ISSN : 2350-0743

International Journal of Recent Advances in Multidisciplinary Research

Vol. 11, Issue 12, pp. 10520-10530, December, 2024

RESEARCH ARTICLE

TARGET GENE-CAPTURE STRATEGY FOR THE *CICHLID* SUBFAMILY *PSEUDOCRENILABRINAE*

Mojekwu, Tonna, O. [1,2], Cunningham, Michael, J. [1,4], and Hoareau, Thierry, B. [1,3]

[1]Department of Biochemistry, Genetics and Microbiology, University of Pretoria, X20, Hatfield 0028, Pretoria, South Africa; [2]Nigerian Institute for Oceanography and Marine Research, P.M.B. 12729, Victoria Island, Lagos State, Nigeria; [3]Boobook Ecology, Roma Qld, in the Southern Brigalow Belt Bioregion. Australia
[4]Present address: Reneco International Wildlife Consultants LLC, Sky Tower, Offices 3902 and 3903 - Al Reem Island P.O. Box 61741 - Abu Dhabi, United Arab Emirates

ARTICLE INFO

Article History
Received 20th September, 2024
Received in revised form
16th October, 2024
Accepted 27th November, 2024
Published online 29th December, 2024

Keywords:
Conservation, Exome,
Gene Capture,
Introgression,
Pseudocrenilabrinae, SNP.

Corresponding author:
Mojekwu, Tonna

ABSTRACT

The genome availability for cichlid species of the subfamily Pseudocrenilabrinae, including Oreochromis niloticus can help develop robust molecular tools to address questions related to the conservation and management of *O. niloticus* in its invasive range and *O. mossambicus* in its natural range. The objectives of this study include (1) identifying the best candidate orthologous genes among Pseudocrenilabrinae using gene ontology and a dedicated filtering strategy, (2) mining coding sequences for these candidate genes from *O. niloticus* reference genome, and (3) in-silico testing of the gene-capture custom panel for Pseudocrenilabrinae. Combining a literature survey and sequence mining from the Genbank database based on 14 ontologies, we identified 2,040 candidate genes. We excluded genes not annotated in the *O. niloticus* genome, without orthologue in the Human Genome, with exons smaller than 300 bp, that were on scaffolds that failed to map to a chromosome or presented gaps in the alignment. The coding sequence of remaining 247 genes was mined from the *O. niloticus* genome and transferability across available species of Pseudocrenilabrinae was successful (coverage ~99%). This panel represents an important molecular resource for the family and should open new perspectives for the assessment of introgressive hybridisation between *O. niloticus* and *O. mossambicus.*

Citation: Mojekwu, Tonna, O. Cunningham, Michael, J., and Hoareau, Thierry, B. 2024. "Target gene-capture strategy for the cichlid subfamily *Pseudocrenilabrinae", International Journal of Recent Advances in Multidisciplinary Research,* 11, (12), 10520-10530.

INTRODUCTION

The publication of five cichlid genomes within the subfamily Pseudocrenilabrinae (*Haplochromis (Astatotilapia) burtoni*, *Maylandia zebra*, *Neolamprologus brichardi*, *Oreochromis niloticus,* and *Pundamilia nyererei*; Brawand *et al.* 2014) make the development of low cost genomic molecular resource possible (Conte *et al.* 2017; Jaser *et al.* 2017; Lind *et al.* 2017). Besides, a growing number of studies have used available reference genomes to target candidate genes of interest in flatfishes, Nile tilapia and Rainbow trout (Van Bers *et al.* 2012; Xia *et al.* 2014; Neto *et al.* 2019; Mwaura *et al.* 2023; Atta *et al.* 2022). A large number of SNPs have already been detected in *O. niloticus*, including some belonging to candidate genes and genomic regions involved in pathways under selection (Van Bers *et al.* 2012; Palaiokostas *et al.* 2013; Xia *et al.* 2015; Ciezarek *et al.* 2022; Mwaura *et al.* 2023). SNP panels have also been developed in farmed *O. niloticus* for the determination of complex genetic traits (Yáñez *et al.* 2020; Wenne, 2023). It should then be possible to make use of these fully sequenced genomes of cichlid species to identify candidate SNPs in a related nonmodel species like *Oreochromis mossambicus*. Exon capture is an approach used to capture single-copy coding sequences across a range of species either wild or domesticated (Bi *et al.* 2012; Li *et al.* 2013; Hughes *et al.* 2021). It requires the available genomic resource of one of the species as a reference (Hodges *et al.* 2007). It has been used to investigate evolution in cichlids using phylogenomic approaches (Ilves and Lopez-Fernandez, 2014; Ilves *et al.* 2018; Jiamei *et al.* 2019), to identify candidate genes associated with diseases (Cosart *et al.* 2011), or to identify fitness traits in the wild (Cosart *et al.* 2011; Roffler *et al.* 2016).

It has been applied in phylogenetic and evolution studies on cichlids (Ilves and Lopez-Fernandez 2014; Song, *et al.* 2017). It is yet to be used to study introgressive hybridization, which is a major problem for *O. mossambicus* in the south-eastern region of Africa. It would be ideal to design an automated, inexpensive, and reliable molecular tool with a broad range of applications in cichlid breeding programs, especially when native species are under threat of genetic extinction (Bradbeer *et al.* 2019; Tibihika *et al.* 2020; Mojekwu *et al.* 2024). Using exome capture methods, the amount of genomic data (exon sequences) is smaller than when using data from full genomes, which makes it easier to manipulate and analyse. Moreover, there is higher confidence in the data as the coverage becomes very high with a smaller number of markers. Finally, as these regions express segments of the genome that code for proteins, exons are generally highly conserved, which highly reduces the chance of missing genotypes, even when the target species are phylogenetically distant from the focal species (Cosart *et al.* 2011; Mason *et al.* 2011). The aim of the study is to design an exome gene capture panel that would work within the cichlid subfamily Pseudocrenilabrinae, and help us address the question of introgressive hybridization between *O. mossambicus* and *O. niloticus*. To achieve this, we used the genomic resource available for the genus Oreochromis (*O. niloticus* reference genome, and *O. mossambicus* transcriptome data) to address three objectives: (1) identifying multiple candidate orthologous genes among Pseudocrenilabrinae using gene ontology search from literatures and mining of the NCBI gene database ; (2) mining coding sequences for these candidate genes from *O. niloticus* reference genome; 3) *in-silico* testing of the gene-capture custom panel for transferability of these markers across the cichlid subfamily Pseudocrenilabrinae . This genomic tool will help, among other things, with the detection of introgressed genes within the Pseudocrenilabrinae subfamily in the cichlid family.

MATERIALS AND METHODS

Identification and Data capture of potential target genes: Our initial approach was to identify potential target genes from the Genbank database and literature searches of previous studies. In the NCBI gene database, the gene ontology (GO) and functional search term used was (*Oreochromis niloticus* [orgn]) ("term") combined to GO search terms that included Carbohydrate metabolism, Cold, Disease, Drug resistance, Growth, Heat shock, Lipid metabolism, Protein metabolism, Salt, Sex, Stress. Other search terms in the database were included for Conserved single copy, Regulators and, Ubiquitous, sorted in our spreadsheet with the prefix "Z-" (Table 1). From the literature search, additional genes for comparability across vertebrates were added and tagged as "V-". These were comprised of genes previously studied in *O. mossambicus* (Cnaani *et al.* 2007; Hiroi *et al.* 2005) and genes previously used in vertebrate phylogenies (Brawand *et al.* 2014). The metadata recovered from the NCBI Gene database was as follows: Gene (name) abbreviation, Gene ID, description, GO keywords (used in searches) or comparative source (Orenil1.1), Chromosomal Scaffold, Sequence ID, coordinates & genomic span, Number of annotated orthologues (for the human equivalent), Number of exons, CDS length in first 15 exons, first 15 exons start and endpoints, lengths of each of the first 15 exons (Table S.1).

