Received: 27 June 2020 | Revised: 15 October 2020 | Accepted: 6 November 2020

DOI: 10.1111/mec.15743

FROM THE COVER

MOLECULAR ECOLOGY WILEY

The strength and form of natural selection on transcript abundance in the wild

Freed Ahmad[1] | Paul V. Debes[2] | Ilkka Nousiainen[3] | Siim Kahar[1,3] | Lilian Pukk[3] | Riho Gross[3] | Mikhail Ozerov[1,4] | Anti Vasemägi[3,4]

[1]Department of Biology, University of Turku, Turku, Finland

[2]Department of Aquaculture and Fish Biology, Hólar University, Sauðárkrókur, Iceland

[3]Department of Aquaculture, Institute of Veterinary Medicine and Animal Sciences, Estonian University of Life Sciences, Tartu, Estonia

[4]Department of Aquatic Resources, Swedish University of Agricultural Sciences, Drottningholm, Sweden

Correspondence
Anti Vasemägi, Department of Aquatic Resources, Swedish University of Agricultural Sciences, 17893 Drottningholm, Stångholmsvägen 2, Sweden.
Email: anti.vasemagi@slu.se

Funding information
Estonian Ministry of Education and Research, Grant/Award Number: IUT8-2; Eesti Teadusagentuur, Grant/Award Number: PRG852; Academy of Finland, Grant/Award Number: 266321; Deutsche Forschungsgemeinschaft, Grant/Award Number: DE 2405/1–1; Ella & Georg Ehrnrooth foundation; Sihtasutus Archimedes

Abstract

Gene transcription variation is known to contribute to disease susceptibility and adaptation, but we currently know very little about how contemporary natural selection shapes transcript abundance. Here, we propose a novel analytical framework to quantify the strength and form of ongoing natural selection at the transcriptome level in a wild vertebrate. We estimated selection on transcript abundance in a cohort of a wild salmonid fish (*Salmo trutta*) affected by an extracellular myxozoan parasite (*Tetracapsuloides bryosalmonae*) through mark–recapture field sampling and the integration of RNA-sequencing with classical regression-based selection analysis. We show, based on fin transcriptomes of the host, that infection by the parasite and subsequent host survival is linked to upregulation of mitotic cell cycle process. We also detect a widespread signal of disruptive selection on transcripts linked to host immune defence, host–pathogen interactions, cellular repair and maintenance. Our results provide insights into how selection can be measured at the transcriptome level to dissect the molecular mechanisms of contemporary evolution driven by climate change and emerging anthropogenic threats. We anticipate that the approach described here will enable critical information on the molecular processes and targets of natural selection to be obtained in real time.

KEYWORDS

climate change, contemporary natural selection, gene expression, host–parasite relationships, selection differential and gradient

1 | INTRODUCTION

Understanding how natural selection acts on traits and eventually on organisms represents a fundamental challenge in biology (Mayr, 1982). Using a now classical regression-based approach (Lande & Arnold, 1983), ecologists have generated thousands of phenotypic selection estimates over the past 35 years; these estimates help to understand the contemporary selection processes in nature and enable comparisons of the strength and mode of selection across traits and species (Kingsolver et al., 2001; Kingsolver & Pfennig, 2007; Siepielski et al., 2017). However, despite this wealth of phenotypic selection estimates and a large number of studies that indirectly infer the roles of different evolutionary forces in shaping gene expression patterns (Fraser et al., 2010; Gilad et al., 2006), we know very little about how natural selection affects transcript abundance in the wild (Miller et al., 2011). This is remarkable given that

variation in transcript abundance is of central importance to evolution (Emilsson et al., 2008; Fraser, 2013; Fraser et al., 2010; Gilad et al., 2006; Miller et al., 2011), linking molecular functions to performance and Darwinian fitness.

