1 2	Supporting Information . Pages S1 to S21. Supplementary Text. Supplementary Tables: Tables S1 to S10. Supplementary Figures: Figures S1 to S8.
3	Transcriptional changes in the ovaries of perch from
4	Chernobyl
5	^{1,2} *Adélaïde Lerebours, ³ Samuel Robson, ² Colin Sharpe, ⁴ Liubov Nagorskaya, ⁵ Dmitri Gudkov,
6	² Charlotte Haynes-Lovatt and ¹ Jim T. Smith
7	
8	¹ School of the Environment, Geography and Geosciences, University of Portsmouth,
9	Portsmouth, PO1 3QL, United Kingdom
10	² School of Biological Sciences, University of Portsmouth, Portsmouth, PO1 2DY, United
11	Kingdom
12	³ Centre for Enzyme Innovation, University of Portsmouth, Portsmouth, PO1 2DT, United
13	Kingdom
14	⁴ Applied Science Center for Bioresources of the National Academy of Sciences of Belarus,
15	Minsk, 220072, Belarus
16	⁵ Institute of Hydrobiology of the National Academy of Sciences of Ukraine, Kiev, UA-04210,
17	Ukraine
18	
19	
20	
21	Supplementary Text

22 Methods

Assembly and Annotation of the European Perch. The parameters used in Trim Galore were:
"--illumina -q 20 --stringency 5 -e 0.1 --length 20 --trim-n". The parameters used in Trinity were:
"--seqType fq --max_memory 100G --CPU 24 --min_contig_length 200 --min_kmer_cov 1 -SS_lib_type RF --verbose --full_cleanup".

27

28 **Results**

29 General health condition. The whole-body weight and the total length were not statistically significantly different between fish from different gonad phenotypes (p = 0.468 and 0.159 30 31 respectively) (Table S1). The Fulton condition index was significantly lower in fish with 32 developed/irradiated and underdeveloped gonads from the Cooling Pond (mean = 1 ± 0.7 , p =33 0.005 and mean = 1 ± 0.2 , p = 0.016, respectively) and in fish with developed/irradiated gonads 34 from Yanovsky lake (mean = 1 ± 0.1 , p = 0.021) as compared to fish with developed gonads 35 (mean = 1.3 ± 0.1) (Table S1). The Fulton condition index of fish with underdeveloped gonads 36 from Yanovsky lake (mean = 1.1 ± 0.1 , p = 0.052) and developed/irradiated gonads from 37 Svyatoye (mean = 1.2 ± 0.1 , p = 0.371) were not statistically significantly different to fish with 38 developed gonads (Table S1). No difference was observed between the liver weight and the 39 hepatosomatic index of fish with different gonad phenotypes (p = 0.372 and p = 0.0840 respectively) (Table S1).

41

42 De novo transcriptome assembly. To assess completeness, the assembled transcriptome was
43 compared against a database of universal *vertebrata* single copy orthologs using BUSCO v2.0
44 (Simão et al, 2015). This showed that 91.1 % of the universal single copy orthologs were present

45 in a complete form in the transcriptome, the majority of which (84.1 %) were found duplicated 46 suggesting redundancy amongst the assembled transcripts. Of those that were not found in a complete form, 4.8 % were present in a fragmented form, and only 4.1 % were missing entirely. 47 48 To assess the level of redundancy in the transcriptome, CD-HIT was used to cluster the 49 transcripts based on sequence similarity. Clustering of the transcriptome resulted in removal of 50 25,419 (4.8 %), 74,487 (13.9 %) and 99,724 (18.7%) transcripts when using sequence identity 51 thresholds of 100 %, 99 % and 98 % respectively, suggesting the presence of redundant 52 transcript sequences in the assembly. However, to avoid removing biologically distinct but 53 highly similar transcripts we proceeded with the unclustered transcripts, and instead controlled 54 for redundancy downstream. The majority (88.6 %) of the transcripts identified with an open 55 reading frame were annotated, whilst those with no open reading frame were generally not 56 annotated (Figure S3). In addition, the majority (73.8 %) of transcripts with no open reading 57 frame represented short fragments of less than 500 bases. Transcripts were filtered to remove 58 those annotated to bacterial, viral or archaeal species, those shorter than 500 bases in length, 59 those with a normalised abundance (FPKM value) lower than 1 in all samples, and those with no 60 open reading frame detected".

