

MOLECULAR ECOLOGY RESOURCES

Supplemental Information for:

Development of epigenetic biomarkers for the identification of sex and thermal stress in fish using DNA methylation analysis and machine learning procedures

Valdivieso A¹, Anastasiadi, D^{1,2}, Ribas L.1, Piferrer F.^{1*}

1 Institut de Ciències del Mar, Consejo Superior de Investigaciones Científicas (CSIC), Barcelona, Spain

2 Current address: The New Zealand Institute for Plant and Food Research Limited, Nelson Research Centre, 293–297 Akersten Street, Port Nelson, Nelson 7010, New Zealand

* Correspondence: Dr Francesc Piferrer: piferrer@icm.csic.es

Table of Contents:

Table S1	Page 2
Table S2	Page 3-6
Table S3	Page 7
Table S4	Page 7
Table S5	Page 8-9
Table S6	Page 10
Figure S1	Page 11
Figure S2	Page 12
Figure S3	Page 13
Figure S4	Page 14

MOLECULAR ECOLOGY RESOURCES

Supplementary tables

Table S1. Information of samples used in this study at 90 days post-fertilization. Number of fish per temperature and sex and the number of samples for DNA methylation analysis

Family	Temperature (°C)	Phenotypic sex	Number fish	Number samples for analysis
1	28	Male	39	11
		Female	29	12
	35	Male	77	12
		Female	0	0
2	28	Male	17	11
		Female	22	10
	35	Male	32	5
		Female	18	13
Total			234	74

MOLECULAR ECOLOGY RESOURCES

Table S2. Information of the ten genes selected for this study. Gene name, symbol, detailed sequences for the bisulfite primers and information related to the amplicon: positions in the genome, number of CpG sites, CpG island and annealing temperature

Gene name	anti-Müllerian hormone	double-sex and mab-3 related transcription factor 1	double-sex and mab-3 related transcription factor 3a
Gene symbol	<i>amh</i>	<i>dmrt1</i>	<i>dmrt3a</i>
Reference sequence	NM_001007779	NM_205628	NM_001005779
Chromosome	22	5	5
Strand	-	+	+
Forward primer (5' to 3')	TTTTATTTTATTAATAAATTAAGGTATGTGA	AATATTTTTTATTGTGGTGTGGTT	TTTTGTAATTTTTATGTTGTAATAAGTTA
Reverse primer (5' to 3')	AACACTCCTCCAAAATATAAAAAC	CAAAATAACCTCTCATAACCTTATC	CACACTTAAAAATCCTCTAAAAAAAAC
Start coordinate amplicon	20736331	44344542	44400282
End coordinate amplicon	20736847	44345024	44400732
Amplicon size bp and relative position from TSS	517 (-217, +300)	483 (-249, +234)	451 (-366, +85)
Total number of CpG	15	24	18
Number of CpG before TSS	7	10	12
Number of CpG islands	1	2	2
Temperature annealing (°C)	51	53	53

MOLECULAR ECOLOGY RESOURCES

Table S2 (Continued)

Gene name	cytochrome P450, family 19, subfamily A, polypeptide 1a	hydroxysteroid (11- β) dehydrogenase 2	forkhead box L2a
Gene symbol	<i>cyp19a1a</i>	<i>hsd11b2</i>	<i>foxl2a</i>
Reference sequence	NM_131154	NM_212720	NM_001045252
Chromosome	18	7	15
Strand	-	-	+
Forward primer (5' to 3')	TATTTTTGTTTGTAGGTTTGAT	GTAAAGTTTGGTTATTGGGTTTG	TTTATAAATAGGTTTAATTTATAAAAA
Reverse primer (5' to 3')	CAACAATATTATAACTATAACTCCACAC	CAACCATACCACCACTAAAAACATA	ACTATTCTACCATCCCTTCTTATTCTTCT
Start coordinate amplicon	39655109	34854510	7060278
End coordinate amplicon	39655578	34854994	7060715
Amplicon size bp and relative position from TSS	470 (-326, +144)	485 (-362, +123)	438 (-159, +279)
Total number of CpG	15	9	26
Number of CpG before TSS	9	5	5
Number of CpG islands	1	0	2
Temperature annealing ($^{\circ}$ C)	53	59	50

MOLECULAR ECOLOGY RESOURCES

Table S2 (Continued)

