



Characterizing the genetic structure of introduced Nile tilapia (*Oreochromis niloticus*) strains in Tanzania using double digest RAD sequencing

Mbiru Moses^{1,2} · Matern S. P. Mtolera¹ · Leonard J. Chauka¹ ·
Fernando A. Lopes, Pinto³ · Dirk Jan de Koning³ · Ross D. Houston⁴ ·
Christos Palaiokostas^{3,4} 

Received: 5 March 2019 / Accepted: 26 September 2019 / Published online: 4 November 2019
© The Author(s) 2019

Abstract

Tilapia hatcheries in Tanzania rely heavily on importing germplasm. Nevertheless, the genetic structure of the imported stocks is poorly understood. In the current study, the level of genetic diversity and differentiation of eight populations of Nile tilapia (*Oreochromis niloticus*) strains imported in Tanzania was investigated. Four of the studied strains originated from Thailand, three from Uganda, and one from the Netherlands. Double-digest restriction site-associated DNA sequencing (ddRAD-seq) was applied to identify and genotype single nucleotide polymorphisms (SNPs). In total, 2214 SNPs passed all the quality control steps and were utilized for downstream analysis. Mean heterozygosity estimates were higher for the Thailand strains (H_o , 0.23) compared with the strains from Uganda (H_o , 0.12). Low genetic distance was observed amongst populations from the same geographic origin (F_{st} , 0.01–0.04). However, genetic distance between populations from different geographic origins was substantial (F_{st} , 0.24–0.44). Bayesian model-based clustering (STRUCTURE) and discriminant analysis of principal components (DAPC) grouped the studied animals into three distinct clusters. A cross-validation approach (where 25% of animals from each population were considered of unknown origin) was conducted in order to test the efficiency of the SNP dataset for identifying the population of origin. The cross-validation procedure was repeated 10 times resulting in approximately 97% of the tested animals being allocated to the correct geographic population of origin. The breeding history and hatchery practices used to manage these stocks prior and after import appear to be the main factors for the genetic diversity observed in this study. Our study will help inform hatchery stock management and future breeding program designs in Tanzania.

Keywords Nile tilapia · ddRAD-seq · Genetic diversity

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10499-019-00472-5>) contains supplementary material, which is available to authorized users.

✉ Christos Palaiokostas
christos.palaiokostas@slu.se

Extended author information available on the last page of the article

Background

Nile tilapia (*Oreochromis niloticus*) is the focal farmed fish in Tanzania accounting for over 95% of the country's aquaculture production volume (Kaliba et al. 2006). Farming is mainly conducted in small ponds. In addition, cage farming is also practiced in the country's main lakes: Victoria, Tanganyika, and Nyasa (URT 2016, 2017). Nevertheless, the contribution of farmed tilapia to the fisheries sector and its total economic value is still limited. In 2016, aquaculture (mainly farmed tilapia) accounted for only 2.9% (11,000 tonnes; URT 2017) of the total fish production. Due to the need for increased production (400,000 tonnes per year to meet the FAO recommendation of an average 17 Kg/capita/year; URT 2014), private hatchery operators have been importing tilapia strains from Asia, Europe, and within Africa (mainly Uganda).

The development of genetically improved farmed tilapia (GIFT) from a base population of 7 Nile tilapia strains by ICLARM (currently WorldFish) was a major achievement in the history of tilapia aquaculture (FAO 2011; ADB 2004; Azhar et al. 2004; ADB 2005). Undoubtedly, the GIFT strain has been instrumental in enhancing the worldwide production of Nile tilapia (ADB 2005; Li and Cai 2008; Ansah et al. 2014). Similar approaches have been practiced by several tilapia aquaculture research and development programs, like GET-EXCEL (Tayamen 2004), GenoMar Supreme Tilapia (GST) (Zimmermann and Natividad 2004), Hainan Progift (Thodesen et al. 2013), and FAST (Bolívar 1998).

In Tanzania, no national breeding program exists for tilapia, but some of the aforementioned improved strains have been imported with a view of increasing production output. In particular, the most common imported strains include (i) the Chitralada imported in 2016 by EDEN Agri-Aqua and JAN Aqua Center hatcheries from the Asian Institute of Technology (AIT; Khlong Luang, Pathum Thani, Thailand), (ii) the BIG NIN and GIFT from Thailand, imported in 2017 by Mpanju Fish Farm from the Nam Sai Farms Co. Ltd (Bangrabow, A.Bansang, Prachinburi, Thailand), and (iii) the *silver* strain/YY super males strain imported in 2017 by Rift Valley AquaCulture Farm from Til-Aqua International (Velden, The Netherlands).

Interestingly, the aforementioned genetically improved strains were originally collected from Africa (ADB 2004, 2005). The Chitralada strain originated from a base population of 50 animals from an Egyptian stock that was later introduced in Thailand (Damrongratana and Kessanchai 1966; Pullin and Capili 1988). In comparison, the GIFT strain originated from a diverse tilapia germplasm of more than 2000 individuals from Egypt, Ghana, Kenya, and Senegal (ADB 2004, 2005). Additional tilapia strains derived from the GIFT breeding nucleus include the BIG NIN (<http://tilapiathai.com/project/nile-tilapia/>). Furthermore, Til-Aqua International developed a silver tilapia strain from an all-male GIFT silver strain and a female line from various strains like GIFT and Chitralada from Egypt and the Philippines.

On the other hand, imported tilapia germplasm from Uganda is not derived from selective breeding. Generally, most of the hatcheries in Uganda (~70%) obtained their broodstock from different lakes, i.e., Victoria, Kyoga, and Albert (Mwanja et al. 2015). Common hatchery practices in the absence of selective breeding may result in the reduction of genetic diversity and inbreeding accumulation. For example, consanguineous matings as a consequence of limited numbers of broodstock and mass selection for growth (Aho et al. 2006; Qin et al. 2007; Smallbone et al. 2016) are still relatively common. The adverse effects of inbreeding accumulation can be minimized with selective breeding programs that utilize pedigree records. However, such programs are relatively expensive, and some GIFT and GIFT-derived strains

have already suffered a significant reduction of genetic diversity when compared with the original stocks (McKinna et al. 2010). Therefore, the accurate evaluation of genetic purity and diversity of stocks over time should be a priority.

High-throughput sequencing has revolutionized the field of genetics, enabling cost-effective discovery and characterization of molecular polymorphisms (Mardis 2008a, b; Zhang et al. 2011; Koboldt et al. 2013; Attia and Saeed 2016). Restriction site-associated DNA (RAD) sequencing (Baird et al. 2008) and double-digest RAD sequencing (ddRAD-seq; Peterson et al. 2012) rely in high-throughput sequencing for genotyping multiplexed samples. Both RAD and ddRAD-seq have been widely applied in aquaculture breeding and genetics studies (Robledo et al. 2018). In particular, ddRAD-seq has been applied for genotyping large multiplexed datasets (e.g., Maroso et al. 2018), construction of genetic linkage maps (e.g., Recknagel et al. 2013; Oral et al. 2017), analyzing life history traits (e.g., Pukk 2016), mapping sex determining loci (e.g., Palaiokostas et al. 2015; Brown et al. 2016), genomic predictions and genome-wide association studies (e.g., Barria et al. 2018), assessing genetic diversity (e.g., Hosoya et al. 2018; Tony et al. 2015; Siccha-Ramirez et al. 2018; Torati et al. 2019), phylogeography (e.g., Stobie et al. 2018), and species identification in tilapias (Syaifudin et al. 2019).

