expression levels of the 10 genes were examined using the RNA-sequencing data. The *PI4K2A* gene showed significantly different expression levels between the PO and SJ lineages (upregulated in the PO lineage), whereas significantly different expression levels were also observed in the myoglobin (*MB*) gene (upregulated in the PO lineage; Tables 2 and S8). Although mitochondria-related terms were not allocated to the *MB* gene, myoglobin is known to localize in the mitochondria and to interact with mitochondrial complex IV, which contains COI and COIII proteins and regulates OXPHOS activity (Kamga et al. 2012; Yamada et al. 2016).

POSSIBLE INVOLVEMENT OF AN miRNA GENE IN MITONUCLEAR INCOMPATIBILITY

It should also be noted that, among the four individuals that had the PO-TaroB-minor mitogenomic haplotype, one individual (*TaroB06*) belonged to the PO-TaroB-major and SJ-TaroB individuals cluster in the PCA plot (the major cluster in Fig. 3A). Therefore, we examined the same 20 SNP loci in TaroB06, and found that TaroB06 and the other three individuals with the PO-TaroB-minor mitogenome only had the 20th, 21st, and 22nd SNP loci in common (Table 1). At these sites, all four PO-TaroB-minor individuals had GT-GA-CA alleles, where the T-A-A alleles were derived from the PO lineage (Fig. S8). On the other hand, all individuals with the PO-TaroB-major mitogenome had

Table 1. Twenty SNP sites contributing to the principal component 2 of PCA of RAD-sequencing data.

Individual	Mitogenome	Site 20	21	22	308	509	528	1150	1152	1496	1602	1747	1748	1749	2368	2401	2606	3007	3142	3143	3688
PO consensus (Onagawa)	PO	G/T	G/A	C/A	A/T	G/C	G/A	C	A/C	C	T/C	T	A/C	T/C	T/C	G	T/G	G	G/A	T/A	C/T
SJ-TaroB consensus (Aomori)	SJ-TaroB	G	G	C	A	G	G	C/T	A	A/C	T	T/C	C	T	T	G/T	G	G	G	T	C
TaroB01	PO-TaroB-major	GG	GG	CC	AT	GG	GG	CC	AA	AC	TT	CC	AA	CC	TC	GT	GT	GG	GG	TT	CC
TaroB02	PO-TaroB-major	GG	GG	CC	AA	GG	GG	CC	AA	AC	TT	CC	AA	CC	TT	na	GG	GG	GG	TT	CC
TaroB03	SJ-TaroB	GG	GG	CC	na	GG	GG	CC	AA	AC	TC	CC	AA	CC	na	na	GG	GG	GG	TT	CC
TaroB04	PO-TaroB-major	GG	GG	CC	AA	GG	GG	CC	AA	AC	TT	CC	AA	CC	TT	na	GG	GG	GG	TT	CC
TaroB05	PO-TaroB-major	GG	GG	CC	AA	GG	GG	CC	AA	AC	TT	CC	AA	CC	TT	GG	GG	GG	na	na	CC
TaroB06	PO-TaroB-minor	GT	GA	CA	na	GG	GG	na	AA	AA	TT	CC	AA	CC	TT	na	GG	GG	GG	TT	CC
TaroB07	SJ-TaroB	GG	GG	CC	na	GC	GG	na	AA	AC	TC	CC	AA	CC	TT	na	GG	GG	na	na	CT
TaroB08	PO-TaroB-major	GG	GG	CC	na	GC	GG	CC	AA	AA	TT	CC	AA	CC	TC	GG	GG	GA	AA	AA	CC
TaroB09	SJ-TaroB	GG	GG	CC	AA	GG	GG	CC	AA	AC	TT	CC	AA	CC	TC	GT	GG	GA	GG	TT	CC
TaroB10	PO-TaroB-major	GG	GG	CC	AA	GC	GG	CC	AA	CC	TT	CC	AA	CC	TT	GG	GT	GG	GG	TT	CC
TaroB11	SJ-TaroB	na	na	na	na	GG	GG	CC	AA	CC	TT	CC	AA	CC	TT	na	GG	GG	GG	TT	CC
TaroB12	PO-TaroB-major	GG	GG	CC	AA	GC	GG	CC	AA	AA	TC	CC	AA	CC	TC	na	GT	GA	GA	TA	CC
TaroB13	SJ-TaroB	GG	GG	CC	na	GG	GG	na	AA	AC	TC	CC	AA	CC	TC	na	GG	GG	na	na	CC
TaroB14	PO-TaroB-major	na	na	na	AT	GG	GG	na	AA	AA	TT	CC	AA	CC	TT	na	GG	GG	na	na	CC
TaroB15	SJ-TaroB	GG	GG	CC	na	GG	GG	CT	AC	AC	TT	TC	CA	TC	TT	na	GG	GG	GG	TT	CT
TaroB16	PO-TaroB-major	GG	GG	CC	AT	GG	GG	CC	AA	CC	TT	CC	AA	CC	TT	GG	GG	GG	GG	TT	CC
TaroB17	SJ-TaroB	GG	GG	CC	AA	GC	GG	CC	AA	CC	TT	CC	AA	CC	TC	GG	GG	GG	GG	TT	CC
TaroB18	PO-TaroB-major	GG	GG	CC	AA	GC	GG	CC	AA	AA	TC	CC	AA	CC	TC	GG	GG	GA	GA	TA	CC
TaroB19	SJ-TaroB	GG	GG	CC	AA	GG	GG	CC	AA	AC	TC	CC	AA	CC	TT	na	GG	GG	GG	na	CC
TaroB20	SJ-TaroB	GG	GG	CC	AA	GG	GG	CC	AA	CC	TC	CC	AA	CC	TC	GG	GG	GA	GG	TT	CC
TaroB21	PO-TaroB-major	na	na	na	na	GG	GG	na	AA	AC	TC	CC	AA	CC	na	na	GG	GG	na	na	CC
TaroB22	PO-TaroB-major	GG	GG	CC	na	GG	GG	CC	AA	AA	TT	TC	CA	TC	TT	na	GG	GG	GG	TT	CC
TaroB23	SJ-TaroB	GG	GG	CC	AA	GG	GG	CC	AA	AC	TT	CC	AA	CC	TT	GG	GG	GG	GG	TT	CC
TaroB24	PO-TaroB-major	GG	GG	CC	AT	GG	GG	na	AA	CC	TC	CC	AA	CC	TT	GG	GG	GG	GA	TA	CC

