

# Mapping and Validation of the Major Sex-Determining Region in Nile Tilapia (*Oreochromis niloticus* L.) Using RAD Sequencing

Christos Palaiokostas<sup>1</sup>✉, Michaël Bekaert<sup>1</sup>✉, Mohd G. Q. Khan<sup>1,2</sup>, John B. Taggart<sup>1</sup>, Karim Gharbi<sup>3</sup>, Brendan J. McAndrew<sup>1</sup>, David J. Penman<sup>1</sup>\*

**1** Institute of Aquaculture, School of Natural Sciences, University of Stirling, Stirling, Scotland, United Kingdom, **2** Department of Fisheries Biology and Genetics, Bangladesh Agricultural University, Mymensingh, Bangladesh, **3** The GenePool, School of Biological Sciences, University of Edinburgh, Edinburgh, Scotland, United Kingdom

## Abstract

Sex in *Oreochromis niloticus* (Nile tilapia) is principally determined by an XX/XY locus but other genetic and environmental factors also influence sex ratio. Restriction Associated DNA (RAD) sequencing was used in two families derived from crossing XY males with females from an isogenic clonal line, in order to identify Single Nucleotide Polymorphisms (SNPs) and map the sex-determining region(s). We constructed a linkage map with 3,802 SNPs, which corresponded to 3,280 informative markers, and identified a major sex-determining region on linkage group 1, explaining nearly 96% of the phenotypic variance. This sex-determining region was mapped in a 2 cM interval, corresponding to approximately 1.2 Mb in the *O. niloticus* draft genome. In order to validate this, a diverse family (4 families; 96 individuals in total) and population (40 broodstock individuals) test panel were genotyped for five of the SNPs showing the highest association with phenotypic sex. From the expanded data set, SNPs *Oni23063* and *Oni28137* showed the highest association, which persisted both in the case of family and population data. Across the entire dataset all females were found to be homozygous for these two SNPs. Males were heterozygous, with the exception of five individuals in the population and two in the family dataset. These fish possessed the homozygous genotype expected of females. Progeny sex ratios (over 95% females) from two of the males with the “female” genotype indicated that they were neomales (XX males). Sex reversal induced by elevated temperature during sexual differentiation also resulted in phenotypic males with the “female” genotype. This study narrows down the region containing the main sex-determining locus, and provides genetic markers tightly linked to this locus, with an association that persisted across the population. These markers will be of use in refining the production of genetically male *O. niloticus* for aquaculture.

**Citation:** Palaiokostas C, Bekaert M, Khan MGQ, Taggart JB, Gharbi K, et al. (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

**Editor:** Laszlo Orban, Temasek Life Sciences Laboratory, Singapore

**Received:** February 13, 2013; **Accepted:** May 29, 2013; **Published:** July 11, 2013

**Copyright:** © 2013 Palaiokostas et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The authors are grateful for support from the Marine Alliance for Science and Technology for Scotland, a Commonwealth PhD scholarship to MGQK (CSC reference BDCS-2007-28), and University of Stirling PhD scholarship to CP. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** One of the authors (BJM) is a shareholder and technical director of FishGen, a tilapia breeding and genetic improvement company. The authors do not think that there is a conflict with respect to the submitted manuscript as no relevant information from the research has been withheld. The authors are happy to discuss this further if requested. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

\* E-mail: d.j.penman@stir.ac.uk

✉ These authors contributed equally to this work.

## Introduction

Patterns of sex determination and differentiation in fish are very varied, with a wide range of gonochoristic and hermaphroditic species. Among the gonochoristic species, genetic and/or environmental factors determine sex [1–3]. The first sex-determining gene isolated in a fish species was *DMY/dmrt1bY* in *Oryzias latipes* (medaka) [4,5]. More recently, several other fish sex-determining genes have been isolated: *Gsdf(1)* in *Oryzias luzonensis* (Luzon ricefish) [6,7]; *amhy* in *Odontesthes hatcheri* (Patagonian pejerrey) [2,8]; *Amhr2* in *Takifugu rubripes* (tiger pufferfish) [1]; and *sdY* in *Oncorhynchus mykiss* (rainbow trout) [4]. Four of these genes come from those involved in sexual differentiation, while one derives from an immune-related gene (the *sdY* protein in *O. mykiss* is similar to part of the interferon regulatory factor 9). In all of

these species the sex determination system is principally XX/XY, but mismatches between sexual genotype and phenotype are not uncommon [6], sex reversal can be induced, YY males are viable in several species, and differentiated sex chromosomes are relatively uncommon [2]. The consequences of such variation are that sex determination genes in fish may likely require identification at the species or genus level. Furthermore, while we might search for genes from the sex differentiation pathway in a region where a sex determination gene has been mapped, it is by no means certain that we can easily identify a master switch gene.

Since many farmed species of fish exhibit sexual dimorphism in a range of traits of interest like growth or age at maturity, clarification of the sex-determining system of such fish is beneficial for the aquaculture industry towards the production of mono-sex stocks. *Oreochromis niloticus* (Nile tilapia) is one of the most important

farmed species with a production exceeding 2.8 million metric tonnes in 2010 [9]. Intensive commercial production generally requires all-male stocks, not only because males grow faster but also to avoid uncontrolled reproduction before harvest.

In tilapias, evidence so far suggests the existence of two different major sex-determining systems. In some tilapia species, including *O. niloticus* and *Oreochromis mossambicus* (Mozambique tilapia), sex is primarily determined by an XX/XY system on linkage group (LG) 1, whereas in others, for example *Oreochromis aureus* (blue tilapia), sex is primarily determined by a WZ/ZZ system on LG 3 [10]. However, other factors may influence sex determination and differentiation. In *O. niloticus*, genes on LG 3 [11] and LG 23 [12,13], and temperature [14] can affect sex ratio. Crosses between YY males and XX females generally give less than to 100% male progeny predicted from a simple XX/XY system [15]. Many of the studies on sex determination in *O. niloticus* have been carried out on fish derived from Lake Manzala in Egypt, the subject of the present study, and it is clear that both non-LG 1 genes and temperature affect sex ratios in at least some families in this population.

The current linkage map for *O. niloticus* is based on more than 500 markers, mostly microsatellites [16]. The sex-determining region has been previously mapped close to microsatellite *UNH995* on LG 1 [17]. This region contains two genes implicated in vertebrate sexual differentiation, *wt1b* and *cyp19a*, but further mapping ruled these out as candidates for the major sex-determining locus [18]. Restriction site associated DNA (RAD) sequencing [19] offers the possibility to construct much higher density linkage maps in a cost-efficient manner. In this study we used RAD sequencing to identify single nucleotide polymorphisms (SNPs) in two crosses between XY males and females from an isogenic clonal line. A genetic map was constructed based on 3,802 SNP markers. A quantitative trait locus (QTL) analysis was conducted based on these SNPs and was followed by an association analysis for the SNPs that showed the highest association with phenotypic sex using a diverse dataset of both family and population data. Altogether these data located the sex-determining QTL in a region of approximately 1.2 Mb on LG 1.

