Non-invasive sex genotyping of paiche *Arapaima gigas* by qPCR: An applied bioinformatic approach to identify sex differences

**Edgar A. López-Landavery^{1§}, Guillermo A. Corona-Herrera^{1§}, Luis E. Santos-Rojas¹,
Nadhia M. Herrera-Castillo¹, Tomás H. Delgadin¹, Sandra Tapia-Morales¹, Sophia
González-Martínez¹, Lorenzo E. Reyes-Flores¹, Alan Marín¹, Carmen G. Yzásiga-
Barrera¹, Juan I. Fernandino^{1,2*}, Eliana Zelada-Mázmela^{1*}**

¹ Laboratorio de Genética, Fisiología y Reproducción, Facultad de Ciencias, Universidad Nacional del Santa, Chimbote, Perú.

² Laboratorio de Biología del Desarrollo - Instituto Tecnológico de Chascomús. INTECH (CONICET-UNSAM), Argentina.

§ Equal contribution.

* Equal contribution and corresponding authors: EZM (ezelada@uns.edu.pe), JIF (fernandino@intech.gov.ar)

Keywords

bioinformatics, DNA, sexual markers, mucus, duplex qPCR.

Highlights of the manuscript

- Implementation of a novel approach, called Genome Differences by Unmapped Reads, to identify DNA sex markers.
- Finding of a Male-Specific Region (MSR), present only in males.
- Development of a duplex qPCR to identify genotypic sex through non-invasive sampling.

Abstract

Arapaima gigas, one of the largest freshwater fish in the world, is suffering from high fishing pressure and habitat loss, which threaten the conservation status of its natural populations. Of great cultural importance to Amazonian people, the paiche or pirarucu *A. gigas* is in high demand for its meat, ornamental uses and other byproducts such as scales. Aquaculture is a feasible solution to this dilemma. However, the fact that *A. gigas* presents no sexual dimorphism until it is 5 years old and its long period to sexual maturity are major obstacles for brood-stock management and fingerling production. Thus, the aim of this study was to develop a molecular tool for non-invasive genotypic sexing of paiche throughout its life cycle. We collected samples from gonads, fins and gill mucus of juvenile specimens from local facilities for histological and molecular analysis. Based on the recently available genome sequence of the paiche and making use of current NGS method, we implemented a novel approach, called Genome Differences by Unmapped Reads, to identify DNA sex markers. We found a Male-Specific Region (MSR), identified as MSR_3728, to be

present only in males. Next, we designed two specific sets of primers on this region to identify genotypic sex by qPCR assays. Both primer sets, MSR_107 and MSR_129, detected males with 100% accuracy. Then we developed a duplex qPCR reaction for each primer set along a reference gene, analyzed the melting curves and detected males by observing two distinct peaks, one for MSR_107 or MSR_129 and one for the reference, while females only presented the reference peak. The same results were obtained for gonads, fins and interestingly, a non-invasive source from gill mucus samples. Finally, the gonads were evaluated histologically in a double-blind test, showing 100% accuracy with qPCR assay for identifying males and females. Data clearly demonstrated a novel pipeline approach for identifying DNA sex markers, followed by a quick, non-invasive, cost-effective duplex qPCR method for sexing *A. gigas*. These results may be valuable to efficient paiche aquaculture and conservation studies, helping to reduce the fishing pressure on its natural populations.

Introduction

Arapaima gigas, also known as paiche in Peru and pirarucu in Brazil, is one of the largest freshwater fish species in the world. It inhabits floodplains, rivers and lagoons in the Amazon basin (Nelson et al., 2016; Carvalho et al., 2017). Its natural distribution encompasses the South American countries Brazil, Colombia, Ecuador, Guyana, Peru, Suriname, and Venezuela (Chu-Koo et al., 2008; Hrbek et al., 2005; Farias et al., 2019, de Oliverira et al., 2019), and it was introduced to the Bolivian Amazon in the late 1970s (Miranda-Chumacero et al., 2012). Owing to its high value as an ornamental fish species, it can also be found in Southeastern Asian countries (e.g., Indonesia; Marková et al., 2020). The paiche has great cultural and historical importance for the Amazon region. It has been used as a meat supply for local people for decades, and as consequence of high fishing pressure and habitat loss, its natural population has been declining (Hrbek et al., 2005; Castello and Stewart, 2010; Castello et al., 2015; Farias et al., 2019). Importantly, *A. gigas* has been listed in Appendix II of CITES for protection since 1975. However, the declining population size of *A. gigas* contrasts with the high demand for its meat and other byproducts from local and potential new foreign consumers. In this conflictive scenario, *A. gigas* aquaculture

appears as an ideal solution, which may benefit its conservation status, consumer demand and local producers.

In addition to the biological importance of *A. gigas* from an ecological and conservation standpoint, in recent decades it has been used for aquaculture due to its striking features (Saint-Paul, 1986; Pereira-Filho, 2003). It has one of the highest growth rates in fish, easily reaching a body weight of 10 kg during the first year of life and as much as 200kg as an adult (Saint-Paul, 1986; de Oliveira et al., 2012). Moreover, it is an obligate air-breathing fish with a large, highly vascularized swim-bladder that functions as an air-breathing organ and enables it to inhabit the hypoxic floodplain waters in the Amazon basin, which would facilitate rearing individuals at high densities in ponds (Pereira-Filho, 2003; de Oliveira et al., 2012; Val and Almeida-Val, 2012). Nevertheless, not much is known about its reproductive biology.

A. gigas attains sexual maturity at 5 years of age (Fontenele, 1948; Guerra Flores, 1980; Sandoval, 2007), when males and females may display different body color patterns (Saavedra Rojas et al., 2005). After a pair is formed, the female prepares a nest in sandy soils to spawn the eggs, which are immediately fertilized by the male. Both male and female adults display parental care during the first days after hatching, secreting a characteristic

mucus from the “secretory organ” located at the top of the head, which is thought to provide nutrients to the fries, although only the male continues this behavior until fingerlings are left free (Noakes, 1979; Sandoval, 2007; Torati et al., 2017).

Other than color pattern during the breeding period, *A. gigas* presents no evident sexual dimorphism. The lack of a quick, reliable means to identify its sex before sexual maturation hampers the breeding management and fingerling production involved in rearing it in captivity. Although there have been recent efforts to resolve this situation (Chu-Koo et al., 2008; Torati et al., 2017), one of which resulted in an immune assay method for vitellogenin and sexual steroid determination, there is still a need for a quick method. Taking advantage of the recent availability of the *A. gigas* genome sequence and based on newly-developed bioinformatic methods for determining sexual markers, we aimed to develop a duplex qPCR method for a rapid, non-invasive, robust genotypic sex identification of *A. gigas* at any developmental stage.

