# Detailed kinship estimation for detecting bias among breeding families in a reintroduced population of the endangered bagrid catfish Tachysurus ichikawai

Received: 20 August 2023 | Revised: 17 January 2024 | Accepted: 25 January 2024

DOI: 10.1002/1438-390X.12183

Population Ecology WILEY

SPECIAL FEATURE

Ecological Perspectives of Pedigree Reconstruction with Genome-wide Data

# Detailed kinship estimation for detecting bias among breeding families in a reintroduced population of the endangered bagrid catfish *Tachysurus ichikawai*

**Hinano Mizuno**[1] | **Kouji Nakayama**[1] | **Tetsuya Akita**[2] | **Yasuyuki Hashiguchi**[3] | **Tomonori Osugi**[4] | **Hirohiko Takeshima**[5]

[1]Division of Applied Biosciences, Graduate School of Agriculture, Kyoto University, Kitashirakawa Oiwake-cho, Kyoto, Japan

[2]Fisheries Resources Institute, Japan Fisheries Research and Education Agency, Yokohama, Kanagawa, Japan

[3]Department of Biology, Faculty of Medicine, Osaka Medical and Pharmaceutical University, Osaka, Japan

[4]Environment Research Division, Water Resources Environment Center, Tokyo, Japan

[5]Research Center for Marine Bioresources, Faculty of Marine Bioscience, Fukui Prefectural University, Fukui, Japan

**Correspondence**
Kouji Nakayama, Division of Applied Biosciences, Graduate School of Agriculture, Kyoto University, Kitashirakawa Oiwake-cho, Sakyo, Kyoto 606-8502, Japan.
Email: nakayama.kouji.8z@kyoto-u.ac.jp

**Present address**
Hinano Mizuno, Toyota Central R&D Labs., Inc., Nagakute, Aichi, Japan.

## Abstract

In the context of initiatives focused on captive breeding and reintroduction of endangered animal species, it is crucial to minimize any bias in reproductive success during the reintroduction phase in order to preserve genetic diversity. One population of *Tachysurus ichikawai*, a critically endangered bagrid catfish endemic to Japan, faces a threat from the construction of a dam. To address this, a captive breeding program followed by translocation is being implemented. Multiple breeding families are involved in this process; however, if there is a bias in reproductive success among them after release, it will result in a decline in genetic diversity. To ascertain potential biases of reproductive success among released individuals, we attempted to identify the familial lineage of individuals born at the release site. Due to the unavailability of samples from the released individuals themselves, we reconstructed the pedigree of three generations using distant kinship relationships, such as grandparent–grandchild and uncle–aunt–nephew–niece relationships, with data of 2230–5674 single-nucleotide polymorphisms obtained from whole genome re-sequence, and three different software. Our findings indicate no bias between lineages in the first year after reintroduction, but a significant bias in the second year, emphasizing the need for continuous management and monitoring of reintroduced populations. This study demonstrates that monitoring kinship after reintroduction can correct lineage bias, which is critical for the prompt restoration of genetic diversity.

KEYWORDS

endangered species, kinships estimation, reintroduction, reproductive success bias, whole-genome resequencing

## 1 | INTRODUCTION

Captive breeding is one of the effective measures for conserving threatened populations. Conservation of endangered species must consider genetic diversity, not just population growth (Frankham et al., 2002). However, captive breeding of endangered species often involves only a few individuals, which can lead to inbreeding (Wajiki et al., 2015) and reintroduction of them may decrease the genetic diversity of a natural population (Philippart, 1995; Ryman & Laikre, 1991). Because low genetic diversity can negatively affect the future viability of a population (Charlesworth & Charlesworth, 1987; Keller & Waller, 2002), captive breeding programs are designed to conserve genetic diversity as much as possible (Fraser, 2008). Moreover, even with many individuals released or used as breeding parents within the reintroduction program, biased breeding after release can lead to a decline in genetic diversity. Nevertheless, there have been only a few investigations into reproductive bias after reintroduction (e.g., Jamieson, 2010), and as far as our knowledge goes, there are no documented instances in fishes. In stock enhancement programs of fishes, there have been instances of substantial reproductive bias in the seed production process (e.g., Sekino et al., 2003), and it is plausible that a comparable phenomenon is occurring after the reintroduction of endangered species.

To investigate whether any bias has occurred after reintroduction of endangered fish species, it is necessary to conduct a kinship analysis in the wild for 1–2 generations after release. The most basic scheme for captive breeding and subsequent reintroduction of fish involves capturing wild fish ($F_0$), breeding them to produce offspring ($F_1$), and releasing these offspring into the wild. The released $F_1$ individuals breed in the field and produce the next generation ($F_2$). To evaluate the reproductive contribution of each $F_0$ individual, it is sufficient to observe the proportion of $F_2$ individuals with a grandparent–grandchild (GG) relationship with the $F_0$ individual. However, as released fish are typically small, it is often not feasible to conduct genetic analyses of the released $F_1$ individuals themselves, so it is difficult to examine the parent–offspring relationships of the $F_0$–$F_1$ or $F_1$–$F_2$ generations. It is therefore necessary to directly examine the GG relationships between $F_0$ and $F_2$ generations. Additionally, it would be valuable to investigate the uncle–aunt–nephew–niece (UANN) relationships among the siblings of the released $F_1$ individuals and the wild-bred $F_2$ individuals, particularly in cases where genetic analysis of the siblings of $F_1$ is feasible, for instance, when they are preserved for subsequent breeding purposes. Typically, more genetic markers are required to elucidate more distant relationships such as GG or UANN, compared to those required for examining parent–offspring relationships (Snedecor et al., 2022). However, this is difficult in populations with low genetic diversity because there are few genetic differences among individuals; general methods of genetic analysis often do not yield a sufficient number of polymorphisms to estimate kinship (Collevatti et al., 2007).

In recent years, kinships have begun to be analyzed using single-nucleotide polymorphisms (SNPs) obtained by next-generation sequencers. For example, in rainbow trout, a panel of SNPs at 95 loci has been used to estimate parent–offspring relationships and to examine life-history characteristics such as age of participation in reproduction (Abadía-Cardoso et al., 2013). In addition, SNPs at 6437 loci in red hammerhead sharks obtained by diversity array technology sequencing (DArTseq) (Melville et al., 2017), an SNP acquisition method similar to restriction site-associated DNA sequencing (RAD-seq), has been used to estimate full-sibling (FS) and half-sibling (HS) relationships and discuss reproductive patterns (Marie et al., 2019). Nevertheless, there are few previous studies estimating GG or UANN relationships to determine familial lineages in wild fish, except for that by Delomas and Campbell (2021).

*Tachysurus ichikawai* (Siluriformes: Bagridae), the subject species of this study, is endemic to Japan, exclusively inhabiting the rivers that flow into Ise Bay and Mikawa Bay (Nakamura, 1963; Niwa, 1967). Human interference in river systems has resulted in the destruction of this species' natural habitat, leading to a decline in population size and placing the species at risk of extinction (Watanabe, 1998). Specifically, the construction of dams and weirs has had a significant detrimental impact on their habitat, both directly and indirectly (Mie Prefecture, 2005). Consequently, this species has been classified as an endangered species with a high risk of near-future extinction in the wild (Ministry of the Environment, Japan, 2020).

Recently, a population of this species has encountered a threat due to dam construction. As one of the conservation measures, an attempt is underway to collect parental fish from the area slated to be submerged by dam construction, breed them into several groups in a facility, and release the young fish at another location to establish an alternative population (Figure 1). The release of fish began in 2017, and since 2018 the successful reproduction of released individuals at the transfer sites has been confirmed (Shitara Dam Construction Office, 2023). If reproduction at the release site only involved individuals from a particular family, this would lead to a loss of genetic diversity and inbreeding depression. It is therefore important to estimate the reproductive contribution of released $F_1$ individuals in the reintroduced population. However,

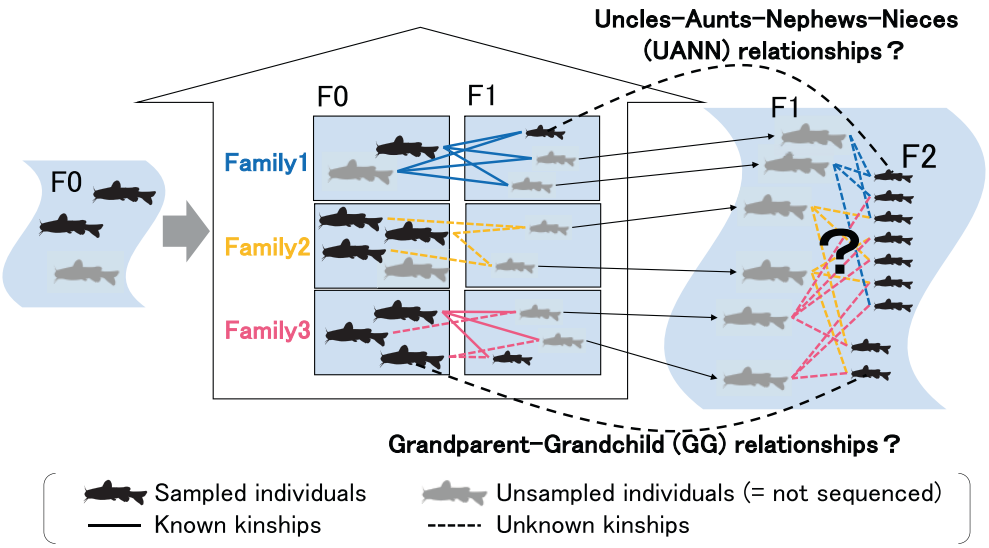


**FIGURE 1** Schematic diagram of captive breeding, reintroduction, and reproduction in the field subsequent to release. Wild adult fish ($F_0$ generation) obtained from the construction site of the dam are housed in multiple tanks (families), and the resultant offspring ($F_1$ generation) are released. The released $F_1$ individuals reproduce in natural conditions, giving birth to the succeeding generation ($F_2$ generation). Black silhouette denotes individuals with available samples for genetic analysis, while gray represents individuals without samples. Note that samples of the released $F_1$ fish themselves are not available (see text). Solid lines indicate known relationships based on pedigree information, and dashed lines indicate unknown relationships.