Functional classification of transcriptome. In total, genes were found associated with 28, 39 and 53 distinct terms of the three main functional categories of "cellular component", "molecular function" and "biological process" respectively. The 60 GO classes with more than 50 genes are shown in Supplementary Figure S4.

65

66 Supplementary Tables

67 **Supplementary Table S1:** Table showing the hydrological parameters of the lakes and the total 68 dose rate to perch. The total dose rate is the sum of the external (^{137}Cs) and internal (^{137}Cs and 69 ^{90}Sr) dose rate.

Lake	Maximum depth (m)	Surface (km ²)	Distance from CNPP (km)	¹³⁷ Cs deposition (kBq/m ²) "corrected to 6/05/1986"	Total dose rate (µGy/h) to perch in 2015
Yanovsky (H)	11	0.8	2.3	18650	9.8
Cooling Pond (H)	18	22	1.5	19180	7.9
Svyatoye (M)	5.1	0.25	225	1778	1.7
Stoyacheye (R)	17	0.46	167	293	0.1
Dvoriche (R)	3.5	0.128	95	100	0.1
Gorova (R)	2	0.05	88	10	0

70

71

Supplementary Table S2. A) Nutrient concentrations (NO₃⁻, NO₂⁻ and PO₄³⁻ in μ g/L, mean \pm SD, n = 3 per lake). B) Physicochemical parameters (Dissolved Oxygen, DO, in %, pH, temperature, conductivity in μ S/cm). C) Major alkali and alkali-earth element water concentrations (Na, Mg, S, K and Ca in mg/L, mean \pm SD, n = 3 per lake). D) Trace elements concentrations (As, Sr, Cd, Cs, Pb and U in μ g/L, mean \pm SD, n = 3 per lake) determined by ICP-MS. All the measurements were performed in the surface water samples of the 6 lakes in September 2014 and 2015.

79 A

Lakes	NO ₃	NO ₂ ⁻	PO ₄ ³⁻
Yanovsky (H)	76.1 ± 44.7	9.8 ± 14.7	6.0 ± 1.1
Cooling Pond (H)	57.9 ± 21.9	11.9 ± 3.9	15.0 ± 15.5
Svyatoye (M)	159.3 ± 81.8	2.0 ± 1.0	1.7 ± 0.5
Stoyacheye (R)	259.2 ± 182.7	1.5 ± 0.5	1.4 ± 0.9
Dvoriche (R)	201.0 ± 39.7	2.1 ± 0.2	1.4 ± 0.4
Gorova (R)	79.6 ± 68.7	2.5 ± 0.1	5.6 ± 1.0

8.3

55

17.2

255

82

83

84

85 C

Lake	Na	Mg	S	K	Ca
Yanovsky (H)	16.1 ± 1.3	4.9 ± 0.2	13.8 ± 1.8	4.8 ± 0.7	25.4 ± 1.4
Cooling Pond (H)	11.1 ± 1.4	7.9 ± 0.2	8.1 ± 1.1	3.2 ± 0.6	37.0 ± 1.3
Svyatoye (M)	2.02 ± 0.2	2.7 ± 0.1	0.3 ± 0.1	4.0 ± 0.2	18.0 ± 0.3
Stoyacheye (R)	2.7 ± 0.1	5.6 ± 0.3	0.8 ± 0.2	6.0 ± 0.3	33.2 ± 1.7
Dvoriche (R)	3.6 ± 0.1	5.7 ± 0.1	1.0 ± 0.1	2.4 ± 0.03	28.2 ± 0.5
Gorova (R)	6.2 ± 1.3	7.0 ± 0.3	2.4 ± 1.2	3.6 ± 0.7	37.8 ± 6.4

