npg

ORIGINAL ARTICLE

Copy number variation of lipocalin family genes for male-specific proteins in tilapia and its association with gender

A Shirak[1], M Golik[1], B-Y Lee[2], AE Howe[2], TD Kocher[2], G Hulata[1], M Ron[1] and E Seroussi[1]

[1]Institute of Animal Science, Agricultural Research Organization, Volcani Center, Beit Dagan, Israel and [2]Department of Biology, University of Maryland, College Park, MD, USA

Lipocalins are involved in the binding of small molecules like sex steroids. We show here that the previously reported tilapia male-specific protein (MSP) is a lipocalin encoded by a variety of paralogous and homologous genes in different tilapia species. Exon–intron boundaries of MSP genes were typical of the six-exon genomic structure of lipocalins, and the transcripts were capable of encoding 200 amino-acid polypeptides that consisted of a putative signal peptide and a lipocalin domain. Cysteine residues are conserved in positions analogous to those forming the three disulfide bonds characteristic of the ligand pocket. The calculated molecular mass of the secreted MSP (20.4 kDa) was less than half of that observed, suggesting that it is highly glycosylated like its homologue tributyltin-binding protein. Analysis of sequence variations revealed three types of paralogs MSPA, MSPB and MSPC. Expression of both MSPA and MSPB was detected in

testis. In haploid Oreochromis niloticus embryos, each of these types consisted of two closely related paralogs, and asymmetry between MSP copy numbers on the maternal (six copies) and the paternal (three copies) chromosomes was observed. Using this polymorphism we mapped MSPA and MSPC to linkage group 12 of an F_2 mapping family derived from a cross between O. niloticus and Oreochromis aureus. Females with high MSP copy number were more frequent by more than twofold than males. Gender–MSPC combinations showed significant deviation from expected Mendelian segregation ($P = 0.009$) suggesting elimination of males with MSPC copies. We discuss different hypotheses to explain this elimination, including possibility for allelic conflict resulted by the hybridization.
Heredity (2008) 101, 405–415; doi:10.1038/hdy.2008.68; published online 23 July 2008

Keywords: sex determination; CNV; QTL; cichlidae; cichlid family; orosomucoid

Introduction

Sex determination (SD) in fish is poorly characterized and might frequently undergo evolutionary changes bringing the sex-determining cascade under new master sex regulators (Volff et al., 2007). True tilapias (Tilapiini) are a group of approximately 50 species of perch-like fishes, taxonomically divided into three genera Oreochromis, Sarotherodon and Tilapia (Trewavas, 1983; Majumdar and McAndrew, 1986). Several studies have identified genetic markers linked to sex in tilapia. Genetic mapping of genes for SD in purebred tilapia species indicated that SD may be regarded as a quantitative trait (Lee et al., 2004) and several linkage groups (LGs) were implicated. Lee et al. (2003) identified an XY system on LG1 in a strain of Oreochromis niloticus, and Lee et al. (2004) identified two epistatically interacting quantitative trait loci (QTL)—one on LG1 and the other on LG3 (Lee et al., 2004). QTL for sex-specific mortality (SSM) were detected on LG2, 6 and 23 in an inbred line of O. aureus (Palti et al., 2002; Shirak et al., 2002). Two distinct QTL for SD were also reported on LG23 in an F_2 hybrid cross

between Oreochromis aureus and Oreochromis mossambicus (Cnaani et al., 2003, 2004).

A genetic linkage map in tilapia, based on an F_2 mapping family derived from a cross between two tilapia species: O. aureus and O. niloticus was previously reported (Lee et al., 2005). Male to female ratio in this family was skewed around two to three. In this family a major QTL for sex was mapped to a wide region of more than 13 cM on LG3 (Lee et al., 2005), which did not provide complete explanation for gender variance nor for the skewed sex ratio.

Distortions from 1:1 sex ratio are common in tilapia interspecific crosses (Avtalion and Hammerman, 1978). These distortions were explained by (1) increased sensitivity of hybrids to environmental influence, including thermosensitivity (Desprez et al., 2006); (2) SSM (Shirak et al., 2002); and (3) imbalance in polygenic sex determination (Lee et al., 2004).

Cytochrome P450 aromatase catalyzing conversion of androgens to estrogens probably plays a critical role for ovarian differentiation in tilapia (Baroiller and D'Cotta, 2001). Exogenous androgens, estrogens, antiestrogens and aromatase inhibitors trigger sex reversal in tilapias, suggesting that they are sensitive to the levels of sex steroids at a critical period of development, like other fish species. In this period, environmental factors like temperature, pH, density and social interactions influence sex ratio (Baroiller and D'Cotta, 2001).

Another agent having endocrine-disrupting effect in fish and other aquatic organisms is tributyltin (TBT). This trialkyl organotin compound can induce masculinization in genetically females of Japanese flounder (*Paralichthys olivaceus*; Shimasaki *et al.*, 2003) and affect sexual behavior and reproduction in medaka (*Oryzias latipes*; Nakayama *et al.*, 2004). TBT accumulates in the serum of Japanese flounder as a complex with TBT-binding protein (TBT-bp; Shimasaki *et al.*, 2002). Mature TBT-bp is heavily glycosylated (Hano *et al.*, 2007) with a molecular mass of approximately 46.5 kDa, and only 191 amino acids with a lipocalin-like sequence pattern.

A similar protein was detected in the serum of sexually active males in different tilapia species as a male-specific electrophoretic band corresponding to a 41 kDa protein (Avtalion *et al.*, 1975; Avtalion, 1982). In dominant *Sarotherodon galilaeus* males the level of this male-specific protein (MSP) could achieve concentration of $5.2 \, \text{mg ml}^{-1}$, which is more than 1% of total serum proteins (Kirsh, 1991) and about fourfold higher than the concentration in average sexually active males. The level of MSP reaches a maximum at spawning time when MSP is also detectable in females (Trombka and Avtalion, 1993).

Adopting a candidate gene approach, we have systematically mapped genes that are potentially involved in SD and/or SSM of tilapias (Shirak *et al.*, 2006b). The resemblance of sequence motifs between TBT-bp and MSP prompted us to characterize the latter genes in tilapia and study their possible role in SD.

Materials and methods

Mapping family

A red *O. niloticus* male from the University of Stirling stock (McAndrew and Majumdar, 1983) was crossed with a normally colored *O. aureus* female from an Israeli Mehadrin strain (Lee *et al.*, 2004). The F_2 family consisted of 156 offspring (61 males, 91 females and 4 unsexed individuals) of which 90 individuals were used for genotyping to create the second generation map (Lee *et al.*, 2005).

Collection of samples from other tilapia species and DNA extraction

Fin samples of 40 individuals from different tilapia species and strains (detailed in Table 1) were collected for DNA extraction and sequencing in order to determine and compare *MSP* copies. DNA was extracted from fin tissue by the salting out procedure (Ma *et al.*, 1996).

Production and analysis of haploid embryos

UV-irradiated *Tilapia zillii* sperm was used for the induction of gynogenetic haploids in eggs of a single Ghana-BIU *O. niloticus* female carrying the three *MSP* types according to procedure previously described (Shirak and Avtalion, 2001). This was followed by lysis of embryonic tissue from the haploids and DNA extraction (Gates *et al.*, 1999). Tracefiles and Genome Assembly Program 4 (GAP4) database including assembled and annotated sequences of type-specific PCR products obtained from the haploid embryos is downloadable at http://cowry.agri.huji.ac.il/MSP/. PCR amplification reactions (10 μl) were performed using Super-Therm polymerase (0.5 U; JMR Holdings, St Louis, MO, USA) according to the instruction of the manufacturer in the following conditions: 33 cycles for 40 s at 92 °C, 30 s at 61 °C and 1 min at 72 °C. PCR products were separated and sequenced as described in the next section.

