

RESEARCH ARTICLE

Open Access



Comparative analysis of a sex chromosome from the blackchin tilapia, *Sarotherodon melanotheron*

William J. Gammerdinger¹, Matthew A. Conte¹, Jean-François Baroiller², Helena D'Cotta² and Thomas D. Kocher^{1*}

Abstract

Background: Inversions and other structural polymorphisms often reduce the rate of recombination between sex chromosomes, making it impossible to fine map sex-determination loci using traditional genetic mapping techniques. Here we compare distantly related species of tilapia that each segregate an XY system of sex-determination on linkage group 1. We use whole genome sequencing to identify shared sex-patterned polymorphisms, which are candidates for the ancestral sex-determination mutation.

Results: We found that *Sarotherodon melanotheron* segregates an XY system on LG1 in the same region identified in *Oreochromis niloticus*. Both species have higher densities of sex-patterned SNPs, as well as elevated number of ancestral copy number variants in this region when compared to the rest of the genome, but the pattern of differentiation along LG1 differs between species. The number of sex-patterned SNPs shared by the two species is small, but larger than expected by chance, suggesting that a novel Y-chromosome arose just before the divergence of the two species. We identified a shared sex-patterned SNP that alters a *Gata4* binding site near *Wilms tumor protein* that might be responsible for sex-determination.

Conclusions: Shared sex-patterned SNPs, insertions and deletions suggest an ancestral sex-determination system that is common to both *S. melanotheron* and *O. niloticus*. Functional analyses are needed to evaluate shared SNPs near candidate genes that might play a role in sex-determination of these species. Interspecific variation in the sex chromosomes of tilapia species provides an excellent model system for understanding the evolution of vertebrate sex chromosomes.

Keywords: African cichlids, Sex chromosome evolution, Sex-determination

Background

Natural selection alters the local genomic environment around newly evolved sex-determination loci in two ways. First, selection favors the accumulation of sexually antagonistic alleles and their association with the appropriate X- or Y-haplotype. Second, selection also favors a reduction in the rate of recombination between the proto-X and proto-Y to preserve the associations between the sexually antagonistic alleles and the sex-determination locus [1, 2]. The process is cyclic, in that selection will favor the recruitment of additional sexually antagonistic loci within a growing region of reduced

recombination [3]. Once a region of reduced recombination is established, deleterious mutations will begin to accumulate via Muller's Ratchet [4]. Through these mechanisms, the region of differentiation between the X- and Y-chromosome grows, creating a series of evolutionary strata from the oldest and most decayed regions to the newest and least decayed [5]. Structural rearrangements, such as inversions, are one mechanism that reduces recombination and contributes to the creation of evolutionary strata on sex chromosomes [1, 2, 5].

The initial sex-determination loci are buried within the oldest evolutionary strata because the first selectively favored chromosomal rearrangement encompasses both the sex-determination locus and at least one sexually antagonistic locus. Traditional genetic mapping techniques cannot be used to fine map the sex-determination locus

* Correspondence: tdk@umd.edu

¹Department of Biology, University of Maryland, College Park, MD 20742, USA
Full list of author information is available at the end of the article



because of the lack of recombination inside these structural polymorphisms. An alternative approach is to identify shared ancestral polymorphisms inside the deepest and most decayed stratum among species that share a common ancestral sex-determination mechanism.

Cichlid fishes are a model system for studying evolutionary processes [6]. Sex-determination has evolved rapidly among African cichlids. Sex-determination loci have been identified on linkage groups (LG) 1 (XY), 3 (WZ) and 23 (XY) among species of tilapia, and on linkage groups 5 (WZ), 7 (XY) and possibly others in haplochromine cichlids [7–9]. The sex determination locus on LG23 appears to be a duplication of *amh*, but the sex-determination genes on the other chromosomes have not yet been identified [10, 11].

Previous studies have mapped the sex-determination locus on linkage group 1 in the Nile tilapia (*Oreochromis niloticus*) to a region of a few centimorgans [8, 12–14]. We recently identified a high density of sex-patterned SNPs on LG1 in the region from 10.1 to 18.9 Mb [15]. This 8.8 Mb region is significantly enriched for missense mutations and likely corresponds to an inversion that restricts recombination between the X- and Y-chromosomes.

The blackchin tilapia, *Sarotherodon melanotheron*, is a sister group to the genus *Oreochromis* [16] and has received relatively little attention with respect to sex-determination. One study reported that the chromosomes of *S. melanotheron* were homomorphic with a chromosome count consistent with that of other species of *Oreochromis* [17]. In this study we identify the sex-determination region in *S. melanotheron*, characterize the shared regions of differentiation on LG1 between *O. niloticus* and *S. melanotheron* and catalog shared SNPs that may be responsible for sex-determination in these species.

Methods

Materials

Sarotherodon melanotheron were originally collected from Lake Guiers, Senegal. The individuals sampled are the 4th generation progeny of the wild-caught fish. Fins were sampled from a total of 22 male and 22 female fish and preserved in a salt-DMSO preservation solution. The sex ratio of the family was 50:50.

Sequencing

DNA was extracted from each fin-clip by phenol-chloroform extraction using phase-lock gel tubes (5Prime, Gaithersburg, Maryland). DNA concentrations were measured by fluorescence spectrometry and normalized during library preparation. DNA from 21 males and 22 females were then separately pooled (one likely male was excluded due to ambiguous gonads). Sequencing libraries were constructed using the TruSeq DNA PCR-Free LT Kit (Illumina, San Diego, California). A full

lane of Illumina 183 bp paired-end sequencing was performed for each sex.

For comparative analyses, we reanalyzed whole genome sequencing data for male and female *Oreochromis niloticus* from our previous study [15]. These data consist of 100 bp paired-end reads from separate male and female DNA pools.

Read mapping

Reads that did not pass Illumina CASAVA 1.8 filtering were removed. Read qualities for both lanes were analyzed using FastQC [18]. The reads were then mapped to the *O. niloticus* reference assembly with Bowtie2 using the *-very-sensitive* setting and a minimum insert length of 200 bp [19, 20]. Variants were called using GATK in the *UnifiedGenotyper* mode [21]. Read qualities with a PHRED score of at least 20 were required for variant calling.