The objective of this study was to study the level of genetic diversity and differentiation of eight Nile tilapia strains introduced into Tanzania by various sources using ddRAD-seq. The findings of this study will improve knowledge of the genetic structure of existing farmed tilapia strains, and will inform future tilapia breeding program initiatives in Tanzania.

Material and methods

Sample collection, processing, and DNA extraction

A total of 160 Nile tilapia fin clip samples derived from eight strains ($n = 20$ per strain) were collected from seven different hatcheries located in different regions in Tanzania: Silver-YY, BIG NIN, Chitralada-N, Chitralada-E, GIFT, Ruvu Farm-R, Chifive-C and Muleba-M (Table 1; Fig. 1). Regarding the GIFT strain, no information was available, whether it directly originated from the World Fish Center. Fish were reared in labelled plastic tanks of $1.5 \text{ m} \times 1.5 \text{ m} \times 1.5 \text{ m}$ at the Institute of Marine Sciences Mariculture Centre (IMS-MC) at Bweni village, Pangani District in Tanga, Tanzania. Fin clips approximately 1.5-cm long were taken and preserved in 95% ethanol and stored at $-20 \text{ }^\circ\text{C}$. Total genomic DNA was extracted with the QIAasympphony kit using an automated DNA extraction robot (QIAasympphony, QIAGEN, Germany) according to the manufacturer's instructions. Following the DNA extraction, quantification of DNA samples was done using Qubit fluorometer (Thermo Fisher Scientific, USA) and diluted by TE buffer to $20 \text{ ng}/\mu\text{L}$ followed by gel electrophoresis (1% agarose gel) to assess DNA quality.

ddRAD library preparation and sequencing

The ddRAD library preparation protocol was based on the methodology originally reported by Peterson et al. (2012), with modifications described in Palaiokostas et al. (2015). Briefly, for each library, $3\text{-}\mu\text{L}$ DNA samples ($20 \text{ ng}/\mu\text{L}$) were simultaneously digested with two high-fidelity restriction enzymes: *Sbf*I (CCTGCA|GG recognition site) and *Sph*I (GCATG|C

Table 1 Hatchery sources of eight Nile tilapia strains in Tanzania and country of origin used for ddRAD-seq

Name of strains	Name of hatcheries	Districts	Origin of the strains
Silver-YY	Rift Valley	Kinondoni	Til-Aqua Inter, Netherland
BIG NIN	Mpanju Fish Farm	Misungwi	Nam Sai Farm, Thailand
Chitralada-N	JAN Aqua Center	Mkuranga	Asian Institute of Technology, Thailand
Chitralada-E	EDEN Agri-Aqua	Ilala	Asian Institute of Technology, Thailand
Ruvu-Farm-R	Ruvu Farm	Bagamoyo	AA Fisheries and Aqua Cons LTD, Uganda
GIFT	Mpanju Fish Farm	Misungwi	Nam Sai Farm, Thailand
CHIFIVE-C	CHIFIVE Farm	Busega	Unknown Hatchery, Uganda
Muleba-M	J and B Ruhanga	Muleba	Unknown Hatchery, Uganda

recognition site), both sourced from New England Biolabs, (NEB) UK. Digestions were incubated for 40 min at 37 °C in 1× CutSmart Buffer (NEB). Following cooling to c. 22 °C, 3 µL of a premade barcode/adaptor mix was added to each digested DNA sample. The adaptor mix comprised individual-specific barcoded combinations of P1 (*SbfI* compatible) and P2 (*SphI* compatible) adaptors at 6 nM and 72 nM concentrations, respectively, in 1× reaction buffer 2 (NEB). Ligation was performed over 2 h at 22 °C by the addition of 3 µL of a ligation mix comprising 4 mM rATP (Promega, UK), and 2000 cohesive end units of T4 ligase (NEB) in 1× CutSmart Buffer. The ligated samples were then heat denatured at 65 °C for 20 min and cooled to room temperature. Thereafter, samples were combined into one pool. The pooled library was column purified (MinElute PCR Purification Kit, Qiagen, UK), and eluted in 80 µL EB buffer (Qiagen, UK). Size selection of fragments, ranging from approximately 400 bp to 700 bp, was performed by agarose gel separation. Following gel purification (MinElute Gel Extraction Kit, Qiagen, UK), the eluted size-selected template DNA (60 µL in EB buffer) was PCR amplified (13–14 cycles PCR dependent on library: 24 separate 12.5 µL reactions, each with 1 µL template DNA) using a high-fidelity Taq polymerase (Q5 Hot Start High-Fidelity DNA Polymerase, NEB). The PCR reactions were column purified (MinElute PCR Purification Kit) and eluted in 50 µL of EB buffer. A final cleanup step was performed using an equal volume of AMPure magnetic beads (Perkin-Elmer, UK). The final library was eluted in 20 µL EB buffer and QUBIT quantified. The library was sequenced at Edinburgh Genomics Facility, University of Edinburgh, on an Illumina HiSeq 4000 instrument (Illumina, Inc., San Diego, CA) using 150 base-paired end reads (v4 chemistry).

Genotyping ddRAD alleles

Sequenced reads were aligned to the Nile tilapia reference genome assembly version GCA_001858045.2 (Conte et al. 2017) using bowtie2 (Langmead and Salzberg 2012). Uniquely aligned reads were retained for downstream analysis. The aligned reads were sorted into RAD loci, and SNPs were identified from both P1 and P2 reads using the Stacks software v2.4 (Catchen et al. 2011; Rochette et al. 2019). SNPs were detected using gstacks (*-var-alpha 0.001 -gt-alpha 0.001 -min-mapq 40*). SNPs with minor allele frequency (MAF) below 0.05, heterozygosity above 0.7, and those for which > 25 % of individuals had missing genotypes were discarded. Moreover, SNPs showing substantial coverage differences (3 or more standard deviations) from the mean obtained coverage were removed. Finally, only animals with less than 25% missing genotypes were retained for downstream analysis. The aligned reads in the format of bam files were deposited in the National Centre for Biotechnology Information (NCBI) repository under project ID PRJNA518067 (File_S1).

