Marine Genomics 56 (2021) 100809

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Data article

The Atlantic salmon whole blood transcriptome and how it relates to major locus maturation genotypes and other tissues

Samuel C. Andrew [a], Craig R. Primmer [a,b], Paul V. Debes [a,c], Jaakko Erkinaro [d], Jukka-Pekka Verta [a,*]

[a] Organismal and Evolutionary Biology Research Programme, University of Helsinki, Finland
[b] Institute of Biotechnology, University of Helsinki, Finland
[c] Department of Aquaculture and Fish Biology, Hólar University College, Sauðárkrókur, Iceland
[d] Natural Resources Institute Finland (Luke), Oulu, Finland

ARTICLE INFO

Keywords:
Salmo salar
vgll3
Transcriptome comparison
Gene expression
Maturation

ABSTRACT

The Atlantic salmon (*Salmo salar*) is important to many ecosystems and local economies and has therefore become the focus of a broad range of research questions that have benefited from the availability of high-quality genomic resources. Albeit gene expression studies have been extensive for this species, the transcriptome information for Atlantic salmon whole blood has been lacking. A transcriptome of Atlantic salmon blood would be a valuable resource for future studies, especially those wishing to take non-lethal samples. Here, we report a whole blood transcriptome for Atlantic salmon constructed from twelve 8-month old salmon parr using RNA-seq. We identify transcriptomic proxies for the genotype at the major maturation timing locus *vestigial-like 3* (*vgll3*). Differentially expressed genes between the *early* and *late* maturing genotypes showed overrepresented Gene Ontology (GO) terms with the strongest result linked to 13 ribosomal subunit genes. To assess how the whole blood gene expression profile relates to other tissues, we compare the blood transcriptome to the reference transcriptome of fourteen other tissue types using both a common PCA method and a novel method. The novel method compares transcriptomes when gene expression is visualised as a layer using thin-plate spline smoothers. Both methods found similar patterns with the blood transcriptome being quite unique compared to the transcription profiles of other tissues.

1. Introduction

The Atlantic salmon (*Salmo salar*) is a key species in marine and freshwater ecosystems and has a high economic value to ecotourism and fisheries. The species economic importance as well as other factors have led to it becoming a model species for a number of biological questions. Major research focuses include the process of genome rediploidization after a whole genome duplication (Lien et al., 2016; Warren et al., 2014) and the regulation of maturation timing (Barson et al., 2015; Czorlich et al., 2018). Recent research on the Atlantic salmon has benefited from the availability of a wide range of genomic resources including SNP arrays (Barson et al., 2015; Bourret et al., 2013) reference transcriptomes (Mohamed et al., 2018) and a reference genome (Lien et al., 2016; Macqueen et al., 2017).

Despite the wide range of genomic resources available for the Atlantic salmon there is still no reference transcriptome available for the blood. The RNA transcripts present in the whole blood can be used to quantify physiological responses non-lethally (Liew et al., 2006), and as potential proxies for life history phenotypes such as age at maturity. Although cell-type specific transcriptomes can provide more detail, whole tissue (or in this case, whole blood) transcriptomes can be important first steps for understanding tissue specific processes and may be the only viable option when sampling in field conditions. The most abundant group of transcripts in the blood are the haemoglobin alpha and beta subunits that influence the properties of oxygen transport (Quinn et al., 2010). The third whole genome duplication in the common ancestor of teleost fish led to two haemoglobin gene clusters. In salmonids haemoglobin cluster two was lost before the salmonid-specific fourth genome duplication that resulted in a duplication of haemoglobin cluster one (Quinn et al., 2010). Haemoglobin diversity could allow salmon to adapt to changing metabolic demands and O_2 availability throughout the Atlantic salmon life cycle, through

* Corresponding author.
E-mail address: jukka-pekka.verta@helsinki.fi (J.-P. Verta).

https://doi.org/10.1016/j.margen.2020.100809
Received 3 May 2020; Received in revised form 23 July 2020; Accepted 24 July 2020
Available online 31 July 2020

regulating the expression of different paralogs (Opazo et al., 2013). Consistently, Quinn et al., (2010) found that alpha and beta subunit diversity was highest in the Atlantic salmon when compared to four other teleostean species without a fourth genome duplication. However, the expression profiles of haemoglobin subunit paralogs have not been explored.