Materials and Methods

Experimental organisms, biometrics and biological samples

Juvenile (18-month-old) *A. gigas* males (n=27) and females (n=28) were obtained from aquaculture farms in the city of Pucallpa, department of Ucayali, Peru. Their total length and mass, and gonad length and mass, were measured to calculate condition factor (Htun-Han, 1978) and gonadosomatic index (Anderson et al., 1983). Fish were handled in accordance with the Universities Federation for Animal Welfare Handbook on the Care and Management of Laboratory Animals (www.ufaw.org.uk) and internal institutional regulations.

For molecular and histological identification, gonads (from 28 females and 27 males) and pectoral fin clip (16 females and 13 males) samples were taken from each specimen and preserved in 100% ethanol until molecular analysis. Additionally, gonad samples were taken and preserved in 10% neutral buffered formalin (NBF 10%) for phenotypic sex identification by histological procedures. Finally, gill mucus (16 females and 13 males) was sampled with sterile drugstore cotton swabs (Cody et al., 2018) and preserved in DNA lysis buffer solution until molecular analysis.

Gonad histology

The gonad samples, fixed in NBF 10%, were rinsed and transferred to 70% alcohol, and processed in the automatic dehydrating and paraffin inclusion device Thermo Scientific™ STP 120 Spin Tissue Processor. The tissue was embedded in a modular Tissue Embedding Center EC 350, and 6 µm sections were prepared with a Rotary microtome/ Semi-automatic CUT 5062 SLEE Medical and stained with Hematoxylin-Eosin standard protocol in the Thermo Scientific Varistain® V24-4 AUTOMATIC SLIDE STAINER. The histological slides were mounted and observed under a Motic BA310 Compound Microscope and microphotographs were taken with a Moticam 10.0 MP.

De Novo genome assembly

Illumina NGS sequences (250 bp) were obtained from *A. gigas* females (n = 2) and males (n = 2) (Du et al., 2019), hosted in the database SRA (SRX7250390, SRX7250391) from NCBI. As a first step, the reads were analyzed with the FASTQC program (Andrews, 2010) to determine whether they met the minimum quality necessary to carry out the assembly.

Subsequently, the low-quality reads ($Q < 28$), adapters and “N” bases were removed with the Trimmomatic program (Bolger et al., 2014). Then duplicates were eliminated with the Deduped program (Gregg and Eder, 2019) and corrected with the SOAPdenovo Corrector or SOAPec program (Li et al., 2010). Next, a second FASTQC quality control analysis (Andrews, 2010) and a results report were performed with the MultiQC software (Ewels et al., 2016) to corroborate the quality of the reads before the assembly process. Subsequently, the genomic assembly was carried out with the Discovardenovo program (Weisenfeld et al., 2014). Finally, the male and female genomic assembly quality was evaluated with BUSCO (Simão et al., 2015), Genome Scope (Vurture et al., 2017), Quast (Gurevich et al., 2013) and Bowtie2 (Langmead and Salzberg, 2012).

Male-Specific Region (MSRs) identification by the Genome Differences by Unmapped Reads (GDUR) pipeline

After the male and female genome assemblies were obtained, a search was performed for sex-specific candidate regions. For this purpose, we implemented a bioinformatic pipeline, which we called “Genome Differences by Unmapped Reads (GDUR)”. It is based on multiple programs that are currently used for mapping and filtering NGS sequences, but which have not been used together to obtain sex-specific sequences.

As a first step, we mapped independently the forward and reverse male reads vs. the female *A. gigas* genome with the Fastq Screen program (Wingett and Andrews, 2018). We focused on the unmapped male reads because they have a greater probability of being male-specific sequences. Subsequently, the pair sequences (Forward and Reverse) of the unmapped male reads were searched using the Fastq Pair program (Edwards and Edwards, 2019). Once the sequences with their

corresponding pairs were obtained, the cleaning and preparation steps of the reads were performed with the Trimmomatic (Bolger et al., 2014), Deduped (Gregg and Eder, 2019), SOAPec (Li et al., 2010), FastQC (Andrews, 2010) and MultiQC (Ewels et al., 2016) programs. After that, the genomic sex sequences of the *A. gigas* male were assembled with Discovardenovo and evaluated with BUSCO, Genome Scope, Quast and Bowtie2, as explained in the previous section. Then, local BLAST (Altschul et al., 1990) alignments of the male-specific sequences vs. the female genomic assembly were carried out to corroborate that male sex-specific sequences are not present in the female genome.

Similarly, a local BLAST (Altschul et al., 1990) of the sex-specific sequences vs. the male genomic assembly was carried out to corroborate that the sequences are found in the genome assembly of *A. gigas* males. In addition, with some of the identified sequences, forward and reverse pairs (Fw-Rv) of Male-Specific Region (MSR) and reference gene primers were designed with Primerquest and the oligo analyzer tools of IDT (www.idtdna.com). The designed primer sets were used to perform experimental validation by PCR and qPCR analysis.

DNA extraction and qPCR experimental validation of Male-Specific Regions (MSRs)

A. gigas male and female DNA was extracted from gonads, fins and mucus by the standard phenol-chloroform method added with proteinase K and RNase (Mirimin et al., 2015; Dominguez et al., 2018). The integrity of DNA was observed on a 1% agarose gel electrophoresis with TBE 0.5x and 100 V for 60 min (Simbolo et al., 2013). The purity (A260/A280) and concentration (ng/μl) values were obtained with an Epoch Microplate Spectrophotometer (www.biotek.com) and a Qubit fluorometer® 3.0 using a Qubit™ dsDNA BR assay kit (<https://www.thermofisher.com/>), respectively. Preliminary PCRs were carried out with DNA from

male and female gonads with a PCR Master Mix 2x (<https://www.thermofisher.com/>) and a male-specific or reference gene final primer concentration of 0.5 μ M and 100 ng of total DNA. Under a cycling protocol of initial denaturation step 5 min at 95°C, final extension 5 min at 72°C and 35 cycles of amplification (Denaturation 94°C, Annealing 60°C, Extension 72°C x 20 s) in a Veriti 96-Well Thermal Cycler (<https://www.thermofisher.com/>). Finally, the presence of one specific PCR product in males and no product in females (MSR primers), and one product in both sexes (reference gene primers) were corroborated with standard 1.5% agarose gel electrophoresis (TBE 0.5x, 100 V, 45 min).

