1  **Sex differentiation in grayling (Salmonidae) goes through an all-male stage and is**
2  **delayed in genetic males who instead grow faster**
3

4  Diane Maitre[1*], Oliver M. Selmoni[1*], Anshu Uppal[1*], Lucas Marques da Cunha[1], Laetitia G.
5  E. Wilkins[1], Julien Roux[1,2], Kenyon B. Mobley[1], Susanne Knörr[3#], Marc Robinson-
6  Rechavi[1,2#], Claus Wedekind[1#]
7  *[1] Department of Ecology and Evolution, Biophore, University of Lausanne, 1015 Lausanne,*
8  *Switzerland*
9  *[2] Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland*
10 *[3] Aquatic Ecology and Toxicology Group, Center of Organismic Studies, University of*
11 *Heidelberg, Heidelberg, Germany*
12 *\* equal contributors*
13 *[#] shared senior authors*
14
15 Present addresses:
16 O. Selmoni: Swiss Federal Institute of Technology (EPFL), Bâtiment GC, 1015 Lausanne,
17 Switzerland
18 L. Wilkins: Department of Environmental Sciences, Policy and Management, 130 Mulford
19 Hall #3114, University of California, Berkeley, CA 94720, USA
20 K. Mobley: Max-Planck Institute for Evolutionary Biology, Department of Evolutionary
21 Ecology, August Thienemann Str. 2, 24306 Plön, Germany
22 J. Roux: Department of Biomedicine, University of Basel, Hebelstr. 20, 4031 Basel,
23 Switzerland
24
25 Correspondence: Claus Wedekind, claus.wedekind@unil.ch, orcid.org/0000-0001-6143-4716
26

27 **Abstract**
28 Fish can be threatened by distorted sex ratios that arise during sex differentiation. It is
29 therefore important to understand sex determination and differentiation, especially in river-
30 dwelling fish that are often exposed to environmental factors that may interfere with sex
31 differentiation. However, sex differentiation is not sufficiently understood in keystone taxa
32 such as the Thymallinae, one of the three salmonid subfamilies. Here we study a wild
33 grayling (*Thymallus thymallus*) population that suffers from distorted sex ratios. We found
34 sex determination in the wild and in captivity to be genetic and linked to the sdY locus. We
35 therefore studied sex-specific gene expression in embryos and early larvae that were bred and
36 raised under different experimental conditions, and we studied gonadal morphology in five
37 monthly samples taken after hatching. Significant sex-specific changes in gene expression
38 (affecting about 25,000 genes) started around hatching. Gonads were still undifferentiated
39 three weeks after hatching, but about half of the fish showed immature testes around seven
40 weeks after hatching. Over the next few months, this phenotype was mostly replaced by the
41 "testis-to-ovary" or "ovaries" phenotypes. The gonads of the remaining fish, i.e.
42 approximately half of the fish in each sampling period, remained undifferentiated until six
43 months after fertilization. Genetic sexing of the last two samples revealed that fish with
44 undifferentiated gonads were all males, who, by that time, were on average larger than the
45 genetic females (verified in 8-months old juveniles raised in another experiment). Only 12%
46 of the genetic males showed testicular tissue six months after fertilization. We conclude that
47 sex differentiation starts around hatching, goes through an all-male stage for both sexes
48 (which represents a rare case of "undifferentiated" gonochoristic species that usually go
49 through an all-female stage), and is delayed in males who, instead of developing their gonads,
50 grow faster than females during these juvenile stages.
51

1

**Author contribution**

MRR and CW initiated the project. DM, OS, AU, LMC, LW, and CW sampled the adult fish, did the experimental *in vitro* fertilizations, and prepared the embryos for experimental rearing in the laboratory. All further manipulations on the embryos and the larvae were done by DM, OS, AU, LMC, and LW. The RNA-seq data were analyzed by OS, JR, and MRR, the histological analyses were done by DM, supervised by SK, and the molecular genetic sexing was performed by DM, OS, AU, and KBM. DM, OS, and CW performed the remaining statistical analyses and wrote the first version of the manuscript that was then critically revised by all other authors.

**1. Introduction**

Fishes show a great diversity of gonadal development and differentiation that can be classified into five categories: (i) gonochoristic species with individuals developing either testes or ovaries, (ii) sequential hermaphrodites that mature as males (protandrous) or (iii) as females (protogynous) and may change sex later in life, (iv) simultaneous hermaphrodites, and (v) all-female species that reproduce gynogenetically [1,2]. In the gonochoristic species, primordial germ cells are typically formed during embryo or early larval development and subsequently differentiate into male or female gonads under the influence of genetic mechanisms and/or endocrine, environmental, or behavioural signals [1]. This process can be direct, as in the so-called "differentiated" gonochoristic species [3] where primordial germ cells develop without any detour into testicular or ovarian tissues, for example, in Arctic charr (*Salvelinus alpinus*) [4] and some cyprinid fishes [5,6]. In "undifferentiated" gonochoristic species, the typical pattern is that individuals first develop ovarian tissues that may subsequently degenerate, followed by a masculinization of the gonads that finally leads to normal testes, as in zebrafish (*Danio rerio*) and some other cyprinids [7,8]. Other species, so-called "secondary" gonochoristics [9], seem to first develop into simultaneous hermaphrodites before most individuals mature as only females or males, as in many eel populations *Anguilla* sp. [10,11]. We know of no example of an "undifferentiated" gonochoristic species where all individuals first develop immature testicular tissue followed by a feminization of the gonads that finally leads to normal ovaries.

Fishes also show a great diversity in sex determination systems that range from purely genetic to purely environmental, and different types of environmental sex reversals have been described in many different orders of the teleosts [1,12]. The fact that sex determination in fish can be very labile has several practical consequences: it can (i) be exploited in aquaculture where one-sex cultures can sometimes be more profitable [13], (ii) be used to control problem species, e.g. invasive species [14,15], (iii) be used to boost population growth in the wild [16], and (iv) cause some species to be sensitive to environmental changes, especially to different types of endocrine-disrupting pollutants [17,18]. It is therefore important to understand the diversity of sex determination and sex differentiation among fishes.

We study a population of European grayling (*Thymallus thymallus* L.) that uses the pre-alpine lake Thun as its feeding habitat, and that spawns in spring at the outlet of this lake. A yearly monitoring program initiated in the 1940s revealed significantly distorted adult sex ratios (an excess of males) that coincide with an abrupt temperature change in Europe [19,20] and that may contribute to the continuous decline of the population [21]. Grayling belong to the salmonids that include the three subfamilies Salmoninae (e.g. Pacific and Atlantic salmon, trout, char), Thymallinae (grayling), and Coregoninae (whitefish). Salmonids are usually keystone species of their respective habitat and of considerable economic and cultural importance for local communities. Sex determination seems mostly genetic in Salmoninae and Thymalinae, where it may be driven by a master sex-determining gene [22]. With regard to sex differentiation, much variation is observed among the Salmoninae [1], and little is known about the Thymalinae, including the European grayling (*Thymallus thymallus*).