(*Continued*)

Table 1. (Continued).

Individual	Mitogenome	Site 20	21	22	308	509	528	1150	1152	1496	1602	1747	1748	1749	2368	2401	2606	3007	3142	3143	3688
TaroB25	PO-TaroB-minor	GT	GA	CA	AT	GC	GA	CT	AC	AA	CC	TC	CA	TC	CC	GT	TT	GA	GA	TA	CT
TaroB26	PO-TaroB-major	GG	GG	CC	na	GC	na	CC	AA	CC	TT	CC	AA	CC	na	na	na	GA	GG	TT	CC
TaroB27	SJ-TaroB	GG	GG	CC	AA	GG	GG	CT	AC	AA	TT	TC	CA	TC	TC	na	GG	GG	GA	TA	CC
TaroB28	PO-TaroB-major	GG	GG	CC	na	GG	GG	CC	AA	AC	TT	CC	AA	CC	TC	GG	GG	GA	GG	TT	CC
TaroB29	SJ-TaroB	GG	GG	CC	AA	GG	GG	CC	AA	AC	TT	CC	AA	CC	TT	na	GT	GG	GG	TT	CC
TaroB30	PO-TaroB-major	na	na	na	na	GG	GG	na	AA	AC	TT	CC	AA	CC	TC	GG	GG	GG	na	na	CT
TaroB31	PO-TaroB-major	GG	GG	CC	AA	GG	GG	CC	AA	AC	TT	CC	AA	CC	TT	GG	GT	GG	GA	TA	CC
TaroB32	PO-TaroB-major	GG	GG	CC	AA	GG	GG	CC	AA	CC	TT	CC	AA	CC	TC	GG	GG	GG	na	na	CC
TaroB33	SJ-TaroB	GG	GG	CC	AA	GG	GG	CC	AA	AC	TC	TC	CA	TC	TC	GT	GT	GG	GA	TA	CC
TaroB34	SJ-TaroB	GG	GG	CC	AA	GG	GG	CC	AA	AC	TT	CC	AA	CC	TT	GG	GG	GG	na	na	CC
TaroB35	PO-TaroB-major	GG	GG	CC	AA	GG	GG	CC	AA	CC	TT	CC	AA	CC	TT	GG	GT	GG	GG	TT	CC
TaroB36	SJ-TaroB	GT	GA	CA	AA	GG	GG	na	AA	CC	TC	CC	AA	CC	TT	GG	GG	GG	GG	TT	CC
TaroB37	PO-TaroB-minor	GT	GA	CA	TT	GC	GA	CT	AC	AA	CC	TC	CA	TC	CC	TT	TT	na	na	na	CT
TaroB38	SJ-TaroB	GG	GG	CC	AA	GC	GA	CC	AA	AC	TT	CC	AA	CC	TT	na	GG	GG	na	na	CC
TaroB39	PO-TaroB-major	na	na	na	na	GG	GG	na	AA	AC	TT	CC	AA	CC	TT	GG	GG	GG	na	na	CC
TaroB40	PO-TaroB-major	GG	GG	CC	na	GG	GA	CC	AA	AC	TC	CC	AA	CC	TT	GG	GT	GG	GG	TT	CC
TaroB41	SJ-TaroB	GT	GA	CA	na	GG	GG	CC	AA	AA	TT	TC	CA	TC	TT	na	GG	GG	GA	TA	CC
TaroB42	PO-TaroB-minor	GT	GA	CA	TT	GC	GA	CT	AC	AA	CC	TC	CA	TC	CC	TT	TT	AA	AA	AA	TT
TaroB43	SJ-TaroB	GT	GA	CA	AA	GG	GG	CC	AA	AC	TC	CC	AA	CC	TT	na	GG	GG	GA	TA	CC
TaroB44	SJ-TaroB	GG	GG	CC	na	GG	GG	na	AA	AA	CC	TC	CA	TC	na	na	GT	GG	GG	TT	CC
TaroB45	PO-TaroB-major	GG	GG	CC	AA	GG	GG	CC	AA	AC	TT	CC	AA	CC	TT	GG	GG	GG	GG	TT	CC
TaroB46	PO-TaroB-major	GG	GG	CC	na	GG	GG	CC	AA	AC	TT	CC	AA	CC	TT	GG	GG	GG	GG	TT	CC
TaroB47	PO-TaroB-major	GG	GG	CC	AA	GG	GG	na	AA	CC	TT	CC	AA	CC	TT	na	GG	GG	na	na	CC

na = no genotype data.

Table 2. Mitochondria-related genes exist around the 20 SNP loci that contributed the principal component 2 in the PCA analysis (Fig. 4A) and were associated with the types of mitogenomes.

