Fine mapping using whole-genome sequencing confirms anti-Müllerian hormone as a major gene for sex determination in farmed Nile tilapia (*Oreochromis niloticus* L.)

Giovanna Cáceres[1], María E. López [1,2], María I. Cadiz[1], Grazyella M. Yoshida[1,3], Ana Jedlicki[1], Ricardo Palma-Véjares[4,5], Dante Travisany[4,5], Diego Díaz-Domínguez[4,5], Alejandro Maass[4,5], Jean P. Lhorente[3], Jose Soto[6], Diego Salas[6], José M. Yáñez [1,7*]

[1] Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, Santiago, Chile.

[2] Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden.

[3] Benchmark Genetics Chile, Puerto Montt, Chile

[4] Centro para la Regulación del Genoma, Universidad de Chile, Santiago, Chile

[5] Centro de Modelamiento Matemático UMI CNRS 2807, Universidad de Chile, Santiago, Chile

[6] Grupo Acuacorporación Internacional (GACI), Cañas, Costa Rica

[7] Núcleo Milenio INVASAL, Concepción, Chile

Correspondence:

Dr. José Manuel Yáñez

jmayanez@uchile.cl

ABSTRACT

Nile tilapia (*Oreochromis niloticus*) is one of the most cultivated and economically important species in world aquaculture. Faster male development during grow-out phase is considered a major problem that generate heterogeneous sizes of fish at harvest. Identifying genomic regions associated with sex determination in Nile tilapia is a research topic of great interest. The objective of this study was to identify genomic variants associated with sex determination in three commercial populations of Nile tilapia. Whole-genome sequencing of 326 individuals was performed, and a total of 2.4 million high-quality bi-allelic single nucleotide polymorphisms (SNPs) were identified. A genome-wide association study (GWAS) was conducted to identify markers associated with the binary sexual trait (males = 0; females = 1). A mixed logistic regression GWAS model was fitted and a genome-wide significant signal comprising 36 SNPs, located on chromosome 23 spanning a genomic region of 536 kb, was identified. Ten out of these 36 genetic variants, intercept the anti-Müllerian hormone gene. Other significant SNPs were located in the neighboring *Amh* gene region. This gene has been strongly associated with sex determination in several vertebrate species, playing an essential role in the differentiation of male and female reproductive tissue in early stages of development. This finding provides useful information to better understand the genetic mechanisms underlying sex determination in Nile tilapia.

Keywords: SNP, sex control, quantitative trait loci, WGS, GWAS

INTRODUCTION

In aquaculture, many fish species of commercial interest exhibit sexual dimorphism in a variety of economically important traits such as growth rate, age at sexual maturity and carcass quality traits (Díaz et al., 2013; Martínez et al., 2014; Purcell et al., 2018). When the sexual dimorphism is relevant for production, the identification of genomic regions or markers associated to these traits is of great interest as they can be used to develop more efficient methodologies for generating monosex population. For instance, in Nile tilapia, the current methods for monosex (all-male) production populations only rely on the use of hormones (Baroiller et al., 2001; Beardmore et al., 2001; El-Sayed et al., 2012; Alcántar et al., 2014).

Teleost fish have developed a variety of sex determination mechanisms, including i) strict control due to genetic factors, ii) control by environmental factors only, or iii) interactions of both factors (Barroiller et al., 2006; Kijas et al., 2018). Genetic sex-determining systems may be chromosomal which involve one gene or master region involved in sex determination, or they may be polygenic involving several genes or multiple genomic regions (Martínez et al., 2014; Palaiokostas et al., 2015). Genomic regions associated with sex determination have been identified in some aquaculture species, including chinook salmon (*Oncorhynchus tshawytscha*) (Devlin et al., 2001), rainbow trout (*Oncorhynchus mykiss*) (Felip et al., 2005), yellow catfish (*Pelteobagrus fulvidraco*) (Wang et al., 2009), and Atlantic salmon (*Salmo salar*) (Kijas et al., 2018). In recent years, at least five genes have been identified as key factors in the gonadal differentiation pathway. For instance, the *Dmy* gene regulates sex differentiation in medaka (*Oryzias latipes*) (Matsuda et al., 2002), *Amhr2* in puffer fish fugu (*Takifugu rubripes*) (Shirak et al., 2006), *Amhy* in pathogonic silverside (*Odontesthes hatcheri*) (Hattori et al., 2012), *gsdf Y* in Luzon ricefish (*Oryzias luzonensis*) (Myosho et al., 2012),, and *sdy* - *irf9* in rainbow trout (*Oncorhynchus mykiss*) (Yano et al., 2012a). The first four genes have been implicated in the signaling pathways for sexual differentiation of vertebrates (Herpin et al., 2005; Pan et al, 2016), while the *sdY* gene described for rainbow trout has been proposed as the master gene for sex differentiation in salmonids, which evolved from the immune system-related i*rf9* gene and participates in the modulation of the interferon-9 signaling pathway (Yano et al., 2012b; Pan et al, 2016).

In Nile tilapia, it is suggested that genetic control for sex is determined primarily by a heterogamous XX/XY male system (Beardmore, 2001). However, other genetic factors and environmental variables such as temperature may intervene in sex determination (Baroiller et al., 2001; Cnaani et al., 2008;

Palaiokostas et al., 2013, Wessels et al., 2014; Eshel et al., 2014). To date, different sex-linked genomic regions have been identified in Nile tilapia, including associated regions in linkage groups (LG) 1, 3, 20, and 23 (Lee et al., 2003, 2005; Shirak et al., 2006; Eshel et al., 2010; Cnaani 2013; Polaikostas et al., 2015). Most of the studies published to date have reported that the sex-determining region would be found in linkage group 1 (Lee et al., 2003; Ezaz et al., 2004; Lee et al., 2005; Lee et al., 2011; Palaiokostas et al., 2013; Palaiokostas et al., 2015). The presence of genes involved in the cascade of sexual differentiation of vertebrates have been described and mapped in this region, including Wilms tumor suppressor protein 1b (*wt1b*) and cytochrome P450 family 19 subfamily A member 1 (*cyp19a*) (Lee et al., 2007). The latter is a strong candidate for involvement in sex determination as its final product is the aromatase enzyme, which plays a crucial role in ovarian differentiation in vertebrates (Herpin et al., 2005; Ma et al., 2016).

Through quantitative trait loci (QTL) mapping by using linkage analysis Eshel et al. (2010-2012) identified a sex-determining region in LG23, which hosts the anti-Müllerian hormone (*Amh*) and the doublesex- and mab-3 related transcription factor 2 (*Dmrt2*) genes (Shirak et al., 2006). The *Amh* gene is the mediator of the regression of Müller's ducts in mammals, birds, and reptiles (Rehman et al., 2017). Müller's ducts are responsible for the development of the uterus and fallopian tubes in females during fetal development (Jamin et al., 2003; Pfenning et al., 2015), while the *Dmrt2* gene, which is a member of the *Dmrt* family of transcription factors has been suggested as an essential regulator of male development in vertebrates (Herpin et al., 2015).

