The Phylogenetic Position of Zebrafish (*Danio Rerio*) from South African Pet Shops

E. Blom
University of the Free State - Bloemfontein Campus: University of the Free State
Willem G. Coetzer (coetzerwg@outlook.com)
University of the Free State https://orcid.org/0000-0003-2189-5539
S-R. Schneider
University of the Free State - Bloemfontein Campus: University of the Free State
J.P. Grobler
University of the Free State - Bloemfontein Campus: University of the Free State

Research Article

Keywords: zebrafish, cytochrome b, microsatellites, genetic diversity, phylogenetic origin

Posted Date: January 17th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1227383/v1

Abstract

Zebrafish (*Danio rerio*), a small freshwater fish that originates from India, Bangladesh, Nepal, Bhutan and northern Myanmar, have been widely used as a model organism for studies of developmental biology and genetics. The current study aimed to determine the origin of South African pet shop stock that are currently being used to establish a laboratory population founded from diverse sources available locally. Zebrafish DNA was extracted from 65 specimens housed at the University of the Free State (UFS) Department of Genetics. For phylogenetic analysis, *cytb* sequences were generated from all samples. A further 178 sequences were downloaded from the GenBank database, including sequences of an outgroup species (*Danio kyathit*). Five microsatellite markers were used to further assess the genetic diversity of the UFS zebrafish specimens. A maximum likelihood analysis was performed for the *cytb* data. Results of the phylogenetic analyses divided the sequences into three major genetic groups, which was congruent with a previous study on laboratory zebrafish provenance. The SA pet shop fish grouped with the strains from the northern and north-eastern regions of India. High levels of microsatellite genetic diversity were observed for the pet shop sourced population, correlating to what has previously been observed in zebrafish. These results can be used to guide the future development of laboratory strains suited to the needs at the UFS.

Introduction

The zebrafish (*Danio rerio*) is a widely used model organism in biomedical research, developmental genetics, and neurophysiology (Lieschke and Currie 2007; Spence *et al.* 2008) and increasingly also in environmental studies (Paull *et al.* 2008; Scholz *et al.* 2008). Zebrafish have several qualities that make them suitable for manipulation and use in research experiments. They are small, robust fish that can be kept in large numbers. Females spawn every 2-3 days, and a single batch may contain several hundred eggs. The generation time is relatively short at 3-4 months, making these fish suitable for genetic selection experiments (Spence *et al.* 2008).

The natural distribution range of the zebrafish is centred around the Ganges and Brahmaputra River basins in Bangladesh, Nepal, and north-eastern India. However, specimens have also been collected in the Indus, Pennar, Mahanadi, Cauvery, and Godavari River basins (Fig. 1) (Spence *et al.* 2008; Whiteley *et al.* 2011). Zebrafish have been described to typically inhabit slow-moving or standing water bodies, and the edges of streams and ditches especially adjacent to rice-fields (Sterba 1962; Talwar and Jhingran 1991; Jayaram 1999).

Over 400 laboratories worldwide use zebrafish in fundamental and applied research (The Zebrafish International Resource Center, www.zebrafish.org). Most laboratory strains used are the product of many generations bred in captivity. One of the most well-known zebrafish laboratory strains is the AB line, which was established from unknown zebrafish source stocks bought from two different pet shops in Albany, Oregon (USA) in the early 1970s (Staff 2016). Haploid offspring from AB females were crossed with random males for about 70 generations up until the early 1990s. At that time, six diploid progeny

stock populations (each originating from a distinct haploid female) were crossed to produce the current AB strain. The current AB strain is being maintained through large group spawning crosses (Holden and Brown 2018).

While long established laboratory lines are suitable for many lines of research, such a history could also have resulted in unplanned selection for certain gene variants that are favoured in the laboratory environment (Robison and Rowland 2005; Wright *et al.* 2006). Specifically, such selection could then render laboratory stock unsuitable for studies involving the interaction between environmental stressors and responses coded by diverse alleles. For this reason, a new line founded from diverse sources is currently being established in our laboratory at the University of the Free State, South Africa.

Genetic diversity within wild and captive bred zebrafish populations have previously been studied by Gratton *et al.* (2004), Coe *et al.* (2009), Whiteley *et al.* (2011) and Balik-Meisner *et al.* (2018). According to Whiteley *et al.* (2011), there exists a high level of genetic diversity within wild zebrafish populations. In contrast, genetic variation in laboratory strains has been shown to be significantly lower than that of wild populations (Whiteley *et al.* 2011). This difference is most likely due to selective breeding for specific traits in laboratory populations as well as genetic drift.

