iScience, Volume *24*

Supplemental Information

fshr: a fish sex-determining locus shows variable incomplete penetrance across flathead grey mullet populations

Serena Ferraresso, Luca Bargelloni, Massimiliano Babbucci, Rita Cannas, Maria Cristina Follesa, Laura Carugati, Riccardo Melis, Angelo Cau, Manos Koutrakis, Argyrios Sapounidis, Donatella Crosetti, and Tomaso Patarnello

Transparent methods

Ethical statement

Mugil cephalus samples were received from commercial catches in the Mediterranean sea. No fish were handled while alive for the purpose of this project. All fish were dead when they were selected for the study. Thus, the research did not involve animal experimentation or harm, and required no ethical permits.

Samples origin, sex phenotyping, and DNA extraction

Adult *M. cephalus* female and male individuals employed in this study belonged to four different Mediterranean origins: Cabras (**CAB**) in West Sardinia, Tortolì (**TOR**) in East Sardinia, Orbetello Lagoon (**ORB**) in Tuscany and Bay of Kavala (**KAV**) in East Macedonia (Figure 1).

The sex was recorded by visual inspection of the gonads or by light microscope when macroscopic observation was ambiguous. As mentioned before, sexually mature individuals were opportunistically sampled at processing plants, where fish roe is collected. Mature individuals were first assessed by compressing the body to observe whether the animal emitted sperm. Secondly, all animals were dissected and mature gonads were isolated. In mature individuals recognition of males (emitting sperm and showing characteristic gonad morphology) and females (non emitting sperm and with clearly identifiable gonads) is never ambiguous (see Figure 4). This method for sex phenotyping was used for almost all samples (Table S1). Only for three putative males with ambiguous identification (non emitting sperm, not fully differentiated gonads), it was necessary to using standard histology to confirm putative sex. An example of the microscope picture of male gonads is shown in Figure S1.

The reliable identification of phenotypic sex was possible for a total of 330 individuals; 109 CAB (57 males and 52 females), 92 TOR (32 males and 60 females), 67 ORB (34 males and 33 females) and 62 KAV (31 males and 31 females).

Fin clips were employed for DNA extraction using Invisorb® DNA Tissue HTS 96 Kit (Invisorb, Germany) following the manufacturer's instructions. DNA concentration was determined using a Qubit fluorimeter with a dsDNA BR Assay (Invitrogen, USA) and DNA quality was assessed by loading a 100ng sample onto a 1 % agarose gel electrophoresis.