Here we present the blood transcriptomes from twelve 8-month old parr. To illustrate the utility of the blood transcriptome as a proxy for life-history phenotypes, we identified Differentially Expressed Genes (DEGs) in the blood transcriptome for fish carrying either the *early*, or the *late vgll3* transcription cofactor maturation allele associated with the sexual maturation (Ayllon et al., 2015; Barson et al., 2015). We aim to compare these two genotypes at this early stage before maturation. The presented transcriptomes will also be a publicly available genomic resource for identifying genes that are likely to be expressed in the blood, to design targeted gene expression assays. Aside from providing these detailed data we will present two methods for visualising overall similarity between the whole blood transcriptome presented here and transcriptomes previously published for fourteen other Atlantic salmon tissue types.

2. Data description

2.1. Sample collection

Blood samples were taken from immature 8-month old Atlantic salmon juveniles that were being destructively sampled for tissue collection for other purposes (Table 1). The fish were reared at the University of Helsinki, Finland as part of a larger study where the *vgll3* genotype of each fish was known from controlled fertilisations (Debes et al., 2020). Animal ethics approval for this experiment was given by the national authorities (permit number ESAVI/2778/2018). For sample collection fish were euthanised with an overdose of MS222 (250 mg/L), after which an incision behind the gills was made and 50–100 μL of blood was drawn from the ventral aorta with a 200 μL pipette tip. The blood samples were immediately snap frozen in liquid nitrogen before being stored at −80 °C.

2.2. RNA extraction and sequencing

RNA extractions were done within a week of sample preservation using the TRIzol Reagent (ThermoFisher Scientific) according to the manufacturer's instructions (using 1 mL of TRIzol per sample). Whole RNA quality and quantity was assessed on a Bioanalyzer 2100 with the RNA 6000 Nano kit (Agilent Technologies).

Library preparation were done using the Illumina TruSeq Stranded mRNA kit and indexes (Illumina, San Diego, CA) following the standard protocol with the alteration that RNA fragmentation was done for 4 (not 8) minutes at 94 °C to yield the fragment length required for sequencing. All libraries were run on the same flow cell of an Illumina NextSeq 500 (pair-end 2 × 75 bp reads) at the Institute of Biotechnology at the University of Helsinki. All 12 sequenced individuals were homozygous at the *vgll3* locus: six for the *early* allele and six for the *late* allele.

Table 1
Sequencing information (MIxS) for Atlantic salmon blood transcriptome.

Item	Description
Classification	Animalia; Chordata; Salmoniformes; Salmonidae; *Salmo salar*
Investigation type	Eukaryote
Project name	Atlantic salmon blood transcriptome
Geographic location (latitude and longitude)	64°58′56.4"N 25°36′15.4″E
Geographic location (country and/or sea, region)	Oulujoki, Finland
Collected by	Samuel C. Andrew, Jukka-Pekka Verta
Collection date	2018-08-08
Environment (biome)	aquarium (ENVO:00002196)
Environment (feature)	fresh water aquarium (ENVO:00002198)
Environment (material)	fresh water (ENVO:00002011)
Environmental package	Water
Sequencing method	Illumina NextSeq 500, paired-end
Estimated Size	3–7 g
BioProject number	PRJNA626461
SRA accession numbers	SRR11570046, SRR11570047, SRR11570048, SRR11570049, SRR11570050, SRR11570051, SRR11570052, SRR11570053, SRR11570054, SRR11570055, SRR11570056, SRR11570057

2.3. Aligning reads and novel transcript identification

Quality assessment and filtering of reads was done with *fastp* using default settings (base phred quality ≥ Q15 is qualified and reads with greater than 40% unqualified bases are removed, Chen et al., 2018). The number of reads returned per library varied between 10 and 58 million paired-end reads (mean ± s.d. = 28 ± 15 million). With 4 to 11% of reads filtered due to low quality (mean ± s.d. = 6.44 ± 2.25%) by *fastp*. The filtered reads were then aligned to the Atlantic salmon reference genome using *STAR* (Dobin et al., 2013). To make the genome index files for STAR, the reference genome (.fna) and annotation (.gff) files used were from assembly version GCF_000233375.1 downloaded from NCBI. The annotation ".gff" file was converted to ".gtf" format using the cufflinks function gffread (Trapnell et al., 2012). In the annotation file about 500 exons did not have a gene id and they were given a gene_id by using the transcript ID, a large number of reads mapped to these exons and these reads would have been excluded from further analysis without a gene_id. For the STAR alignment we used mostly default options (including: -outSAMtype BAM SortedByCoordinate –quantMode GeneCounts –twopassMode Basic –outWigType –alignEndsProtrude 10 ConcordantPair –outFilterIntronMotifs RemoveNoncanonicalUnannotated –chimSegmentMin 1 –limitOutSJcollapsed 5,000,000).