## Materials and Methods

### Sample Collection and Preparation

The fish used in this study came from the Tropical Aquarium Facilities of the Institute of Aquaculture at the University of Stirling. They originated from a population that was established in 1979 from Lake Manzala, Egypt (31°16'N, 32°12'E). All working procedures complied with the Animals Scientific Procedures Act [20]. Fish were reared in recirculating water systems at 27–28°C, and fed on commercial trout diet (Trouw Aquaculture Nutrition, UK; manufacturer Skretting, UK). To set up the families used in this study, mature females were held in glass aquaria and eggs were manually stripped following ovulation. Milt was manually stripped from male fish and used to fertilise the eggs *in vitro*. Eggs were incubated in downwelling incubators until the larvae had absorbed the yolk sac. Fry from families 1–6 (Table 1) were then transferred to tanks in recirculating systems and reared for 3–4 months before being killed and sexed by microscopic examination of the gonads [21]. A sample of fin tissue was taken and fixed in 100% ethanol for DNA extraction. Family 7 was split at yolk sac absorption: one group of 80 fry was reared at 36°C for ten days [22] in a static 5 L tank, while a control group (80 fry) was reared at 28°C. The survival of the two groups was 88% and 91% respectively. Subsequent rearing and sexing was as for families 1–6.

Families 1 and 2 (68 offspring and 20 offspring respectively) with dams from an isogenic XX clonal line and XY sires (as judged from balanced sex ratios in crosses to clonal line and outbred females and a high association between phenotypic sex and the LG 1 *UNH995* marker) were used to prepare RAD Libraries (Table 1). The available genome draft of *O. niloticus* is based on females from this clonal line. The sex associated SNPs were further validated by genotyping four further families with balanced sex ratios (Families 3–6: 24 offspring each) and broodstock (40 individuals; Table 1). These SNPs were finally used for genotyping a family in which elevated temperature induced a change in sex ratio (Family 7) to test whether the above SNPs could be useful in distinguishing neomales (XX males) from normal males. The sex ratio of the control group (reared at 28°C) did not show any deviations from the expected 1:1 ratio, while in the high temperature (36°C) treated group the proportion of males exceeded 96%.

### RAD Library Preparation and Sequencing

DNA was extracted from fin tissue of the fish using the REALPure genomic DNA extraction kit (Duzviz S.L.) and treated with RNase to remove residual RNA from the sample. Each sample was quantified by spectrophotometry (Nanodrop) and quality assessed by agarose gel electrophoresis, and was finally diluted to a concentration of 50 ng/μL in 5 mmol/L Tris, pH 8.5. The RAD library preparation protocol followed essentially the methodology originally described in Baird et al. [19] and comprehensively detailed in Etter et al. [23], with the minor modifications described in Houston et al. [24]. The RAD specific P1 and P2 paired-end adapters and library amplification PCR primer sequences used in this study are detailed in Baxter et al. [25].

Each sample (0.72 μg parental DNA/0.24 μg offspring DNA) was digested at 37°C for 40 minutes with *SbfI* (recognising the CCTGCA|GG motif) high fidelity restriction enzyme (New England Biolabs; NEB) using 6 U *SbfI* per μg genomic DNA in 1 × Reaction Buffer 4 (NEB) at a final concentration of c. 1 μg DNA per 50 μL reaction volume. The reactions (12 μL final volumes) were then heat inactivated at 65°C for 20 minutes. Individual specific P1 adapters, each with a unique 5 bp barcode (Table 1), were ligated to the *SbfI* digested DNA at 22°C for 60 minutes by adding 1.8/0.6 μL 100 nmol/L P1 adapter, 0.45/0.15 μL 100 mmol/L rATP (Promega), 0.75/0.25 μL 10 × Reaction Buffer 2 (NEB), 0.36/0.12 μL T4 ligase (NEB, 2 M U/mL) and reaction volumes made up to 45/15 μL with nuclease-free water for each parental/offspring sample. Following heat inactivation at 65°C for 20 minutes, the ligation reactions were slowly cooled to room temperature (over 1 hour) then combined in appropriate multiplex pools (Table S1). Shearing (Covaris S2 sonication) and initial size selection (250–550 bp) by agarose gel separation [24] was followed by gel purification, end repair, dA overhang addition, P2 paired-end adapter ligation, library amplification, exactly as in the original RAD protocol [19,23]. A total of 150 μL of each amplified library (16–18 PCR cycles, library dependent) was size selected (c. 350–650 bp) by gel electrophoresis [24]. Following a final gel elution step into 20 μL EB buffer (MinElute Gel Purification Kit, Qiagen), the libraries were sent to The GenePool Genomics Facility at the University of Edinburgh, UK, for quality control and high-throughput sequencing. Libraries were accurately quantified by qPCR (Kapa Library) and run in two lanes of an Illumina HiSeq 2000, one run using 100 base paired-end reads, the other 100 base single reads (v3 chemistry). Raw reads were processed using RTA 1.12.4.2 and Casava 1.6 (Illumina). The reads were deposited at the NCBI BioProject under the accession SRP017804.

**Table 1.** Summary of fish used in the study.

ID	Use	Sire strain	Dam strain	No. of females	No. of males	Total
Family1	RAD Libraries	Red <sup>†</sup>	Clonal	33	35	68
Family2	RAD Libraries	Red <sup>†</sup>	Clonal	10	10	20
Family3	SNP Assays	Red <sup>†</sup>	Wild*	12	12	24
Family4	SNP Assays	Red <sup>†</sup>	Wild*	12	12	24
Family5	SNP Assays	Wild*	Clonal	12	12	24
Family6	SNP Assays	Red <sup>†</sup>	Clonal	12	12	24
Family7 (28°C)	SNP Assays	Red <sup>†</sup>	Wild*	34	33	67
Family7 (36°C)	SNP Assays	Red <sup>†</sup>	Wild*	4	66	70
Clonal line	SNP Assays	–	–	0	2	2
Red strain	SNP Assays	–	–	6	15	19
Wild strain	SNP Assays	–	–	12	5	17

\*"wild" refers to wild type coloration;

<sup>†</sup>"red" refers to red body colour, which is controlled by a single gene.

doi:10.1371/journal.pone.0068389.t001

### Genotyping RAD Alleles

Reads of low quality (score under 30, while the average quality score was 37), missing the restriction site or with ambiguous barcodes were discarded. Retained reads were sorted into loci and genotyped using Stacks software 0.9995 [26]. The likelihood-based SNP calling algorithm [27] implemented in Stacks evaluates each nucleotide position in every RAD-tag of all individuals, thereby differentiating true SNPs from sequencing errors. A minimum stack depth of at least 30 and a maximum of 2 mismatches were allowed in a locus in an individual and up to 1 mismatch between alleles. The pair-ends were assembled using Stacks and Velvet version 1.2.08 [28] and used to separate RAD-tag sequence with or without potential SNP but belonging to separate loci (duplication products). Polymorphic RAD-tags may contain more than one SNP, but the vast majority (over 99%) showed only two allelic versions; the very small proportion of RAD-tags with more than two alleles were excluded.