Filtering targets: Several steps were followed to obtain single copy loci with orthologues across vertebrates. These steps include: (a) Including only fully annotated genes from Orenil1.1 genome assembly (#PRJNA59571**)** with a verified human orthologue (i.e. different from the form LOCXXXX), with theexceptions of genes included for comparative reasons tested through Blast and confidently identified as single copy; (b) Excluding genes with fewer than 50 annotated vertebrate orthologues out of a total of 709 (available on NCBI as of [3rd December 2015]; (c) Excluded genes found on scaffolds that failed to map to a chromosome of Orenil1.1; (d) Excluded genes with smaller than 300 bp of CDS. The remaining candidate genes were then sorted by length of the first exon (Table S.1).

Conservation across vertebrate taxa: Danio, Xenopus, Gallus & Homo For the design of the BED file of variant sites and for identifying conserved regions to target (300 bp segments), tests to verify the conservation across cichlid fish and vertebrates were performed following three approaches: 1) BLAST algorithm, 2) Mapping to four cichlid fish genomes, 3) Mapping to four other vertebrate genomes. Target exons were BLASTed against the NCBI nucleotide database to check for errors in assembly coordinates and exon annotation. The results were saved in MEGA 7 (Kumar *et al.* 2017). Then aligned the target Orenil1.1 exon sequence against cichlid fish genomes and four other vertebrates (Danio, Xenopus, Gallus & Homo). All the target exon alignments were checked to verify their reading frame. Furthermore, the sequence coordinates on NCBI (nucleotide) were checked; with coordinates and alignment adjusted/extended/reduced as needed for coding. The condition for the reference sequence was the absence of gaps in alignments. If any alignments showed the presence of gaps in Orenil1.1, it was corrected for the reading frame when possible (to be inframe) and saved as <gene_Orenil Xbp gap.mas>. The Orenil1.1 gap codon were then deleted from the alignment but left any gaps in other species. This modified alignment was saved as <gene.mas>. The variable nucleotides (nVar), variable amino acids (AAvar), and sites with a gap (Gaps_bp) were recorded in the excel data file. All alignments were also saved in FASTA format <gene.fas>. We designed several procedures for checking alignments and genome coordinates of targets and specifying variants across the alignments (details in Table S.1).

Checking target annotations in other cichlid genomes: Initial attempts to align targets across vertebrates failed to identify conserved regions to target (300 bp segments). Subsequently, all target sequences were aligned across all five available cichlid genomes. The input made in NCBI (nucleotide) were, the coordinates for the target exon (small to large) and the coding strand (reverse complement or not) to check if the resulting sequence is annotated as a coding sequence. Then, blast by clicking the Run BLAST Under Analyse this Sequence on the right-hand side (RHS) to go to a new page, blast was against nucleotide nr/nt database (for all species or optionally Pseudocrenilabrinae to restrict hits to related organisms) using default parameters. Followed by a Search for sequences from the five-reference species

(Pseudocrenilabrinae) according to Brawand *et al*. (2014): *Pundamilia nyererei* (Lake Victoria), *Maylandia zebra* (Lake Malawi), *Astatotilapia* [=*Haplochromis*] *burtoni* (Riverine East African, Malagarasi, etc.), *Neolamprologus brichardi* (Lake Tanganyika), *Oreochromis niloticus* (Riverine – Nile/Niger systems) in phylogenetic order as successive sister groups. Where there are multiple RNA variants, select the best match (first on the list).

RESULTS AND DISCUSSION

Identification of target genes and CDSs: Our search based on databases and previous studies (Brawand *et al.* 2014; Cnaani *et al.* 2007; Hiroi *et al.* 2005) helped identify a total of 2,040 potential target genes, which represent 14 Ontologies (Table 1). These genes have been associated with specific adaptation and susceptibility (Cold (14), Disease (63), Drug resistance (17), Growth (487), Heat shock (160), Salt (7), Sex (27), Stress (49)), have been shown to be involved in various metabolic pathways (Carbohydrate metabolism (65), Lipid metabolism (129), Protein metabolism (476)), or have been linked to various functions and structure throughout the genome (Conserved single copy (81), Regulators (431), Ubiquitous (19)). Additional genes were included for comparability across vertebrates ranging from *O. mossambicus* (3) to vertebrate phylogenies (12). Filtering steps excluded 1,224 genes that were not fully annotated and without a human orthologue. We also excluded 107 genes that were unmapped to scaffolds and genes that were annotated in fewer than 50 vertebrate genomes. We finally excluded gene predictions found in the Orenil1.1 genome assembly but without a human orthologue (only fully annotated genes, not of the form LOCXXXX). This was done to overcome the issues of stop codon caused by poor genome annotation (Jiamei *et al.* 2019) and paralogue (Li *et al.* 2012) which can lead to difficulty in identifying open reading frame(s) (ORF). Exceptions involve slc12a2, sox14 genes studied in *O. mossambicus* and 12 genes for vertebrate phylogenies, included for comparative reasons but were all initially confirmed through BLAST search. This search excluded possible paralogues to confidently target single copy genes (Yuan *et al.* 2016). There were 260 target genes left after removing 449 genes that had any exon with a CDS below 300bp (CDS <300 bp). Exceptions concern two target genes with short exon lengths (foxred1 and itbp3 with exon lengths 255 and 273, respectively) that were kept to compensate for the low representation of target genes on chromosome LG10 (Table 1). The final 247 gene targets were obtained after removing 13 genes with sites that were highly variable > 6%, noncoding, or with the presence of gaps. The percentage distribution of target genes on the chromosomes of the reference genome of *Oreochromis niloticus* showed the presence of more than one gene on each chromosome (Figure 1). Majority of the target genes (67%) were distributed to 15 chromosomes while the remainder (33%) were distributed among nine chromosomes. The selected 247 genes were represented with an ideogram which showed that the number of genes ranged from three for LG3 to 21 for LG7 (Figure 2). Targeted genes for disease resistance/susceptibility were found on eight chromosomes while genes for drug resistance were only present on LG10 and, LG5 (Table 2). Genes for growth were present on all the chromosomes except LG3 and LG8.

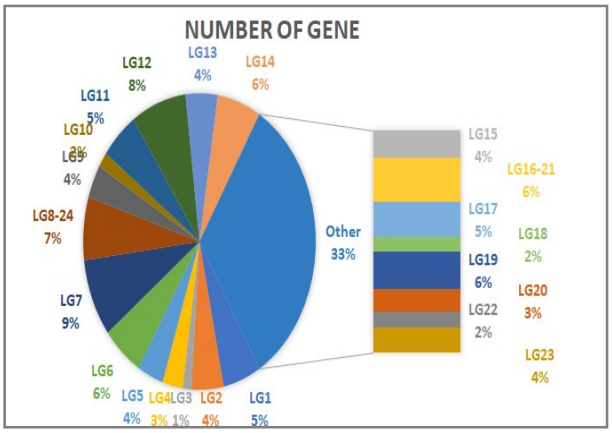

Figure 1. The percentage distribution of target genes on the chromosomes of the reference genome of *Oreochromis niloticus*. Majority of the target genes (67%) were distributed to 15 chromosomes while the remainder (33%) were distributed among nine chromosomes.LG7 (9%) has the highest percentage of the target genes while LG3 has the least number of target genes

However, some of the genes for growth also represents other ontologies for Conserved single copy), Disease, Regulator, Sex determination, Stress and vertebrate phylogeny (Table 2). Some of the genes represents more than one ontologies as the number of genes associated with each ontology is indicated on Table 2; Heat shock (23), Protein metabolism (35), Protein metabolism/Carbohydrate (11), Protein metabolism /Lipid (16), Protein metabolism /Regulator (2), Salt tolerance (2), Sex determination (10), Stress (8), genes found in *O.mossambicus* (sox14 and slc12a2) for comparability, Vertebrate Phylogeny (8), Conserved single copy (7), Regulator (12), Ubiquitous (7), and Ubiquitous/ Regulator (11).