Gene counts were taken from the "_STARReadsPerGene.out.tab" files that are generated by the option "–quantMode GeneCounts". The fourth column of the output file was used because these are the counts for stranded RNA-seq. *STAR* successfully aligned 77.78 ± 1.70% (mean ± s.d.) of reads with the remainder of reads being mapped to multiple loci (17.06 ± 1.59%), or were too short to map (2.92 ± 0.94%), and the rest were unmapped reads (1.52 ± 0.85%).

Cufflinks tools (Trapnell et al., 2012) were used to identify novel transcripts that are not included in the reference transcriptome previously published (version GCF_000233375.1). Using the default settings, the cufflinks function was used to quantify the mapped reads from STAR. Then the cuffmerge function was used to combine the transcripts from all 12 individuals and then the cuffcompare function was used to compare the merged transcriptome to the reference transcriptome. The merged blood transcriptome had 132,921 complete matches ("=" code) to the reference and 17,837 novel splice junctions ("j" code, 11.8% of total transcripts in merged transcriptome) and no other unannotated transcripts or classifications.

The haemoglobin subunit genes made up a large proportion of all expressed transcripts (mean ± s.e. = 35.0 ± 1.6%). Interestingly, the expression of the paralogs of the alpha and beta haemoglobin subunits was highly variable (Fig. S1), while the ratio of alpha to beta subunits was balanced and close to unity (mean ± s.e. = 1.06 ± 0.004).

2.4. Differentially expressed genes and links with life-history phenotype proxies

To test for DEGs between the two groups of individuals with alternative homozygote *vgll3* genotypes, two common methods were used to compare result. The first method *DESeq2* (Love et al., 2014) was performed using the default settings. For the second method *limma + voom* (Law et al., 2014), the analysis was performed on all genes with more

than 0.5 counts per million in two or more libraries and *t*-statistics were calculated with the *eBayes* function. The *DESeq2* analysis found 179 DEGs out of 20,494 genes, which were tested. The *limma + voom* analysis found no significantly DEGs out of 25,699 genes, which were tested after filtering (Table S2). Although there is no overlap in the DEGs for the two methods the relationship between the *t*-values for the two methods was very strong (Fig. S2, linear model: Estimate = 0.998, $t_{20492} = 626.2$, $R^2 = 0.95$), suggesting that the difference between the methods is due to general sensitivity in detecting small but significant expression differences.

To test for overrepresented GO terms from the list of DEGs the *topGO R* package (Alexa and Rahnenfuhrer, 2019) was used. The *topGO* analysis used only genes included in *DESeq2* analysis and had Atlantic salmon GO terms from the *AnnotationHub* package (Morgan, 2019). All DEG and GO analyses were performed in R version 3.6.1 (Core R team, 2017, supplementary data and R code files in Appendix A and B). The 179 DEGs from the *DESeq2* analysis yielded overrepresented GO terms (Table S1). The most overrepresented GO term was "structural constituent of ribosome" (GO:0003735) encompassing 13 differentially expressed genes, all of which were ribosomal proteins and most were 40S and 60S ribosomal subunits. Of these 13 genes 12 were up regulated in the *late* genotype fish. It has been observed previously that Atlantic salmon going through the parr-smolt transformation have an upregulation of ribosomal subunit genes in the liver, gill, and olfactory rosette (Robertson and McCormick, 2012). We might expect a similar up regulation of ribosomal subunit genes in salmon going through other critical phases of development. These ribosomal subunit genes would then be potential candidates for a time-series analysis of the maturation process using a targeted gene-expression assay method.

Out of the top 10 over represented GO terms there were five GO terms related to transcription, RNA and DNA binding: ("telomeric DNA binding [GO:0042162]", "DNA binding [GO:0003677]", "DNA-binding transcription factor activity, RNA polymerase II-specific [GO:0000981]", "rRNA binding [GO:0019843]"). Out of these five related GO terms, 22 out of 27 contributing genes were upregulated in the *late* genotype fish. Average age at maturation differs among Atlantic salmon genotypes. Therefore, extrapolating physiological differences between stages of maturation from gene expression profiles could be used to understand variation for maturation rates in wild populations across years and variable environmental conditions.