Here, we present an integrative approach investigating how contemporary natural selection shapes transcriptomic variation by combining analyses of selection differentials and gradients (Lande & Arnold, 1983) with the high-throughput screening of molecular phenotypes at the gene transcription level. Such use of the so-called molecular phenotypes has been highly successful in medical science for discovering the mechanisms underlying complex human diseases (e.g.,Chaussabel et al., 2008; Cobb et al., 2005), but we currently know very little about how within-generation natural selection in the wild translates to changes at the RNA and protein levels (Husak, 2016). However, regression-based and distributional selection differentials and gradients (Henshaw & Zemel, 2016; Lande & Arnold, 1983), which measure the effect of a trait on relative fitness in standard deviation trait units, can be used to estimate the form and strength of contemporary natural selection on any quantitative trait, including transcript abundances, allowing direct comparisons among traits, populations and species (Lande & Arnold, 1983).

We focus on a host–parasite system consisting of brown trout (*Salmo trutta*) as the host and a myxozoan parasite (*Tetracapsuloides bryosalmonae*), the causative agent of temperature-dependent proliferative kidney disease (PKD) in salmonid fishes (Okamura et al., 2011). Recent work has demonstrated that *T. bryosalmonae* is widespread in Europe and North America (Dash & Vasemägi, 2014; Debes et al., 2017; Mo & Jørgensen, 2016; Skovgaard & Buchmann, 2012; Vasemägi et al., 2017). At elevated temperatures (>15°C–18°C), this parasite causes high mortality in wild and farmed salmonids (Hari et al., 2006; Hedrick et al., 1993; Tops et al., 2006). The parasite has a complex two-stage life cycle in which freshwater bryozoans and salmonid fishes are consecutive hosts (Okamura et al., 2011). Mass release of *T. bryosalmonae* spores from bryozoans occurs from spring to early summer (Hedrick et al., 1993), resulting in the synchronized infection of young-of-the-year fish through gills and/or skin (Longshaw et al., 2002). Inside the salmonid host, the parasite multiplies and induces an inflammatory response and tumour-like chronic lymphoid hyperplasia in the kidney (Bettge et al., 2009; Hedrick et al., 1993). The impairment to the kidney, the major organ responsible for haematopoiesis in fish, results in anaemia (Clifton-Hadley et al., 1987; Hedrick et al., 1993), which decreases oxygen transportation capacity, lowering the maximum metabolic rate and upper thermal tolerance (Bruneaux et al., 2016). The reduction in aerobic and renal capacity, combined with decreased oxygen solubility and increased oxygen demand at higher temperatures, makes PKD a serious threat to cold-water salmonid populations in regions affected by warm summers, which are expected to become more frequent under global warming (Okamura et al., 2011). Compared to many other host–parasite systems, brown trout and *T. bryosalmonae* represents a highly suitable model for studying contemporary natural selection on host gene expression in the wild because the parasite shows a high prevalence (Hedrick et al., 1993) and imposes a strong temperature-dependent effect on host physiology, performance (Bruneaux et al., 2016) and survival (Hedrick et al., 1993). Furthermore, many challenges associated with field data, such as differences in host age, infection onset and conspecific co-infection dynamics (Bishop et al., 2012; Doeschl-Wilson et al., 2012) or host exposure avoidance (Graham et al., 2010), are minimal or absent.

To quantify the strength and form of within-generation selection on transcript abundance, we collected small fin biopsies from wild juvenile trout in August, when we expected all individuals to be infected, for transcriptome and multilocus fingerprint profiling, after which they were released back into their native environment. Approximately 1 month later, after the anticipated period of parasite-associated mortality, we recaptured and identified survivors based on multilocus genotype information and tested whether fin-tissue transcript abundances measured in August correlate with survival. To further elucidate the transcriptional signatures linked with the observed mortality, we also measured the *T. bryosalmonae* load in kidney tissue among survivors to identify transcripts and protein–protein interaction (PPI) networks associated with both survival and parasite load (PL).