### Genetic Map Construction

The genetic map was constructed using R/Onemap [29] and TMAP [30]. The allocation of markers into linkage groups was conducted using R/Onemap. This package uses Hidden Markov Models (HMM) algorithms for outbred species while in parallel implements the methodology described in Wu et al. [31] for calculating the most probable linkage phase. Linkage groups were formed using a minimum LOD value of 6. TMAP was used to order the markers in every linkage group. By using an HMM maximum likelihood model and taking into account potential genotypic errors it reduces the tendency to erroneously derive oversized linkage groups, a phenomenon which is often observed in dense maps. Map distances were calculated in centiMorgans (cM) using the Kosambi mapping function (Table S2). The linkage group name (number) was subsequently matched with the Broad Institute of MIT and Harvard genome assembly Oren11.1 (NCBI Assembly GCA\_000188235.2). The genetic map was drawn and aligned using Genetic-Mapper v0.3 [32].

### QTL Mapping

The QTL analysis was performed using R/qtl [33]. With the dam originating from a clonal line and by inferring the most probable phase of the genetic markers of the sire the cross had the

same properties as a backcross and was analysed as such. Initially existence of single QTLs was tested (R/qtl function: *scanone*). The model used for the analysis was based on interval mapping. The algorithm used considers the phenotype to follow a mixture of Bernoulli distributions and uses a form of the expectation-maximisation algorithm for obtaining maximum likelihood estimates [33]. Permutation tests (10,000 permutations) were conducted in order to correct for the multiple testing. A multidimensional approach towards QTL mapping was adopted by using models from R/qtl that accounted for the QTL in LG 1, for two-QTLs and for multiple-QTLs simultaneously (R/qtl functions: *makeqtl*, *addqtl*, *scantwo*, *fitqtl*, *stepwiseqtl*). With this approach greater power can be achieved in the analysis allowing for detection of QTLs that would be remained undetected in the one-dimensional approach above. Approximate Bayesian 95% density intervals were calculated. An approximate estimate of the phenotypic variance explained by the QTL is obtained from the following equation:  $1 - 10^{-2LOD/n}$ . It has to be stressed that while the estimated variance may be reasonable for additive QTL, problems can be caused in the case of linked QTL [33].

### SNP Assays

We designed SNP assays using the KASP genotyping system (KBioscience UK Ltd) for five SNPs (*Oni20117*, *Oni61067*, *Oni23063*, *Oni28137*, *Oni22734*, NCBI dbSNP accession 748775078, 748775079, 748775085, 748775081 and 748775082 respectively; Table S3) that showed the highest association with sex in the two families that were used for the RAD-seq (Table S4). Allele-specific primers and other assay components were supplied by KBioscience UK Ltd, based on the supplied marker sequences (Table S3, Data S1). PCR reactions were carried for 10 µL final volume reactions. The cycling conditions were the following: 94°C for 15 min, 94°C for 20 sec, touchdown over 61°C to 55°C for 60 sec (10 cycles dropping 0.8°C each cycle) and an extra 34 cycles at 55°C.

### Association Analysis in Family and Population Data

Family data (offspring from four different families; 96 individuals in total) and 40 unrelated broodstocks (Table 1) were genotyped for the above five SNP markers. A family showing high response in elevated temperature was also genotyped for the same SNPs in order to check whether those SNPs could be used in

distinguishing neomales from normal males. An association analysis was performed using R/SNPAssoc [34]. In the case of family data, association was tested both in separate families and across all families together. A Bernoulli generalised linear model was applied in order to test the magnitude of association between the SNP genotypes and phenotypic sex using this package (function *association*). Both the Bonferroni and permutation tests (10,000 permutations) were used in order to correct for multiple testing.

## Results

### RAD Sequencing

Two crosses with 68 and 20 offspring and their parents (including clonal females) were sampled (Table S1). The samples were barcoded, pooled and sequenced in two lanes of an Illumina HiSeq 2000 sequencer (see Materials and Methods). In total, 459,543,173 raw reads (101 bases long) were produced (171,201,863 paired-end and 117,139,447 single-end reads, NCBI BioProject SRP017804). After removing low quality sequences (quality score under 30), ambiguous barcodes and orphaned paired-end reads, 87.03% of the raw reads were retained (399,918,024 reads). The Stacks package [26] was then used to make the assembly of the sampled loci from each individual: 95,791 and 127,014 RAD-tags were retrieved for Families 1 and 2 respectively, covering 152,916 RAD-tags in total including 69,889 of these shared between the two families (Figure 1). The number of reads and RAD-tags for each sample are reported in Table S1.

### Genetic Map

In order to maximise the number of informative markers and minimise the amount of missing or erroneous data, we used SNP markers retrieved in at least 75% of the samples in each family, and carrying one or two SNPs. Since Family 2 had only 20 offspring, the genetic map was constructed with the Family 1 data only (68

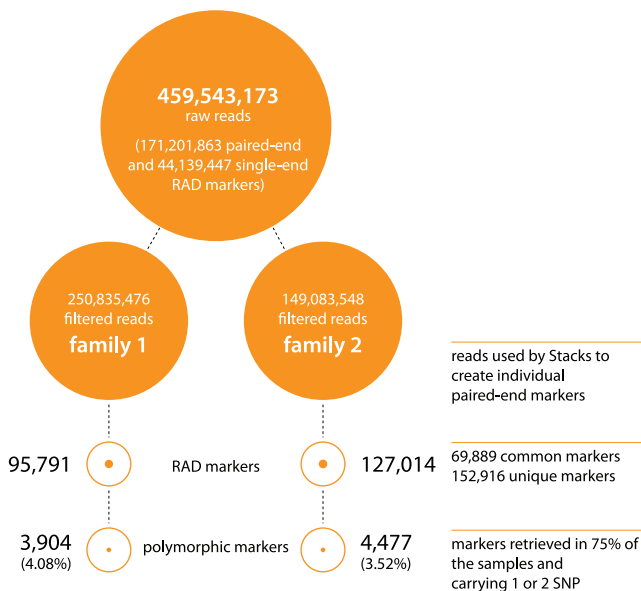
offspring), while Family 2 was used to validate this. The sire-based map consists of 3,802 SNP markers that were grouped in 23 linkage groups, with an average spacing of 0.7 cM and spanning a total distance of 1,176 cM (Figure 2 and Tables 2 & S2). A unique genotypic pattern was observed in 3,280 of the above markers (522 showed identical inheritance patterns to other SNPs). In the second family (Family 2, 20 offspring) 724 of the above markers were heterozygous. The linkage groups were named according to the Broad Institute of MIT and Harvard genome assembly Orenill.1 (NCBI Assembly GCA\_000188235.2). We were not able to join the markers into 22 linkage groups corresponding to the 22 chromosomes expected from the karyotype [35]. However, by comparing our map to the draft tilapia genome sequence, two linkage groups were coalesced to form LG 3 in Figure 2. LG 3 contains a broad region of recombination suppression [36].