Gene symbol	Gene name	Nonsynonymous SNPs	Differential expression (brain)
MB	Myoglobin	0	PO > SJ
AGTPBP1[*]	ATP/GTP binding protein 1	1	
EFHD1[*]	EF-hand domain family member D1	0	
NLRX1[*]	NLR family member X1	1	
SLC25A42[*]	Solute Carrier Family 25 Member 42	0	
IREB[*]	Iron responsive element binding protein 2	0	
PI4K2A[*]	Phosphatidylinositol 4-Kinase Type 2 Alpha	0	PO > SJ
FASTKD5[*]	FAST Kinase Domains 5	1	
SERHL2[*]	Serine hydrolase-like protein 2	0	
SMDT1[*]	Essential MCU regulator, mitochondrial	0	
MRPS27[*]	28S ribosomal protein S27, mitochondrial	0	

[*] **Mitochondria-related functions were annotated by DAVID analysis. For the full list of genes that include those without annotated mitochondria-related functions, see Table S8.**

homozygous GG-GG-CC alleles (Table 1). Thus, these loci were assumed to be particularly important in ameliorating mitonuclear incompatibility in individuals with the PO-TaroB-minor mitogenome.

The 20th, 21st, and 22nd SNP loci were located within a 27-bp region that was sequenced by one RAD-tag on the largest scaffold (scaffold1) and was likely to be under strong linkage (Table 1). We found another (fourth) uncalled SNP locus that was also linked to the 20th, 21st, and 22nd SNP loci just 22-bp away from the 22nd SNP (G and A were linked with the G-G-C and T-A-A alleles, respectively). We could not find any protein-coding genes around these four SNP loci, but found an approximately 120-bp region that was approximately 850-bp away from them and is highly conserved across many fish genomes (e.g., in the stickleback genome [Group II: 1,774,680-1,774,799; Ensembl release 94]), as a possible conserved *cis*-regulatory element. To search for protein-coding genes potentially regulated by this highly conserved region, we surveyed other fish genomes; however, no such gene was found in the regions around (<50 kb) to the conserved region in the *C. annularis* genome or in other fish genomes.

The number of SNP loci within this RAD-tag was confirmed to be significantly large when regions with mapped RAD-seq reads were used as a null distribution (35,299 of 2,810,696 RAD-tags showed three or more called SNP loci; $P = 0.012$). The significant accumulation of the SNP loci within approximately 50 bp indicates the existence of a lineage-specific selection pressure on this small nuclear genomic region. BLASTN searches of this 50-bp region against the NCBI *nt* database identified 3' untranslated regions (UTRs) for 19 *Boleophthalmus pectinirostris* genes, which also belongs to the Gobiidae family (*E*-value < 0.01; Table S10). DAVID analysis inferred mitochondria-related functions for four of these proteins, including Electron transfer flavoprotein dehydrogenase (*ETFDH*), autophagy and beclin 1 regulator 1 (*AMBRA1*), hepatocyte growth factor-like protein (*HGF*), and Ras-related protein Rab-35 (*RAB35*), based on the Gene Ontology and UniProt databases (Table S10). Secondary structure analysis of the reverse complement sequence of the 50-bp region by miRNAFold predicted a 96-nt RNA with a hairpin structure (Fig. 6), and such a structure is typically found in precursor miRNA genes (He and Hannon 2004). Furthermore, miRBoost predicted this region to be an miRNA gene. Finally, by conducting small RNA-sequencing of *C. annularis* brain samples (Table S11), we detected the expressions of seven small RNAs that contained the same 19-nt reverse complement sequence, including the fourth (G or A) SNP locus in that region (Fig. 6). Although expression of the 50-bp region itself was not detected in our brain samples, the aforementioned small RNAs were also predicted to be miRNAs by miRBoost and likely target the same mRNA molecule using the shared sequence as their seed sequences (Broughton et al. 2016). We conducted a BLASTN search of that candidate seed sequence against the *C. annularis* genome (*E*-value < 0.0000001) by referring to the RNA-sequencing results. As a result, two perfect-match hits were retrieved on 3'-UTR regions of the stonustoxin subunit beta (g51326) and ubiquitin associated and SH3 domain containing B (*UBASH3B*, g55929) genes. The UBASH3B protein is known to interact with solute carrier family 25 member 48 protein (*SLC25A48*), which is a proton transporter protein localized in the inner mitochondrial membrane (Ramsden et al. 2012). These findings may suggest that the 50-bp region may encode an miRNA gene that is

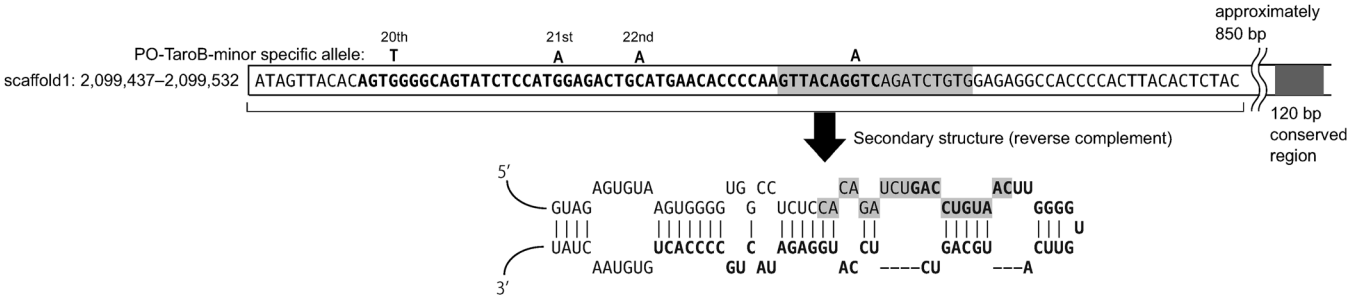

Figure 6. A predicted lineage-specific miRNA gene likely to be involved in mitonuclear incompatibility. The scaffold 1 (2,099,437-2,099,532) sequence containing the 51-bp RAD locus (shown by bold), as well as the PO-TaroB-minor-specific 20th, 21st, 22nd, and additional SNP loci, is shown. The 19-nt sequence shared by seven small RNAs is shown in gray as a candidate seed sequence. A reverse complement secondary structure of the sequence predicted by miRNAfold is also shown, as the small RNA sequences were mapped to the reverse strand.

involved in mitonuclear incompatibility and is transcribed by the conserved *cis*-regulatory element, and that the fourth SNP locus within the probable seed sequence could affect its regulatory activity (Mencía et al. 2009).