2 | MATERIALS AND METHODS

2.1 | Study population, temporal abundance and temperature records

The coastal river Altja (length 17.6 km, catchment area 46.1 km^2) flows into the Gulf of Finland in the Baltic Sea (Figure S1) and supports a small wild anadromous brown trout population with a high prevalence of *Tetracapsuloides bryosalmonae* (Dash & Vasemägi, 2014). Records on young-of-the-year trout abundance in the river Altja across 13 years (2005–2017) were obtained from the Estonian Ministry of Environment Fisheries Monitoring Program. River water temperature was measured twice per day (at noon and midnight) over two years (2014 and 2015) using automatic temperature loggers (HOBO 8K Pedant Temperature/Alarm Data Logger, Onset Computer Corporation). Mean monthly air temperature records over a 73-year period (1945–2017, Kunda Coastal Meteorological station, 59°31′17″N, 26°32′29″E, 25 km from river Altja) were obtained from the Estonian Weather Service (Environmental Agency). The studied population showed a strong negative correlation between mean summer air temperature and young-of-the-year density as well as pervasive temperature dependence of disease severity, consistent with experimental work (Bettge et al., 2009) (Figure 1).

2.2 | Field sampling, phenotyping and genetic mark–recapture analysis

On 30 August 2015, we electrofished 278 young-of-the-year trout in the river Altja from the same five areas along a 330-m stretch that were sampled in 2014 (Table S1, Figure S1, area