86

87 **D**

Lake	As	Sr	Cd	Cs	Pb	U
Yanovsky (H)	0.7 ± 0.2	117.5 ± 6.8	0.008 ± 0.009	0.005 ± 0.0007	0.02 ± 0.03	0.2 ± 0.1
Cooling P. (H)	0.8 ± 0.1	188.4 ± 14.2	0.002 ± 0.001	0.004 ± 0.0006	0.1 ± 0.3	0.6 ± 0.06
Svyatoye (M)	0.3 ± 0.001	64.4 ± 1.1	nd	0.01 ± 0.0009	nd	0.0006 ± 0.00005

80

81

B

Gorova (R)

Stoyacheye (R)	0.7 ± 0.05	61.5 ± 2.9	0.001 ± 0.001	0.006 ± 0.0002	nd	0.03 ± 0.003
Dvoriche (R)	0.9 ± 0.05	71.4 ± 1.8	0.0004 ± 0.002	0.02 ± 0.0007	0.01 ± 0.028	0.003 ± 0.0006
Gorova (R)	0.9 ± 0.2	159.8 ± 5.7	0.002 ± 0.001	0.002 ± 0.0007	0.006 ± 0.014	0.2 ± 0.03

- 89 **Supplementary Table S3:** Table showing the composition of the 6 different pools used for next
- 90 generation sequencing analyses. Sampling sites show either reference (R), medium (M) or high
- 91 (H) levels of radiation.

Pool Id.	Lakes	Gonad Phenotype	Number of fish
1	Vanavalay (II)	Developed/irradiated	3
2	i anovský (h)	Undeveloped	3
3	Cooling Dond (II)	Developed/irradiated	3
4	Cooling Pond (H)	Undeveloped	3
5	Svyatoye (M)	Developed/irradiated	3
	Gorova (R)		3
6	Dvoriche (R)	Developed	3
	Stoyacheye (R)		3

- 92 **Supplementary Table S4:** List of the target genes and the associated primers and probe used to
- 93 perform the qPCR validation.

Gene name	Primers	Probe
ska1	F: TCCAGCACATTAAGGACAACGT	CCTGCACACATGCCCCGGAA
	R: GCTTCCATTTGCTGGCTCTT	
cp2g1	F: GGTGGACCAGGCTGATGACT	CACAGGCAGAGCGCCAATGCC
	R: ACCGTAGCCCTTGGTAACTTTG	
hus1	F: TCTCTCATGGACGATATTGCACAT	ACCTGCTGACCCATGAGGAACTGCTG
	R: GGCCAGGGTGGACATCAG	
gthb1	F: ACCGAATGTAACCCCGAAAAC	CGTACTGCGGGCACTTTCACGGAC
	R: TTAAAATGAGACACAGCTGGATACG	
star	F: CGTCAATGTTAAGAAGCCATGTG	AGCGGCTTTTTCCTGTATCCTCGTCCA
	R: ACCAACCTGCATCATCATTCAG	
actb	F: CGTGACCTCACAGACTACCTCATG	ATCCTGACAGAGCGTGGTTACTCATTCACC
	R: CCTTGATGTCACGCACGATT	

95 Supplementary Table S5: Number of reads from *Perca fluviatilis* sequencing. Counts are given
96 for the raw data, the quality trimmed data and the data mapped back to the de novo assembled
97 transcriptome. For each library, the percent of G and C nucleotides, and the percentage of bases
98 with a quality score greater than 20 (Q20) or 30 (Q30) are shown.

Phenotype (Exposure Level)	Total Raw PE Reads	%GC	%Q20	%Q30	Total Trimmed PE Reads	Total Mapped Reads	Percent Mapped Reads
Developed/irradiat ed (High Level 1)	24,529,588	48	97.70%	95.60%	24,499,661	19,390,487	79.00%
Undeveloped (High Level 1)	24,350,401	49	97.80%	95.90%	24,320,952	20,127,478	82.70%
Developed/irradiat ed (High Level 2)	20,898,790	47	98.10%	96.50%	20,872,080	15,395,334	73.70%
Undeveloped (High Level 2)	29,294,698	49	97.80%	96.00%	29,257,591	24,930,050	85.10%
Developed/irradiat ed (Medium Level)	28,909,427	48	98.00%	96.30%	28,873,161	22,765,363	78.70%
Developed	23,802,568	47	97.80%	96.00%	23,766,562	18,650,271	78.40%
99							

100

101 **Supplementary Table S6:** Transcriptome assembly statistics from Trinity.