it is almost impossible to identify family structure by actually observing reproduction under natural conditions, so it is necessary to use genetic information to identify the family to which each individual belongs. There have been only a few genetic studies of *T. ichikawai*; one is a study of polymorphisms in the control region of mitochondrial DNA, and another uses microsatellite markers. In the study of mitochondrial DNA, all 75 individuals from 8 rivers analyzed were found to have a single haplotype (Watanabe & Nishida, 2003). And by surveying microsatellite markers, the population in the Toyokawa River system was reported to have lower genetic diversity than populations of the same species in other rivers (Mie Prefecture, 2006). This information suggests that the genetic diversity of the reintroduced population focused on in this study is expected to be low. Consequently, elucidating kinships within this population using reduced representation genome sequencing, such as RAD-seq, is deemed challenging due to the limited abundance of its polymorphic genetic markers.

In this study, we utilized whole-genome resequencing, the most informative method currently available in genetic research, to obtain numerous SNPs in the species with low genetic diversity. We attempted to estimate detailed kinship relationships within three generations in the reintroduced population to examine the reproductive contribution of each family based on more distant GG and UANN relationships. Our objective was to assess the potential occurrence of post-release reproductive bias in the captive breeding and reintroduction process of fishes.

## 2 | MATERIALS AND METHODS

### 2.1 | Target population and samples

*Tachysurus ichikawai* is a nocturnal species, residing in pools and flats with slow currents. It does not exhibit significant migratory patterns over its lifespan, and it is not uncommon for individuals to reside in a single pool for multiple years (Watanabe, 1995). The average lifespan for this species ranges from 3 to 6 years, with sexual maturity typically occurring at approximately 2–3 years (Watanabe, 1994a, 2008). The sex ratio within the breeding population is skewed toward females (Watanabe, 1994a), likely due to the higher mortality rate of adult males. Breeding is concentrated during June and July, during which males establish breeding territories and females visit them, thus demonstrating continuous polygamy (Watanabe, 1994a, 1994b, 2008).

The area 70 km upstream from the mouth of the Toyokawa River, which flows into Mikawa Bay, is an important habitat for a population of this species. In this location, the Shitara Dam is under construction (impounded water surface area: approximately 3 $km^2$), and when the dam is completed, the habitat of the population will disappear. Therefore, surveys and consideration

of specific conservation measures have been ongoing for about 20 years since 2003. The Shitara Dam Construction Office has conducted an environmental impact assessment (EIA) of the dam project based on the EIA Law; the assessment report was published in 2007. This report recommends conservation measures for *T. ichikawai* because of the impact of the project, and translocation experiments are being conducted (Figure 1).

Samples were collected from three successive generations in a single breeding population (Figure 1): the $F_0$ generation, collected near the dam construction site and used for breeding at the facility; the $F_1$ generation, produced by breeding between $F_0$ males and females at the facility; and the $F_2$ generation, produced by breeding between $F_1$ individuals under natural conditions at the translocated site. $F_0$ rearing and breeding at the facility were conducted in multiple tanks containing multiple $F_0$ males and females or one $F_0$ male and one $F_0$ female each, with each tank used for a single breeding season. For this study, the group of fish used for breeding in each tank will be referred to as a "family" for simplicity because in the majority of cases offspring from a given tank are genetically related (full or half-sibs). Some of the $F_1$ individuals born in each family were released to the transfer sites in the spring and fall, resulting in $F_1$ individuals from 12 families having been released by spring 2019 (families A1, A2, A3, A4, A5, A6A7A9, A10, A11, A12, A15, A16 and A17: A6A7A9 were initially treated as a separate group, but were later mixed. A8, A13, A14 were not released within the period covered by this study). Table S1 contains a summary of information about the $F_0$ individuals in each family.

The reproductive contribution of individual families was estimated by determining the number of $F_2$ individuals derived from each family. The simplest way to achieve this is to check the parent–offspring relationships between $F_1$ and $F_2$ generations. However, as previously stated, the $F_1$ individuals that were released were not sampled for genetic analysis. Therefore, it was imperative to investigate GG relationships between $F_0$ and $F_2$ generations, as well as UANN relationships between siblings of the released $F_1$ individuals and wild-bred $F_2$ individuals (Figure 1).

The selection criteria for individuals used in determining relationships were as follows. First, if samples from all $F_0$ individuals from a particular family were available, all $F_0$ individuals from that family were included for analysis. This applied to families A2, A3, A6A7A9, and A17. Second, if a sample from at least one $F_0$ individual in a family was not obtainable, the $F_1$ individuals (the siblings of the released $F_1$ individuals) from that family were analyzed instead. This approach was applied to families A1, A4, A5, A12, A15, and A16. For families where a single male and female pair were used for breeding, a single $F_1$ individual was selected, given that all $F_1$ individuals within the family were FS. For families where breeding involved multiple $F_0$ individuals, two $F_1$ individuals were analyzed. As for the $F_2$ generation, individuals born in the NN6 pool, one of the release sites and the pool that showed the highest reproductive success after release, were selected for analysis. Samples from the $F_2$ generation consisted of two year-classes: the 2018 year-class ($F_2$ individuals born during the breeding season around June–July 2018, sampled in June 2019 at approximately 1 year old, identified based on body length) and the 2019 year-class ($F_2$ individuals born during the breeding season around June–July 2019, sampled in June 2020). A total of 70 individuals were chosen for analysis based on these criteria: 15 $F_0$, 8 $F_1$, and 47 $F_2$ individuals (35 from the 2018 year-class and 12 from the 2019 year-class).

## 2.2 | Sequencing the reference genome of *T. ichikawai*

To obtain a reference genome for genome-wide population genetic analysis, we determined a draft genome for *T. ichikawai*. Genomic DNA was extracted from a clip of the adipose fin from one $F_0$-generation individual reared at the Aichi Prefectural Fisheries Experimental Station, Inland Fisheries Research Institute Mikawa-Ichinomiya Guidance Center in Toyokawa City, Aichi Prefecture, Japan, using the standard phenol–chloroform method. The quality of the genomic DNA sample was assessed by using an automated gel electrophoresis system (2200 TapeStation; Agilent Technologies, Santa Clara, CA, USA). The library of the genomic DNA was prepared by the single-tube long fragment read (stLFR) method (O. Wang et al., 2019), and sequenced on a paired-end $2 \times 100$ nt lane high-throughput sequencer (MGISEQ-2000RS; MGI Tech, Shenzhen, China).

## 2.3 | Genome size estimation

To estimate the *T. ichikawai* genome size, k-mer frequency analysis was performed using KmerGenie (Chikhi & Medvedev, 2014). First, stLFR barcodes in the MGISEQ reads were trimmed by the stLFR2Supernova pipeline (O. Wang et al., 2019). Second, the best k-mer length was estimated from the barcode-trimmed reads. Third, genome size was estimated from the frequency distribution of the best k-mer length.

## 2.4 | Assembly of the *T. ichikawai* genome

A scaffold-level assembly of the *T. ichikawai* genome was generated under the stLFR2Supernova pipeline (O. Wang et al., 2019) with Supernova 2.1.1 (Weisenfeld et al., 2017). Parameters for the Supernova run were set to MAX_READS = 400,000,000 and MINSIZE = 200. The assembly was polished automatically with the Pilon software tool (Walker et al., 2014). The quality of the genome assembly was evaluated using the completeness estimated by BUSCO version 5.4.4 (Simão et al., 2015) with a database of ray-finned fishes (Actinopterygii odb10). Transposable elements (TEs) and other repeat sequences of the *T. ichikawai* genome were identified using RepeatModeler version 2.0.3 (Flynn et al., 2020) with Dfam 3.7 repeat database. Second, the TEs and repeats within the *T. ichikawai* genome were soft-masked by RepeatMasker version 4.1.2-p1 (http://www.repeatmasker.org).