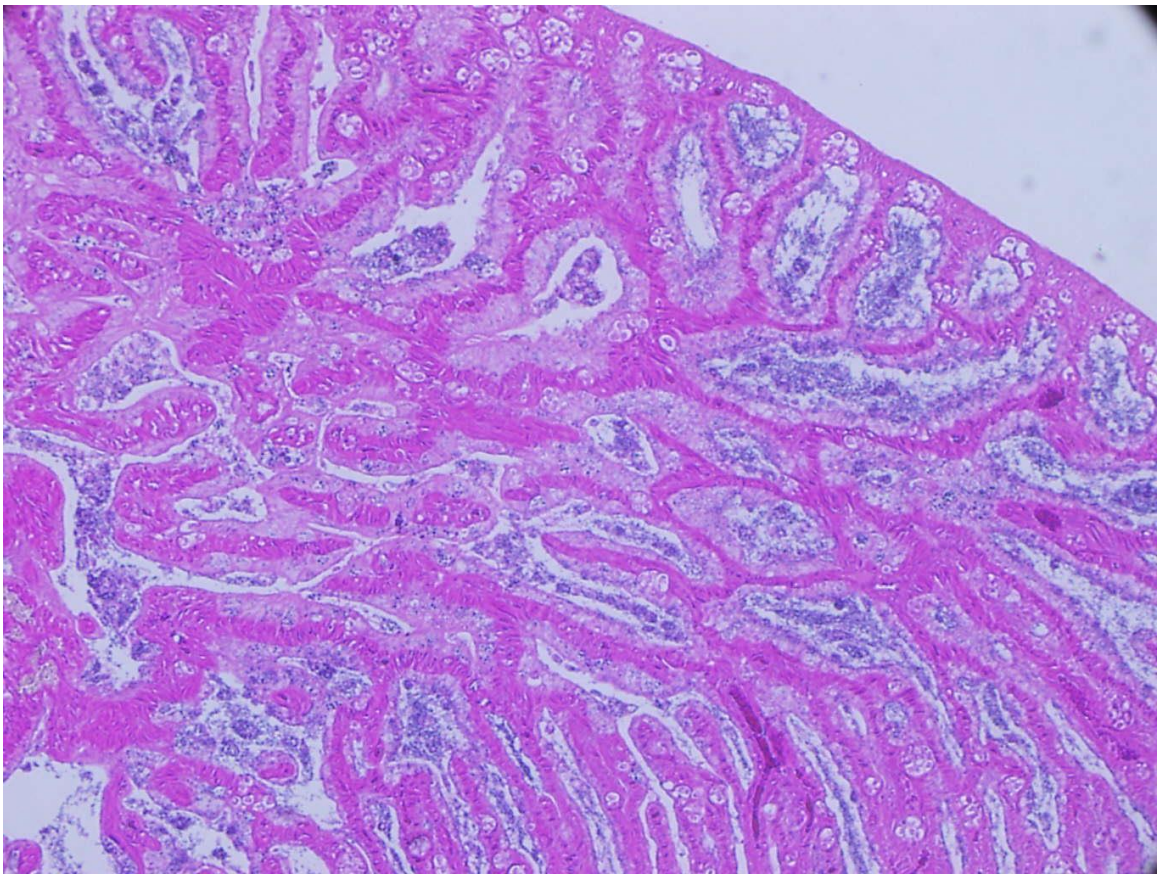

Figure S1. Microscope picture of male gonads. Related to Figure 4. Hematoxilin-eosin preparation of male gonad.

Sequencing and assembly of *M. cephalus* draft genome

In order to construct a draft assembly of *M. cephalus* genome, the DNA of a single TOR female was used for whole-genome sequencing. A total of three the standard protocol of the TruSeq DNA sample preparation kit (Illumina, CA, USA) and sequenced on an Illumina HiSeq4000 instrument following a 150 paired-end (PE) strategy. The total amount of reads obtained after sequencing was 709 millions corresponding to an estimated coverage of around 120X.

Quality of Illumina raw reads was analyzed with the FastQC program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Subsequently, low-quality regions and adapters were trimmed using Trimmomatic 0.361. Global assembly of the Illumina reads was accomplished with the software CLC Genomics Workbench v10 (https://www.qiagenbioinformatics.com/) using a minimum contig size of 500 bp and default settings for the other parameters. Quantitative assessment of the genome assembly was assessed with BUSCO v.3 software (Simão et al., 2015). The Actinopterygii dataset, containing 4,584 well-conserved genes, was employed to investigate the completeness of the assembly.

PoolSeq

For both sexes, 60 DNA samples (belonging to 30 TOR and 30 CAB individuals) were equimolarly pooled for sequencing. To minimize the contribution of pipetting errors and PCR biases, three independent technical replicates of pooling and library construction were carried out for both sex. A total of six genomic libraries were thus constructed and sequenced as described above. Details on obtained reads before and after quality trimming are reported on Table S2. The total amount of reads obtained after sequencing were 185 and 181.5 millions corresponding to an estimated coverage of 63X and 62X for female (MuCe_F) and male (MuCe_M) pool, respectively. Raw Illumina reads were deposited in the SRA repository under the accession numbers PRJNA657721 and BioSample SAMN15835396 to SAMN15835401.