Gene name	cytochrome P450, family 11, subfamily A, polypeptide 1	cytochrome P450, family 11, subfamily C, polypeptide 1	hydroxysteroid (17- β) dehydrogenase 1
Gene symbol	<i>cyp11a1</i>	<i>cyp11c1</i>	<i>hsd17b1</i>
Reference sequence	NM_152953	NM_001080204	NM_205584
Chromosome	25	16	3
Strand	+	-	-
Forward primer (5' to 3')	GGGGTTAAATTTGTTATTGATATATTTTG	TTTTTTTTGATTTGTGTTTTTTTATAGTAA	ATTGTGATTAATAATTTATAAGATGTATGA
Reverse primer (5' to 3')	ATAAACATTCCTAAAACCTCCCAT	AACTCCTCCATATATTTATACATAC	CAATAATCATACATTAACACAAAATAATA
Start coordinate amplicon	22224239	54645580	17897321
End coordinate amplicon	22224672	54645995	17897746
Amplicon size bp and relative position from TSS	434 (-230, +204)	416 (-201, +215)	426 (-394, +30)
Total number of CpG	8	19	6
Number of CpG before TSS	0	2	6
Number of CpG islands	0	1	0
Temperature annealing ($^{\circ}$ C)	53	57	53

MOLECULAR ECOLOGY RESOURCES

Table S2 (Continued)

Gene name	hydroxysteroid (17- β) dehydrogenase 3
Gene symbol	<i>hsd17b3</i>
Reference sequence	NM_200364
Chromosome	8
Strand	+
Forward primer (5' to 3')	TTTTGGTGGATTTATATAATAATAGG
Reverse primer (5' to 3')	ACTCTATCAAACCTCTCTTAATTTCTACTA
Start coordinate amplicon	29856058
End coordinate amplicon	29856474
Amplicon size bp and relative position from TSS	417 (-370, +47)
Total number of CpG	4
Number of CpG before TSS	4
Number of CpG islands	0
Temperature annealing (°C)	53

MOLECULAR ECOLOGY RESOURCES

Table S3 Nextera Adapters sequences and the primers sequences with complementary to Nextera Adapters use to validate the amplification of region of interest by Sanger sequencing

Nextera Adapters sequences	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG+[forward primer]	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG+[reverse primer]
Primer for Sanger sequencing	CGGCAGCGTCAGATGTGTAT	GTGGGCTCGGAGATGTGTAT

Table S4. Primer sequences used for qPCR gene expression analysis

Gene	Gene name	Gene symbol	Reference sequence	Forward (5' to 3')	Reverse (5' to 3')	Reference
Target genes	anti-Müllerian hormone	<i>amh</i>	NM_001007779	ACAACCCGAAGGTCAACCCGC	GTGGCATGTTGGTCAGTTGGCTG	[1]
	cytochrome P450, family 11, subfamily A, polypeptide 1	<i>cyp11a1</i>	NM_152953	GAGGCCTCAGGAATGTCCAC	GGTCCACGCGTCTACATTGA	[2]
	cytochrome P450, family 11, subfamily C, polypeptide 1	<i>cyp11c1</i>	NM_001080204	CCTCGGGCCCATATACAGAGA	CGTCCCCTTCTTGAGGAAGA	[1]
	cytochrome P450, family 19, subfamily A, polypeptide 1a	<i>cyp19a1a</i>	NM_131154	GATATTTGCTCAGAGCCATGGA	GCTCTGGCCAGCTAAAACACT	[1]
	doublesex and mab-3 related transcription factor 1	<i>dmrt1</i>	NM_205628	TGCCAGGTGGCGTTACGG	CGGGTGATGGCGTCTGAG	[1]
	hydroxysteroid (11- β) dehydrogenase 2	<i>hsd11b2</i>	NM_212720	GGGGGTCAAAGTTTCCACTA	TGGAAGAGCTCCTTGGTCTC	[3]
	hydroxysteroid (17- β) dehydrogenase 1	<i>hsd17b1</i>	NM_205584	GTCTGATGGGTCCTCTGGAA	TGCCGTGTCTTTTTCTTCA	[4]
Reference genes	eukaryotic translation elongation factor 1 α 1 like 1	<i>eef1a1l1</i>	NM_131263	GGAGGCTGCCAACTTCAACGC	GGCGATGTGAGCAGTGTGGC	[5]
	ribosomal protein L13a	<i>rpl13a</i>	NM_212784	TCTGGAGGACTGTAAGAGGTATGC	AGACGCACAATCTTGAGAGCAG	[5]

