The phylogenetic position of zebrafish (*Danio rerio*) from south african pet shops

Elmarie Blom¹ · Willem G. Coetzer¹,2 · Sue-Rica Schneider¹ · J. Paul Grobler¹

Received: 4 January 2022 / Accepted: 26 April 2022 / Published online: 20 May 2022 © The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

Background  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.

Methods and results  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 lines 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.

Conclusion  These results can be used to guide the future development of laboratory lines suited to the needs at the UFS.

Keywords  Zebrafish · Cytochrome b · Microsatellites · Genetic diversity · Phylogenetic origin

Introduction

The zebrafish (*Danio rerio*) is a widely used model organism in biomedical research, developmental genetics, and neurophysiology [1, 2] and increasingly also in environmental studies [3, 4]. Zebrafish have several qualities that make them suitable for manipulation and use in research experiments. They are small (~ 2.5–4 cm long), 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 [2].

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) [2, 5]. 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 [6–8].

Over 400 laboratories worldwide use zebrafish in fundamental and applied research (The Zebrafish International Resource Center, www.zebrafish.org). Most laboratory lines used are the product of many generations bred in captivity. One of the most well-known zebrafish laboratory lines 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 [9]. 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
Genetic diversity within wild and captive bred zebrafish populations have previously been studied by Gratton et al. [13], Coe et al. [14], Whiteley et al. [5] and Balik-Meisner et al. [15]. According to Whiteley et al. [5], there exists a high level of genetic diversity within wild zebrafish populations. In contrast, genetic variation in laboratory line has been shown to be significantly lower than that of wild populations [5]. 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
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/) [21], EMBL (www.ebi.ac.uk/emb1) [22], and ZFIN [23], with the latter database specific to zebrafish. Primers developed for these loci can be cross amplified between related species. Rico et al. [24] 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 [25]. 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 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). These three groups forms the UFS zebrafish populations used for downstream breeding of new progeny. These fish showed no notable deviations from the normal wild-type zebrafish, and did not show notable morphological differences between each other.

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. [26]. Samples from dead fish (stored at -20 °C) were taken by means of tail cuttings of roughly 4 mm x 2 mm.

For comparative purposes, a further 178 cytb sequences from Whiteley et al. [5] were downloaded from the GenBank database (Accession numbers, JN234180–JN234356), including the cytb sequence of an outgroup species (Danio kymnath; Accession number, EF452733). See Fig. 1 for sampling localities from the Whiteley et al. [5] 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. [5]. For the forward primer we used fishcytbzaF from Fang et al. [27], with HA-danio from Mayden et al. [28] used as reverse primer (Table 1). This specific gene region was selected to compare our sample set to the dataset generated by Whiteley et al.
Nucleotide diversity ($\pi$) results were rounded to two significant digits.

For the zebrafish, 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$_2$O, 0.375 µl of each primer (10 µM 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, and a final extension step at 72 °C for 5 min.

The closely related Danio kyathit [28] was used as the outgroup (Accession number, EF452733).

A Maximum likelihood (ML) analysis of the identified haplotypes was performed using the online PhyML platform [34] (http://www.atgc-montpellier.fr/phyml/) to assess the relationship between our sequences and previously published data from Whiteley et al. [5] (Accession numbers JN234180–JN234356). Automatic model selection by Smart Model Selection (SMS) [35] was selected, and branch support was estimated by performing 1 000 bootstrap iterations. The best model was identified as GTR+$G$.

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 [36]. A paired t test was performed to determine if null alleles has a significant effect on the haplotypes (h), haplotype diversity (HD) and nucleotide diversity ($\pi$). Nucleotide diversity ($\pi$) 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 [32]. PhiPT is an analogue to $F_{ST}$, which is shown to be an ideal method for codominant data [33].