After duplex PCR experiments were designed with the MSRs and reference gene primers,

corroborating the absence of dimers and unspecific products, the best duplex primers were tested by qPCRs with the SYBR-Green master I protocol, in a Roche LightCycler 480 II real-time PCR instrument (<https://lifescience.roche.com/>). The qPCR reactions were optimized with 30 ng of total DNA, and a final primer concentration of 0.05 μ M, with a protocol of initial denaturation step 5 min at 95°C, final extension 5 min at 72°C, 35 cycles of amplification (Denaturation 94°C, Annealing 62°C, Extension 72°C x 20 s) and a final melting curve analysis of the PCR products.

Statistical analysis

For biometric comparisons, the statistical differences between values were determined by

Figure 1. A) Summary of bioinformatic methods for finding male-specific regions in *Arapaima gigas*; B) Dot-plot alignment of male-specific regions vs. male and female genome scaffolds of *A. gigas*; C) Male-specific region (MSR_3728) used to validate the methodology by qPCR vs. Female genome alignment. (FMA) Female genome scaffolds alignment; (MA) Male genome scaffolds alignment.

Student's t-test ($P \leq 0.05$) with the GraphPad prism V. 9.0 software.

Results

Biometrics, condition factor and gonadosomatic index (GSI) of A. gigas

Fish sampled at the aquaculture facilities did not show any differences between males and females regarding total length, total mass and condition factor (**Supplementary Table 1**). Nevertheless, there were statistical differences in gonad mass, gonad length and GSI between males and females, with females having the largest gonads ($p < 0.05$; **Supplementary Table 1**).

Genome assemblies and Male-Specific Region identification by GDUR pipeline

The genomic assembly generated a total length of bases of 657.53 Mb for females and 659.05 Mb for males. The evaluation of assemblies from both sexes indicated N50 values of approximately 75,000 bp for females and 81,000 bp for males. This meant a difference between assemblies of about 6,000 bp between sexes. The L50 values were 2,112 bp for females and 1,962 bp for males (**Supplementary Table 2**).

The coverage for the genome was 42.9x for females and 42.6x for males. In addition, genome scope analyses indicated an error rate of 0.04% in the genome reads. Comparison of the two sexes showed that the assembly of males presented a greater haploid length (628.16 Mb), a greater length of repetitions (28.077 Mb), a greater length of unique sequences (600.08 Mb), and a greater number of duplications (0.74%) than did the female assembly. Additionally, males presented lower levels of heterozygosity (0.16%) than did females (0.22%, **Supplementary Table 2**). These values suggest that males have a larger genome, with

more duplicated regions, and have lost heterozygosity throughout their evolution. In this case, sex-specific genomic duplications and a possible hemizygous condition could be assumed (they only have one copy of the X chromosome genes, instead of the usual two), as occurs in male mammals with XY sex determination systems.

The evaluation of genomes by mapping the reads with Bowtie2 indicated that assembly of females presents a slightly higher number of reads (64 million) than the males; however, the mapping rate for both genomes was $\sim 98.8\%$ (**Supplementary Table 2**). On the other hand, the analysis of the presence of orthologous groups with BUSCO indicated that assemblies presented similar values in the coincidence of identified orthologous genes (86%), with about 10% fragmented and only 4.3% lost using the “vertebrata” database. Based on statistics and quality control of genome assemblies, the next step was to proceed with the analyses to obtain sex-specific genomic sequences in *A. gigas*.

Paiche sex-specific sequences were generated with the methodology called Genome Differences by Unmapped Reads or GDUR (**Figure 1A**). As the first step, the reads of the males that were not mapped into the female genome were identified. The levels of reads not mapped in the female genome were 2.33% and 5.50% for the forward and reverse sense, respectively. This condition found with the present workflow suggests that most of the sex-specific sequences are found in the antisense strand of the male *A. gigas* genome. Once sequences were identified, their corresponding pairs were obtained from the original cleaned files. Quality control of male sex-specific paired reads generated 897,263 unique reads (91.5%) and 83,571 duplicated reads (8.5%), with average Phred quality 38, 45% GC, and average length 176 bp (**Supplementary Table 3**).

The assembly of sex-specific paired reads generated an N50 value for scaffolds of 3,542 bp, with a total length of approximately 2.08 Mb, a low percentage of chimeras (0.02%) and a coverage of 93.1x. The total number of contigs for this assembly was 903, with a N50 value of 3,417 bp, 43.59% GC and with the presence of 9.82 “Ns” per 100 Kb. The analysis with BUSCO showed that only 0.4% of the orthologous genes in the vertebrata database were found in the sex-specific assembly of *A. gigas* (Supplementary Table 3).

The genome scope analysis indicated that the read error rate for this assembly was 1.67%, with a haploid length of 3.09 Mb, a repeat length of 102.20 bp and unique sequences length of 2.98 Mb. Furthermore, the duplication level was 1.23% and heterozygosity was 1.74%, with a model fit of 97%. Another characteristic was that the percentage of female reads that unmapped in the male sex-specific assembly was 98, suggesting that just 2% of the female reads showed some similarity with the

male sex-specific assembly (Supplementary Table 3).

Finally, an analysis to compare the assembled genomes was performed with the D-Genies software (Cabanettes and Klopp, 2018). Comparison between MSRs and male genome indicated that just 0.13% of the sequences matched between scaffolds (Figure 1B). In contrast, comparison between MSRs vs. female genome clearly showed a match of just 0.02% between scaffolds (Figure 1B). These results suggest very low similarity of the MSRs vs. the female genome, and demonstrate the confidence of the GDUR pipeline implemented in this study to find sex-specific sequences of *A. gigas* males.

qPCR Primers design and experimental validation of Male-Specific Regions in *A. gigas* gonads, fins and mucus

After identifying the MSR_3728 region, two sets of primers on this sequence were designed to be used

Figure 2. A) Summary of *Arapaima gigas* juvenile sampling; B) qPCR melting curves of Male-Specific Regions (MSR_107/MSR_129) and the reference gene TBPL1_205 in *A. gigas* males and females; C) Gel electrophoresis of duplex PCR products of Male-Specific Regions (MSR_107/MSR_129) and the TBPL1_205 gene in *A. gigas* males and females. (L) Ladder 50 bp; (M) Male; (FM) Female.

in the qPCR sexing tool. The first step was to check whether there were female scaffolds that matched the MSR_3728 region, obtaining the FMA_495bp at the beginning and FMA_188bp at the end of this region. Next, the MSR_3728 region was aligned against the male genome, obtaining a 3,222bp region, and corroborating its presence in males and its absence in females. Then, and based on a 3,045bp sequence in the MSR_3728 region, two pairs of qPCR primers were designed to validate the GDUR pipeline in *A. gigas* (**Figure 1C**): MSR_107 (from 2,457bp to 2,564bp) and MSR_129 (from 2,540bp to 2,669bp) (**Figure 1C**). As the present study aimed to develop a duplex qPCR controlling for DNA integrity, four reference genes (HKGs) were tested: TATA-Box Binding Protein Like 1 (TBPL1), 60S Ribosomal Protein L3 (RPL3), Eukaryotic Translation Elongation Factor 1 Alpha 1 (EF1A1), and Phosphoglycerate Kinase 1 (PGK1). Out of all the HKGs, TBPL1 was the best choice, because it produced a 205bp amplicon with no inter- or intra-specific dimeric interactions with itself or between the MSR_107 or MSR_129 primers.