Despite the growth in genomics, two established classes of molecular markers – mitochondrial sequences and microsatellite fragments – can also be applied as cost-effective methods to study genetic diversity and differentiation in zebrafish. Mitochondrial DNA (mtDNA) sequences, such as sequences of the cytochrome *b* (*cytb*) region, is one of the most extensively studied regions used in vertebrates (Irwin *et al.* 1991; Lydeard and Roe 1997; Moore and Defilippis 1997). The *cytb* gene evolves relatively slow and encodes a protein, which is a well characterized molecular system (Esposti *et al.* 1993). An application of whole-mitochondria work on zebrafish was reported by Broughton *et al.* (2001), who studied the entire mitochondrial genome of zebrafish to determine the evolutionary patterns for extrapolation to other vertebrate mtDNA.

Microsatellites remain a useful marker in population genetic studies, due to the high mutation rate of these markers. These markers also have the advantage of being easily detected by polymerase chain reaction (PCR) at a low-cost. Large databases with published loci exist, such as GenBank (www.ncbi.nlm.nih.gov/genbank/) (Clark *et al.* 2016), EMBL (www.ebi.ac.uk/embl) (Kanz *et al.* 2005), and ZFIN (Ruzicka *et al.* 2019), with the latter database specific to zebrafish. Primers developed for these loci can be cross amplified between related species. Rico *et al.* (1996) studied the transferability of microsatellite loci between fish species whose last common ancestor lived 470 million years before present (Ma BP). These authors found that primer pairs designed from microsatellite flanking regions, amplify homologous sequences from these fish. Microsatellite markers can detect both homo- and heterozygous genotypes (Hoshino *et al.* 2012). This characteristic makes it an important marker to help determine the genetic diversity within and between populations.

Here, we report on the genetic characterization of the zebrafish population being established at the University of the Free State (UFS) to serve as stock for future research in population genetics and

response to environmental stressors. We collected zebrafish from several pet shops and ornamental fish suppliers in the Bloemfontein and Johannesburg areas, South Africa. Our objectives were: (i) to determine the possible geographic origin of the South African (SA) pet shop zebrafish stock that is used in the research laboratories of the Department of Genetics, (UFS, South Africa), targeting a segment of the *cytb* gene; and (ii) to determine the level of genetic diversity in zebrafish bought from different sources using the *cytb* data, as well as five microsatellite markers.

Materials And Methods

Ethical approval

The housing of all animals located at the Department of Genetics, (UFS, South Africa) as well as the experimental procedures of this study adheres to the guidelines approved by the Interfaculty Animal Ethics Committee at the University of the Free State (Ethical approval number: UFS-AED2018/0037). Section 20 veterinary authorization was obtained from the South African Department of Agriculture, Forestry, and Fisheries (DAFF).

Sampling

A total of 65 zebrafish were selected from several sources, with localities coded for anonymity. Forty-six specimens were obtained from supplier “1” and supplier “2” in Bloemfontein, Free State Province, South Africa. These specimens were kept in the same quarantine tank and formed sample group “A” (ZFA). Another eight specimens were acquired from supplier “3” in the Bloemfontein area (ZFB), and ten fish were obtained from a large-scale ornamental fish supplier based in Johannesburg, Gauteng Province (supplier “4”) (ZFC).

Samples for DNA extraction were obtained from both living fish and laboratory fish that died from natural causes. Live fish were sampled by swabbing as described by Le Vin *et al*. (2011). Samples from dead fish (stored at -20°C) were taken by means of tail cuttings of roughly 4mm x 2mm.

For comparative purposes, a further 178 *cytb* sequences from Whiteley *et al.* (2011) were downloaded from the GenBank database (Accession numbers, JN234180–JN234356), including the *cytb* sequence of an outgroup species (*Danio kyathit*; Accession number, EF452733). See Fig. 1 for sampling localities from the Whiteley *et al.* (2011) data. Abbreviations for the sampled areas are indicated in Supplementary Table S1.

DNA extraction and PCR amplification

DNA was extracted from the zebrafish samples using the Roche© High-Pure PCR Template Preparation Kit (Roche Diagnostics, Indianapolis, IN, USA), following the manufacturer’s protocol. An assessment of DNA

quality and quantity was performed on a NanoDrop® Spectrophotometer ND-1000 (ThermoFisher Scientific, Waltham, MA, USA). All DNA extracts were subsequently stored at -20˚C.

A 1 122 bp region of the *cytb* gene was amplified, using the primers utilised by Whiteley *et al.* (2011) (Whiteley *et al.* 2011). For the forward primer we used fishcytbzf-F from Fang *et al.* (2009), with HA-danio from Mayden *et al.* (2007) used as reverse primer (Table 1). PCR reaction mixes (12.5 µl volume) were composed of 6.25 µl Ampliqon TEMPase Hot Start 2X Master Mix (Odense M, Denmark), 3.5 µl dH_2O, 0.375 µl of each primer (10µM stock), and 2 µl DNA. The PCR reaction conditions were as follows: 95°C for 15 min, 35 cycles of 95°C for 30 s, 56.5°C for 40 s, 72°C for 1 min, 72°C for an additional 5 min and 12°C until manually terminated. All PCR products were sequenced on an ABI 3500 Genetic Analyser at the Department of Genetics, UFS. All sequences generated from this study were deposited in GenBank (Accession numbers: MK893921-MK893984).