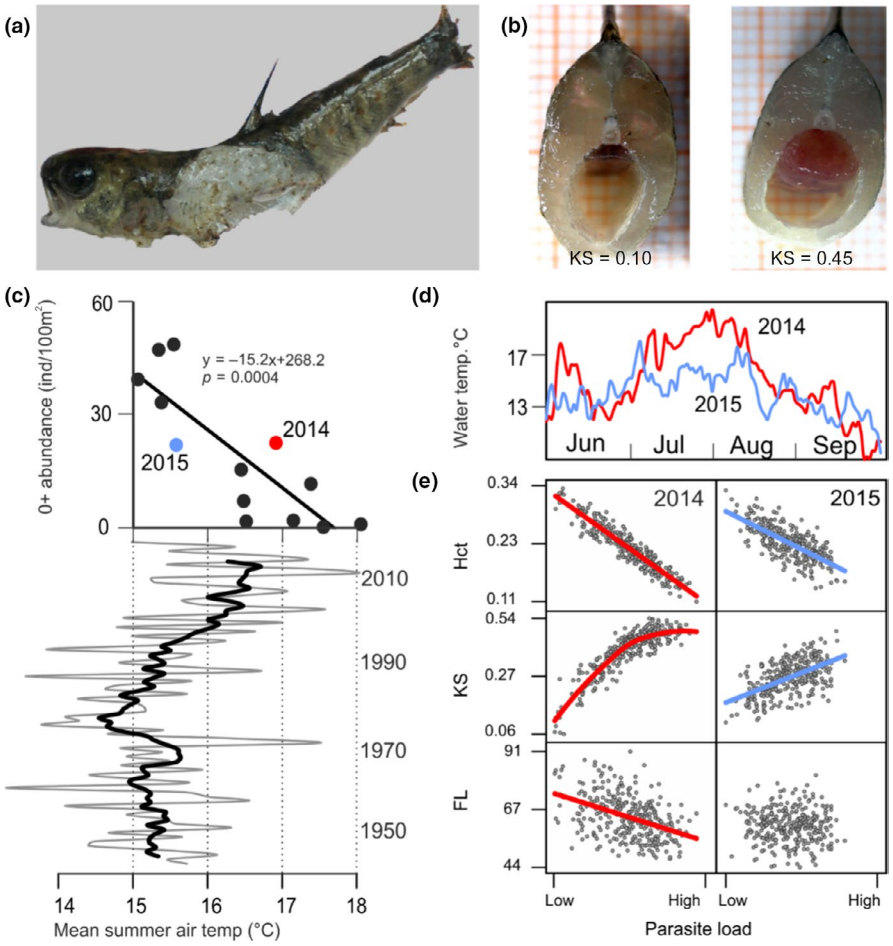

FIGURE 1 Temperature dependence of PKD in wild trout. (a) Dead young-of-the-year brown trout found in the Altja river with putative PKD-associated death symptoms (swollen kidney, a widely opened mouth and flared gills suggestive of anaemia). (b) Body section of trout with normal (left) and swollen (right) kidney. (c) Effect of temperature on juvenile trout abundance during 2005–2017 in the Altja river in relation to average summer air temperature (7-year moving average mean summer air temperature over 73 years is highlighted in bold). (d) Water temperature variation over a 4-month period in 2014 (red) and 2015 (blue) in the Altja river. (e) Relationships between parasite load (PL) and fork length (FL), kidney swollenness (KS) and haematocrit (Hct) in 2014 and 2015. All plotted relationships (model-based regression lines; individual points based on the model output) are significant ($p < .001$), except FL versus PL in 2015 ($p = .933$)

IDs 1–5). Individuals were anaesthetized with buffered MS-222 (SigmaAldrich) and measured for fork length (±1 mm) as a measure of body size. After small biopsies of the right pelvic fin tissue for genetic mark–recapture analysis and 3′ RNA sequencing (see below), we released the recovered trout within their original capture area. As fins regenerate in teleost fishes, a small fin biopsy is unlikely to impair fish survival (Gjerde & Refstie, 1988). Biopsies were immediately frozen in liquid nitrogen and stored at −80°C. We used a fin subsample for DNA analysis and individual identification (see below). Approximately 1 month after initial electrofishing (22–27 September), we caught 685 young-of-the-year trout along a 780-m stretch that included the initial 330-m stretch (Table S1). The five initial catch areas (area IDs 1–5) were electrofished three times (Table S1), and we estimated the capture probability and total number of fish using the depletion method (Zippin, 1958) implemented in the FSA (Fisheries Stock Assessment) package version 0.8.17 (Ogle, 2017) in R version 3.3.3 (R Core Team, 2017). A high recapture probability in all areas (average catchability 0.65, 95% confidence interval [CI] 0.60–0.70) combined with low inferred fish dispersal (based on electrofishing of the extended areas up- and downstream; Table S1) indicated that only a few survivors were not recaptured in September (n = 13.9, 95% CI 7.2–23.1). Among the 685 fish caught in September, we killed 363 via MS-222 overdose. The sampling procedure, microsatellite genotyping, measurement of phenotypic traits (fork length [FL] as a measure of body size; haematocrit [Hct]; kidney-to-muscle ratio as a measure of kidney swollenness [KS]) and quantification of PL were carried out as previously described (Debes et al., 2017). The relationships between PL and PKD-linked phenotypic traits (Hct, KS, FL) were analysed using general linear mixed models in ASREML-R 3.0 (Butler et al., 2009). To control for genetic variation in the expression of the traits, the models accounted for the genetic relationship matrix of the individuals (A) via linking (A^{-1}) to random individual effects. The relationship matrix was obtained by genotyping the sampled individuals (see next section) and then reconstructing their parentage using COLONY version 2.0.6.5 (Jones & Wang, 2010).

2.3 | Microsatellite analysis and identification of individual recaptures

We genotyped individuals using 14 highly variable microsatellite loci as previously described (Debes et al., 2017). Individuals with at least 10 successfully genotyped loci were included in the analysis. Recaptured individuals were defined as having identical genotypes with at most one allele mismatch (at least a 95% match when 10 loci were genotyped) using MICROSATELLITE TOOLKIT (Stephen D. E. Park, Trinity College, Dublin, Ireland). To estimate genotyping error

rates, we amplified and genotyped 440 randomly selected samples twice, which indicated low error rates (mean allelic dropout rate: 0.0107, range 0.0013–0.0292; mean false allele rate: 0.0027, range 0.0010–0.0176).