MOLECULAR ECOLOGY RESOURCES

Table S5. Information of the sequencing data for each sample: quality of the raw data, number of reads before and after trimming, efficiency alignment and bisulfite conversion efficiency

Sample	Quality scores ≥Q30 (%)	Number reads before trimming	number reads after trimming	Alignment efficiency (%)	Bisulfite conversion efficiency (%)
Fam1_1_MHT	84.0	663270	660193	87.4	99.7
Fam1_1_MLT	86.8	579689	576949	89.1	99.7
Fam1_10_FLT	85.6	870587	865560	35.4	99.1
Fam1_10_MHT	70.6	374875	373317	63.2	99.7
Fam1_11_MLT	70.4	85423	84977	53.7	99.5
Fam1_12_FLT	87.0	510934	508119	42.9	99.7
Fam1_12_MHT	77.2	502960	500941	75.3	99.7
Fam1_13_MHT	81.4	414473	412616	84.6	99.7
Fam1_13_MLT	75.7	478134	476238	68.1	99.5
Fam1_14_FLT	85.5	784384	780630	45.6	99.3
Fam1_14_MHT	78.1	654094	651414	84.0	99.6
Fam1_15_FLT	84.2	612595	609742	59.6	99.6
Fam1_15_MHT	82.5	410448	408414	95.1	99.6
Fam1_16_MLT	76.5	522846	520469	73.8	99.6
Fam1_17_FLT	83.6	744496	741491	54.7	99.7
Fam1_17_MHT	84.3	491794	489523	88.5	99.6
Fam1_18_MLT	78.3	661699	658812	74.2	99.6
Fam1_19_MHT	56.5	438897	436493	46.1	99.6
Fam1_19_MLT	73.4	745698	742215	66.3	99.6
Fam1_2_MHT	72.5	473882	472157	82.4	99.7
Fam1_2_MLT	82.8	491815	489408	86.3	99.5
Fam1_20_MLT	84.8	730424	726889	52.4	99.6
Fam1_21_MLT	77.4	561900	560018	61.6	99.7
Fam1_22_MLT	87.7	745466	741832	59.0	99.5
Fam1_3_FLT	89.2	311504	310089	53.0	99.5
Fam1_4_MHT	78.1	288904	287760	78.5	99.7
Fam1_4_MLT	81.6	278121	276585	78.2	99.5
Fam1_5_MHT	75.4	221927	220856	77.9	99.7
Fam1_5_MLT	84.2	242394	241227	84.5	99.6
Fam1_6_MHT	74.3	631565	629024	72.5	99.7
Fam1_6_MLT	79.4	377090	375559	69.7	99.5
Fam1_7_FLT	87.2	333887	332213	57.7	99.3
Fam1_8_FLT	87.3	346344	344728	68.3	99.6
Fam1_9_FLT	86.4	776875	772537	51.7	99.6
Fam2_0_MLT	67.3	619509	615936	66.3	99.7
Fam2_1_MHT	83.9	302069	300544	83.4	99.6
Fam2_10_FHT	83.7	517175	513880	80.3	99.6