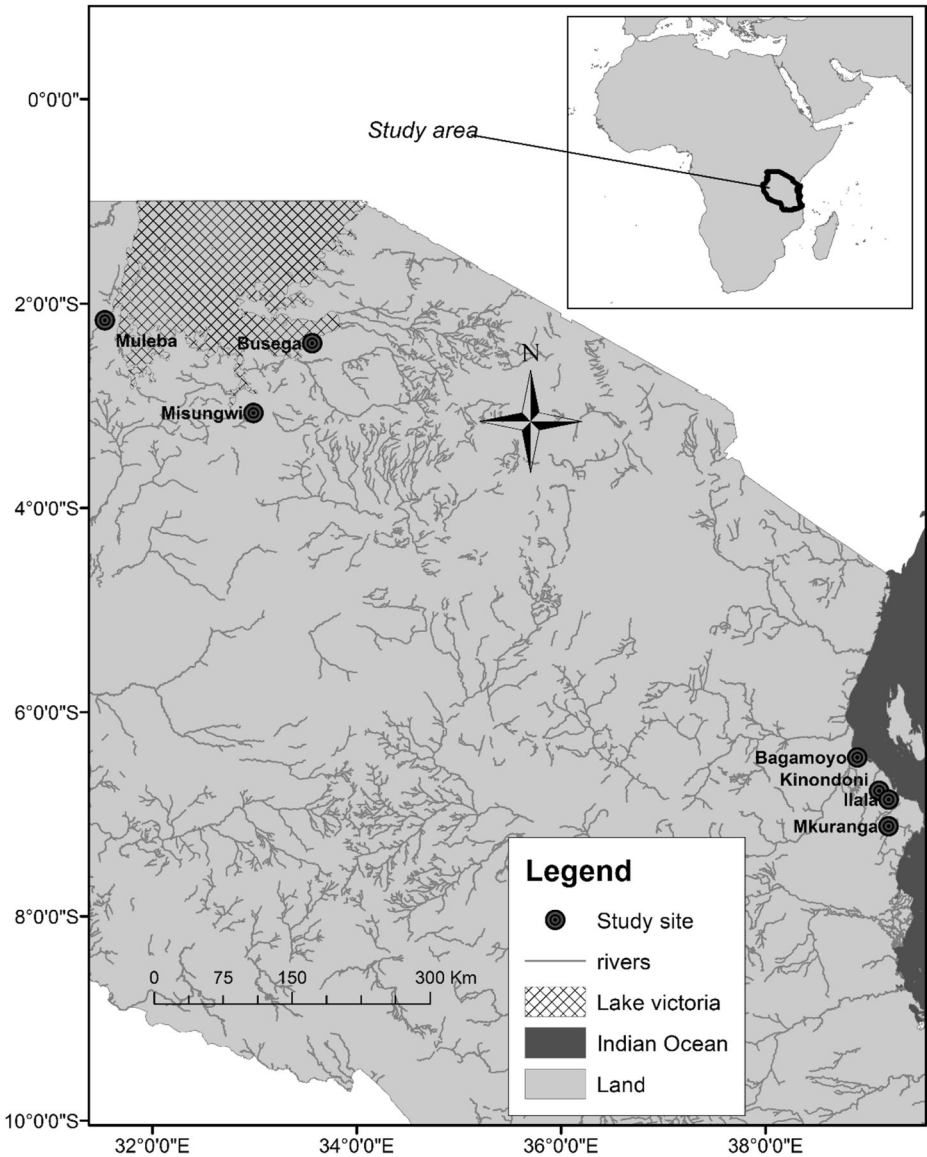


Fig. 1 Locations of sampling sites

Population genetic analyses

Mean observed heterozygosity (H_o), expected heterozygosity (H_e), and average individual inbreeding coefficients (F_{is}) were estimated using the Stacks software v2.4 (Catchen et al. 2011). The R package *StAMPP* (Pembleton et al. 2013) was used to estimate pairwise F_{st} values and the corresponding confidence intervals (using 1000 bootstraps). Principal component analysis (PCA) was implemented using the package *adegenet* v2.1.1 (Jombart 2008) from R-3.5.1 (R Core Development Team 2018). Population admixture was analyzed with

STRUCTURE v.2.3.4 (Pritchard et al. 2000; Falush et al. 2003; Hubisz et al. 2009) using K values from 2 to 5. Markov chain Monte Carlo (MCMC) of 100,000 iterations with a burn-in period of 10,000 was carried out for each K value. For each tested K value, 3 independent MCMC samplings were performed. The obtained posterior probability values (Pritchard et al. 2000) were used to determine the optimal number of clusters. Structure results were interpreted using Structure Harvester (Earl 2012) and CLUMPAK (Kopelman et al. 2015) for identifying the most probable number of clusters. Additionally, the pattern of population structure was also investigated using discriminant analysis of principal components (DAPC) (Jombart et al. 2010). The Bayesian information criterion (BIC) was used for selecting the optimal number of clusters (K) based on the elbow method (Jombart et al. 2010).

Assignment of individuals to population using the SNP dataset

A 4-fold cross validation scheme was performed using the R package adegenet v2.1.1 (Jombart 2008) in order to test the utility of the SNP dataset to correctly identify the geographic population of origin. The origin of 25% animals from each geographic population was masked and formed a test dataset. Predictions regarding population of origin on the aforementioned test set were performed using information obtained through DAPC (*predict.dapc*) on the remaining training data set. The entire procedure was repeated 10 times in order to minimize potential bias due to the stochasticity of sample allocation in the training /test datasets.

Results

SNP identification

The alignment rate of the sequenced reads to the tilapia reference genome (Genbank accession GCA_001858045.2; Conte et al. 2017) ranged between 94 and 97% across the tested animals. In total, 37,683 putative ddRAD loci were identified; out of which, 3791 contained one or more SNPs. The mean sequence coverage of the identified loci was approximately 170X (SD, 70X). In total, 2214 SNPs passed all quality control steps and were retained for downstream analysis. Moreover, 5 animals were removed due to high rate of missing genotypic data (4 animals from the Chifive-C and 1 animal from the Ruvu Farm-R) resulting in a dataset of 155 animals for the subsequent analysis.

Genetic diversity: distance

The Chitralada-N strain showed the highest level of expected heterozygosity (H_e , 0.27 ± 0.003) followed by the other Thai-derived strains Chitralada-E, BIG NIN, and GIFT (H_e , 0.26 ± 0.003). In contrast, the lowest expected heterozygosity was observed in the Ugandan strains (Ruvu Farm-R, CHIFIVE-C, and Muleba-M) (Table 2). Regarding the F_{is} metric, the lowest value was detected in the silver-YY strain and the highest in the Chitralada (Table 2). Low genetic distance was observed amongst populations from the same geographic origin (F_{st} , 0.01–0.04; Table 3). On the contrary, the genetic distance between populations from different geographic origins was substantial (F_{st} , 0.24–0.44). The estimated pairwise F_{st} values revealed a low genetic distance amongst strains imported from Thailand (BIG NIN, Chitralada-

Table 2 Genetic diversity indices for introduced Nile tilapia (*Oreochromis niloticus*) sequenced using ddRAD-seq. *N*, number of samples per locus; *H_o*, observed heterozygosity; *H_e*, expected heterozygosity; *F_{is}*, inbreeding coefficient

Name of strain	<i>N</i>	<i>H_e</i> (mean ± SE)	<i>H_o</i> (mean ± SE)	<i>F_{is}</i> (mean ± SE)
Silver-YY	20	0.16 ± 0.003	0.23 ± 0.006	− 0.14 ± 0.02
BIG NIN	20	0.26 ± 0.003	0.23 ± 0.003	0.10 ± 0.03
Chitralada-N	20	0.27 ± 0.003	0.23 ± 0.003	0.13 ± 0.03
Chitralada-E	20	0.26 ± 0.003	0.24 ± 0.003	0.09 ± 0.03
Ruvu-Farm-R ¹	19	0.14 ± 0.003	0.13 ± 0.003	0.02 ± 0.02
GIFT	20	0.26 ± 0.003	0.24 ± 0.003	0.07 ± 0.02
Chifive-C	16	0.11 ± 0.003	0.11 ± 0.003	0.01 ± 0.02
Muleba-M	20	0.14 ± 0.003	0.13 ± 0.003	0.08 ± 0.02