Prediction of intron–exon borders, PCR primer design and sequence analysis

Genomic and expressed sequence tag sequences related to *MSP* were detected using the Basic Local Alignment Search Tool programs on the National Center for Biotechnology Information/National Institutes of Health server (http://www.ncbi.nlm.nih.gov/BLAST). Primers were designed with Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) and PCR

Table 1 *MSP* gene types in different tilapia species and strains

Species	Strain[a]	Sample size	Females	Males	MSP paralogs[b]
O. niloticus	Stirling Red	3	1	2	MSPA, MSPB
	Swansea	2	1	1	MSPA, MSPB
	Ghana-Dor	6	3	3	MSPA, MSPB
		2	2	0	MSPA, MSPB, MSPC
	Ghana-BIU	2	1	1	MSPA, MSPB
		2	2	0	MSPA, MSPB, MSPC
O. aureus	Mehadrin	5	2	3	MSPA, MSPB
		2	2	0	MSPA, MSPB, MSPC
	Stirling	2	1	1	MSPA, MSPB
	Ein Feshka	2	1	1	MSPA, MSPB
	Lake of Galilee	2	1	1	MSPA, MSPB
	BIU-1 line	2	1	1	MSPA, MSPB
F₁ hybrids	Mapping family	1	0	1	MSPA, MSPB
		1	1	0	MSPA, MSPB, MSPC
O. mossambicus	Natal-Volcani	2	1	1	MSPA
S. galilaeus	Volcani	2	1	1	MSPA
T. zillii	Yarkon River	2	1	1	MSPA
Total		40	22	18	

Abbreviation: MSP, male-specific protien.
[a]Various experimental strains held at the ARO aquaculture research unit, or recently collected from various locations in Israel.
[b]The presence of the paralogs was determined by analyzing sequencing chromatograms (see below in the 'Results' section).

Table 2 PCR primer pairs used for the characterization of *MSP* genes

No.	Forward	No.	Reverse	Amplicon	
				Spans	Size (bp)[a]
1	TCTGGAGGATCGTAGCAAGG	2	TGGCAGTGACATTTCCATGA	Introns 1 and 2	831
3	GGAAATGTCACTGCCACCTTC	4	CCAAGATGCTCATGACTGGA	Intron 3	168
5	TGGAACATCAGACAGCGAAG	2	TGGCAGTGACATTTCCATGA	Intron 2	454
6	tet-AACTCCCTGAGGCCTGAATC	7	CCACATGCATAAACATGCTC	Microsatellite	180

[a]Average size; variants slightly differ in length according to their microsatellite content.

fragments were amplified using BIO-X-ACT Long DNA polymerase kit (Bioline Ltd, London, UK) according to the instructions of the manufacturer and the following conditions: 30 cycles for 30 s at 92 °C, 40 s at 63 °C, and 1 min kb^{-1} at 68 °C. Primer pairs for the amplification of introns are described in Table 2. PCR products were separated on agarose gels, excised from the gel, purified with DNA Montage Gel Extraction Kit (Millipore, Bedford, MA, USA) and then sequenced using an ABI 3700 sequencer. Sequences were assembled and compared using the GAP4 program (Staden *et al.*, 2000). Protein sequences were aligned with ClustalW (http://clustalw. genome.jp), using default settings and Gonnet weight matrix. The graphical image of the multiple alignment was made using BoxShade (http://www.ch.embnet.org/ software/BOX_form.html). Predicted molecular mass was calculated with PeptideMass (http://www.expasy. ch/tools/peptide-mass.html) using default values.

Type-specific PCR amplification in *O. niloticus* and *O. aureus*

To determine and compare sequences of the three *MSP* gene types we used one forward primer (no. 1; Table 2) designed at the end of exon 1 and three different reverse primers with the consensus sequence AAATATCGA CAATTACGCTaTGN when 'N' represents T, C or A nucleotides for *MSPA*, *MSPB* or *MSPC*, respectively. 'N' is positioned at the unique base at the end of the second exon that distinguishes between these genes. Henceforth, this base position, seven nucleotides before the exon end, will be denoted *MSP* reference base (MSPrb). A mismatch at the fourth position from the primers' ends, indicated by lower case letter in the consensus sequence, was introduced to increase the specificity of the PCR amplification. PCR amplification reactions and sequencing was performed as above described in the section of analysis of haploid embryos.

Tissue collection for the expression examination

Two males and four females of *O. niloticus* Ghana-Dor (weight 200–350 g) were separated to a 100 l aquarium. Two days later, when male and females demonstrated the prespawning behavior, they were dissected, and samples of liver, testis and pronephros were collected. Dorsal fin samples were collected and stored in 70% ethanol, whereas internal organs were kept in RNA*later* (Ambion, Austin, TX, USA). Type-specific PCR analysis of genomic DNA (see 'Results') indicated that females and males were carriers of three and two *MSP* paralogs, respectively.

RNA extraction and analysis for gene expressions study

Tissue sections (0.5 g) of liver, testis and pronephros were stabilized in RNA*later* (Qiagen GmbH, Hilden, Germany).

RNA was extracted from samples (50 mg) immediately following homogenization in TRIzol (Gibco-BRL, Gaithersburg, MD, USA), according to the instructions of the manufacturer. Concentrations of 3.5–6 μg RNA per μl in final volume of 400 μl were obtained. *MSP* cDNA was reverse transcribed and PCR amplified using 10 pmol of each primer (no. 1 and no. 2; Table 2). Reverse transcription was performed using 1 μg RNA in a total volume of 20 μl using 200 U SuperScript (Invitrogen, Carlsbad, CA, USA) according to the instructions of the manufacturer. Intronless products were sequenced and analyzed.

Genotyping of microsatellite marker

Genotyping of microsatellite DNA marker in the first intron of *MSP* gene was performed as previously described (Palti *et al.*, 2002) using dye-labeled primers (no. 6 and no. 7; Table 2). To better distinguish between the alleles that differ in one base, the procedure was modified to use high fidelity Accuzyme DNA polymerase (Bioline Ltd, London, UK). Genotyping was performed using as template the PCR product obtained from type-specific PCR amplification. Prior to use, this product (1 μl) was treated by shrimp alkaline phosphatase (SAP 2 U; Promega, Madison, WI, USA) in 5 μl reaction containing SAP buffer, for 1 h at 37 °C following with heat inactivation at 65 °C for 15 min.

Genotyping of MSPrb

Analysis of variation in copy number of the *MSP* genes was performed using DNA MassArray technology (Jurinke *et al.*, 2001). PCR primers were designed using Sequenom's assay design software. For the first PCR amplification ACGTTGGATGGCCTTGAAGGTAGTAC AAATC and ACGTTGGATGCCATACCTGACAGTGA AGAC primers were used. The extension primer CAAAT ATCGACAATTACGCTTTG, positioned downstream of MSPrb was used for the second PCR.