Five highly variable microsatellite loci (Ztri1, Z249, Z6104, Z9230, and Z20450), previously utilised by Coe *et al.* (2009), were used to determine the genetic diversity levels for the zebrafish from different domestic sources. All primer sequences are presented in Table 1. The markers were divided into two multiplex sets (multiplex 1: Ztri1 and Z9230; multiplex 2: Z249, Z6104, and Z20450). The forward primer in each microsatellite pair was fluorescently labelled at the 5’ end. Ampliqon TEMPase Hot Start 2X Master Mix was used for all amplifications (Odense M, Denmark). The PCR reactions for multiplex 1 (12.5 µl) consisted of 4.5 µl dH_2O, 0.375 µl of each Ztri1 primer (10 µM stock), 0.250 µl of each Z6104 primer (10 µM stock), 6.25 µl Ampliqon TEMPase Hot Start 2X Master Mix, and 0.5 µl of template DNA. The PCR reaction for multiplex 2 consisted of 4.25 µl dH_2O, 6.25 µl Ampliqon TEMPase Hot Start 2X Master Mix, 0.25 µl of each primer (10 µM stock), and 0.5 µl DNA to give a final reaction volume of 12.5 µl. Samples that did not amplify at all markers when using multiplex, were then amplified separately, with the PCR reaction (11 µl) composed of 3.5 µl dH2O, 6.25 µl Ampliqon TEMPase Hot Start 2X Master Mix, 0.375 µl for each primer (10 µM stock), and 0.5 µl DNA.

A touchdown PCR was performed for all microsatellite amplifications. The PCR profile consisted of 95°C for 3 min, 8 cycles of 95°C for 30 s, 64°C for 40 s, -1°C per cycle, 72°C for 1 min. Twenty-five cycles of 95°C for 30 s, 56°C for 40 s, 72°C for 1 min, and a final extension step at 72°C for 5 min.

Statistical analysis

All DNA sequences were assembled and aligned in GENEIOUS v4.7.4 (Drummond *et al.* 2009) using the ClustalW option (Thompson *et al.* 1994). DnaSP software (Rozas *et al.* 2003) was used to calculate the number of haplotypes (h), haplotype diversity (HD) and nucleotide diversity (π) .

Nucleotide diversity (π) results were rounded to four decimal places due to the small magnitude of values obtained. Pairwise population PhiPT values among groups were estimated in GenAlEx v6.5 (Peakall and Smouse 2012).

A Maximum likelihood (ML) analysis of the identified haplotypes was performed using the online PhyML platform (Guindon *et al.* 2010) (http://www.atgc-montpellier.fr/phyml/) to assess the relationship between our sequences and previously published data. Automatic model selection by Smart Model Selection (SMS) (Lefort *et al.* 2017) was selected, and branch support was estimated by performing 1 000 bootstrap iterations. The closely related *Danio kyathit* (Mayden *et al.* 2007) was used as the outgroup (Accession number, EF452733).

The genetic variation estimated from the microsatellite loci were quantified in terms of observed and expected heterozygosity, number of alleles observed, polymorphic information content (PIC), allelic richness, conformation of expected numbers of genotypes to expectations under Hardy-Weinberg Test equilibrium (HWE), the inbreeding coefficient, null alleles, and presence of linkage disequilibrium. The expectation maximization (EM) algorithm for detection of null allele frequencies was used as implemented in the software program FreeNA (Chapuis and Estoup 2007). A paired t test was performed to determine if null alleles has a significant effect on F_{ST} estimates by comparing null allele corrected and uncorrected F_{ST} values calculated by the excluding null allele (ENA) method (Chapuis and Estoup 2007). The test for HWE and linkage disequilibrium was done using GENEPOP (Raymond and Rousset 1995). Number of alleles and observed, expected heterozygosity, and pairwise F_{ST} and associated *p*-values, with a significance level of 0.05, were calculated using GenAlEx v6.5 (Peakall and Smouse 2012). Polymorphic information content was determined using Cervus (Kalinowski *et al.* 2007). Allelic richness and inbreeding coefficient were calculated using FSTAT 1.2 (Goudet 1995).

Results

Results from cytb sequences

A 1 122 bp region of the *cytb* gene was successfully sequenced for 65 individuals. A total of seven unique haplotypes were identified for the SA pet shop populations. Seventy haplotypes defined by 176 segregating sites were identified for the combined dataset (Supplementary Table S2), including data from Whiteley *et al.* (2011). There were no gaps in the alignment. The number of haplotypes per population ranged from 1 to 10 (Table 2), with ZFA having the highest number from the SA pet shop populations. The haplotype diversity (*h*) for the SA pet shop populations ranged from 0.429 to 0.733. Nucleotide diversity for the SA pet shop populations ranged from 0.0030 to 0.0049.