### QTL Mapping

The results from the single-QTL model for binary traits provided evidence for the existence of a major QTL in LG 1 for both families (Figure 3A). The highest logarithm of odds (LOD) score for Families 1 and 2 were 18.50 and 6.02 respectively and the QTL was observed in the same location for both families with the same SNPs showing significant linkage (Figure 3B). The difference in the LOD scores reflects the smaller number of meioses in the second family (20 offspring). The fact that the marker phase was the same in the two families (in the QTL region) allowed an additional joint analysis of the two families. The genome wide threshold LOD value ( $\alpha=0.001$ ) was 3.89 as calculated from permutation tests (10,000 permutations). The highest LOD score was observed at the 28.5<sup>th</sup> cM of LG 1. Two adjacent SNPs were located in this position (*Oni23063*, *Oni68581*, from paired RAD-tags on either side of the same *SbfI* restriction site). Only *Oni23063* was analysed subsequently. The two-QTL model did not reveal any significant additional QTL or any evidence for epistasis. Even though models that take into account the existence of a major QTL (as in this study) or ones that test for existence of multiple QTLs simultaneously reduce the residual variation (providing higher power in the analysis for detecting additional QTLs at least of modest effect), no additional QTLs were detected. The calculated 95% Bayesian Density Intervals for the QTL location spanned a region of 2 cM (28–30 cM in LG 1), corresponding to approximately 1.2 Mb in the *O. niloticus* genome. The phenotypic variance explained by the QTL was 96%.

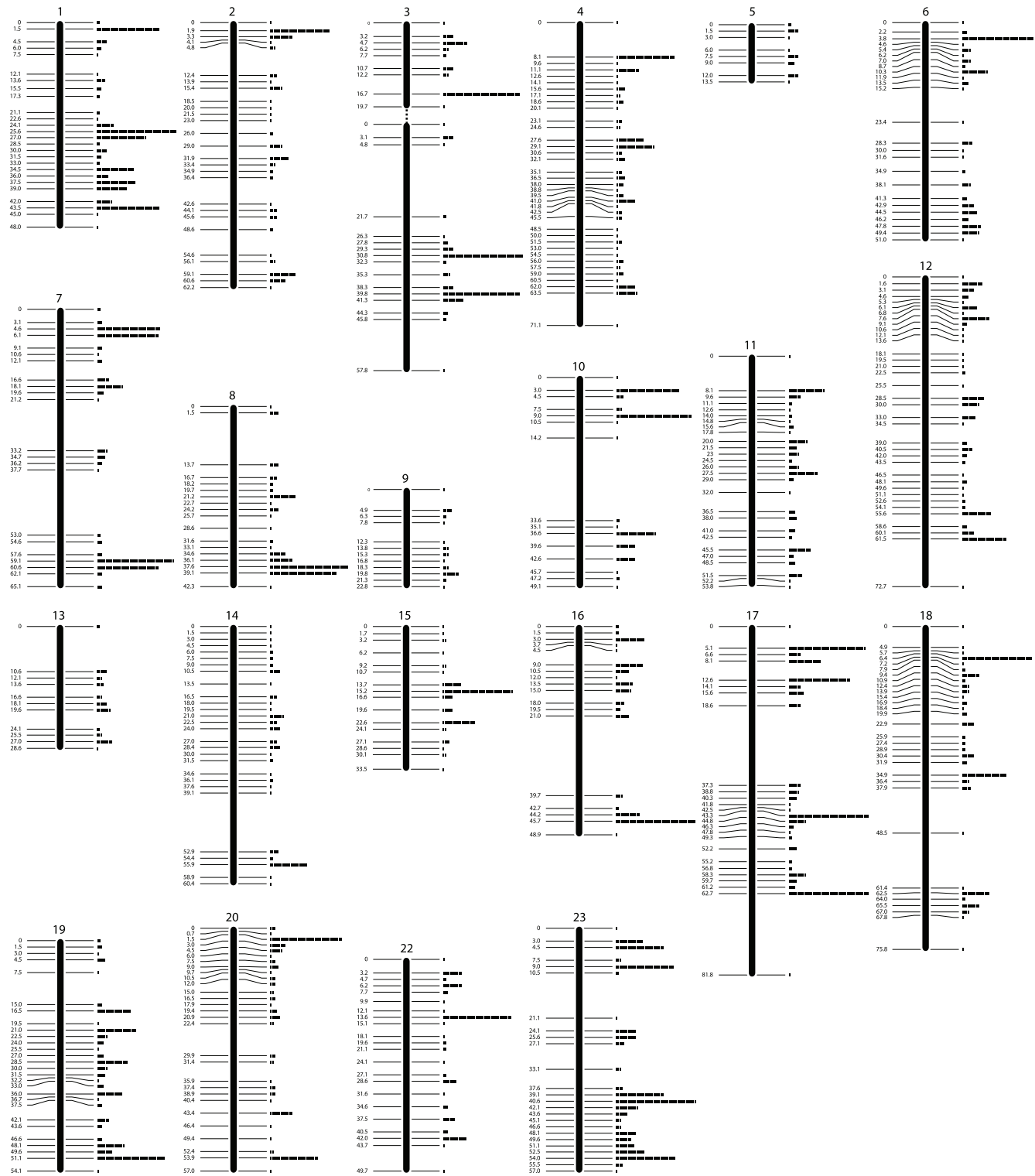
### Association Analysis using Family and Population Samples

The five SNP markers that showed the strongest linkage with sex (Figure 4A) in the two mapping families were tested in a larger panel consisting of family and population samples. The  $p$ -value thresholds ( $\alpha=0.05$ , multiple test correction) for permutation and the Bonferroni correction tests were 0.016 and 0.01 respectively. All five SNP markers were found to be significantly associated with sex in the family data both when tested in each family separately and across all families simultaneously (Figure 4A). For the population data (40 broodstock), four of the five SNP markers showed significant association with phenotypic sex (the exception was *Oni20117*,  $p=0.73$ ). *Oni23063* and *Oni28137* showed the highest association with phenotypic sex for both the population and family data ( $p$ -values of  $1.08 \times 10^{-7}$  and  $3.024 \times 10^{-29}$  respectively). Females were homozygous and males heterozygous for those two SNPs. The only exceptions from the above pattern were found for one male progeny from the four tested families and five male broodstock fish (Table S4). Progeny testing data were available for two of these five male broodstock (in crosses to XX