## 2.5 | Genome-wide SNP heterozygosity estimation

To assess the genetic diversity of the *T. ichikawai* genome, we estimated the genome-wide SNP heterozygosity of the reference individual. Initially, stLFR barcode-trimmed MGISEQ reads were mapped to the reference genome using NextGenMap (Sedlazeck et al., 2013), and a binary format alignment/map (BAM) file was generated. The BAM file was sorted by SAMTools version 1.7 (Li et al., 2009). Next, local realignments of INDELs in the sorted BAM file were conducted by GATK v3.8.1 (McKenna et al., 2010). Then, a genomic variant call format (GVCF) file of the reference individual was generated by GATK HaplotypeCaller with options -hets 0.001 and -indelHeterozygosity 0.001. Finally, SNPs of the reference individual were called and an output variant call format (VCF) file was generated using the GATK GenotypeGVCF tool. For genotyped SNPs, variant filtering was applied using the GATK VariantFilteration tool with cutoff values as follows: MQ >30.00, SOR <4.000, QD >2.00, FS <60.000, MQRankSum >−20.000, ReadPosRankSum >−10.000, and ReadPosRankSum <10.000.

## 2.6 | Library preparation and resequencing

The adipose fins of the 70 individuals selected above that were collected non-invasively and preserved in 99% ethanol were used for DNA extraction. DNA extraction was performed using the DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. DNA concentrations were measured using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Massachusetts, USA). Agarose electrophoresis was performed using KANTO ST (Kanto Chemical, Tokyo, Japan) high gel strength (1% agarose with TAE) to confirm the quality of genomic DNA.

Libraries were prepared from the DNA extracts following the Hackflex protocol (Gaio et al., 2022). For indexes, we used UD Index Adapter (Integrated DNA Technologies, Iowa, USA). Concentrations of the DNA library were measured by a Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific) and the KAPA Library Quantification Kit (Kapa Biosystems) for the StepOnePlus Real-Time PCR System (Thermo Fisher Scientific). Library size was measured using an automated gel electrophoresis system (2200 TapeStation high sensitivity D1000 [Agilent] or 4200 TapeStation high sensitivity D1000 [Agilent]). Fragments in the DNA library were sequenced by using a genome sequencer (HiSeq X; Illumina, California, USA) at 151 bp paired ends.

## 2.7 | Obtaining genome-wide SNPs in multiple samples

After trimming adapter sequences with Trimmomatic v0.39 (Bolger et al., 2014), only paired-end data were mapped to a reference genome using NextGenMap v0.5.2 (Sedlazeck et al., 2013), and BAM files were generated. After the BAM files were sorted using SAMTools v1.7 (Li et al., 2009), each individual was genotyped and VCF files were generated by HaplotypeCaller in Picard v2.25.1 (Broad Institute, 2021) and GATK v4.2.0.0 (McKenna et al., 2010). Then, only biallelic SNPs were identified and hard filtered (MQ <30.00, SOR >4.000, QD <2.00, FS >60.000, MQRankSum <−20.000, ReadPosRankSum <−10.000, and ReadPosRankSum >10.000) with VariantFiltration in GATK. We also determined the sequence of the mitochondrial DNA from $F_0$ individual f12A using GetOrganelle (Jin et al., 2020) and mapped the sequences of all other individuals for this reference and searched for SNPs in the same way as for the whole genome, except that the depth was set to >50 in the filtering conditions for detecting SNPs in mitochondrial DNA.

## 2.8 | Evaluation of SNP filtering criteria for kinship estimation

We next examined the filtering parameters for SNPs using a total of 20 individuals of the $F_0$ and $F_1$

generations, including kinships known from the rearing conditions. The known kinships among the 20 individuals included three parent–offspring relationships, one FS, and a maximum of five HS. We considered the parameters under which all known kinships were correctly reproduced as good parameters that could accurately estimate unknown kinships.

The filtering parameters considered were [1] minimum depth (MIN_DP), 5; [2] maximum depth (MAX_DP), 30–200; [3] minimum mean depth (MIN_MEAN_DP), 15; [4] minimum genotyping quality (MIN_GQ), 20–30; [5] call rate (CR), 0.7–1.0; [6] minor allele frequencies (MAF), 0.01–0.1; [7] deviation from Hardy–Weinberg equilibrium (HWE), 0.00001–0.01; [8] maximum heterozygosity (HET), 0.6–0.8; and [9] linkage disequilibrium (LD), 0.1–0.3. The ranges of values following each parameter are the minimum–maximum values considered. Terms for the parameters and their values were determined based on several previous studies (Barnes & Breen, 2010; Chen et al., 2017; Dou et al., 2017; Miyagawa et al., 2008; Nishida et al., 2008; O'Leary et al., 2018; Roshyara et al., 2014).

We used COLONY 2.0 (Jones & Wang, 2010) and Sequoia (Huisman, 2017) to estimate kinships. First, we varied the values of each of the nine filtering parameters listed above, along with the allele dropout rate and error rate settings in COLONY 2.0, to identify multiple candidate combinations that could accurately reproduce known kinships, and then identified the ones for which Sequoia could also produce correct estimates. Because observations from laboratory experiments suggest that *T. ichikawai* practices "continuous" polygamy (Watanabe, 1994a, 1994b), we assumed random mating in COLONY 2.0. The estimation accuracy was set to very high, and run length was set to 3 out of a range of 1–4. Vcftools (Danecek et al., 2011), Plink (Purcell et al., 2007), and SnpEff (Cingolani et al., 2012) were used for SNP filtering.

## 2.9 | Estimation of kinship between $F_2$ and captive individuals

From the SNP set filtered by the criteria determined above, COLONY 2.0 (Jones & Wang, 2010), Sequoia (Huisman, 2017), and gRandma (Delomas & Campbell, 2021) were used to estimate relationships between $F_2$ and captive individuals. COLONY 2.0 was used to estimate FS and HS relationships between $F_0$ individuals and between $F_2$ individuals. Both COLONY 2.0 and gRandma were used to estimate GG relationships between $F_0$ and $F_2$ individuals. Sequoia was used to estimate UANN relationships between $F_1$ and $F_2$ individuals. Here, HS, GG, and UANN have the same relatedness (Thompson, 1975). Although these three relationships are inherently genetically indistinguishable without utilizing linkage information or employing three or more individuals jointly for the relationship inference (J. Wang, 2007), it is important to note that in this study, $F_0$ consistently represents the grandparent generation, $F_1$ corresponds to the uncle/aunt generation (parent generation), and $F_2$ represents the grandchild generation (Figure 1). Therefore, if a HS relationship is estimated between $F_0$ and $F_2$ individuals using COLONY 2.0, it actually indicates a GG relationship (J. Wang, pers. comm.). Specifically, all 62 individuals from the $F_0$ and $F_2$ generations were included as candidate offspring in COLONY 2.0, and if an HS relationship was estimated between $F_0$ and $F_2$ individuals, it was considered a GG relationship. Likewise, all 55 individuals from the $F_1$ and $F_2$ generations were entered as candidate offspring in COLONY 2.0, and UANN relationships were considered as HS relationships.

For the COLONY 2.0 settings, both the allele dropout rate and error rate were set to 0.01 based on the results of the preliminary considerations alongside other filtering conditions, and inbreeding was considered. Other settings were the same as those used in the consideration of filtering parameters (random mating was assumed, estimation accuracy was very high, and run length was set to 3). From a prior study, those relationships with a support probability of 0.8 or greater in COLONY 2.0 were considered to be true kinship (Thow et al., 2022).

gRandma can provide estimates for trio relationships between paternal and maternal grandparents and one grandchild. The estimated GG trios are then output with the Mendelian incompatibility (MI: the closer to 0, the less genetic inconsistency) and the log likelihood ratio (LLR) assuming that the trio is unrelated. It is also possible to calculate the false positive and false negative rates for each MI when the LLR threshold is set to a certain value. In this study, the false positive and false negative rates for each MI were calculated for LLR = 30, 50, 60, 70, 80, and 90 as candidates for the threshold, and LLR = 70 was adopted because at this value the false positive rate was 0.06 and the false negative rate was 0 when MI = 0 (Table S2). The results of the kinship estimation are presented as "robust" if LLR ≥70 and MI = 0, and "weak" if LLR <70 and MI = 0 or LLR ≥70 and MI >0. A total of 62 individuals ($F_0$–$F_2$) were used, the same as in the estimation of GG using COLONY 2.0, and the filtering conditions were also the same (MIN_DP = 5, MAX_DP = 50, MIN_MEAN_DP = 15, MIN_GQ ≥30, CR >0.9, MAF ≥0.03, HWE <0.00001, HET ≤0.7, LD ≤0.2). The allele dropout rate and error rate were both set to 0.01, the same as in COLONY 2.0.

### 2.10 | Identification of families for $F_2$ individuals and investigation of reproductive bias

The results of the estimations of GG or UANN relationships were combined to identify the families to which each $F_2$ individual belonged. To identify the family, we also indirectly used the FS relationships between $F_2$ individuals. For example, if a GG relationship is estimated between $F_0$ individual A and $F_2$ individual B, and an FS is estimated between $F_2$ individual B and $F_2$ individual C, then individual A and individual C are also considered to have a GG relationship, even if a GG relationship is not detected between A and C.