Linkage mapping

Genotype data for the *MSP* genes (*MSPA1*, *MSPA2*, *MSPC1* and *MSPC2*) was added to genotype data of the mapping family for 545 markers, and mapping was performed using *JoinMap* software (3.0) as previously described (Lee *et al.*, 2005). We were not able to detect sequence variation between MSPB homologues and therefore this group of genes was not mapped.

Statistical comparison of models for SD and SSM

In order to evaluate and compare the effects of the *MSPA-MSPC* locus on SD and SSM, we built two statistical models. For this analysis, out of the 156 F2 individuals of the mapping family, we considered the

Table 3 Expected distribution of *MSP* genotypes between females and males under the SD model

$MSPA_{214}$-MSPC	Females	Males
Present	$75-0.5m+n$	$0.5m-n$
Absent	$75-0.5m$	$0.5m$

Table 4 Expected distribution of *MSP* genotypes between females and males under the SSM model

$MSPA_{214}$-MSPC	Females	Males
Present	$(75-0.5\,m)\,(150/(150\text{-}n))$	$(0.5m-n)\,(150/(150\text{-}n))$
Absent	$(75-0.5\,m)\,(150/(150\text{-}n))$	$0.5m\,(150/(150\text{-}n))$

Abbreviation: MSP, male-specific protien.

150 sexed individuals that were nonsergeants for *MSPA-MSPC* genes. We hypothesized that prior effect of this locus the numbers of males (m) and females ($150-m$) should be equally distributed between the two classes—with or without the $MSPA_{214}$-*MSPC* genotype. Under a SD model, the effect of the $MSPA_{214}$-*MSPC* genes may be increasing the number of females with $MSPA_{214}$-*MSPC* by number n, resulting in $((75-0.5m)+n)$ females with this genotypic combination, on account of the number of males with the same genotypic combination, which becomes $(0.5m-n)$ (Table 3).

Under a SSM model, *MSPA-MSPC* genes cause the elimination of n males that carried the $MSPA_{214}$-*MSPC* genotype. The expected segregations into the four classes, in a sample of 150 individuals, are calculated using the $(150/(150-n))$ coefficient (Table 4).

In both models, observed and expected segregations were used to solve the equations for minimal χ^2, n and m values by Newton's method of optimization (Tseng, 1998).

Results

MSP has a lipocalin signature motif typical of transporters of steroids

In a search for genes that may affect SD and sex reversal in tilapia, we encountered recent sequence information for male-specific protein (GenBank accession number, AAR19269) that was significantly similar (E-value of 2e−09) to TBT-bps (30% identity and 46% similarity to poTBT; 25% identity and 41% similarity to fhTBT; Figure 1). The alignment of the 200 predicted amino-acid residues of MSP with seven orthologous lipocalins (Figure 1) indicated that MSP had a similar domain structure including a signal peptide and a lipocalin domain with the three main structurally conserved regions (SCRs). The calculated molecular mass of MSP without the peptide leader (181 residues) was 20.4 kDa. Cysteine residues were present in the expected positions to form the three sulfide bonds of the lipocalin subfamily of retinol-binding protein-like proteins (Yamada *et al.*, 2006). These cysteines were implicated in the folding of a ligand pocket that mediates steroid transport by lipocalins such as hsORM1, which significantly resembled MSP (17% identity and 34% similarity; Figure 1) and binds to progesterone (Albani, 2006).

Exon–intron boundaries of the *MSP* gene are typical to the lipocalin six-exon gene structure

In order to generate genetic markers that would allow studying possible association of *MSP* with gender in tilapia we searched this gene for polymorphisms using the genomic DNA of the grandparents and parents of the tilapia mapping family (Lee *et al.*, 2005). We predicted the position of the intron–exon boundaries by aligning

the partial *MSP* transcript (GenBank accession number, EF661585) with a transcript derived from *Tetraodon nigroviridis* (GenBank accession number, CR693561) capable of encoding a similar lipocalin (40% identity and 65% similarity; Figure 1) with six exons that have been previously determined (Ensemble gene, GSTENT00028694001). We amplified the first, second and third introns (331, 301 and 87 bp, respectively) using PCR primers designed in the flanking exons. All exon–intron boundaries contained the canonical GT and AG intronic dinucleotides at the donor and acceptor splice sites, respectively (GenBank accession numbers, AM886141, AM886142 and AM886143). The position of these exon–intron boundaries was in agreement with the six-exon structure typical of the lipocalin genes (Figure 1).

Copy number variation and allele polymorphism of *MSP* genes in the tilapia mapping family