2

103    Here we first verified that the *sdY* locus that Yano et al. [22] tested on 54 graylings from
104    a fish farm in France can be used to predict phenotypic sex in our wild study population. We
105    searched for genotype-phenotype mismatches (i) in wild-caught breeders, (ii) in F1 progeny
106    that were raised to maturity in captivity, and (iii) in juvenile fish that were raised in the
107    laboratory under warm or cold conditions. We then experimentally produced half-sib groups
108    and raised the embryos first singly in 24-well plates, then group-wise during their larval and
109    juvenile stages. We extracted mRNA from embryos, hatchlings, and early larvae, a time
110    window that corresponds to the onset of sexual differentiation in rainbow trout
111    (*Oncorhynchus mykiss*) [23], to detect sex-specific gene expression patterns. We also used
112    histological techniques to describe sex differentiation in relation to genetic sex markers over a
113    period of several months.
114
115    **2. Methods**
116    *2.1 Verification of genetic sex determination and genetic sexing of larvae*
117    For determining the sex of larvae and juveniles, genomic DNA was extracted from tissue
118    samples (tails or fins) using the DNEasy Blood and Tissue Kit (Qiagen, Hombrechtikon,
119    Switzerland), following manufacturer's instructions. To verify the efficacy of the sex typing
120    primers, fin clips were taken of in total 192 adults that were either caught from the study
121    population [21] (92 males and 29 females) or F1s that had been raised in captivity until sexual
122    maturity (20 males and 49 females). All these fish showed the typical sexual dimorphism of
123    this species. They were stripped for their gametes, i.e. phenotypic sex could be verified
124    without dissection. For adult tissue, DNA was extracted using a BioSprint 96 robot tissue
125    extraction kit (Qiagen).
126    Polymerase chain reaction (PCR) was conducted following protocols by Yano et al. [22]
127    with modifications. Briefly, each 15μL PCR contained 1.5μL 10X PCR Buffer, 210 μM
128    dNTPs, 1.5 mM $MgCl_2$, 0.3 μM each of primers *sdY E1S1* and *sdY E2AS4* (male-specific
129    amplification [22]), 0.075 μM each of primers *18S S* and *18S AS* ribosomal RNA (positive
130    amplification control [22]), 0.75 units of Taq Polymerase (Qiagen or Promega GoTaq), and 100
131    ng of DNA. Thermal cycling consisted of denaturing for 3 min at 95°C followed by 40
132    amplification cycles of 94°C/30s, 62°C/30s, and 72°C/30s, with a final extension of 10 min at
133    72°C. PCR products were visualized on a 1.5% agarose gel at 100V for 50min
134    (Supplementary Figure S1).
135    A multiplex reaction with three microsatellite markers previously used to characterize
136    population differentiation in grayling: *BFRO005, BFRO006* [24] *and Ogo2* [25] and with
137    *ThySex225* was used to analyze the tissue samples of the adults. We designed a small
138    amplicon (~225bps) of the *sdY* locus from sequences specific to grayling [22] using Primer3 [26]
139    (*ThySex225*: 5□-AGCCCAGCACTCTTTTCTTATCTC-3 □; genbank probe DB accession #:
140    Pr032825786). The 5' end of the *ThySex225* was labelled with a fluorescent label (ATTO532)
141    to aid viewing using capillary gel electrophoresis and to multiplex with microsatellites for
142    high-throughput genotyping. We used the reverse primer *sdY E2AS4* of Yano et al. [22]. Each
143    polymerase chain reaction (PCR) was accomplished in a10μl reaction volume containing
144    1.5μl water, 5μl of Qiagen Hotstar Taq Mix (final concentration 0.5 units of HotStarTaq
145    DNA Polymerase, 1XPCR buffer, 1.5mM $MgCl_2$ and 200uM of each dNTP), 0.3μl of each
146    primer (10μM), and 2μl of template DNA (1-5ng/μl). The thermal cycling profile consisted of
147    an initial denaturation for 15 min at 95°C followed by 35 cycles of 94 °C (30 sec), 56°C
148    reannealing temperature (90 sec) an extension phase at 72°C (60 sec), and a final extension at
149    72°C for 30 min. Fragment lengths were visualized on an Applied Biosystems® (ABI, Life
150    Technologies GmbH, Darmstadt, Germany) 3730 capillary sequencer and alleles were scored
151    using GeneMarker Version 2.6.4 software (SoftGenetics, LLC, State College, PA, USA).
152
153

154 *2.2 Breeding experiments*
155 Two sets of breeding experiments contribute to the present analyses. For the first breeding
156 experiment, gametes were stripped from 16 wild genitors (8 males and 8 females) and used
157 for block-wise full-factorial breeding with two males and two females per block. The embryos
158 of these 16 half-sib groups were distributed to 24-well plates (Falcon, Becton-Dickinson),
159 with two eggs per well that had been filled with 2 mL of chemically standardized water [27].
160 After hatching, in total 532 fish were about equally distributed to eight 200L aquaria filled
161 with filtered lake water (closed system). Fish were randomly assigned to 4 aquaria at 12°C
162 and 4 aquaria at 18°C in separate climate chambers. At day 174 after fertilization (i.e. day 145
163 after peak hatching), two of the aquaria of each temperature treatment were exposed to the
164 parasite *Tetracapsuloides bryosalmonae* in the course of another study (Uppal *et al.*,
165 unpublished manuscript). Exposure to *T. bryosalmonae* had no effect on any analysis
166 presented here (data not shown).
167 Fish were fed *ad libitum*, initially on a live zooplankton and then on dry food
168 (Skretting, Nutra Brut 3.0, 2.0, T-1.1). At days 63-71 post exposure, all fish were euthanized
169 with an overdose of Koi Med Sleep (Ethylenglycolmonophenylether) and sexed by visual
170 inspection of the gonads or via the sdY genotype. In total 60 individuals were both
171 morphologically and genetically sexed to test for possible genotype-phenotype mismatches at
172 this stage.
173 The second breeding experiment was performed similarly to the first breeding
174 experiment, with the following modifications. The genitors were sampled from a captive
175 breeding stock (F1 progeny of the wild population), and their gametes stripped and used in
176 two full-factorial breeding blocks with 4 females crossed with 5 males each, resulting in 40
177 different half-sib groups. The embryos were raised singly in 2 mL wells of 24-well plates at
178 7°C. At 14 days post fertilization (*dpf*), embryos were exposed either to 1 ng/L 17α-
179 ethinylestradiol ("EE2"), to *Pseudomonas fluorescens* ("PF"; $10^6$ bacterial cells per well),
180 simultaneously to EE2 and *P. fluorescens* ("EPF"), or sham-treated ("control") in the course
181 of parallel studies on the evolutionary potential of the population to adapt to environmental
182 stress during embryogenesis (Marques da Cunha, L., Maitre D. *et al.* unpublished manuscript)
183 and on EE2-effects on gene expression at different developmental stages [28].
184 In order to induce and synchronize hatching, the temperature was raised to 10°C at 27
185 *dpf* and to 11.5°C the next day. At 40 *dpf*, i.e. 11 days after peak hatching, a random sample
186 per treatment was transferred to 8 tanks filled with 200 L lake water. Two of the 8 tanks each
187 were stocked with fish that had been either exposed to EE2, *P. fluorescens*, EE2 and *P.
188 fluorescens*, or nothing, respectively. Once per week, 40L of water per tank was replaced with
189 either filtered lake water (groups "controls" and "PF") or with filtered lake water to which
190 EE2 had been added to reach a concentration of 1ng/L (groups "EE2" and "EPF"). The larvae
191 were fed with live *Artemia*, then live copepods and later also dry food as in the first breeding
192 experiment. Temperature was gradually increased to 18°C towards the end of the study at 163
193 *dpf* (in order to simulate the increase of mean temperature in the wild). The treatment with *P.
194 fluorescens* during embryogenesis did not show any significant effects during larval stages
195 while treatment with EE2 delayed sex differentiation [28] and seems responsible for 2 cases of
196 ovarian tissues in genetic males. EE2-treated individuals were therefore excluded from the
197 present analyses.
198 Larval length was determined by digital analysis of photos taken from freshly killed
199 fish (first breeding experiment) or from the fixed and stained sections that were used for
200 histological analyses (second breeding experiment; see below) using ImageJ [29].
201
202 *2.3 Sampling and preparations for gene expression analyses*
203 A subset of 5 half-sib groups (1 female crossed with 5 males) from the treatment groups
204 "controls" and "EE2" of the second breeding experiment was used to also study gene