The study by Jiamei *et al*. (2019) suggested that more appropriate annotation of the tilapia genome should be performed to ascertain if targeted exons are coding regions or not. To only keep guenuine CDS in thepresent panel, we only selected genes that were fully annotated with more than 50 annotated vertebrate orthologues. This has also helped us 1) checking if the CDSs are inframe, 2) verify the coordinates of the targeted exons in NCBI (See Table 3 protocols), and 3) eliminate selected sequences that are untranslated or duplicated. Paralogue genes could be erroneously used as orthologs, especially with phylogenomic data, and there is no set method to validate loci orthology assembled from Next Generation Sequence (NGS) (McCormack, *et al.* 2013; Chakrabarty, *et al.* 2017). However, recent studies were able to eliminate potential paralogues by finding the best reciprocal hits between assembled contigs and a reference genome of *Oreochromis niloticus* with the Perl script reblast.pl. (Yuan *et al.* 2016; Atta *et al.* 2022).

Conservation across vertebrates: BLAST searches from the Orenil 1.1 query often failed to find matches in Danio, Xenopus, Gallus, and Homo genomes. Our primary reference genome was *Oreochromis niloticus* (Cichlidae: Pseudocrenilabrinae) Orenil1.1 (GCA_000188235.2 from Broad Institute, USA). After filtering, a single exon coding sequence (each ~300 bp) was selected from each of the 247 genes. The selected exons must be atleast 300bp upwards, with exceptions of two genes foxred1 (255bp) and ltbp3 (273bp) included to increase the coverage of LG10.

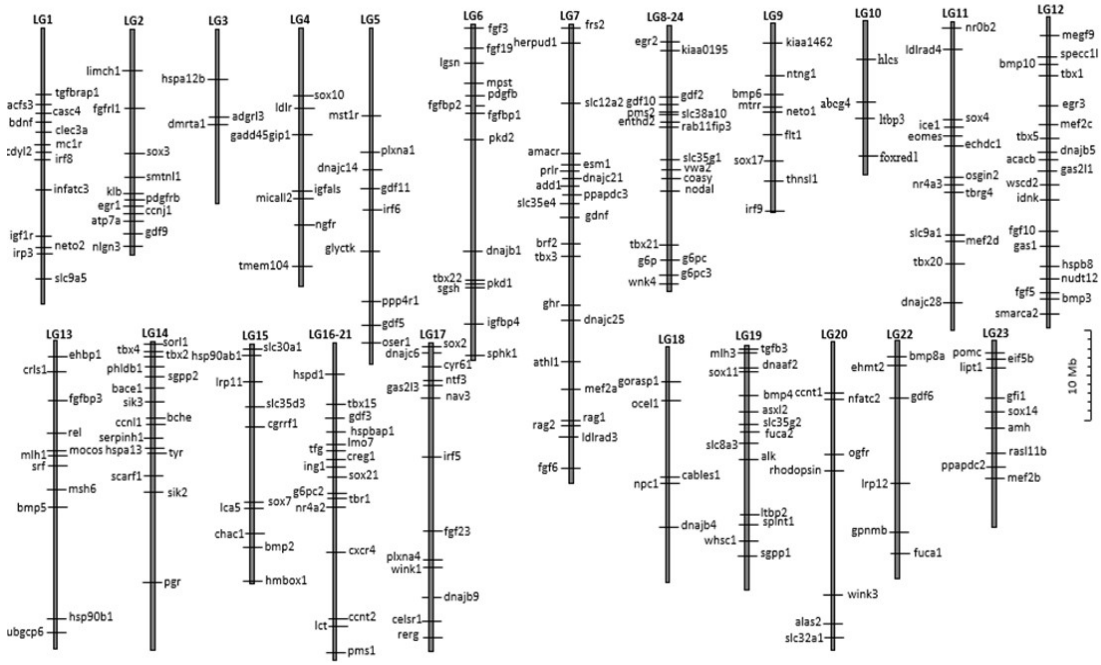

Figure 2. Ideogram illustrating the distribution of the 247 selected target genes on the 23 chromosomes of O. niloticus

Figure 3 (a). Alignment of the sequences of .gene abcg4 (Drug resistance) LG10, from *Oreochromis niloticus* with the four reference species (Pseudocrenilabrinae; after Brawand *et al.* 2014). (b).Translation of the gene into protein-coding sequences to verify that they are in-frame and without gaps

Table 1. Potential target gene ontologies captured in the *O. niloticus* genome (#PRJNA59571). "Z-" is a prefix used for other search terms. V- is a tag used for genes added for comparability across vertebrates

S/N	Gene ontology	Initial Number of genes checked	Single copy Genes with orthologues across vertebrate	Unplaced genes (UNK)	Genes with an exon > 300 bp CDS exceptions to 2 genes in LG10 after filtering	Removed genes with high variable sites >6%, noncoding and gaps	Final Targets captured
1	Carbohydrate metabolism	65	34	5			
2	Cold tolerance	14	9	2			
3	Disease	63	26	0	11		11
4	Drug resistance	17	6	1	2		2
5	Growth	487	191	22	83	4	79
6	Heat shock	160	88	17	25	2	23
7	Lipid metabolism	129	60	6			
8	Protein metabolism	476	242	33	67	3	64
9	Salt tolerance	7	2	0	2		2
10	Sex determination	27	17	2	12	2	10
11	stress	49	19	3	8		8
12	Z- Conserved single copy	81	35	6	8	1	7
13	Z- Regulators	431	59	7	14	1	13
14	Z- Ubiquitous	19	28	3	18		18
15	V- *O. mossambicus*	3			2		2
16	V- Vertebrate phylogeny	12			8		8
Total		2040	816	107	260	13	247