The sequence of PCR-amplified products spanning the second intron and parts of the adjacent exons displayed several ambiguous nucleotides in the matriarch (Mehadrin stock of *O. aureus*) and patriarch (Stirling Red *O. niloticus*) of the mapping family (Figure 2a, lanes *Oa* and *On*, respectively). At the end of exon 2 of the maternal sequence trace at the base denoted MSPrb, we observed an overlap of three nucleotides, that is A, G and T (Figure 2b), suggesting the presence of three *MSP* paralogs. Using type-specific PCR, which was based on MSPrb, we sequenced three distinct amplicons (GenBank accession numbers, AM886144, AM886145 and AM886146). The polymorphic nature of these sequences is exemplified in a stretch of 50 bases at the end of the second exon (Figure 2b) and in its corresponding translation (Figure 2c), where multiple nonconservative changes of amino acids were recorded. A model of inheritance with three types of paralogs explained all ambiguous nucleotides in the coding region. The existence of multiple variants within these types was evident from sequencing of the intronic regions, as was best exemplified by a region of the first intron that was rich in nucleotide repeats which we designated as MSP microsatellite. To better distinguish between these variants we used the products of the type-specific PCR amplification as a template for genotyping the microsatellite. This two-step procedure indicated that different combinations of length variants were associated with the *MSPA* (177, 178, 179 and 214 bp), the *MSPB* (177 and 178 bp) and the *MSPC* (180 and 181 bp) types. *ShiftDetector* analysis (Seroussi *et al.*, 2002) of the trace files obtained from sequencing of type-specific PCR products from individuals carrying different combinations of variants was in agreement with the observed length polymorphism. Arabic numerals were used to denote the closely related paralogs within the types according to the inheritance model suggested in discussion.
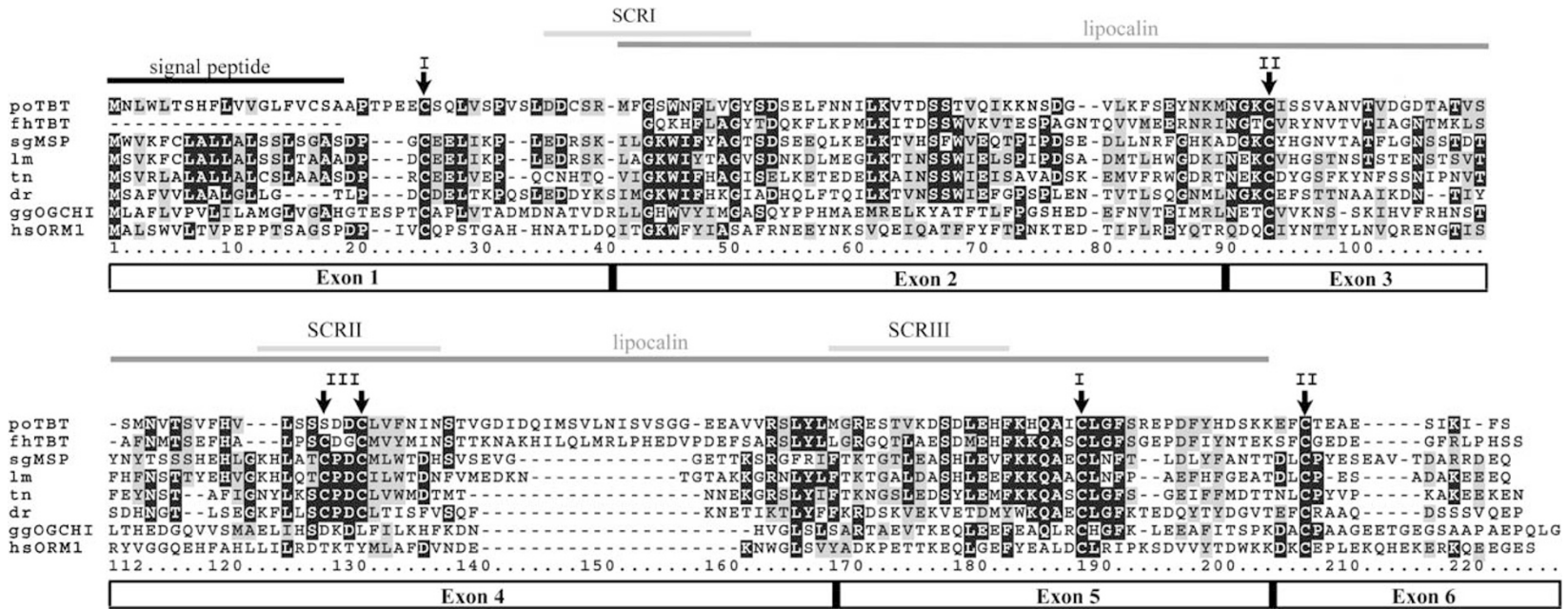
Figure 1 Alignment of male-specific protein (MSP) with similar lipocalins. The ClustalW alignment of predicted amino-acid sequences of eight orthologous lipocalins from *Homo sapiens* (hsORM1, orosomucoid 1, CAA29229); *Gallus gallus* (ggOGHI, α₁-acid ovoglycoprotein, ABU24464); *Danio rario* (dr, hypothetical protein XP_678073 isoform 1, TGIR Tentative Consensus TC315060); *T. nigroviridis* (tn, CR693561); *Lithognathus mormyrus* (lm, DQ887348); *S. galilaeus* (sgMSP, AAR19269); *Fundulus heteroclitus* (fhTBT, TBT-binding protein, AAU21488); and *P. olivaceus* (poTBT, BAB83525) is shown. Identity and similarity between the amino-acid sequences are indicated by black and grey boxes, respectively. White boxes indicate nonconservative amino-acid changes between the proteins. Dashes indicate gaps introduced by the alignment program. Two putative protein domains are marked above the alignment by horizontal bars according to the human protein sequence: a signal peptide and the lipocalin domain. Further above and following (Flower, 1996) the three main structurally conserved regions (SCRs) of the fold, SCR1, SCR2 and SCR3, are delineated. The three (I, II, III) sulfide bonds implicated in the folding of the lipocalin group of proteins (Yamada *et al.*, 2006) are denoted by arrows above the alignment. Below the alignment, the exon numbers of the exons capable of encoding the respective polypeptides above appear on the delineation of the lipocalin transcript. Black boxes mark borders between the coding exons where amino acids are encoded by both adjacent exons.
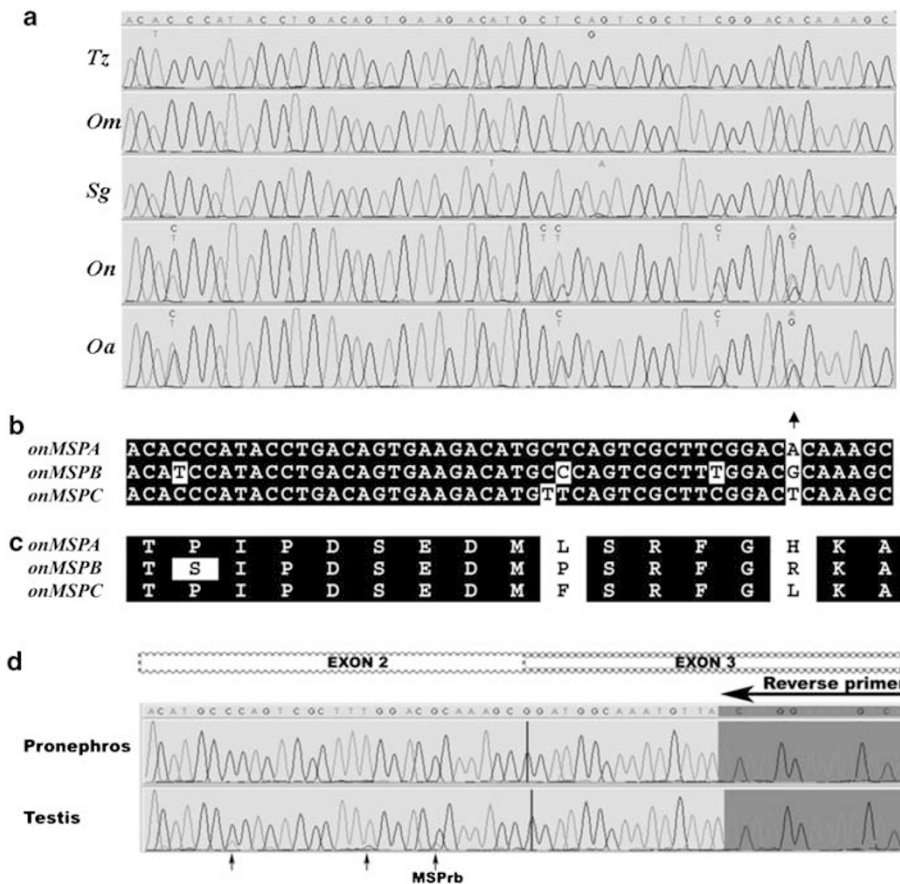
Figure 2 Sequencing of the second exon exemplifies the polymorphic nature of male-specific protein (*MSP*) genes. (**a**) PCR amplifications of 436 bp products from the MSP loci were performed using oligonucleotide primers (no. 5 and no. 2; Table 2) designed according to the sequence of the second and third exons of *S. galilaeus* (GenBank accession number, AAR19269), respectively. The genomic DNAs of five tilapia species were used as template for this amplification: *T. zillii* (Tz); *O. mossambicus* (Om); *S. galilaeus* (Sg); *O. niloticus* (On); *O. aureus* (Oa). Above the lanes the sequence of the last 50 bp of the second exon as recorded for *Om* is shown in one-letter color code (A, green; C, blue; G, black; T, red). Positions where the bases deviated from the *Om* sequence are indicated with the nucleotide code above the corresponding peaks. (**b**) Type-specific PCR was performed using DNA from the matriarch of the mapping family and the PCR products were sequenced. The sequence of the last 50 bp of the second exon that corresponds to the chromatogram above is shown. Identity between the nucleotide sequences is indicated by black boxes. White boxes indicate nucleotide changes between variants. Arrow points to the polymorphism used to discriminate the three *MSP* paralogs (MSPrb). (**c**) Amino-acid translation of the sequences displayed in (**b**). Box shading follows the legend of Figure 1. (**d**) PCR amplifications of 142 bp products from the *MSP* transcripts were performed using oligonucleotide primers (no. 1 and no. 2; Table 2). The presented chromatograms were obtained from the sequencing of reverse transcribed total RNA extracted from pronephros and testis of Ghana-Dor *O. niloticus* male. Above the lanes, similarly to what presented in (**a**), the sequence spanning the last 28 bp of the second exon and 30 bp of the third exon is shown in one-letter color code. Arrows below the chromatogram point to ambiguous bases including MSPrb where the simultaneous expression of *MSPA* and *MSPB* is detectable. Above the chromatogram boxes denote the exons and an arrow marks the sequence of the reverse primer (shaded).