For qPCR validation, DNA was extracted from female and male *A. gigas* gonads, fins and mucus (**Figure 2A**). All samples presented an A260/A280

ratio of 1.8 and concentrations between 946.5 and 1,999.6 ng/ μ l. Analysis of the PCR products showed two amplicons in males and only one amplicon in females, corresponding to the MSR_107 or MSR_129 product plus TBPL1_205 product in males and TBPL1_205 product in females (**Figure 2B**). Afterwards, melting curve analyses showed that the best qPCR condition was 35 amplification cycles with a primer combination of 0.1 μ M for MSR and 0.05 μ M for HKG at annealing temperature of 62°C (**Figure 2C**). PCR efficiencies obtained by the standard curve method were 109.19% and 101.32% for MSR_107 and MSR_129 primers, respectively. Efficiencies for TBPL1_205 were 96.16% for males and 98.22% for females (**Supplementary Table 4**). The qPCR melting curve analysis showed temperatures of 76.6, 77.5 and 84.7°C for the qPCR products of MSR_107, MSR_129 and TBPL1_205, respectively (**Figure 2C**). Data of melting curves and qPCR products for 14 males and 14 female gonads is shown in Figure 3, and the melting temperature and Cq values for all the samples are shown in **Supplementary Table 5**.

Finally, the set of primers was tested in double-blind mode to corroborate the correlation between gonadal phenotype and genotypic sex. For 100% of

MSR	Source	Male	Female	Total n	Percentage of sex identification
MSR_107	Gonad	27	28	55	100 %
	Fin	13	16	29	100 %
	Mucus	13	16	29	100 %
MSR_129	Gonad	27	28	55	100 %
	Fin	13	16	29	100 %
	Mucus	13	16	29	100 %

Table 1: Percentage of sex identification with the Male-Specific Regions (MSR_107/MSR_129) in *Arapaima gigas*. (n) number of individuals.

the tested samples, males and females were identified at molecular level with the primers MSR_107 and MSR_129 (Table 1). In some mucus samples, the qPCR analysis showed a degree of DNA cross-contamination, mainly due to the nature of the samples.

Histological observations on gonads of *A. gigas* juveniles

Finally, gonad phenotype was identified in order to validate the double-blind performed previously with the sexual genotype. In this regard, the histological observations identified female ovaries at primary stages of development (early previtellogenesis), characterized by the presence of a clear ovarian cavity, oogonium, perinucleolar oocytes and early lipid oocytes with a nucleus, nucleolus and cortical alveoli (Figure 3 A, B). The testes of males were found at primary developmental stages, characterized by clear spermatocytes with spermatogonium, testicular lumen and in some cases a slightly presence of spermatids (Figure 3 C, D). The histological observations and the GSI data clearly indicate that 18-month-old *A. gigas* juveniles are at immature gonadal stages, as described in Yong et al. (2020), for fish at same age.

Discussion

External sexual dimorphism is present in many fish species, but absent in many others (Nelson et al., 2016). From the standpoint of aquaculture and conservation, easy sex identification is needed to manage reproduction and measure the health of a fish population. There is great interest in the aquaculture and conservation of *A. gigas* (paiche) (Hrbek et al., 2005; Castello and Stewart, 2010; Castello et al., 2015; Farias et al., 2019), which is a very important Amazonian fish. It presents no sexual growth dimorphism, and as far as we know,

Figure 3. Histological observations of gonads from *Arapaima gigas* juveniles. A; Early growing ovary, B; Growing stage ovary, C; Early growing testis, D; Growing stage testis. (OC) Ovarian cavity; (OG) Oogonium; (PO) Perinucleolar oocyte; (ELO) Early lipid oocyte; N (Nucleus); (n) Nucleolus; (ca) Cortical alveoli; (SC) Spermatocytes; (SG) Spermatogonium; (SP) Spermatids.

there is no other published data reliable enough to identify sex throughout its life cycle. Thus, the lack of convenient methods for sex identification, which impairs breeding management and fingerling production, is a major obstacle for paiche culture. The *A. gigas* genome has recently been published (Du et al., 2019), providing the opportunity to search for sex-specific genetic markers which would enable genotypic sex to be identified at any stage in the life cycle by means of a simple method. As mentioned above, one of the main difficulties for efficient breeding management of *A. gigas* in captivity is the lack of an adequate method for identifying sex. Thus, after identifying a male-specific region (MSR) for *A. gigas*, we developed a duplex qPCR to identify genotypic sex, which provided highly accurate sex identification for all samples, with 100% match between genotypic sex (PCR) and biological sex by morphological identification of the gonad (histology). Moreover, this qPCR method was applied successfully not only to tissue samples (i.e., gonads and fins) but also to gill mucus, which makes this a truly non-invasive, novel technique for sexing fish.

The identification of genotypic sex in fish is being increasingly used thanks to the availability of NGS techniques for species for which aquaculture and conservation are important. Recently, an approach using NGS method and PCR validation was used on the mandarin fishes *Siniperca scherzeri* and *Siniperca kneri* (Han et al., 2020). These studies designed three set of primers on a MSR which identified males with a success rate of 100%. Similar PCR protocols based on sex-specific markers have been developed for other fish species, for instance in the carps *Hypophthalmichthys nobilis*, *Hypophthalmichthys molitrix* (Liu et al., 2018) and the snakehead *Channa argus* (Ou et al., 2017). However, all these methods were based on end-point PCR, which does not enable the simultaneous detection of multiple genes and requires tissue samples for DNA extraction, affecting the desired non-invasive attribute. Importantly, our PCR-based sex identification tool is a duplex Real Time PCR, where we can measure both the sex-specific marker and the HKG in the same reaction well, a desirable feature for minimizing testing costs. In fact, we were able to detect two peaks in the melting curve analysis with either combination of HKG plus the sex-specific primer set, which indicates that the duplex reaction can efficiently amplify both amplicons in every well.