Discussion

In this study, *C. annularis* was investigated to clarify the hybridization process of two diverged lineages. Our key findings are that (1) significantly large numbers of nonsynonymous substitutions occurred in most haplotypes of the PO mitogenomes in the Taro B hybrid population, (2) two nonsynonymous substitutions occurred in divergent sites of the *ND4* mitochondrial gene between the PO and SJ mitogenomes, which make the PO mitogenome haplotypes (PO-TaroB-major) similar to the SJ mitogenome haplotypes, (3) a few individuals with the PO mitogenome without such substitutions (PO-TaroB-minor) tend to have genomic sequences that are specific to the PO lineage in a few nuclear regions, and (4) the SJ mitogenome, on the other hand, did not evolve in the hybrid population. These results suggest that the PO mitogenomes evolved to become more similar to the SJ mitogenome under selective pressure due to mitonuclear incompatibility, which was between the SJ nuclear genomes that were dominant in the hybrid population. The genome sequence and the integrative genomic phylogeographic datasets obtained by RNA-sequencing, RAD-sequencing, whole mitogenome sequencing, amplicon sequencing, genome resequencing, and small RNA-sequencing in this study will constitute a rich resource for future studies focusing on this important but enigmatic evolutionary process.

It should be noted that these findings are consistent with the mitonuclear compensatory coevolution hypothesis (Hill 2019), which assumes that mitonuclear coevolution is driven by deleterious mutations in one genome and subsequent compensatory mutations in the other genome. Given asexual and error prone mitogenome inheritance, it is theoretically predicted that mitogenomes likely accumulate deleterious mutations first, and then mutations on nuclear genomes compensate these mutations (Hill et al. 2019). However, empirical data have shown that either of the nuclear and mitochondrial genomes can trigger the process by accumulating deleterious mutations (Hill 2019). Our study demonstrates that mitogenomic mutations can compensate for deleterious nuclear genomic mutations through hybridization in a natural environment.

We could not rule out possibilities that genetic drift affected by local substructures and/or inbreeding resulted in the genotype changes within the Taro B population. However, the excess of heterozygosities in the PO-TaroB-minor individuals and the enrichment of mitochondria-related genes in the associated nuclear regions also support the existence of selective pressures due to mitonuclear incompatibility. It is difficult to say whether the substitutions in the PO mitogenomes arose *de novo* after hybridization or were standing variations. However, given that the PO-TaroB-major haplotypes belonged to the clades that are only distributed in the Sanriku coast, the novel mutations may have arisen *de novo* after secondary contact.

Among the *de novo* mutations in the *C. annularis* mitogenomes, the convergent substitutions in the *ND4* gene were particularly notable. We compared the *ND4* gene sequences with those on the mitogenome sequences of related species (*C. gulosus*, *Luciogobius* spp., and *Gymnogobius* spp.), and found that the PO-lineage sequence is ancestral (i.e., the 184th and 270th codons in the SJ-lineage mitogenome were lineage-specific derived mutations). The SJ lineage of *C. annularis* experienced a previous bottleneck followed by sudden expansion events in the Pleistocene (Hirase et al. 2012; Hirase and Ikeda 2014; Hirase et al. 2016). During the glacial periods, the Sea of Japan was characterized by cold water temperature caused by the closing of the south Tsushima Strait and an inflow of cold freshwater from rivers (Oba et al. 1991; Tada 1994; Gorbarenko and Southon

2000). A previous study reported that *ND4* mutations can contribute to adaptation to colder environments through increased heat production by the OXPHOS process in an Australian bird, *Eopsaltria australis* (Morales et al. 2015). Also, *ND4* and other *ND* genes, which encode elements of Complex I, have been suggested to be frequently under adaptive evolution, for example, for thermal tolerance in fish (Li et al. 2013; Silva et al. 2014; Garvin et al. 2015; Consuegra et al. 2015; Wang et al. 2016; Hulsey et al. 2016). Thus, if the *ND4* mutations were fixed in the SJ lineage because they were adaptive to the cold environment and needed to be subsequently compensated by nuclear mutations, it may be possible that the cold environment in the coastal area around Taro (Hirase and Ikeda 2015) increased the frequencies of SJ-lineage mitonuclear interactions with the *ND4* mutations and led to the convergent substitutions observed on the PO-TaroB-major mitogenomes. Collectively, these substitutions in the PO-lineage mitogenomes may not only resolve the mitonuclear incompatibility with SJ-lineage nuclear genes but also aid in the adaptation of the host. Relationships between environmental effects and nuclear/mitochondrial genome evolution in natural hybrid populations should be further investigated in the future (Arnqvist et al. 2010; Zhang et al. 2017; Hill 2019).