N, Chitralada-E, and GIFT) with values ranging from 0.01 to 0.04. Likewise, strains from Uganda (Ruvu Farm-R, Chifive-C, and Muleba-M) showed *F_{st}* values ranging from 0.01 to 0.03. Regarding the strain from Netherlands (silver-YY strain), higher genetic distance was found with the strains from Uganda (*F_{st}*, 0.38–0.44), compared with the strains from Thailand (*F_{st}*, 0.26–0.30). The lowest genetic distance amongst all tested populations was observed between the Thai strains of Chitralada-N and E and the Ugandan strains of Ruvu Farm-R and

Table 3 *F_{st}* values and corresponding confidence intervals (95%; 1000 bootstraps) between eight farmed strains of Nile tilapia (*Oreochromis niloticus*)

Population 1	Population 2	<i>F_{st}</i>	Confidence interval
Silver-YY	Big-Nin	0.29	0.28–0.31
Silver-YY	Chitralada-N	0.26	0.25–0.27
Silver-YY	Chitralada-E	0.28	0.26–0.29
Silver-YY	Ruvu-Farm-R	0.40	0.38–0.42
Silver-YY	GIFT	0.30	0.28–0.31
Silver-YY	Chifive-C	0.44	0.42–0.45
Silver-YY	Muleba-M	0.38	0.36–0.39
Big-Nin	Chitralada-N	0.02	0.02–0.03
Big-Nin	Chitralada-E	0.03	0.02–0.03
Big-Nin	Ruvu-Farm-R	0.31	0.30–0.33
Big-Nin	GIFT	0.02	0.02–0.03
Big-Nin	Chifive-C	0.35	0.33–0.36
Big-Nin	Muleba-M	0.31	0.30–0.33
Chitralada-N	Chitralada-E	0.01	0.00–0.01
Chitralada-N	Ruvu-Farm-R	0.24	0.23–0.26
Chitralada-N	GIFT	0.03	0.02–0.03
Chitralada-N	Chifive-C	0.28	0.26–0.29
Chitralada-N	Muleba-M	0.24	0.23–0.26
Chitralada-E	Ruvu Farm-R	0.29	0.28–0.31
Chitralada-E	GIFT	0.04	0.03–0.04
Chitralada-E	Chifive-C	0.33	0.31–0.34
Chitralada-E	Muleba-M	0.29	0.28–0.31
Ruvu Farm-R	GIFT	0.32	0.30–0.33
Ruvu Farm-R	Chifive-C	0.03	0.02–0.04
Ruvu Farm-R	Muleba-M	0.01	0.01–0.02
GIFT	Chifive-C	0.35	0.33–0.36
GIFT	Muleba-M	0.32	0.30–0.33
Chifive-C	Muleba-M	0.03	0.03–0.04

Muleba-M (F_{st} , 0.01), while the highest was found between the silver-YY and the Chifive-C (F_{st} , 0.44; Table 3).

Population structure

Principal component analysis (PCA) was used to visualize individual relationships within and between populations. The first and second principal components accounted for 24% and 10% of the variation respectively. Overall, PCA indicated the existence of 3 genetic groups (Fig. 2). STRUCTURE analysis provided evidence regarding the genetic structure amongst the tested Nile tilapia populations indicating that $K = 4$ was the most probable number of clusters (File_S2). Indications for admixture between the strains from Thailand and the Ugandan strains were suggested particularly for the Chitralada-N and E strains (Fig. 3). Furthermore, DAPC supported a close relationship of strains based on their geographical origin. The difference of BIC values between the most probable K values (3 to 5) was less than 1%, with lowest value for $K = 4$ (File_S3). Regarding the silver-YY strain from Netherlands, both STRUCTURE and DAPC suggested the existence of an isolated cluster, with the exception of one animal from the Ugandan strains that appeared indistinguishable from the silver-YY strain (Fig. 4).

Population prediction using the SNP dataset

Predictions regarding the geographic origin were performed using DAPC on the test set. Using a 4-fold cross validation scheme, a successful population assignment rate of approximately 97% was obtained (Fig. 5). The mean correct assignment rate for the Netheland strain was

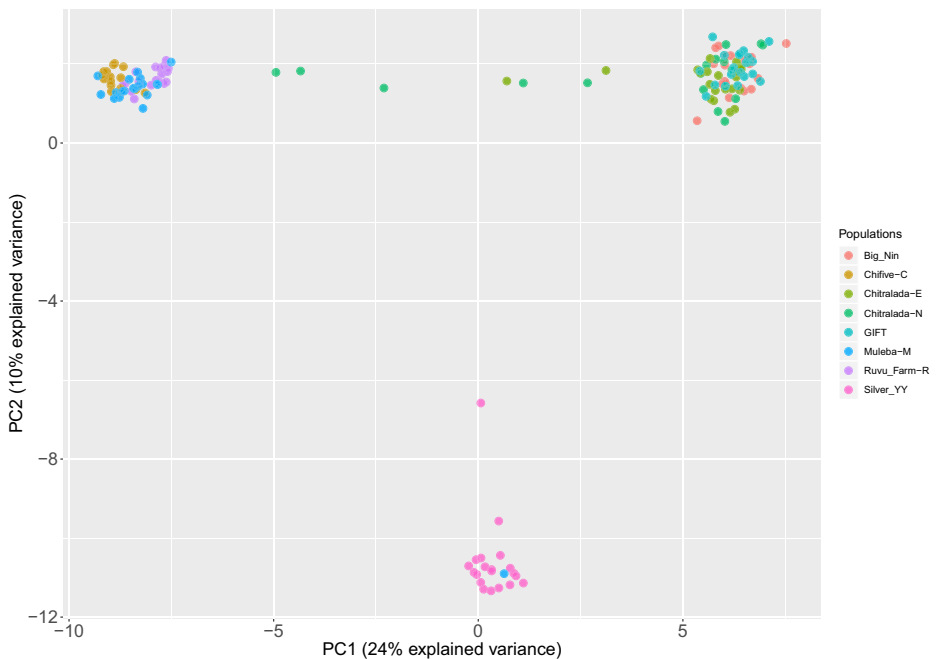
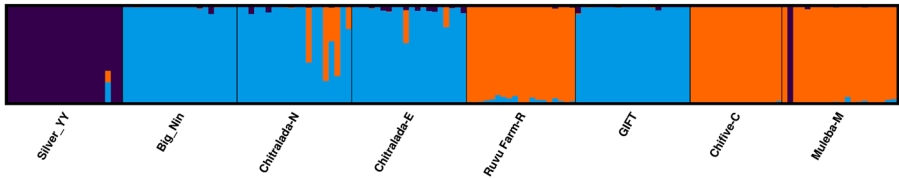
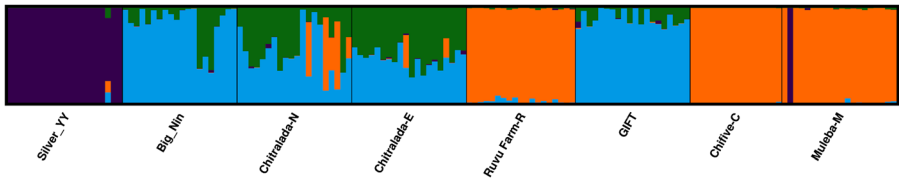