The multiple sex-determining regions described for Nile tilapia support the evidence that sex differentiation is a complex trait, and it is not yet clear which specific putative causative variants are involved in regulating sex differentiation in this species. In this study, we perform a genome-wide association analysis for phenotypic sex using genotypes from a whole-genome resequencing experiment performed in 326 fish belonging to three different commercial populations of Nile tilapia. Our results provide further evidence that the *Amh* gene is a major gene associated with sex determination in this species.

MATERIALS AND METHODS

Fish

For the present study we used individuals from three commercial breeding populations established in Latin America, which are directly or indirectly related to Genetically Improved Farmed Tilapia (GIFT); the most widely cultivated Nile tilapia strain in the world. The GIFT strain was initially established in the Philippines in 1980 by the World Fish Center to initiate the first breeding program in Nile tilapia. The strain was originated using crosses between four Asian cultured strains from Israel, Singapore, Taiwan and Thailand and four strains from wild populations captured across the natural distribution of this species in Africa (Egypt, Senegal, Kenya, and Ghana) (Neira et al., 2016).

We used 59 samples from POP_A breeding population belonging to AquaAmerica (Brazil), and 126 and 141 samples from the POP_B and POP_C breeding populations, respectively, both belonging to Acuacorporación Internacional (ACI, Costa Rica). The Brazilian strain used to establish POP_A was introduced to Brazil from a Malaysian breeding population in early 2005 for breeding and production purposes. The POP_B population was generated with individuals from Asian populations in Israel, Singapore, Taiwan and Thailand present in the Philippines in the late 1980s. The POP_B strain was imported to Costa Rica by ACI in 2005 from the aquaculture station Carmen Aquafarm (Philippines). The POP_C was established in Philippines from the mixture of genetic material of the best available stock of GIFT populations with two original African strains that founded GIFT.

DNA extraction and whole-genome sequencing

Genomic DNA was extracted from a total of 326 fish samples, using the Wizard Genomic DNA purification kit, (Promega) according to manufacturer's specifications. DNA quality was evaluated by agarose gel electrophoresis and quantified by a Qubit fluorimeter (Thermo Scientific, USA). After normalization, the sequencing libraries were prepared and barcoded with the 200-cycle TruSeq sample preparation kit in pair-end format and sequenced through 66 lanes of an Illumina HiSeq 2500 machine (Illumina, USA) by a commercial supplier.

Variant discovery and filtering of SNPs

The SNP calling workflow was carried out as described in by Yáñez et al. (submitted). Briefly, the sequencing reads of each sample was quality controlled using FASTQC (Andrews, 2014), and then aligned to the Nile tilapia genome (Conte et al., 2017) using the BWA *mem* (Li et al., 2009; Li et al.,

2010) tool (predefined parameters). BAM files generated with BWA were further processed with the GATK pipeline (https://www.broadinstitute.org/gatk/) (McKenna et al., 2010) in order to get the set of raw SNPs.

Raw SNPs were filtered out using Vcftools software v. 0.1.15 (Denecek et al., 2011). All INDELs were removed, and SNPs that did not meet the following criteria were discarded: (1) Quality > 40, (2) non-biallelic and (3) percentage of missing genotypes > 50% across all individuals. Additionally, the following filters were applied using the GenAbel R package (Aulchenko et al., 2007): (1) minor allele frequency (MAF) < 0.05, (2) Hardy-Weinberg equilibrium (HWE) p-value > 1x 10-9, (3) SNP call rate < 0.90. Finally, samples with more than 80% of missing genotypes were also discarded. The genomic regions containing the filtered SNPs were remapped to the actual reference assembly (GenBank accession GCF_001858045.2), using the protocol provided in Yáñez et al. (submitted).

Basic population genetic statistics and differentiation

The population genetic diversity and differentiation was investigated among the three populations. First, genetic differentiation between populations was measured with pairwise F_{ST} (Weir & Cockerham'S F_{ST}) estimates using Vcftools software v. 0.1.15 (Danecek et al. 2011). Second, an individual-based principal component analysis (PCA) was carried out using PLINK v1.9 (Purcell et al., 2007). Finally, the nucleotide diversity was estimated using Vcftools. We used 20 kb genomic bins with a 10 kb step window (--window-pi 20000 --window-pi-step 10000) (Kijas et al., 2018). Genetic differentiation between male and female sub-populations was analyzed using F_{ST} estimates throughout the genome using filtered SNP variants and within the region involved in the sexual determination, obtained from the association analysis, using all unfiltered SNP variants. The heterozygosity of each SNP in the sex-associated region was assessed using PLINK v1.9 (Purcell et al., 2007).

Genome-wide association study

GWAS was conducted using the GenAbel R package (Aulchenko et al., 2007). The phenotype for sex determination was recorded as “0” for female and “1” for male. To identify the association between SNPs and sex-determining region in Nile tilapia, a mixed logistic regression model was used, accounting for the binary nature of the sex trait (male/female). The general formula used for the logistic regression model is as follows:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 * SNP + \beta_2 * S}}{1 + e^{\beta_0 + \beta_1 * SNP + \beta_2 * S}}$$

Where, $\pi(x)$ corresponds to the probability that the phenotype is male given, β_0: intercept, β_1*SNP is the SNP effect, β_2*S is the effect of the Nile tilapia strain (with three levels). The -log10 (p-value) for each SNP across the genome was plotted to summarize the GWAS results. The significance threshold was determined by Bonferroni correction. A SNP was considered significant if its p-value was < 0.05/N, where N is the number of total markers analyzed in the GWAS. At the chromosome level a SNP was considered significant if its p-value was <0.05/Nc, where Nc is the number of markers on each particular chromosome. The proportion of heritability explained by each significant marker was obtained by comparing estimated heritability with polygenic function and estimated heritability with the inclusion of the significant SNP genotype as a factor (Korte and Farlow, 2013).

RESULTS

Quality control

The whole-genome sequencing (WGS) and posterior alignment of the 326 fish generated an average of 76.9 million raw reads (SD = 65.0 million reads) and 76.3 million mapped reads (SD = 64.6 million mapped reads). The average coverage for each individual was 8.7X (SD = 8.9 X). From the discovery phase a total of 38,454,943 genetic variants were identified in the 326 animals. After the quality control steps, which included discarding indels, low-quality variants, exclusion of SNPs other than biallelic and genotypes called below 50% across all individuals, a total of 4,762,199 SNPs were retained. After QC including MAF, HWE, SNP and sample call rate filtering, a total of 2.4 million high-quality bi-allelic SNPs were retained in 302 samples (144 females and 158 males) for further analysis.