MOLECULAR ECOLOGY RESOURCES

Fam2_10_FLT	80.6	239459	238005	74.6	99.6
Fam2_11_FLT	54.7	393647	391249	41.1	99.6
# Fam2_11_MHT	#86.6	#471489	#469075	#82.9	#96.1
Fam2_12_FHT	88.3	236425	235230	82.1	99.6
Fam2_12_FLT	60.2	712937	708940	53.3	99.7
Fam2_13_FLT	71.0	818943	815503	60.1	99.7
Fam2_13_MHT	89.7	232734	231501	83.8	99.3
# Fam2_14_FHT	#67.4	#86	#86	#73.3	#97.7
Fam2_14_MLT	71.1	545839	541353	75.0	99.5
Fam2_15_FHT	80.3	533056	527861	86.0	99.6
Fam2_15_MLT	70.6	589836	586707	73.1	99.7
Fam2_16_FHT	87.0	373490	371253	84.4	99.5
Fam2_16_FLT	84.5	508395	505826	75.5	99.6
Fam2_17_MHT	87.1	240551	238992	84.0	99.4
Fam2_17_MLT	74.0	367667	365510	72.0	99.6
Fam2_18_MHT	85.6	513416	510763	82.9	99.5
Fam2_19_FLT	83.8	3199	311175	83.9	99.6
Fam2_19_MHT	91.1	379215	377245	89.4	99.5
Fam2_2_MHT	90.4	685302	681900	92.3	99.6
Fam2_2_MLT	72.3	306695	304636	71.1	99.5
Fam2_20_FHT	87.9	410936	408532	84.9	99.5
Fam2_20_FLT	73.7	602202	599546	68.9	99.6
Fam2_3_FHT	75.1	549351	547091	69.0	99.6
Fam2_3_MLT	81.7	333823	332216	83.7	99.5
Fam2_4_FHT	81.7	725085	721156	75.9	99.5
Fam2_4_FLT	69.5	68151	67667	70.1	99.7
Fam2_5_FHT	86.1	428220	426024	87.4	99.6
Fam2_5_MLT	58.6	510120	503954	53.0	99.6
Fam2_6_FHT	80.5	344182	342546	76.1	99.5
Fam2_6_FLT	72.2	318457	316750	76.2	99.7
Fam2_7_FHT	83.5	383269	381043	81.4	99.6
# Fam2_7_MLT	#38.3	#759440	#755330	#22.6	#97.2
Fam2_8_MHT	83.6	719377	715464	78.5	99.5
Fam2_8_MLT	68.9	166694	165697	63.2	99.7
Fam2_9_FHT	84.2	277488	276009	78.0	99.6
Fam2_9_MLT	77.6	258539	256769	73.9	99.7
Mean	79.3	469779	471392	71.9	99.6
SD	8.08	185638	186554	13.63	11.38
SEM	0.97	22188	22297	1.63	1.36

Values not computed because they did not pass the quality control check.

* Information of samples: Fam1 or Fam2= family 1 and 2. Number followed of fam correspond number assigned to sample. M= male; F = female; LT= low temperature (28°C) and HT = high temperature (35°C).

MOLECULAR ECOLOGY

RESOURCES

Table S6. Mean DNA methylation levels for each gene according to family, treatment and sex. Notice that for Family 1 at 35°C we did not have females due to full masculinization of the fish

	Family							
	1				2			
	28°C		35°C	28°C		35°C		
Gene	Male	Female	Male	Male	Female	Male	Female	
<i>cyp19a1a</i>	95.53	32.63	95.80	32.42	39.73	93.86	33.62	
<i>cyp11a1</i>	67.44	40.20	N/A	41.92	43.21	69.24	46.56	
<i>cyp11c1</i>	86.43	63.38	87.08	64.57	61.84	86.36	66.41	
<i>hsd11b2</i>	89.11	25.50	89.24	20.07	23.58	85.86	24.17	
<i>hsd17b1</i>	79.60	32.16	79.44	37.38	33.12	78.84	32.83	
<i>hsd17b3</i>	97.35	90.84	97.02	89.45	90.33	96.38	91.33	
<i>amh</i>	85.93	33.28	88.45	34.77	26.77	88.84	31.49	
<i>dmrt1</i>	3.15	29.81	3.36	23.76	25.98	5.03	27.21	
<i>dmrt3a</i>	1.35	2.69	1.23	2.22	2.29	1.68	2.49	
<i>foxl2a</i>	0.60	0.87	0.58	0.89	1.03	0.77	1.20	
Mean	60.65	35.13	60.24	34.74	34.79	60.69	35.73	
SD	39.38	25.09	41.66	25.47	25.25	38.77	25.85	