Pool-Seq mapping and variant calling

PoolSeq sequence data were trimmed to remove adaptors and low quality regions by means of Trimmomatic 0.361. Filtered sequences were then mapped to the reference *M. cephalus* genome with the software CLC Genomics Workbench v10 by setting the following parameters: match score=1, mismatch cost=2, insertion cost=3, deletion cost=3, length fraction=0.8, similarity fraction=0.8. PICARD tools v2.6 were then employed for alignment sorting and duplicate removing, and the resultant bam files were further filtered to remove sequences with mapping quality lower than 20 and unpaired reads by means of SAMTOOLS 1.9

PoPoolation2 v1.201 (Kofler et al., 2011) was then employed for SNP calling, FST calculation and Fisher's exact test to assess SNP differentiation between MuCe_M and MuCe_F pools. Regions surrounding indels (--indel-window 5) were excluded from the analysis and only SNPs with min coverage above 20X (max coverage 200X) were retained.

Identification of sex-patterned SNPs

Results obtained from Popoolation2 were further analyzed in order to identify sex-patterned SNPs, sites that are fixed or nearly fixed in the homogametic sex and in a frequency between 0.4 and 0.6 in the heterogametic sex. To reach such a goal, the Sex_SNP_finder_now.pl script developed by Gammerdinger et al (2016, https://github.com/Gammerdinger/sex-SNP-finder) was employed by setting --fixed_threshold=0.95, --minimum_polymorphic_freq=0.4, --maximum_polymorphic_freq=0.6 and --read_depth=30.

Sequence similarity analysis to detect duplicated *fshr* copies

Extensive sequence similarity analysis were carried out against the two flathead grey mullet genome assemblies to identify putative duplicated *fshr* copies. TBlastN using as query the Nile tilapia protein

fshr sequence as well as *M.cephalus fshr* were run against both genome assemblies. The nucleotide sequence of *M. cephalus fshr* was also used as a query with BlastN.

Sanger sequencing of *fshr*

A primer pair spanning a region of 286 nt that encompasses MuCe179, MuCe266 and MuCe322 variants on contig_111122 was designed using Primer3 software (Primer Forward: TGCTCCTCCTCAACATCCTG; Primer Reverse: AAGAAGGCGTACAGGAAGGG). Polymerase Chain Reaction (PCR) was then performed on DNAs extracted from ORB, TOR, CAB and KAV populations. For each population, 30 to 33 individuals for both sexes were investigated.

Cycling conditions were: initial incubation at 95° C for 2 min followed by 37 cycles at 95°C for 30 s, 60°C for 30 s and 72°C for 1 min. A final extension step at 72°C for 5 min was added at the end of the last cycle. PCR products were purified with ExoSAP-IT™ (Thermofisher, USA) followed by Sanger sequencing. All obtained sequences were analysed with Chromas Lite 2.0 software.

Statistical analysis

Fisher Exact tests and chi-square tests were performed using an online calculator (https://www.socscistatistics.com/tests/).

Sanger sequencing of nuclear receptor nr2f5 gene

PCR amplification of a 290 nt fragment of the nuclear receptor nr2f5 gene was carried out for 15 individual female fish and 25 males to assess the presence of a single nucleotide variant. The observed genotype was as follows: 14 females were homozyogous for the wt allele, 1 was wt/m; 10 males were heterozygous (wt/m), 15 homozygous (wt/wt).

Supplemental references

Gammerdinger WJ, Conte MA, Baroiller JF, D'Cotta H, Kocher TD. Comparative analysis of a sex chromosome from the blackchin tilapia, Sarotherodon melanotheron. BMC Genomics. 2016; 7(1):808.

Kofler R, Pandey RV, Schlötterer C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics. 2011; 27(24):3435-6.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015 ;31(19):3210-2.

Supplemental Data S1. Nucleotide sequence of contig_111122. Related to Figure 3. MuCe179 (red), MuCe206 (green) and MuCe322 (yellow) variants are highlighted in male nucleotide sequence.