Copy number variation of *MSP* genes among tilapia species

MSPrb nucleotides of A, G and T, which are markers for *MSPA*, *MSPB* and *MSPC*, respectively, were detected by sequencing in different tilapia species (Figure 2a; Table 1). In these samples we observed individuals with three MSPrb combinations: (1) singlet (A); (2) couplet (A and G) as in the grandfather and father of the mapping family, and (3) triplet (A, G and T) as in the grandmother and the mother of this family. Members of three different genera: *O. mossambicus*, *S. galilaeus* and *T. zillii* had only one copy of *MSP* (singlet, *MSPA* type). *O. aureus* and *O. niloticus* stocks and lines mostly had doublets except those that had triplets—*O. niloticus* stocks that originated from Ghana (Dor and BIU; Table 1) and *O. aureus* from the Mehadrin stock. All seven individuals with triplets were females and

none of the 18 males in the analyzed stocks had such a triplet (Table 1).

Copy number variation of *MSP* genes in Ghana *O. niloticus*

By type-specific PCR, we showed the existence of the three types of *MSP* paralogs in haploid gynogenetic embryos derived from a female of Ghana *O. niloticus*, which had all three base variants for MSPrb, like the matriarch of the mapping family (Figure 3). Three of the embryos carried all three types of paralogs, whereas the other three carried only *MSPA* and *MSPB*. This is in concordance with the model suggesting that the diploid matriarch carries two sets of chromosomes, which differ in *MSP* copy number due to the presence or absence of *MSPC* type of genes. Sequencing of the PCR products of the first and second haploid embryos (Figure 3) revealed
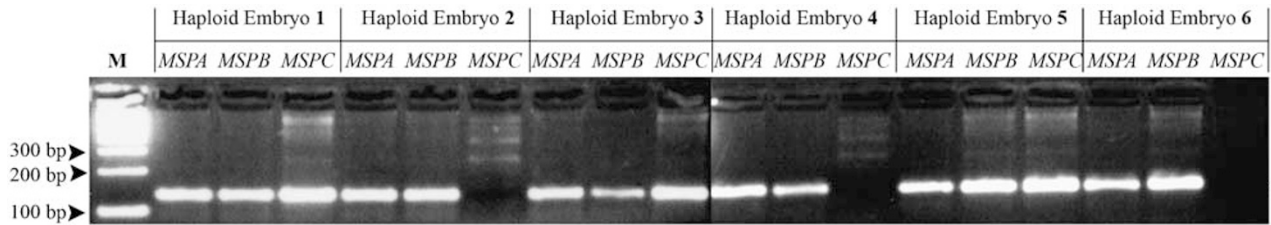
Figure 3 Type-specific PCR amplification of male-specific protein (*MSP*) genes from haploids. Ultraviolet (UV)-irradiated *T. zillii* sperm was used for the induction of gynogenetic haploids in eggs Ghana *O. niloticus* female carrier of the three *MSP* paralogs. Genomic DNA extracted from six embryos was used as a template for copy-specific amplification of 133 bp PCR products using primers specific for *MSPA*, *MSPB* and *MSPC*, as indicated above the lanes. Arrows point to the molecular size standard (M), the PCR products.

Table 5 Detection of *MSP* paralogs in haploid embryos by sequencing and *ShiftDetector* analysis

Type-specific PCR	Haploid embryo1 (MSPC carrier)		Haploid embryo2 (MSPC noncarrier)	
	No. of variants	Microsatellite sizes (bp)	No. of variants	Microsatellite sizes (bp)
MSPA	2	214, 214[a]	1	179
MSPB	2	177, 178	2	177, 178
MSPC	2	180, 181		

Abbreviation: MSP, male-specific protien.
[a]The presence of a second variant was indicated by variation in sequence outside the microsatellite.

that these *MSP* types included at least six paralogs as indicated in Table 5. Existence of at least two paralogous variants was evident from ambiguous nucleotides along the sequence or from ambiguous traces following the indel site due to the difference in length of the *MSP* microsatellite. Asymmetry between the *MSP* copy number on the maternal chromosome represented by haploid embryo 1 and paternal chromosome (haploid embryo 2) was suggested (Table 5). Three extra copies on the maternal chromosome were evident, characterized by the presence of the *MSPA* 214 bp variants ($MSPA_{214}$) in combination with *MSPC* variants.

Expression of *MSP* paralogs
To further assess if the *MSP* paralogs are functional genes, we examined the expression of *MSP* transcripts in three organs: liver, testis and pronephros using type-specific PCR in sexually active females and males (see details in 'Material and methods') with two and three *MSP* genomic copies, respectively. We sequenced the copy-specific PCR products and confirmed that they were derived from spliced transcripts corresponding to the expected exon structure (GenBank accession numbers, AM886144, AM886145, AM886146). *MSPA* was only detected in testis, whereas *MSPB* was observed in testis and in pronephros of both the genders (Figure 2D). *MSPC* was not detected in this experiment.

Genetic mapping of *MSPA* and *MSPC*
For linkage mapping of *MSPA1* we used the microsatellite variations in the first intron of this gene. Mapping of *MSPA2* gene was performed by defining the 214 bp allele as absent or present. Similarly, *MSPC* was mapped by defining two genotypes (T absent or present) according to the status of MSPrb detected by mass spectrometry. This analysis detected simultaneously the three possible MSPrb variants. All genotypes

were either couplet or triplets corresponding to the model of the patriarch and matriarch having *MSPA* and *MSPB* or *MSPA*, *MSPB* and *MSPC*, respectively. Genotypes were obtained for all 156 individuals of the F_2 mapping population. Among the 65 that inherited *MSPC*, 63 individuals also inherited $MSPA_{214}$, indicating tight linkage between these genes. The two groups of genes were mapped to two locations on LG12 with an interval of 2 cM (Figure 4).

Association between gender and variability in *MSP* genes in the mapping family
Females with high *MSP* copy number were more frequent by more than twofold than males. The numbers for the four possible combinations of $MSPA_{214}$-*MSPC* state with gender deviated significantly from the expected 1:1:1:1 Mendelian segregation ($P = 0.009$). The class of males with $MSPA_{214}$-*MSPC* was clearly deficient compared to all other combinations (Table 6).

Statistical comparison of models for SD and SSM
The best solution of expected segregations under the SD model was obtained with values of 69 and 11 for *m* and *n*, respectively, and yielded insignificant χ^2 value ($\chi^2 = 3.844$ and $P = 0.7207$). On the other hand, the best solution of expected segregations under the SSM model was obtained with values of 71 and 18 for *m* and *n*, respectively, was significantly more probable ($\chi^2 = 0.182$ and $P = 0.0192$). Hence, we suggest that the $MSPA_{214}$-*MSPC* genotype induced SSM that resulted in eliminating 18 of the presumable 35.5 (50.7%) males carrying this genotype.