### Table 1

| Locus   | Forward Sequence (5’–3’) | Reverse Sequence (5’–3’) | Ref |
|---------|--------------------------|--------------------------|-----|
| fishcytbzf | ACCACTGTGG- TAGTCCAACCTACAAGAC | - | Fang et al. (2009) |
| HA-danio | - | CTGCCGATCTCTGCAGTGGTTGACAG | Mayden et al. (2007) |
| Ztri1   | 6FAM-AACTCAAACAAACAGGCTG | ATAAACACTTGCTCAGTGGTACAG | Coe et al. (2008) |
| Z249    | 5HEX-TCTTTCCCTACAGGCAGTGC | ATGATACGCAGCTAACCAGTATTTCGCAGTGC | Coe et al. (2008) |
| Z6104   | 6FAM-GGCTTTTCTCCAGTTGAGTGCTGC | CATGTCGCTATTGCTAACTGGGa | Coe et al. (2008) |
| Z9230   | 5HEX-TAATCTACAACCGGCACAGG | ATGGTTAATCTACAAACCAGGCC | Coe et al. (2008) |
| Z20450  | 6FAM-CCAAATCTACGCCATGCTCT | CGACCTCTGAATCTGCTT | Coe et al. (2008) |

[5]. 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$_2$O, 0.375 µl of each primer (10 µM 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. [14], were used to determine the genetic diversity levels for the zebrafish from different domestic sources. These microsatellite markers were selected so direct comparisons can be made between our results and that observed by Coe et al. [14] of wild and laboratory zebrafish lines. 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$_2$O, 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$_2$O, 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 dH$_2$O, 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 [29] using the ClustalW option [30]. DnaSP software [31] was used to calculate the number of haplotypes (h), haplotype diversity (HD) and nucleotide diversity ($\pi$). Nucleotide diversity ($\pi$) 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 [32]. PhiPT is an analogue to $F_{ST}$, which is shown to be an ideal method for codominant data [33].
effect on $F_{ST}$ estimates by comparing null allele corrected and uncorrected $F_{ST}$ values calculated by the excluding null allele (ENA) method [36]. The test for HWE and linkage disequilibrium was done using GENEPOP [37]. 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 [32]. Polymorphic information content was determined using Cervus [38]. Allelic richness and inbreeding coefficient were calculated using FSTAT 1.2 [39].

### Results

#### Results from cytb sequences

A 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. [5]. 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.

### Table 2

Genetic diversity estimates obtained from *cytb* sequences of SA pet shop populations (indicated in bold) and reference groups [5], expressed as haplotype frequency; haplotype diversity; nucleotide diversity; number of segregating sites; and with the number of sequences used indicated

| Haplotypes | 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

The pairwise population PhiPT values for the zebrafish *cytb* sequences generated from this study and sequences sourced from Whiteley et al. [5]. The PhiPT values are presented below the diagonal, with the $p$-values above the diagonal. South African pet shop fish are highlighted in grey.
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 (Table 3) 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 line 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). A strong 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 likeness 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. [5]. The phylogenetic analyses of the 70 haplotypes assessed revealed three major genetic clades. Membership of clades was as follows: Clade 1, the laboratory line, SA pet shop fish, Northern India, and western...
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 (Supplementary Table S3). The per locus estimated null allele frequencies ranged from 0.000 to 0.1860. Null allele presence can inflate F<sub>ST</sub> values [40]. However, no significant difference was observed between the ENA corrected (F<sub>ST</sub> = 0.019) and the uncorrected ENA (F<sub>ST</sub> = 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 (Table S3). All five loci showed significant levels of heterozygosity (Table S3), with two loci showing negative F<sub>IS</sub> values. No linkage disequilibrium was observed (Table S3). All three zebrafish lines at the UFS had high levels of heterozygosity, H<sub>O</sub> ranged from 0.596 to 0.720 and H<sub>E</sub> ranged from 0.674 to 0.743. ZFB was the only population to show a negative F<sub>IS</sub> value, suggesting an excess of heterozygotes (Table 4). No genetic differentiation was seen between the ZFA and ZFB populations (F<sub>ST</sub> = 0.000; p-value = 0.435). Low, but significant genetic differentiation was observed between the ZFA and ZFC populations (F<sub>ST</sub> = 0.029; p-value = 0.013), and similarly low genetic differentiation was seen between populations ZFB and ZFC (F<sub>ST</sub> = 0.053; p-value = 0.014).

Discussion

Origin of SA pet shop populations

Although founded independently from established laboratory line, the UFS population shows a high level of shared ancestry with these lines following ML analysis. Approximately 60% of the SA pet shop fish grouped with haplotypes 1 and 2 similarly to the established laboratory lines. The phylogenetic grouping containing the pet shop fish and the laboratory lines also share ancestry with the northern and north-eastern wild populations. This observation is in line with known populations histories, with established laboratory lines also historically bred from pet shop fish [41] 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 [5]. 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. [42] found wild populations from West Bengal to be most closely related to the tested laboratory lines. 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 lines. The populations from Nepal (KHA) and Bangladesh (CHT) represent distinct lines of zebrafish that diverged from the West Bengal populations before the laboratory and pet shop fish were established [43].