**Figure 1. Sequencing and RAD-tag summary.** Details of the number of reads before and after filters (orange disk) followed by the reconstructed number of RAD markers and polymorphic RAD markers (orange circles).

doi:10.1371/journal.pone.0068389.g001



**Figure 2. Genetic linkage map.** Map with linkage group assignment determined using syntenic markers with previously published *O. niloticus* maps. The positions on the left side of the chromosomes are in cM. The rectangles on the right hand side represent the number of markers at this position. The numbering of the linkage groups corresponds to that in the Broad Institute genome anchored assembly OreNil1.1 [16,47]. Detailed data are provided in Table S2. doi:10.1371/journal.pone.0068389.g002

females) and gave nearly all-female progeny (over 95% females), which would explain the homozygous SNP genotypes for these sires.

The sex-determining region spanned a distance of 2 cM, corresponding to approximately 1.2 Mb in the *O. niloticus* genome, with the peak (highest LOD score) located at the position of



**Table 2.** *O. niloticus* Genetic Map.

Linkage Group	No. of markers	No. of informative markers	Length (cM)
1	284	248	48.0
2	136	123	62.2
3a	109	96	19.7
3b	153	138	57.8
4	191	177	71.1
5	27	25	13.5
6	161	134	51.0
7	232	197	65.2
8	170	136	42.3
9	32	32	22.8
10	143	111	49.1
11	135	119	53.8
12	173	155	72.7
13	47	41	28.6
14	90	86	60.4
15	192	157	33.5
16	154	124	48.9
17	323	277	81.5
18	188	160	75.8
19	246	210	54.1
20	155	136	57.0
22	117	111	49.7
23	344	287	57.0
Total	3,802	3,280	1,176

doi:10.1371/journal.pone.0068389.t002

*Oni23063* (Figure 3B). The two markers most strongly associated with sex, *Oni28137* and *Oni23063*, are 400 kb apart in the *O. niloticus* genome. The 10 annotated genes within this region of the genome are therefore potential candidates for the sex-determining gene (Figure 4B and Data S2).

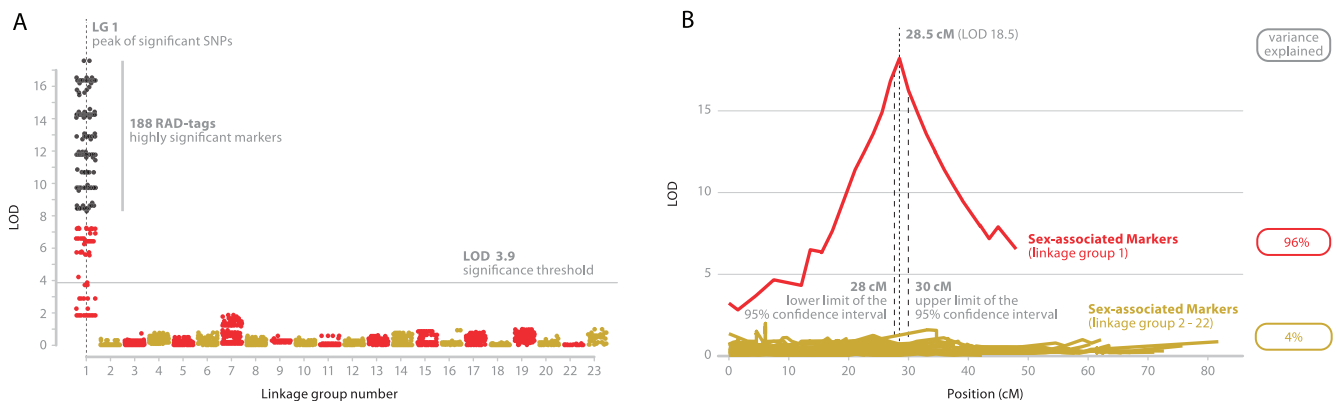
## Temperature Sex Reversal and Genotype

To explore the effects of experimental sex reversal on the association between genotype and phenotypic sex, since family-specific QTLs involved in temperature sex determination have been recently identified [37], a family that showed a significant effect of raised temperature on sex ratio was genotyped for the two SNPs showing the strongest association (*Oni23063* and *Oni28137*). While these showed highly significant association ( $p = 4.05 \times 10^{-14}$  for both markers) in the control group, in the temperature treated group neither marker was significantly associated after correction for multiple testing ( $p = 0.039$  for both markers). Five phenotypic males were found to deviate from the expected association in the control group, while in the treated group 28 phenotypic males deviated from the expected association for *Oni23063* and *Oni28137* (Figure 5).