Identification of sex-patterned SNPs

A custom script was used to identify sites showing a sex-patterned signature. *Sex_SNP_Finder_GA.pl* is an expansion of the *Sex_SNP_finder_now.pl* script released previously [15, 22]. Sex-patterned SNPs are sites that are fixed or nearly fixed in the homogametic sex and in a frequency between 0.3 and 0.7 in the heterogametic sex. We required a minimum read depth of ten and a minimum allele count of two. We assessed the overall density of sex-patterned SNPs in 10 kb non-overlapping windows. We also counted regions enriched for sex-patterned SNPs by identifying non-overlapping 10 kb windows containing at least 10 sex-patterned SNPs.

In addition to finding sex-patterned SNPs, the updated script calculates F_{ST} , d_{xy} , d_a , Nei's D and C_p for each nucleotide position across the genome. F_{ST} is calculated in accordance with the method used in PoPoolation2, except that we calculated values at each site instead of by windows [23]. Due to numeric difficulties handling the upper bound of Nei's D, the calculation is modified for alternatively fixed positions to make the calculation based upon the maximum coverage level specified by the user. C_p is a metric of population differentiation that ranges from 0 (no population differentiation) to 1 (complete population differentiation), where x_i and y_i are the frequencies of each allele in populations x and y [24].

$$C_p = \sum_{i=1}^n \frac{1}{2} |x_i - y_i|$$

Functional annotation

The functional significance of sex-patterned SNPs was evaluated with SnpEff and SnpSift using the gene annotations obtained from NCBI (RefSeq release 70) [25, 26].

We excluded mRNA models that did not have full-length protein coding sequences. Non-synonymous substitutions were subsequently evaluated with PROVEAN to predict functional impacts on protein structure [27]. Three missense mutations, R203C in XM_003448054.2 (LG6), E1235V in XM_003438386.2 (LG13) and G215A in XM_005466093.1 (LG22), were excluded from the PROVEAN analysis due to low quality gene annotations as defined by PROVEAN. Missense mutations harboring PROVEAN scores less than the recommended threshold of -2.5 were considered deleterious.

Identification of shared sex-specific SNPs

Sex-patterned SNPs in *S. melanotheron* were compared to the sex-patterned SNPs identified in *O. niloticus*, which also segregates a sex-determination system on LG1. The null hypothesis for the expected number of shared sex-patterned SNPs was calculated by multiplying the frequency of sex-patterned SNPs within the previously identified region from 10.1 to 18.9 Mb on LG1 in *O. niloticus*, by the frequency of sex-patterned SNPs within the same region in *S. melanotheron* by the size of the region.

Transcription factor binding site analysis

SNPs that were sex-patterned and shared between *O. niloticus* and *S. melanotheron* were inspected to ensure that they shared common X- and Y-alleles. Flanking regions for the SNPs passing this criteria were extracted using Bedtools and compared to the JASPAR CORE Vertebrata 2016 database to identify possible transcription factor binding sites [28, 29]. A JASPAR relative score threshold of 0.80 was used to assess the significance of putative binding sites.

Identification of copy number variants

Copy number variation within the *S. melanotheron* and *O. niloticus* datasets was assessed with VarScan 2, using a minimum window size of 100 bp, a maximum window size of 1 kb and *amp* and *del* thresholds equal to 0.2 [30]. A custom script (Varscan_multiple.pl) was developed to find conserved copy number variants between *n*-number of VarScan 2 comparisons [31]. We then utilized a non-overlapping window approach, which excluded assembly gaps from the windows, to quantify the density of these conserved copy number variants within 10 kb windows across the genome.

Statistical assessment

Mann–Whitney U-tests were used to determine if a region was enriched for sex-patterned SNPs or conserved copy number variants. All testing used 10 kb non-overlapping windows as samples. A Bonferroni correction was applied to control for 12 comparisons resulting in a significance threshold of $\alpha=0.004167$.

Results

Sequencing and mapping of reads

We obtained a total of ~160 million and ~162 million 183 bp paired-end reads from the female and male pools of *S. melanotheron*, respectively. The alignment rates to the *O. niloticus* reference sequence were 85.50 and 85.69 %, respectively. The mean coverage in the female pool was 35.12 reads, while the mean coverage in the male pool was 35.75 reads.

We obtained a total of ~219 million and ~202 million 100 bp paired-end reads from the female and male pools of *O. niloticus*, respectively. The alignment rates to the *O. niloticus* reference were 90.49 and 89.96 %, respectively. The mean coverage in the female pool was 27.17 reads, while coverage in the male pool was 26.84 reads.