2.5. Multi-tissue comparison

We present two methods for comparing the similarity between transcriptomes from different Atlantic salmon tissue types. Transcriptomes for 14 Atlantic salmon tissues were downloaded from NCBI Sequence Read Archive (SRA) from the *Salmo salar* Genome (BioProject: PRJNA72713; Lien et al., 2016), these 14 tissues were from the same juvenile male except for the ovary tissue. The first method used a common PCA ordination method to visualise variation between the transcriptomes in multidimensional space with the first two dimensions explaining 39% of the variance (Fig. 1a).

The second method for visualising similarity between transcriptomes aimed to convert gene expression data into a two-dimensional layer that can be used in a GAM model to fit a thin-plate spine smoother using the interaction between the two axes (Wood, 2017). For this method genes were grouped by GO terms and the GO terms were arranged in principle component space using GO term similarity values from the *GOSim* package (Lerman and Shakhnovich, 2007) to run PCA (Fig. S3). Furthermore, we conducted pairwise comparisons between tissues by including two tissues in a model that fits separate thin-plate spline smoothers for each tissue to test how much the model is improved by fitting independent relationships for the two tissue layers. This factor for the comparison between tissues can be interpreted by a high *F*-value and significant *P*-value, meaning a significant improvement to the fit of the model due to differences between the layers (Fig. 1b). The individual

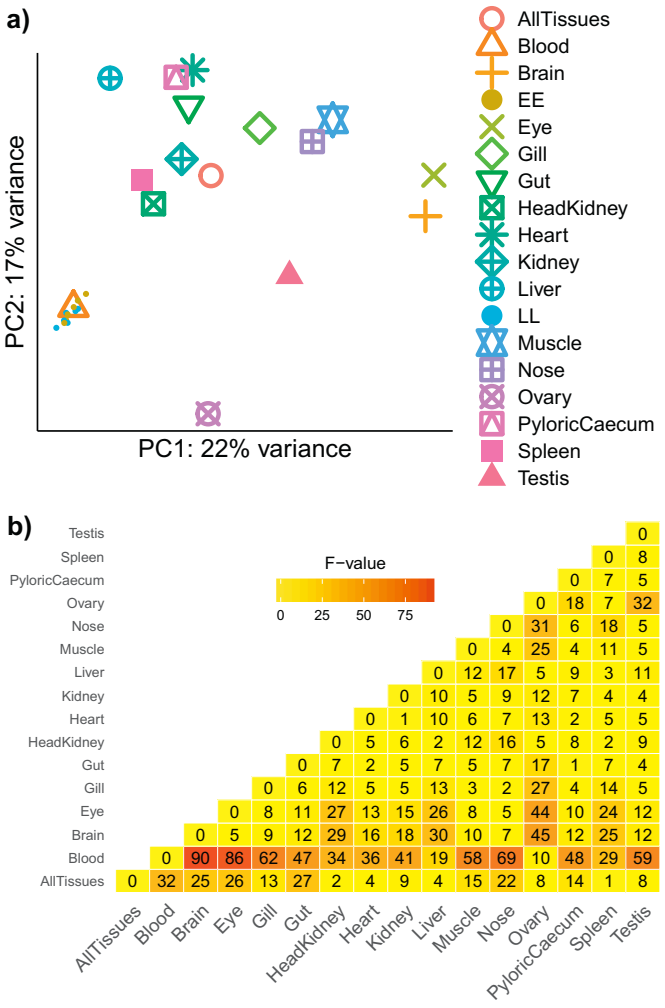

Fig. 1. Comparison of salmon transcriptomes. a) PCA plot with all tissue types and individual blood transcriptomes. The mean blood transcriptome is plotted as an orange triangle and the individual transcriptomes as small points (*early* = EE and *late* = LL in legend). The individual blood transcriptomes were projected after the PCA was fitted with all the individual tissue types. b) Pair-wise matrix of tissue comparisons using GAM method. Higher *F*-values mean a stronger difference in the fit between tissues (All *F*-values greater than 3 have significant *P*-values). These comparisons use the individual tissue layers from Fig. 2, in these layers individual gene expression values are grouped by GO terms that have been arranged by similarity.

tissue layers from the GAM approach (Fig. 2) provide a visualisation of more subtle functional differences between the transcriptomes as well as the statistical comparisons between transcriptomes unlike the PCA visualisation. The full GAM model analysis is presented in the supplementary R code file (Appendix B).