2.4 | Quantitative real-time polymerase chain reaction (qPCR)

PL was determined from kidney tissues collected in September 2015 by qPCR using previously described protocols (Debes et al., 2017). For each sample, we quantified two DNA sequences per run: a 166-bp fragment of *T. bryosalmonae* 18S rDNA sequence (GenBank accession U70623) and 74-bp fragment of the *Salmo salar* nuclear DNA sequence (GenBank accession BT049744.1). Simultaneous quantification of both DNA fragments enabled us to standardize the amount of parasite DNA relative to brown trout DNA. We ran 10 plates (384-well format) on the QuantStudioTM 12K Flex Real-Time PCR System (Thermo Fisher Scientific). Each 10 µl reaction contained 200 nm of each primer, 1 × HOT FIREPol EvaGreen qPCR Supermix master mix (Solis BioDyne) and 2.5 µl of template DNA (10 ng/µl). Each sample was run in quadruplicate per gene and included four nontemplate controls per gene to detect possible contamination. To determine the quantification cycle (Cq), we used the online tool REAL-TIME PCR MINER (Zhao & Fernald, 2005). PL was defined as the difference between the Cq values (on the $\log_2$ scale) of the two target genes ($Cq_{S.\ trutta}$–$Cq_{T.\ bryosalmonae}$, lower values indicate low PL). Our 2015 qPCR results were calibrated to 2014 results using 10 2014 samples that we repeated along with the 2015 samples (linear regression, PL2014 = −0.193 + 1.031 × PL2015). To estimate technical bias among plates, we included a $\log_{10}$ dilution series (50, 5, 0.5, 0.1 and 0.05 ng/µl) from a reference sample in quadruplicate per gene on every plate. Standardized amplification efficiency was calculated among plates, using plate- and gene-specific efficiencies estimated from the Cq versus $\log_{10}$-dilution slopes (Debes et al., 2017). Subsequently, we fitted a linear mixed model to estimate PL for each individual that accounted for technical bias among plates and wells (Debes et al., 2017).

2.5 | RNA isolation and library preparation for Illumina sequencing

Total RNA was successfully extracted from pelvic fin tissue of 238 individuals (85.6%) out of 278 collected in August 2015 (i.e., survivors and nonsurvivors) using the NucleoSpin RNA kit (Macherey-Nagel) and ensuring RNA quality using the Agilent 2100 Bioanalyzer. We prepared a barcoded library using Lexogen QuantSeq 3′ mRNA-Seq Library Prep Kit FWD for Illumina according to the manufacturer's recommendations (Lexogen). QuantSeq provides an efficient and cost-effective protocol for generation of strand-specific next-generation sequencing libraries close to the 3′ end of polyadenylated RNAs (Moll et al., 2014). This approach is analogous to other tag-based RNA sequencing (RNAseq) approaches, such as TagSeq (Meyer et al., 2011), which have been shown to generate more accurate estimates of transcript abundances than standard RNAseq with a fraction of the sequencing effort (Lohman et al., 2016). We inspected all barcoded libraries for quality with FRAGMENT ANALYZER (Advanced Analytical, AATI) using the High Sensitivity NGS Fragment Analysis Kit. The three pooled barcoded libraries, consisting of 64, 91 and 96 individuals, were single-end sequenced using an Illumina HiSeq2500 in 14 lanes. For the first two pooled libraries, we generated 125-bp-long reads in two lanes. For the remaining 12 lanes, we generated 100-bp reads. Overall, we obtained 2.21 billion raw reads, with 1.5–34.6 million reads per individual (median 8.9 million reads). In addition, conventional Illumina mRNA paired-end sequencing (100 bp) was carried out for two fin-clip mRNA pools both consisting of seven individuals from the River Altja (29.5 and 25.9 million reads). The data have been deposited with links to BioProject accession no. PRJNA517427 in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/). The library preparation for this was done according to the Illumina TruSeq Stranded mRNA Sample Preparation Guide. For adapter trimming and read preprocessing, we used TRIMMOMATIC (version 0.35; Bolger et al., 2014) (parameters: ILLUMINACLIP = TruSeq3-SE.fa; HEADCROP = 10; SLIDINGWINDOW = 4:20; LEADING = 5; TRAILING = 5; MINLEN = 40). A total of 44.6 million quality-controlled paired-end reads were retained (23.8 and 20.8 million reads per pool). For the QuantSeq, we used TRIMMOMATIC with slightly different settings (HEADCROP = 12 and MINLEN = 70). We subsequently used CUTADAPT version 1.10 (Martin, 2011) to inspect and trim longer runs of poly-As at the end of the QuantSeq reads (parameters: q = 10; b = A{20}; b = A{30}; m = 40) and discarded sequences shorter than 40 bp. We assessed the quality of reads before and after trimming using FASTQC (version 0.11.2; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). A total of 1.1–27.1 million QuantSeq reads per sample remained after quality filtering (median 6.8 million reads).