The pairwise population PhiPT values (Supplementary Table S3) showed that the fish in the ZFA group differs significantly from ZFB and ZFC, but with no significant differentiation between ZFB and ZFC. No significant differences were observed between laboratory strain SJA, studied by Whiteley *et al.* (2011), and either of the South African sourced fish groups (*p*-value = 0.097 to 0.374). All three SA pet shop populations (ZFA, ZFB, and ZFC) were genetically most similar to populations from north-eastern India (UTR). Similarity was also observed between the SA pet shop fish and fish from southern India (WYD), although these results should be considered with caution due to the small samples size of the WYD group (n = 2). Furthermore, ZFA and ZFC showed genetic similarity to SRN from southern India, but this

group was also presented by a small sample size (n = 3). Populations ZFB and ZFC did not differ significantly from SHK and PGM sampled from northern India.

The topology of the ML tree obtained from the current study (Fig 2.), closely resembles the tree resolved by Whiteley *et al.* (2011). The phylogenetic analyses of the 70 haplotypes assessed revealed three major genetic clades. Membership of clades was as follows: Clade 1, the laboratory strains, SA pet shop fish, Northern India, and western and eastern Nepal; Clade 2, Bangladesh and southern India; and Clade 3, Central Nepal (Fig. 2). These three clades were well supported, as indicated by the bootstrap values.

Results from microsatellite analysis

A total of 65 zebrafish was successfully genotyped across five microsatellite loci to determine the genetic diversity of the SA pet shop populations. The per locus estimated null allele frequencies ranged from 0.000 to 0.1860. Null allele presence can inflate F_{ST} values (Carlsson 2008). However, no significant difference was observed between the ENA corrected (F_{ST} = 0.019) and the uncorrected ENA (F_{ST} = 0.018) values (p-value < 0.05). Analyses were therefore performed using data from all loci. All loci were observed to be informative, with PIC values ranging from 0.590 to 0.891. Numbers of genotypes at two loci in ZFA deviated significantly (p-value < 0.05) from expected HWE (Ztri1 and Z9230). A total of 29 alleles were detected in all loci across all three populations. The mean number of alleles per locus ranged from 4.333 (Z249) to 7.333 (Z9230). Allelic richness values ranged from 4.02 to 6.583 across all loci (Supplementary Table S4). All five loci showed significant levels of heterozygosity (Supplementary Table S4), with two loci showing negative F_{IS} values. No linkage disequilibrium was observed (Supplementary Table S4). All three UFS groups had high levels of heterozygosity, H_O ranged from 0.596 to 0.720 and H_E ranged from 0.674 to 0.743. ZFB was the only population to show a negative F_{IS} value, suggesting a excess of heterozygotes (Table 3). No genetic differentiation was seen between the ZFA and ZFB populations (F_{ST} = 0.000; p-value = 0.435). Low, but significant genetic differentiation was observed between the ZFA and ZFC populations (F_{ST} = 0.029; p-value = 0.013), and similarly low genetic differentiation was seen between populations ZFB and ZFC (F_{ST} = 0.053; p-value = 0.014).

Discussion

Origin of SA pet shop populations

Although founded independently from established laboratory strains, the UFS population shows a high level of shared ancestry with these strains following ML analysis. Approximately 60% of the SA pet shop fish grouped with haplotypes 1 and 2 similarly to the established laboratory strains. The phylogenetic grouping containing the pet shop fish and the laboratory strains also share ancestry with the northern and north-eastern wild populations. This observation is in line with known populations histories, with established laboratory strains also historically bred from pet shop fish (Howe *et al.* 2013) and it suggests

that the northern and north-eastern populations might be a favoured source for the collection of fish for the pet shop trade internationally.

The majority of recorded occurrences of wild zebrafish are in northern India, Nepal, and Bangladesh (Whiteley *et al.* 2011). This higher level of occurrence in these areas further supports the notion that the northern and north-eastern regions may form the main sources for the ornamental fish trade. Suurväli *et al*. (2019) found wild populations from West Bengal to be most closely related to the tested laboratory strains. Studying the West Bengal wild populations could potentially reveal a broader spectrum of genetic effects for specific mutations, than when only studying the inbred laboratory strains. The populations from Nepal (KHA) and Bangladesh (CHT) represents a distinct linage of zebrafish that diverged from the West Bengal populations before the laboratory and pet shop fish were established (Suurväli *et al.* 2019).

Overall, the ML-based phylogenetic results obtained were congruent with the results reported by Whiteley *et al*. (2011). This provides confidence in the phylogenetic analyses from the current study.