4

205 expression during early developmental stages. The first samples (12 embryos per family and
206 treatment, i.e. 120 in total) were taken at 21 *dpf*. Embryos were immediately transferred to
207 RNAlater (Thermo Scientific, Reinach, Switzerland) and snap frozen at -80°C. The second
208 sampling occurred the day of hatching (31 *dpf*) with 8 larvae per family and treatment (i.e. 80
209 in total). The third sampling was at day 52 *dpf* (i.e. 21 days after hatching) with 5 larvae per
210 family and treatment (40 in total). Larvae of the second and third sampling were euthanized
211 with KoiMed (0.5 mL/L for five minutes) and then decapitated. The heads were immediately
212 stored in RNAlater at -80°C for later analyses. We sampled heads because the neuroendocrine
213 system is known to play a crucial role in the sexual differentiation [1]. For example, sex
214 differentiation of the brain is strongly dependent on the local action of estrogenic compounds
215 [30].
216   RNA extractions were performed using the QIAgen 96 RNeasy Universal Tissue Kit
217 (Qiagen) following the manufacturer instructions, except that the centrifugation was done at
218 half of the protocol speed for double the amount of the time (Eppendorf 5804 R centrifuge
219 with an A-2-DWP rotor; Eppendorf, Schönenbuch, Switzerland). In total, three distinct runs
220 of extractions (up to 96 samples each) were performed. Samples under the same treatment,
221 from the same family and collected at the same developmental stage were assigned to the
222 same run of extraction. RNA was extracted from the whole egg for the first sampling date and
223 from heads only in subsequent samples. Samples were eluted in 100 µL of RNase free water.
224   The RNA extraction protocol we used did not include a DNase treatment. Therefore,
225 we amplified DNA traces inside the RNA samples to determine the sdY genotype of eggs and
226 fry. We used two amplification protocols using the 18s gene as an internal control. The first
227 method was used in multiplex for samples with a high amount of DNA. For the samples with
228 low DNA content, the second PCR protocol was used in single reactions with half the
229 amounts of the respective primers each. After genetic sexing, one female and one male per
230 maternal half-sib group and time point was haphazardly chosen for sequencing. In one family,
231 two females were used for the second time point because no male was found in the respective
232 family. In another family, two males were used for the third time point each because no
233 female could be found in the respective family.
234   The 60 samples designed for sequencing were checked for quality (absorbance ratios
235 and RNA-Quality-Number, RQN) and concentration using both Nanodrop (Thermo
236 Scientific, Reinach, Switzerland) and a Fragment Analyser (Advanced Analytical, Ankeny,
237 USA). All samples were provided for library preparation in an equimolar concentration of 6
238 ng/µL in 100 µL of RNAse-free water. For each library, 50 µL were used (*i.e.* 300 ng of
239 RNA). The libraries were prepared in one batch on a robot using the Truseq Stranded RNA
240 protocol (Illumina, Part# 15026495 Rev. A) and multiplexing adaptors. Libraries were then
241 sequenced on an Illumina HiSeq 2500 machine to produce 2*100 bp paired -end reads. The
242 60 samples were sequenced in ten lanes with six samples per lane. Both the library
243 preparation and sequencing steps were performed at the Genomic Technologies Facility at the
244 University of Lausanne. See Supplementary Tables S1 and S2 for further details.
245
246 *2.4  Bioinformatics RNA-seq processing*
247 Pairs of reads of 100 bps each were quality trimmed using fastq-mcf (ea-utils, version 1.1.2;
248 Aronesty, 2013). Low quality reads (mean Phred quality score below 20) and reads containing
249 adapter sequences were trimmed. All the retained reads were truncated and returned with a
250 length of exactly 2*90 bps. An additional quality check using FastQC (version 0.11.2;
251 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) showed abnormal k-mer
252 frequencies, so we removed ten additional bps at the 5' end of reads, returning 2*80 bps long
253 reads. This last step was performed with a custom Python script (all scripts available at
254 https://github.com/Oselmoni/GSD_MS), which also corrected the headers of the sequence