F_{ST} differentiation in *S. melanotheron*

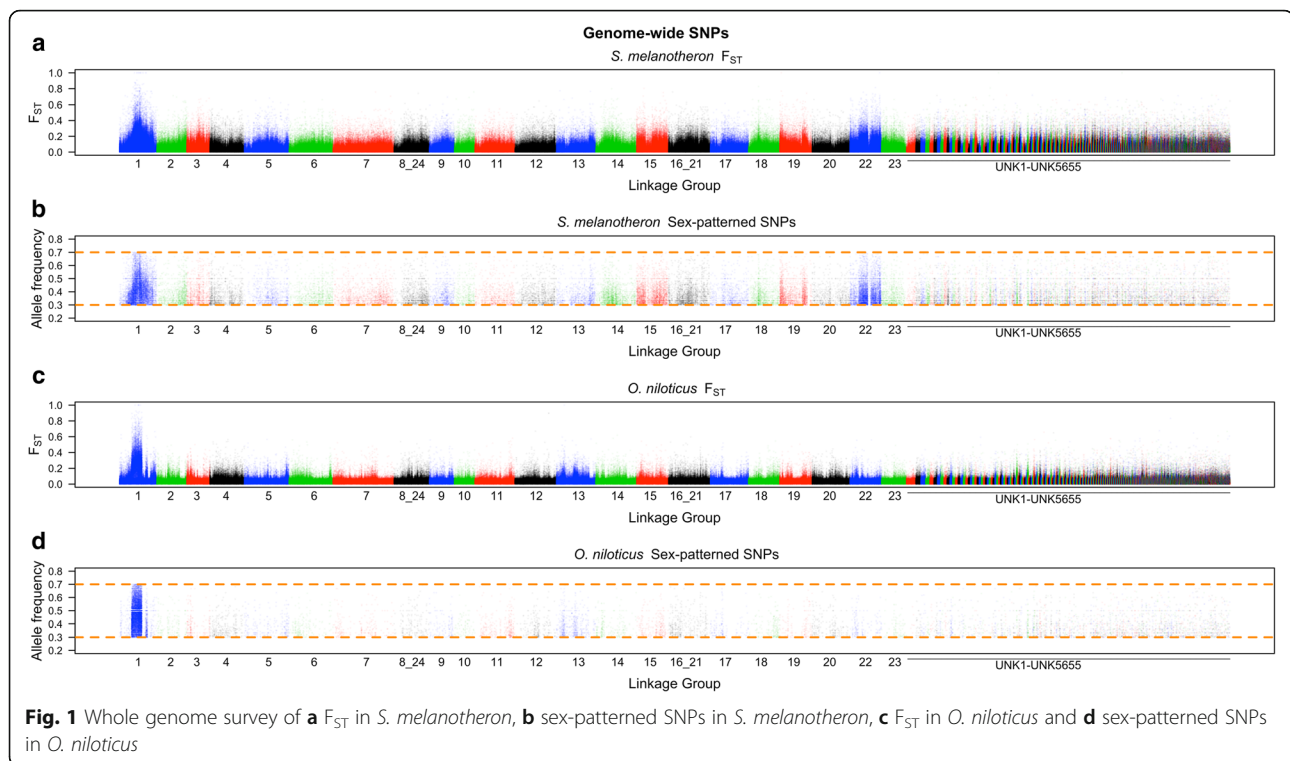
Examination of the whole genome F_{ST} plot comparing male and female *S. melanotheron* identifies a strong signal on LG1 (Fig. 1). A closer examination of LG1 reveals that this region overlaps with the previously identified XY sex-determination region in *O. niloticus* (Fig. 2). The region of divergence on LG1 is broader in *S. melanotheron* than in *O. niloticus* and spans from approximately 10.1 to 28 Mb. The boundaries of the differentiated region in *S. melanotheron* are gradual, unlike the sharp boundaries found in *O. niloticus* on LG1 from 10.1 to 18.9 Mb and from 21.7 to 23.6 Mb.

S. melanotheron also shows indications of divergence between the sexes on LG22. There is no previous evidence to suggest that this region is associated with sex in any cichlid species and the signal is not as strong as on LG1 (Fig. 3 and Table 1). Importantly, *S. melanotheron* shows no strong sex-patterned signal from LG3, a region known to harbor a sex-determination locus in *Pelmatotilapia mariae* and some members of *Oreochromis* [8].

Sex-patterned SNPs

We found a total of 50,088 SNPs in *S. melanotheron* and 19,793 SNPs in *O. niloticus* fitting the sex-patterned criteria. There were 448 and 539 non-overlapping 10 kb windows with at least 10 sex-patterned SNPs in *S. melanotheron* and *O. niloticus*, respectively (Table 1). The highest densities of sex-patterned SNPs occurred between 10.1 and 18.9 Mb in both *O. niloticus* and *S. melanotheron*. Sex-patterned SNPs found on LG1 between 10.1 and 18.9 Mb are at significantly higher densities in this region than any other region in either species ($p < 0.0001$, for all five comparisons to other regions noted in the table).

In *S. melanotheron*, the region on LG1 from 18.9 to 28 Mb, as well as LG22, both showed a significantly higher density of sex-patterned SNPs when compared to



the rest of the genome ($p < 0.0001$, for both comparisons). The region on LG1 from 18.9 to 28 Mb also demonstrated a significantly higher density of sex-patterned SNPs than LG22 ($p < 0.0001$). In *O. niloticus*, the region on LG1 from 21.7 to 23.6 Mb showed a significantly higher density of sex-patterned SNPs than the rest of the genome ($p < 0.0001$).

Functional impacts

We evaluated the functional impacts of the sex-patterned SNPs and found a total of 810 missense and 1231 synonymous mutations in *S. melanothereon*. Table 1 shows the average F_{ST} across all polymorphic SNPs within each region as well as the genomic distribution of each subset of the sex-patterned SNPs: total sex-patterned SNPs, 10 kb windows enriched for sex-patterned SNPs, missense SNPs, missense SNPs with a PROVEAN score less than -2.5 and synonymous mutations. The density of each class of sex-patterned SNPs in *S. melanothereon* was consistently from highest to lowest: LG1 between 10.1 and 18.9 Mb, LG1 between 18.9 and 28 Mb, LG22, then across the rest of the genome. There were seven stop codon gains or losses, but only one mapped to a region of divergence on LG1 (Additional file 1).

In *O. niloticus* we identified a total of 331 missense and 442 synonymous mutations. The density of each class of sex-patterned SNPs in *O. niloticus* was consistently from highest to lowest: LG1 between 10.1 and