Overall, the ML-based phylogenetic results obtained were congruent with the results reported by Whiteley et al. [5]. 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. [42] and Whiteley et al. [5]. 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 line at UFS. This high level of genetic diversity could be the result of different source populations used to supplement the current pet shop populations [44]. Other examples of genetic diversity analyses performed on commercially traded fish include studies on red and white koi carp (Cyprinus carpio L) [45], guppies (Poecilia reticulata) [46], and freshwater angelfish (Pterophyllum scalare) [47]. High levels

| Population | N | Na | H<sub>O</sub> | H<sub>E</sub> | F<sub>ST</sub> | F<sub>IS</sub> |
|------------|---|----|-------------|-------------|-------------|-------------|
| ZFA        | 46| 6.800 (0.663) | 0.596 (0.064) | 0.726 (0.026) | 0.175       |
| ZFB        | 9 | 4.600 (0.245) | 0.667 (0.061) | 0.674 (0.048) | -0.019      |
| ZFD        | 10| 6.000 (0.707) | 0.720 (0.073) | 0.743 (0.035) | 0.021       |
of genetic diversity were reported in breeding populations of red and white koi carp [45]. 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 line. Importing existing zebrafish laboratory lines can be expensive from a South African perspective and establishing a new, highly diverse, laboratory line from pet shop fish will be more affordable. A genetically diverse line can be advantageous during population studies, for example: to study the effects of bottlenecks on diversity at specific genes. The established laboratory lines studied by Whiteley et al. [5] 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 line with restricted genetic diversity. This selective breeding is possibly the leading cause in the decline of genetic diversity seen in established laboratory lines. Selective breeding can also lead to a loss of fecundity and a skewed sex ratio, which will negatively impact the maintenance of genetic diversity [48]. To prevent such a decline, new fish will have to continuously be introduced to counteract the effects of selective breeding. The Cooch Behar (CB) line is an example of a newly established laboratory line from zebrafish that were taken directly from the wild [49]. These fish showed a reduced sex dimorphism with females being favoured (28 females and eight males) [49]. Further studies into the impact on the diversity for the CB line could help determine the implications of establishing a laboratory line directly from the wild.

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. [14] for the two commercial lines and the wild zebrafish population (\(H_O\) range = 0.525–0.714). In contrast, the \(H_O\) estimates for the majority of the lab lines studied by Coe et al. [14] 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) [45] and angelfish (average PIC = 0.587) [47]. 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. [14] 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 [50]. \(F_{IS}\) values obtained by Suurväli et al. [42] for their laboratory zebrafish line, did not deviate far from zero. The inbreeding seen in the current study, as well as in the study performed by Suurvali et al. [42], can be caused by reduced population sizes [50] 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 [47] 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 lines studied by Bleakley et al. [46] were all positive but did not statistically deviate from zero [46]. 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. South African pet shops are doing a good job at keeping the diversity levels high and should continue their current practices and not start to breed their own fish which might promote inbreeding and a loss of diversity. Further studies on additional genes could be performed to strengthen the current results. The SA pet shop zebrafish can therefore serve as a good source for the development of laboratory lines suited to the needs at the University of the Free State.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1007/s11033-022-07522-x.

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. Coetzee, 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.
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.