5

255  names in order to allow compatibility with downstream tools. After these processing steps,
256  the quality of each library was rechecked using FastQC.
257      Twenty-four libraries, i.e. two individuals randomly chosen across each of all 12
258  possible combinations of developmental stage, sex and treatment, were chosen (Table S3),
259  and duplicated reads were removed with fastq-mcf. The 24 de-duplicated read sets were then
260  merged and used for assembly with Trinity (version 2.0.3; Grabherr et al., 2011). To speed up
261  the assembly, we ran Trinity with the read coverage normalization option set to 50. The
262  assembled transcriptome statistics were provided by Trinity and custom Python scripts.
263      To exclude spurious transcripts from the assembly, we blasted the assembled
264  sequences against the UniProtKB/Swiss-Prot database (version of October the 16th 2015) [31]
265  using the blastx command line application (version 2.2.26)[32]. For each transcript, only the
266  best hit was kept using a cut-off E-value set at $10^{-6}$. The transcripts that did not have any
267  match were filtered out of the assembly. In total, 228,417 transcripts were kept, distributed
268  across 52,353 genes.
269      We used Kallisto (version 0.42) [33] to pseudo-map reads to the transcriptome. Next, we
270  summed the estimated counts of the isoforms of each gene by using a custom Python script.
271  The estimated counts per gene were used for differential expression analysis.
272      Normalization factors for each library were calculated using the Trimmed Mean
273  Method (TMM) [34] of the EdgeR package (version 3.12.0) [35]. A log transformation of count-
274  per-million (log2(cpm)) was then applied to raw expression values. Log transformed count-
275  per-million values were used for a Principal Component Analysis (PCA) of samples [36].
276  Identified outlier samples (s24, s25) were filtered out. Additionally, visual inspection of the
277  distribution of the normalized expression values showed that one sample (s58) did not follow
278  the trend of the other samples and was therefore filtered out.
279      To process the remaining 57 samples, only genes showing an expression of at least 2
280  cpm in at least five samples were kept (in total, 35,348 genes). Two more PCAs were
281  performed: one to investigate the role of experimental factors (developmental stage, sex,
282  sibgroup, sequencing lane, batch of RNA extraction) in the expression measures of control
283  individuals, and a second one applied to the whole expression matrix (i.e. with both control
284  and EE2 treated individuals). This last PCA allowed us to select the most important factors to
285  include in our model. We considered developmental stage, sex and treatment as a combined
286  variable (with twelve possible levels) and the sib-group as an independent variable. This
287  model included sufficient replication for each comparison: including sib-group inside the
288  "combined block" would not have allowed comparisons between the different levels of this
289  block. For this same reason, technical factors (sequencing lane and batch of RNA
290  purification) were not included in the model.
291      The differential gene expression analysis was performed using the Bioconductor
292  limma-voom package (version 3.26.3) [37,38]. We used the *voomWithQualityWeight*, which
293  calculated sample quality weights in addition to the observational weights of limma-voom [39].
294  We used a modified version of the *voomWithQualityWeight* function performing the cpm
295  transformation using the cpm function from the edgeR package, resulting in a reduced range
296  of minimal expression values across samples (since the minimal cpm values for each sample
297  depend on the size of the library). *voomWithQualityWeight* was run with a cyclic-loess
298  normalization step and the distribution of the expression values was manually checked
299  (Figure S2) before proceeding to the differential gene expression analysis.
300      A linear model was fit for each gene, and coefficients and standard errors were
301  computed for all contrasts of interest using limma. At each developmental stage, we obtained
302  p-values for each gene for the comparison between control males and control females. Q-
303  values [40] were obtained from the vector of p-values. A threshold of q=0.15 was applied to
304  select differentially expressed genes, i.e. a 15% false discovery rate.
305      In order to have a biologically meaningful annotation of Gene Ontology (GO; Ashburner

306 *et al.*, 2000) terms, we performed a similarity search via blastx command line application
307 (version 2.2.26; National Center for Biotechnology Information, 2008) of the remaining
308 transcripts against a customized database. This customized database contained all the protein
309 sequences from Uniprot (version of December the 15th 2015; Boeackmann *et al.*, 2003)
310 associated to Zebrafish. According to the transcript identifiers of the Trinity assembly, we
311 were able to identify isoforms from the same gene and to keep all the associated Uniprot
312 entries. This information was then processed by a custom Python script: for each gene, we
313 retrieved a list of all the GO terms associated to all blast hits of the respective isoforms. Our
314 annotation concerned only the Biological Function domain of the GO term classifications
315 (Ashburner *et al.*, 2000). The enrichment analysis of GO terms was performed using the
316 goseq package (version 1.22.0; Young *et al.*, 2010) of the R Bioconductor (version 1.20.1;
317 Huber *et al.*, 2015). The list of genes differentially expressed in a contrast were checked for
318 enrichment of GO terms using a Wallenius hypergeometric distribution inspired method
319 (Young *et al.*, 2010). A key feature of goseq is that it takes into account the length of the gene
320 to calculate the enrichment scores (Young *et al.*, 2010). Length of genes was measured by a
321 custom Python script. When a gene showed multiple isoform, the median isoform length was
322 used as length of the gene. For each GO term, goseq returns a p- value associated to the score
323 of the enrichment test. These p-values were used to filter and rank GO terms significantly
324 overrepresented in the gene list of interest. The results of the enrichment analysis where then
325 visualized with REVIGO (Supek *et al.*, 2011). For each GO terms enrichment analysis, the
326 REVIGO visualization concerned only the GO terms with a $p<0.05$. To assure output
327 readability, if the number of terms having a $p<0.05$ was above 150, only the 150 GO terms
328 with the lowest p-value were kept for the REVIGO visualization.

329
330 *2.5 Gonadal development*
331 In total 256 fish were randomly sampled 51, 79, 107, 135, and 159-163 *dpf*, i.e. about
332 monthly over the course of 5 months. These fish were euthanized with an overdose of
333 KoiMed Sleep, the heads stored in RNAlater at -80°C for further analyses, and the rest of the
334 bodies fixated for two weeks in Davidson solution (AppliChem product No. A3200) for
335 histology. They were transferred to embedding cassettes and dehydrated for 48h using a Leica
336 TP1020 tissue processor (Leica, Tempe, USA). Dehydrated tissues were flowed in hot
337 paraffin wax (Histoplast P, Serva, Heidelberg, Germany) using a paraffin dispenser
338 embedding (Leica EG1150H, Leica, Tempe, USA) and paraffin wax was finally cooled to
339 obtain a solid paraffin block. Sections were cut at 4 µm ventrally, from anterior to posterior,
340 floated in a water-bath, and collected onto glass slides. Sections were stained with standard
341 Mayer's haematoxylin and eosin staining (HE-stain [41]) and cover slipped to be conserved.
342 Fish sections were analysed by light microscopy using a Leitz Aristoplan microscope (Leitz,
343 Wetzlar, FRG) and analysed with an associated digital camera (Color View, Soft Imaging
344 Systems, Münster, FRG) supported by "Analysis software" (Soft Imaging Systems, Münster,
345 FRG).
346
347 **3. Results**
348 *3.1 Verification of genetic sex determination*
349 Using the microsatellite multiplex with the *Thysex225* locus, we assigned sex to 121 wild-
350 caught adults and 71 adults from the captive breeding stock. Males have a single peak at 228
351 bps at this locus whereas females only showed signals for the microsatellite loci. There was
352 perfect alignment of sex assigned in the field based on morphology and on the presence of
353 eggs and sperm with the multiplex protocol, i.e. there was no ambiguous assignment of sex
354 based on this procedure.

355    Genetic sexing of 237-245 *dpf* old juveniles (first breeding experiment) based on the
356 multiplex PCR using *sdY* primers, and 18S primers as a positive amplification control, also
357 provided a perfect alignment between gonad morphology and the multiplex protocol with 60
358 individuals.
359
360 *3.2 Gene expression*
361 RNA quality check, sexing PCR results, RNA sequencing reads quality check, transcriptome
362 assembly statistics and data preparation for differential gene expression analysis are given in
363 the Supplementary Material. Figure 1 shows the results of a principal component analysis
364 (PCA) on the normalized expression levels matrix of the 28 samples. Gene expression
365 clustered first by developmental stage, and second by maternal half-sibgroups within each
366 developmental stage (Fig. 1). Inside each developmental stage, sex appeared to best explain
367 the variance in gene expression in the first two principal components in hatchlings (Fig. 1a).
368 This is confirmed in Table 1 that reports the number of genes differentially expressed between
369 males and females at each developmental stages. While only 15 genes were differentially
370 expressed at the embryo stage (Fig. S3), a strong increase of differently expressed genes could
371 be observed at the day of hatching (72% of total genes, Fig. S4). The sex difference remained
372 important at the first feeding stage (3% of total genes, Fig. S5). The GO enrichment analysis
373 of the genes differentially expressed between males and females are shown in the
374 supplementary figures S6-S9.
375
376 *3.3 Gonadal development*