Basic population statistics and genetic structure

The population genetic structure was explored through Principal Component Analysis (PCA). The first two principal components represented 25% of the genetic variation. A clear differentiation is observed among the three populations analyzed in this study (Figure 1A). The PCA allowed us to distinguish between the farmed populations from Brazil (POP_A) and Costa Rica (POP_B and POP_C), and also between the two populations from the latest location. The global F_{ST} among the three commercial

208 populations was 0.049. The lowest genetic differentiation was observed between the POP_B and POP_C
209 (F_{ST} 0.044). The analysis of nucleotide diversity among the populations revealed that the Costa Rica
210 populations (POP_B average $\pi = 8.78 \times 10\text{-}4$; POP_C $\pi = 8.71 \times 10\text{-}4$), are slightly less diverse than the
211 POP_A population (average $\pi = 7.93 \times 10\text{-}4$) (Figure 1B).
212
213 **Genome-wide association study**
214 A genome-wide significant association signal was detected in a genomic region within chromosome 23
215 (see Figure 2). The genomic region strongly associated with phenotypic sex identified by the mixed
216 logistic regression model revealed 36 genome-wide significant SNPs associated with sex determination
217 in Nile tilapia. The significant SNPs were located in a single genomic region which spanned ~ 536 kb in
218 linkage group 23 (Table 1). The proportion of the genetic variance explained for phenotypic sex based
219 on the significantly associated SNPs present in this region ranged from 0.4 to 0.7 (Table 1).
220
221 The genomic region of ~ 536 kb in linkage group harboring SNPs significantly associated with
222 phenotypic sex, contains about 30 annotated genes. Some are strong candidates to play an important role
223 in sex determination in Nile tilapia. More interesting, is that the three most significantly associated SNPs
224 are located within the anti-Müllerian hormone gene (*Amh*), which has been linked to the differentiation
225 process of male and female reproductive tissue in early stages of development of vertebrates and various
226 fish species (Shirak et al., 2006; Hattori et al., 2012; Pan et al., 2016). Moreover, SNPs
227 NC_031986.2:34500823, NC_031986.2:34500954 and NC_031986.2:34501082 are found in a region of
228 the second exon of the *Amh* gene. Markers NC_031986.2:34502582 and NC_031986.2:34502748 also
229 intercept the anti-Mullerian hormone gene in exon 7, while NC_031986.2:34502034 does so in exon 5
230 of the *Amh* gene. The SNP NC_031986.2:34590018 is found in an intronic region of the Elongation
231 Factor *ELL* (Eleven-Nineteen Lysine-Rich Leukemia), which has been described as a selective co-
232 regulator for steroid receptor functions (Pascual-Le Tallec et al., 2005; Zhou et al., 2009). SNP
233 NC_031986.2:34433907 is found downstream of the *Pias4* gene, protein inhibitor of activated STAT
234 (signal transducer and activator of transcription), which inhibits the LRH-1 receptor (liver receptor
235 homolog-1), a gene abundantly expressed in the ovary and shown to activate the transcription of steroid
236 genes including *Cyp11a1* in granulose cells (Hsieh et al., 2009).
237
238 **Genetic and heterozygosity differentiation between males and females**
239 The overall F_{ST} estimate between males and females across all populations was 0.0003, indicating a very
240 low genetic differentiation between both sub-populations. In the 536 Kb genomic region associated to

the sex determination located in chromosome 23 identified in this study the average F_{ST} was one order of magnitude higher (0.004) (see Figure 3). These results suggest that there is a higher degree of genetic differentiation between male and female sub-populations in this genomic region compared with the average genetic differentiation across the whole genome.

To search for highly heterozygote loci in males (potentially XY) compared to females (potentially XX), the average heterozygosity differences (AHD) between males and females were locally calculated for each unfiltered SNP in the associated region within linkage group 23. Interestingly, we found co-localization of highly heterozygote variants in males when compared to females in the region harboring the most significant SNPs within and near the *Amh* gene (Figure 4)

DISCUSSION

The main objective of the present study was to identify genomic regions involved in sex determination in Nile tilapia by using a genome-wide association analysis approach based on a high density SNP panel derived from whole-genome sequencing of fish from different commercial populations. Previous studies have emphasized the complexity of determining and identifying the genomic regions associated with sex determination in Nile tilapia (*Oreochromis niloticus*). For instance, most of the markers reported to be associated with phenotypic sex have been located in different linkage groups including LG1, LG33, LG20, and LG23 (Lee et al., 2003; Ezaz et al., 2004; Lee et al., 2011; Eshel et al., 2010; Eshel et al., 2012; Palaiokostas et al., 2015). Palaiokostas et al. (2013) concluded that one of the main limitations of studies to detect sex-determining regions in *Oreochromis niloticus* was the limited number of genetic markers used in previous studies. Also, these studies only considered specific farmed Nile tilapia strains (Lee et al., 2003; Ezaz et al., 2004; Lee et al., 2005; Eshel et al., 2010).

In this study we analyzed three different farmed Nile tilapia populations which shown a low level of genetic differentiation based on the global F_{ST} estimate. This low differentiation is probably because the three Nile tilapia populations share common ancestors and are related due to the common origin of GIFT strain. The lowest genetic differentiation was observed between the POP_B and POP_C (F_{ST} 0.044), which is concordant with the common geographical origin of these farmed populations (Costa Rica). However, the three populations clustered in three clear different groups in the PCA, indicating a patent genetic structure among the three populations.

The approach used in this study, which was based on whole-genome sequencing of males and females from three different farmed populations provided a highly dense distribution of markers covering the whole genome and a good representation of the genetic variation of this species, which in turn gave an appropriate statistical power to identify significant markers associated with the trait of interest. GWAS revealed a single genomic region in linkage group 23 that is associated with sex determination in Nile tilapia. These results are consistent with those previously reported by other authors, who have suggested that the sex determination region in Nile tilapia would be in linkage group 23 (Shirak et al., 2006; Cnaani et al., 2008; Lühmann et al. 2012; Wessels et al., 2014). Eshel et al. (2010, 2014), described a sex-associated QTL by microsatellite markers in this linkage group. While Joshi et al. (2018), using a panel of 58K SNP markers, reported that the most likely position of the sex locus would be found in the LG23. None of these studies have either the sufficient resolution or statistical power to narrow down the region to a single gene. In this study, we provide consistent evidence that polymorphism within or nearby *Amh* gene can be partially controlling sex determination in Nile tilapia.

The GWAS we performed identified 36 markers significantly associated with sex on a genome-wide basis in a single genomic region in chromosome 23. Seven markers intercept the anti-Müllerian hormone gene, which mediates the regression of Müller's ducts in several vertebrate species. The Müller ducts are responsible for the development of the uterus and fallopian tubes in females during fetal development (Jamin et al., 2013; Pfenning et al., 2015). The anti-Müllerian hormone gene has previously been considered a candidate gene for sex determination in Nile tilapia, as it is found in the sex locus (Eshel et al., 2014). Li et al. (2015) isolated a specific duplicate of *Amh*, which was designed as *Amhy*, located immediately downstream of *Amh*, whose expression was identified in males with genotype XY - YY and only during the critical sex determination period in Nile tilapia.