Fig. 2 Principal component analysis (PCA) based on allele frequencies for Nile tilapia (*Oreochromis niloticus*) strains introduced and farmed in Tanzania

K=3



K=4



K=5

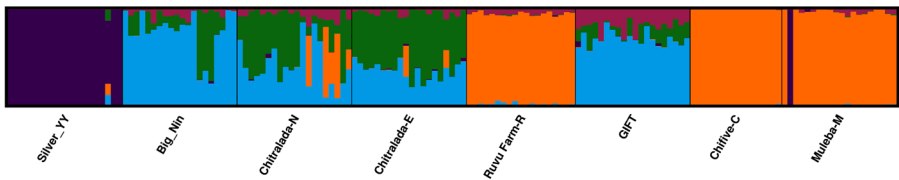


Fig. 3 The admixture analysis assigned individuals in clusters ($K = 3-5$) using eight strains. Each single vertical bar represents an individual, and each color represents a probability that the individual is assigned to each genepool

98%. Moreover, the correct assignment rates for the Thailand and Ugandan populations were 97% and 98% respectively.

Discussion

Our study revealed higher genetic variation in the case of animals that originated from selective breeding programs. According to the obtained estimate of H_e , both the silver-YY and the Thai strains (Chitralada-N, Chitralada-E, BIG NIN, and GIFT) had higher genetic variation in comparison with the Ugandan strains (Ruvu Farm-R, CHIFIVE-C, and Muleba-M). This observation is in agreement with findings of previous studies which compared improved strains with local farmed strains by using microsatellite markers (Romana-Eguia et al. 2004; Baggio et al. 2016). Hatchery practices and genetic management of fish stocks constitute key factors affecting the genetic diversity and long-term sustainability of the breeding populations. Both the silver-YY and the Thai strains of our study are based on selective breeding programs (ADB 2004, 2005; Mair 1997) with appropriate hatchery management (Bhujel 2013). However, to the best of our knowledge, the utilized broodstock of the Ugandan strains lack the availability of pedigree records that could be of value for avoiding crossings amongst close relatives (Mwanja et al. 2015). Furthermore, both the Thai and the silver-YY strains utilized in our study have been imported recently to Tanzania (2016–2017)

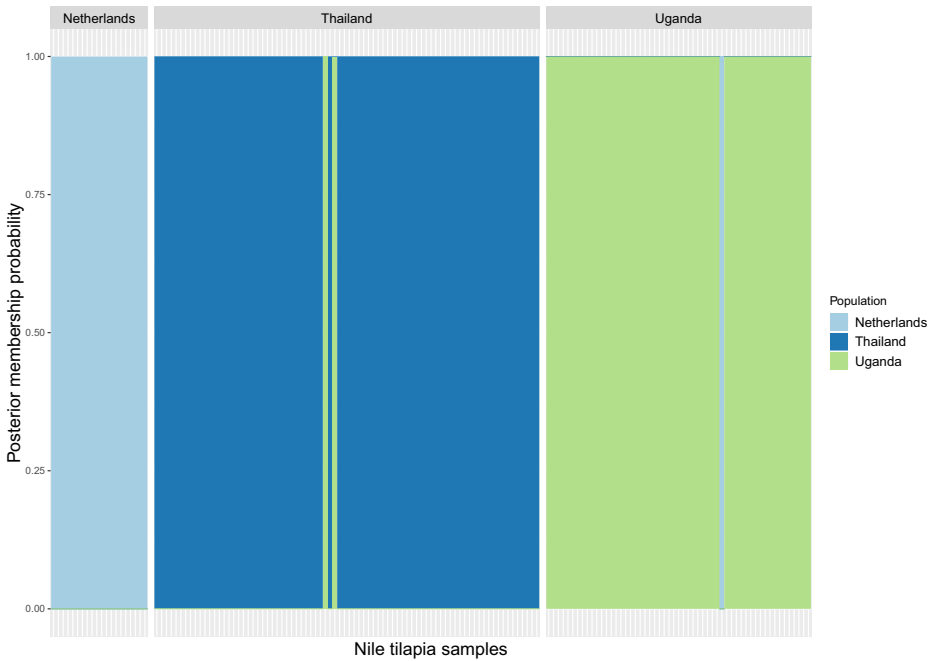


Fig. 4 Discriminant analysis of principal components (DAPC) of eight farmed populations of Nile tilapia *Oreochromis niloticus* strains introduced in Tanzania

indicating that crossing of close relatives has been most likely avoided. However, a controversy appears in our study based on the Fis estimates where the strains from Uganda appear to have a lower heterozygosity deficit compared with the Thai strains. Fis is negative in case of an

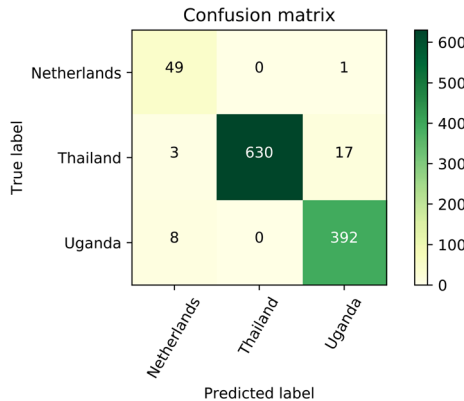


Fig. 5 Confusion matrix for prediction efficiency of the SNP dataset using cross validation. 4-fold cross validation was performed where 5 randomly chosen animals on each population were considered of unknown origin. The entire procedure was repeated 10 times in order to minimize potential bias due to sample allocation in the training/test datasets. The diagonal contains the number of correct population assignments for the overall sum of the cross-validation scheme. Off-diagonals contain the number of erroneously population allocations for each particular case. The confusion matrix is colored according to the number assignments on each populations as indicated by the color bar scale on the right

excess of heterozygosity and positive in the case of heterozygosity deficiency (Allendorf and Luikart 2007). Nevertheless, by construction, F_{IS} values depend on the absolute difference between observed and expected heterozygosity. As such, the F_{IS} values should be interpreted in connection with the relevant heterozygosity values, which in the case of the strains from Uganda are less than half compared with the Thai and the silver-YY strains.

Interestingly, the overall level of heterozygosity observed in our analysis is lower compared with prior tilapia studies (Fuerst et al. 2000; Bhassu et al. 2004; Hassanien and Gilbey 2005; Briñez et al. 2011; Baggio et al. 2016; Silva 2015). However, since the aforementioned studies utilized microsatellites, a direct comparison of heterozygosity magnitude may not be appropriate. Compared with other ddRAD studies in fish species, the Thai and the silver-YY strains appear to have heterozygosity values within the reported range (0.18–0.28; Saenz-Agudelo et al. 2015; Leitwein et al. 2016). Nevertheless, in the case of the Ugandan strains, the estimated genetic variation is particularly low. It has to be stressed that since most of the studied strains are descendants of small founder populations, genetic drift could be a main causative factor behind the low genetic diversity (Gilpin and Soulé 1986; Soulé 1987; Was and Wenne 2002; Sekino et al. 2002; Romana-Eguia et al. 2005; Aho et al. 2006; McKinna et al. 2010). Furthermore, particularly in the absence of pedigree records, severe reductions in genetic variation can occur due to crossings between close relatives (Westemeier et al. 1998; Madsen et al. 1999; Romana-Eguia et al. 2005; Aho et al. 2006; Appleyard and Ward 2006; Mbiru et al. 2015; Mapenzi and Mmochi 2016; Qin et al. 2007; Smallbone et al. 2016). In addition, the family structure of the sampled animals was unknown at the time of sampling, and this may affect estimates of genetic variation in a particular population.