Discussion

The alignment of the amino-acid sequence of *MSP* with orthologous proteins (Figure 1) indicated that it belongs to the lipocalin family of proteins. Among the
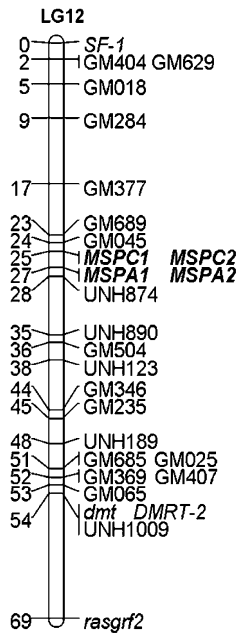
412

Figure 4 Genetic mapping of the male-specific protein A (*MSPA*) and the *MSPC* groups of genes. Genotype data for the markers in the *MSP* genes (*MSPA1*, *MSPA2*, *MSPC1* and *MSPC2*) and for the genetic markers of the mapping family were used as input for JoinMap software. The output for linkage group 12 (LG12) is presented in Centimorgans.

Table 6 Association between gender and high copy number of *MSP* genes in the mapping family

MSPA₂₁₄-MSPC	Females	Males
Present	43	20
Absent	47	40

Abbreviation: MSP, male-specific protien.

characterized lipocalin proteins of higher vertebrates it best resembled the highly glycosylated chicken ovoglycoprotein (ggOGCHI) and human orosomucoid 1 (ORM1). Indeed, high glycosylation level may explain the twofold incompatibility between calculated (20.4 kDa) and observed (41 kDa; Avtalion *et al.*, 1975) molecular masses of *MSP*. Human ORM1 was also referred to as α_1-acid glycoprotein that was implicated in mediating immune response by activated neutrophils (Theilgaard-Monch *et al.*, 2005). The cluster of the three orosomucoid genes on mouse chromosome 4 is tightly linked to the cluster of another group of five lipocalins named major urinary proteins (Mups). ORM1 and *Mup3* are also similar on the protein level (18% identity and 39% similarity). The sex-dependent expression, association with sexual activity and extraordinary high level of *MSP* in tilapia serum (Avtalion *et al.*, 1975, 1976) support the possibility that MSP is excreted in the urine or milt close to or during sexual intercourse. All the mentioned MSP properties highly resemble rat and mice Mups, whose expression is strongly stimulated up to 150-fold by androgens, principally by testosterone (Knopf *et al.*, 1983). Mups play a role in pheromone binding and

also are considered pheromones on their own (Flower, 1996).

Existence of genes orthologous to tilapia *MSP* genes in other species of fish is evident from similarity to poorly characterized genes in seabream (lm), tetraodon (tn) and zebrafish (dr) and to TBT-bps of flounder and killifish (Figure 1). Recent results in Japanese flounder showed that TBT-bp was also duplicated and that both paralogs were expressed in this species (Oba *et al.*, 2007). Tandem triplication of the *MSP*-like genes is present in tetraodon (GenBank accessions, CR693561, CR702458 and CR719634). TBT induction of masculinization in genetic females of flounder (Shimasaki *et al.*, 2003) further implies that the *MSP* genes, which encode TBT-bp-like proteins are associated with sexual functions and may influence SD.

Analysis of *MSP* sequence variation obtained by sequencing and genotyping of MSPrb and of the *MSP* microsatellite in the mapping family indicated that the four major types of gene combinations can be explained by a simple model of inheritance (Figure 5). Combining the genotyping data of *MSPA* microsatellite and copy status of *MSPC* with the available genotype data of our tilapia mapping family mapped *MSPA* and *MSPC* groups of genes to the same LG but 2 cM apart. Furthermore, the high frequency of the nonconservative amino-acid exchange (4/5 = 80%) relative to DNA sequence variations (Figure 2c) may indicate that they have diverted from their ancestral gene long enough to acquire different functions. This is also supported by the different patterns of *MSP* expression that were detected. Although *MSPA* was only detected in testis, *MSPB* was observed in the testis and in the pronephros of both genders and *MSPC* was undetectable.

MSP genes were mapped within a region of QTL for SSM that has not been previously described. Early studies relating SSM detected QTL on LG2, 6 and 23 in the different genetic background of an inbred line of *O. aureus* (Palti *et al.*, 2002; Shirak *et al.*, 2002). Previous study of the mapping family detected the observed effect; however, it was ignored as the study did not focus on the analysis of secondary QTL for SD (Lee *et al.*, 2005). Analysis of association between the presence and absence of additional *MSP* copies and gender revealed that males with high copy number were twice as rare as expected. Comparison of observed segregations with those of statistical models supported this observation. Only the SSM model provides good explanation for the reduction in males with high *MSP* copy number, as a result of their specific mortality.

We propose two possible explanations for the association of male-specific mortality with the *MSP* copy number: (1) the additional *MSP* copies encode transporters for sex steroids which can directly connect SD and maturation with viability; and (2) male-specific mortality may be caused by another gene closely linked to the *MSP* genes on LG12.

The first hypothesis is in line with the observation that gene dosage in the evolution and function of vertebrate sex was associated with master-key genes for SD (Ferguson-Smith, 2007; Quinn *et al.*, 2007; Volff *et al.*, 2007). The domain structure of the *MSP* genes suggests that they are carriers of sex steroids and thus influence the most fundamental pathway of SD. Tilapia SD is highly sensitive to exogenous steroids. In tilapia hybrids,
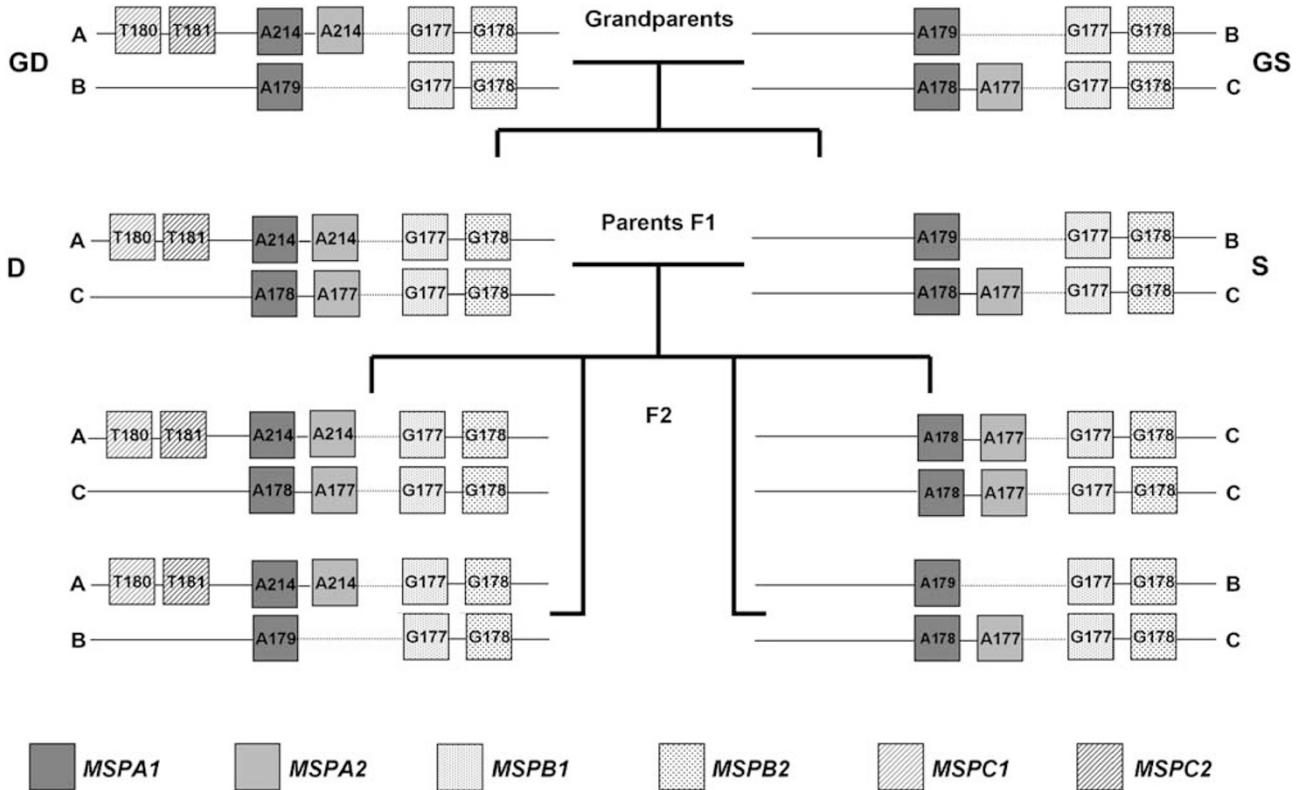
Figure 5 Inheritance of male-specific protein (MSP) genes within the mapping family. A simplified model for inheritance of the observed MSP-variant combinations in a mapping family derived from a cross between O. niloticus and O. aureus was based on genotypes obtained by sequencing, genotyping of MSPrb and of MSP microsatellite. Excluding three recombinants, four major types of F_2 gene combinations (AC, AB, BC, CC) were observed and denoted according to their origin: grand dam (GD, 29 individuals, AB), grand sire or sire (GS&S, 46 individuals, BC), dam (D, 34 individuals, AC) and a new type (41 individuals, CC). Each combination is described by delineation of the two homologous chromosomes on which the haplotypes consisting of MSPrb followed by the MSP microsatellite status are denoted. Shaded boxes present gene identities. Solid lines indicate mapped proximity and dashed ones suggest proximity yet to be detected. Bold solid lines show the gene flow between the generations.