Genetic diversity

South African pet shop sourced fish displayed a high level of genetic diversity for the *cytb* gene, comparable to the wild populations. This high level of genetic diversity in the wild population was also seen by Suurväli *et al*. (2019) and Whiteley *et al*. (2011). The high level of diversity seen in the overall pet shop zebrafish population indicates that it is a good source population to establish a new laboratory strain at UFS. This high level of genetic diversity could be the result of different source populations used to supplement the current pet shop populations (Mäkinen *et al.* 2014). Other examples of genetic diversity analyses performed on commercially traded fish include studies on red and white koi carp (*Cyprinus carpio L*) (Shi *et al.* 2020), guppies (*Poecilia reticulata*) (Bleakley *et al.* 2008), and freshwater angelfish (*Pterophyllum scalare*) (Pandolfi *et al.* 2021). High levels of genetic diversity were reported in breeding populations of red and white koi carp (Shi *et al.* 2020). These koi breeding populations serve as a source for commercial trading. If similarly high levels of diversity (relative to wild fish) are seen in zebrafish breeding populations used for commercial trading, it would be a good indicator that pet shop fish can be used to establish a laboratory strain. Importing existing zebrafish laboratory strains can be expensive from a South African perspective and establishing a new, highly diverse, laboratory strain from pet shop fish will be more affordable. A genetically diverse strain can be advantageous during population studies, for example: to study the effects of bottlenecks on diversity at specific genes. The established laboratory strains studied by Whiteley *et al.* (2011) show very low levels of genetic diversity for the *cytb* gene and will thus not be as useful for population genetic studies. By selective breeding with healthy fish, it is possible to select against lethal mutations and establish a laboratory strain with restricted genetic diversity. This selective breeding is possibly the leading cause in the decline of genetic diversity seen in established laboratory strains. To prevent such a decline, new fish will have to continuously be introduced to counteract the effects of selective breeding.

The microsatellite data also showed comparatively high levels of genetic diversity in the South African groups. The H_O estimates observed for the three UFS groups (H_O range = 0.596 – 0.720) is in the same range as that observed by Coe *et al*. (2009) for the two commercial strains and the wild zebrafish population (H_O range = 0.525 – 0.714). In contrast, the H_O estimates for the majority of the lab strains studied by Coe *et al*. (2009) were below 0.500, except for a WIK group sourced from the University of Exeter in 2006. The genetic differentiation estimates showed little to no genetic difference between the three pet shop populations from the current study. On average the SA pet shop zebrafish had higher PIC values (average PIC = 0.67) than observed in red and white koi carp (average PIC = 0.557) (Shi *et al.* 2020) and angelfish (average PIC = 0.587) (Pandolfi *et al.* 2021). The Ar values from the current study, ranging from 5.851 to 4.600, differed only slightly from the values reported on zebrafish by Coe *et al*. (2009) ranging 5.478 to 1.967, with the wild population being an exception with an Ar value of 14.126. The positive F_{IS} value in the ZFA population is indicative of a deficit of heterozygotes, which indicates inbreeding is taking place. The F_{IS} values for the ZFB and ZFC populations did not deviate far from zero. The differences seen between the three SA pet shop populations could be due to the small population sizes of ZFB and ZFC (Reed and Frankham 2003). F_{IS} values obtained by Suurväli *et al.* (2019) for their laboratory strain, did not deviate far from zero. The inbreeding seen in the current study, as well as in the study performed by Suurvali *et al*. (2019), can be caused by reduced population sizes (Reed and Frankham 2003) when compared to wild populations. Another possibility is that perhaps only a few individuals, in the already reduced population, actively breed. The mean F_{IS} for the angelfish populations (Pandolfi *et al.* 2021) were close to zero, despite being wild populations. These authors speculated that a possible explanation for this observation is overfishing at the collection site. Similarly, the mean F_{IS} values for all guppy strains studied by Bleakley *et al*. (2008) were all positive but did not statistically deviate from zero (Bleakley *et al.* 2008). It is expected for designer guppies to show some level of inbreeding, since selective breeding is taking place.

Conclusion

The analyses offered good insight into the phylogenetic origin of SA pet shop fish and the high levels of diversity indicates a diverse founding population and suggests that there is still a steady inflow of new genetic material into the country. This will ensure that the gene pool remains diverse and will not become fixed for mutations that might be detrimental to the population. The SA pet shop zebrafish can therefore serve as a good source for the development of laboratory strains suited to the needs at the University of the Free State. Further studies on additional gene regions could be performed to strengthen the current results.

Declarations

Funding

Research funding was provided by the 50% Special Projects: Central Research Fund (CRF) program of the Faculty of Natural and Agricultural Sciences of the University of the Free State.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Acknowledgements

We want to thank all members of the UFS Zebrafish unit and the Department of Genetics who provided assistance during this project. Willem G. Coetzer, Sue-Rica Schneider and J. Paul Grobler contributed to the study conception and design. This paper forms part of the MSc studies of Elmarie Blom. Material preparation, data collection and analysis were performed by Elmarie Blom. The first draft of the manuscript was written by Elmarie Blom and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. Research funding was awarded to J. Paul Grobler through the 50% Special Projects: Central Research Fund (CRF) program of the Faculty of Natural and Agricultural Sciences of the University of the Free State.