MOLECULAR ECOLOGY RESOURCES

Supplementary figures

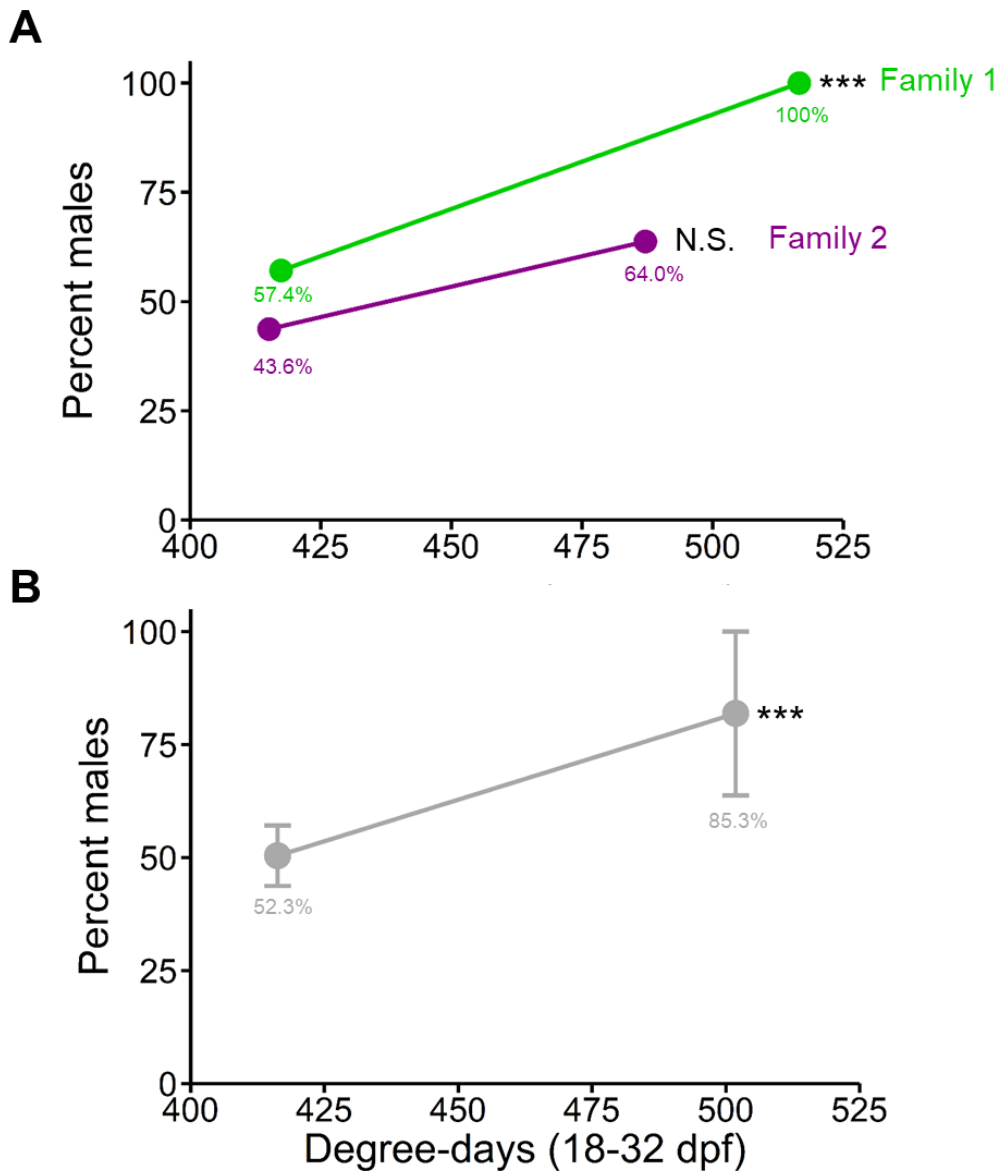
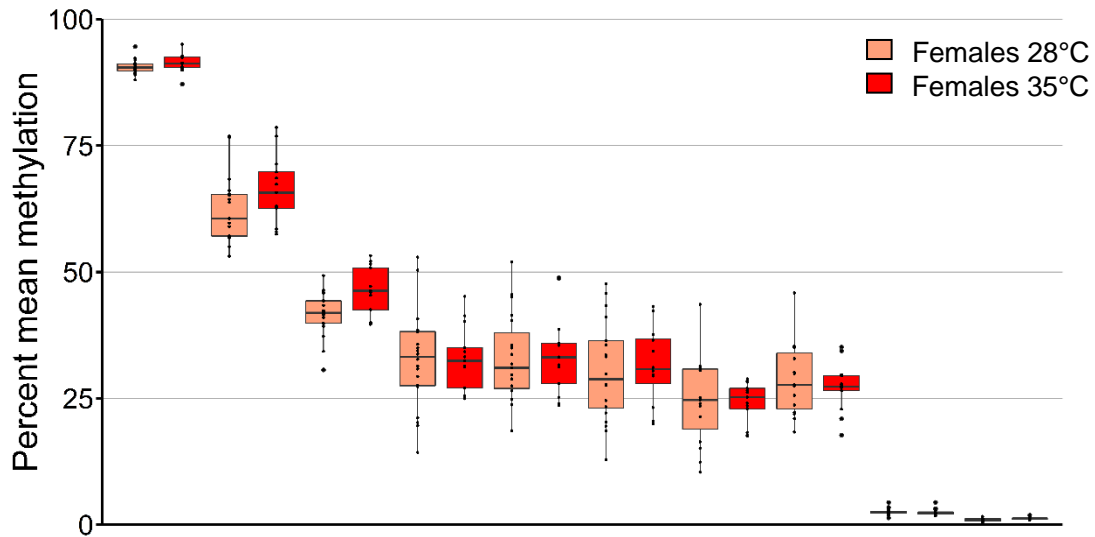