it may have happened that driving MSP genes out of copy balance produced an effect of master-key regulator of SD for these genes. For example, at the second and third weeks after hatching in the hormone sensitive period for SD, transport of estrogens could have increased in individuals with high MSP copy number and thus skewing SD toward females. Appearance of new QTL for SD due to hybridization was predicted by an autosomal theory based on analysis of sex ratios in tilapia hybrids (Avtalion and Hammerman, 1978; Hammerman and Avtalion, 1979). In the present work we demonstrate the appearance of a new QTL for SSM due to hybridization, possibly due to MSP copy imbalance. Mortality resulting in reduced egg, embryo and larval survival also could have been brought about through TBT-bp, MSP homologue of medaka treated by TBT (Nirmala et al., 1999).

In support for the second hypothesis it should be noted that other strong candidate genes are located near the MSP locus, and that the six upper markers on LG12 (Figure 5) showed significant association with gender (data not shown). Most lipocalins are clustered on the same chromosome in several genomes, for example human chromosome 9 and mouse chromosome 2 (Chan et al., 1994; Suzuki et al., 2004). In humans this cluster consists of 16 characterized genes including ORM1 and ORM2 (9q32), odorant-binding proteins (9q34), progesta-gen-associated endometrial protein (9q34) and human prostaglandin D2 synthase (9q34.2–q34.4). The latter was

proposed to be a key regulator of SD in mammals regulating the import and export of SOX9 transcription factor from the nucleus, which are critical steps in the cascade of testis determination initiated by SRY (Malki et al., 2005). Another human gene NR5A1 (9q33) involved in the binding of steroid hormones, which is orthologous to the tilapia steroidogenic factor (SF1) implicated in primary SD (Shirak et al., 2006a, b), does not belong to the lipocalin family but was mapped close to this cluster both in human and tilapia.

One MSP copy of type MSPA was detected in species of three different genera—O. mossambicus, S. galilaeus and T. zillii, and two MSP copies were detected in most O. niloticus and O. aureus stocks from different origins. These results indicate that MSP duplication has probably occurred before the evolutionary divergence of these species. Three MSP copies were detected in females of O. aureus from Mehadrin stock and two stocks of O. niloticus both originated from Ghana. In a previous study we detected a high number (36%) of O. niloticus alleles in the O. aureus Mehadrin stock, and suggested the possibility that this stock is not purebred (Shirak et al., 2006a). We hypothesize that MSP was primarily triplicated in the Ghana O. niloticus and later was introgressed into the O. aureus Mehadrin stock. Therefore, to identify the native role of all three MSP types of paralogs they should be studied in the Ghana O. niloticus strain.

Acknowledgements

References

Albani JR (2006). Progesterone binding to the tryptophan residues of human alpha1-acid glycoprotein. *Carbohydr Res* **341**: 2557–2564.

Avtalion RR (1982). Genetic markers in *Sarotherodon* and their use for sex and species identification. In: Pullin RSV, McConnel RHL (eds). *The Biology and Culture of Tilapias*. ICLARM: Manila. pp 269–277.

Avtalion RR, Duczyminer M, Wojdani A, Pruginin Y (1976). Determination of allogeneic and xenogeneic markers in genus of tilapia—identification of *T aurea*, *T vulcani* and *T nilotica* by electrophoretic analysis of their serum-proteins. *Aquaculture* **7**: 255–265.

Avtalion RR, Hammerman IS (1978). Sex determination in *Sarotherodon* (tilapia).1. Introduction to a theory of autosomal influence. *Bamidgeh* **30**: 110–115.

Avtalion RR, Pruginin Y, Rothbard S (1975). Determination of allogeneic and xenogeneic markers in the genus of Tilapia. I. Identification of sex and hybrids in Tilapia by electrophoretic analysis of serum proteins. *Bamidgeh, Bull Fish Cult Israel* **27**: 8–13.

Baroiller JF, D'Cotta H (2001). Environment and sex determination in farmed fish. *Comp Biochem Phys C* **130**: 399–409.

Chan P, Simonchazottes D, Mattei MG, Guenet JL, Salier JP (1994). Comparative mapping of lipocalin genes in human and mouse—the 4 Genes for complement C8-Gamma-chain, prostaglandin-D-synthase, oncogene-24P3, and progestagen-associated endometrial protein Map to HSA9 and MMU2. *Genomics* **23**: 145–150.

Cnaani A, Hallerman EM, Ron M, Weller JI, Indelman M, Kashi Y *et al.* (2003). Detection of a chromosomal region with two quantitative trait loci, affecting cold tolerance and fish size, in an F-2 tilapia hybrid. *Aquaculture* **223**: 117–128.

Cnaani A, Zilberman N, Tinman S, Hulata G, Ron M (2004). Genome-scan analysis for quantitative trait loci in an F-2 tilapia hybrid. *Mol Genet Genomics* **272**: 162–172.

Desprez D, Briand C, Hoareau MC, Melard C, Bosc P, Baroiller JF (2006). Study of sex ratio in progeny of a complex *Oreochromis* hybrid, the Florida red tilapia. *Aquaculture* **251**: 231–237.

Ferguson-Smith M (2007). The evolution of sex chromosomes and sex determination in vertebrates and the key role of DMRT1. *Sex Dev* **1**: 2–11.