To test the significance of reproductive bias, we performed Pearson's $\chi$-square goodness-of-fit tests with the null hypothesis that all families contributed equally to breeding regardless of the number of $F_1$ individuals released, that is, each family was responsible for the birth of an equal number of $F_2$ individuals.

The families taken into consideration were determined based on the $F_1$ birth year and release status. Table S3 shows the number of $F_1$ individuals released from each family from spring 2017, when releases to the NN6 pool began, to spring 2019, when all individuals that could have contributed to the birth of the 2019 year-class had been released. Here, only three $F_1$ groups—A1, A2, and A3—born at the facility in 2016 and released in spring of 2017 and 2018, could be the parents of the 2018 year-class. This is because age at first maturity of *T. ichikawai* is 2 years old or later. In addition, individuals listed as spring releases were released around May, just before the June–July breeding season of this species, so individuals released in the spring of 2018 may have also contributed to the birth of the 2018 year-class. In the same way, the families that could be parents of the 2019 year-class are the A1, A2, and A3 groups released in spring of 2017, 2018, and 2019, and the A4, A5, and A6A7A9 groups, born at the facility in 2017 and released in fall 2017, spring 2018, and spring 2019, for a total of six groups. Although the possibility of their reproducing is unlikely, the A10, A11, A12, A15, A16, and A17 families were also included in the analysis as they were released at the same location.

## 3 | RESULTS

### 3.1 | Draft genome assembly of *T. ichikawai*

In the *T. ichikawai* individual used for genome sequencing, 100.6 Gb raw reads were generated by sequencing the stLFR library using MGISEQ-2000RS. The draft genome sequences are available in the DDBJ Sequence Read Archive (DRA) database (DRA ID DRA017372, BioProject ID PRJDB16846). In the k-mer analysis for genome size estimation, the best k-mer length was 77 and the haploid genome size was estimated as 586.8 Mb.

The draft genome of *T. ichikawai* was assembled into 65,543 scaffolds. The assembly size was 693.4 Mb, which was slightly larger than that estimated by k-mer analysis. The N50 of the genome assembly was 302,574 bp and the average scaffold length was 10,580 bp. Nucleotide length of the largest scaffold was 5,257,662 bp. The completeness of the protein-coding genes in the genome assembly was estimated to be 75.8% (74.7% single and 1.1% duplicated copies). We consider this assembly to be sufficient for the reference genome of the SNP detection in this study, and it was used for further analysis.

### 3.2 | Genome-wide SNPs from multiple samples and assessment of filtering parameters

We identified the following numbers of SNPs: 173,375 for $F_0$ alone (15 individuals); 291,546 for $F_2$ alone (47 individuals); 317,193 for $F_0$ (15 individuals) + $F_2$ (47 individuals); 310,249 for $F_1$ (8 individuals) + $F_2$ (47 individuals); and 335,104 for $F_0$ (15 individuals) + $F_1$ (8 individuals) + $F_2$ (47 individuals). Approximately 54% of the SNPs were singletons, occurring only once in the entire population, as calculated for 20 individuals including $F_0$ and $F_1$.

The parameters investigated for SNP filtering are shown in Table S4, and filtering condition set no. 20, which accurately reproduced the known kinships, was adopted as the filtering combination for estimating unknown kinships. First, we identified eight sets of conditions (numbers 13, 18–23, and 25 in Table S4) that accurately reproduced all of the known kinships using COLONY 2.0 for the dataset from 20 individuals. The higher the sequential number among these eight sets of conditions, the "looser" they are; that is, the greater the number of loci left after filtering. Filtering condition set no. 20 was selected as the filtering conditions for estimating unknown relationships (MIN_DP = 5, MAX_DP = 50, MIN_MEAN_DP = 15, MIN_GQ ≥30, CR >0.9, MAF ≥0.03, HWE <0.00001, HET ≤0.7, LD ≤0.2) because these yielded a probability of 1.000 (= maximum) for all known kinships, and the number of loci left was the largest among those with the same results. However, the MAF value was adjusted based on the number of individuals analyzed simultaneously, so that only heterozygous

SNP loci that occurred in a single individual would be filtered out (the threshold value of MAF ≥0.03 shown here is for the case of 20 individuals).

The number of loci in each combination of generations after filtering was 2479 for $F_0$ alone (15 individuals), 2230 for $F_2$ alone (47 individuals), 2541 for $F_0$ (15 individuals) + $F_2$ (47 individuals), 5674 for $F_1$ (8 individuals) + $F_2$ (47 individuals), and 3045 loci for $F_0$ (15 individuals) + $F_1$ (8 individuals) + $F_2$ (47 individuals).

### 3.3 | Kinship between $F_2$ and captive individuals

We first estimated kinships by COLONY 2.0 and Sequoia for each of the filtered SNP sets. COLONY 2.0 estimated the following kinships with a support probability of 0.8 or higher: 1 pair for FS within $F_0$, 9 pairs for HS within $F_0$, 18 pairs for FS within $F_2$, 78 pairs for HS within $F_2$, 26 pairs for GG between $F_0$ and $F_2$, and 15 pairs for UANN between $F_1$ and $F_2$. In COLONY 2.0, we treated HS relationships between generations as GG or UANN. The detailed results from COLONY 2.0 are shown in Tables S5–S10 for all outputs regardless of the support probability. The results from Sequoia were 1 pair for FS within $F_0$, 2 pairs for HS within $F_0$, 19 pairs for FS within $F_2$, 9 pairs for HS within $F_2$, 9 pairs for GG between $F_0$ and $F_2$, and 9 pairs for UANN between $F_1$ and $F_2$ (shown in Tables S11–S16). The GG relationships estimated from gRandma were 42 robust trios (Table S17) and 20 weak trios (Table S18). The results from these three programs are presented in Figure 2 as solid lines.

### 3.4 | Family identification for $F_2$ individuals and reproductive bias

The families from which all 47 $F_2$ individuals originated, as determined from the kinship estimation, are shown in Table 1. Thirteen, twenty-four, and thirty-seven $F_2$ individuals were identified as descendants of the A1, A2, and A3 families, respectively. The total number of $F_2$ individuals exceeds 47, since one $F_2$ individual may have originated from two families (maternal and paternal). Conversely, no $F_2$ individuals belonging to families other than those named above were found. Thus, only three families were identified in this study: A1, A2, and A3. Three family lineages, denoted as A1, A2, and A3, were potentially implicated in the 35 $F_2$ individuals in the 2018 year-class, and kinship analysis revealed 13, 17, and 25 involvements, respectively (Table 2). No significant reproductive bias was detected among the lineages (Table S19). On the other hand, 12 individuals of the 2019 year-class showed significant reproductive bias, with 0, 7, 12, 0, 0, and 0 involvement detected among six family lineages, namely A1, A2, A3, A4, A5, and A6A7A9. For the $F_2$ individuals for which only one family lineage was estimated, there exists the potential presence of an additional familial lineage that eluded estimation. Consequently, we assessed the scenario wherein the undetermined familial lineage was uniformly distributed among all lineages for $F_2$ individuals with a sole familial lineage estimation. Even under these conditions, no statistically significant bias was observed in the 2018 year-group whereas a significant bias was evident in the 2019 year-group (data not shown).

We excluded two cases from our identification of families as they were clearly false positives: UANN 190619-NN6-59 and $F_1$5A (which would have indicated that 190,619-NN6-59 was from the A6A7A9 family, but $F_1$ in A6A7A9 could not be the parent of the 2018 year-class because of age), and GG between 190,619-NN6-63 and f9A (which would have indicated that 190,619-NN6-63 was from the A17 family, but an $F_1$ individual of A17 could not be the parent of the 2018 year-class because of age).

### 3.5 | Mitochondrial DNA sequence variation

We determined a mitochondrial DNA sequence of 16,528 bases with an average depth of 385.5. The minimum depth for all bases was greater than 50, so no SNPs were excluded by filtering with MIN_DP = 50. Only 2 biallelic SNPs were identified from all 70 $F_0$, $F_1$, and $F_2$ individuals. These two loci in the mitochondrial DNA sequence were in the ND6 gene region (gene encoding NADH dehydrogenase 6) and the Cytb gene region (gene encoding cytochrome *b*) in the first and second codon positions, respectively. Only two haplotypes were found at these loci. The haplotypes of all individuals are shown in Table S20.