The possible involvement of miRNAs in posttranscriptional mechanisms helping to resolve mitonuclear incompatibility in a hybrid population is also a new concept. Recent studies have provided evidence that miRNAs regulate nuclear-encoded mitochondrial proteins at the posttranscriptional level by affecting the OXPHOS and ATP synthesis systems (Sripada et al. 2012). For example, *miR-107* inhibits mitochondrial β-oxidation, which is closely linked to the OXPHOS system (Bhatia et al. 2016). Considering that the study species is currently unculturable in laboratories, making physiological and molecular experiments difficult, our integrative genomic phylogeography approach provides a powerful means of highlighting its key evolutionary processes.

AUTHOR CONTRIBUTIONS

SH and WI designed the study. SH, AT, and AN performed wet laboratory experiments and collected data. SH and WI wrote the manuscript with the help of MS, SH, and KK. All authors approved the final version of the manuscript.

ACKNOWLEDGMENTS

The authors thank Y. Ikeda and T. Shimizu for providing specimens and M. Ikeda, S. Hayasaka, S. Kondo, R. Kawaguchi, J. Aoyama, Y. Minegishi, S. Yoshizawa, and K. Kogure for supporting sampling or experiments. The authors are grateful to members of the Iwasaki and Kikuchi laboratories for helpful comments on this research. This work was supported by the Ministry of Education, Culture, Sports, Science, and Technology (KAKENHI 221S0002 and Project "Construction of the platform for intellectual cooperation"), the Japan Society for the Promotion of Science (KAKENHI 23710231, 23370041, 26850131, 18H02493, 17K19280, and 16H06279, 15J06937), the Japan Science and Technology Agency (CREST), the Mikimoto Fund for Marine Ecology, Fujiwara Natural History Foundation, and the Canon Foundation. This study was partly supported by the Interdisciplinary Collaborative Research Program and the Cooperative Program of the Atmosphere and Ocean Research Institute, the University of Tokyo. Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA ARCHIVING

The *Chaenogobius annularis* genome sequences are available from DNA Data Bank of Japan (DDBJ) with the accession numbers BJKM01000001 to BJKM01001897. All mitochondrial sequence data are available from DDBJ under accession numbers: AP019419-AP019513. The DDBJ accession numbers of the data of the whole genome resequencing (DRR174910-DRR174921), RAD sequencing (DRR174875-DRR174909, DRR175781-DRR175955, and DRR177663-DRR177719), amplicon sequencing (DRR174926-DRR174972), RNA-sequencing (DRR174749-DRR174770), and small RNA-sequencing (DRR174922-DRR174925) are shown in Tables S3, S4, and S11. Other data for analysis are available on Dryad: https://doi.org/10.5061/dryad.5qfttdz3m.

LITERATURE CITED

Akihito, K. Sakamoto, Y. Ikeda, and K. Sugiyama. 2002. Gobioidei. In: Nakabo T. (ed) Fishes of Japan with pictorial keys to the species. English ed. Tokai Univ. Press.Tokyo.

Alexander, D. H., and K. Lange. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12:246.

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

Arnqvist, G., D. K. Dowling, P. Eady, L. Gay, T. Tregenza, M. Tuda, and D. J. Hosken. 2010. Genetic architecture of metabolic rate: environment specific epistasis between mitochondrial and nuclear genes in an insect. Evolution 64:3354–3363.

Asahida, T., T. Kobayashi, K. Saitoh, and I. Nakayama. 1996. Tissue preservation and total DNA extraction from fish stored at ambient temperature using buffers containing high concentration of urea. Fish. Sci 62:727–730.

Baris, T. Z., D. N. Wagner, D. I. Dayan, X. Du, P. U. Blier, N. Pichaud, M. F. Oleksiak, and D. L. Crawford. 2017. Evolved genetic and phenotypic differences due to mitochondrial-nuclear interactions. PLos Genet 13:e1006517.

Barreto, F. S., E. T. Watson, T. G. Lima, C. S. Willett, S. Edmands, W. Li, and R. S. Burton. 2018. Genomic signatures of mitonuclear coevolution across populations of *Tigriopus californicus*. Nat. Ecol. Evol 2:1250–1257.

Bhatia, H., B. Pattnaik, and M. Datta. 2016. Inhibition of mitochondrial β-oxidation by miR-107 promotes hepatic lipid accumulation and impairs glucose tolerance in vivo. Int. J. Obes 40:861–869.

Boetzer, M., C. V. Henkel, H. J. Jansen, D. Butler, and W. Pirovano. 2010. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27:578–579.

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

Boratyn, G. M., J. Thierry-Mieg, D. Thierry-Mieg, B. Busby, and T. L. Madden. 2018. Magic-BLAST, an accurate DNA and RNA-seq aligner for long and short reads. bioRxiv https://doi.org/10.1101/390013

Broughton, J. P., M. T. Lovci, J. L. Huang, G. W. Yeo, and A. E. Pasquinelli. 2016. Pairing beyond the seed supports microRNA targeting specificity. Mol. Cell 64:320–333.

Burton, R. S., and F. S. Barreto. 2012. A disproportionate role for mtDNA in Dobzhansky–Muller incompatibilities? Mol. Ecol 21:4942–4957.

Burton, R. S., R. J. Pereira, and F. S. Barreto. 2013. Cytonuclear genomic interactions and hybrid breakdown. Annu. Rev. Ecol. Evol. Syst 44:281–302.

Calabrese, V., G. Scapagnini, A. G. Stella, T. Bates, and J. Clark. 2001. Mitochondrial involvement in brain function and dysfunction: relevance to aging, neurodegenerative disorders and longevity. Neurochem. Res 26:739–764.

Chen, B., J. W. Cole, and C. Grond-Ginsbach. 2017. Departure from Hardy Weinberg equilibrium and genotyping error. Front Genet 8:167.

Consuegra, S., E. John, E. Verspoor, and C. G. de Leaniz. 2015. Patterns of natural selection acting on the mitochondrial genome of a locally adapted fish species. Genet. Sel. Evol 47:58.