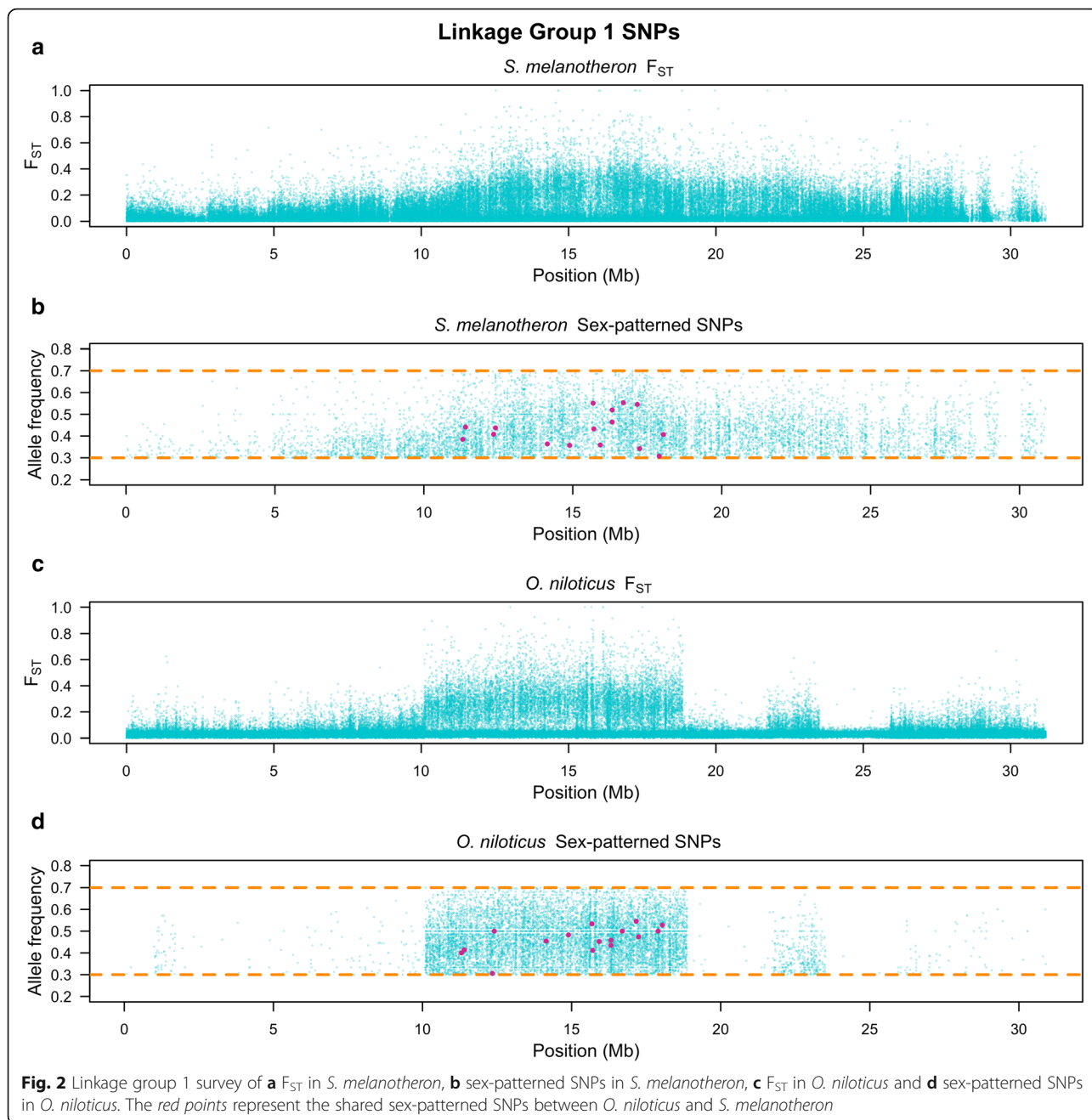
18.9 Mb, LG1 between 21.7 and 23.6 Mb, then across the rest of the genome. There were six stop codon gains or losses, but none mapped to the regions of divergence (Additional file 2). These counts differ slightly from Gammerdinger et al., 2014 due to a difference in the gene annotation versions used. There were no conserved stop codon gains or losses between *S. melanothereon* and *O. niloticus*.

Copy number variants

We identified a significantly higher density of conserved duplications and deletions between *O. niloticus* and *S. melanothereon* on both LG1 from 10.1 to 18.9 Mb ($p < 0.0001$) and LG1 from 21.7 to 23.6 Mb ($p < 0.0002$) when each is compared to the rest of the genome. However, we did not detect any significant difference in the density of duplications and deletions conserved between the species in these divergent regions on LG1.

Shared SNPs

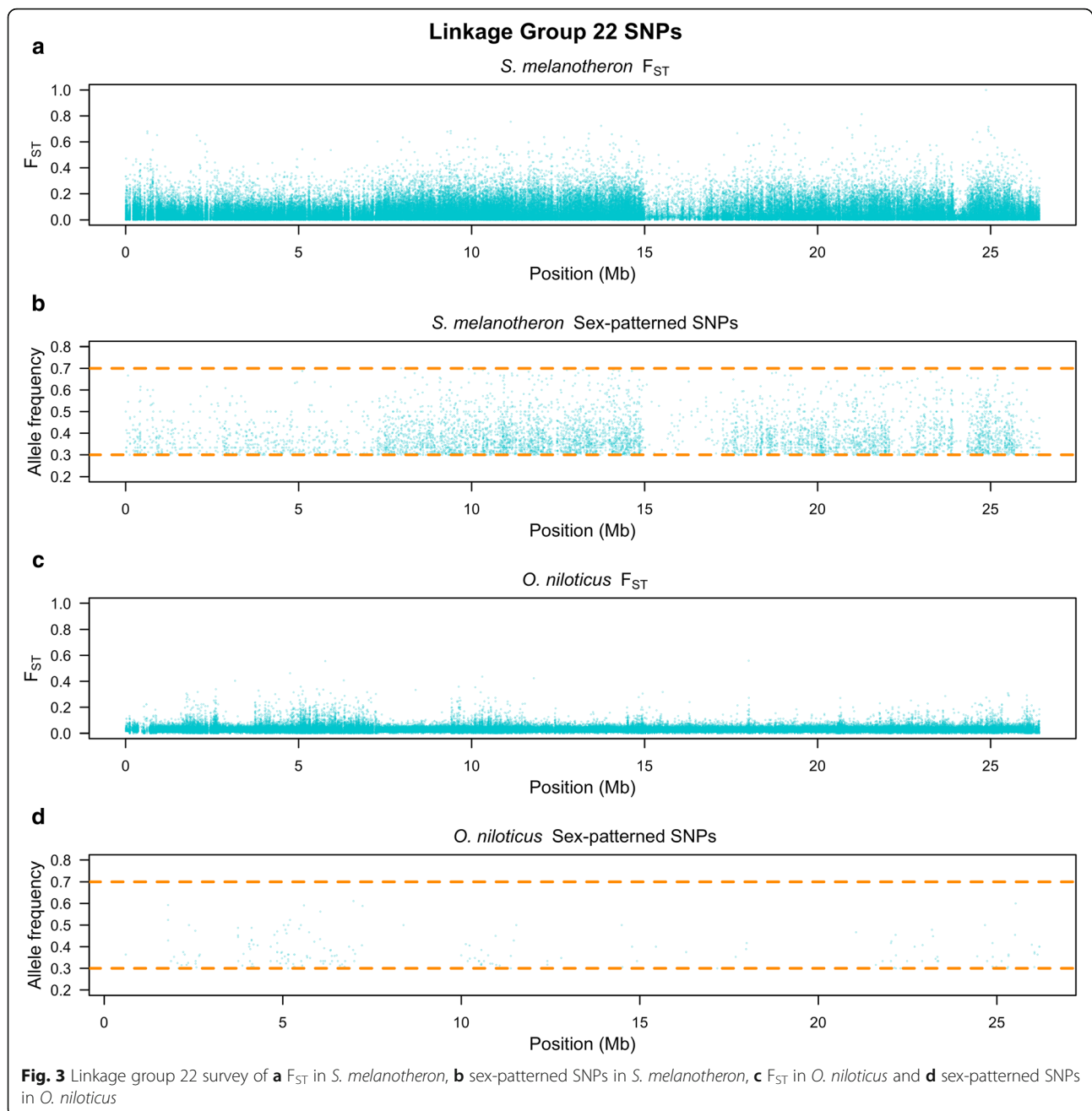
There are 42 sex-patterned SNPs conserved between *S. melanothereon* and *O. niloticus*, compared to the null expectation of 6.55 conserved SNPs. We examined all 42 positions to confirm that the X- and Y-alleles were consistent in the two species. There were 13 SNPs where the X- and Y-alleles are switched between species (e.g. when the X-allele in *O. niloticus* is the Y-allele in *S. melanothereon*). Eight sites showed a pattern where the Y-alleles are different but the X-alleles were the same in