Some authors, including Ijiri et al. (2008) and Eshel et al. (2014), have evaluated the gene expression of the *Amh* gene during gonadal differentiation and early development in undifferentiated Nile tilapia gonads from 2 days post-fertilization; concluding that from day 5 post-fertilization the expression of the *Amh* gene increases sharply in male gonads until day 35 post-fertilization, demonstrating a crucial role of this gene in sexual differentiation. This overexpression of *Amh* in male gonads in the early stages of development has been reported in other fish species such as Japanese sole (*Paralichthys olivaceus*) (Yoshinaga et al., 2004). In the same way, the *Amh* gene in fugu pufferfish (Shirak et al., 2006) and

pathogonic silverside (Hattori et al., 2012) is considered the determinant genetic sex marker for both species.

It is interesting to note that the seven SNPs that intercept the *Amh* gene explain a relatively high proportion of genetic variance. SNP NC_031986.2:34501082 intercepts the second exon in the *Amh* gene and explains 0.69 of the additive genetic variance, which confirms that the *Amh* gene is a major gene controlling sex differentiation in Nile tilapia. Taking into account the relatively high proportion of genetic variance explained by the significant markers, the findings obtained in this study suggest an important potential to incorporate molecular information to optimize methods for producing all-male populations, thereby decreasing the use of hormones that is currently the only methodology that can produce over 99% of through confirmation of genotypic sex of sex reversed animals.

CONCLUSIONS

This study provides further evidence to better understand the genetic architecture of sex determination in commercial Nile tilapia in strains established in Latin America. Because there is one genomic region explaining a high proportion of the genetic variance associated to phenotypic sex our results indicate that the Nile tilapia sex determination trait can be considered an oligogenic trait. Seven SNPs highly associated with sex determination intercept the anti-Müllerian hormone gene (*Amh*) providing strong evidence of *Amh* being a major gene controlling sex differentiation in Nile tilapia farmed populations.

Acknowledgements

This study was funded by CORFO grant 14EIAT-28667. GC, MIC & MEL want to acknowledge the Nacional Commission of Scientific and Technologic Research (CONICYT) for the funding through the Nacional PhD funding program (21150346-21171369). AM would like to acknowledge Conicyt-PIA Program AFB 170001 y Fondap Nº 15090007 grants. Powered@NLHPC: This research was partially supported by the supercomputing infrastructure of the NLHPC (ECM-02). The authors would like to acknowledge to Aquacorporación Internacional and AquaAmerica for providing the samples used in this study. We would like to thank to Gabriel Rizzato and Natalí Kunita from AquaAmerica for their kind contribution of samples from Costa Rica and Brazil, respectively. JMY is supported by Núcleo Milenio INVASAL funded by Chile's government program, Iniciativa Científica Milenio from Ministerio de Economía, Fomento y Turismo.

Competing Interest

Jean-Paul Lhorente and Grazyella Yoshida were employed by company Benchmark Genetics Chile during the course of the study. All other authors declare no competing interests.

Data Availability

The sequencing data will be deposited in public database upon acceptance.

Ethics Statement

DNA sampling was carried out in accordance with the commercial practice and norms by Aquacorporación Internacional and Aquamerica.

Abbreviations

AMH	Anti-Müllerian hormone
GIFT	Genetically Improved Farmed Tilapia
FST	Weir & Cockerham'S F_{ST}
LG	Linkage group
GWAS	Genome Wide Association Study
PCA	Principal component analysis
SNP	Single nucleotide polymorphism

Authors' contributions

GC performed DNA extractions, GWAS analysis and wrote the initial version of the manuscript. MEL and GMY contributed with genetic analyses and writing. MIC and AJ contributed with sample processing DNA extractions and writing. AM, DD, DT and RP contributed with bioinformatic analysis and writing. JPL contributed with study design and writing. JS and DS contributed with sampling and phenotyping. JMY conceived and designed the study, supervised the work of GC and contributed with the analysis, discussion and writing.

Contributor Information

Giovanna Cáceres, Email: pdcaceres@gmail.com

María E. López, Email: me.lopez.dinamarca@gmail.com

Grazyella Yoshida, Email: grazyoshida@hotmail.com

María I. Cádiz, Email: mariaignaciacadizescobar@gmail.com

Ana Jedlicki, Email: jedlicki@gmail.com

Jean P. Lhorente, Email: jean.lhorente@bmkgenetics.com

Alejandro Maass, Email: amaass@dim.uchile.cl

Ricardo Palma, Email: rpalmavejares@gmail.com

Dante Travisany, Email: dtravisany@gmail.com

José M. Yáñez, Email: jmayanez@uchile.cl

José Soto, Email: jsoto@tilapia.com

Diego Salas, Email: dsalas@tilapia.com

Diego Díaz-Domínguez: diediaz@dcc.uchile.cl

References

Alcántar, JP .; Santos, C.; Moreno R.; Antonio, C. 2014. Manual para la Producción de supermachos de tilapia del Nilo (Oreochromis niloticus). UNPA- PIFI, Oaxaca. México. 81 pp.

Andrews, S. 2014. FastQC A Quality Control tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Aulchenko, Y.; De Koning, D.; Haley, C. 2007. Genome wide Rapid Association Using Mixed Model and Regression: A Fast and Simple Method for Genome-wide Pedigree□Based Quantitative Trait Loci Association Analysis. Genetics 177: 577□85.

Baroiller, J.; D'Cotta, H. 2001. Environment and sex determination in farmed fish. Comparative Biochemistry and Physiology. C: Comparative Pharmacology, 130: 399-409.

Baroiller, J.; D'Cotta, H.; Bezault, E.; Wessels, S.; Hörstegen-Schawark, G. 2006. Tilapia sex determination: Where temperature and genetics meet. Comparative Biochemistry and Physiology, Part A 153: 30–38.

Beardome, J.; Mair, G.; Lewis, R. 2001. Monosex male production in finfish as exemplified by tilapia: applications, problems, and prospects. Aquaculture 197: 283-301.

Cnnani, A.; Lee, B.; Zilberman, N.; Ozouf-Costaz, C.; Hulata, G.; Ron, M.; D'Hont, A.; Baroiller, J.; D'Cotta, H.; Penman, D.; Tomasio, E.; Coutanceau, J.; Perrey, E.; Shirak, A.; Kocher, T. 2008. Genetics of Sex Determination in Tilapine Species. Sex Dev 2:43-54.