Population structure analysis grouped all strains in 3 main clusters which corresponded to their countries of origin. Both STRUCTURE and DAPC can detect the underlying genetic structure of the tested animals, being able to suggest a most probable number of genetic clusters. In the case of STRUCTURE, the most probable number of clusters (K) is assigned by comparing the mean likelihoods of Bayesian-based models of a priori determined K . Regarding DAPC, the most probable K is derived through comparisons of BIC values. In our study, both approaches suggested that the most probable number of clusters is between 3 and 5, with the model selection criteria in both analysis suggesting a $K = 4$ as the most probable. Nevertheless, we need to point out that only a minute difference of the utilized model selection criteria was observed between K values ranging from 3 to 5, as opposed to when models of $K = 2$ were tested against $K = 3$ (File_S2, File_S3).

Regardless of the exact value of most probable K , our analysis suggested the existence of admixture between the Thai origin Chitralada strains and the Ugandan-derived strains. Furthermore, no gene flow was suggested between the GIFT or the BIG NIN strain and the Ugandan strains. As such, our results indicate that the detected similarity of the Chitralada and the Ugandan strains is not due to the common founder stock (African origin), but most probably due to recent admixture. Taking into account that the Thai strains were imported in Tanzania after 2016 and that stock exchange between fish farms in Tanzania is particularly common, we suggest that admixture probably occurred in recent generations. Interestingly, our data clustered one animal from the Ugandan strains (Muleba-M) with the silver-YY strain from Netherlands. However, since our analysis suggested that the silver-YY strain forms a uniquely isolated cluster with no apparent gene flow amongst any of the other tested strains, sample mislabeling could be the underlying reason for clustering an animal of putative Ugandan origin with the silver-YY strain.

Overall, our SNP dataset proved highly efficient in identifying the geographic origin of more than 97% of “putative” unknown samples. The ability to predict the population of origin is most valuable both for aquaculture and for conservation purposes of wild populations. Hybridization and introgression are fairly common in tilapias (Shirak et al. 2009; Wu and Yang 2012). Taking into account that Tanzania is a hot spot for wild cichlid populations, it is easily apparent that introgression with farmed strains could jeopardize the local adaptivity of the wild populations. As such, it is easily apparent that our approach could be also of value in conservation management of wild tilapia populations in Tanzania.

As part of an initiative to facilitate a national tilapia breeding program, the genomic resources of this study offer the required resolution for detecting the genetic diversity and differentiation of introduced tilapia strains in Tanzania. Importing tilapia strains is a valid option for the promotion of tilapia farming in Tanzania. Although GIFT and GIFT-derived strains showed significant impact to food security, rural income, and employment in both Asia and Africa (i.e., Egypt and Ghana) (ADB 2005; Hussain et al. 2013; Ansah et al. 2014), the argument to import the GIFT in Tanzania is still being debated in view of how best to conserve biodiversity and aquaculture germplasm for future use. Lind et al. (2012) suggested a zoned aquaculture system based on large water body catchments, while Mbiru et al. (2015) and Mapezi and Mmochi (2016) suggest the usage of all-male hybrid populations. However, since improvements using hybridization are done through trial and error (Chattopadhyay 2017), the outcomes are usually a reduced fingerling production due to potential reproductive incompatibilities between crossed species (Popma and Lovshin 1995).

In view of the aforementioned challenges, decisions related to importing strains like GIFT as a main seed source in tilapia farming should be informed by data originating from population genetics, growth performance, economic analysis, and potential biodiversity threats for the local wild populations. A thorough assessment of the “exotic” strain growth performance in comparison with the native ones will reveal the potential of each strain. Research on developing a new strain from the native germplasm for future use could be of utmost importance. Modern genomic tools like ddRAD-seq could play a pivotal role in assisting the formation of a Nile tilapia breeding nucleus of high genetic diversity, upon which future stock improvement will be based.

Author Contributions MM and FP carried out DNA extraction. CP and RH performed ddRAD library preparation and sequencing. DK, MM, and LC framed the study and contributed to designing the experiments. MM and CP performed the statistical and genetic analysis. MM wrote the manuscript. DK and CP revised the manuscript. All authors approved the final draft of the manuscript

Funding information Open access funding provided by Swedish University of Agricultural Sciences. We would like to acknowledge financial support from the Swedish International Development Agency (Sida) and BBSRC Institute Strategic Program Grants (BBS/E/D/20002172 and BBS/E/D/30002275) from Roslin Institute (University of Edinburgh). Edinburgh Genomics is partly supported through core grants from NERC (R8/H10/56), MRC (MR/K001744/1), and BBSRC (BB/J004243/1).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- ADB (Asian Development Bank) (2004) An impact evaluation of the development of genetically improved farmed tilapia and their dissemination in selected countries. The Asian Development Bank, Manila, Philippines 90
- ADB (Asian Development Bank) (2005) An impact evaluation of the development of genetically improved farmed tilapia and their dissemination in selected countries. The Asian Development Bank, Manila, Philippines 137
- Aho T, Rönn J, Piironen J, Björklund M (2006) Impacts of effective population size on genetic diversity in hatchery reared Brown trout (*Salmo trutta* L.) populations. *Aquaculture* 253:244–248
- Allendorf FW, Luikart G (2007) Conservation and the genetics of populations. Blackwell Publishing, Oxford UK
- Ansah YB, Frimpong EA, Hallerman EM (2014) Genetically improved tilapia strains in Africa: potential benefits and negative impacts. *Sustainability* 6(6):3697–3721
- Appleyard SA, Ward RD (2006) Genetic diversity and effective population size in mass selection lines of Pacific oyster. *Aquaculture* 254:148–159
- Attia TH, Saeed MA (2016) Next generation sequencing technologies: Next Generat Sequenc and Applic. S1: 006
- Azhar H, Ponzoni R, Nurhidayat K, Masazurah AR, Roslina AN (2004) Genetic selection of farmed tilapia: the performance of the 9th generation of the GIFT strain in different farm environments. *Malaysian Fish J* 3(2): 74–80
- Baggio RA, Ribeiro RO, Boeger WA (2016) Identifying Nile tilapia strains and their hybrids farmed in Brazil using microsatellite markers. *Pesq. agropec. bras. Brasília* 51(10):1744–1750
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3(10):e3376
- Barria A, Christensen KA, Yoshida GM, Correa K, Jedlicki A, Lhorente JP, Davidson WS, Yáñez JM (2018) Genomic predictions and genome-wide association study of resistance against *Piscirickettsia salmonis* in coho salmon (*Oncorhynchus kisutch*) using ddRAD sequencing. 8:41183–119
- Bhujel RC (2013) On-farm feed management practices for Nile tilapia (*Oreochromis niloticus*) in Thailand. In M.R. Hasan and M.B. New, eds. On-farm feeding and feed management in aquaculture. FAO Fisheries and Aquaculture Technical Paper No. 583. Rome, 159–189
- Bhassu S, Yusoff K, Panandam JM, Embong WK, Oyyan S, Tan SG (2004) The genetic structure of *Oreochromis* spp. (Tilapia) populations in Malaysia as revealed by microsatellite DNA analysis. *Biochemical Genetics* Vol. 42, Nos. 7/8
- Bolivar RB (1998) Estimation of response to within-family selection for growth in Nile tilapia (*Oreochromis niloticus*). PhD. Dissertation, Dalhousie University, Halifax, N.S., Canada 166
- Briñez BR, Caraballo XO, Salazar MV (2011) Genetic diversity of six populations of red hybrid tilapia, using microsatellites genetic markers. *Rev.MVZ. Córdoba* 16(2):2491–2498
- Brown JK, Taggart JB, Bekaert M, Wehner S, Palaiokostas C, Setiawan AN, Penman DJ (2016) Mapping the sex determination locus in the Hapuku (*Polyprion oxygeneios*) using ddRAD sequencing. *BMC Genomics* 17(1):448
- Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3-Genes Genomes Genet* 1:171–182
- Chattopadhyay NR (2017) Induced fish breeding: a practical guide for hatcheries. Elsevier 61-104
- Conte MA, Gammerding WJ, Bartie KL, Penman DJ, Kocher TD (2017) A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. *BMC Genomics* 18:341
- Damrongratana S, Kessanchai P (1966) *Tilapia niloticus* Linnaeus. Presentation to the His Majesty, the King of Thailand, March 17, 1966, Department of Fisheries. Bangkok, Thailand 29
- Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4:359–336
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567–1587
- FAO (Food and agriculture Organization) (2011) Fisheries and aquaculture technical paper. World aquaculture. Technical Paper No. 500/1. Rome 105
- Fuerst PA, Mwanja WW, Kaufman L (2000) The genetic history of the introduced Nile tilapia of lake Victoria (Uganda - E. Africa): the populations structure of *Oreochromis niloticus* (Pisces:Cichlidae) revealed by DNA microsatellite markers. In: International Symposium on Tilapia in Aquaculture, 5., 2000. Proceedings. Rio de Janeiro: Ista, 2000, 30–40