References

1. Balik-Meisner, M., L. Truong, E.H. Scholl, R.L. Tanguay and D.M. Reif, 2018 Population genetic diversity in zebrafish lines. Mammalian Genome 29**:** 90–100.
2. Bleakley, B.H., A.C. Eklund and I. Brodie, E.D., 2008 Are Designer Guppies Inbred? Microsatellite Variation in Five Strains of Ornamental Guppies, Poecilia reticulata, Used for Behavioral Research. Zebrafish 5.
3. Broughton, R.E., J.E. Milam and B.A. Roe, 2001 The Complete Sequence of the Zebrafish (Danio rerio) Mitochondrial Genome and Evolutionary Patterns in Vertebrate Mitochondrial DNA. Cold Spring Harbor Laboratory Press**:** 1958-1967.
4. Carlsson, J., 2008 Effects of microsatellite null alleles on assignment testing. J Hered 99**:** 616-623.
5. Chapuis, M.-P., and Estoup, 2007 Microsatellite null alleles and estimation of population differentiation. Molecular Biology and Evolution 24**:** 621-631.
6. Clark, K., I. Karsch-Mizrachi, D. Lipman, J. Ostell and E. Sayers, 2016 GenBank. Nucleic Acids Res 44**:** D67–D72.
7. Coe, T.S., P.B. Hamilton, A.M. Griffiths, D.J. Hodgson, M.A. Wahab *et al.*, 2009 Genetic variation in strains of zebrafish (Danio rerio) and the implications for ecotoxicology studies. Ecotoxicology 18**:** 144–150.
8. Drummond, A.J., B. Ashton, M. Cheung, J. Heled, M. Kearse *et al.*, 2009 Geneious v4.7, pp.

9. Esposti, M.D., S. De Vries, M. Crimi, A. Ghelli, T. Patarnello *et al.*, 1993 Mitochondrial cytochrome b: evolution and structure of the protein. Biochem. Biophys. Acta 1143**:** 243–271.
10. Fang, F., M. Noren, T.Y. Liao, M. Kallersjo and S.O. Kullander, 2009 Molecular phylogenetic interrelationships of the south Asian cyprinid genera Danio, Devario, and Microrasbora (Teleostei, Cyrinidae, Danioninae). Zoologica Scripta 38**:** 237–256.
11. Goudet, J., 1995 FSTAT (version 1.2): a computer program to calculate F-statistics. J Hered 86**:** 485-486.
12. Gratton, P., G. Allegrucci, M. Gallozzi, C. Fortunato, F. Ferreri *et al.*, 2004 Allozyme and microsatellite genetic variation in natural samples of zebrafish, Danio rerio. J. Zool. Syst. Evol. Research 42**:** 54–62.
13. Guindon, S., J.F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk *et al.*, 2010 New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology 59**:** 307-321.
14. Holden, L.A., and K.H. Brown, 2018 Baseline mRNA expression differs widely between common laboratory strains of zebrafish. Scientific Reports DOI:10.1038/s41598-018-23129-4**:** 8:4780.
15. Hoshino, A.A., J.P. Bravo, P.M. Nobile and K.A. Morelli, 2012 Microsatellites as Tools for Genetic Diversity Analysis, pp. ISBN: 978-953-951-0064-0065 in *Genetic Diversity in Microorganisms*, edited by P.M. Caliskan. InTech.
16. Howe, D.G., Y.M. Bradford, T. Conlin, A.E. Eagle, D. Fashena *et al.*, 2013 ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics, pp.
17. Irwin, D.M., T.D. Kocher and A.C. Wilson, 1991 Evolution of the cytochrome b gene of mammals. J. Mol. Evol. 32**:** 128–144.
18. Jayaram, K.C., 1999 *The freshwater fishes of the Indian region*. Narendra Publishing House, Delhi.
19. Kalinowski, S., M. Taper and T. Marshall, 2007 Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology 22**:** 1099-1106.
20. Kanz, C., P. Aldebert, N. Althorpe, W. Baker, A. Baldwin *et al.*, 2005 The EMBL Nucleotide Sequence Database. Nucleic Acids Res 33**:** D29–D33.
21. Le Vin, A.L., A. Adam, A. Tedder, K.E. Arnold and B.K. Mable, 2011 Validation of swabs as a non-destructive and relatively non-invasive DNA sampling method in fish. Molecular Ecology Resources 11**:** 107-109.
22. Lefort, V., J.-E. Longueville and O. Gascuel, 2017 SMS: Smart Model Selection in PhyML. Molecular Biology and Evolution**:** msx149.
23. Lieschke, G., and P. Currie, 2007 Animal models of human disease: zebrafish swim into view. Nature Reviews Genetics 8**:** 353-367.
24. Lydeard, C., and K.J. Roe, 1997 The phylogenetic utility of the mitochondrial cytochrome b gene for inferring relationships among Actinopterygian fishes, pp. 285–303 in *Molecular systematics of fish*,

edited by T.D. Kocher and C.A. Stepien. Academic Press, New York.