each lineage. This narrowed the list to 21 SNPs where the X- and Y-alleles were shared between the two species. We filtered these 21 SNPs by genomic position. Sixteen fell within the region on LG1 between 10.1 and 18.9 Mb, one lay on LG14 and four fell on unanchored scaffolds (Additional files 3 and 4). Of these shared SNPs, none were missense mutations and only one was a synonymous mutation. The rest were located in non-coding regions. We identified shared sex-patterned SNPs within 20 kb of two previously identified candidate genes for sex-determination on LG1 [15]. One was 19 kb

downstream of *Wilms' tumor protein* and the other was 5 kb downstream of *Ras association domain containing protein 10*.

We used JASPAR to examine a short region flanking each of the 21 shared sex-patterned SNPs for transcription factor binding sites (Additional file 5). A sex-patterned SNP at position 14,895,959 on LG1, near *Wilms' tumor protein*, resulted in a loss of a *Gata4* binding site. The sex-patterned SNP at 11,400,015 on LG1, near *Ras association domain containing protein 10* did not alter binding sites for any transcription factor



binding sites known to be involved in the sex-determination network (Additional file 5).

Discussion

Evidence for a shared LG1 sex-determination mechanism between 10.1 Mb and 18.9 Mb

The initial line of evidence that *S. melanothereon* shares an ancestral sex-determination mechanism with *O. niloticus* is the overlap of the region harboring sex-determination on LG1 (Figs. 1 and 2). Convergence for sex-determination within the same 8.8 Mb region within

a 927 Mb genome, while possible, seems unlikely. This region also contains the highest density of sex-patterned SNPs across the genome in both species. The high density of sex-patterned SNPs explains the elevated sex-related F_{ST} witnessed within this region. Furthermore, the copy number variant analysis demonstrates that the region on LG1 from 10.1 to 18.9 Mb is not significantly more divergent than the region from 21.7 to 23.6 Mb, but it does have significantly more conserved insertions and deletions compared to the rest of the genome. Lastly, the number of shared sex-patterned SNPs is

Table 1 Average F_{ST} across all polymorphic sites along with total counts and densities of sex-patterned SNPs across the regions of differentiation in *S. melanothereon* and *O. niloticus*

	Average F_{ST} across polymorphic sites	Sex-patterned SNPs (per Mb)	Enriched 10 kb windows (per Mb)	Missense SNPs (per Mb)	PROVEAN SNPs < -2.5 (per Mb)	Synonymous SNPs (per Mb)
<i>S. melanothereon</i>						
LG1: 10.1 Mb-18.9 Mb	0.111	5342 (607.05)	193 (21.93)	61 (6.93)	6 (0.68)	121 (13.75)
LG1: 18.9 Mb-28 Mb	0.081	2702 (296.92)	52 (5.71)	38 (4.18)	5 (0.55)	40 (4.40)
LG22	0.057	5709 (216.16)	80 (3.03)	42 (1.59)	9 (0.34)	108 (4.09)
Rest of the genome	0.034	36,335 (41.13)	123 (0.14)	669 (0.76)	103 (0.12)	962 (1.09)
Total		50,088	448	810	123	1231
<i>O. niloticus</i>						
LG1: 10.1 Mb-18.9 Mb	0.039	10,792 (1226.36)	517 (58.75)	159 (18.07)	25 (2.84)	256 (29.09)
LG1: 21.7 Mb-23.6 Mb	0.024	399 (210.00)	3 (0.33)	5 (2.63)	1 (0.53)	6 (3.16)
Rest of the genome	0.021	8602 (9.38)	19 (0.02)	167 (0.18)	27 (0.03)	180 (0.20)
Total		19,793	539	331	53	442

several times more than expected by chance (6.55 shared sex-patterned SNPs). Collectively, the evidence strongly suggests the idea that *O. niloticus* and *S. melanothereon* share a common ancestral sex-determination mechanism on LG1.

Evidence for a shared sexually antagonistic locus on LG1 between 21.7 Mb and 23.6 Mb

The region from 21.7 to 23.6 Mb on LG1 has a moderately strong signal of differentiation between males and females in both lineages. The region from 18.9 to 28 Mb in *S. melanothereon* includes this region and is the second most densely sex-patterned SNP region in the genome. The region from 21.7 to 23.6 Mb does not have significantly more conserved insertions and deletions than the region from 10.1 to 18.9 Mb, but it does have significantly more conserved insertions and deletions when compared to the rest of the genome. This suggests that the block of differentiation between 21.7 and 23.6 Mb has a shared deep history with the sex-determination region.

Within this region, we found no shared SNPs with a consistent allelic pattern between *O. niloticus* and *S. melanothereon*, but there was a sex-patterned locus in both species where the X- and Y-alleles were switched. This suggests that there could have been an ancestral polymorphism that was alternatively sorted during speciation and the emergence of the sex chromosome. The evidence suggests that the diverged region on LG1 from 21.7 to 23.6 Mb was present during the initial divergence of sex-determination ancestral to both *O. niloticus* and *S. melanothereon* on LG1 from 10.1 to 18.9 Mb.