377 Figure 2 gives an overview of the sex differentiation over the 5 sampling periods of the
378 second breeding experiment (Fig. 2). The figure also gives the frequency of genetic males and
379 females in the first and the second breeding experiment. All gonads were undifferentiated at
380 the first sampling period (51 *dpf*). The rate of undifferentiated gonads dropped to 30.4% at the
381 second sampling period (79 *dpf*). The other fish first showed early testicular tissues while
382 there were no signs of ovarian tissue. From the second to the third sampling day (107 *dpf*), the
383 rate of fish with testes only dropped from 69.6% to 20.8%, while 45.8% of the fish showed
384 testis-to-ovary or ovaries only. From then until the fifth sampling, the rate of fish with
385 undifferentiated gonads did not decline further, and there was no clear change in the rates of
386 fish with different types of gonadal tissues varied. All fish that remained undifferentiated at
387 the fourth and fifth sampling period turned out to have the male genotype, while all except
388 three of the remaining fish showed the females genotype (Table 2). See Figure 3 for examples
389 of the various phenotypes.
390
391 *3.4 Sex difference in growth*
392 In the first breeding experiment (sampling between days 237 and 245 *dpf*, i.e. 208-216 post
393 hatching peak), males were larger than females (on average 56.5 mm vs 54.9 mm; 95% CI =
394 ±1.8 each, $N_{total}$ = 436) while rearing temperature had no significant effect on size (effects of
395 sex: t = 2.4, p = 0.016; of rearing temperature: t = -0.8, p = 0.45). Males were also heavier
396 than females (t = 1.9, p = 0.05). These findings were confirmed in the second experiment
397 were genetic males were significantly larger than genetic females at the last two sampling
398 periods: male trunks were on average 40.6 mm (95% CI = ±5.3) long, while female trunks
399 were on average 35.3 mm (95% CI = ±5.8) (t = 2.7, p = 0.008).
400
401 **4. Discussion**
402 It is important to understand sex determination and sex differentiation especially in fish that
403 play key roles in their respective environment [42], that are, or could potentially be, important in
404 aquaculture [13], or that already suffer from distorted sex ratios as repeatedly observed in river-
405 dwelling salmonids [43]. All of this is true for the European grayling, and especially for the

406     population we study [21]. So far, little was known about sex determination and differentiation in
407     this species [22,44]. As a rule, sex determination is generally more diverse and also more labile
408     in ray-finned fishes and amphibians than it is in birds and mammals [45,46]. Sometimes,
409     variation in sex determination can even be found within a species or a genus [47,48]. Therefore,
410     even though Yano et al. [22] found genetic sex determination in 54 grayling sampled from a fish
411     farm in France, it seems necessary to verify their finding in a geographically distinct wild
412     population. We therefore first tested whether sex determination in a wild and in a related
413     captive population is indeed genetic. Our multiplex protocols resulted in perfect alignment of
414     sex phenotype and genotype in 192 adults of wild and captive origin, and in 60 juveniles that
415     had been raised at cold and warm temperatures for 8 months. This demonstrates that sex
416     determination has a strong genetic basis in our study population. It also supports Yano et al.
417     [22]'s hypothesis that the sdY locus is conserved among two of the three subfamilies of the
418     Salmonidae, namely the Salmoninae and the Thymallinae, and it supports the conclusion of
419     Pompini et al. [44] that distorted sex ratios in grayling are not due to temperature-induced sex
420     reversal under ecologically relevant conditions. Because of the clear pattern we found, we
421     were able to use the sdY locus to study sex-specific gene expression at embryonic stages, and
422     sex-specific gonadal development at early juvenile stages.
423        Sex differentiation is expected to be largely controlled by steroids and gonadotropins,
424     produced mainly by the brain in early life, and later also by gonads [1]. We found that only a
425     few genes show sex-specific expression in late embryogenesis. However, we cannot exclude
426     the possibility that, by sampling whole embryos instead of heads only, we diluted, and
427     therefore potentially missed, some sex-specific gene expression in the brain. At the time of
428     hatching and when sampling heads only, we found that a very high number of genes showed
429     sex-specific expression. The number of differentially expressed genes dropped towards first
430     feeding but was still high around that developmental stage. Our findings support Baroiller et
431     al. [49] who concluded for another salmonid, the coho salmon (O. kisutch), that the maximum
432     sensitivity to an exogenous estrogenic treatment was around hatching time.
433        We analyzed gene expression in five different paternal half-sib groups that had been
434     experimentally bred and raised to differ only in their paternal contribution, i.e. they differed
435     only genetically (they all shared the same mother and were raised in the same environment).
436     As we sampled only one female and five male breeders, a reliable quantification of the
437     additive genetic effects on gene expression is not yet possible. However, our principle
438     component analysis suggests that family effects are important. This suggests that there is
439     significant heritability in gene expression around hatching. It remains to be tested whether
440     rapid evolution in response to anthropogenic changes of the environment is therefore possible.
441        The first microscopically detectable characteristic of gonadal differentiation is the
442     migration of the primordial germ cell [50]. At the time of hatching, the number of primordial
443     germ cells in salmonid gonads is small [50]. When we analyzed gonadal tissue three weeks after
444     hatching, all gonads were still undifferentiated. Four weeks later, about half of the juveniles
445     showed testicular tissues while the remaining fish showed undifferentiated gonads. The rate
446     of undifferentiated fish remained approximately constant over the next three sampling periods
447     up to about the 19th week after hatching, i.e. the rate of differentiated fish remained about
448     constant, too. Surprisingly, however, the male phenotype was increasingly replaced by female
449     phenotypes among the differentiated fish over the 14 weeks that were covered by the second
450     to the fifth sampling periods.
451        Within the Salmoninae, differentiated gonochorism where primordial germ cells
452     develop directly into testis or ovarian tissues has been described in, for example, Arctic charr
453     (Salvelinus alpinus) [4], brook charr (Salvelinus fontinalis) [50], brown trout (Salmo trutta) [51],
454     rainbow trout (O. mykiss) [52], and coho salmon (O. kisutch) [53]. The European grayling as
455     representative of the Thymallinae instead shows a rare form of undifferentiated gonochorism,
456     since undifferentiated gonochoristic species usually go through an all-female stage before