Declarations

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

References

1. Lieschke G, Currie P (2007) Animal models of human disease: zebrafish swim into view. Nat Rev Genet 8:353–367
2. Spence R, Gerlach G, Lawrence C, Smith C (2008) The behaviour and ecology of the zebrafish, *Danio rerio*. Biol Rev 83:13–34
3. Scholz S, Fischer S, Güdel U, Küster E, Luckenbach T, Voelker D (2008) The zebrafish embryo model in environmental risk assessment—applications beyond acute toxicity testing. Environ Sci Pollut Res 15:394–404
4. Paul G, Van Look KJW, Santos EM, Filby AL, Gray DM, Nash JP 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. Aquat Toxicol 87:115–126
5. Whiteley AR, Bhat A, Martins EP, Mayden RL, Conway KW (2016) Phylogenetic relationships and the implications for ecotoxicology studies. Ecotoxicol Environ Safety 142:394–400
6. Sterba G (1962) Freshwater fishes of the world. Vista Books, Longacre Press, London
7. Talwar PK, Jhingran AG (1991) Inland fishes of India and adjacent countries. Oxford & I. B. H. Publishing, Calcutta
8. Jayaram KC (1999) The freshwater fishes of the Indian region. Narendra Publishing House, Delhi
9. Staff Z (2016) Mutation Details Curation of Older Features - Wild Type Line: AB. ZFIN Historical Data.
10. Holden LA, Brown KH (2018) Baseline mRNA expression differs widely between common laboratory strains of zebrafish. Sci Rep. DOI: 10.1038/s41598-018-23129-4:8-4780
11. Robison BD, Rowland W (2005) A potential model system for studying the genetics of domestication: behavioral variation among wild and domesticated strains of zebra danio (*Danio rerio*). Mol Ecol 20:4259–4276
12. Wright D, Nakamichi R, Krause J, Butlin RK (2006) QTL analysis of behavioural and morphological differentiation between wild and laboratory zebrafish (*Danio rerio*). Behav Genet 36:271–284
13. Gratton P, Allegrenucci G, Gallozzio M, Fortunato C, Ferreri F, Sbordoni V (2004) Allozyme and microsatellite genetic variation in natural samples of zebrafish, *Danio rerio*. J Zool Syst Evol Resarch 42:54–62
14. Cote TS, Hamilton PB, Griffiths AM, Hodgson DJ, Wahab MA, Tyler CR (2009) Genetic variation in strains of zebrafish (*Danio rerio*) and the implications for ecotoxicology studies. Ecotoxicology 18:144–150
15. Balik–Meisner M, Truong L, Scholl EH, Tanguay RL, Reif DM (2018) Population genetic diversity in zebrafish lines. Mamm Genomes 29:90–100
16. Irwin DM, Kocher TD, Wilson AC (1991) Evolution of the cytochrome b gene of mammals. J Mol Evol 32:128–144
17. Lydeard C, Roe KJ (1997) The phylogenetic utility of the mitochondrial cytochrome b gene for inferring relationships among Actinopterygian fishes. In: Kocher TD, Stepnie CA (eds) Molecular systematics of fish. Academic Press, New York, pp 285–303
18. Moore WS, DeFilippis VR (1997) The window of taxonomic resolution for phylogenies based on mitochondrial cytochrome b. In: Mindell DP (ed) Avian molecular evolution and systematics. Academic Press, New York, pp 83–119
19. Espositi MD, De Vries S, Crimi M, Ghelli A, Patarnello T, Meyer A (1993) Mitochondrial cytochrome b: evolution and structure of the protein. Biochem Biophys Acta 1143:243–271
20. Broughton RE, Milam JE, Roe BA (2001) The Complete Sequence of the Zebrafish (*Danio rerio*) Mitochondrial Genome and Evolutionary Patterns in Vertebrate Mitochondrial DNA. Cold Spring Harbor Laboratory Press, pp 1958–1967
21. Clark K, Karsch-Mizrachi I, Lipman D, Ostell J, Sayers E (2016) GenBank Nucleic Acids Res 44:D67–D72
22. Kanz C, Aldebert P, Althorpe N, Baker W, Baldwin A, Bates K et al (2005) The EMBL Nucleotide Sequence Database. Nucleic Acids Res 33:D29–D33
23. Ruizicza L, Howe D, Ramachandran S, Toro S, Van Slyke C, Bradford Y et al (2019) The Zebrafish Information Network: new support for non-coding genes, richer Gene Ontology annotations and the Alliance of Genome Resources. Nucleic Acids Res 8:E867–D73
24. Ricco C, Rico I, Hewitt G (1996) 470 Million Years of Conservation of Microsatellite Loci among Fish Species. Proceedings: Biological Sciences,. London: Royal Society; p. 579–57