>MuCe204_contig_111122_FEMALE

GTTCGTGCAGGTGAGCATCTGCTTGCCCATGGATGTGGAGGATCTGGTGTCCCAGGTCTA
TGTAGTGTCCCTGCTCCTCCTCAACATCCTGGCCTTCCTCTGTGTGTGCGGCTGCTACCT
CAGCATCTACCTGACCTTCCGCAATCCCTCTTCGGTGCCGGCCCACGCCGACACGCGC**G**T
GGCTCAACGCATGGCCGTCCTCATC**T**TCACCGACTTCATCTGCATGGCCCCGATCTCCTT
CTTCGCCGTCTCGGCCGCGCTCAAGACCCCCCTCATCACCGTCTCGGAATCTAAGGTCCT
CCTGGTCCTGTTCTACCCCAT**C**AACTCGTGCGCCAACCCCTTCCTGTACGCCTTCTTCAC
CCGCACCTTCCGGCGGGACTTCTTTTTCCTGGCGGCTCGCTTCGGCCTGTTTAAGACTCG
GGCGCAGATTTACCGGACAGAGACCTCTTCCTGTCAGCAGCCAGCATGGACCTCTTCGAG
GAGCAGCCGCGTGACAATGTACTCCTTGGCCAACACCTTGAGCCTGGACGCGTGCGTAGA
CTCCCAGTCGTCCAAGTCATGGTAGACCAAA

>MuCe204_contig_111122_MALE

GTTCGTGCAGGTGAGCATCTGCTTGCCCATGGATGTGGAGGATCTGGTGTCCCAGGTCTA
TGTAGTGTCCCTGCTCCTCCTCAACATCCTGGCCTTCCTCTGTGTGTGCGGCTGCTACCT
CAGCATCTACCTGACCTTCCGCAATCCCTCTTCGGTGCCGGCCCACGCCGACACGCGC**A**T
GGCTCAACGCATGGCCGTCCTCATC**G**TCACCGACTTCATCTGCATGGCCCCGATCTCCTT
CTTCGCCGTCTCGGCCGCGCTCAAGACCCCCCTCATCACCGTCTCGGAATCTAAGGTCCT
CCTGGTCCTGTTCTACCCCAT**T**AACTCGTGCGCCAACCCCTTCCTGTACGCCTTCTTCAC
CCGCACCTTCCGGCGGGACTTCTTTTTCCTGGCGGCTCGCTTCGGCCTGTTTAAGACTCG
GGCGCAGATTTACCGGACAGAGACCTCTTCCTGTCAGCAGCCAGCATGGACCTCTTCGAG
GAGCAGCCGCGTGACAATGTACTCCTTGGCCAACACCTTGAGCCTGGACGCGTGCGTAGA
CTCCCAGTCGTCCAAGTCATGGTAGACCAAA

Protein Alignment

```
                    10        20        30        40        50        60
                    |         |         |         |         |         |
Female        FVQVSICLPMDVEDLVSQVYVVSLLLLNILAFLCVCGCYLSIYLTFRNPSSVPAHADTRV
Male          FVQVSICLPMDVEDLVSQVYVVSLLLLNILAFLCVCGCYLSIYLTFRNPSSVPAHADTRM
              ************************************************************:
                    70        80        90       100       110       120
                    |         |         |         |         |         |
Female        AQRMAVLIFTDFICMAPISFFAVSAALKTPLITVSESKVLLVLFYPINSCANPFLYAFFT
Male          AQRMAVLIVTDFICMAPISFFAVSAALKTPLITVSESKVLLVLFYPINSCANPFLYAFFT
              ********.***************************************************
                   130       140       150       160       170       180
                    |         |         |         |         |         |
Female        RTFRRDFFFLAARFGLFKTRAQIYRTETSSCQQPAWTSSRSSRVTMYSLANTLSLDACVD
Male          RTFRRDFFFLAARFGLFKTRAQIYRTETSSCQQPAWTSSRSSRVTMYSLANTLSLDACVD
              ************************************************************
```

MuCe179
MuCe206
MuCe322