404 Cnaani, A. 2013. The Tilapias Chromosomes Influencing Sex Determination. Cytogenet Genome Res 141:195–
405 205.
406 Conte, M.; Gammerdinge, W; Bartie, K.; Penman, D.; Kocher, T. 2017. A High-Quality Assembly of the Nile
407 Tilapia (Oreochromis niloticus) Genome Reveals the Structure of Two Sex Determination Regions.
408 bio.Rxiv 099.564;.doi: https://doi.org/10.1101/099564.
409 Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.; Banks, E.; Depristo, M.; Handsaker, R.; Lunter, G.; Marth, G.;
410 Sherry, S.; Mcvean, G.; Durbin, R. 2011. The variant call format and VCFtools. Bioinformatics (Oxford,
411 UK), 27:2156–2158.
412 Delvin, R.; Biagi, C.; Smailus, D. 2001. Genetic mapping of Y-chromosomal DNA markers in Pacific salmon.
413 Genetica 111: 43–58.
414 Díaz, N.; Ribas, L.; Piferrer, F. 2013. The relationship between growth and sex differentiation in the European sea
415 bass (*Dicentrarchus labrax*). Aquaculture 408: 191–202.
416 El-Saved, A.; Abdel-Aziz, E.S.; Abdel-Ghani, H. 2012. Effects of phytoestrogens on sex reversal of Nile tilapia
417 (*Oreochromis niloticus*) larvae fed diets treated with 17α-Methyltestosterone. Aquaculture 360–361: 58–63.
418 Eshel, O .; Shirak, A.; Weller, J.; Slossman, T.; Hulata, G.; Cnaani, A.; Ron, M. 2010. Fine-mapping of a locus on
419 linkage group 23 for sex determination in Nile tilapia (*Oreochromis niloticus*). Animal Genetics, 42: 222–
420 224.
421 Eshel, O.; Shirak, A.; Weller, J.; Hulata, G.; RON, M. 2012. Linkage and Physical Mapping of Sex Region on
422 LG23 of Nile Tilapia (*Oreochromis niloticus*), G3 2, pp. 35–42.
423 Eshel, O.; Shirak, A.; Dor. L.; Band. M.; Zak, T.; Markovich-Gordon, M.; Chalifa-Caspi, V.; Feldmesser, E.;
424 Weller, J.; Seroussi, E.; Hulata, G.; Ron, M. 2014. Identification of male-specific a*mh* duplication, sexually
425 differentially expressed genes and microRNAs at early embryonic development of Nile tilapia
426 (*Oreochromis niloticus*). BMC Genomics 15:774.
427 Ezaz, M.; Harvey, S.; Boonphakdee, C.; Teale, A.; Mcandrew, B.; Penman, D. 2004. Isolation and physical
428 mapping of sex-linked AFLP markers in Nile tilapia (*Oreochromis niloticus* L.). Mar. Biotechnol. 6: 435–
429 445.
430 Felip, A.; Young, W.; Wheeler, P.; Thorgaard, G. 2005. An AFLP-based approach for the identification of sex-
431 linked markers in rainbow trout (*Oncorhynchus mykiss*). Aquaculture 247: 35–43.
432 Hattori, R.; Murai, Y.; Oura, M.; Masuda , S.; Majhi, S. el al., 2012. A Y- linked anti-Müllerian hormone
433 duplication takes over a critical role in sex determination. Proc Natl Acad Sci USA 109: 2955–2959.
434 Herpin, A.; Schartl, M. 2015. Plasticity of gene-regulatory networks controlling sex determination: of masters,
435 slaves, usual suspects, newcomers, and usurpators. Embo Reports 16: 1260–1274.
436 Hsieh, H.; Wang, C.; Wu, M.; Yang, F.; Tai, Y el al. 2009. PIASy inhibits LRH-1-dependent CYP11A1
437 expression by competing for SRC-1 binding. Biochem. J. 419: 201–209.

438 Ijiri, S.; Kaneko, H.; Kobayashi, T.; Wang, D.; Sakai, F.; Prasanth, B.; Nakamura, M.; Nagahama, Y. 2008.
439 Sexual Dimorphic Expression of Genes in Gonads During Early Differentiation of a Teleost Fish, the Nile
440 Tilapia *Oreochromis niloticus*. Biology of Reproduction 78: 333-341.
441 Jamin, S.; Arango, N.; Mishina, Y.; Hanks, M.; Behringer, R. 2013.Genetic studies of the AMH/MIS signaling
442 pathway for Müllerian duct regression. Molecular and Cellular Endocrinology 211, 15–19.
443 Joshi, R.; Arnyasi, M.; Lien, S.; Gjoen, H.; Alvarez, A.; Kent, M. 2018. Development and validation of 58K SNP-
444 array and high-density linkage map in Nile tilapia (*O. niloticus*). doi: http://dx.doi.org/10.1101/322826.
445 Kijas, J.; Mcwilliam, S.; Sanchez, M.; Kube, P.; King, H.; Evans, B.; Nome, T.; Lien, S.; Verbyla, K. 2018.
446 Evolution of Sex Determination Loci in Atlantic Salmon. Scientific Reports 8:5664 DOI:10.1038/s41598-
447 018-23984.
448 Korte, A.; Farlow, A. 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant
449 Methods. 9:29.
450 Lee, B.; Penman, D.; Kocher, T. 2003. Identification of a sex-determining region in Nile tilapia (*Oreochromis*
451 *niloticus*) using bulked segregant analysis. Anim. Genet. 34: 379–383.
452 Lee, B.; Lee, W.; Streelman, J.; Carleton, K.; Howe, A. et al,. 2005. A second-generation genetic linkage map of
453 tilapia (Oreochromis spp.). Genetics 170: 237–244.
454 Lee, B., Kocher, T. 2007. Exclusion of Wilms tumors (WT1b) and ovarian cytochrome P450 aromatase
455 (CYP19A1) as candidates for sex determination genes in Nile tilapia (*Oreochromis niloticus*). Anim Genet
456 38: 85–86.
457 Lee, B.; Coutanceau, J.; Ozouf-Costaz, C.; D'Cotta, H.; Barroiller, JF.; Kocher, T. 2011. Genetic and Physical
458 Mapping of Sex-Linked AFLP Markers in Nile Tilapia (*Oreochromis niloticus*). Mar Biotechnol 13: 557–
459 562.
460 Li, H.; Durbin, R. 2009. Fast and accurate short read alignment with Burrows- Wheeler transform.
461 Bioinformatics: 25: 1754-1760.
462 Li, H.; Durbin, R. 2010. Fast and accurate long-read alignment with Burrows- Wheeler transform. Bioinformatics.
463 26: 589-595.
464 Li, M.; Sun, Y.; Zhao, J.; Shi, H.; Zeng, S.; Ye, K.; Jiang, D.; Zhou, L.; Sun, L.; Tao, W.; Nagahama, Y.; Kocher,
465 T.; Wang, D. 2015. A Tandem Duplicate of Anti-Müllerian Hormone with a Missense SNP on the Y
466 Chromosome Is Essential for Male Sex Determination in Nile Tilapia, *Oreochromis niloticus*. PLOS
467 Genetics DOI: 10.1371/journal.pgen.1005678.
468 Luhmann, L.; Knorr, C.; Horstgen-Schwark, G.; Wessels, S. 2012. First evidence for family-specific QTL for
469 temperature-dependent sex reversal in Nile tilapia (*Oreochromis niloticus*). Sex. Dev. 6: 247–256.
470 Ma, K.; Liao, M.; Liu, F.; Ye, B.; Sun, F.; Huayue, G. 2016. Charactering the ZFAND3 gene mapped in the sex-
471 determining locus in hybrid tilapia (Oreochromis spp.). Sci. Rep. 6, 25471; doi: 10.1038/srep25471.