2.6 | Trout fin-specific splice sites, reference genome modifications and mapping

To identify brown trout fin-specific splice sites, we mapped quality-filtered paired-end reads from two pooled fin libraries (total 44.5 million paired and 9.1 million unpaired reads) and three 3′ mRNA-Seq samples with the highest read depth (total 58.4 million reads) to the Atlantic salmon genome (GCF_000233375.1) (Lien et al., 2016) using spliced aligner HISAT2 (version 2.1.0 (Kim et al., 2015). Subsequently, the resulting spliced alignment and the salmon genome annotation file were used as an input for STRINGTIE (Pertea et al., 2015) to assemble full-length transcripts expressed in the trout fins. The splice-site locations of the STRINGTIE output were extracted using the extract_splice_sites.py script provided with HISAT2. Furthermore, single nucleotide variants were called from the pooled fin paired-end alignment using mpileup in SAMTOOLS (Li et al., 2009). The Atlantic salmon genome was modified with the alternative alleles. Finally, all

quality-controlled reads from QuantSeq 3′ mRNA-Seq were splice aligned to the modified reference genome using HISAT2 with trout fin-specific splice sites as a guide.

2.7 | Estimation of transcript abundance and batch effect correction

All alignment data were loaded into R as the RangedSummarizedExperiment object returned by the summarizeOverlaps function available in the R package GENOMICALIGNMENTS (Lawrence et al., 2013). The original salmon-genome-annotation GFF file was used to dissect exons on the basis of gene information, and union mode was selected for assigning the reads within the exons while considering strand information. Read counts from separate lanes, runs and replicate files were summed to individual counts using collapseReplicates implemented in the R package DESEQ2 (Love et al., 2014). The resulting data object contained a raw read count matrix and phenotypes for each sample. For subsequent analysis, we selected only protein-coding, nuclear genes with an average of >10 reads per individual. To make the gene expression data compatible for linear modelling, the raw read counts were converted into quantile-normalized $\log_2$-counts per million (logCPM) using the voom method (Law et al., 2014) implemented in the LIMMA package (Ritchie et al., 2015). Pooled library batch effects were removed by employing the ComBat function implemented in the SVA package (Leek et al., 2012), and corrected gene counts were used in differential gene expression analysis.

2.8 | Differential expression (DE)

To describe the relationships between the continuous phenotype (PL) and transcript abundance, generalized linear models were fitted using the glm function available in R. Each gene in the corrected gene count matrix was used as a predictor against the phenotype (response variable) assuming the normal distribution for both.