Although previous work based on immunoassays for vitellogenin and sexual steroids hormones has been conducted for *A. gigas* with a high degree of success (Chu-Koo et al., 2008), it requires the handling of animals, which in the case of *A. gigas*, due to its large size, requires very laborious logistics to obtain blood samples. Moreover, it is only effective once the fish has begun its reproductive stage, which happens as from three years of age, making this tool unviable when early identification is needed for managing breeding stock. Other methods, such as endoscopies or surgery are more difficult because

they are invasive, risky and require sedation, although good progress has been made in this regard (Torati et al., 2017). Importantly, however, none of these paiche sexing approaches can be applied at early developmental stages or juvenile/resting state. Our innovative non-invasive qPCR method enables early sex identification in *Arapaima* and can be used for studying biological traits such as growth during any developmental stage, performing nutrition tests and evaluating natural populations. However, when mucus samples are collected in rearing ponds, maximum care is required in order to minimize the chance of any possible cross-contamination for DNA amplification. Despite this small contamination to be expected given the nature of sampling, sex identification at molecular level is still reliable, because the qPCR product and the shape of the curve/band are intense, clear and well defined. Overall, these results indicate that primers MSR_107 and MSR_129 are male-specific and can feasibly be implemented as molecular markers to differentiate *A. gigas* genotypic males and females.

On the other hand, and according to the statistical results, the quality of the *A. gigas* female and male genomic assemblies was good. Based on the above, the bioinformatics pipeline described here may be considered as preliminarily validated to perform *de novo* genomic assemblies in several species of interest, with high probability of identifying sex-specific sequences to be used as molecular markers in aquaculture and conservation activities. Another interesting aspect of the male genome statistics is that they suggest that *A. gigas* has an XY determination system, where the male has the heterogametic sex. However, further bioinformatic, molecular, and cytogenetic evidence is required to corroborate this hypothesis.

Finally, the modest percentage of reads alignment with the male-specific assembly may be explained by the natural accumulation of repetitive

sequences observed in the sex-specific regions of vertebrates. With this type of sequences, it is difficult to make a complete assembly when there are no long sequencing reads that enable this type of repetitive sequence patterns to be resolved. Despite the percentage of alignment and the lack of identification of orthologous genes, preliminary identification was possible for some protein sequences which may be good candidates for use as sexual molecular markers in *A. gigas*.

Conclusion

The present work has developed a qPCR method for quick, accurate sex determination of *A. gigas* individuals at any stage of development. It was validated for gonad, fin and mucus samples, enabling non-invasive testing, with 100% of fish correctly sexed. This qPCR method is of great value to aquaculture and would be an excellent tool for ecological studies of threatened wild populations of *A. gigas*, as only a sample of mucus or fin suffices to sex the fish accurately.

CRedit author statement

EALL, GACH: Conceptualization, Software, Methodology, Formal Analysis; LESR, NMHC, STM, SGM, LERF, AM, CGYB: Methodology; THD: Writing- Original draft preparation, JIF: Funding acquisition, Conceptualization, Investigation, Supervision; Writing- Original draft preparation, Writing- Reviewing and Editing; EZM: Funding acquisition, Conceptualization, Investigation, Supervision.

Supplementary Files

Supplementary Figure 1: A, B) Gel electrophoresis of duplex PCR products of Male-Specific Regions (MSR_107/MSR_129) and the TBPL1_205 gene in *Arapaima gigas* males; C, D) Gel electrophoresis of duplex PCR products of Male-Specific Regions (MSR_107/MSR_129) and the TBPL1_205 gene in *Arapaima gigas* females; E, F) qPCR melting curves of Male-Specific Regions (MSR_107/MSR_129) and the reference gene TBPL1_205 in *Arapaima gigas* males and females.

Supplementary Table 1: Fish total weight, total length, condition factor, gonad weight and length, and Gonadosomatic Index (GSI) of *Arapaima gigas* analyzed in this study (NA: Not available).

Supplementary Table 2: Statistical summary for the male and female genome assemblies and assessment of their quality in *Arapaima gigas*.

Supplementary Table 3: Statistical summary of assembly from male-specific reads in *Arapaima gigas*.

Supplementary Table 4: Genes of interest⁽¹⁾ and housekeeping genes⁽²⁾, efficiency (E) and R² values from duplex qPCR assays.

Supplementary Table 5: Raw data of duplex qPCR assays for each individual tested in this study (UDT: Undetected; NA: Not available).

Disclosure summary

The authors have no conflict of interest.

Acknowledgments

The authors gratefully acknowledge the support from Universidad Nacional del Santa (UNS).

Funding sources

This work was supported by contributions from the Consejo Nacional de Ciencia y Tecnología (CONCYTEC) and the World Bank, granted through Fondo Nacional de Desarrollo Científico, Tecnológico y de Innovación Tecnológica (FONDECYT) to the project “Application of omics and gene editing tools in the production and research of commercially important aquatic organisms from Peru” under the program “Incorporation of Researchers”.

References

- Almeida, I., Ianella, P., Faria, M., Paiva, S. & Caetano, A. (2013) Bulk segregant analysis of the pirarucu (*Arapaima gigas*) genome for identification of sex-specific molecular markers. *Gen Mol Res* 12, 6299–6308. doi: 10.4238/2013
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–10. doi: 10.1016/S0022-2836(05)80360-2
- Andrews, S. (2010) FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