This linked, differentiated region from 21.7 to 23.6 Mb may harbor a sexually antagonistic locus. However, the sequence between these regions of divergence, from 18.9 to 21.7 Mb, shows differentiation

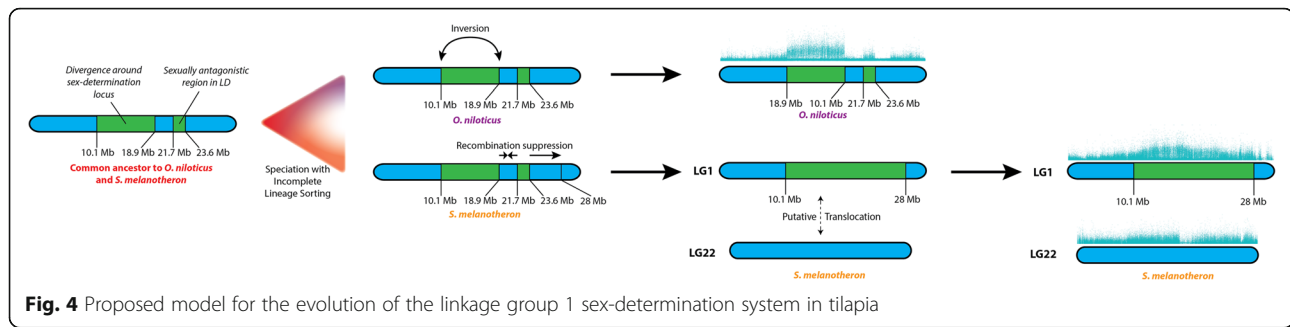
in *S. melanothereon*, but not in *O. niloticus*. Previous theoretical work has suggested that sexually antagonistic loci can be in linkage disequilibrium with the sex-determination locus over large distances while not showing strong signs of differentiation between the two loci [32]. It appears that *O. niloticus* has potentially maintained linkage disequilibrium between these two pockets of differentiation, while *S. melanothereon* has accumulated mutations across this entire region. This theoretical prediction also postulates that the sexually antagonistic region should have a lower level of divergence than the sex-determination region. This proposition is consistent with the data collected in *O. niloticus*.

From our data, it is unclear if the intervening region from 18.9 to 21.7 Mb diverged initially and then recombination was restored in the *O. niloticus* lineage or if these were initially two separate regions of divergence held together by sexually antagonistic selection and recombination was subsequently reduced across the entire region in the *S. melanothereon* lineage. Figure 4 illustrates the latter sequence of events, but the former sequence of events is also plausible. Sequences of more species harboring the LG1 sex-determination system might distinguish these two hypotheses.

Alternatively, this region could represent a misassembly in the *O. niloticus* reference genome or a Y-specific structural rearrangement.

Evidence for speciation soon after sex chromosome emergence

Of the 42 loci that demonstrate a sex-patterned profile in both data sets, 21 loci had the X- and Y-alleles conserved, while 13 loci had X- and Y-alleles switched. The 21 conserved X- and Y-alleles indicate that this sex chromosome had a shared ancestry between *O. niloticus*



and *S. melanotheron*. The 13 loci where the X- and Y-alleles were switched, indicate that this period of shared ancestry was limited, and that some sites that underwent alternative lineage sorting as the proto-X and proto-Y were diverging. This indicates that speciation occurred soon after sex chromosome divergence.

Evidence for structural rearrangements and their timing

The pronounced boundaries of divergence, high densities of sex-patterned SNPs and predicted deleterious mutations above the background level in *O. niloticus*, when taken together with the knowledge that chromosomes of this species are homomorphic in cytogenetic analyses, provide strong evidence for an inversion between the X- and Y-chromosomes in *O. niloticus* [33]. The absence of similar sharp boundaries in *S. melanotheron*, along with lower densities of sex-patterned SNPs and predicted deleterious mutations, suggest that this inversion is likely not present in the *S. melanotheron* lineage. We suggest the most parsimonious scenario is that this inversion arose following the divergence of *O. niloticus* and *S. melanotheron*.

The differentiation between males and females on LG22 of *S. melanotheron* suggests an association between this linkage group and sex-determination. While the levels of differentiation are lower than those witnessed on LG1 between 10.1 and 28 Mb, LG22 has a high density of sex-patterned deleterious mutations and the overall level of differentiation is significantly higher than the genomic background. Currently, this patterned is consistent with four hypotheses: 1) The signal could be the result of a reciprocal translocation of similar sized fragments between LG1 and LG22 after the divergence of *O. niloticus* and *S. melanotheron*. A previous karyotyping study has concluded that the chromosomes are homomorphic and that *S. melanotheron* have the same chromosome count as *O. niloticus*, but this study might have failed to detect such a translocation [17]. 2) LG22 is in linkage disequilibrium with the LG1 XY system. This seems unlikely because of the strong selection needed to overcome linkage disequilibrium between two independently assorting chromosomes. 3) A region on

LG22 epistatically contributes to sex-determination within LG1 in this family. Multiple sex-determination systems have been observed in some cichlid species, however no sex-determination loci been previously reported from LG22 in cichlids [7, 8, 34]. 4) The signal could be an artifact of the sampling structure. This also seems unlikely as it would suppose that two copies of LG22 were segregating in the family and that males disproportionately received more of one copy and females received almost none of that copy by chance meiotic events. We suggest that a reciprocal translocation from LG22 to LG1 following the divergence of *S. melanotheron* and *O. niloticus* is the most likely scenario for explaining the observed patterned of sex-patterned differentiation on LG22. Future cytogenetic studies may provide data to test this hypothesis.

We summarize our findings by proposing a model for the evolution of the LG1 sex chromosome in tilapia (Fig. 4). The sex-determination mechanism resided on LG1 in the common ancestor of *O. niloticus* and *S. melanotheron*. Early stages of sex differentiation encompassed a region on LG1 from 10.1 to 18.9 Mb. The region from either 18.9 or 21.7 to 23.6 Mb was maintained in linkage disequilibrium with sex-determination locus by sexually antagonistic selection [32]. Due to a lack of shared SNPs between 18.9 and 21.7 Mb, we are unable to determine whether this region was originally not part of the diverging region and recombination was reduced only in the *S. melanotheron* lineage (Fig. 4) or this region was originally part of a block of divergence and recombination was restored to it in the *O. niloticus* lineage. *Oreochromis niloticus* and *S. melanotheron* diverged shortly after the emergence of the novel sex-determination locus. We propose that in the *O. niloticus* lineage, the region from 10.1 to 18.9 Mb experienced a chromosomal inversion and began to rapidly accumulate mutations. In the *S. melanotheron* lineage, the region did not experience this inversion and the region of sex differentiation expanded to encompass a region from 10.1 to 28 Mb. In *S. melanotheron*, LG1 may have further experienced a translocation of a region from LG22.