General Assembly Statistics	
Number of assembled transcripts	534,099
Number of assembled "genes"	309,937
Assembly GC content	46.31 %
Total number of raw PE reads	151,785,472
Total number of filtered PE reads	151,590,007
Total number of mapped PE reads	121,258,983
Transcript Statistics	
Median Contig Length	374 bp
Mean Contig Length	871.54 bp
Contig N10	5,937 bp
Contig N20	4,261 bp

Contig N30	3,238 bp
Contig N40	2,483 bp
Contig N50	1,879 bp
Gene Statistics	
Median Contig Length	321 bp
Mean Contig Length	607.55 bp
Contig N10	5,038 bp
Contig N20	3,334 bp
Contig N30	2,283 bp
Contig N40	1,486 bp
Contig N50	909 bp

103 **Supplementary Table S7:** A) Differential expression of some genes between the 104 Undeveloped/Irradiated, Developed/Irradiated and Developed/Reference phenoypes assessed by 105 the NGS and qPCR techniques. \uparrow : up-regulated; \downarrow : down-regulated; =: non significant (-1 < 106 $\log_2 (FC) < 1$). Results show \log_2 Fold Change ($\log_2(FC)$).

	Undeveloped/Irradiated compared to Developed/Reference		Undevelope compa Developed	ed/Irradiated ared to I/Irradiated	Developed/Irradiated compared to Developed/Reference	
	NGS	qPCR	NGS	qPCR	NGS	qPCR
ska1	=	1.8	=	1.9	=	=
gthb1	↓7.6	↓4.8	↓7.1	↓4.3	=	=
cp2g1	=	1.4	=	1.3	=	=
star	√9	√3.8	√8.5	√3.6	=	=
hus1	个10.3	个3.7	=	=	个10.3	个3

107

B) Gene expression results measured in individual gonad samples from the different lakes thatwere used to form the pools for NGS analyses. The bactin gene is used as the reference. The data

110 are expressed as mean
$$(2^{-\Delta}Ct) \pm SD$$
, $n = 3$.

Phenotype	Lake	Gene name				
		ska1	gthb1	cp2g1	star	hus1
Dev_Ref	Gorova	0.14 ± 0.02	0.011 ± 0.003	0.004 ± 0.002	0.009 ± 0.003	0.02 ± 0.008
	Dvoriche	0.11 ± 0.04	0.011 ± 0.006	0.001 ± 0.002	0.005 ± 0.002	0.02 ± 0.01
	Stoyacheye	0.07 ± 0.03	0.007 ± 0.003	0.004 ± 0.003	0.004 ± 0.0005	0.01 ± 0.005
Dev_Irr	Yanovsky	0.08 ± 0.06	0.005 ± 0.003	0.00002 ± 0.00002	0.005 ± 0.003	0.3 ± 0.6
	Cooling Pond	0.08 ± 0.02	0.009 ± 0.003	0.001 ± 0.001	0.005 ± 0.001	0.03 ± 0.003
	Svyatoye	0.13 ± 0.07	0.006 ± 0.003	0.004 ± 0.004	0.005 ± 0.001	0.01 ± 0.003
Undev_Irr	Yanovsky	0.26 ± 0.04	0.0004 ± 0.0004	0.005 ± 0.007	0.0004 ± 0.0004	0.07 ± 0.02
	Cooling Pond	0.45 ± 0.18	0.0003 ± 0.0002	0.004 ± 0.006	0.0004 ± 0.0002	0.4 ± 0.5

Supplementary Table S8: Excel sheet giving the differential expression analysis results from DESeq2 for comparisons between the phenotypes a) undeveloped gonads vs developed gonads, b) undeveloped gonads vs developed/irradiated gonads, c) developed/irradiated gonads vs developed gonads. Samples from different sampling sites were combined as replicates. BaseMean, log2FoldChange, lfcSE, stat, pvalue and padj represent standard output from the DESeq2 package. FPKM values represent the mean over all sampling sites. Transcripts were classified as being significantly differentially expressed if they showed a fold change greater than 2-fold (upregulated or downregulated), an adjusted p value less than 0.05, and a mean

normalized FPKM expression value greater than 1 for at least one of the two conditions. The raw
read counts and normalized FPKM values for all 534,099 transcripts are provided.