Mckenna, A.; Hanna, M.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; Depristo, M. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20(9):1297-303.

Martínez, P.; Viñas, A.; 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.

Matsuda, M.; Nagahama, Y.; Shimomiya, A.; Sato, T.; Matsuda, C.; Kobayashi, T.; Morrey, C.; Shibata, N. et al., 2002. DMY is a Y- specific DM-domain gene required for male development in the medaka fish. Nature 417: 559–563.

Myosho, T.; Otake, H.; Masuyama, H.; Matsuda, M.; Kuroki, Y. et al. 2012. Tracing the emergence of a novel sex-determining gene in medaka, *Oryzias luzonensis*. Genetics 191: 163–170.

Neira, R.; García, X.; Lhorente, J.; Filp, M.; Yáñez, J.; Cascante, A. 2016. Evaluation of the growth and carcass quality of diallel crosses of four strains of Nile tilapia (*Oerochromis niloticus*). Aquaculture 451, 213–222.

Palaiokostas, C.; Bekaert, M.; Khan, M.; Taggart, J.; Gharbi, K.; Mcandrew, B.; Penman, D. 2013. Mapping and Validation of the Major Sex-Determining Region in Nile Tilapia (*Oreochromis niloticus* L.) Using Rad Sequencing. PLoS ONE 8(7): e68389. doi: 10.1371/journal.pone.0068389.

Palaiokostas, C.; Bekaert, M.; Khan, M.; Taggart, J.; Gharbi, K.; Mcandrew, B.; Penman, D. 2015. A novel sex-determining QTL in Nile tilapia (*Oreochromis niloticus*). BMC Genomics 16:171 DOI 10.1186/s12864-015-1383.

Pan, Q.; Anderson, J.; Bertho, S.; Herpin, A.; Wilson, C.; John, H.; Postlethawait, J.; Schartl, M.; Guiguen, Y. 2016. Vertebrate sex-determining genes play musical chairs**.** C R Biol. 339(7-8): 258–262. doi:10.1016/j.crvi.2016.05.010.

Pascual-Le Tallec, L.; Simone, F.; Viengchareun, S.; Meduri, G.; Thirman, M.; Lombe, M. 2005.The Elongation Factor ELL (Eleven-Nineteen Lysine- Rich Leukemia) Is a Selective Coregulator for Steroid Receptor Functions. Molecular Endocrinology 19(5):1158–1169.

Pfenning, F.; Standke, A.; Gutzeit, H. 2015. The role of *Amh* signaling in teleost fish – Multiple functions not restricted to the gonads. General and Comparative Endocrinology 223: 87–107.

Purcell, C.; Seetharam, A.; Snodgrass, O.; Ortega-Garcia, S.; Hyde, J.; Severin, A. 2018. Insights into teleost sex determination from the *Seriola dorsalis* genome assembly. BMC Genomics DOI 10.1186/s12864-017-4403

Rehman, Z.; Worku, T.; Davis, J.; Talpur, H.; Bhattarai, D.; Kadariya, I.; Hua, G.; Cao, J.; Dad, R.; Farmanullah, Hussain, T.; Yang, L. 2017. Role and mechanism of AMH in the regulation of Sertoli cells in mice. Journal of Steroid Biochemistry and Molecular Biology 174, 133–140.

Shirak, A.; Seroussi, E.; Cnaani, A.; Howe, A.; Domokhovsky, R.; Zilberman, N.; Kocher, T.; Hulata, G.; Ron, M. 2006. *Amh* and Dmrta2 Genes Map to Tilapia (*Oreochromis spp.*) Linkage Group 23 Within Quantitative Trait Locus Regions for Sex Determination. Genetics 174: 1573–1581.

Wang, D.; Mao, H.; Chen, H.; Liu, H.; Gui, J. 2009. Isolation of Y- and X-linked SCAR markers in yellow catfish and application in the production of all-male populations. Anim. Genet. 40: 978–981.

Wessels, S.; Sharifi, R.; Luehmann, L.; Rueangri, S.; Krause, I.; Pach, S.; Knorr, C.; Horstgen- Schawark, G. 2014. Allelic variant in the Anti-Müllerian Hormone gene leads to autosomal and temperature-dependent sex reversal in a selected tilapia line. PLOS ONE 9 (8), e104795.

Yano, A.; Guyomard, R.; Nicol, B.; Jouanno, E.; Quillet, E.; Klopp, C.; Cabau, C.; Bouchez, O. 2012a. An immune-related gene evolved into the master sex determining gene in rainbow trout, Oncorhynchus mykiss. Curr. Biol. 22: 1423–1428.

Yano, A.; Nicol, B.; Jouanno, E.; Quillet, E.; Fostier, A.; Guymard, R.; Guiguen, Y. 2012b. The sexually dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. Evolutionary applications vol. 6,3 (2012): 486-96.

Yoshinaga, N.; Shiraishi, E.; Yamamoto, T.; Iguchi, T.; Abe, S.; Kitano, T. 2004. Sexually dimorphic expression of a teleost homologue of Mullerian inhibiting substance during gonadal sex differentiation in Japanese flounder, *Paralichthys olivaceus*. Biochemical and Biophysical Research Communications 322: 508–513.

Zhou, J.; Feng, X.; Ban, B.; Liu, J.; Wang, Z el al. 2009 Elongation factor ELL (eleven-nineteen lysine-rich leukemia) acts as a transcription factor for direct thrombospondin-1 regulation. J. Biol. Chem. 284: 19142–19152.

Table 1. SNP and genes significantly associated with phenotypic sex in Nile tilapia (*Oreochromis niloticus*).