## 4 | DISCUSSION

### 4.1 | Genetic diversity of *T. ichikawai*

The genome-wide SNP heterozygosity of *Tachysurus ichikawai* ($1.6 \times 10^{-4}$/bp) estimated from the individual of the reference genome was considerably lower than its congeneric species *T. fulvidraco* ($4.5 \times 10^{-3}$/bp; Gong et al., 2018). Additionally, the heterozygosity of *T. ichikawai* was much lower than that of some critically

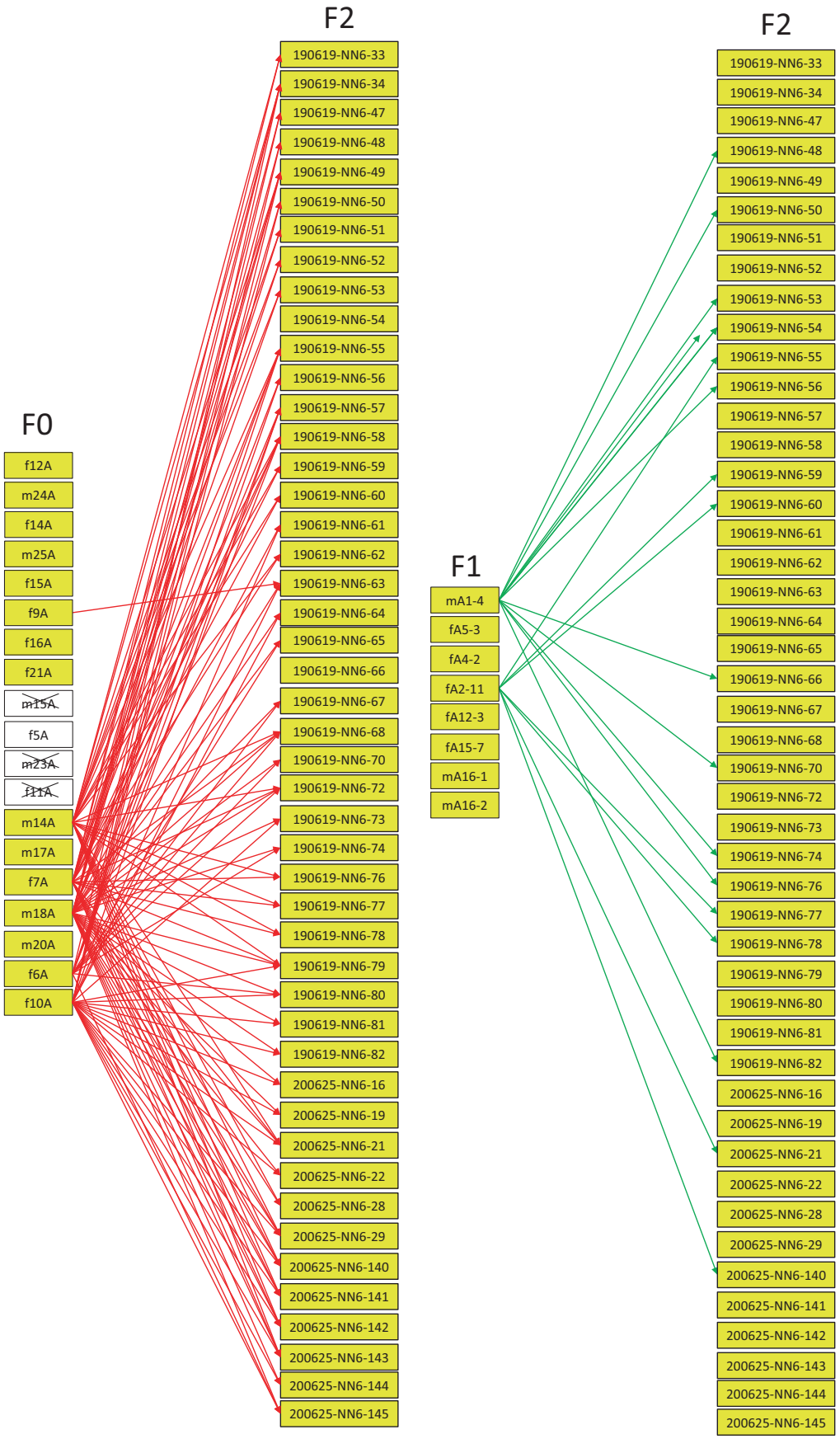


**FIGURE 2** Legend on next page.

endangered teleost species listed on the IUCN Red List, such as large yellow croaker ($3.58 \times 10^{-3}$/bp; Wu et al., 2014) or European eel ($1.48–1.59 \times 10^{-2}$/bp; Jansen et al., 2017). However, it remains uncertain whether the reduction in genetic diversity is a recent consequence of declining population size or has historical origins. Furthermore, approximately 54% of the SNPs determined in this study for a group of 20 individuals including $F_0$ and $F_1$ generations, were singletons. Singleton SNPs exhibit mutations in only one allele among all individuals examined, so they do not provide useful information for kinship inference. Thus, despite our acquisition of a substantial number of SNPs, we estimated that approximately half of them were uninformative loci for kinship analysis.

Despite the challenge posed by limited genetic diversity, frequent occurrence of singleton SNPs, and high level of inbreeding demonstrated by frequent HS and FS occurrence among $F_0$s (Tables S8 and S9), we successfully accomplished a kinship analysis by employing whole-genome resequencing. An alternative method incorporating reduced genome analysis, such as RAD-seq, would have yielded an insufficient quantity of SNPs to facilitate a comprehensive kinship analysis. Although it is imperative to carefully select suitable loci for kinship analysis, we deem the use of whole-genome resequencing for pedigree analysis in wild animals exhibiting low genetic diversity and inbreeding to be effective.

## 4.2 | Reproductive bias in the reintroduced population

By combining the kinship estimates from COLONY 2.0 and Sequoia, we were able to identify at least one family for all $F_2$ individuals. This is a significant accomplishment of our study, as there have been few instances where detailed kinships, such as GG and UANN relationships, have been identified in a highly inbred population. In families where breeding involved several $F_0$ individuals and at least one $F_0$ individual was not accessible, we used two $F_1$ individuals alongside. If not all $F_0$ individuals in the tank contributed to producing the two sampled $F_1$ individuals, it could have resulted in underestimating the number of descendants in this family. On the other hand, for some $F_2$ individuals, we indirectly estimated the familial lineage. That is, we considered the $F_2$ individual, in a full-sibling relationship with another $F_2$ individual whose familial lineage was identified, to belong to the same family. This is a two-step estimation, which might have caused a proliferation of inference errors. However, it was thought to be unlikely that these would have remarkably affected the results, as familial lineage was successfully estimated for all $F_2$, with the combination of three different software.

We next investigated whether there was any bias in reproductive contribution among the identified families. In the 2018 year-class, there was no statistically significant difference between the contributions from the three families that were likely involved in reproduction. This is a desirable outcome, as outlined in the introduction, because bias among families can result in a loss of genetic diversity (Willoughby et al., 2017). However, in the 2019 year-class, we observed a significant bias in breeding contribution among families. This was largely influenced by the fact that no $F_2$ individuals were found from the A4, A5, or A6A7A9 families. We anticipated that $F_1$ individuals from these families would likely have had few offspring in 2019. It is probable that a reduction in the water level of the NN6 pool, caused by the collapse of a downstream weir due to a flood in fall 2018, had a negative impact on the growth and survival rates of $F_1$ individuals in families A4, A5, and A6A7A9 (Shitara Dam Construction Office, unpublished information). Although *T. ichikawai* typically reaches sexual maturity at age 2–3 years (Watanabe, 1994a, 2008), it is possible that only a few $F_1$ individuals from the A4, A5, and A6A7A9 families had reached maturity during the 2019 breeding season because of delayed growth. Even if some $F_1$ individuals from these families had reached maturity, they would have been smaller than the $F_1$ individuals from the A1, A2, and A3 families, which were 1 year older, and therefore at a disadvantage in breeding competition. Moreover, the low water level may have decreased the overall reproductive rate in 2019. The number of F1 individuals able to produce offspring in 2019 was small, as only 12 F2 samples from year-class 2019 were obtained, despite a vigorous sampling effort, of which six individuals had the same estimated grandparents (m14A and f7A, m18A and f10A), and four individuals had the same grandparents (m18A and f10A), indicating that the number of $F_1$ individuals that were able to produce

**FIGURE 2** The estimated relationships by COLONY 2.0, Sequoia, and gRandma. Left panel: Grandparent–grandchild (GG) relationships between $F_0$ and $F_2$ generations. Right panel: Uncle–aunt–nephew–niece (UANN) relationships between $F_1$ and $F_2$ generations. Individuals shaded in yellow represent the samples used for genetic analyses. Individuals marked with a cross were not sampled for genetic analysis. GG relationship between 190,619-NN6-63 and f9A was presumed to be false positive (see text for details). See also Table 1 for detailed relationships.