Crusius, J., T. F. Pedersen, S. E. Calvert, G. L. Cowie, and T. Oba. 1999. A 36 kyr geochemical record from the Sea of Japan of organic matter flux variations and changes in intermediate water oxygen concentrations. Paleoceanography 14:248–259.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, and S. T. Sherry. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Dobzhansky, T. 1937. Genetics and the origin of species. Columbia Univ. Press, New York.

Doiron, S., L. Bernatchez, and P. U. Blier. 2002. A comparative mitogenomic analysis of the potential adaptive value of arctic charr mtDNA introgression in brook charr populations (Salvelinus fontinalis Mitchill). Mol. Biol. Evol 19:1902–1909.

Elgvin, T. O., C. N. Trier, O. K. Tørresen, I. J. Hagen, S. Lien, A. J. Nederbragt, M. Ravinet, H. Jensen, and G.-P. Sætre. 2017. The genomic mosaicism of hybrid speciation. Sci. Adv 3:e1602996.

Ellison, C. K., and R. S. Burton. 2006. Disruption of mitochondrial function in interpopulation hybrids of *Tigriopus californicus*. Evolution 60:1382–1391.

Emanuelsson, O., H. Nielsen, S. Brunak, and G. Von Heijne. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol 300:1005–1016.

Gagnaire, P.-A., E. Normandeau, and L. Bernatchez. 2012. Comparative genomics reveals adaptive protein evolution and a possible cytonuclear incompatibility between European and American eels. Mol. Biol. Evol 29:2909–2919.

Garvin, M. R., J. P. Bielawski, and A. J. Gharrett. 2011. Positive Darwinian selection in the piston that powers proton pumps in complex I of the mitochondria of Pacific salmon. PLoS One 6:e24127.

Garvin, M. R., J. P. Bielawski, L. A. Sazanov, and A. J. Gharrett. 2015. Review and meta-analysis of natural selection in mitochondrial complex I in metazoans. J. Zool. Syst. Evol. Res 53:1–17.

Glémet, H., P. Blier, and L. Bernatchez. 1998. Geographical extent of Arctic char (*Salvelinus alpinus*) mtDNA introgression in brook char populations (*S. fontinalis*) from eastern Quebec. Canada. Mol. Ecol 7:1655–1662.

Gompert, Z., and C. Buerkle. 2012. bgc: software for Bayesian estimation of genomic clines. Mol. Ecol. Resour 12:1168–1176.

Gompert, Z., L. K. Lucas, C. A. Buerkle, M. L. Forister, J. A. Fordyce, and C. C. Nice. 2014. Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. Mol. Ecol 23:4555–4573.

Good, J. M., D. Vanderpool, S. Keeble, and K. Bi. 2015. Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. Evolution 69:1961–1972.

Gorbarenko, S., and J. Southon. 2000. Detailed Japan Sea paleoceanography during the last 25 kyr: constraints from AMS dating and δ18O of planktonic foraminifera. Palaeogeogr. Palaeoclimatol. Palaeoecol 156:177–193.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, and Q. Zeng. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat. Biotechnol 29:644–652.

Haas, B. J., A. L. Delcher, S. M. Mount, J. R. Wortman, R. K. Smith Jr, L. I. Hannick, R. Maiti, C. M. Ronning, D. B. Rusch, and C. D. Town. 2003. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic. Acids. Res 31:5654–5666.

Hartman, P. S., N. Ishii, E.-B. Kayser, P. G. Morgan, and M. M. Sedensky. 2001. Mitochondrial mutations differentially affect aging, mutability and anesthetic sensitivity in *Caenorhabditis elegans*. Mech. Ageing Dev 122:1187–1201.

He, L., and G. J. Hannon. 2004. MicroRNAs: small RNAs with a big role in gene regulation. Nat. Rev. Genet 5:522–531.

Healy, T. M., and R. S. Burton. 2020. Strong selective effects of mitochondrial DNA on the nuclear genome. Proc. Natl Acad. Sci. USA 117:6616-6621.

Hill, G. E. 2017. The mitonuclear compatibility species concept. Auk 134:393–409.

———. 2019. Mitonuclear ecology. Oxford Univ. Press, Oxford, U.K.

Hill, G. E., J. C. Havird, D. B. Sloan, R. S. Burton, C. Greening, and D. K. Dowling. 2019. Assessing the fitness consequences of mitonuclear interactions in natural populations. Biol. Rev 94:1089–1104.

Hirase, S., and M. Ikeda. 2014. Long-term vicariance and post-glacial expansion in the Japanese rocky intertidal goby *Chaenogobius annularis*. Mar. Ecol. Prog. Ser 499:217–231.

———. 2015. Hybrid population of highly divergent groups of the intertidal goby *Chaenogobius annularis*. J. Exp. Mar. Biol. Ecol 473:121–128.

Hirase, S., M. Ikeda, M. Kanno, and A. Kijima. 2012. Phylogeography of the intertidal goby *Chaenogobius annularis* associated with paleoenvironmental changes around the Japanese Archipelago. Mar. Ecol. Prog. Ser 450:167–179.

Hirase, S., H. Takeshima, M. Nishida, and W. Iwasaki. 2016. Parallel mitogenome sequencing alleviates random rooting effect in phylogeography. Genome Biol. Evol 8:1267–1278.

Huang, D. W., B. T. Sherman, and R. A. Lempicki. 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc 4:44–57.

Hulsey, C. D., K. L. Bell, F. J. García-de-León, C. C. Nice, and A. Meyer. 2016. Do relaxed selection and habitat temperature facilitate biased mitogenomic introgression in a narrowly endemic fish? Ecol. Evol 6:3684–3698.