SNP	LG	Position (LG[1])	Binary sex (p-val[2])	PVAR[3]	Gene
NC_031986.2:34501082	LG 23	34501082	3.03e-13	0.691	*Amh*[4]
NC_031986.2:34502865	LG 23	34502865	1.28e-12	0.656	
NC_031986.2:34502864	LG 23	34502864	4.32e-12	0.625	
NC_031986.2:34510529	LG 23	34510529	8.98e-12	0.619	
NC_031986.2:34503003	LG 23	34503003	2.15e-11	0.596	
NC_031986.2:34502748	LG 23	34502748	2.43e-11	0.578	*Amh*
NC_031986.2:34502756	LG 23	34502756	3.79e-11	0.573	*Amh*
NC_031986.2:34491574	LG 23	34491574	6.54e-11	0.578	
NC_031986.2:34503006	LG 23	34503006	1.08e-10	0.550	
NC_031986.2:34512000	LG 23	34512000	1.38e-10	0.557	
NC_031986.2:34492336	LG 23	34492336	1.46e-10	0.540	
NC_031986.2:34500954	LG 23	34500954	2.65e-10	0.524	*Amh*
NC_031986.2:34492141	LG 23	34492141	3.29e-10	0.512	
NC_031986.2:34689299	LG 23	34689299	3.43e-10	0.537	
NC_031986.2:34510584	LG 23	34510584	3.88e-10	0.517	
NC_031986.2:34509215	LG 23	34509215	4.03e-10	0.531	
NC_031986.2:34500823	LG 23	34500823	4.61e-10	0.507	*Amh*
NC_031986.2:34501194	LG 23	34501194	5.84e-10	0.500	*Amh*
NC_031986.2:34501574	LG 23	34501574	7.71e-10	0.489	*Amh*
NC_031986.2:34502034	LG 23	34502034	8.58e-10	0.498	*Amh*
NC_031986.2:34491477	LG 23	34491477	1.18e-09	0.485	
NC_031986.2:34787936	LG 23	34787936	1.21e-09	0.496	
NC_031986.2:34502582	LG 23	34502582	1.53e-09	0.477	*Amh*
NC_031986.2:34500543	LG 23	34500543	1.58e-09	0.486	*Amh*
NC_031986.2:34433907	LG 23	34433907	3.42e-09	0.469	*Pias4*[5]
NC_031986.2:34510652	LG 23	34510652	3.64e-09	0.453	
NC_031986.2:34504122	LG 23	34504122	4.28e-09	0.467	
NC_031986.2:34505919	LG 23	34505919	6.91e-09	0.452	
NC_031986.2:34504349	LG 23	34504349	1.00e-08	0.444	
NC_031986.2:34503967	LG 23	34503967	1.08e-08	0.443	
NC_031986.2:34510699	LG 23	34510699	1.10e-08	0.428	
NC_031986.2:34969977	LG 23	34969977	1.48e-08	0.421	
NC_031986.2:34642986	LG 23	34642986	1.50e-08	0.419	
NC_031986.2:34596057	LG 23	34596057	1.60e-08	0.438	
NC_031986.2:34590018	LG 23	34590018	1.83e-08	0.419	*ELL*[6]
NC_031986.2:34502993	LG 23	34502993	2.07e-08	0.413	

[1] Linkage group

[2] P-value

545 [3] Proportion of the genetic variance explained by SNP

546 [4] Anti-Müllerian hormone

547 [5] Protein inhibitor of activated STAT 4

548 [6] Elongation factor for RNA polymerase II

549

550

551 **Figure 1. Population genetic structure of males and females, and nucleotide diversity from three**
552 **Nile tilapia (*Oreochromis niloticus*) farmed populations.** (A) Principal Components Analysis between
553 POP_A (green), POP_B (red) and POP_C (orange). Female are represented by unfilled squares and
554 males as triangles. (B) Nucleotide diversity of POP_A (red line), POP_B (green line) and POP_C (blue
555 line).
556
557 **Figure 2. Manhattan plot for GWAS results for sex determination measured as a binary trait**
558 **(male/female) in Nile tilapia (*Oreochromis niloticus*).** (A) The black line indicates the Bonferroni
559 corrected threshold for genome-wide significance. Evidence for genome-wide significant association
560 involving 36 SNPs is observed on chromosome 23. (B) The QQ-plot graph shows the relationship of the
561 normal theoretical quantiles of the probability distributions between the expected (x-axis) and observed
562 (y-axis) $-\log_{10}$(p-values) plotted for each SNP associated with sex determination in Nile tilapia (dots)
563 and the null hypothesis of no association (diagonal solid line).
564
565 **Figure 3. Genetic differentiation between male and female sub-populations in the sex-determining**
566 **region.** F_{ST} estimates between males and females within chromosome 23, dashed red line represents the
567 threshold for the 0.025% highest of the F_{ST} values.
568
569 **Figure 4. Regional plot of SNP associated with sex determination on chromosome 23.** (A)
570 Unfiltered SNPs are plotted by position on the chromosome (x-axis) and the average heterozygosity
571 difference (AHD) between males and females across all populations (y-axis). The significance (-
572 log10(P-value)) of SNPs associated with phenotypic sex (red dots) and harboring the *Amh* gene (gray
573 dots) is also shown (secondary y-axis). (B) The *Amh* gene and SNPs significantly associated with sex
574 determination.

575

Figure 1.

A)

B)