**TABLE 1** Families of all 47 F2 individuals identified from the kinship analysis.

| Name of $F_2$ | Family | Results used to identify family[a] | Individuals of $F_0$ or $F_1$ generations with estimated GG or UANN relationships |
|---|---|---|---|
| 190619-NN6-54 | A1 | (UANN by Se) | mA1-4 |
| 190619-NN6-66 | A1 | (UANN by Se) | mA1-4 |
| 190619-NN6-51 | A1 A2 | (UANN by Co), (rob GG by gR) | m14A, f7A, mA1-4 |
| 190619-NN6-52 | A1 A2 | (we GG by gR), (Comb) | m14A, f7A |
| 190619-NN6-53 | A1 A2 | (UANN by Co), (rob GG by gR) | m14A, f7A, mA1-4 |
| 190619-NN6-76 | A1 A2 | (UANN by Co), (rob GG by gR) | m14A, f7A, mA1-4 |
| 190619-NN6-48 | A1 A3 | (UANN by Co), (we GG by gR) | m18A, f10A, mA1-4 |
| 190619-NN6-50 | A1 A3 | (UANN by Co), (we GG by gR) | m18A, f10A, mA1-4 |
| 190619-NN6-56 | A1 A3 | (GG by Co), (UANN by Se), (rob GG by gR) | m18A, f6A, mA1-4 |
| 190619-NN6-61 | A1 A3 | (we GG by gR), (Comb) | m18A, f10A |
| 190619-NN6-70 | A1 A3 | (GG by Co), (UANN by Co), (UANN by Se), (rob GG by gR) | m18A, f6A, mA1-4 |
| 190619-NN6-74 | A1 A3 | (GG by Co), (UANN by Co), (rob GG by gR) | m18A, f6A, mA1-4 |
| 190619-NN6-82 | A1 A3 | (UANN by Co), (rob GG by gR) | m18A, f10A, mA1-4 |
| 190619-NN6-33 | A2 | (rob GG by gR) | m14A, f7A |
| 190619-NN6-59 | A2 | (GG by Co), (UANN by Se), (rob GG by gR) | m14A, f7A, fA2-11, f15A |
| 190619-NN6-60 | A2 | (GG by Co), (UANN by Se), (rob GG by gR) | m14A, f7A, fA2-11 |
| 190619-NN6-62 | A2 | (rob GG by gR) | m14A, f7A |
| 190619-NN6-77 | A2 | (UANN by Co), (rob GG by gR) | m14A, f7A, fA2-11 |
| 190619-NN6-78 | A2 | (UANN by Co), (UANN by Se), (rob GG by gR) | m14A, f7A, fA2-11 |
| 200625-NN6-28 | A2 | (rob GG by gR) | m14A, f7A |
| 190619-NN6-34 | A2 A3 | (GG by Co), (rob GG by gR) | m14A, f7A, m18A, f10A |
| 190619-NN6-49 | A2 A3 | (we GG by gR) | m14A, f7A |
| 190619-NN6-55 | A2 A3 | (GG by Co), (UANN by Co), (rob GG by gR), (we GG by gR) | m14A, f7A, m18A, f6A, fA2-11 |
| 190619-NN6-58 | A2 A3 | (GG by Co), (rob GG by gR) | m14A, f7A, m18A, f6A |
| 190619-NN6-68 | A2 A3 | (GG by Co), (we GG by gR) | m14A, f7A, m18A, f6A |
| 190619-NN6-72 | A2 A3 | (GG by Co), (rob GG by gR) | m14A, f7A, m18A, f6A |
| 190619-NN6-79 | A2 A3 | (GG by Co), (rob GG by gR) | m14A, f7A, m18A, f10A |
| 200625-NN6-21 | A2 A3 | (GG by Co), (UANN by Co), (UANN by Se), (we GG by gR) | m14A, f7A, m18A, f10A, fA2-11 |
| 200625-NN6-29 | A2 A3 | (rob GG by gR), (we GG by gR) | m14A, f7A, m18A, f10A |
| 200625-NN6-140 | A2 A3 | (GG by Co), (UANN by Co), (we GG by gR) | m14A, f7A, m18A, f10A, fA2-11 |
| 200625-NN6-141 | A2 A3 | (GG by Co), (rob GG by gR) | m14A, f7A, m18A, f10A |
| 200625-NN6-142 | A2 A3 | (rob GG by gR) | m14A, f7A, m18A, f10A |
| 200625-NN6-143 | A2 A3 | (rob GG by gR), (we GG by gR) | m14A, f7A, m18A, f10A |
| 190619-NN6-47 | A3 | (we GG by gR) | m18A, f6A |
| 190619-NN6-63 | A3 | (GG by Co), (rob GG by gR) | m18A, f10A |
| 190619-NN6-64 | A3 | (we GG by gR) | m18A, f10A |
| 190619-NN6-65 | A3 | (GG by Co), (rob GG by gR) | m18A, f6A |
| 190619-NN6-67 | A3 | (GG by Co), (rob GG by gR) | m18A, f6A |
| 190619-NN6-73 | A3 | (we GG by gR) | m18A, f10A, m20A |
| 190619-NN6-81 | A3 | (we GG by gR) | m18A, f10A |

(Continues)

TABLE 1 (Continued)

| Name of $F_2$ | Family | Results used to identify family[a] | Individuals of $F_0$ or $F_1$ generations with estimated GG or UANN relationships |
|---|---|---|---|
| 200625-NN6-16 | A3 | (GG by Co), (rob GG by gR) | m18A, f10A |
| 200625-NN6-19 | A3 | (GG by Co), (rob GG by gR) | m18A, f10A |
| 200625-NN6-22 | A3 | (we GG by gR) | m18A, f10A |
| 200625-NN6-144 | A3 | (we GG by gR) | m18A, f10A |
| 190619-NN6-57 | A3 A3 | (rob GG by gR) | m18A, f6A, f10A |
| 190619-NN6-80 | A3 A3 | (GG by Co), (rob GG by gR) | m18A, f6A, f10A |
| 200625-NN6-145 | A3 A3 | (GG by Co), (rob GG by gR) | m18A, f6A, f10A |

*Note*: (GG by Co) GG (grandparent–grandchild) relationship estimated by COLONY 2.0, (UANN by Co) UANN (uncle–aunt–nephew–niece) relationship estimated by COLONY 2.0, (UANN by Se) UANN relationship estimated by Sequoia, (rob GG by gR) Robust GG relationship estimated by gRandma, (we GG by gR) Weak GG relationship estimated by gRandma, (Comb) Whose family was indirectly identified by the estimation of FS relationships by COLONY 2.0 with individuals whose family was identified by COLONY 2.0, Sequoia, or gRandma (other methods failed to identify the family).
[a]The following numbers indicate the results of kinship estimation used to identify the family from which each $F_2$ was born.

TABLE 2 Familial lineage composition for each year-class.

| | A1 family | A2 family | A3 family | A4 family | A5 family | A6A7A9 family | Total |
|---|---|---|---|---|---|---|---|
| 2018 year-group | 13 | 17 | 25 | – | – | – | 55 |
| 2019 year-group | 0 | 7 | 12 | 0 | 0 | 0 | 19 |

offspring in 2019 was small. However, the observed bias in 2019 could be a short-term result of demographics and is not necessarily indicative of a loss in genetic diversity at the population scale.

## 4.3 | SNP filtering criteria suitable for kinship estimation

Although some studies have shown that using more SNPs can increase the number of identified kinships (Mendes et al., 2022), we chose to prioritize the creation of an SNP dataset with minimal errors by applying filtering conditions, even if it meant a reduction in the number of SNPs. This is because when a large number of SNPs were used under loose filtering conditions, the estimation of kinships did not yield satisfactory results when the assumed error rate was high in COLONY 2.0. We finally found what were probably the best filtering conditions (MIN_DP = 5, MAX_DP = 50, MIN_MEAN_DP = 15, MIN_GQ ≥30, CR >0.9, MAF ≥0.03, HWE <0.00001, HET ≤0.7, LD ≤0.2) that accurately reproduced known kinships in both COLONY 2.0 and Sequoia. The subtle adjustments of HET or HWE values, which resulted in successful kinship estimation using both COLONY 2.0 and Sequoia, likely played a crucial role in the success of our study; we used this combination to estimate unknown kinship relationships. We are confident that we have identified the conditions required to generate a set of SNPs that can produce relatively reliable results for kinship analysis in the target population.

## 4.4 | Analysis of relationships by COLONY 2.0 and Sequoia

We conducted kinship analysis using COLONY 2.0 and Sequoia and found that Sequoia produced lower reliability of kinships for all types of relationships compared to COLONY 2.0. This is possibly because COLONY 2.0 estimates kinships by calculating the likelihood for the entire set of individuals (J. Wang, 2004, 2012), whereas Sequoia identifies the most likely kinship based on a threshold likelihood ratio between the first and second most likely relationships for a given pair (Huisman, 2017), which is less powerful than the COLONY 2.0 calculation. Moreover, the target population in this study was considered to be highly inbred, making COLONY 2.0 more suitable for this study as it can account for inbreeding. Nevertheless, almost all the pairs estimated as FS or HS by Sequoia were similarly estimated by COLONY 2.0, whereas four out of eight UANN pairs were estimated only by Sequoia. This can be attributed to COLONY 2.0 originally lacking a function to estimate UANN relationships.

We identified 10 pairs of FS or HS among the 15 $F_0$ individuals, indicating that the natural population was

highly inbred, making it challenging to determine the detailed kinships of a reintroduced population. Despite these difficulties, we were able to identify almost all of the $F_2$ families by combining the estimated GG, UANN, and FS relationships between $F_2$ individuals. However, despite sampling all possible grandparents for the $F_2$ individuals analyzed, GG relationships were not estimated for all $F_2$ individuals using COLONY 2.0. When multiple individuals in $F_2$ are in an FS relationship, COLONY 2.0 would not estimate the relationship between them and $F_0$ as HS "incorrectly" (J. Wang, 2007). In this study, the abundance of FS and HS relationships in $F_2$ may have led to a reduced frequency of HS relationships (interpreted as GG relationships) between $F_0$ and $F_2$. Moreover, COLONY 2.0 assumes hypothetical parents when estimating kinship (Steyaert et al., 2012), resulting in inconsistencies when attempting to estimate individuals that originally have a GG relationship by regarding them as HS.