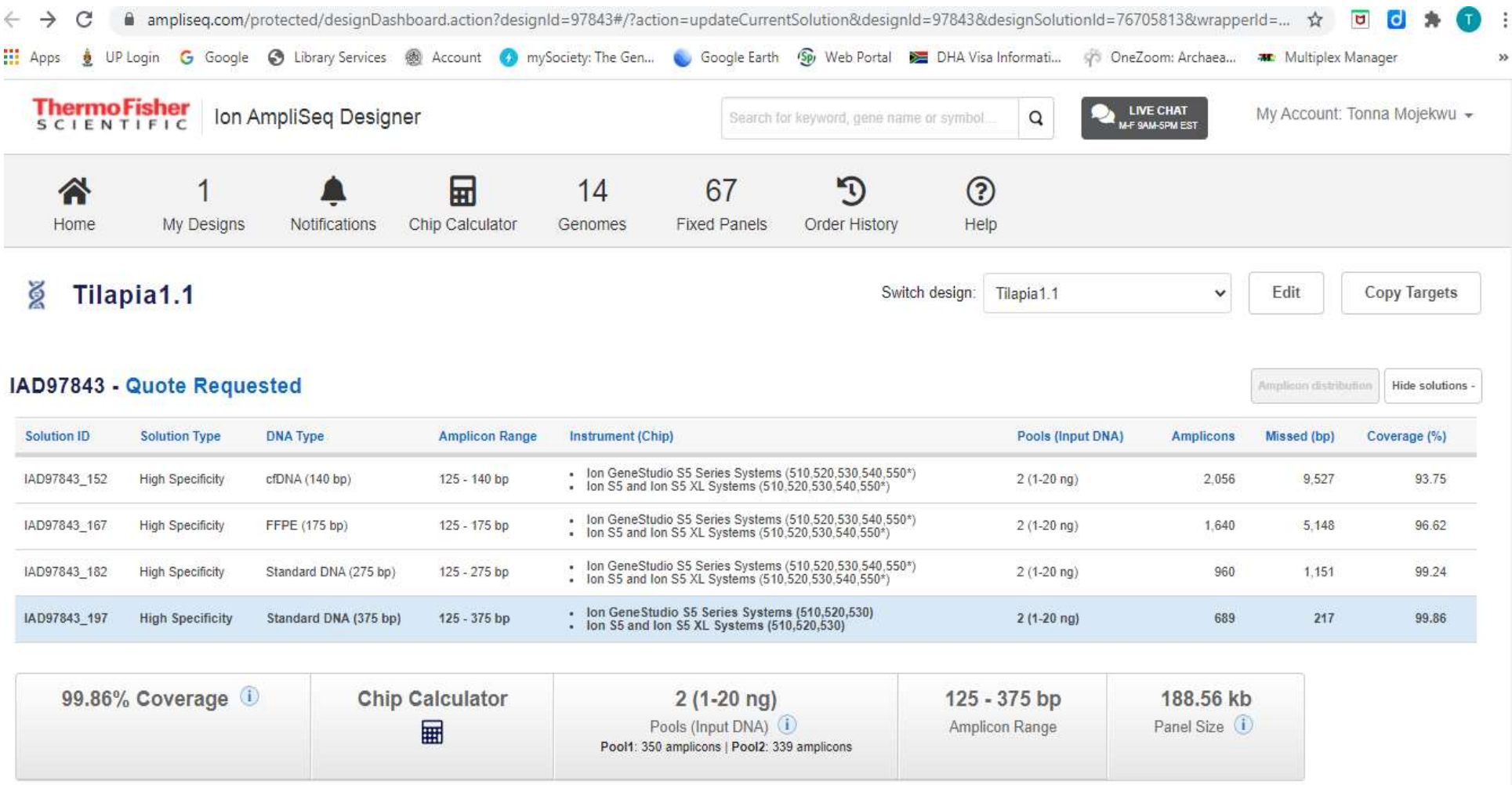

Figure 4. Output example of the Ion AmpliSeq Designer website. We illustrate various amplicons for which we selected high specificity, standard DNA type with the highest coverage of 99.86% for the solution ID IAD97843_197 (highlighted in blue)

Table 2. The number of targets genes captured on each chromosome in the *O. niloticus* genome and gene names withmore than one ontology, details at supplementary Table S.1.

S/N	Gene ontology	Number of target Genes	Chromosomes	Gene names with more than one ontology*
1	Disease resistance/susceptibility	11	LG6, LG8-24, LG9, LG11, LG14, LG15, LG18 and LG22	Disease (lrp11) *, Growth(whsc1) *, ice1,bace1,phldb1,Ica5,npc1,gpnmb,pkd2,pkd, rab11fip3, and kiaa1462
2	Drug resistance	2	LG10 and,LG5	abcg4 and ppp4r1
3	Growth	79	Present in all except LG3 and LG8	Conserved single copy (gas2l1, gas2l3) *, Disease (whsc1) * Regulator (nr4a2) * Sex determination (amh) * Stress(osgin2) *, Vertebrate phylogeny(prlr, tyr) *,
4	Heat shock	23	Present in all except LG1, LG2, LG4, LG9, LG10, LG22, LG23	dnajc28,hspb8, dnajb5,hsp90b1,mlh1,hspa13,serpinh1,hsp90ab1,hspd1,hspbap1,pms1,dnajb9,dnajc6,dnajb4,gorasp1,mlh3,dnajc22,hspa12b,dnajc14, dnajb1,dnajc25,dnajc21,pms2,
5	Protein metabolism (35) /Carbohydrate /Lipids /Regulator	64	Present in all except LG17	Carbohydrate (11) *, Lipids (16) *, Regulator (2)(srf andmef2a) *
6	Salt tolerance	2	LG14	sik2 and sik3
7	Sex determination	10	LG2, LG3, LG4, LG9,LG11,LG15, LG16-21, LG17, LG19	sox4,sox7,sox21,sox2,sox11,asxl2,sox3,dmrta1,sox10,sox17
8	Stress	8	LG5, LG7, LG8-24, LG12, LG13, LG17, LG19, LG20	Heat shock (dnaaf2)* wscd2,rel,wink1,wink3,oser1,herpud 1,wnk4,
9	Conserved single copy	7	LG2, LG4, LG12, LG13,LG16-21, LG17	specc1l,ehbp1,lmo7,nav3,smtnl1,limch1,micall2
10	Regulators	13	LG2, LG7, LG11, LG12, LG14, LG15, LG16-21, LG18, LG20, LG22, LG23	nr0b2,mef2d,mef2c,pgr,ccnl1,hmbox1,ccnt2,cables1,ccnjl,ccnt1,ehmt2,mef2b,brf2
11	Ubiquitous	18	LG1, LG5, LG6, LG7, LG9, LG11, LG12, LG14, LG16-21, LG17, LG20, LG8-24,	Regulator (eomes,tbx20,tbx1,tbx5,tbx2,tbx4,tbr1,tbx15,tbx22,tbx3,tbx21) *, infatc3,irf8,creg1,irf5,nfatc2,irf6,irf9
12	*O. mossambicus*	2	LG7 and LG23	sox14 and slc12a2
13	Vertebrate phylogeny	8	LG1, LG16-21, LG7, LG19, LG20, LG23, LG8-24	mc1r,cxcr4,slc8a3,rhodopsin,pomc,rag1, rag2,g6p
Total		247		

Table 3. Protocol for checking alignments and genome coordinates of targets

A	**Finding the Reference Genome sequence** 1.Record the coordinate of the target gene in the *O. niloticus* reference genome in NCBI. 2.Record the coordinate of the nucleotide sequences of the CDS in the FASTA format for the target exons, ensuring its functionality. Save this fasta sequence as. fas or. mas.
B	**Check the translation frame of the target fragment and update coordinates** 1.Open the saved NCBI fasta sequence in MEGA alignment explorer, translate to Protein Sequences using the Standard Genetic Code Table, confirm if it is in Frame (in a frame: the first triplet is the first codon) and if the initial or final AA sequences positions are not ambiguous (or had?), otherwise correct the target exon coordinates for the frame. 2.After correcting the selected region coordinates, check that the translated sequence matches that in the GenBank record and correct the exon boundaries in the record datasheet (decrease start coordinates CDS#a, and/or increase end coordinate CDS#z).
C	**Finding homologues in selected relatives with BLAST** 1.BLAST against NCBI Genomes database (all species or optionally Pseudocrenilabrinae to restrict hits to related organisms) using default parameters. Harvest hits to the selected comparative genomes and If there are multiple hits for a species, choose only the first of these (the best match, lowest E value) as this is evidence for gene duplication. Copy each FASTA format sequence and paste these as new sequences in the MEGA alignment. 2.If no hit was found on one or more of the target taxa, BLAST against the Refseq genomic or Change the Database to nucleotide collection (nr/nt). There will often be multiple transcript variants; choose the first of these from the target species (best match).
D	**Searching directly for genomic sequences of this CDS in a related species** 1. Alternatively, go directly for the genomic sequence of this gene (if it is annotated), check the first CDS feature and calculate whether the exon length roughly matches our reference target in MEGA and adjusts the coordinates to match ends. 2. Copy the FASTA sequence and paste this into our alignment panel.
E	**Alignment of related sequences in MEGA** 1.For closely related sequences initial alignment is best done by eye or use either Clustal or Muscle algorithms. Translate the sequence without any gaps and are in a frame. 2.Save the alignment session in MEGA as <genename.mas> (e.g. "dnajb1.mas") and export in a FASTA format as <genename.mas> (e.g. "dnajb1.mas").
F	**Capturing data on variance across the alignment** 1.Record the number of variable nucleotide sites, gap sites and variable AA sites from the MEGA data explorer.
	I have now finished checking this gene! Hence this was done for all the selected genes targets.This was the protocol used in this chapter for alignment, nucleotide database check for errors in assembly, coordinates, exon annotation and variance alignment.