Supplementary Table S9: Excel sheet giving the gene ontology results for the comparisons between the phenotypes for a) undeveloped gonads vs developed gonads, b) undeveloped gonads vs developed/irradiated gonads, c) developed/irradiated gonads vs developed gonads. Gene ontology terms are collected into three classes; biological process, molecular function and cellular component.

133 **Supplementary Table S10.** Genes related to gonad development and pathway previously found responding to radiation exposure (*).

Gene	Comparison	Bibliographical Notes	
Nltp	Dev/irr vs Dev/ref (+21)	Encodes the non-specific lipid transfer protein. Involved in intracellular transport and signalling of lipids (Peretti <i>et al.</i> , 2020).	
Ldah	Dev/irr vs Dev/ref (+3) Undev/irr vs Dev/ref (+3)	Dev/irr vs Dev/ref (+3) Undev/irr vs Dev/ref (+3) Encodes the lipid droplet associated hydrolase.	
At11b	Dev/irr vs Dev/ref (-10)	Encodes the phospholipid transporting ATPase IF	
Odo1	Dev/irr vs Dev/ref (+9)	Catalyzes the conversion of 2-oxoglutarate, a substrate increased during the eggs development of the parasite wor in response to a higher energetic need (Beis and Barrett, 1975).	
	Undev/irr vs Dev/ref (+9)		
Cyh5	Dev/irr vs Dev/ref (-2)	Encodes an hemoprotein that modulates the synthesis of androgens by regulating the activity of CYP17A (Dharia et	
Cybs	Undev/irr vs Dev/ref (+11)	<i>al.</i> , 2004 ; Nakamura <i>et al.</i> , 2010,	
Cyp17a	Undev/irr vs Dev/ref (-7)		
Ovos	Undev/irr vs Dev/ref (-7)	Identified in the egg white of birds (Stevens 1991; Saxena and Tayyab, 1997), in cancerous ovaries of hens (Lim <i>et al.</i> , 2011)	
		Encodes a protein which is produced in granulosa and theca cells in humans (Agrawal et al., 2002)	
Vegfc	Undev/irr vs Dev/ref (-23)	Regulates oocyte quality (Bokal et al., 20005)	
		Regulates oocyte development (Devjak et al., 2012)	
Prgr	Undev/irr vs Dev/ref (-9)	Interact plays a major role in regulating female reproductive system (Gadkan-Sable et al., 2005)	
		Involved in oocyte maturation in the African clawed frog (Xenopus laevis) (Bayaa et al., 2000; Tian et al., 2000)	
K1C18	Dev/irr vs Dev/ref (+5)		