- Anders, S., Huber, W. (2010) Differential expression analysis for sequence count data. *Gen Biol* 11(10):R106. doi: 10.1186/gb-2010-11-10-r106
- Arantes, C.C., Castello, L., Cetra, M., Schilling, A. (2013) Environmental influences on the distribution of arapaima in Amazon floodplains. *Environ Biol Fish* 96: 1257–1267. doi: 10.1007/s10641-011-9917-9
- Bruce, J., Walker, T.A., Terrance S., et al. (2014) Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* 9(11): e112963. doi: 10.1371/journal.pone.0112963
- Bairoch, A., Apweiler, R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res* 28(1):45–48. doi: 10.1093/nar/28.1.45
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120. doi.org/10.1093/bioinformatics/btu170
- Cabanettes F, Klopp C. (2018) D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* 6:e4958. doi: 10.7717/peerj.4958
- Carreiro, C. R. P., Furtado-Neto, M. A. A., Mesquita, P. E. C. & Bezerra, T. A. (2011) Sex determination in the Giant of Amazon Basin *Arapaima gigas* using laparoscopy. *Acta Amazonica* 41. doi: 10.1590/S0044-59672011000300012
- Carvalho, F., Power, M., Forsberg, B.R., Castello, L., Martins, E.G., Freitas, C.E. (2018). Trophic Ecology of *Arapaima* sp. in a ria lake—river—floodplain transition zone of the Amazon. *Ecol Freshw Fish* 27(1), 237–246. doi: 10.1111/eff.12341
- Castello, L., Stewart, D.J. (2010). Assessing CITES non-detriment findings procedures for *Arapaima* in Brazil. *J Appl Ichthyol* 26(1), 49–56. doi:10.1111/j.1439-0426.2009.01355.x
- Castello, L., Arantes, C.C., Mcgrath, D.G., Stewart, D.J., Sousa, F.S.D. (2015) Understanding fishing-induced extinctions in the Amazon. *Aquat Conserv* 25, 587–598. doi: 10.1002/aqc.2491
- Cavole, L.M., Arantes, C.C. Castello, L. (2015) How illegal are tropical small-scale fisheries? An estimate for arapaima in the Amazon. *Fish Res* 168, 1–5. doi: 10.1016/j.fishres.2015.03.012
- Clark, S.C., Egan, R., Frazier, P.I., Wang, Z. (2013) ALE: a generic assembly likelihood evaluation framework for assessing the accuracy of genome and metagenome assemblies. *Bioinformatics* 29(4):435–43. doi: 10.1093/bioinformatics/bts723
- Conte, M.A., Gammerding, W.J., Bartie, K.L., Penman, D.J., Kocher, T.D. (2017) A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. *BMC Genomics* 18:341. doi:10.1186/s12864-017-3723-5
- Costa, G.W., Cioffi, M.B., Bertollo, L.A., Molina, W.F. (2013) Transposable elements in fish chromosomes: a study in the marine cobia species. *Cytogenet Genome Res* 141(2–3):126–32. doi: 10.1159/000354309
- Chu-Koo, F., Dugué, R., Alván Aguilar, M., Casanova Daza, A., Alcántara Bocanegra, F., Chávez Veintemilla, C., et al. (2008) Gender determination in the Paiche or Pirarucu (*Arapaima gigas*) using plasma vitellogenin, 17 β -estradiol, and 11-ketotestosterone levels. *Fish Physiol Biochem* 35(1), 125–136. doi:10.1007/s10695-008-9211-8
- de Oliveira, E. G. et al. (2012) Effects of stocking density on the performance of juvenile pirarucu (*Arapaima gigas*) in cages. *Aquaculture* 370, 96–101. doi: 10.1016/j.aquaculture.2012.09.027
- de Oliveira, E.A., Bertollo, L.A.C., Rab, P., Ezaz, T., Yano, C.F., Hatanaka, T., et al. (2019) Cytogenetics, genomics and biodiversity of the South American and African *Arapaimidae* fish family (Teleostei, Osteoglossiformes). *PLoS ONE* 14(3): e0214225. doi: 10.1371/journal.pone.0214225
- Dillies, M.A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., et al (2013) A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform* 14(6): 671–683. doi: 10.1093/bib/bbs046
- Dominguez Del Angel, V., Hjerde, E., Sterck, L., et al. (2018) Ten steps to get started in Genome Assembly and Annotation. *F1000Research* 7: ELIXIR-148. doi: 10.12688/f1000research.13598.1
- Du, K., Wuertz S., Adolphi M., Kneitz S., Stöck M., Oliveira M., Nóbrega R., Ormanns J., Kloas W., Feron R., Klopp C., et al. (2019) The genome of the arapaima (*Arapaima gigas*) provides insights into gigantism, fast growth and chromosomal sex determination system. *Sci Rep* 9: 5293. doi:10.1038/s41598-019-41457-x
- Ewels, P., Magnusson, M., Lundin, S., Käller, M. (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32(19): 3047–3048, doi: 10.1093/bioinformatics/btw354
- FAO (2012). Cultured Aquatic Species Information Programme. *Arapaima gigas*. Cultured Aquatic Species Information Programme. Text by Nuñez, J. In: FAO Fisheries and Aquaculture Department. Rome.
- Farias, I. P., Willis, S., Leão, A., Verba, J. T., Crossa, M., Foresti, F., et al. (2019). The largest fish in the world's biggest river: Genetic connectivity and conservation of *Arapaima gigas* in the Amazon and Araguaia-Tocantins drainages. *PLoS ONE* 14(8), e0220882. doi: 10.1371/journal.pone.0220882

- Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucl Acid Res.* 239: W29-W37. doi:10.1093/nar/gkr367
- Flutre T, Duprat E, Feuillet C, Quesneville H (2011) Considering Transposable Element Diversification in De Novo Annotation Approaches. *PLoS ONE* 6(1): e16526. doi: 10.1371/journal.pone.0016526
- Fontenele, O. (1948) Contribution to the biology of the arapaima *Arapaima gigas* (Cuvier) in captivity (Actinopterygii, Osteoglossidae). *Revista Brasileira de Biologia*, 8, 445-459.
- Gao, S., Bertrand, D., Chia, B.K.H., Nagarajan, N. (2016) OPERA-LG: efficient and exact scaffolding of large, repeat-rich eukaryotic genomes with performance guarantees. *Genome Biol* 2016: 17:102. doi:10.1186/s13059-016-0951-y
- Gamble, T., Zarkower, D. (2014) Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. *Mol Ecol Res* 14, 902–913. doi: 10.1111/1755-0998.12237
- Grabherr, M.G., Haas, B.J., Yassour, M., et al. (2011) Trinity: Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotech* 29(7):644-652. doi: 10.1038/nbt.1883
- Gregg, F., Eder, D. (2019) Dedupe. <https://github.com/dedupeio/dedupe>.
- Guerra Flores, H. (1980) Desarrollo sexual del paiche (*Arapaima gigas*) en la zonas reservadas del estado (Ríos Pacaya y Samiria) 1971–1975. Informe IMARPE 67. Instituto del Mar del Perú, Callao, Peru
- Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29(8): 1072-5. doi: 10.1093/bioinformatics/btt086
- Haas BJ, Papanicolaou A, Yassour M, et al. (2013) De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat Prot* 8(8):10.1038/nprot.2013.084. doi: 10.1038/nprot.2013.084
- Hamada, M., Ono, Y., Asai, K. (2017) Training alignment parameters for arbitrary sequencers with LAST-TRAIN. *Bioinformatics* 33,6: 926–928. doi: 10.1093/bioinformatics/btw742
- Han, C., Zhu, Q., Zhou, X., Ouyang, H., Han, L., Chen, J., ... & Zhang, Y. (2020). A PCR-based genetic sex identification method in spotted mandarin fish (*Siniperca scherzeri*) and big eye mandarin fish (*Siniperca kneri*). *Aquac Rep* 18, 100552. doi: 10.1016/j.aqrep.2020.100552
- Hrbek, T., Farias, I. P., Crossa, M., Sampaio, I., Porto, J. I., & Meyer, A. (2005). Population genetic analysis of *Arapaima gigas*, one of the largest freshwater fishes of the Amazon basin: implications for its conservation. *Animal Conserv* 8(3), 297-308. doi: 10.1017/S1367943005002210
- Honeycutt JL, Deck CA, Miller SC, Severance ME, Atkins EB, Luckenbach JA, Buckel JA, Daniels HV, Rice JA, Borski RJ & Godwin J (2019) Warmer waters masculinize wild populations of a fish with temperature-dependent sex determination. *Sci Rep* 9: 6527. doi: 10.1038/s41598-019-42944-x
- Humanson, G.L. (1972) *Animal Tissue Techniques*, 3rd ed. W. H. Freeman and company, San Francisco.
- Hunt, M., Kikuchi, T., Sanders, M. et al. (2013) REAPR: a universal tool for genome assembly evaluation. *Genome Biol* 14(5):R47. doi: 10.1186/gb-2013-14-5-r47
- Hrbek, T., Farias, I. P. (2008) The complete mitochondrial genome of the pirarucu (*Arapaima gigas*, Arapaimidae, Osteoglossiformes). *Genet Mol Biol* 31, 293–302. doi: 10.1590/S1415-47572008000200024
- Ilha, P., Schiesari, L., Yanagawa, F.I., Jankowski, K., Navas, C.A. (2018). Deforestation and stream warming affect body size of Amazonian fishes. *PloS ONE* 13(5), e0196560. doi: 10.1371/journal.pone.0196560
- Iwasaki, Y., Nishiki, I., Nakamura, Y., et al. (2016) Effective de novo assembly of fish genome using haploid larvae. *Gene* 576,2,1; 644-649. doi: 10.1016/j.gene.2015.10.015
- Jacobina, U.P., Cioffi, M.B., Rodrigues Souza, L.G., et al. (2011) Chromosome Mapping of Repetitive Sequences in *Rachycentron canadum* (Perciformes: Rachycentridae): Implications for Karyotypic Evolution and Perspectives for Biotechnological Uses. *J Biomed Biotech.* 218231,8. doi: 10.1155/2011/218231
- Khan, A.R., Pervez, M.T., Babar, M.E. (2018) A Comprehensive Study of De Novo Genome Assemblers: Current Challenges and Future Prospective. *Evolut Bioinform Online* 14: 1176934318758650. doi: 10.1177/1176934318758650
- Kamiya, T., Kai, W., Tasumi, S., et al. (2012) A Trans-Species Missense SNP in *Amhr2* Is Associated with Sex Determination in the Tiger Pufferfish, *Takifugu rubripes* (Fugu). *PLoS Genet* 8(7): e1002798. doi: 10.1371/journal.pgen.1002798
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., Tanabe, M. (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucl Acid Res* 40: D109-D114. doi: 10.1093/nar/gkr988
- Koren, S., Walenz, B.P., Berlin, K., et al. (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research.* 2017;27(5):722-736. doi: 10.1101/gr.215087.116
- Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden

- Markov model: application to complete genomes. *J Mol Biol* 305(3): 567-80. doi: 10.1006/jmbi.2000.4315
- Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmmer (2007) consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35(9): 3100–3108. doi: 10.1093/nar/gkm160
- Langmead, B., Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4): 357-9. doi: 10.1038/nmeth.1923
- Lapointe, D., Cooperman, M.S., Chapman, L.J., Clark, T.D., Val, A.L., Ferreira, M.S., et al (2018) Predicted impacts of climate warming on aerobic performance and upper thermal tolerance of six tropical freshwater fishes spanning three continents. *Conserv Physiol* 6(1): coy056. doi: 10.1093/conphys/coy056
- Lavoué S, Sullivan JP (2004) Simultaneous analysis of five molecular markers provides a well-supported phylogenetic hypothesis for the living bony-tongue fishes (Osteoglossomorpha: Teleostei). *Mol Phylogenet Evol* 33: 171–185. doi: 10.1016/j.ympev.2004.04.021
- Li H, Handsaker B, Wysoker A, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16): 2078-2079. doi: 10.1093/bioinformatics/btp352
- Li B and Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12: 323. doi: 10.1186/1471-2105-12-323
- Liu H, Pang M, Yu X, et al. (2018) Sex-specific markers developed by next-generation sequencing confirmed an XX/XY sex determination system in bighead carp (*Hypophthalmichthys nobilis*) and silver carp (*Hypophthalmichthys molitrix*). *DNA Res* 25(3):257-264. doi: 10.1093/dnares/dsx054.
- Marková, J., Jerikho, R., Wardiatno, Y., Kamal, M. M., Magalhães, A. L. B., Bohatá, L., ... Patoka, J. (2020). Conservation paradox of giant arapaima *Arapaima gigas* (Schinz, 1822) (Pisces: Arapaimidae): endangered in its native range in Brazil and invasive in Indonesia. *Knowl Manag Aquat Ecosyst* 421: 47. doi: 10.1051/kmae/2020039
- Martínez, P., Viñas, A.M., Sánchez, L., Díaz, N., Ribas, L., Piferrer, F. (2014) Genetic architecture of sex determination in fish: applications to sex ratio control in aquaculture. *Front Genet* 5: 340. doi: 10.3389/fgene.2014.00340
- Miranda-Chumacero, G., Wallace, R., Calderón, H., Calderón, G., Willink, P., Guerrero, M., et al. (2012). Distribution of arapaima (*Arapaima gigas*) (Pisces: Arapaimidae) in Bolivia: implications in the control and management of a non-native population. *BioInvasions Record*, 1(2): 129–138. doi: 10.3391/bir.2012.1.2.09
- Mirimin, L. and Roodt-Wilding, R. (2015) Testing and validating a modified CTAB DNA extraction method to enable molecular parentage analysis of fertilized eggs and larvae of an emerging South African aquaculture species, the dusky kob *Argyrosomus japonicus*. *J Fish Biol* 86: 1218–1223. doi: 10.1111/jfb.12639
- Molina WF, Alves DEO, Araújo WC et al. (2010) Performance of human immunostimulating agents in the improvement of fish cytogenetic preparations. *Genet Mol Res* 9:1807–1814. doi: 10.4238/vol9-3gmr840
- Nelson, J.S., Grande, T.C., Wilson, M.V.H. (2016) Fishes of the world. Fifth edition. Hoboken, New Jersey: John Wiley & Sons, 707 p.
- Noakes, D. L. (1979). Parent-touching behavior by young fishes: incidence, function and causation. *Environ Biol Fish* 4(4), 389-400. doi: 10.1007/BF00005529
- Núñez, J. et al. (2011) Reproductive success and fry production of the paiche or pirarucu, *Arapaima gigas* (Schinz), in the region of Iquitos, Perú. *Aquaculture Res* 42, 815–822. doi: 10.1111/j.1365-2109.2011.02886.x
- Ospina-Alvarez N, & Piferrer F (2008) Temperature-dependent sex determination in fish revisited: prevalence, a single sex ratio response pattern, and possible effects of climate change. *PLoS ONE* 3(7), e2837. doi: 10.1371/journal.pone.0002837
- Ou, M., Yang, C., Luo, Q., Huang, R., Zhang, A. D., Liao, L. J., ... & Wang, Y. P. (2017). An NGS-based approach for the identification of sex-specific markers in snakehead (*Channa argus*). *Oncotarget* 8(58), 98733. doi: 10.18632/oncotarget.21924
- Pereira-Filho, M., Cavero, B. A. S., Roubach, R., Ituassú, D. R., Gandra, A. L., Crescêncio, R. (2003). Cultivo do pirarucu (*Arapaima gigas*) em viveiro escavado. *Acta Amazonica*, 33(4), 715-718.
- Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8(10):785-6. doi: 10.1038/nmeth.1701.
- Powell S, Szklarczyk D, Trachana K, et al. (2012) eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucl Acid Res* 40: D284-D289. doi: 10.1093/nar/gkr1060
- Punta M, Coggill PC, Eberhardt RY, et al. (2012) The Pfam protein families database. *Nucl Acid Res* 40: D290-D301. doi: 10.1093/nar/gkr1065
- Roberts A and Pachter L. (2013) Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat Meth.* 10(1):71-73. Doi: 10.1038/nmeth.2251