- Gilpin ME, Soulé ME (1986) Minimum viable populations: the processes of species extinctions. In conservation biology: the science of scarcity and diversity (ed. Soulé ME). Sinauer Associates, Sunderland, MA pp. 13–34
- Hassanien HA, Gilbey J (2005) Genetic diversity and differentiation of Nile tilapia (*Oreochromis niloticus*) revealed by DNA microsatellites. *Aquac Res* 36:1450–1457
- Hosoya S, Kiyoshi K, Hiroshi N, Junichi O, Kouichi S, Kou S, Keisuke M, Masaki Y, Atsushi JN, Akira K, Kenichi U, Tadahide K (2018) Assessment of genetic diversity in Coho salmon (*Oncorhynchus kisutch*) populations with no family records using ddRAD-seq. *BMC Res Notes* 11:548. <https://doi.org/10.1186/s13104-018-3663-4>
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour* 9:1322–1332
- Hussain MG, Barman BK, Karim M, Keus E (2013) Progress and the future for tilapia farming and seed production in Bangladesh. *Selected Papers, ISTA10.1059*, 8
- Jombart T (2008) *adeigenet: A R package for the multivariate analysis of genetic markers*. *Bioinformatics* 24:1403–1405
- Jombart T, Sébastien D, François B (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* 11:94
- Kaliba A, Osewe K, Senkondo E, Mnembuka B, Quagrainie K (2006) Economic analysis of Nile tilapia (*Oreochromis niloticus*) production in Tanzania. *J World Aquacult Soc* 37:464–473
- Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER (2013) The next-generation sequencing revolution and its impact on genomics. *Cell* 155:27–38
- Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour* 15:1179–1191
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359
- Leitwein M, Gagnaire PA, Desmarais E, Guendouz S, Rohmer M, Berrebi P, Guinand B (2016) Geonome-wide nucleotide diversity of hatchery-reared Atlantic and Mediterranean strains of brown trout *Salmo trutta* compared to wild Mediterranean populations. *J Fish Biol* 89:2717–2734
- Li SF, Cai WQ (2008) Contribution of genetic improved strains to Chinese tilapia industry. 8th International Symposium on Tilapia in Aquaculture 213–218
- Lind CE, Brummett RE, Ponzoni RW (2012) Exploitation and conservation of fish genetic resources in Africa: issues and priorities for aquaculture development and research. *Rev Aquac* 4:125–141
- Madsen T, Shine R, Olsson M, Wittzell H (1999) Restoration of an inbred adder population. *Nature* 402:34–35
- Mapenzi LL, Mmochi AJ (2016) Role of salinity on growth performance of *Oreochromis niloticus* ♀ and *Oreochromis urolepis urolepis* ♂ hybrids. *J Aquac Res Development* 7:431
- Mair GC (1997) Genetic manipulations for improved tilapia – technology adaptation and development II. Final Report 120
- Mardis ER (2008a) The impact of next-generation sequencing technology on genetics. *Trends Genet* 24:133–141
- Mardis ER (2008b) Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 9:387–402
- Maroso F, Hillen JEJ, Pardo BG, Gkagkavouzis K, Coscia I, Hermida M, Franch R, Hellemans B, Van Houdt J, Simionati B, Taggart JB, Nielsen EE, Maes G, Ciavaglia SA, Webster LMI, Volckaert FAM, Martinez P, Bargelloni L, Ogden R (2018) Performance and precision of double digestion RAD (ddRAD) genotyping in large multiplexed datasets of marine fish species. *Mar Genomics* 39:64–72.
- Mbiru M, Limbu SM, Chenyambuga SW, Lamtane HA, Tamatamah R, Madalla NA, Mwandya AW (2015) Comparative performance of mixed-sex and hormonal-sex-reversed Nile tilapia *Oreochromis niloticus* and hybrids (*Oreochromis niloticus* × *Oreochromis urolepis hornorum*) cultured in concrete tanks. *Aquac Int* 24:557–566
- McKinna EM, Nandlal S, Mather PB, Hurwood DA (2010) An investigation of the possible causes for the loss of productivity in genetically improved farmed tilapia strain in Fiji: inbreeding versus wild stock introgression. *Aquac Res* 41:e730–e742
- Mwanja M, Justus R, Constantine O, Richard D, Casius A (2015) Current fish hatchery practices in Uganda: the potential for future investment. *Int J Fish Aquat Stud* 2(4):224–232
- Oral M, Colléter J, Bekaert M, Taggart JB, Palaiokostas C, McAndrew BJ, Vandeputte M, Chatain B, Kuhl H, Reinhardt R, Peruzzi S, Penman DJ (2017) Gene-centromere mapping in meiotic gynogenetic European seabass. *BMC Genomics* 18:449
- Palaiokostas C, Bekaert M, Khan MG, Taggart JB, Gharbi K, Penman DJ (2015) Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) using RAD sequencing. *PLoS One*. 2013;8:e68389.
- Pembleton LW, Cogan NO, Forster JW (2013) StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Mol Ecol Resour* 13:946–952

- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for *De Novo* SNP discovery and genotyping in model and non-model species. *PLoS One* 7(5): e37135
- Popma TJ, Lovshin LL (1995) Worldwide prospects for commercial production of tilapia, Auburn University, Alabama 42
- Pritchard JK, Stephens M, Donnelly PJ (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Pukk L (2016) Analysis of molecular genetic and life history traits in Eurasian perch (*Perca fluviatilis* L.), PhD. Dissertation, Estonian University of Life Sciences, Tartu, Estonia 116
- Pullin RSV, Capili JB (1988) Genetic improvement of tilapias: problems and prospects, pp. 259–266. In Pullin RSV, Bhukaswan T, Tonguthai K, Maclean, JL (eds.). The Second International Symposium on Tilapia in Aquaculture ICLARM Conference Proceedings 15
- Qin Y, Liu X, Zhang H, Zhang G (2007) Effect of parental stock size on F1 genetic structure in the bay scallop, *Argopecten irradians* (Lamarck, 1819). *Aquac Res* 38:174–181
- R Core Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Recknagel H, Elmer KR, Meyer A (2013) A hybrid genetic link age map of two ecologically and morphologically divergent Midas cichlid fishes (*Amphilophus* spp.) obtained by massively parallel DNA sequencing (ddRADSeq). *G3 Genes Genomes. Genetics* 3:65–74
- Robledo D, Palaiokostas C, Bargelloni L, Martinez P, Houston R (2018) Applications of genotyping by sequencing in aquaculture breeding and genetics. *Rev Aquac* 10:670–682
- Rochette NC, Rivera-Colon AG, Catchen JM (2019) Stacks: analytical methods for paired-end sequencing improve RADseq-based population genomics. *bioRxiv*. <https://doi.org/10.1101/615385>
- Romana-Eguia MRR, Ikeda M, Basiao ZU, Taniguchi N (2004) Genetic diversity in farmed Asian Nile and red hybrid tilapia stocks evaluated from microsatellite and mitochondrial DNA analysis. *Aquaculture* 236:131–150
- Romana-Eguia MRR, Ikeda M, Basiao ZU, Taniguchi N (2005) Genetic changes during mass selection for growth in Nile tilapia *Oreochromis niloticus* (L.), assessed by microsatellites. *Aquac Res* 36:69–78
- Saenz-Agudelo P, Dibattista JD, Piatek MJ, Gaither MR, Harrison HB, Nanninga GB, Berumen ML (2015) Seascape genetics along environmental gradients in the Arabian Peninsula: insights from ddRAD sequencing of anemonefishes. *Mol Ecol* 24:6241–6255
- Sekino M, Hara M, Taniguchi N (2002) Loss of microsatellite and mitochondrial DNA variation in hatchery strains of Japanese flounder *Paralichthys olivaceus*. *Aquaculture* 213:101–122
- Shirak A, Cohen-Zinder M, Barroso RM, Serousi E, Ron M (2009) DNA barcoding of Israeli indigenous and introduced cichlids. *Isr. J. Aquac Bamidgeh* 61:83–88
- Siccha-Ramirez ZR, Francesco M, Belén GP, Carlos F, Paulino M (2018) SNP identification and validation on genomic DNA for studying genetic diversity in *Thunnus albacares* and *Scomberomorus brasiliensis* by combining RADseq and long read high throughput sequencing. *Fish Res* 198:189–194
- Silva MPKSK (2015) Genetic diversity of genetically improved farmed tilapia (GIFT) broodstocks in Sri Lanka. *Int J of Scientific Research and Innovative Tech* 2:3
- Smallbone W, Oosterhout C, Cable J (2016) The effects of inbreeding on disease susceptibility: Gyrodactylus Tumbulli infection of guppies, *Poecilia reticulata*. *Exp Parasitol* 167:32–37
- Soulé M ed. (1987) Viable populations for conservation. Cambridge Univ. Press, Cambridge, U.K
- Stobie CS, Carel JO, Michael JC, Paulette B (2018) Exploring the phylogeography of a hexaploid freshwater fish by RAD sequencing. *Ecol Evol* 8:2326–2342
- Syaifudin M, Bekaert M, Taggart JB, Bartie KL, Wehner S, Palaiokostas C, Khan MGQ, Selly SLC, Huluta G, D’Cotta H, Baroiller J-F, McAndrew BJ, Penman DJ (2019) Species-Specific Marker Discovery in Tilapia. *Sci Rep* 9:13001
- Tayamen MM (2004) Nationwide dissemination of GET-EXCEL tilapia in the Philippines. *Proceedings of the Sixth International Symposium on Tilapia in Aquaculture* 12–16
- Thodesen J, Rye M, Wang YX, Li SJ, Bentsen HB et al (2013) Genetic improvement of tilapias in China: genetic parameters and selection responses in growth, survival and external color traits of red tilapia (*Oreochromis* spp.) after four generations of multi-trait selection. *Aquaculture* 416–417:354–366
- Tony K, Jeffrey G, Fiona H, Elizabeth GB (2015) Low-cost ddRAD method of SNP discovery and genotyping applied to the periwinkle *Littorina saxatilis*. *J Molluscan Stud* 82:104–109
- Torati LS, John BT, Eduardo SV, Juliana A, Stefanie W et al (2019) Genetic diversity and structure in *Arapaima gigas* populations from Amazon and Araguaia-Tocantins river basins. *BMC Genet* 20:13
- URT (United Republic of Tanzania) (2014) Ministry of Livestock and Fisheries Development. Scoping exercise for prioritization (to achieve big results) in Tanzania, livestock and fisheries sector. Fisheries-Aquaculture sub-sector report, Dar es Salaam, Tanzania, ver 4

- URT (United Republic of Tanzania) (2016) The Tanzania fisheries sector, challenges and opportunities. Dar es Salaam, Tanzania
- URT (United Republic of Tanzania) (2017) Ministry of Agriculture, Livestock and Fisheries. Annual Report, Dodoma, Tanzania
- Was A, Wenne R (2002) Genetic differentiation in hatchery and wild sea trout (*Salmo trutta*) in the Southern Baltic at microsatellite loci. *Aquaculture* 204:493–506
- Westemeier RL, Brawn JD, Simpson SA, Esker TL, Jansen RW (1998) Tracking the long-term decline and recovery of an isolated population. *Science* 282:1695–1698
- Wu L, Yang J (2012) Identifications of captive and wild Tilapia species existing in Hawaii by mitochondrial DNA control region sequence. *PLoS One* 7:e51731
- Zhang J, Chiodini R, Badr A, Zhang G (2011) The impact of next-generation sequencing on genomics. *J Genet Genomics*
- Zimmermann S, Natividad JM (2004) Comparative pond performance evaluation of genoma supreme tilapia GST 1 and GST3 Groups. In Sixth International Symposium on Tilapia in Aquaculture 89

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

Affiliations

Mbiru Moses^{1,2} · **Matern S. P. Mtolera**¹ · **Leonard J. Chauka**¹ · **Fernando A. Lopes, Pinto**³ · **Dirk Jan de Koning**³ · **Ross D. Houston**⁴ · **Christos Palaiokostas**^{3,4}

¹ Institute of Marine Sciences, UDSM, Dar es Salaam, Tanzania

² Aquaculture Department, Ministry of Fisheries and Livestock, Dar es Salaam, Tanzania

³ Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, 750 07 Uppsala, Sweden

⁴ The Roslin Institute, University of Edinburgh, Edinburgh, UK