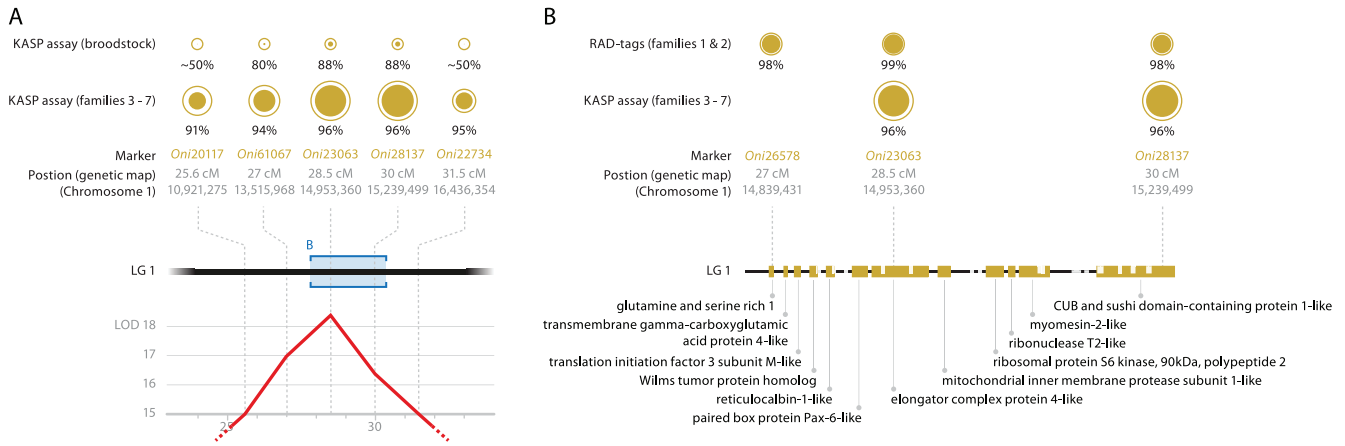
## Discussion

Current evidence suggests that *O. niloticus* possess an XY/XX male heterogametic system complicated by genetic variance at this and other loci, environmental factors and probably by their interaction [36]. While previous family-based studies provided evidence for the existence of a major sex-determining region on LG 1 in *O. niloticus* [10,17,38,39], various anomalies, including inconsistencies among families, have been observed. For example, Lee et al. [17] demonstrated the existence of a sex-determining region on LG 1, using microsatellite markers, which was mapped to an interval of 10 cM. However, the association between this region and phenotypic sex was only observed in two out of the three crosses studied. In the third cross, no association was observed with any genomic region. Frequent departures have been observed from the 100% male progeny in YY  $\times$  XX crosses predicted from a single XX/XY locus [15,40]. Eshel et al. [12,13] found that markers in LG 23 showed the highest association with phenotypic sex in a cross in a population of *O. niloticus* also derived from Lake Manzala in Egypt. However no information was provided whether the above QTL on LG 23 persists in other crosses as well.

In the present study, our strategy was to use RAD sequencing to develop a much higher density linkage map than that of Lee et al. [16], based on carefully selected mapping families (isogenic female crossed to normal male; balanced sex ratio; progeny already tested for association between LG 1 microsatellite markers and phenotypic sex), then validate a set of tightly sex-linked markers in further family and population samples to test for population-



**Figure 3. Results from QTL-Analysis.** (A) Association results for genotyped SNPs. SNPs with  $P$ -values achieving genome-wide significance ( $p < 7.2 \times 10^{-8}$ ) are shown in black. (B) Regional analysis of the QTL. Plot of the LOD score (sex-association QTL search) along the linkage groups. doi:10.1371/journal.pone.0068389.g003



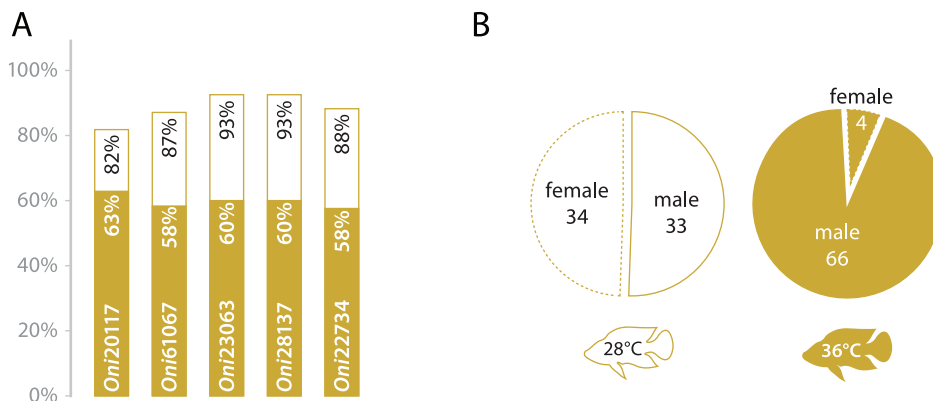
**Figure 4. KASP assay and fine gene mapping on LG 1.** (A) Details of the five markers tested by KASP assay. From bottom to top: LOD score (QTL sex-association); Location of the five markers (in the genetic map in cM and anchored draft genome in bp); KASP assay results. The outer circle diameters are proportional to the number of alleles tested. The inner (solid) disks represent the marker association with the phenotypic sex. Detailed data are provided in Table S4. (B) Details of the region of higher association. The bottom half is a schematic (to scale) of the chromosome LG 1, scaffold 17. It includes 14 gaps (white gaps) in the genome and 10 annotated genes (orange boxes). The sex-determining factor is located between “glutamine and serine rich 1” and “CUB and sushi domain-containing protein 1- like”. Detailed data are provided in Data S2. doi:10.1371/journal.pone.0068389.g004

wide association. SNP markers *Oni61067*, *Oni23063* and *Oni28137* showed the highest and most consistent association with phenotypic sex in all our data. *Oni23063* was the marker with the highest LOD score in the two families that were used for the QTL analysis, while the additional families that were genotyped had the same magnitude of association with *Oni28137*, with no recombination observed between these two markers. *Oni61067* gave the third largest association. The above provides evidence that the most probable location for the master sex-determining gene is in this region on LG 1, which spans around 1.2 Mb in scaffold 17 in the *O. niloticus* draft genome (Figure 4).

In all tested families, at each of the two loci (*Oni23063* and *Oni28137*) females were homozygous for the same allele, while males were heterozygous, apart from two males (one each from Families 1 and 4) that were homozygous for the allele found in females. The results from the unrelated adult individuals showed again the same pattern, with the exception of five males that were homozygous for these two SNPs. Of these five, two were shown to give >95% female progeny in crosses to normal females (while no progeny testing data was available for the other three). No

phenotypic female was mis-assigned in our entire dataset. Our data support the hypothesis that the mis-assigned (homozygous) males were in fact neomales, genetically female fish (as defined by the LG1 sex-determining locus) that have undergone sex reversal due to unknown genetic or environmental factor(s). The highly significant population-based linkage disequilibrium observed strongly supports the view that these two SNP markers are in very close proximity to the causative locus.

When a family responsive to sex reversal by exposure to elevated temperature was genotyped for *Oni23063* and *Oni28137*, the temperature-treated group showed a significant number of phenotypic males with the female compatible genotype, while the typically strong association between SNP genotype and phenotypic sex was found in the control group. Rearing at high temperatures (36°C) results in significant masculinisation in some progenies if started around 10 days post fertilisation and applied for at least 10 days [41–43]. Wessels & Hörstgen-Schwark [22] showed that temperature-dependent sex ratio is a heritable trait. Lühmann et al. [37] mapped family-specific QTL for high temperature masculinisation effects to LG 1, LG 3 and LG 23. There is also



**Figure 5. Sex reversion tests (family 7).** (A) KASP assay results in family 7 offspring in response to elevated temperature (36°C) compare to standard 28°C. Each bar represent the ratio of male/female at 36°C (solid orange) and at 28°C (white background). (B) Detail of the phenotypic sex in response to elevated temperature (male have continuous border, female are presented with a discontinuous border). doi:10.1371/journal.pone.0068389.g005

evidence for feminisation by high temperature, from crosses expected to produce all-male progeny [44–46].