Hence, the selected exons ranged from 255-2571bp, distributed across the primary reference genome, and representing a total of around 154,335bp (Table S.1). The target sequences could not be recovered from these study models, making it difficult to align sequences of the study models with the *Oreochromis niloticus* genome to verify the utility of the marker across vertebrates. This is surprising since designed primers from a single reference species have been successful in capturing sequences that are orthologs across divergent sets of species (Hedtke *et al.* 2013; Ilves *et al.* 2018; Jiamei *et al.* 2019; Atta *et al.* 2022).

Conservation among cichlids: For the comparative genomic approach, the reference genome of several cichlid species of the subfamily Pseudocrenilabrinae have been used. These includes the species *Neolamprologus brichardi* (NeoBri1.0; GCA_000239395.1), *Haplochromis (Astatotilapia) burtoni* (AstBur1.0; GCA_000239415.1), *Pundamilia nyererei* (PunNye1.0; GCA_000239375.1), *Maylandia zebra* (M_zebra_UMD1; GCA_000238955.3), all generated in a previous study on the radiation of African cichlids (Brawand *et al.* 2014). All targets were highly conserved across the five reference genomes of cichlids and the fraction of sites that are variable across these exons (average variable sites) were 3.3%. For each of the five species and 247 targeted genes, the FASTA file was downloaded, combined, and aligned in MEGA 7 (Figure 3; Table S.2). Targets were selected from the conserved region across five cichlid genomes to avoid data from non-coding flanking regions and problems of alignment difficulty, which probably increase the chance of success of the panel design. The further away the sequences are from the conserved core region, the higher the polymorphism (Faircloth *et al.* 2012). This conservative procedure also addresses some issues raised when using whole-genome sequencing data such as collecting and processing SNPs data that occur in noncoding regions (Brown and Lemmon 2007; Fan *et al.* 2011; Jiamei *et al.* 2019).

Custom Panel Design: The panel was designed using the Ampliseq website www.lifetechnologies.com/ ampliseqcustom) by uploading the reference genome Orenil1.1 together with the list of specified target exons including 500bp flanking regions and a BED file. The BED file is the list of exons with the start (ExonTa), and end coordinates (ExonTz). From the output the best amplicon panel provided (IAD97843_197) had high specificity standard DNA solutions a panel size of 188.56 kb, covering 99.86% of the target sequencesand divided into 689 amplicons. These would be split into two pools (~350 & 339) and with standard DNA amplicon ranging between 125-375 bp (Figure 4).The amplicons was checked using in-silico PCR available from UCSC websites (https://genome.ucsc.edu/) to test the design against a range of customs and our standard reference genome. This was done to test the transferability of the markers across other families of cichlid but was only successful for the Pseudocrenilabrinae. Steps for the captured data on variance across the alignment are in the protocol (Table 3) and saved on the Spreadsheet (Table S.1), with sequence data in Table S.2.

Application for the conservation of *O. mossambicus* in its natural habitats: The *Oreochromis niloticus* genome, a major invasive species involved in the threat on *Oreochromis mossambicus* due to aquaculture utilization represents an excellent resource (Marr *et al.,* 2017; Bills, 2019). The existence of this genome allowed the accurate mining of coding sequences with exact exon coordinates of genes with known functions (Brawand *et al.* 2014). Selecting multilocus genes from its genome will be more efficient in detecting the presence of those genes in other species and address introgressive hybridization (Ciezarek*et al.* 2022). SNPs have been used to discriminate between tilapia species and analyse hybridization. Several studies have shown that SNPs has been applied to detect hybrids, capture the genetic diversity in diverse populations of both wild and aquaculture Oreochromis species (Joshi *et al.* 2018; Syaifudin *et al.* 2019; Peñaloza *et*

al. 2020; Yáñez *et al.* 2020). However, some of the techniques like RAD and SNP arrays, applied in these studies are either showing under-representation of heterozygotes (Davey *et al.* 2013) or limited comparative applicability across with cumbersome SNP selection criteria (Davey *et al.* 2013; Robledo *et al.* 2018). Recent study using genome wide SNPs established hybridization between farm and native species of *Oreochromis* in Tanzania (Ciezarek*et al.* 2022). Though, the study was limited to a country's geographical location as some of the genus diversity from other native areas such as the *O. mossambicus* was not captured on the SNP panel. The 247 genes captured in this study will be a molecular resource in addressing introgressed genes of *Oreochromis niloticus* into the *O. mossambicus* in South Africa where the former has been invasive (D'Amato*et al.* 2007; Moralee *et al.* 2000;Deines *et al.* 2014; Mashaphu *et al.* 2024; Mojekwu *et al.* 2024). Our panel targeted only a single copy of fully annotated genes with more than 50 annotated vertebrate orthologues and checked if the CDS of the exons captured are in the right coordinates. Eliminating selecting sequences that are paralogues, which could be mistakenly used as orthologs since there was no set method to validate loci orthology assembled from Next Generation Sequence (NGS) at the start of this project (McCormack, *et al.* 2013; Chakrabarty, *et al.* 2017). However, current studies have applied bioinformatics pipelines to address the challenges of paralogues (Yuan, *et al*. 2019; Hughes, *et al.* 2021; Ciezarek *et al.* 2022). Secondly, our targets were selected from the conserved region across five cichlid genomes to avoid data from non-coding flanking regions giving more credence to the panel designed. This conservative procedure also addresses some issues raised when using whole-genome sequencing data such as collecting and processing SNPs data that occur in noncoding regions (Brown and Lemmon 2007; Fan *et al.* 2011; Jiamei *et al.* 2019). Perhaps, our targets can be applied to other *Oreochromis* species and broadly across other cichlids.The captured targets will also serve as a good tool for more panel design to detect *O. niloticus* genes in both wild and cultured stocks. Hence, the determination of diversity across wild stocks and safeguarding pure refugial areas will be of importance to ecologists, conservation biologists, and the aquaculture industry. There is still a lot to do but this panel produced synthetic amplicons with high coverage and specificity, that will give improved, efficient results using NGS (Figure 4).

Data Availability: The datasets generated are provided in the supplementary material and on request (for Table S.2.) using the link https://www.researchgate.net/publication/358923535

Ethics and permit approval: Not Applicable

Funding statement: Open access funding enabled withindividual financial contributions.

Conflict of interest disclosure: The authors declare that there is no confict of interest.

Permission to reproduce material from other sources: Not applicable.

ACKNOWLEGDEMENT

We appreciate the support, time, bench space and encouragement, from the University of Pretoria, Nigerian Institute For Oceanography and Marine Research, my family, Mrs Ifeoma Lilian Umeokafor and Dr. Emeka Obumneme Okoro.

AUTHOR CONTRIBUTIONS

Mojekwu TO: literature search, mining gene database, evaluation among different groups of vertebrates, panel design and write-up. Cunningham M: conceived and supervise the work. Hoareau T :Completed the work supervision and write-up

REFERENCES

AttaCJ, Yuan H, Li C, Arcila D, Betancur-R R, Hughes LC, Ortí G, Tornabene L. 2022 Exon-capture data and locus screening provide new insights into the phylogeny of flatfishes (*Pleuronectoidei*). *Molecular Phylgenetic and Evolution* 166: 1055-7903. doi: 10.1016/j.y mpev.2021.107315.

Azaiez A, Pavy N, Gérardi S, Laroche J, Boyle B, Gagnon F, Mottet M-J, Beaulieu J, Bousquet J. 2018. A catalog of annotated high-confidence SNPs from exome capture and sequencing reveals highly polymorphic genes in Norway spruce (*Picea abies*). *BMC Genomics* 19: 942

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS One*3: e3376.

Bi K, VanderpoolD, Singhal S, LinderothT, Moritz, C, Good, JM. 2012. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales.*BMC Genonomics*13:403.

Bills R. 2019. *Oreochromis mossambicus*. The *IUCN Red List of Threatened Species* version 2019: e. T63338A3124798. Available at https://dx.doi.org/ 10. 2305/ IUCN.UK.2019-3. RLTS. T63338A3124798.en. (last accessed 5 May 2020).