Figure S1. Percent males as a function of temperature (expressed as accumulated degree-days) during the sex differentiation period (18–32 days post-fertilization, dpf) in **A)** Separate families 1 and 2 and **B)** Grouping families 1 and 2. Sex ratio between groups was analyzed by χ^2 test. *** = $P < 0.001$; N.S = no significant. The sample size in each family is shown in Table S2.

MOLECULAR ECOLOGY RESOURCES

A



B

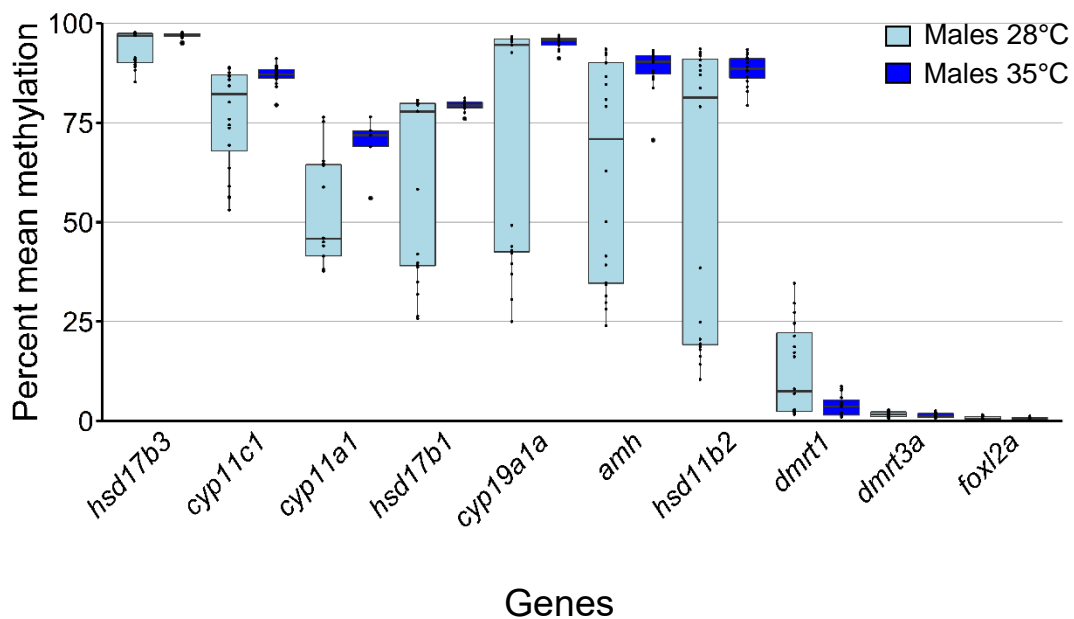


Figure S2. DNA methylation levels per gene in **A)** females and **B)** males exposed to control (low) and high temperature. The boxes include the values distributed between the lower and upper quartiles. The black dots in the boxplots correspond to individual samples and the black line inside the box represents the median.