25. Mäkinen, H., A. Vasemägi, P. McGinnity, T. Cross and C. Primmer, 2014 Population genomic analyses of early-phase Atlantic Salmon (Salmo salar) domestication/captive breeding. Evolutionary Applications 8**:** 93–107.
26. Mayden, R.L., K.L. Tang, K.W. Conway and e. al, 2007 Phylogenetic relationships of Danio within the order cypriniformes: a framework for comparative and evolutionary studies of a model species. Journal of Experimental Zoology Part B-Molecular and Developmental Evolution 308B**:** 642-654.
27. Moore, W.S., and V.R. DeFilippis, 1997 The window of taxonomic resolution for phylogenies based on mitochondrial cytochrome b, pp. 83–119 in *Avian molecular evolution and systematics*, edited by D.P. Mindell. Academic Press, New York.
28. Pandolfi, V.C.F., A.L. Yamachita, F.P. de Souza, S.M. de Godoy, E.C.S. de Lima *et al.*, 2021 Development of microsatellite markers and evaluation of genetic diversity of the Amazonian ornamental fish Pterophyllum scalare. Aquaculture International
29. Paull, G.C., K.J.W. Van Look, E.M. Santos, A.L. Filby, D.M. Gray *et al.*, 2008 Variability in measures of reproductive success in laboratory-kept colonies of zebrafish and implications for studies addressing population-level effects of environmental chemicals. Aquatic Toxicology 87**:** 115-126.
30. Peakall, R., and P.E. Smouse, 2012 GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research - an update. Bioinformatics 28**:** 2537-2539.
31. Raymond, M., and F. Rousset, 1995 GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J. Heredity 86**:** 248-249.
32. Reed, D.H., and R. Frankham, 2003 Correlation between Fitness and Genetic Diversity. Conservation Biology 17**:** 230–237.
33. Rico, C., I. Rico and G. Hewitt, 1996 470 Million Years of Conservation of Microsatellite Loci among Fish Species, pp. 579-557 in *Proceedings: Biological Sciences ,*. Royal Society, London.
34. Robison, B.D., and W. Rowland, 2005 A potential model system for studying the genetics of domestication: behavioral variation among wild and domesticated strains of zebra danio (Danio rerio). Canadian Journal of Fisheries and Aquatic Sciences 62**:** 2046–2054.
35. Rozas, J., J.C. Sánchez-DelBarrio, X. Messeguer and R. Rozas, 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19**:** 2496-2497.
36. Ruzicka, L., D. Howe, S. Ramachandran, S. Toro, C. Van Slyke *et al.*, 2019 he Zebrafish Information Network: new support for non-coding genes, richer Gene Ontology annotations and the Alliance of Genome Resources. Nucleic Acids Res 8**:** D867-D873.
37. Scholz, S., S. Fischer, U. Gündel, E. Küster, T. Luckenbach *et al.*, 2008 The zebrafish embryo model in environmental risk assessment—applications beyond acute toxicity testing. Environmental Science and Pollution Research 15**:** 394–404.
38. Shi, D., H. Zhu, W. Li, S. Wang and J. Qu, 2020 Genetic Biodiversity of the Breeding Population of the Red and White Koi Carp. IOP Conf. Ser.: Earth Environ. Sci. 495**:** 012060.

39. Spence, R., G. Gerlach, C. Lawrence and C. Smith, 2008 The behaviour and ecology of the zebrafish, Danio rerio. Biological Reviews 83**:** 13-34.
40. Staff, Z., 2016 Mutation Details Curation of Older Features - Wild Type Line: AB. ZFIN Historical Data, pp.
41. Sterba, G., 1962 *Freshwater fishes of the world*. Vista Books, Longacre Press, London.
42. Suurväli, J., A.R. Whiteley, Y. Zheng, K. Gharbi, M. Leptin *et al.*, 2019 The Laboratory Domestication of Zebrafish: From Diverse Populations to Inbred Substrains. Mol. Biol. Evol. 37**:** 1056–1069.
43. Talwar, P.K., and A.G. Jhingran, 1991 *Inland fishes of India and adjacent countries*. Oxford & I. B. H. Publishing, Calcutta.
44. Thompson, J.D., D.G. Higgins and T.J. Gibson, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22**:** 4673-4680.
45. Whiteley, A.R., A. Bhat, E.P. Martins, R.L. Mayden, M. Arunachalam *et al.*, 2011 Population genomics of wild and laboratory zebrafish (Danio rerio). Molecular Ecology 20**:** 4259–4276.
46. Wright, D., R. Nakamichi, J. Krause and R.K. Butlin, 2006 QTL analysis of behavioural and morphological differentiation between wild and laboratory zebrafish (Danio rerio). Behavior Genetics 36**:** 271-284.