Candidate sex-determination mutations

We identified 21 shared SNPs with consistent sex-patterned profiles in *S. melanotheron* and *O. niloticus*, which represent candidate sex-determination mutations for this lineage. Two of these are located near previously identified sex-determination candidate genes, so we evaluated how these shared sex-patterned SNPs might alter transcription factor binding sites. The mutation near *Ras association domain containing protein 10* did not affect any transcription factor binding sites known to play a role in sex-determination. The mutation near *Wilms tumor protein* altered a *Gata4* transcription factor binding site. *Gata4* is a transcription factor that has been associated with the sex-determination pathway [35]. *Gata4* has been shown to bind in the promoter of the *Wilms tumor protein* in mammals and teleosts and *Wilms tumor protein* is a strong activator of *Amhr*, a known sex-determination gene [36, 37]. A Y-specific loss of a *Gata4* binding site would decrease transcription of *Wilms tumor protein* in males, which in turn would not activate *Amhr* and result in masculinization. This variant currently represents the strongest SNP candidate for sex-determination on LG1 and further functional analysis is necessary to assess its role in sex-determination.

Our analysis is limited by the inability of the short-read Illumina data to clearly resolve the large number of deletions and insertions that are likely present on the Y-chromosome. Future studies should consider using longer read sequencing technologies to assemble X- and Y-specific sequences for both species. It may then be possible to more clearly identify conserved insertion and deletion events on the proto-Y chromosome in both species.

Conclusions

This study utilized a comparative approach in order to identify the ancestral state of the LG1 sex-determination locus. We determined that *S. melanotheron* shares an ancestral XY sex-determination mechanism with *O. niloticus*. In both species, the highest differentiation between the X- and Y-chromosomes is found in the region from 10.1 to 18.9 Mb on LG1. We identified 21 sex-patterned SNPs shared between the two species. One of these SNPs alters a *Gata4* transcription factor binding site near *Wilms tumor protein*, which might alter the function of the sex-determination pathway. Future studies should assess the role of this candidate SNP in sex-determination. Investigation of additional species segregating this sex-determination system on LG1 will provide a clearer understanding of evolutionary processes during the early stages of sex chromosome divergence.

Additional files

Additional file 1: Stop codon altering sex-patterned SNPs in *S. melanotheron*. (XLSX 39 kb)

Additional file 2: Stop codon altering sex-patterned SNPs in *O. niloticus*. (XLSX 37 kb)

Additional file 3: Shared sex-patterned SNPs between *O. niloticus* and *S. melanotheron* within intergenic region and the genes on either side of each shared sex-patterned SNP. *These shared sex-patterned SNPs had a gene on one side and the end of the scaffold on the other side (XLSX 32 kb)

Additional file 4: Shared sex-patterned SNPs between *O. niloticus* and *S. melanotheron* within intragenic regions and their position within those genes. (XLSX 47 kb)

Additional file 5: JASPAR analysis of transcription factor binding sites in the regions encompassing the shared SNPs. The uppercase letters in the sequence columns represent the SNP in question. The bolded SNPs are the SNPs within 20 kb of previously identified candidate genes for sex-determination. (XLSX 81 kb)

Abbreviation

LG: Linkage group

Acknowledgements

The authors acknowledge the University of Maryland supercomputing resources (<http://www.it.umd.edu/hpcc>) made available for conducting the research reported in this paper. The authors would like to thank Aston Belew for his programming advice.

Funding

This work was supported in part by the National Science Foundation under Grant Number DEB-1143920 and through a U.S. Department of Education Graduate Assistance in Areas of National Need fellowship offered through the University of Maryland Biology Department under Award Number P200A150160. Partial funding for open access provided by the UMD Libraries' Open Access Publishing Fund.

Availability of data and materials

The *Oreochromis niloticus* male and female reads are available from NCBI under SRR1606298 and SRR1606304, respectively. The *Sarotherodon melanotheron* male and female reads are available NCBI under SRR3473459 and SRR3473461, respectively.

Authors' contributions

WJG helped conceive the study and was involved in DNA extraction, DNA quantification, computational analyses and drafting of the manuscript. MAC was involved in computational analyses and editing of the manuscript. HD raised the specimens, collected the fin clips and editing of the manuscript. J-FB raised the specimens and collected the fin clips. TDK helped conceive the study, constructed the pooled DNA libraries, guided the analysis and helped draft the manuscript. All authors have read and approved the manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable

Ethics approval

All animal procedures were conducted in accordance with University of Maryland IACUC Protocol #R-10-73. All the experimental procedures took place in CIRAD facilities in Montpellier under the Laboratory agreement for animal experimentation number A-34-172-24 and the author's personal authorization for animal experimentation N° 35-15, both delivered by the French Government.

Author details

¹Department of Biology, University of Maryland, College Park, MD 20742, USA. ²UMR-ISEM, CIRAD, Montpellier, France.