### 4.5 | Analysis of relationships by gRandma

By utilizing kinship estimates from COLONY 2.0 or Sequoia, including GG, UANN, and FS relationships, we were able to identify families for 36 out of 47 $F_2$ individuals. However, the families for the remaining 11 $F_2$ individuals were still unknown. In contrast, gRandma, a specialized software for estimating GG trio relationships, was able to estimate grandparents for almost all $F_2$ individuals. gRandma failed to infer any GG relationship for only two of the $F_2$ individuals, but Sequoia estimated that they had a UANN relationship with an $F_1$ individual (mA1-4) of the A1 family, which supports the hypothesis that these $F_2$ individuals were produced by breeding between $F_1$ individuals of the A1 family.

The algorithm used by gRandma is unique because it considers GG trios instead of just GG pairs. Originally developed to estimate GG relationships between wild-caught grandchildren and hatchery grandparents, gRandma was designed to detect crossbreeding when adult sampling in the wild is limited (Delomas & Campbell, 2021). As a result, it was initially intended for cases where only one set of grandparents, paternal or maternal, was sampled. However, in this study, where both paternal and maternal grandparents were sampled, gRandma was able to estimate both sets of grandparents (T. A. Delomas, pers. comm.). However, because Delomas and Campbell (2021) initially assumed a low probability of false positives from trios other than GG, it is possible for a false-positive trio to yield a high LLR when, for example, the potential grandparents include siblings (Delomas, pers. comm.). As the COLONY 2.0 analysis revealed that the target population in this study contained many siblings in the grandparents' generation, we incorporated a MI threshold in addition to the LLR value. This is because a trio with a high LLR and low MI is more likely to be reliable, as a true GG trio would have both a high LLR and low MI, whereas a false-positive trio would have a high LLR and high MI. Moreover, we have information on the relationships between grandparents, such as if they are FS and HS, and which individuals were possibly interbred, so we were able to identify whether the inferred trios were false positives.

## 5 | CONCLUSIONS

In this investigation, we conducted a whole-genome SNP analysis to investigate detailed kinships and identify potential biases among families. Our findings indicate that family bias might be associated with the age and timing of release. Additionally, the high proportion of siblings among the $F_0$ cohort highlights the severely depleted genetic diversity of the natural population. We have also made novel contributions to the development of methods for characterizing detailed kinships in populations with extremely low genetic diversity, and we anticipate that our approach will prove beneficial for kinship estimation and genealogical analysis for other endangered species.

### ACKNOWLEDGMENTS

We thank the Shitara Dam Construction Office of the Chubu Regional Development Bureau, Ministry of Land, Infrastructure, Transport and Tourism, Japan, for providing the data on the captive breeding, reintroduction, and reproduction program of *T. ichikawai*. This study was conducted as part of the endangered freshwater fishes conservation project by the Shitara Dam Construction Office. We thank IDEA Consultants, Inc.; and Aichi Fisheries Research Institute, for sample collection. We are also grateful to Katsutoshi Watanabe, Yo Y. Yamasaki, Keisuke Onuki, Ryoya Yanada, Masato Kinoshita, Jinliang Wang, Thomas A. Delomas, and Jisca Huisman for their helpful comments.

### CONFLICT OF INTEREST STATEMENT

This research was conducted in part with a research fund from the Water Resources Environment Center, which the fifth author belongs to.

### ORCID

*Hinano Mizuno* https://orcid.org/0009-0007-3826-4561
*Kouji Nakayama* https://orcid.org/0000-0002-9996-8702

## REFERENCES


Abadía-Cardoso, A., Anderson, E. C., Pearse, D. E., & Carlos Garza, J. (2013). Large-scale parentage analysis reveals reproductive patterns and heritability of spawn timing in a hatchery population of steelhead (*Oncorhynchus mykiss*). *Molecular Ecology*, *22*, 4733–4746. https://doi.org/10.1111/mec.12426

Barnes, M. R., & Breen, G. (2010). *Genetic variation: Methods and protocols*. Springer.

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

Broad Institute. (2021). *Picard Tools*. Version 2.25.1 GitHub Repository. http://broadinstitute.github.io/picard/

Charlesworth, D., & Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics*, *18*, 237–268. https://doi.org/10.1146/annurev.ecolsys.18.1.237

Chen, B., Cole, J. W., & Grond-Ginsbach, C. (2017). Departure from Hardy Weinberg equilibrium and genotyping error. *Frontiers in Genetics*, *8*, 167. https://doi.org/10.3389/fgene.2017.00167

Chikhi, R., & Medvedev, P. (2014). Informed and automated k-mer size selection for genome assembly. *Bioinformatics*, *30*, 31–37. https://doi.org/10.1093/bioinformatics/btt310

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, *6*, 80–92. https://doi.org/10.4161/fly.19695

Collevatti, R. G., Leite, K. C., de Miranda, G. H., & Rodrigues, F. H. (2007). Evidence of high inbreeding in a population of the endangered giant anteater, *Myrmecophaga tridactyla* (Myrmecophagidae), from Emas National Park, Brazil. *Genetics and Molecular Biology*, *30*, 112–120. https://doi.org/10.1590/S1415-47572007000100020

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R., Lunter, G., Marth, G., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Delomas, T. A., & Campbell, M. (2021). Grandparent inference from genetic data: The potential for parentage-based tagging programs to identify offspring of hatchery strays. *North American Journal of Fisheries Management*, *42*, 85–95. https://doi.org/10.1002/nafm.10714

Dou, J., Sun, B., Sim, X., Hughes, J. D., Reilly, D. F., Tai, E. S., Liu, J., & Wang, C. (2017). Estimation of kinship coefficient in structured and admixed populations using sparse sequencing data. *PLoS Genetics*, *13*, e1007021. https://doi.org/10.1371/journal.pgen.1007021

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 9451–9457. https://doi.org/10.1073/pnas.1921046117

Frankham, R., Ballou, J. D., & Briscoe, D. A. (2002). *Introduction to conservation genetics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511808999

Fraser, D. J. (2008). How well can captive breeding programs conserve biodiversity? A review of salmonids. *Evolutionary Applications*, *1*, 535–586. https://doi.org/10.1111/j.1752-4571.2008.00036.x

Gaio, D., Anantanawat, K., To, J., Liu, M., Monahan, L., & Darling, A. E. (2022). Hackflex: Low cost Illumina Nextera Flex sequencing library construction. *Microbial Genomics*, *8*, 000744. https://doi.org/10.1099/mgen.0.000744

Gong, G., Dan, C., Xiao, S., Guo, W., Huang, P., Xiong, Y., Wu, J., He, Y., Zhang, J., Li, X., Chen, N., Gui, J. F., & Mei, J. (2018). Chromosomal-level assembly of yellow catfish genome using third-generation DNA sequencing and Hi-C analysis. *GigaScience*, *7*, giy120. https://doi.org/10.1093/gigascience/giy120

Huisman, J. (2017). Pedigree reconstruction from SNP data: Parentage assignment, sibship clustering and beyond. *Molecular Ecology Resources*, *17*, 1009–1024. https://doi.org/10.1111/1755-0998.12665

Jamieson, I. G. (2010). Founder effects, inbreeding, and loss of genetic diversity in four avian reintroduction programs. *Conservation Biology*, *25*, 115–123. https://doi.org/10.1111/j.1523-1739.2010.01574.x

Jansen, H. J., Liem, M., Jong-Raadsen, S. A., Dufour, S., Weltzien, F. A., Swinkels, W., Koelewijn, A., Palstra, A. P., Pelster, B., Spaink, H. P., van den Thillart, G. E., Dirks, R. P., & Henkel, C. V. (2017). Rapid de novo assembly of the European eel genome from nanopore sequencing reads. *Scientific Reports*, *7*, 7213. https://doi.org/10.1038/s41598-017-07650-6

Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., DePamphilis, C. W., Yi, T. S., & Li, D. Z. (2020). GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology*, *21*, 241. https://doi.org/10.1186/s13059-020-02154-5

Jones, O. R., & Wang, J. (2010). COLONY: A program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, *10*, 551–555. https://doi.org/10.1111/j.1755-0998.2009.02787.x

Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, *17*, 230–241. https://doi.org/10.1016/S0169-5347(02)02489-8

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Marie, A. D., Stockwell, B. L., & Rico, C. (2019). DNA analysis of juvenile scalloped hammerhead sharks *Sphyrna lewini* (Griffith, 1834) reveals multiple breeding populations and signs of adaptive divergence in the South Pacific. *Frontiers in Marine Science*, *6*, 718. https://doi.org/10.3389/fmars.2019.00718

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110

Melville, J., Haines, M. L., Boysen, K., Hodkinson, L., Kilian, A., Smith Date, K. L., Potvin, D. A., & Parris, K. M. (2017). Identifying hybridization and admixture using SNPs: Application of the DArTseq platform in phylogeographic research on vertebrates. *Royal Society Open Science*, *4*, 161061. https://doi.org/10.1098/rsos.161061

Mendes, B., Sampaio, T., Antunes, M. A., Magalhães, H., Costa e Silva, F., Borges, C., Simões, F., Usié, A., Almeida, M. H., & Ramos, A. M. (2022). Kinship analysis and pedigree reconstruction of a natural regenerated cork oak (*Quercus suber*) population. *Forests*, *13*, 226. https://doi.org/10.3390/f13020226

Mie Prefecture. (2005). *Guidelines for the protection and management of the Natural Monument Nekogigi* (p. 54). Mie Prefectural Board of Education. http://www.bunka.pref.mie.lg.jp/bunkazai/papers/pdfs/nekoconguidv5-3s.pdf

Mie Prefecture. (2006). *Report on the project for the protection and propagation of Nekogigi, a natural monument of the Katsuben River system* (pp. 16–17). Mie Prefectural Board of Education. https://www.bunka.pref.mie.lg.jp/common/content/000146720.pdf

Ministry of the Environment. Ministry of the Environment Red List. (2020). *Office for conservation of rare species*. Wildlife Division, Nature Conservation Bureau, Ministry of the Environment.