Kajitani, R., K. Toshimoto, H. Noguchi, A. Toyoda, Y. Ogura, M. Okuno, M. Yabana, M. Harada, E. Nagayasu, and H. Maruyama. 2014. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res 24:1384–1395.

Kamga, C., S. Krishnamurthy, and S. Shiva. 2012. Myoglobin and mitochondria: a relationship bound by oxygen and nitric oxide. Nitric Oxide 26:251–258.

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

Krogh, A., B. Larsson, G. Von Heijne, and E. L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes1. J. Mol. Biol 305:567–580.

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

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

Li, Y., Z. Ren, A. M. Shedlock, J. Wu, L. Sang, T. Tersing, M. Hasegawa, T. Yonezawa, and Y. Zhong. 2013. High altitude adaptation of the schizothoracine fishes (Cyprinidae) revealed by the mitochondrial genome analyses. Gene 517:169–178.

Librado, P., and J. Rozas. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452.

Luu, K., E. Bazin, and M. G. Blum. 2017. pcadapt: an R package to perform genome scans for selection based on principal component analysis. Mol. Ecol. Resour 17:67–77.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, and M. Daly. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303.

Meirmans, P. G., and P. H. Van Tienderen. 2004. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Mol. Ecol. Notes 4:792–794.

Mencía, Á., S. Modamio-Høybjør, N. Redshaw, M. Morín, F. Mayo-Merino, L. Olavarrieta, L. A. Aguirre, del I. Castillo, K. P. Steel, and T. Dalmay. 2009. Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. Nat. Genet 41:609–613.

Mishmar, D., E. Ruiz-Pesini, P. Golik, V. Macaulay, A. G. Clark, S. Hosseini, M. Brandon, K. Easley, E. Chen, and M. D. Brown. 2003. Natural selection shaped regional mtDNA variation in humans. Proc. Natl Acad. Sci. USA 100:171-176.

Morales, H. E., A. Pavlova, L. Joseph, and P. Sunnucks. 2015. Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. Mol. Ecol 24:2820–2837.

Muller, H. J. 1940. Bearing of the Drosophila work on systematics. Pp. 185–268 *in* J. Huxley, ed. The new systematics. Clarendon Press, Oxford, U.K.

Nagano, A. J., M. N. Honjo, M. Mihara, M. Sato, and H. Kudoh. 2015. Detection of plant viruses in natural environments by using RNA-Seq. Methods Mol. Biol 1236:89–98.

Nei, M., and S. Kumar. 2000. Molecular evolution and phylogenetics. Oxford Univ. Press, Oxford, U.K.

Oba, T., M. Kato, H. Kitazato, I. Koizumi, A. Omura, T. Sakai, and T. Takayama. 1991. Paleoenvironmental changes in the Japan Sea during the last 85,000 years. Paleoceanography 6:499–518.

Parra, G., K. Bradnam, and I. Korf. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23:1061–1067.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7:e37135.

Pons, J., S. Sonsthagen, C. Dove, and P. Crochet. 2014. Extensive mitochondrial introgression in North American Great Black-backed Gulls (Larus marinus) from the American Herring Gull (Larus smithsonianus) with little nuclear DNA impact. Heredity 112:226–239.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. De Bakker, and M. J. Daly. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet 81:559–575.

Puritz, J. B., C. M. Hollenbeck, and J. R. Gold. 2014. dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ 2:e431.

Ramsden, D. B., P. W. L. Ho, J. W. M. Ho, H. F. Liu, D. H. F. So, H. M. Tse, K. H. Chan, and S. L. Ho. 2012. Human neuronal uncoupling proteins 4 and 5 (UCP4 and UCP5): structural properties, regulation, and physiological role in protection against oxidative stress and mitochondrial dysfunction. Brain Behav. 2:468–478.

Rand, D. M. 2017. Fishing for adaptive epistasis using mitonuclear interactions. PLos Genet. 13:e1006662.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406–425.

Sakaguchi, S., T. Sugino, Y. Tsumura, M. Ito, M. D. Crisp, D. M. Bowman, A. J. Nagano, M. N. Honjo, M. Yasugi, and H. Kudoh. 2015. High-throughput linkage mapping of Australian white cypress pine (*Callitris glaucophylla*) and map transferability to related species. Tree Genet. Genomes. 11:121.

Sánchez-Caballero, L., S. Guerrero-Castillo, and L. Nijtmans. 2016. Unraveling the complexity of mitochondrial complex I assembly: a dynamic process. Biochim. Biophys. Acta Bioenerg. 1857:980–990.

Sato, M., S. Hosoya, S. Yoshikawa, S. Ohki, Y. Kobayashi, T. Itou, and K. Kikuchi. 2019. A highly flexible and repeatable genotyping method for aquaculture studies based on target amplicon sequencing using next-generation sequencing technology. Sci. Rep. 9:6904.

Silva, G., F. P. Lima, P. Martel, and R. Castilho. 2014. Thermal adaptation and clinal mitochondrial DNA variation of European anchovy. Proc. Royal Soc. B 281:20141093.

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212.

Sloan, D. B., J. C. Havird, and J. Sharbrough. 2017. The on-again, off-again relationship between mitochondrial genomes and species boundaries. Mol. Ecol 26:2212–2236.

Sripada, L., D. Tomar, and R. Singh. 2012. Mitochondria: one of the destinations of miRNAs. Mitochondrion. 12:593–599.

Stanke, M., and S. Waack. 2003. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19:215–225.

Tada, R. 1994. Paleoceanographic evolution of the Japan Sea. Palaeogeogr. Palaeoclimatol. Palaeoecol. 108:487–508.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28:2731–2739.