25. Hoshino AA, Bravo JP, Nobile PM, Morelli KA (2012) Microsatellites as Tools for Genetic Diversity Analysis. In: Caliskan PM, editor. Genetic Diversity in Microorganisms: InTech; p. ISBN: 978-53-51-0064-5
26. Le Vin AL, Adam A, Tedder A, Arnold KE, Mable BK (2011) Validation of swabs as a non-destructive and relatively non-invasive DNA sampling method in fish. Mol Ecol Resour 11:107–109
27. Fang F, Noren M, Liao TY, Kallersjo M, Kullander S (2009) Molecular phylogenetic interrelationships of the south Asian cyprinid genera *Danio*, Devario, and Microrasbora (Teleostei, Cyprinidae, Danioninae). Zoolog Scr 38:237–256
28. Mayden RL, Tang KL, Conway KW (2010) Populations and the Alliance of Genome Resources. Nucleic Acids Res 8:E867–D73
29. Mayden RL, Tang KL, Conway KW (2010) Genetic 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 42:292–300
30. Drummond AJ, Ashton B, Cheung M, Heled J, Karsem M, Gous M (2009) Geneious v4.7 et al (2009)
31. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680
32. Rozas R, Sánchez-DelBarrio JC, Meseguer X, Rozas R, DnaSP (2003) DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497
33. Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research - an update. Bioinformatics 28:2527–2539
34. Ang CC, O’Brien MJ, Ng KKS, Lee PC, Hector A, Schmid B et al (2016) Genetic diversity of two tropical tree species of the Diparopcarpaceae following logging and restoration in Borneo: high genetic diversity in plots with high species diversity. Plant Ecol Divers 9:459–469
35. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst Biol 59:307–321
36. Lefort V, Longueville J-E, Gascuel O (2017) SMART Model Selection in PhyML. Molecular Biology and Evolution. :msx149
37. Chapuis M-P (2007) Microsatellite null alleles and estimation of population differentiation. Mol Biol Evol 24:621–631
37. Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Heredity 86:248–249
38. Kalinowski S, Taper M, Marshall T (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 22:1099–1106
39. Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. J Hered 86:485–486
40. Carlsson J (2008) Effects of microsatellite null alleles on assignment testing. J Hered 99:616–623
41. Howe DG, Bradford YM, Conlin T, Eagle AE, Fashena D, Frazer K et al (2013) ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics.
42. Suurväli J, Whiteley AR, Zheng Y, Gharbi K, Leptin M, Wiehe T (2020) The Laboratory Domestication of Zebrafish: From Diverse Populations to Inbred Substrains. Mol Biol Evol 37:1056–1069
43. Suurväli J, Whiteley AR, Zheng Y, Gharbi K, Leptin M, Wiehe T (2019) The Laboratory Domestication of Zebrafish: From Diverse Populations to Inbred Substrains. Mol Biol Evol 37:1056–1069
44. Mäkinen H, Vasmägi A, McGinnity P, Cross T, Primmer C (2014) Population genomic analyses of early-phase Atlantic Salmon (Salmo salar) domestication/captive breeding. Evol Appl 8:93–107
45. Shi D, Zhu H, Li W, Wang S, Qu J (2020) Genetic Biodiversity of the Breeding Population of the Red and White Koi Carp. IOP Conf Ser: Earth Environ Sci 495:012060
46. Bleakley BH, Eklund AC, Brodie I (2008) E.D. Are Designer Guppies Inbred? Microsatellite Variation in Five Strains of Ornamental Guppies, Poecilia reticulata, Used for Behavioral Research. Zebrafish. ; 5
47. Pandolfi VCF, Yamachita AL, de Souza FP, de Godoy SM, de Lima ECS, Feliciano DC et al (2021) Development of microsatellite markers and evaluation of genetic diversity of the Amazonian ornamental fish Pterophyllum scalare. Aquaculture International
48. Crim MJ, Lawrence C (2021) A fish is not a mouse: understanding differences in background genetics is critical for reproducibility. Lab Anim 50:19–25
49. Wilson CA, High SK, McCluskey BM, Amores A, Yan Y-l, Titus TA et al (2014) Wild sex in zebrafish: loss of the natural sex determinant in domesticated strains. Genetics 198:1291–1308
50. Reed DH, Frankham R (2003) Correlation between Fitness and Genetic Diversity. Conserv Biol 17:230–237
51. GBIF.org (2022) GBIF Occurrence Download.
52. Hijmans RJ, Guarino L, Cruz M, Rojas E (2001) Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. Plant genetic resources newsletter. :15 – 9

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.