Blåhed I-M, Königsson H, Ericsson G, Spong G. 2018. Discovery of SNPs for individual identification by reduced representation sequencing of moose (Alces alces). *PLoS ONE* 13:e0197364. https://doi.org/10.1371/ journal.pone. 0197364

Bradbeer SJ, Harrington J, Watson H, Warraich A, Shechonge S, Smith A, Tamatamah R, Ngatunga BP, Turner GF, Genner MJ. 2019. Limited hybridization between introduced and critically endangered indigenous tilapia fishes in northern Tanzania. *Hydrobiologia* 832:257-268. https://doi.org/10.1007/s10750-018-3572-5.

Brawand D, WagnerCE, Li YI, Malinsky M, KellerI, FanS *et al*. 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* 513: 375–381. https://doi.org/10.1038/nature13726.

Brooks A, Creighton EK, Gandolfi B, Khan R, Grahn RA, Lyons LA. 2016.SNP Miniplexes for Individual Identification of Random-Bred Domestic *Cats Journal of Forensic Science.* 61:594–606.https://doi.org/10.1111/ 1556- 4029. 13026

Brown JM, Lemmon AR. 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Systems Biology* 56: 643-655.

Cai L, HouM, Xu C, Xia Z, Wang J. 2020. Development of Novel SNP Assays for Genetic Analysis of Rare Minnow

(*Gobiocypris rarus*) in a Successive Generation Closed Colony. *Diversity* 12: 483. doi:10.3390/d12120483.

Cnaani A, Lee BY, Ozouf-Costaz C, Bonillo C, Baroiller JF, D'Cotta H, Kocher T. 2007. Mapping of sox2 and sox14 in tilapia (*Oreochromis spp*.).*SexualDevelopment*1:207–210. https://doi.org/10.1159/000102109.

Chakrabarty P, FairclothBC, Alda F, Ludt WB, McMahan CD, Near TJ, Dornburg A, Albert JS, Arroyave J, Stiassny ML. 2017. Phylogenomic Systematics of Ostariophysan fishes: Ultraconserved Elements Support the Surprising Non-monophyly of Characiformes. *SystemBiology* 66: 881–895.

Ciezarek *et al*., 2022.Whole genome resequencing data enables a targeted SNP panel for conservation and aquaculture of Oreochromis cichlid fihes. *Aquaculture* 548 – 737637.

Conte AM, Gammerdinger JW, Bartie KL, Penman JD, Kocher T. 2017. A high-quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. *BMC Genomics*18: 341.https://doi.org/10.1186/s12864-017-3723-5

Cosart T, Beja-Pereira A, Chen S, Sarah B, Shendure J, Luikart G. 2011. Exome-wide DNA capture and next-generation sequencing in domestic and wild species. *BioMed Central Genomics*12 347. https://doi.org/ 10.1186/1471-2164-12-347

D'Amato ME, Esterhuyse MM, van der Waal BCW, BrinkD,Volckaer,P. 2007.Hybridization and phylogeography of the Mozambique tilapia *Oreochromis mossambicus* in southern Africa evidenced by mitochondrial and microsatellite DNA genotyping. *Conservation Genetics* 8: 475-488.

Davey JW. *et al*. 2013.Special features of RAD Sequencing data: implications for genotyping. Mol Ecol 22: 3151–64.

Deines AM, Bbole I, Katongo C, Feder JL,Lodge DM. 2014. Hybridisation between native Oreochromis species and introduced Nile tilapia *O. niloticus* in the Kafue River, Zambia. Afri *Journal of Aquatic Science*1–12. http://dx.doi.org/10.2989/16085914. 2013.864965.

Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT,Glenn TC. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *System Biology* 61: 717–726.

Fan Y, Wu R, Chen MH, Kuo L,Lewis PO. 2011. Choosing among partition models in Bayesian phylogenetics. *Molecular BiologyEvolution*28: 523–532.

Galeano E, Bousquet J, Thomas BR. 2021. SNP-based analysis reveals unexpected features of genetic diversity, parental contributions and pollen contamination in a white spruce breeding program. *Scientific Reports* 11: 4990 https://doi.org/10.1038/s41598-021-84566-2

Gutierrez AP, Turne, F, GharbiK, Talbot R, Lowe NR, Penaloza C, McCullough M, Prodohl PA, Bean TP, Houston RD. 2017. Development of a medium density combined-species SNP array for Pacific and european oysters(*Crassostreagigas*and*Ostreaedulis*).*G3-Genes,Genomes,Gentics* 7:2209–2218. https://doi.org/ 10.1534/g3.117.041780.

Hamilton MG, Mekkawy W, Kilian A, Benzie JAH. 2019. Single Nucleotide Polymorphisms (SNPs) Reveal Sibship Among Founders of a Bangladeshi Rohu (*Labeo rohita*) Breeding Population. *Frontiers in Genetics* 10: 597. doi: 10.3389/fgene.2019.00597

Hedtke SM, Morgan MJ, Cannatella DC, Hillis DM. 2013. Targeted enrichment: maximizing orthologous gene comparisons across deep evolutionary time. *PloS one* 8: e67908. https://doi.org/10.1371/journal.pone.0067908

Hiroi J, McCormickSD, Ohtani-Kaneko R, Kaneko T. 2005. Functional classification of mitochondrion-rich cells in euryhaline *Mozambique tilapia* (*Oreochromis mossambicus*) embryos, by means of triple immunofluorescence staining for Na+/K+-ATPase, Na+/K+/2Cl- cotransporter and CFTR anion channel. *The Journal of experimental biology* 208: 2023–2036. https://doi.org/10.1242/jeb.01611

Hodges E, Xuan Z, Balija V, Kramer M, Molla M, Smith S, Middle C, Rodesch M, AlbertT, Hannon G, McCombieWR. 2007. Genome-wide in situ exon capture for selective resequencing. *NatureGenetics*39:1522–1527. https://doi.org/10.1038/ng.2007.42.

Houston RD, Taggart JB, Cezard T, Bekaert M, Lowe NR, Downing A,Talbot R, Bishop SC, Archibald AL, Bron JE, Penman DJ, Davassi A, Brew F, Tinch AE, Gharbi K,Hamilton A. 2014. Development and validation of a high density SNP genotyping array for Atlantic salmon (Salmo salar). BioMed Central Genome 15: Doi.org/ 10.1186/1471-2164-15-90.

Howe GT, Jayawickrama K, Kolpak SE, Kling J, Trappe M, Hipkins V,Ye T,Guida S, Cronn R, Cushman SA *et al*. 2020. An Axiom SNP genotyping array for Douglas-fir. *BMC Genomics* 21: 9.

Hughes LC, Ortí G, Saad H, Li C, WhiteWT, Baldwin CC, Crandall KA, Arcila D, Betancur-RR. 2021. Exon probe sets and bioinformatics pipelines for all levels of fish phylogenomics. *Molecular Ecology Resource* 21: 816–833.

Hyun DY, Raveendar S, Lee KJ, Lee GA, Myoung JS, Kim SH, Lee JR,Cho GT. 2020. Genotyping-by-Sequencing Derived Single Nucleotide Polymorphisms Provide the First Well-Resolved Phylogeny for the Genus *Triticum* (Poaceae). *Frontiers in Plant Science* 11. https://www.DOI.10.3389/fpls.2020.00688

Ilves KL, Lopez-Fernandez H. 2014. A targeted next-generation sequencing toolkit for exon-based cichlid phylogenomics. *MolecularEcologyResource* 14:802-811

Ilves KL, TortiD, López-Fernández H. 2018.Exon-based phylogenomics strengthens the phylogeny of Neotropical cichlids and identifies remaining conflicting clades (Cichlomorphae: Cichlidae: Cichlinae). *Molecular Phylogenetic and Evolution*118:232-243. https://doi.org/ 10.1016/j.ympev.2017.10.008

Jaser SKK, Dias MAD, LagoADA, Reis NRV, Hilsdorf AWS. 2017. Single nucleotide polymorphisms in the growth hormone gene of *Oreochromis niloticus* and their association with growth performance. Aquatic Resource48: 5835–5845. https://doi.org/10.1111/are.13406

Jiamei J, Yuan H, Zheng X, Wang Q, Kuang T, Li J, Liu J, Shuli S, Wang W, Cheng Li, FH, Huang J, Li C. 2019.Gene markers for exon capture and phylogenomics in ray-finned. *Ecology&Evololution* 9:3973–3983. https:// doi.org/10.1002/ece3.5026.