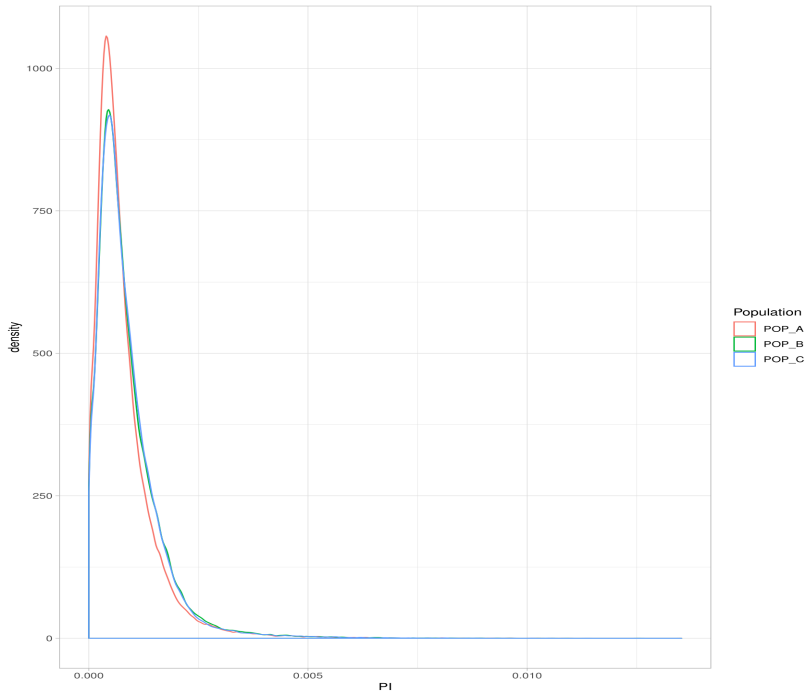
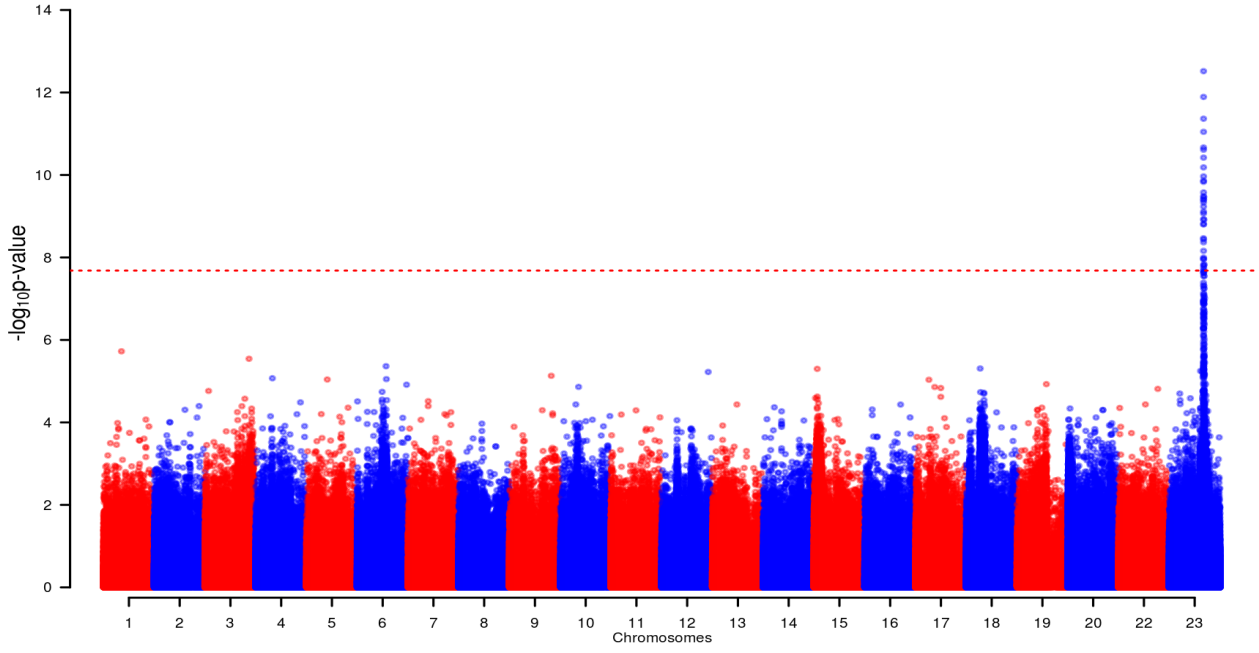


Figure 2.

A)



B)

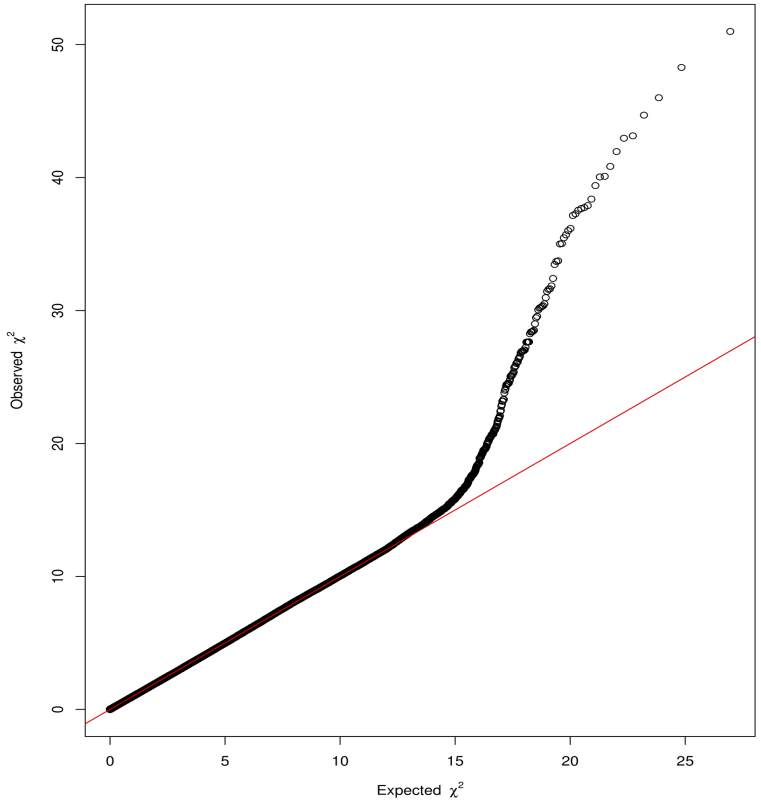


Figure 3.

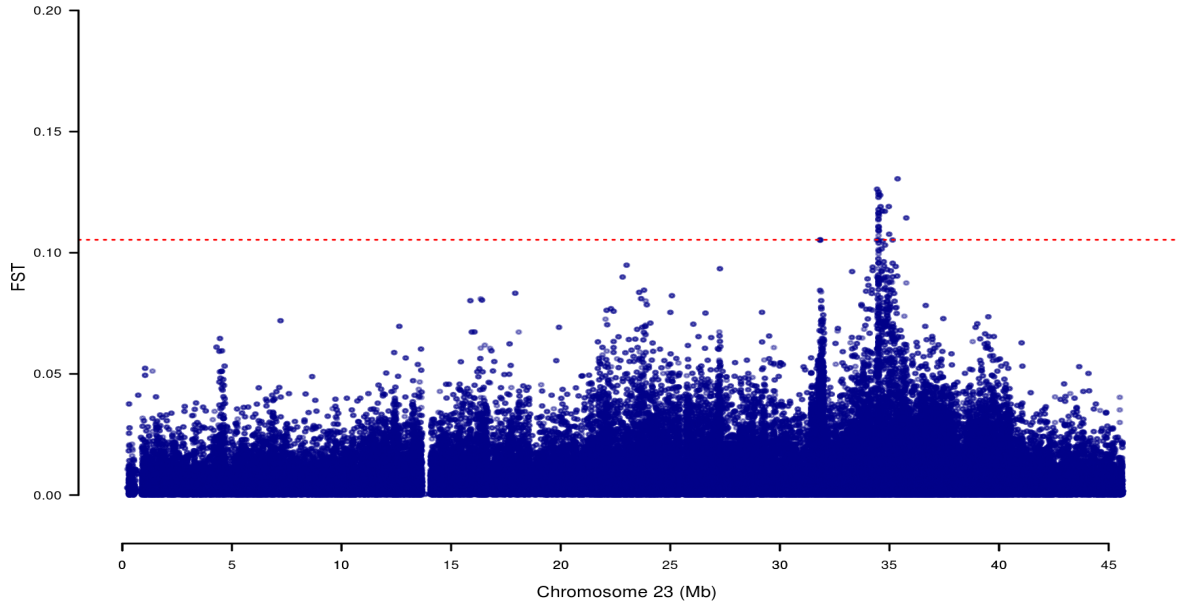


Figure 4.