	Undev/irr vs Dev/ref (-6)		
Nell2	Undev/irr vs Dev/ref (-5)	Involved in the female reproductive cycle in mammals (Ryu <i>et al.</i> , 2011). Regulate the GnRH involved in timing of female puberty and gonad development (Ha <i>et al.</i> , 2008).	
Tcp4	Undev/irr vs Dev/ref (-10)	Encodes a transcription factor involved in cell development and timing of reproductive maturation in plants (Sarvepalli and Nath, 2011).	
Mod5	Dev/irr vs Dev/ref (+10)	Tumour superson conc	
	Undev/irr vs Dev/ref (+10)	rumour suppressor gene	
Cng2	Dev/irr vs Dev/ref (+10)	Tumour suppressor gong	
Gps2	Undev/irr vs Dev/ref (+9)	Tumour suppressor gene	
Rergl	Dev/irr vs Dev/ref (+6)	Encodes the Ras-related and estrogen-regulated growth inhibitor protein. Found downregulated in human high s	
	Undev/irr vs Dev/ref (+6)	ovarian serous cancer tissue samples (Lou et al., 2018)	
Eif3	Undev/irr vs Dev/ref (-10)	EIF3I initiates cell proliferation is involved in embryonic development (Sadato et al. 2018)	
		eif3 gene expression has been associated to gonadal development in freshwater crustaceans (Jiang et al., 2014)	
Ggnb2	Undev/irr vs Dev/ref (+1)	Involved in differentiation and in placental development (Li et al. 2016)	
	Dev/irr vs Dev/ref (+3)		
Rad51*	Undev/irr vs Dev/ref (+4)	Upregulated in marine mussels environmentally exposed to radionuclide discharges from the Sellafield nuclear facility, Ravenglass estuary, UK (AlAmri <i>et al.</i> , 2012).	
	Dev/irr vs Dev/ref (+10)		
Hus1*	Undev/irr vs Dev/ref (+10)		
W.J. J1*	Dev/irr vs Dev/ref (+10)		
wana1**	Undev/irr vs Dev/ref (+11)	Initiates DNA replication (Chen et al., 2017)	

Dmat1*	Dev/irr vs Dev/ref (+10)	
	Undev/irr vs Dev/ref (+11)	

135 Supplementary Figures

Supplementary Figure S1: Filtering approach for assembled transcripts from Trinity *de novo* transcriptome assembly. Transcripts were filtered to keep only those greater than or equal to 500bp in length with a minimum normalised abundance (FPKM) greater than 1 across all samples and an identified open reading frame. Transcripts that showed a match to a non-eukaryotic gene from the UniProt/SwissProt database were also removed.



- 151
- 152 **Supplementary Figure S2:** Proportion of reads for each sample that map to the European perch
- 153 (Perca fluviatilis) genome (blue), other related species predominantly stickleback (Gasterosteus
- 154 *aculeatus*), and reads that have no hit against any of the ~50 species analysed.



Supplementary Figure S3: Proportion of annotated transcripts split based on whether or not an open reading frame was identified for the transcript, and based on the transcript size. Those transcripts with a significant hit against genes in the Uniprot/Swissprot database are coloured blue.



181 Supplementary Figure S4: Number of unique genes present in each of the top GO classes for 182 cellular component, biological process and molecular function. GO classes with more than 50 183 genes are shown.



- 186
- 187
- 188
- 189
- 190

191 Supplementary Figure S5: Principal component analysis showing the distribution of the 192 different gonad phenotypes (developed/reference, developed/irradiated and 193 undeveloped/irradiated) and associated exposure levels (Reference, Medium, High) based on 194 their normalised gene expression profiles. The largest source of variation in the gene expression 195 data was found between the undeveloped phenotypes (blue) and both reference and irradiated 196 developed phenotypes (green and red), which accounts for 37.7% of the variance in the data.



Supplementary Figure S6: Heatmap showing the normalized abundance levels (FPKM) across all 33,047 filtered transcripts. Abundance levels are normalized on a row-by-row basis such that each row has mean 0 and standard deviation 1, so that transcripts with expression above the

201 mean across all samples appear red whilst transcripts with expression below the mean appear

202 blue.



- 204 Supplementary Figure S7. Venn diagram showing the overlap between the expressed
- 205 transcripts (FPKM > 1) for each sample.



213 Supplementary Figure S8: Results of differential expression analysis between different 214 conditions for a) undeveloped gonads vs developed/reference gonads (1,825 transcripts), b) 215 undeveloped gonads vs developed/irradiated gonads (1,994 transcripts), c) developed/irradiated 216 gonads vs developed/reference gonads (610 transcripts). Heatmaps show the normalized 217 abundance levels (FPKM) across all differentially regulated genes identified when using all 218 exposure levels as pseudo-replicates. Abundance levels were further normalized such that each 219 row has mean 0 and standard deviation 1, so that transcripts with expression above the mean 220 appear red whilst transcripts with expression below the mean appear blue.