One way to test the effectiveness of the GAM method for comparing transcriptomes is to compare the F-values for pair-wise comparisons to distances between points in the PCA ordination. The relationship between the methods were moderate with Pearson correlation coefficients of 0.64 and 0.6 for 2D and 3D PCA distance respectively (Fig. S4).

The supplementary results file (Appendix C) also includes interactive heatmaps of gene expression levels for different biological process GO terms for all tissue types. For each tissue the mean expression of each GO term is divided by the median of all tissues to help visualise GO terms that have higher or lower than average gene expression levels. For most tissues the more highly expressed GO terms are close to the expression levels of the median of all tissues (values between 0.8 and 1.2) with some tissue specific GO terms being up regulated. One exception to this

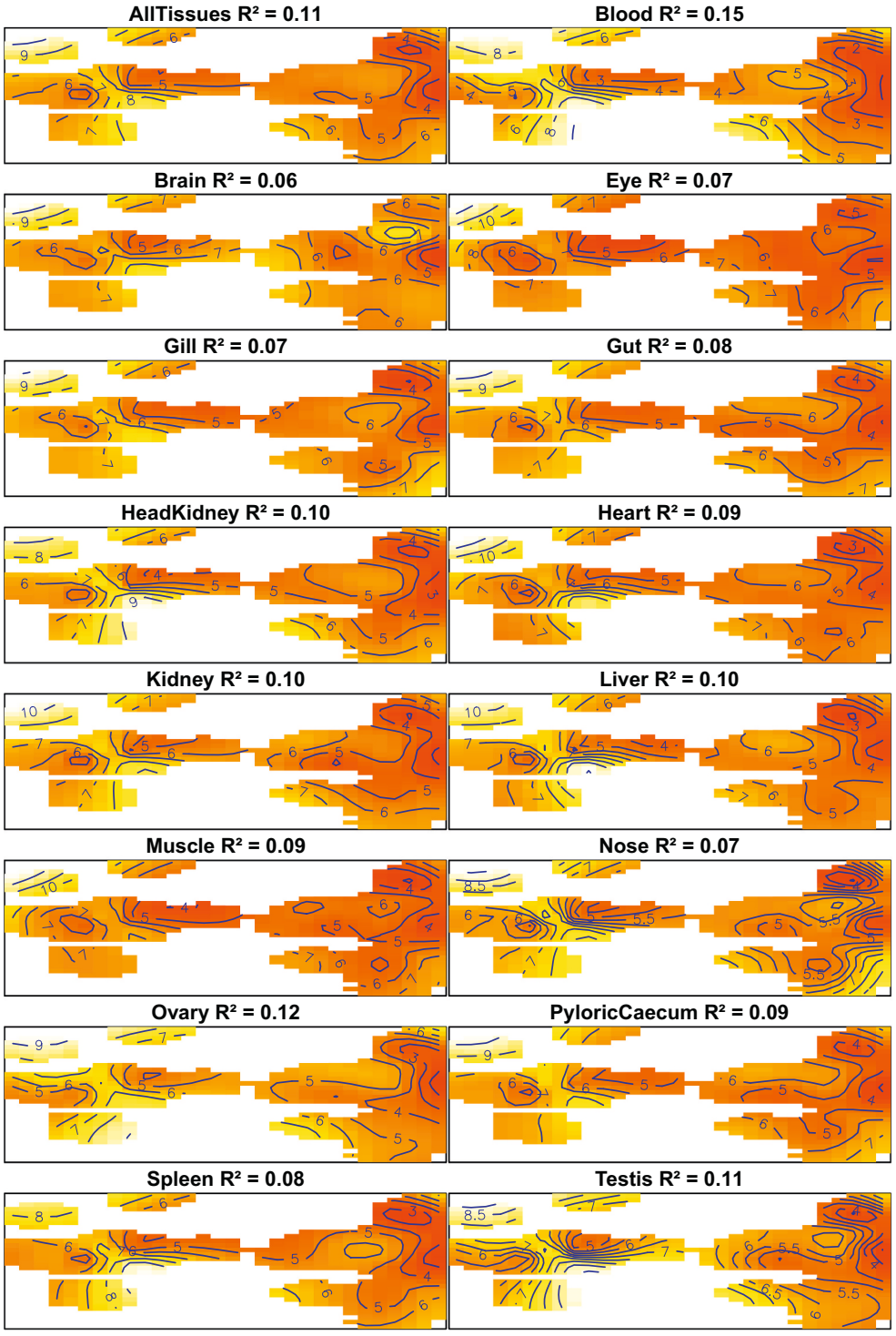

Fig. 2. Gene expression layers when genes are ordered by GO terms. These GAM models use the expression values of individual genes grouped by GO terms that are arranged by similarity. These layers are shown as contoured heat-maps of the smoother fitted using the interaction between PC1 and PC2. With these layers it is possible to see more subtle differentiation in the transcriptome profile of different tissues. This analysis includes ~12,000 genes from 263 GO terms that had 10 or more genes on the reference genome. The R^2 values show how much variance in gene expression is explained by PC1 and PC2.

pattern is the liver with many GO terms having expression that is higher than the median. Interestingly, the blood transcriptome had one of the lowest ranges of expression levels for its heatmap. The PCA and GAM comparisons also show the blood transcriptome to be quite unique to other reference transcriptomes from other tissue types, but many transcripts are still present in Atlantic salmon whole blood. The differentiation in the blood transcriptome profile could also be partly explained by the blood transcriptome being produced independently in this study, and the comparison reference transcriptomes coming from a different study (BioProject: PRJNA72713; Lien et al., 2016). It is unlikely that the confounding factors of different experimental individuals and study batch effects can be disentangled here, but we expect effects to be largest among tissue types. Hopefully our comparison to other reference transcriptomes can inspire future experimental work.