MOLECULAR ECOLOGY RESOURCES

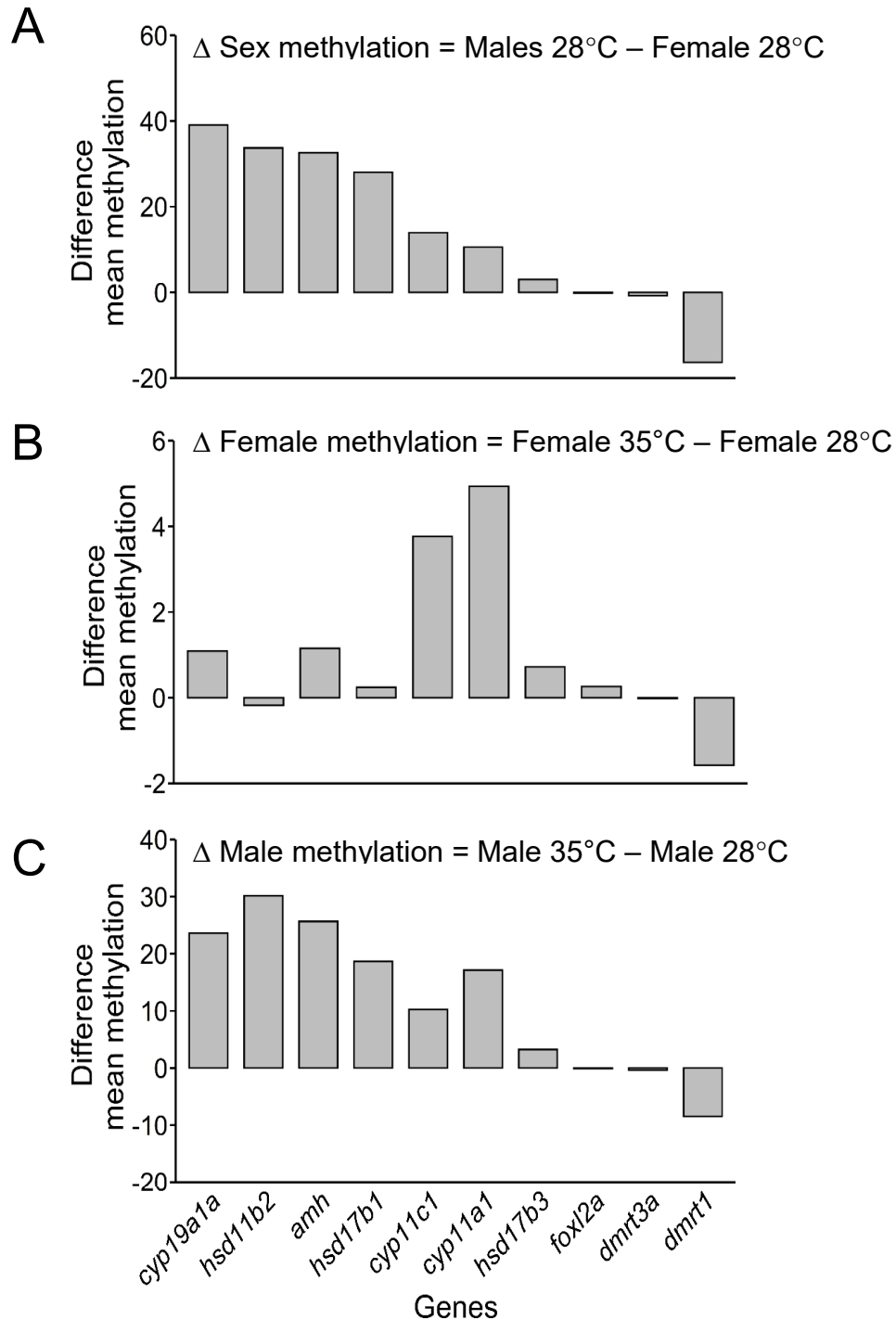
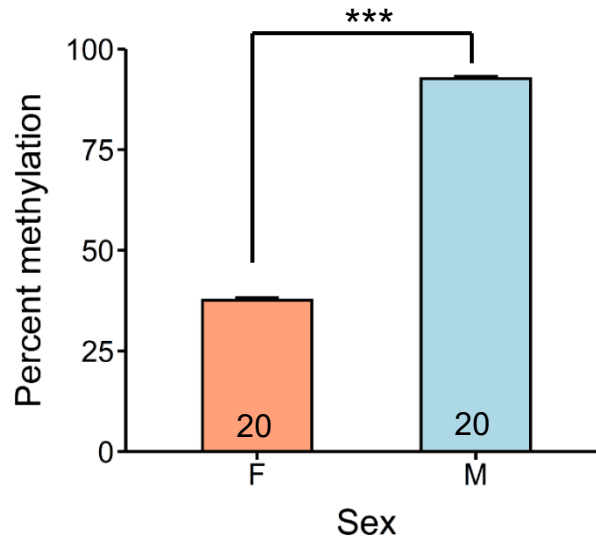