Q-values were calculated using the QVALUE package implemented in R. To identify genes associated with survival, we used an iterative random forest (RF) classification approach using the RANGER (Wright & Ziegler, 2017) R package. The corrected gene count matrix and survival status (dependent variable) were used as an input for the classification. After each RF iteration (100,000 trees), genes with permutation importance value <0 were eliminated for the next iteration. The iterations ended when all the genes in the input matrix have permutation importance values ≥0. After 64 iterations, a final set of 1270 genes classified individuals into survivors and nonsurvivors with a 16% error rate. RF misclassified 25 (10.5%) recaptured individuals as nonsurvivors, and 13 (5.5%) uncaptured individuals as survivors. While misclassification of the recaptured individuals may reflect their poor physiological status, it is likely that some proportion of uncaptured individuals survived. This was further supported by mark–recapture analysis, which indicated that a small number (n = 13.9, 95% CI 7.2–23.1) of surviving fish that were marked in August were not recaptured in September (Table S1). For subsequent analysis, we therefore treated the 13 putatively uncaptured individuals as survivors based on their transcript profiles (Table S4), but the main findings remained unchanged irrespective of the classification (Figures S2–S5). For example, we observed considerable overlap (n = 171) among the top 416 genes (Figure S2) between the top lists of observed survivors and corrected survivors, both of which showed highly significant enrichment for mitotic cell cycle genes (GO:0000278, Figure 2; Figure S3). Furthermore, DE analysis based on uncorrected survival produced a hill-shaped, rather than uniform, p-value distribution, indicating that misclassification of individuals probably resulted in violation of the statistical test assumptions (Figure S2). Thus, corrected survival status was used in the DE analysis using DESEQ2. First, the raw read counts, library size factors and dispersion were estimated using estimateSizeFactors and estimateDispersions, respectively and then the differential gene expression was performed using nbinomWaldTest along with the three pooled library IDs as a covariate. However, given that this work primarily aimed to generate new hypotheses rather than validate earlier findings and focused on pathways rather than single transcript detection, we adopted a relatively relaxed significance threshold (unadjusted $p < .01$) for DE and survival analysis.

2.9 | Discriminant analysis of principal components (DAPC)

We performed DAPC on the corrected gene expression matrix using the dapc function implemented in the R package ADEGENET (Jombart, 2008; Jombart et al., 2010). As the DAPC function requires categorical data, PL values were split into three groups: low (PL < −1.27, n = 24), intermediate (PL = −1.27 to 0.58, n = 39) and high (PL > 0.58, n = 48) using the cut command in R.

2.10 | Weighted gene co-expression network analysis (WGCNA)

Genes associated with survival were subjected to automatic network construction and module detection using the blockwiseModules function implemented in the R package WGCNA (Langfelder & Horvath, 2008). The cut-off for the minimum scale-free topology-fitting index was set to 0.8 (power = 7), and we used biweight midcorrelation (bicor) to estimate correlations (other parameters: networkType = "signed," minKMEtoStay = 0.2). For the analysis of quadratic selection differentials (see below), similar parameters were used (except cut-off for the minimum scale-free topology fitting index = 0.61 [power = 6]).

2.11 | Gene Ontology (GO) and protein–protein interaction network analysis

Atlantic salmon orthologue gene symbols, entrez IDs, and descriptions in humans (86.8%), zebrafish (3.6%) or other organisms (9.6%)