Joshi R, Árnyasi M,Lien S,Gjøen HM, Alvarez, AT *et al*.2018. Development and Validation of 58K SNP-Array andHigh-Density Linkage Map in Nile Tilapia (*O. niloticus*). *Frontier of Genet*ics9: 472.10.3389/fgene.2018.00472

Kaiser SA, Taylor SA, Chen N, Sillett TS, Bondra ER,Webster MS. 2017. A comparative assessment of SNP

and microsatellite markersfor assigning parentage in a socially monogamous bird. *Molecular Ecology Resources* 17: 183–193. https://doi.org/ 10.1111/1755-0998.12589.

Klápšte J, Ashby RL,Telfer EJ, Graham NJ, Dungey HS, Brauning R, Clarke SM, Dodds KG . 2021. The Use of "Genotyping-by-Sequencing" to Recover Shared Genealogy in Genetically Diverse Eucalyptus Populations. *Forests* 12: 904. https://doi.org/10.3390/f12070904

Kumar S, Stecher G, Tamura K. 2017. Mega 7: Molecular Evolution Genetic Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*33:1870–1874. https:// doi.org/ 10.1093/ molbev/msw054

Lang T, Abadie P, Léger V, Decourcelle T, Frigerio JM, Burban C, Bodenes C, Guichoux E, Le Provost G, Robin C, Tani N, Léger P, Lepoittevin C, El Mujtar VA, Hubert F, Tibbits J, Paiva J, Franc A, Raspail F, Mariette S, Reviron MP, Plomion C, Kremer A, Desprez-Loustau ML, Garnier-Gere P. 2020. High-quality SNPs from genic regions highlight introgression patterns among European white oaks (*Quercus petraea and Q. robur*). *bioRxiv* 388447. https://doi.org/10.1101/388447

Li C, Riethoven JJ, NaylorGJ. 2012 Evol Markers: a database for mining exon and intron markers for evolution, ecology and conservation studies. *MolecularEcologyResource* 12: 967-971

LiC, HofreiterM, Straube N,Corrigan S, Naylor GJ. 2013. Capturing protein-coding genes across highly divergent species. *BioTech* 54:321-326.

Liu J, Shen Q, Bao H. 2022. Comparison of seven SNP calling pipelines for the nextgeneration sequencing data of chickens. *PLoS ONE* 17: e0262574. https://doi.org/ 10.1371/journal.pone.0262574

Liu S, Sun L, Li Y, Sun F, Jiang Y, Zhang Y, Zhang J, Feng J, Kaltenboeck L, Kucuktas H, Liu ZZ. 2014. Development of the catfish 250 K SNP array for genome-wide association studies. *BioMed Central Resource Notes* 7: 135. Doi.org /10.1186/1756-0500-7-135

Lind CE, Kilian A, Benzie JAH. 2017.Development of diversity arrays technology markers as a tool for rapid genomic assessment in Nile tilapia, *Oreochromis niloticus*. Animal Genetics 48:362–364.

Lorenzini R, Fanelli R, Tancredi F, Siclari A, Garofalo L. 2020. Matching STR and SNP genotyping to discriminate between wild boar, domestic pigs and their recent hybrids for forensic purposes. *Scientific Reports* 10:3188. https://doi.org/10.1038/s41598-020-59644-6

Marr SM, Ellender BR, Woodford DJ, AlexanderME, Wasserman RJ, Ivey P, Zengeya T, Weyl OLF. 2017. Evaluating invasion risk for freshwater fishes in South Africa' *Bothalia*47(2), a2177. https://doi.org/10.4102/abc.v47i2.2177

Mashaphu, MF., Downs, CT., Burnett, M., O'Brien, G.,& Willows-Munro, S. (2024). Genetic diversity and population dynamics of wild Mozambique tilapia (*Oreochromis mossambicus*) in South Africa. *Global Ecology and Conservation*, 54, e03043,ISSN 2351 -9894. https://doi.org/10.1016/j.gecco.2024.e03043

Mason VC, Li G, Helgen KM,Murphy WJ. 2011. Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. *Genome Research* 21:1695–1704. https://doi.org/10.1101/gr.120 196.111

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecularphylogenetics and evolution* 66:526-538.

Mojekwu, T. O., & Hoareau, T. 2024. Mozambique Tilapia Oreochromis mossambicus (Peters 1852) and the threat from Oreochromis niloticus (Linnaeus 1758) in South Africa: A review. *Agriculture, Food, and Natural Resources Journal*, 3(1), 88–100. https://doi.org/10.5281/zenodo.13894309

Moralee RD, Van der Bank FH, Van der Waal BCW. 2000. Biochemical genetic markers to identify hybrids between the endemic *Oreochromis mossambicus* and the alien species, *O. niloticus* (Pisces: Cichlidae). *Water SA*26: 263–268.

Mwaura, J.G., Wekesa, C., Kelvin, K., Paul, A., Ogutu, P. A., & Okoth, P. 2023. Pangenomics of the cichlid species (Oreochromis niloticus) reveals genetic admixture ancestry with potential for aquaculture improvement in Kenya. *Journal of Basic and Applied Zoology* 84, 28. https://doi.org/10.1186/s41936-023-00346-6

Neto RV, Yoshida GM, Lhorente JP,Yáñez JM. 2019. Genome-wide association analysis for bodyweight identifies candidate genes related to development and metabolism in rainbow trout (*Oncorhynchus mykiss*). *Molecular Genetics and Genomics* 294:563-571. https://doi.org/10.1007/s00438-018-1518-2.

Neumann, GB, Korkuć, P, Arends D *et al*.,2021 Design and performance of a bovine 200 k SNP chip developed for endangered German Black Pied cattle (DSN). *BMC Genomics* 22: 905 https://doi.org/10.1186/s12864-021-08237-2.

Nugent CM, Leong JS, Christensen KA, Rondeau EB, Brachmann MK,Easton AA, Ouellet-Fagg CL, Crown M TT, DavidsonWS, Koop B, Danzmann RG, Ferguson MM. 2019.Design and characterization of an 87k SNP genotyping array for Arctic charr (*Salvelinusalpinus*).PLoSOne 14: Article e0215008. Doi.org/10.1371/journal.pone.0215008

Palaiokostas C, Bekaert M, Khan MGQ, Taggart JB, Gharbi K, McAndrew BJ,Penman DJ2013. Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) using RAD sequencing. *PLoS One* 8: e68389

Palti Y, Gao G, Liu S, Kent MP, Lien S, Miller MR, Rexroad III CE, Moen T. 2015.The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. *Molecular Ecology Resources* 15: 662–672. https://doi. org/10.1111/1755-0998.12337.

Peñaloza C, Robledo D, Barría A,Trịnh T,Mahmuddin M, Wiener P, Benzie JAH,MM, HoustonRD. 2020. Development and Validation of an Open Access SNP Array for Nile Tilapia (*Oreochromis niloticus*), G3 Genes| Genomes| Genetics10: 2777-2785. https://doi.org/ 10.1534/g3.120.401343.

Qi H, Song K, Li C, Wang W, Li B, Li L, Zhang G(2017) Construction and evaluation of a high-density SNP array for the Pacific oyster (*Crassostrea gigas*). PLoSOne 12: e0174007. https://doi.org/10.1371/journal.pone.0174007.