Tanabe, A. S. 2008. Phylogears version 1.5. Available via http://www.fifthdimension.jp/.

Tav, C., S. Tempel, L. Poligny, and F. Tahi. 2016. miRNAFold: a web server for fast miRNA precursor prediction in genomes. Nucleic. Acids. Res. 44:W181–W184.

Tempel, S., B. Zerath, F. Zehraoui, and F. Tahi. 2015. miRBoost: boosting support vector machines for microRNA precursor classification. RNA 21:775–785.

Torres-Bacete, J., E. Nakamaru-Ogiso, A. Matsuno-Yagi, and T. Yagi. 2007. Characterization of the NuoM (ND4) subunit in Escherichia coli NDH-1: conserved charged residues essential for energy-coupled activities. J. Biol. Chem. 282:36914–36922.

Toews, D. P. and A. Brelsford. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol. Ecol. 21:3907–3930.

Trapnell, C., L. Pachter, and S. L. Salzberg. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111.

Trapnell, C., A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, and L. Pachter. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7:562–578.

Trier, C. N., J. S. Hermansen, G.-P. Sætre, and R. I. Bailey. 2014. Evidence for mito-nuclear and sex-linked reproductive barriers between the hybrid Italian sparrow and its parent species. PLoS Genet 10:e1004075.

Wallace, D. C. 1999. Mitochondrial diseases in man and mouse. Science 283:1482–1488.

Wang, Y., Y. Shen, C. Feng, K. Zhao, Z. Song, Y. Zhang, L. Yang, and S. He. 2016. Mitogenomic perspectives on the origin of Tibetan loaches and their adaptation to high altitude. Sci. Rep. 6:1–10.

Wilson, C. C., and L. Bernatchez. 1998. The ghost of hybrids past: fixation of arctic charr (Salvelinus alpinus) mitochondrial DNA in an introgressed population of lake trout (*S. namaycush*). Mol. Ecol. 7:127–132.

Wilson, G. A., and B. Rannala. 2003. Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163:1177–1191.

Woolley, S., J. Johnson, M. J. Smith, K. A. Crandall, and D. A. McClellan. 2003. TreeSAAP: selection on amino acid properties using phylogenetic trees. Bioinformatics 19:671–672.

Yamada, T., H. Takakura, T. Jue, T. Hashimoto, R. Ishizawa, Y. Furuichi, Y. Kato, N. Iwanaka, and K. Masuda. 2016. Myoglobin and the regulation of mitochondrial respiratory chain complex IV. J. Physiol. 594:483–495.

Ye, C., C. M. Hill, S. Wu, J. Ruan, and Z. S. Ma. 2016. DBG2OLC: efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. Sci. Rep. 6:31900.

Zhang, C., K. L. Montooth, and B. R. Calvi. 2017. Incompatibility between mitochondrial and nuclear genomes during oogenesis results in ovarian failure and embryonic lethality. Development 144:2490–2503.

Zieliński, P., K. Nadachowska-Brzyska, B. Wielstra, R. Szkotak, S. Covaciu-Marcov, D. Cogălniceanu, and W. Babik. 2013. No evidence for nuclear introgression despite complete mtDNA replacement in the Carpathian newt (*Lissotriton montandoni*). Mol. Ecol.s 22:1884–1903.

Associate Editor: S. Collins
Handling Editor: T. Chapman

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supplementary Figure 1. Genetic diversity of each population based on nuclear genome SNP data obtained by RAD-sequencing.
Supplementary Figure 2. Individual hybrid indices of putative hybrid individuals between PO (Onagawa, Osaka, and Iyo) and SJ (Shimonoseki, Majima, and Asamushi) lineages based on SNP data obtained by RAD-sequencing.
Supplementary Figure 3. Cross validation scores when one to 10 clusters (K) were assumed in the ADMIXTURE analysis.
Supplementary Figure 4. Individual admixture proportions (q values) of the Taro B population, estimated by ADMIXTURE and SNP data obtained by RAD-sequencing ($K = 3$).
Supplementary Figure 5. Distribution of differences between observed (H_O) and expected (H_E) heterozygosity based on 1,000 bootstrap resampling of three individuals in the Taro B population, excluding those in clusters X and Y (Figure 5).
Supplementary Figure 6. Principal component analysis (PCA) based on divergent SNP loci of OXPHOS genes in the Taro B population.
Supplementary Figure 7. Pairwise genetic differentiation (weighted F_{ST}) between the PO and SJ lineages in non-overlapping 15 kb windows.
Supplementary Figure 8. Ratio of individuals that had the T-A-A alleles at the 20th, 21st, and 22nd SNP loci at different sampling locations.
Supplementary Table 1. *Chaenogobius annularis* samples used in this study
Supplementary Table 2. Summary of *de novo* genome sequencing performed on *Chaenogobius annularis*
Supplementary Table 3. Summary of RNA sequencing performed on *Chaenogobius annularis*
Supplementary Table 4. Summary of whole genome re-sequencing
Supplementary Table 5. Summary of *de novo* genome sequencing assembly performed on *Chaenogobius annularis*
Supplementary Table 6. Migrants per generations estimated by BayesAss
Supplementary Table 7. Nuclear oxidative phosphorylation (OXPHOS) genes genotyped in this study by amplicon sequencing
Supplementary Table 8. Genes located close to the 20 SNP loci that contributed to principal component 2 in the PCA analysis
Supplementary Table 9. Functional terms enriched for the genes in Supplementary Table 8
Supplementary Table 10. *Boleophthalmus pectinirostris* genes whose 3'UTR sequences contained the 51-bp region
Supplementary Table 11. Summary of small RNA sequencing