3. Data deposition

The raw RNA-seq data have been deposited in the NCBI Sequence Read Archive under BioProject number PRJNA626461, accession numbers available in Table 1 for MIxS data.

Declaration of Competing Interest

None.

Acknowledgments

We thank members of Primmer lab for assistance with the common garden experiment and support in the lab. We also thank the hatchery staff of the Natural Resources Institute Finland (Luke) Taivalkoski hatchery for assistance with the initial crosses. We would like to acknowledge the Institute of Biotechnology staff at the University of Helsinki for their rapid turnaround time to produce high quality RNA-seq data. This work was supported by the Academy of Finland (project numbers 284941, 286334, 314254, and 314255), the European Research Council under the European Articles Union's Horizon 2020 research and innovation program (grant agreement number 742312), and the University of Helsinki.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.margen.2020.100809.

References

Alexa, A., Rahnenfuhrer, J., 2019. topGO: Enrichment Analysis for Gene Ontology. R package version 2.37.0.

Ayllon, F., Kjærner-Semb, E., Furmanek, T., Wennevik, V., Solberg, M.F., Dahle, G., Taranger, G.L., Glover, K.A., Almén, M.S., Rubin, C.J., Edvardsen, R.B., Wargelius, A., 2015. The vgll3 locus controls age at maturity in wild and domesticated Atlantic Salmon (Salmo salar L.) males. PLoS Genet. 11, 1–15. https://doi.org/10.1371/journal.pgen.1005628.

Barson, N.J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G.H., Fiske, P., Jacq, C., Jensen, A.J., Johnston, S.E., Karlsson, S., Kent, M., Moen, T., Niemelä, E., Nome, T., Næsje, T.F., Orell, P., Romakkaniemi, A., Sægrov, H., Urdal, K., Erkinaro, J., Lien, S., Primmer, C.R., 2015. Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. Nature 528, 405–408. https://doi.org/10.1038/nature16062.

Bourret, V., Kent, M.P., Primmer, C.R., Vasemägi, A., Karlsson, S., Hindar, K., McGinnity, P., Verspoor, E., Bernatchez, L., Lien, S., 2013. SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (Salmo salar). Mol. Ecol. 22, 532–551. https://doi.org/10.1111/mec.12003.

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

Czorlich, Y., Aykanat, T., Erkinaro, J., Orell, P., Primmer, C.R., 2018. Rapid sex-specific evolution of age at maturity is shaped by genetic architecture in Atlantic salmon. Nat. Ecol. Evol. 2, 1800–1807. https://doi.org/10.1038/s41559-018-0681-5.