Robledo D, Palaiokostas C, Bargelloni L, MartínezP, HoustonR (2018)Applications of genotyping by sequencing in aquaculture breeding and genetics.Rev in Aquac 10: 670-682.

Roffler GH, Amish S J, Smith S, Cosart T, Kardos M, Schwartz MK, Luikart G (2016) SNP discovery in candidate adaptive genes using exon capture in a free-ranging alpine ungulate. Mol Ecol Resour 16: 1147–64.

Ruanjaichon V, Khammona K, Thunnom B, Suriharn K, Kerdsri C, Aesomnuk W, Yongsuwan A, Chaomueang N, Thammapichai P, Arikit S, Wanchana S, Toojinda T (2021) Identification of Gene Associated with Sweetness in Corn (*Zea mays* L.) by Genome-Wide Association Study (GWAS) and Development of a Functional SNP Marker for Predicting Sweet Corn. *Plants* 10:1239. https://doi.org/10.3390/plants10061239

Sang VVu, Premachandra HKA, O'Connor W, Nguyen NTH, Dove M, Van Vu I, Le TS,Vendrami DLJ, KnibbW(2021) Development of SNP parentage assignment in the Portuguese oyster *Crassostrea angulata,* AquaRept19: 100615, ISSN 2352-5134,https://doi.org/ 10.1016/ j.aqrep.2021.100615.

Seo D, Cho S, Manjula P, Choi N, Kim Y-K, Koh YJ, Lee SH, Kim H-Y, Lee JH (2021) Identification of Target Chicken Populations by Machine Learning Models Using the Minimum Number of SNPs. *Animals*11:241. https://doi.org/10.3390/ani11010241

Silva PIT, Silva-Junior OB, Resende LV, Sousa VA, Aguiar AV, Grattapaglia D (2020) A 3K Axiom SNP array from a transcriptome-wide SNP resource sheds new light on the genetic diversity and structure of the iconic subtropical conifer tree Araucaria angustifolia (Bert.) Kuntze. *PLoSONE* 15(8): e0230404. https://doi.org/10.1371/journal.pone.0230404

Song S, Zhao J, Li C. 2017Species delimitation and phylogenetic reconstruction of the sinipercids 307 (Perciformes: Sinipercidae) based on target enrichment of thousands of nuclear coding 308 sequences. *Molecular phylogenetic and evolution*111: 44-55.

Suekawa Y, Aihara H, Araki M, Hosokawa, D, Mannen H, Sasazaki S. 2010. Development of breed identification markers based on a bovine 50K SNP array. Meat Science. 85:285–288

Sveistiene R, Tapio M 2021.SNPs in Sheep: Characterization of Lithuanian Sheep Populations. *Animals*, 11: 2651. https:// doi.org/10.3390/ani11092651

Syaifudin M, Bekaert M, Taggart JB, Bartie KL, Wehner S, Palaiokostas C, Khan MGQ, Selly SLC, Huluta G, D'Cotta H, Baroiller JF, McAndrew BJ, Penman DJ2019. Species-specific marker discovery in tilapia. *Scientific Report*9: 13001.

Tibihika PD, CurtoM, Alemayehu E, Waidbacher H, Masembe C, Akoll P,Meimberg H. 2020. Molecular genetic diversity and differentiation of Nile tilapia (*Oreochromis niloticus*, L. 1758) in East African natural and stocked populations. *BMC Evolionary Biology*20*:*16. https://doi.org/10.1186/s12862-020-1583-0.

Trindade FJ, Rodrigues MR, Figueiró HV, Li G, Murphy WJ, Eizirik E. 2021. Genome-Wide SNPs Clarify a Complex Radiation and Support Recognition of an Additional Cat Species. Molecular Biology and Evolutiov 38: 4987–4991. https://doi.org/10.1093/molbev/msab222

Van Bers NE,Crooijmans RP, Groenen MA, Dibbits BW, KomenJ. 2012.SNP marker detection and genotyping in tilapia. *Molecular Ecology Resources* 12: 932-941.

Wang L, Liu Y, Gao L, Yang X, Zhang X, Xie S, Chen M, Wang Y-H, Li J, Shen Y. 2022. Identification of Candidate Forage Yield Genes in Sorghum (*Sorghum bicolor L.*) Using Integrated Genome-WideAssociation Studies and RNA-Seq. *Frontiers in Plant Science* 12:788433.doi: 10.3389/fpls.2021.788433

Wenne, R. 2023. Single nucleotide polymorphism markers with applications in conservation and exploitation of aquatic natural populations. *Animals*, 13(6), 1089. https://doi.org/10.3390/ani13061089

Xia JH, Wan ZY, Ng ZL, Wang L, Fu GH, Lin G, Liu F,Yue GH 2014. Genome-wide discovery and in silico mapping of gene-associated SNPs in Nile tilapia.*Aqua* 432: 67-73. 10.1016/j.aquaculture.2014.04.028

Xia JH, Bai Z, Meng Y, Zhang L, Wang L, Liu F, Jing W, Lin G, Wan ZY, Li J, Lin H, Yue GY. 2015. Signatures of selection in tilapia revealed by whole-genome resequencing. *Sci entific Reports* 5:1-10. https://doi.org/10.1038/srep14168.

Xu J, Zhao Z, Zhang X, Zheng X, Li J, Jiang Y, Kuang Y, Zhang Y, Feng J, Li C, Yu J, Li Q, Zhu Y, Liu Y, Xu P, Sun X. 2014. Development and evaluation of the first high-throughput SNP array for common carp (*Cyprinus carpio*). *BioMed CentralGenomics* 15:307. https://doi.org/10. 1186/1471-2164-15-307.

Yanez JM, Naswa S, Lopez ME, Bassini L, Correa K, Gilbey J, Bernatchez L, Norris LA, Neira R, Lhorente J P, Schnable PS, Newman S, Mileham A, Deeb N, Di Genova A, Maass A. 2016.Genomewide single nucleotide polymorphism discovery in Atlantic salmon (Salmo salar): validation in wild and farmed American and European populations. *Molecular Ecology Resources* 16, 1002-1011. Doi.org/10.1111/1755-0998.12503

Yáñez J M, Yoshida G, Barria A, Palma-Véjares R, Travisany D, Díaz D, Cáceres G, Cádiz MI, López ME, Lhorente JP, Jedlicki A, Soto J, Salas D, Maass A. 2020. High-throughput Single Nucleotide Polymorphism (SNP) discovery and validation through whole-genome resequencing in Nile Tilapia (*Oreochromis niloticus*). *Marine Biotechnology* 22:109–117. https://doi.org/10.1007/s10126-019-09935-5

Yuan H, Jiang J, Jimenez FA, Hoberg EP, Cook JA, Galbreath KE, Li C. 2016. Target gene enrichment in the cyclophyllidean cestodes, the most diverse group of tapeworms. *Molecular Ecology Resources* 16: 1095-1106.

Yuan H, Atta C, Tornabene L, Li C. 2019. Assexon: Assembling Exon Using Gene Capture Data. *Evolutionary Bioinformatics* 15: 1–13. doi.org/10.1177/ 11769343 19874792.

Zeng Q, Fu Q, Li Y, Waldbieser G, Bosworth B, Liu S, Yang Y, Bao L, Yuan Z, Li, N, Liu Z 2017. Development of a 690 K SNP array in catfish and its application for genetic mapping and validation of the reference genome sequence. *Scientific Report* 7: 40347. https://doi.org/10. 1038/srep40347.

Zou K, Kim K-S, Kang D, Kim M-C, Ha J, Moon J-K, Jun T-H. 2022. Genome-Wide Association Study of Leaf Chlorophyll Content Using High-Density SNP Array in Peanuts (*Arachis hypogaea L.*). *Agronomy* 12: 152. https://doi.org/10.3390/agronomy12010152

****