Received: 11 June 2016 Accepted: 13 October 2016

Published online: 18 October 2016

References

- Graves J. Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. *Annu Rev Genet.* 2008;42:565–86.
- Bachtrog D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet.* 2013;14:113–24.
- Rice WR. Evolution of the Y sex in animals: Y chromosomes evolve through the degeneration of autosomes. *Bioscience.* 1996;46:331–43.
- Muller H. The relation of recombination to mutational advance. *Mutat Res.* 1964;1:2–9.
- Lahn BT, Page DC. Four evolutionary strata on the human X chromosome. *Science.* 1999;286:964–7.
- Kocher TD. Adaptive evolution and explosive speciation: the cichlid fish model. *Nat Rev Genet.* 2004;5:288–98.
- Ser JR, Roberts RB, Kocher TD. Multiple interacting loci control sex determination in lake Malawi cichlid fish. *Evolution.* 2010;64:486–501.
- Cnaani A, Lee B-Y, Zilberman N, Ozouf-Costaz C, Hulata G, Ron M, D'Hont A, Baroiller J-F, D'Cotta H, Penman DJ, Tomasino E, Coutanceau JP, Pepey E, Shirak A, Kocher TD. Genetics of sex determination in tilapia species. *Sex Dev.* 2008;2:43–54.
- Parnell NF, Streebman JT. Genetic interactions controlling sex and color establish the potential for sexual conflict in Lake Malawi cichlid fishes. *Heredity.* 2013;110:239–46.
- Eshel O, Shirak A, Dor L, Band M, Zak T, Markovich-Gordon M, Chalifa-Caspi V, Feldmesser E, Weller JI, Seroussi E, Hulata G, Ron M. Identification of male-specific amh duplication, sexually differentially expressed genes and microRNAs at early embryonic development of Nile tilapia (*Oreochromis niloticus*). *BMC Genomics.* 2014;15:774.
- Li M, Sun Y, Zhao J, Shi H, Zeng S, Ye K, Jiang D, Zhou L, Sun L, Tao W, Nagahama Y, Kocher TD, Wang D. A tandem duplicate of anti-müllerian hormone with a missense SNP on the Y chromosome is essential for male sex determination in Nile Tilapia, *Oreochromis niloticus*. *PLoS Genet.* 2015;11:e1005678.
- Lee B-Y, Penman DJ, Kocher TD. Identification of a sex-determining region in Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis. *Anim Genet.* 2003;34:1–5.
- Ezaz MT, Harvey SC, Boonphakdee C, Teale AJ, McAndrew BJ, Penman DJ. Isolation and physical mapping of sex-linked AFLP markers in Nile tilapia (*Oreochromis niloticus* L.). *Mar Biotechnol.* 2005;6:435–45.
- Palaiokostas C, Bekaert M, Khan MGQ, Taggart JB, Gharbi K, McAndrew BJ, Penman DJ. Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) using RAD sequencing. *PLoS One.* 2013; 8:e68389.
- Gammerding WJ, Conte MA, Acquah EA, Roberts RB, Kocher TD. Structure and decay of a proto-Y region in Tilapia, *Oreochromis niloticus*. *BMC Genomics.* 2014;15:975.
- Nagl S, Tichy H, Mayer WE, Samonte IE, McAndrew BJ, Klein J. Classification and phylogenetic relationships of African tilapia species inferred from mitochondrial DNA sequences. *Mol Phylogenet Evol.* 2001;20:361–74.
- Harvey SC, Powell SF, Kennedy DD, McAndrew BJ, Penman DJ. Karyotype analysis of *Oreochromis mortimeri* (Trewavas) and *Sarotherodon melanotheron* (Rüppell). *Aquac Res.* 2002;33:339–42.
- FASTQC. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, Simakov O, Ng AY, Lim ZW, Bezaul E, Turner-Maier J, Johnson J, Alcazar R, Noh HJ, Russell P, Aken B, Alföldi J, Amemiya C, Azzouzi N, Baroiller J-F, Barloy-Hubler F, Berlin A, Bloomquist R, Carleton KL, Conte MA, D'Cotta H, Eshel O, Gaffney L, Galibert F, Gante HF, Gnerre S, Greuter L, Guyon R, Haddad NS, Haerty W, Harris RM, Hofmann HA, Hourlier T, Hulata G, Jaffe DB, Lara M, Lee AP, MacCallum I, Mwaiko S, Nikaido M, Nishihara H, Ozouf-Costaz C, Penman DJ, Przybylski D, Rakotomanga M, Renn SCP, Ribeiro FJ, Ron M, Salzburger W, Sanchez-Pulido L, Santos ME, Searle S, Sharpe T, Swofford R, Tan FJ, Williams L, Young S, Yin S, Okada N, Kocher TD, Miska EA, Lander ES, Venkatesh B, Fernald RD, Meyer A, Ponting CP, Streebman JT, Lindblad-Toh K, Seehausen O, Di Palma F. The genomic substrate for adaptive radiation in African cichlid fish. *Nature.* 2014;513:376–81.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2013;9:357–9.
- DePristo MA, Banks E, Poplin RE, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43:491–8.
- Sex_SNP_Finder. <https://github.com/Gammerding/sex-SNP-finder>.
- Kofler R, Pandey RV, Schlötterer C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics.* 2011;27:3435–6.
- Nei M. *Molecular evolutionary genetics.* New York: Columbia University Press; 1987.
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Ruden DM, Lu X. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly.* 2012;6:80–92.
- Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, Lu X. Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Front Genet.* 2012;3:35.
- Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One.* 2012;7:e46688.
- Quinlan AR, Hall IM. Genome analysis BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26:841–2.
- Mathelier A, Fornes O, Arenillas DJ, Chen C-Y, Denay G, Lee J, Shi W, Shyr C, Tan G, Worsley-Hunt R, Zhang AW, Parcy F, Lenhard B, Sandelin A, Wasserman WW. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 2016;44:D110–5.
- Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 2012;22:568–76.
- Varscan_multiple. https://github.com/Gammerding/Varscan_multiple.
- Kirkpatrick M, Guerrero R. Signatures of sex-antagonistic selection on recombining sex chromosomes. *Genetics.* 2014;197:531–41.
- Oliveira C, Wright JM. Molecular cytogenetic analysis of heterochromatin in the chromosomes of tilapia, *Oreochromis niloticus* (Teleostei: Cichlidae). *Chromosome Res.* 1998;6:205–11.
- Lee B-Y, Hulata G, Kocher TD. Two unlinked loci controlling the sex of blue tilapia (*Oreochromis aureus*). *Heredity.* 2004;92:543–9.
- Munger SC, Aylor DL, Syed HA, Magwene PM, Threadgill DW, Capel B. Elucidation of the transcription network governing mammalian sex determination by exploiting strain-specific susceptibility to sex reversal. *Genes Dev.* 2009;23:2521–36.
- Miyamoto Y, Taniguchi H, Hamel F, Silversides DW, Viger RS. A GATA4/WT1 cooperation regulates transcription of genes required for mammalian sex determination and differentiation. *BMC Mol Biol.* 2008;9:e1002798.
- Kamiya T, Kai W, Tasumi S, Oka A, Matsunaga T, Mizuno N, Fujita M, Suetake H, Suzuki S, Hosoya S, Tohari S, Brenner S, Miyadai T, Venkatesh B, Suzuki Y, Kikuchi K. A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (Fugu). *PLoS Genet.* 2012;8:e1002798.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