The genomic region, which was derived from the calculated density interval, includes 10 identified genes, none of which have yet been identified as the major sex-determining gene of any species. Nevertheless, it would not be prudent to exclude any of these as candidates at this time. A diversity of sex-determining systems has been documented in fish, implicating a variety of different, and in some cases unexpected, genes. For example, in *O. mykiss*, a gene associated with immune-related functions in other organisms was found to be its master sex-determining gene [4]. However, despite the fact that the results of this study showed that the major sex-determining factor is in close proximity with the tested SNPs, we cannot be confident that the major sex-determining gene is necessarily one of these ten genes. If the master sex-determining gene is male-specific (as found in *O. latipes* [3] and *O. mykiss* [7]), its presence will not be detectable from the current draft tilapia genome, which is derived from an isogenic female line. Finally there is also the possibility that the major sex-determining gene remains un-annotated in the current genome draft.

This study did not address the reasons why YY × XX crosses yield varying proportions of females. However, the SNP markers developed will help to tackle this issue, a key constraint on the commercial development of this technique.

## Conclusions

This study provides a linkage map of the *O. niloticus* genome that is several-fold denser than the existing one, a reduced candidate region for the sex-determining gene(s) and a set of tightly sex-linked SNP markers. Although we could not identify the causative gene(s), the fact that no female was mis-assigned using our sex-associated SNPs means that those SNPs could be also of high practical value towards the production of all male stocks for the *O. niloticus* aquaculture industry.

## References

- Kamiya T, Kai W, Tasumi S, Oka A, Matsunaga T, 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: e1002798. doi:10.1371/journal.pgen.1002798.
- Devlin RH, Nagahama Y (2002) Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture* 208: 191–364. doi:10.1016/S0044-8486(02)00057-1.
- Penman DJ, Piferrer F (2008) Fish Gonadogenesis. Part I: Genetic and Environmental Mechanisms of Sex Determination. *Rev Fish Sci* 16: 16–34. doi:10.1080/10641260802324610.
- Yano A, Guyomard R, Nicol B, Jouanno E, Quillet E, et al. (2012) An Immune-Related Gene Evolved into the Master Sex-Determining Gene in Rainbow Trout, *Oncorhynchus mykiss*. *Curr Biol* 22: 1423–1428. doi:10.1016/j.cub.2012.05.045.
- Matsuda M, Nagahama Y, Shinomiya A, Sato T, Matsuda C, et al. (2002) *DMT* is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* 417: 559–563. doi:10.1038/nature751.
- Nanda I, Hornung U, Kondo M, Schmid M, Schardt M (2003) Common spontaneous sex-reversed XX males of the medaka *Oryzias latipes*. *Genetics* 163: 245–251.
- 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. doi:10.1534/genetics.111.137497.
- Hattori RS, Murai Y, Oura M, Masuda S, Majhi SK, et al. (2012) A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *Proc Natl Acad Sci USA* 109: 2955–2959. doi:10.1073/pnas.1018392109.
- FAO Fisheries and Aquaculture Department (2012) FishStat database. Available: <http://faostat.fao.org/site/629/default.aspx>.
- Cnaani A, Lee B-Y, Zilberman N, Ozouf-Costaz C, Hulata G, et al. (2008) Genetics of Sex Determination in Tilapiine Species. *Sex Dev* 2: 43–54. doi:10.1159/000117718.
- Karayücel İ, Ezaz T, Karayücel S, McAndrew BJ, Penman DJ (2004) Evidence for two unlinked “sex reversal” loci in the Nile tilapia, *Oreochromis niloticus*, and for linkage of one of these to the red body colour gene. *Aquaculture* 234: 51–63. doi:10.1016/j.aquaculture.2003.12.016.
- Eshel O, Shirak A, Weller JL, Slossman T, Hulata G, et al. (2010) Fine-mapping of a locus on linkage group 23 for sex determination in Nile tilapia (*Oreochromis niloticus*). *Anim Genet* 42: 222–224. doi:10.1111/j.1365-2052.2010.02128.x.
- Eshel O, Shirak A, Weller JL, Hulata G, Ron M (2012) Linkage and Physical Mapping of Sex Region on LG23 of Nile Tilapia (*Oreochromis niloticus*). *G3* 2: 35–42. doi:10.1534/g3.111.001545.
- Baroiller J-F, D’Cotta H, Bezault E, Wessels S, Hoerstgen-Schwark G (2009) Tilapia sex determination: Where temperature and genetics meet. *Comp Biochem Physiol A Mol Integr Physiol* 153: 30–38. doi:10.1016/j.cbpa.2008.11.018.
- Beardmore JA, Mair GC, Lewis RI (2001) Monosex male production in finfish as exemplified by tilapia: applications, problems, and prospects. *Aquaculture* 197: 283–301. doi:10.1016/S0044-8486(01)00590-7.
- Lee B-Y, Lee W-J, Streebman JT, Carleton KL, Howe AE, et al. (2005) A second-generation genetic linkage map of tilapia (*Oreochromis* spp.). *Genetics* 170: 237–244. doi:10.1534/genetics.104.035022.
- Lee B-Y, Penman DJ, Kocher TD (2003) Identification of a sex-determining region in Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis. *Anim Genet* 34: 379–383. doi:10.1046/j.1365-2052.2003.01035.x.
- Lee B-Y, Kocher TD (2007) Exclusion of *Wilms tumour (WT1)* and *ovarian cytochrome P450 aromatase (CYP19A1)* as candidates for sex determination genes in Nile tilapia (*Oreochromis niloticus*). *Anim Genet* 38: 85–86. Available: <http://dx.doi.org/10.1111/j.1365-2052.2007.01563.x>.

## Supporting Information

**Table S1 Sample origins and RAD barcodes.** Details each sample used: sample ID, family, gender, RAD barcode (index) used, number of raw reads (paired-ended or single-ended) and number of RAD-tags. (CSV)

**Table S2 Genetic maps.** Ordered markers: marker ID, linkage group and position (cM). (CSV)

**Table S3 KASP assay primer sequences.** List of the allele-specific primers and common primer designed for the allele-specific PCR genotyping assay of the five markers as well as their NCBI dbSNP accession numbers. (CSV)

**Table S4 Details of the KASP assay results.** Genotypes of the 280 assays. (CSV)

**Data S1 Marker sequences.** Details of SNP alleles and RAD-tag allele sequences of the five markers. (FASTA format) (TXT)

**Data S2 Details of the physical location of the markers and the neighbouring annotated genes.** (GFF3 format) (TXT)

## Acknowledgments

We thank staff at The GenePool Genomics Facility, especially Urmi Trivedi and Marian Thomson, for assistance with RAD library preparation and sequencing.

## Author Contributions

Conceived and designed the experiments: CP MB MGQK JBT KG BJM DJP. Performed the experiments: CP MB MGQK JBT KG. Analyzed the data: CP MB KG DJP. Contributed reagents/materials/analysis tools: CP MB MGQK JBT KG. Wrote the paper: CP MB DJP. Contributed to editing: MGQK JBT KG BJM.



19. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3: e3376. doi:10.1371/journal.pone.0003376.
20. Parliament of the United Kingdom (1986) Animals (Scientific Procedures) Act 1986. Available: <http://www.legislation.gov.uk/ukpga/1986/14>.
21. Guerrero RD III, Shelton WL (1974) An aceto-carminine squash method for sexing juvenile fishes. *The Progressive Fish-Culturist* 36: 56–56. doi:10.1577/1548-8659(1974)36[56:AASMF5]2.0.CO;2.
22. Wessels S, Hörstgen-Schwark G (2007) Selection experiments to increase the proportion of males in Nile tilapia (*Oreochromis niloticus*) by means of temperature treatment. *Aquaculture* 272: S80–S87. doi:10.1016/j.aquaculture.2007.08.009.
23. Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA (2011) SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Methods Mol Biol* 772: 157–178. doi:10.1007/978-1-61779-228-1\_9.
24. Houston RD, Davey JW, Bishop SC, Lowe NR, Mota-Velasco JC, et al. (2012) Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC Genomics* 13: 244. doi:10.1186/1471-2164-13-244.
25. Baxter SW, Davey JW, Johnston JS, Shelton AM, Heckel DG, et al. (2011) Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PLoS One* 6: e19315. doi:10.1371/journal.pone.0019315.
26. Catchen JM, Amores A, Hohenlohe PA, Cresko WA, Postlethwait JH (2011) Stacks: building and genotyping loci *de novo* from short-read sequences. *G3* 1: 171–182. doi:10.1534/g3.111.000240.
27. Hohenlohe PA, Bassham S, Etter PD, Stüffler N, Johnson EA, et al. (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet* 6: e1000862. doi:10.1371/journal.pgen.1000862.
28. Zerbino DR, Birney E (2008) Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* 18: 821–829. doi:10.1101/gr.074492.107.
29. Margarido GR, Souza AP, Garcia AA (2007) OneMap: software for genetic mapping in outcrossing species. *Hereditas* 144: 78–79. doi:10.1111/j.2007.0018-0661.02000.x.
30. Cartwright DA, Troggio M, Velasco R, Gutin A (2007) Genetic mapping in the presence of genotyping errors. *Genetics* 176: 2521–2527. doi:10.1534/genetics.106.063982.
31. Wu R, Ma CX, Wu SS, Zeng ZB (2002) Linkage mapping of sex-specific differences. *Genet Res* 79: 85–96. doi:10.1017/S0016672301005389.
32. Bekaert M (2012) Genetic-Mapper. Available: <https://code.google.com/p/genetic-mapper/>.
33. Broman KW, Sen S (2009) *A Guide to QTL Mapping with R/qlt*. New York, USA: Springer. 1 pp. Available: <http://link.springer.com/book/10.1007/978-0-387-92125-9>.
34. González JR, Armengol L, Solé X, Guinó E, Mercader JM, et al. (2007) SNPpassoc: an R package to perform whole genome association studies. *Bioinformatics* 23: 644–645. doi:10.1093/bioinformatics/btm025.
35. Majumdar KC, McAndrew BJ (1986) Relative DNA content of somatic nuclei and chromosomal studies in three genera, *Tilapia*, *Sarotherodon*, and *Oreochromis* of the tribe Tilapiini (Pisces, Cichlidae). *Genetica* 68: 175–188. doi:10.1007/BF02424441.
36. Cnaani A, Hulata G (2008) Tilapia. In: Kocher TD, Kole C, editors. *Genome Mapping and Genomics in Fishes and Aquatic Animals*. Berlin Heidelberg: Springer. 101–116. Available: [http://link.springer.com/chapter/10.1007/978-3-540-73837-4\\_4](http://link.springer.com/chapter/10.1007/978-3-540-73837-4_4).
37. Lüthmann LM, Knorr C, Hörstgen-Schwark G, Wessels S (2012) First evidence for family-specific QTL for temperature-dependent sex reversal in Nile tilapia (*Oreochromis niloticus*). *Sex Dev* 6: 247–256. doi:10.1159/000339705.
38. Lee B-Y, Kocher TD (2007) *Comparative Genomics and Positional Cloning*. Aquaculture Genome Technologies. Oxford, UK: Wiley-Blackwell. 323–339.
39. Lee B-Y, Coutanceau J-P, Ozouf-Costaz C, D’Cotta H, Baroiller J-F, et al. (2011) Genetic and physical mapping of sex-linked AFLP markers in Nile tilapia (*Oreochromis niloticus*). *Mar Biotechnol* 13: 557–562. doi:10.1007/s10126-010-9326-7.
40. Mair GC, Abucay JS, Abella TA, Beardmore JA, Skibinski DOF (1997) Genetic manipulation of sex ratio for the large-scale production of all-male tilapia *Oreochromis niloticus*. *Can J Fish Aquat Sci* 54: 396–404. doi:10.1139/f96-282.
41. Baroiller J-F, Chourrout D, Fostier A, Jalabert B (1995) Temperature and sex chromosomes govern sex ratios of the mouthbrooding Cichlid fish (*Oreochromis niloticus*). *J Exp Zool* 273: 216–223. doi:10.1002/jez.1402730306.
42. Baroiller J-F, D’Cotta H (2001) Environment and sex determination in farmed fish. *Comp Biochem Physiol C Toxicol Pharmacol* 130: 399–409. doi:10.1016/S1532-0456(01)00267-8.
43. Tessema M, Müller-Belecke A, Hörstgen-Schwark G (2006) Effect of rearing temperatures on the sex ratios of *Oreochromis niloticus* populations. *Aquaculture* 258: 270–277. doi:10.1016/j.aquaculture.2006.04.041.
44. Abucay JS, Mair GC, Skibinski DOF, Beardmore JA (1999) Environmental sex determination: the effect of temperature and salinity on sex ratio in *Oreochromis niloticus* L. *Aquaculture* 173: 219–234. doi:10.1016/S0044-8486(98)00489-X.
45. Kwon JY, McAndrew BJ, Penman DJ (2002) Treatment with an aromatase inhibitor suppresses high-temperature feminization of genetic male (YY) Nile tilapia. *J Fish Biol* 60: 625–636. doi:10.1111/j.1095-8649.2002.tb01689.x.
46. Wessels S, Hörstgen-Schwark G (2011) Temperature dependent sex ratios in selected lines and crosses with a YY-male in Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 318: 79–84. doi:10.1016/j.aquaculture.2011.04.039.
47. Guyon R, Rakotomanga M, Azzouzi N, Coutanceau J-P, Bonillo C, et al. (2012) A high-resolution map of the Nile tilapia genome: a resource for studying cichlids and other percomorphs. *BMC Genomics* 13. doi:10.1186/1471-2164-13-222.