Debes, P.V., Piavchenko, N., Erkinaro, J., Primmer, C.R., 2020. Genetic growth potential, rather than phenotypic size, predicts migration phenotype in Atlantic salmon. Proc. Royal Soc. B 287, 20200867.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Law, C.W., Chen, Y., Shi, W., Smyth, G.K., 2014. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15, 1–17. https://doi.org/10.1186/gb-2014-15-2-r29.

Lerman, G., Shakhnovich, B.E., 2007. Defining functional distance using manifold embeddings of gene ontology annotations. PNAS 104, 11334–11339. https://doi.org/10.1073/pnas.0702965104.

Lien, S., Koop, B.F., Sandve, S.R., Miller, J.R., Matthew, P., Leong, J.S., Minkley, D.R., Zimin, A., Grammes, F., Grove, H., Gjuvsland, A., Walenz, B., Hermansen, R.A., Von Schalburg, K., Rondeau, E.B., Di Genova, A., Samy, J.K.A., Vik, J.O., 2016. The Atlantic salmon genome provides insights into rediploidization. Nature 533, 200–205. https://doi.org/10.1038/nature17164.

Liew, C.C., Ma, J., Tang, H.C., Zheng, R., Dempsey, A.A., 2006. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. J. Lab. Clin. Med. 147, 126–132. https://doi.org/10.1016/j.lab.2005.10.005.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 1–21. https://doi.org/10.1186/s13059-014-0550-8.

Macqueen, D.J., Primmer, C.R., Houston, R.D., Nowak, B.F., Bernatchez, L., Bergseth, S., Davidson, W.S., Gallardo-Escárate, C., Goldammer, T., Guiguen, Y., Iturra, P., Kijas, J.W., Koop, B.F., Lien, S., Maass, A., Martin, S.A.M., McGinnity, P., Montecino, M., Naish, K.A., Nichols, K.M., Ólafsson, K., Omholt, S.W., Palti, Y., Plastow, G.S., Rexroad, C.E., Rise, M.L., Ritchie, R.J., Sandve, S.R., Schulte, P.M., Tello, A., Vidal, R., Vik, J.O., Wargelius, A., Yáñez, J.M., 2017. Functional Annotation of All Salmonid Genomes (FAASG): an international initiative supporting future salmonid research, conservation and aquaculture. BMC Genomics 18, 1–9. https://doi.org/10.1186/s12864-017-3862-8.

Mohamed, A.R., King, H., Evans, B., Reverter, A., Kijas, J.W., 2018. Multi-tissue transcriptome profiling of North American derived Atlantic Salmon. Front. Genet. 9, 1–9. https://doi.org/10.3389/fgene.2018.00369.

Morgan, M., 2019. AnnotationHub: Client to Access AnnotationHub Resources. R package version 2.18.0.

Opazo, J.C., Butts, G.T., Nery, M.F., Storz, J.F., Hoffmann, F.G., 2013. Whole-genome duplication and the functional diversification of teleost fish hemoglobins. Mol. Biol. Evol. 30, 140–153. https://doi.org/10.1093/molbev/mss212.

Quinn, N.L., Boroevich, K.A., Lubieniecki, K.P., Chow, W., Davidson, E.A., Phillips, R.B., Koop, B.F., Davidson, W.S., 2010. Genomic organization and evolution of the Atlantic salmon hemoglobin repertoire. BMC Genomics 11, 1–17. https://doi.org/10.1186/1471-2164-11-539.

R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Robertson, L.S., McCormick, S.D., 2012. Transcriptional profiling of the parr-smolt transformation in Atlantic salmon. Comp. Biochem. Physiol. 7, 351–360. https://doi.org/10.1016/j.cbd.2012.07.003.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., Pachter, L., 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. Nat. Protoc. 7, 562–578. https://doi.org/10.1038/nprot.2012.016.

Warren, I.A., Ciborowski, K.L., Casadei, E., Hazlerigg, D.G., Martin, S., Jordan, W.C., Sumner, S., 2014. Extensive local gene duplication and functional divergence among paralogs in Atlantic Salmon. Genome Biol. Evol. 6, 1790–1805. https://doi.org/10.1093/gbe/evu131.

Wood, S.N., 2017. Generalized Additive Models: An Introduction with R, 2nd Ed. Chapman & Hall.