Miyagawa, T., Nishida, N., Ohashi, J., Kimura, R., Fujimoto, A., Kawashima, M., Koike, A., Sasaki, T., Tanii, H., Otowa, T., Momose, Y., Nakahara, Y., Gotoh, J., Okazaki, Y., Tsuji, S., & Tokunaga, K. (2008). Appropriate data cleaning methods for genome-wide association study. *Journal of Human Genetics*, *53*, 886–893. https://doi.org/10.1007/s10038-008-0322-y

Nakamura, M. (1963). *Search for freshwater fishes in Japan in primary colors* (p. 258). Hokuryukan.

Nishida, N., Koike, A., Tajima, A., Ogasawara, Y., Ishibashi, Y., Uehara, Y., Inoue, I., & Tokunaga, K. (2008). Evaluating the performance of Affymetrix SNP Array 6.0 platform with 400 Japanese individuals. *BMC Genomics*, *9*, 431. https://doi.org/10.1186/1471-2164-9-431

Niwa, Y. (1967). *Fish of the Kiso River, river fish ecology, middle and lower streams* (p. 293). Gifu.

O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., & Portnoy, D. S. (2018). These aren't the loci you're looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular Ecology*, *27*, 3193–3206. https://doi.org/10.1111/mec.14792

Philippart, J. C. (1995). Is captive breeding an effective solution for the preservation of endemic species? *Biological Conservation*, *72*, 281–295. https://doi.org/10.1016/0006-3207(94)00090-D

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*, 559–575. https://doi.org/10.1086/519795

Roshyara, N. R., Kirsten, H., Horn, K., Ahnert, P., & Scholz, M. (2014). Impact of pre-imputation SNP-filtering on genotype imputation results. *BMC Genetics*, *15*, 88. https://doi.org/10.1186/s12863-014-0088-5

Ryman, N., & Laikre, L. (1991). Effect of supportive breeding on the genetically effective population size. *Conservation Biology*, *5*, 325–329.

Sedlazeck, F. J., Rescheneder, P., & von Haeseler, A. (2013). NextGenMap: Fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics*, *29*, 2790–2791. https://doi.org/10.1093/bioinformatics/btt468

Sekino, M., Saito, K., Yamada, T., Kumagai, A., Hara, M., & Yamashita, Y. (2003). Microsatellite-based pedigree tracing in a Japanese flounder *Paralichthys olivaceus* hatchery strain: Implications for hatchery management related to stock enhancement program. *Aquaculture*, *221*, 255–263. https://doi.org/10.1016/S0044-8486(02)00667-1

Shitara Dam Construction Office. (2023). *Chubu Regional Development Bureau*. Ministry of Land, Infrastructure, Transport and Tourism. https://www.cbr.mlit.go.jp/shitara/05kankyo_hozen/nekogigi/02jikken.html

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Snedecor, J., Fennell, T., Stadick, S., Homer, N., Antunes, J., Stephens, K., & Holt, C. (2022). Fast and accurate kinship estimation using sparse SNPs in relatively large database searches. *Forensic Science International: Genetics*, *61*, 102769. https://doi.org/10.1016/j.fsigen.2022.102769

Steyaert, S. M. J. G., Endrestol, A., Hacklander, K., Swenson, J. E., & Zedrosser, A. (2012). The mating system of the brown bear Ursus arctos. *Mammal Review*, *42*, 12–34. https://doi.org/10.1111/j.1365-2907.2011.00184.x

Thompson, E. A. (1975). The estimation of pairwise relationships. *Annals of Human Genetics*, *39*, 173–188. https://doi.org/10.1111/j.1469-1809.1975.tb00120.x

Thow, C. M., Eadie, J. M., Wells, C. P., & Lyon, B. E. (2022). Pedigree simulations reveal that maternity assignment is reliable in populations with conspecific brood parasitism, incomplete parental sampling and kin structure. *Molecular Ecology Resources*, *22*, 180–198. https://doi.org/10.1111/1755-0998.13466

Wajiki, Y., Kaneko, Y., Sugiyama, T., Yamada, T., & Iwaisaki, H. (2015). Genetic analyses in the Japanese captive population of Japanese crested ibis (*Nipponia nippon*) using pedigree information. *Poultry Science*, *53*, 1–7. https://doi.org/10.2141/jpsa.0150040

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, *9*, e112963. https://doi.org/10.1371/journal.pone.0112963

Wang, J. (2004). Sibship reconstruction from genetic data with typing errors. *Genetics*, *166*, 1963–1979. https://doi.org/10.1534/genetics.166.4.1963

Wang, J. (2007). Parentage and sibship exclusions: Higher statistical power with more family members. *Heredity*, *99*, 205–217. https://doi.org/10.1038/sj.hdy.6800984

Wang, J. (2012). Computationally efficient sibship and parentage assignment from multilocus marker data. *Genetics*, *191*, 183–194. https://doi.org/10.1534/genetics.111.138149

Wang, O., Chin, R., Cheng, X., Wu, M. K. Y., Mao, Q., Tang, J., Sun, Y., Anderson, E., Lam, H. K., Chen, D., Zhou, Y., Wang, L., Fan, F., Zou, Y., Xie, Y., Zhang, R. Y., Drmanac, S., Nguyen, D., Xu, C., … Peters, B. A. (2019).

Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome Research*, *5*, 798–808. https://doi.org/10.1101/gr.245126.118

Watanabe, K. (1994a). Growth, maturity and population structure of the bagrid catfish, *Pseudobagrus ichikawai*, in the Tagiri River, Mie prefecture, Japan. *Japanese Journal of Ichthyology*, *41*, 15–22.

Watanabe, K. (1994b). Mating behavior and larval development of *Pseudobagrus ichikawai* (Siluriformes: Bagridae). *Japanese Journal of Ichthyology*, *41*, 243–251.

Watanabe, K. (1995). *Natural history of Pseudobagrus ichikawai* (Doctoral thesis). Graduate School of Tokyo University of Fisheries.

Watanabe, K. (1998). Nekogigi. In Fisheries Agency (Ed.), *Data book on rare wild aquatic organisms in Japan* (pp. 154–155). Japan Fisheries Resources Conservation Society.

Watanabe, K. (2008). Diel activity and reproductive territory of the Japanese bagrid catfish, *Pseudobagrus ichikawai*. *Environmental Biology of Fishes*, *81*, 77–86. https://doi.org/10.1007/s10641-006-9173-6

Watanabe, K., & Nishida, M. (2003). Genetic population structure of Japanese bagrid catfishes. *Ichthyological Research*, *50*, 140–148. https://doi.org/10.1007/s10228-002-0149-z

Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., & Jaffe, D. B. (2017). Direct determination of diploid genome sequences. *Genome Research*, *27*, 757–767. https://doi.org/10.1101/gr.214874.116

Willoughby, J. R., Ivy, J. A., Lacy, R. C., Doyle, J. M., & DeWoody, J. A. (2017). Inbreeding and selection shape genomic diversity in captive populations: Implications for the conservation of endangered species. *PLoS One*, *12*, e0175996. https://doi.org/10.1371/journal.pone.0175996

Wu, C., Zhang, D., Kan, M., Lv, Z., Zhu, A., Su, Y., Zhou, D., Zhang, J., Zhang, Z., Xu, M., Jiang, L., Guo, B., Wang, T., Chi, C., Mao, Y., Zhou, J., Yu, X., Wang, H., Weng, X., … Liu, Y. (2014). The draft genome of the large yellow croaker reveals well-developed innate immunity. *Nature Communications*, *5*, 5227. https://doi.org/10.1038/ncomms6227

## SUPPORTING INFORMATION

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

**How to cite this article:** Mizuno, H., Nakayama, K., Akita, T., Hashiguchi, Y., Osugi, T., & Takeshima, H. (2024). Detailed kinship estimation for detecting bias among breeding families in a reintroduced population of the endangered bagrid catfish *Tachysurus ichikawai*. *Population Ecology*, 1–16. https://doi.org/10.1002/1438-390X.12183