9

457 they differentiate into testis and ovaries [1]. We conclude from our observations that the
458 European grayling goes through an all-male stage before developing mature testis and
459 ovaries.
460      By the end of the observational period, all fish that still showed undifferentiated
461 gonads had the male genotype and had also grown faster, i.e. genetic males differentiated later
462 and reached larger body lengths and body weights than genetic females. Delayed male gonad
463 development was also observed in sea trout (*Salmo trutta* morpha *trutta* L.), but sex-specific
464 growth was not observed during the early juvenile stages of this species [51]. As we
465 experimentally raised fish at warm and cold temperatures in our first experiment, we could
466 also test for an interaction between sex-specific growth and water temperature. Such an
467 interaction would potentially help explain the sex-specific juvenile mortality that seems to
468 contribute to the observed correlation between increased water temperatures and population
469 sex ratio [21,44]. However, we could not find any such effects of temperature on sex ratio under
470 our laboratory conditions.
471      Among the 6 individuals with testis at the last two sampling times, three individuals
472 were genetic females and would possibly have developed ovaries later. All fish with female
473 gonads (testis-to-ovaries and ovaries) were genetic females. No genetic male showed female
474 gonadal tissue. As mentioned above, our study population suffers from male-biased
475 population sex ratios that may contribute to the continuous decline of the population [21].
476 Similar observations have been made in other salmonid populations [54]. Pollution of aquatic
477 systems by hormone-active substances is probably ubiquitous wherever humans live (e.g.
478 pollution by EE2, the synthetic component of the contraceptive pill), and there is no reason to
479 assume that lake Thun and the spawning place that is located within the city of Thun are an
480 exception. Various cities, villages, and different types of industries discharge their sewage
481 into the lake or into rivers that feed the lake, normally after treatment in sewage plants that
482 can, however, not be fully effective. Moreover, some of the lake's whitefish populations
483 (*Coregonus* sp.) have been observed to suffer from increased prevalence of gonadal
484 misdevelopments [55,56] that may be due to anthropogenic disturbances of the ecosystem.
485 However, we found no phenotype-genotype mismatch in any of the adult fish that had been
486 sampled in the wild and in the F1 breeding stock. This suggests that environmental sex
487 reversal and its possible effects on the next generation [17] do not, by themselves, explain the
488 distorted population sex ratios. It remains to be tested whether pollution by hormone-active
489 substances and other anthropogenic changes of the environment lead to sex-specific mortality
490 during early developmental stages, possibly because of sex-specific life histories, especially
491 the sex-specific growth patterns and timing of gonadal differentiation that we observed here.
492
493 **5. Acknowledgements**
505
506 **6. References**

507  1    Devlin, R. H. & Nagahama, Y. Sex determination and sex differentiation in fish: an
508       overview of genetic, physiological, and environmental influences. *Aquaculture* **208**,
509       191-364, doi:10.1016/s0044-8486(02)00057-1 (2002).
510  2    Wootton, R. J. & Smith, C. *Reproductive biology of teleost fishes*.  (Wiley-Blackwell,
511       2014).
512  3    Yamamoto, T. in *Fish physiology*  (eds W. Hoar & D. Randall)  117-175 (Academic
513       Press, 1969).
514  4    Chiasson, M. & Benfey, T. J. Gonadal differentiation and hormonal sex reversal in
515       arctic charr (*Salvelinus alpinus*). *Journal of Experimental Zoology Part a-Ecological
516       Genetics and Physiology* **307A**, 527-534, doi:10.1002/jez.407 (2007).
517  5    Komen, J., Yamashita, M. & Nagahama, Y. Testicular development induced by a
518       recessive mutation during gonadal differentiation of female common carp (*Cyprinus
519       carpio* L.). *Development Growth & Differentiation* **34**, 535-544 (1992).
520  6    Talikina, M. G. Sex differentiation and gonad development during the first years of
521       life in the bream *Abramis brama* from the Rybinsk water reservoir. *Voprosy
522       Ikhtiologii* **35**, 114-119 (1995).
523  7    Takahashi, H. Juvenile hermaphroditism in the zebrafish *Brachydanio rerio*. *Bulletin
524       of the Faculty of Fisheries Hokkaido University* **28**, 57-65 (1977).
525  8    Takahashi, H. & Shimizu, M. Juvenile inter sexuality in a cyprinid fish the Sumatra
526       barb *Barbus tetrazona tetrazona*. *Bulletin of the Faculty of Fisheries Hokkaido
527       University* **34**, 69-78 (1983).
528  9    Atz, J. W. in *Intersexuality in vertebrates including man*  (eds C. N. Armstrong & A.
529       J. Marshall)  145-232 (Academic Press, 1964).
530  10   Beullens, K. *et al.* Gonadal differentiation, intersexuality and sex ratios of European
531       eel (*Anguilla anguilla* L) maintained in captivity. *Aquaculture* **153**, 135-150,
532       doi:10.1016/s0044-8486(97)00018-5 (1997).
533  11   Colombo, G. & Grandi, G. Histological study of the development and sex
534       differentiation of the gonad in the European eel. *J. Fish Biol.* **48**, 493-512,
535       doi:10.1111/j.1095-8649.1996.tb01443.x (1996).
536  12   Stelkens, R. B. & Wedekind, C. Environmental sex reversal, Trojan sex genes, and
537       sex ratio adjustment: conditions and population consequences. *Mol. Ecol.* **19**, 627-646,
538       doi:10.1111/j.1365-294X.2010.04526.x (2010).
539  13   Piferrer, F. Endocrine sex control strategies for the feminization of teleost fish.
540       *Aquaculture* **197**, 229-281, doi:10.1016/s0044-8486(01)00589-0 (2001).
541  14   Gutierrez, J. B. & Teem, J. L. A model describing the effect of sex-reversed YY fish
542       in an established wild population: the use of a Trojan Y chromosome to cause
543       extinction of an introduced exotic species. *J. theor. Biol.* **241**, 333-341 (2006).
544  15   Cotton, S. & Wedekind, C. Control of introduced species using Trojan sex
545       chromosomes. *Trends Ecol. Evol.* **22**, 441-443 (2007).
546  16   Cotton, S. & Wedekind, C. Introduction of Trojan sex chromosomes to boost
547       population growth. *J. theor. Biol.* **249**, 153-161, doi:10.1016/j.jtbi.2007.07.016
548       (2007).
549  17   Cotton, S. & Wedekind, C. Population consequences of environmental sex reversal.
550       *Cons. Biol.* **23**, 196-206, doi:10.1111/j.1523-1739.2008.01053.x (2009).
551  18   Lange, A., Paull, G. C., Hamilton, P. B., Iguchi, T. & Tyler, C. R. Implications of
552       persistent exposure to treated wastewater effluent for breeding in wild roach (*Rutilus
553       rutilus*) populations. *Environ. Sci. Technol.* **45**, 1673-1679, doi:10.1021/es103232q
554       (2011).
555  19   Wedekind, C. & Küng, C. Shift of spawning season and effects of climate warming on
556       developmental stages of a grayling (Salmonidae). *Cons. Biol.* **24**, 1418-1423 (2010).