Flower DR (1996). The lipocalin protein family: Structure and function. *Biochem J* **318**: 1–14.

Gates MA, Kim L, Egan ES, Cardozo T, Sirotkin HI, Dougan ST *et al.* (1999). A genetic linkage map for zebrafish: Comparative analysis and localization of genes and expressed sequences. *Genome Res* **9**: 334–347.

Hammerman IS, Avtalion RR (1979). Sex determination in *Sarotherodon* (tilapia). 2. Sex-ratio as a tool for the determination of genotype—model of autosomal and gonosomal Influence. *Theor Appl Genet* **55**: 177–187.

Hano T, Oshima Y, Kim SG, Satone H, Oba Y, Kitano T *et al.* (2007). Tributyltin causes abnormal development in embryos of medaka, *Oryzias latipes*. *Chemosphere* **69**: 927–933.

Jurinke C, van den Boom D, Cantor CR, Koster H (2001). Automated genotyping using the DNA MassArray technology. *Methods Mol Biol* **170**: 103–116.

Kirsh N (1991). Quantitative study of the tilapia male-specific protein (MSP) in normal, sex inversed and gynogenetic tilapias. MSc thesis, Bar-Ilan University, Ramat-Gan, Israel.

Knopf JL, Gallagher JF, Held WA (1983). Differential, multihormonal regulation of the mouse major urinary protein gene family in the liver. *Mol Cell Biol* **3**: 2232–2240.

Lee BY, Hulata G, Kocher TD (2004). Two unlinked loci controlling the sex of blue tilapia (*Oreochromis aureus*). *Heredity* **92**: 543–549.

Lee BY, Lee WJ, Streelman JT, Carleton KL, Howe AE, Hulata G *et al.* (2005). A second-generation genetic linkage map of tilapia (*Oreochromis* spp.). *Genetics* **170**: 237–244.

Lee BY, 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.

Ma RZ, Beever JE, Da Y, Green CA, Russ I, Park C *et al.* (1996). A male linkage map of the cattle (*Bos taurus*) genome. *J Hered* **87**: 261–271.

Majumdar KC, McAndrew BJ (1986). Relative DNA content of somatic nuclei and chromosomal studies in 3 genera, *Tilapia*, *Sarotherodon*, and *Oreochromis* of the tribe Tilapiini (Pisces, Cichlidae). *Genetica* **68**: 175–188.

Malki S, Nef S, Notarnicola C, Thevenet L, Gasca P, Mejean C *et al.* (2005). Prostaglandin D2 induces nuclear import of the sex-determining factor SOX9 via its cAMP-PKA phosphorylation. *EMBO J* **24**: 1798–1809.

McAndrew BJ, Majumdar KC (1983). Tilapia stock identification using electrophoretic markers. *Aquaculture* **30**: 249–261.

Nakayama K, Oshima Y, Yamaguchi T, Tsuruda Y, Kang IJ, Kobayashi M *et al.* (2004). Fertilization success and sexual behavior in male medaka, *Oryzias latipes*, exposed to tributyltin. *Chemosphere* **55**: 1331–1337.

Nirmala K, Oshima Y, Lee R, Imada N, Honjo T, Kobayashi K (1999). Transgenerational toxicity of tributyltin and its combined effects with polychlorinated biphenyls on reproductive processes in Japanese medaka (*Oryzias latipes*). *Environ Toxicol Chem* **18**: 717–721.

Oba Y, Shimasaki Y, Oshima Y, Satone H, Kitano T, Nakao M *et al.* (2007). Purification and characterization of tributyltin-binding protein type 2 from plasma of Japanese flounder, *Paralichthys olivaceus*. *J Biochem (Tokyo)* **142**: 229–238.

Palti Y, Shirak A, Cnaani A, Hulata G, Avtalion RR, Ron M (2002). Detection of genes with deleterious alleles in an inbred line of tilapia (*Oreochromis aureus*). *Aquaculture* **206**: 151–164.

Quinn AE, Georges A, Sarre SD, Guarino F, Ezaz T, Graves JAM (2007). Temperature sex reversal implies sex gene dosage in a reptile. *Science* **316**: 411.

Seroussi E, Ron M, Kedra D (2002). ShiftDetector: detection of shift mutations. *Bioinformatics* **18**: 1137–1138.

Shimasaki Y, Kitano T, Oshima Y, Inoue S, Imada N, Honjo T (2003). Tributyltin causes masculinization in fish. *Environ Toxicol Chem* **22**: 141–144.

Shimasaki Y, Oshima Y, Yokota Y, Kitano T, Nakao M, Kawabata S *et al.* (2002). Purification and identification of a tributyltin-binding protein from serum of Japanese flounder, *Paralichthys olivaceus*. *Environ Toxicol Chem* **21**: 1229–1235.

Shirak A, Avtalion RR (2001). Full-sib mating can reduce deleterious effects associated with residual sperm inheritance in gynogenotes. *Isr J Aquacult Bamidgeh* **53**: 15–22.

Shirak A, Palti Y, Cnaani A, Korol A, Hulata G, Ron M *et al.* (2002). Association between loci with deleterious alleles and distorted sex ratios in an inbred line of tilapia (*Oreochromis aureus*). *J Hered* **93**: 270–276.

Shirak A, Seroussi E, Cnaani A, Howe AE, Domokhovsky R, Zilberman N *et al.* (2006a). *Amh* and *Dmrta2* genes map to tilapia (*Oreochromis* spp.) linkage group 23 within quantitative trait locus regions for sex determination. *Genetics* **174**: 1573–1581.

Shirak A, Seroussi E, Zilberman N, Hulata G, Ron M, Cnaani A *et al.* (2006b). Genetic basis of sex determination in fishes: Searching for master key regulator genes in the sex determination pathway of tilapias. *Isr J Aquacult Bamidgeh* **58**: 350.

Staden R, Beal KF, Bonfield JK (2000). The Staden package, 1998. *Methods Mol Biol* **132**: 115–130.

Suzuki K, Lareyre JJ, Sanchez D, Gutierrez G, Araki Y, Matusik RJ *et al.* (2004). Molecular evolution of epididymal lipocalin genes localized on mouse chromosome 2. *Gene* **339**: 49–59.

Theilgaard-Monch K, Jacobsen LC, Rasmussen T, Niemann CU, Udby L, Borup R *et al.* (2005). Highly glycosylated alpha 1-acid glycoprotein is synthesized in myelocytes, stored in secondary granules, and released by activated neutrophils. *J Leukocyte Biol* **78**: 462–470.

Trewavas E (1983). *Tilapiine Fishes of the Genera Sarotherodon, Oreochromis and Danakilia*. British Mus. Nat. Hist.: London, UK, 583 p.

Trombka D, Avtalion R (1993). Sex determination in tilapia—a review. *Isr J Aquacult Bamidgeh* **45**: 26–37.

Tseng CL (1998). A Newton-type univariate optimization algorithm for locating the nearest extremum. *Eur J Oper Res* **105**: 236–246.

Volff JN, Nanda I, Schmid M, Schartl M (2007). Governing sex determination in fish:regulatory putsches and ephemeral dictators. *Sex Dev* **1**: 85–99.

Yamada Y, Nakagawa K, Yajima T, Saito K, Tokushima A, Fujiwara K *et al.* (2006). Structural and thermodynamic consequences of removal of a conserved disulfide bond from equine beta-lactoglobulin. *Proteins* **63**: 595–602.