Tables

Table 1

Primer details for cytochrome b and the five microsatellite loci used to determine origin and diversity in a laboratory population of zebrafish.

Locus	Forward Sequence (5'–3')	Reverse Sequence (5'-3')	Ref
fishcytbzf	ACCACTGTTGTAGTTCAACTACAAGAAC	-	Fang *et al.* (2009)
HA-danio	-	CTCCGATCTTCGGATTACAAG	Mayden *et al.* (2007)
Ztri1	6FAM-AACTCAAACAAACAGAGCTG	ATAACACTTCCAGTTGACTG	Coe *et al.* (2008)
Z249	5HEX-TCTTCCCCTACAGGCACAGT	ATGATACGCAGTCAACGTATCG	Coe *et al.* (2008)
Z6104	6FAM-GGCTTTTCTCCAGTGAGTGC	CATGTGCCTATTGCCAACTG	Coe *et al.* (2008)
Z9230	5HEX-TAATCTACACCCGCAGCAGG	ATGGTTAATCAGCAAACGCC	Coe *et al.* (2008)
Z20450	6FAM-CCAAATCTACGCCCATGTCT	CGACCTCTGAATCTGCCTTT	Coe *et al.* (2008)

Table 2

Genetic diversity estimates obtained from *cytb* sequences of SA pet shop populations (indicated in red) and reference groups (Whiteley *et al.* 2011), expressed as haplotype frequency; haplotype diversity; nucleotide diversity; number of segregating sites; and with the number of sequences used indicated.

	Number of Haplotypes	Haplotype Diversity	Nucleotide Diversity	Number of Segregating Sites	Number of Sequences
ZFA	**5**	**0.681**	**0.0030**	**11**	**46**
ZFB	**2**	**0.429**	**0.0034**	**9**	**8**
ZFC	**4**	**0.733**	**0.0049**	**12**	**10**
AB	1	0.000	0.0000	0	10
SJA	1	0.000	0.0000	0	10
TM1	1	0.000	0.0000	0	10
PAR	7	0.876	0.0087	25	15
KHA	10	0.924	0.0023	13	15
BER	10	0.895	0.0039	24	15
SHK	6	0.889	0.0052	16	10
JOR	5	0.638	0.0026	13	15
PGM	8	0.923	0.0054	17	13
PNS	7	0.829	0.0055	20	15
UTR	8	0.895	0.0062	20	15
RCH	8	0.890	0.0126	67	14
CHT	6	0.571	0.0049	19	15
SRN	1	0.000	0.0000	0	3
WYD	2	1.000	0.0009	1	2

Table 3

Genetic diversity estimates per zebrafish population estimated from five microsatellite loci. Standard error for average number of alleles, observed heterozygosity, and unbiased expected heterozygosity is provided in parentheses. (*N*: Number of individuals, *N*a: number of alleles, H_o: observed heterozygosity, H_e: expected heterozygosity, F_{IS}: inbreeding coefficient)

	N	*Na*	H_o	H_e	F_{IS}
Population 1 (ZFA)	46	6.800 (0.663)	0.596 (0.064)	0.726 (0.026)	0.175
Population 2 (ZFB)	9	4.600 (0.245)	0.667 (0.061)	0.674 (0.048)	-0.019
Population 3 (ZFD)	10	6.000 (0.707)	0.720 (0.073)	0.743 (0.035)	0.021

Figures

Figure 1

Map of the study area and sampling locations of wild zebrafish specimens used by Whiteley *et al.* (2011). Sampling localities are indicated by: Black circles = Indian sites; Grey circles = Nepali sites; Grey squares = Bangladeshi sites. The recorded occurrences of zebrafish, as sourced from the Global Biodiversity Information Facility (GBIF) database, are indicated by triangles. The image was modified from Whiteley *et al.* (2011). See Supplementary Table S1 for abbreviation definitions.

Figure 2

Map of the study area and sampling locations of wild zebrafish specimens used by Whiteley *et al.* (2011). Sampling localities are indicated by: Black circles = Indian sites; Grey circles = Nepali sites; Grey squares = Bangladeshi sites. The recorded occurrences of zebrafish, as sourced from the Global Biodiversity Information Facility (GBIF) database, are indicated by triangles. The image was modified from Whiteley *et al.* (2011). See Supplementary Table S1 for abbreviation definitions.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 20211028SupplementaryData.docx
- BlomSupTableS2.xls