557  20    Reid, P. C. *et al.* Global impacts of the 1980s regime shift. *Global Change Biology* **22**,
558        682-703, doi:10.1111/gcb.13106 (2016).
559  21    Wedekind, C. *et al.* Persistent unequal sex ratio in a population of grayling
560        (Salmonidae) and possible role of temperature increase. *Cons. Biol.* **27**, 229-234,
561        doi:10.1111/j.1523-1739.2012.01909.x (2013).
562  22    Yano, A. *et al.* The sexually dimorphic on the Y-chromosome gene (sdY) is a
563        conserved male-specific Y-chromosome sequence in many salmonids. *Evol. Appl.* **6**,
564        486-496, doi:10.1111/eva.12032 (2013).
565  23    Krisfalusi, M. & Cloud, J. G. Gonadal sex reversal in triploid rainbow trout
566        (*Oncorhynchus mykiss*). *Journal of Experimental Zoology* **284**, 466-472,
567        doi:10.1002/(sici)1097-010x(19990901)284:4<466::aid-jez13>3.0.co;2-g (1999).
568  24    Sušnik, S., Snoj, A. & Dovč, P. Microsatellites in grayling (*Thymallus thymallus*):
569        comparison of two geographically remote populations from the Danubian and Adriatic
570        river basin in Slovenia. *Mol. Ecol.* **8**, 1756-1758 (1999).
571  25    Olsen, J. B., Bentzen, P. & Seeb, J. E. Characterization of seven microsatellite loci
572        derived from pink salmon. *Mol. Ecol.* **7**, 1087-1089 (1998).
573  26    Untergasser, A. *et al.* Primer3 - new capabilities and interfaces. *Nucleic Acids
574        Research* **40**, e115 (2012).
575  27    OECD. *OECD guideline for the testing of chemicals 203 (fish acute toxicity test),
576        Annex 2. 9.*, <http://www.oecd.org/> (1992).
577  28    Selmoni, O. M. *et al.* Ecologically relevant oestrogen pollution affects sex-specific
578        gene expression and delays sex differentiation in grayling (Salmonidae). *Submitted
579        manuscript*.
580  29    Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of
581        image analysis. *Nature Methods* **9**, 671-675 (2012).
582  30    Diotel, N. *et al.* Aromatase in the brain of teleost fish: Expression, regulation and
583        putative functions. *Front. Neuroendocrinol.* **31**, 172--192,
584        doi:10.1016/j.yfrne.2010.01.003 (2010).
585  31    Boeckmann, B. The SWISS-PROT protein knowledgebase and its supplement
586        TrEMBL in 2003. *Nucleic Acids Research* **31**, 365--370, doi:10.1093/nar/gkg095
587        (2003).
588  32    Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein
589        database search programs. *Nucleic Acids Research* **25**, 3389-3402,
590        doi:10.1093/nar/25.17.3389 (1997).
591  33    Bray, N., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq
592        quantification. *Nature Biotechnology* **34**, 525-527, doi:10.1038/nbt.3519 (2016).
593  34    Robinson, M. D. & Oshlack, A. A scaling normalization method for differential
594        expression analysis of RNA-seq data. *Genome biology* **11**, R25, doi:10.1186/gb-2010-
595        11-3-r25 (2010).
596  35    Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for
597        differential expression analysis of digital gene expression data. *Bioinformatics
598        (Oxford, England)* **26**, 139--140, doi:10.1093/bioinformatics/btp616 (2010).
599  36    Yeung, K. Y. & Ruzzo, W. L. Principal component analysis for clustering gene
600        expression data. *Bioinformatics* **17**, 763--774, doi:10.1093/bioinformatics/17.9.763
601        (2001).
602  37    Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear
603        model analysis tools for RNA-seq read counts. *Genome biology* **15**, R29,
604        doi:10.1186/gb-2014-15-2-r29 (2014).
605  38    Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-
606        sequencing and microarray studies. *Nucleic Acids Research* **43**, e47--,
607        doi:10.1093/nar/gkv007 (2015).

608　39　Liu, R. *et al.* Why weight? Modelling sample and observational level variability
609　　　improves power in RNA-seq analyses. *Nucleic acids research* **43**, e97,
610　　　doi:10.1093/nar/gkv412 (2015).
611　40　Storey, J. D. The positive false discovery rate: a Bayesian interpretation and the q-
612　　　value. *The Annals of Statistics* **31**, 2013--2035 (2003).
613　41　Mulisch, M. & Welsch, U. *Romeis - Mikroskopische Technik, 18th Edition.*
614　　　(Spektrum Akademischer Verlag, 2010).
615　42　Wedekind, C. Demographic and genetic consequences of disturbed sex determination.
616　　　*Phil. Trans. R. Soc. B* (in press).
617　43　Thompson, N. F. *et al.* Sex reversal, selection against hatchery females or wild males
618　　　does not explain differences in sex ratio between first generation hatchery and wild
619　　　steelhead, *Oncorhynchus mykiss. Environmental Biology of Fishes* **98**, 113-120,
620　　　doi:10.1007/s10641-014-0240-0 (2015).
621　44　Pompini, M. *et al.* Temperature-induced sex reversal is not responsible for sex-ratio
622　　　distortions in grayling *Thymallus thymallus* or brown trout *Salmo trutta. J. Fish Biol.*
623　　　**83**, 404-411, doi:10.1111/jfb.12174 (2013).
624　45　Beukeboom, L. W. & Perrin, N. *The evolution of sex determination.* (Oxford
625　　　University Press, 2014).
626　46　Perrin, N. Random sex determination: When developmental noise tips the sex balance.
627　　　*Bioessays* **38**, 1218-1226, doi:10.1002/bies.201600093 (2016).
628　47　Rodrigues, N., Vuille, Y., Loman, J. & Perrin, N. Sex-chromosome differentiation and
629　　　'sex races' in the common frog (*Rana temporaria*). *Proc. R. Soc. B Biol. Sci.* **282**,
630　　　doi:10.1098/rspb.2014.2726 (2015).
631　48　Geffroy, B. & Bardonnet, A. Sex differentiation and sex determination in eels:
632　　　consequences for management. *Fish Fish.* **17**, 375-398, doi:10.1111/faf.12113 (2016).
633　49　Baroiller, J. F., Guiguen, Y. & Fostier, A. Endocrine and environmental aspects of sex
634　　　differentiation in fish. *Cell. Mol. Life Sci.* **55**, 910, doi:10.1007/s000180050344
635　　　(1999).
636　50　Sacobie, C. F. D. & Benfey, T. J. Sex differentiation and early gonadal development
637　　　in brook trout. *N. Am. J. Aquac.* **67**, 181-186, doi:10.1577/a04-059.1 (2005).
638　51　Dziewulska, K. & Domagala, J. Testicular development in the sea trout (*Salmo trutta*
639　　　morpha *trutta* L.) after sex differentiation, with a reference to precocious maturation.
640　　　*Journal of Applied Ichthyology* **20**, 282-289, doi:10.1111/j.1439-0426.2004.00553.x
641　　　(2004).
642　52　Xu, G. *et al.* Morphology, sex steroid level and gene expression analysis in gonadal
643　　　sex reversal of triploid female (XXX) rainbow trout (*Oncorhynchus mykiss*). *Fish*
644　　　*Physiology and Biochemistry* **42**, 193-202, doi:10.1007/s10695-015-0129-7 (2016).
645　53　Piferrer, F. & Donaldson, E. M. Gonadal differentiation in Coho salmon,
646　　　*Oncorhynchus kisutch*, after a single treatment with androgen or estrogen at different
647　　　stages during ontogenesis. *Aquaculture* **77**, 251-262 (1989).
648　54　Clews, E., Durance, I., Vaughan, I. P. & Ormerod, S. J. Juvenile salmonid populations
649　　　in a temperate river system track synoptic trends in climate. *Global Change Biology*
650　　　**16**, 3271-3283, doi:10.1111/j.1365-2486.2010.02211.x (2010).
651　55　Bernet, D., Wahli, T., Küng, C. & Segner, H. Frequent and unexplained gonadal
652　　　abnormalities in whitefish (central alpine *Coregonus* sp.) from an alpine oligotrophic
653　　　lake in Switzerland. *Diseases of Aquatic Organisms* **61**, 137-148 (2004).
654　56　Urbach, D. *et al.* Gonadal alterations in male whitefish *Coregonus fatioi*: no evidence
655　　　for genetic damage reducing viability in early life stages. *Diseases of Aquatic*
656　　　*Organisms* **81**, 119-125, doi:10.3354/dao01952 (2008).
657
658