Figure S3. Differences of the percent mean DNA methylation levels in the steroidogenic enzymes and growth and transcription factors combining the methylation data of the two families. **A)** Differences of the percent mean DNA methylation levels between males minus females exposed to control temperature; **B)** Differences of the percent mean DNA methylation levels between females exposed to high minus control temperature. **C)** Differences of the percent mean DNA methylation levels between males exposed to high minus control temperature.

MOLECULAR ECOLOGY RESOURCES

A



B

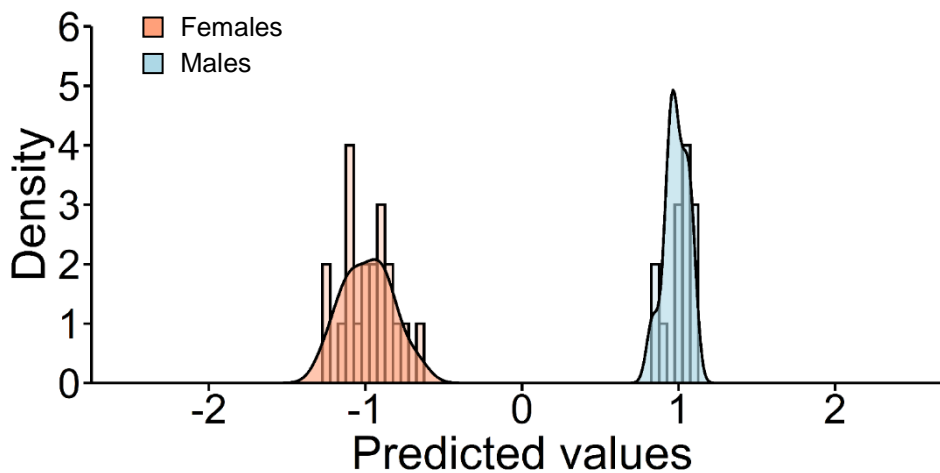


Figure S4. A) DNA methylation levels of the promoter regions of *cyp19a1a* according to sex: female (F) and male (M) of external offspring derived from a female and male (used for cross-validation) raised at 28°C. The numbers of fish analyzed in each group are shown inside the bars. A Kruskal-Wallis test was applied (***) = $P < 0.001$). Data are shown as mean \pm s.e.m. B) Histogram of the 40 samples (20 males and 20 females) from external offspring classified using the training test from machine learning (Flexible Discriminant Analysis) using the methylation levels of the 15 CpGs located in the *cyp19a1a* amplicon.

MOLECULAR ECOLOGY

RESOURCES

References

1. Tzung KW, Goto R, Saju JM, Sreenivasan R, Saito T, Arai K, Yamaha E, Hossain MS, Calvert ME, Orbán L. Early depletion of primordial germ cells in zebrafish promotes testis formation. *Stem Cell Rep.* 2015;4(1):61-73.
2. Hao R, Bondesson M, Singh AV, Riu A, McCollum CW, Knudsen TB, Gorelick DA, Gustafsson JÅ. Identification of estrogen target genes during zebrafish embryonic development through transcriptomic analysis. *PLoS ONE.* 2013;8(11):e79020.
3. Tokarz J, Norton W, Möller G, De Angelis MH, Adamski J. Zebrafish 20 β -hydroxysteroid dehydrogenase type 2 is important for glucocorticoid catabolism in stress response. *PLoS ONE.* 2013;8(1):e54851.
4. Ma Z, Tang S, Su G, Miao Y, Liu H, Xie Y, Giesy JP, Saunders DM, Hecker M, Yu H. Effects of tris (2-butoxyethyl) phosphate (TBOEP) on endocrine axes during development of early life stages of zebrafish (*Danio rerio*). *Chemosphere.* 2016;144:1920-7.
5. Tang R, Dodd A, Lai D, McNabb WC, Love DR. Validation of zebrafish (*Danio rerio*) reference genes for quantitative real-time RT-PCR normalization. *Acta Bioch Bioph Sin.* 2007;39(5):384-90.