- Roberts NB, Juntti SA, Coyle KP, et al. (2016). Polygenic sex determination in the cichlid fish *Astatotilapia burtoni*. *BMC Genomics* 201617:835. doi: 10.1186/s12864-016-3177-1
- Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Gen Biol* 11(3): R25. doi: 10.1186/gb-2010-11-3-r25
- Robinson JT, Thorvaldsdóttir H, Winckler W, et al. (2011) Integrative Genomics Viewer. *Nat Biotech* 29;(1): 24-26. doi: 10.1038/nbt.1754
- Saavedra Rojas E, Quintero Pinto L, Landines Parra M (2005) Aspectos reproductivos. *Biología y cultivo del pirarucú Arapaima gigas* (Schinz, 1822) (Pisces: Arapaimidae): Aspectos reproductivos. Bases para un aprovechamiento sostenible. INCODER-UNC, Bogotá, Colombia
- Saint-Paul, U. (1986). Potential for aquaculture of South American freshwater fishes: a review. *Aquaculture*, 54(3), 205-240. doi: 10.1016/0044-8486(86)90329-7
- Sandoval, M. (2007). Aspectos de manejo, reproducción y alimentación del paiche *Arapaima gigas* en la Amazonía peruana. IIAP/BIODAMAZ (Perú-Finlandia). Documento técnico, 8.
- Schaefer, S., Kloas, W. & Wuertz, S. (2012) *Arapaima*: candidate for intensive freshwater culture. *Global Aquaculture Advocate* 15, 50–51
- Simão FA, Waterhouse RM, Ioannidis P, et al. (2015) BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–2. doi: 10.1093/bioinformatics/btv351
- Simbolo M, Gottardi M, Corbo V, et al. (2013) DNA Qualification Workflow for Next Generation Sequencing of Histopathological Samples. *PLoS ONE* 8(6), e62692. doi: 10.1371/journal.pone.0062692
- Simpson JT, Wong K, Jackman SD et al. (2009) ABySS: A parallel assembler for short read sequence data. *Gen Res* 19(6): 1117-1123. doi: 10.1101/gr.089532.108
- Smeds L, Kawakami T, Burri R, et al. (2014) Genomic identification and characterization of the pseudoautosomal region in highly differentiated avian sex chromosomes. *Nat Comm* 5: 5448. doi: 10.1038/ncomms6448
- Sun S, Li W, Xiao S, et al. (2018) Genetic sex identification and the potential sex determination system in the yellow drum (*Nibea albiflora*). *Aquaculture* 492: 253-258. doi: 10.1016/j.aquaculture.2018.03.042
- Thompson M.M., Coe B.H., Andrews R.M., Stauffer D.F., Cristol D.A., Crossley D.A., Hopkins W.A. (2018) Major global changes interact to cause male-biased sex ratios in a reptile with temperature-dependent sex determination. *Biol Conserv* 222: 64-74. doi: 10.1016/j.biocon.2018.03.006
- Torati, L. S., Migaud, H., Doherty, M. K., Siwy, J., Mullen, W., Mesquita, P. E., & Albalat, A. (2017). Comparative proteome and peptidome analysis of the cephalic fluid secreted by *Arapaima gigas* (Teleostei: Osteoglossidae) during and outside parental care. *Plos ONE*, 12(10), e0186692. doi: 10.1371/journal.pone.0186692
- Trapnell C, Williams BA, Pertea G, et al. (2010) Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nat Biotech*. 28(5):511-515. doi: 10.1038/nbt.1621
- Val, A. L., de Almeida-Val, V. M. (2012). *Fishes of the Amazon and their environment: physiological and biochemical aspects* (Vol. 32). Springer Science & Business Media.
- Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., Schatz, M.C. (2017) GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33(14): 2202-2204. doi: 10.1093/bioinformatics/btx153
- Vialle, R. A., de Souza, J., Lopes, K. P., Teixeira, D. G., Alves Sobrinho, P. A., Ribeiro-Dos-Santos, A. M., et al (2018). Whole Genome Sequencing of the Pirarucu (*Arapaima gigas*) Supports Independent Emergence of Major Teleost Clades. *Genom Biol Evol* 10(9), 2366–2379. doi: 10.1093/gbe/evy130
- Watanabe L, Gomes F, Vianez J, Nunes M, Cardoso J, Lima C, et al. (2018) De novo transcriptome based on next-generation sequencing reveals candidate genes with sex-specific expression in *Arapaima gigas* (Schinz, 1822), an ancient Amazonian freshwater fish. *PLoS ONE* 13(10): e0206379. doi: 10.1371/journal.pone.0206379
- Weisenfeld, N.I., Yin, S., Sharpe, T., Lau, B., Hegarty, R., Holmes, L., Sogoloff, B., et al. (2014) Comprehensive variation discovery in single human genomes. *Nat Genet* 46(12): 1350-5. doi: 10.1038/ng.3121
- Yamaguchi S. and Iwasa Y. (2018) Temperature-dependent sex determination, realized by hormonal dynamics with enzymatic reactions sensitive to ambient temperature. *J Theor Biol* 453: 146-155. doi: 10.1016/j.jtbi.2018.05.023
- Zimin AV, Puiu D, Luo M-C, et al. (2017) Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Res* 27(5): 787-792. doi:10.1101/gr.213405.116