659

660 **Table 1.** Number of genes differentially expressed (q<0.15) between males and females at
661 three developmental stages.
662

| Developmental stage | Number of genes |
|---|---|
| Embryo | 15 |
| Hatchling | 25,372 |
| First feeding | 1,110 |

663

664

665 **Table 2.** Gonads at sampling periods 135 *dpf* and 159-163 *dpf* (both sampling periods pooled)
666 versus result of genetic sexing.
667

| Gonads | Genetic sexing | |
|---|---|---|
| | Males | Females |
| Undifferentiated | 22 | 0 |
| Testicular tissue only | 3 | 3 |
| Testis-to-ovaries | 0 | 14 |
| Ovaries | 0 | 5 |

668

15

669 **Figures**
670
671 **Figure 1.** Principal component analysis of the gene expression matrix. (a) The first two
672 principal components, explaining 35.0% and 34.7% of the observed variance, respectively,
673 and (b) the next two principle components, explaining 9.5% and 3.3% of the variance,
674 respectively. Round symbols represent females, squared symbols males, the colours represent
675 the 5 maternal half-sib families. Ellipses and the grey area in panel b emphasize the clustering
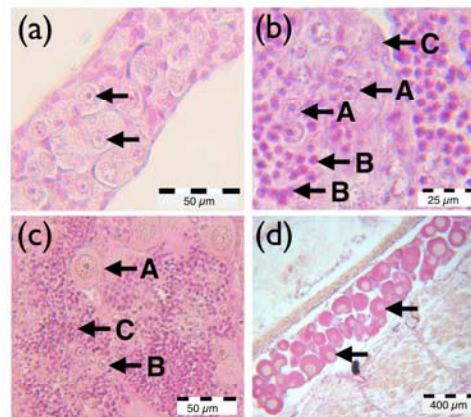676 by developmental stage.
677
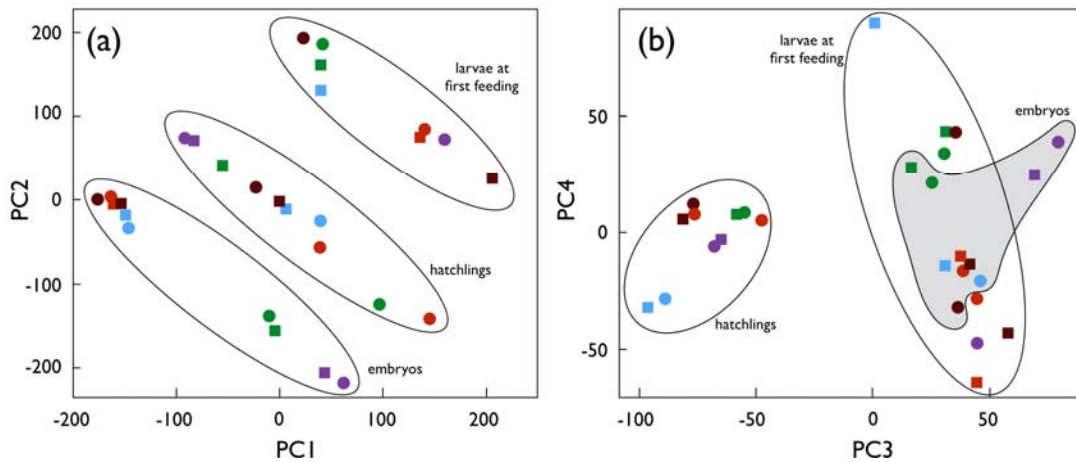


678
679

680    **Fig. 2** Sex differentiation in grayling. Frequencies of fish with undifferentiated gonads (grey
681    bars), gonads with testicular tissue only (blue bars), the testis-to-ovary phenotype (orange
682    bars), and ovaries only (red bars) for the second experiment (sampling periods between 51-
683    163 *dpf*) when phenotypes were determined by histology, and in the first experiment
684    (sampling period at 237-245 *dpf*) when phenotypes were determined by morphology. The
685    numbers in the boxes give the total sample sizes for the second experiment, and the number of
686    fish that were both phenotypically and genetically sexed for the first experiment. The blue and
687    red background colors indicate the overall frequencies of genetic males and females,
688    respectively, for each of the two experiments. Phenotype and genotype matched perfectly for
689    the first experiment (sampling period 237-245 *dpf*). At the last two sampling periods of the
690    second experiment (135 *dpf* and 159-163 *dpf*), all fish with undifferentiated gonads had the
691    male genotype, 3 of 6 individuals with testes had the male genotype, all other individuals had
692    the female phenotype. See table 2 for the match between phenotype and genotype at sampling
693    periods after 135 *dpf* and Fig. 3 for examples of the various developmental stages.
694
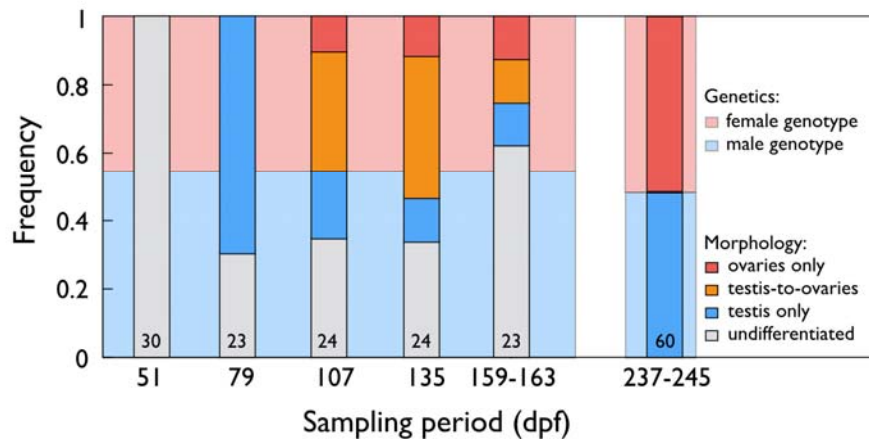


695
696

697  **Fig. 3** Representative examples of (a) undifferentiated gonads at fifth sampling period (arrows
698  mark oogonia), (b) immature testis at second sampling period (A = spermatogonia, B =
699  spermatocytes, C = Sertoli cells), (c) testis-to-ovary at fourth sampling period (A =
700  perinucleolar oocytes, B = spermatogonia, C = spermatocytes), and (d) ovary at fifth sampling
701  period (perinuclear oocytes). Samples were fixed in Davidsons, embedded in paraffin and
702  HE-stain.
703



704