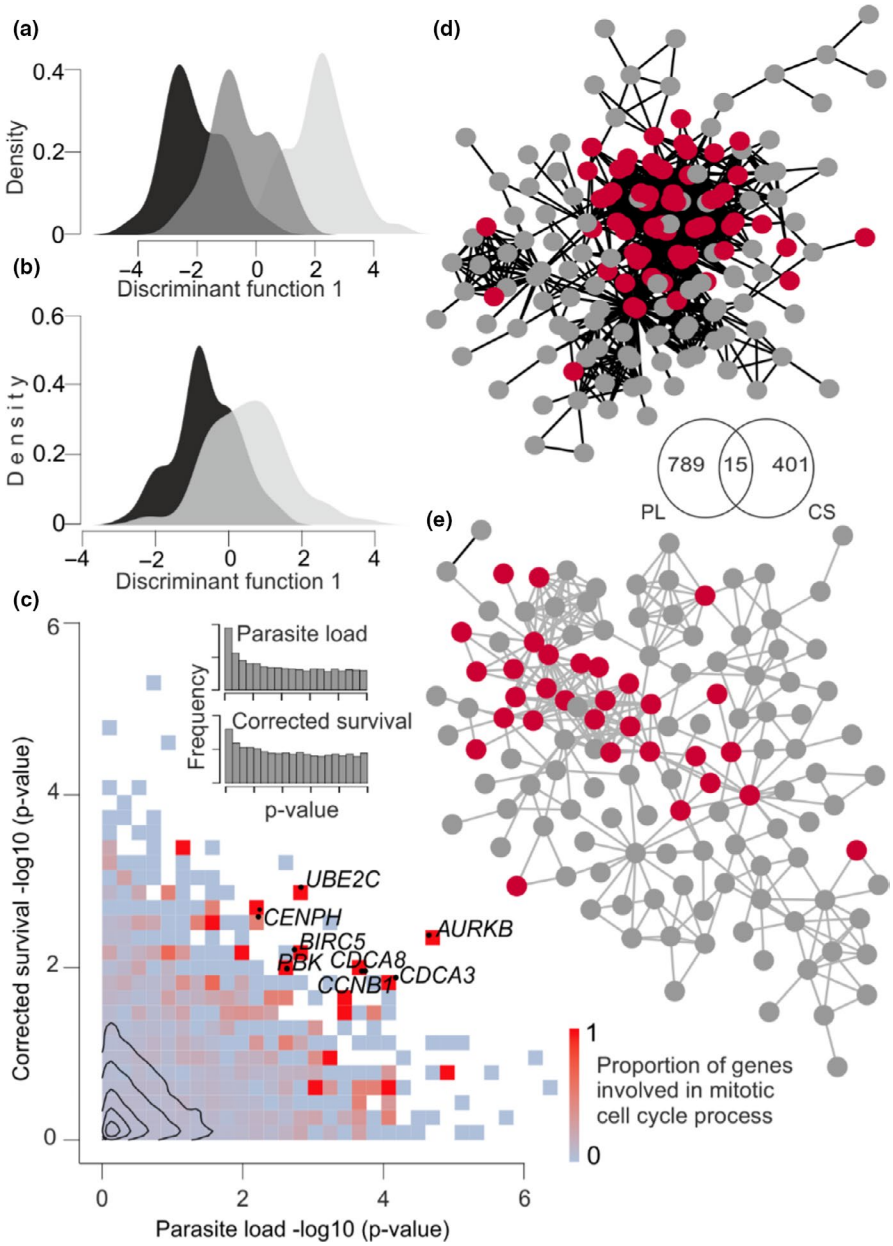

FIGURE 2 Transcriptome responses in relation to parasite load and corrected survival. (a) Density distribution of the first discriminant scores corresponding to low, intermediate and high PLs (black, dark grey and light grey areas, respectively). (b) Density distribution of the first discriminant scores corresponding to survivors and nonsurvivors (light grey and black areas, respectively). (c) Proportion of genes involved in mitotic cell cycle presented as a heatmap. The inserted histograms reflect excess transcripts associated with PL and survival. Contours reflect the density of individual transcripts. (d,e) Protein–protein interaction (PPI) network with transcripts positively correlated with PL (d) and survival (e). Mitotic cell cycle genes (GO:0000278) within the PPI networks are shown as red circles. The overlap between parasite load (PL) and survival (CS) are shown in the inserted Venn diagram

were searched using complete gene names in NCBI using the RENTREZ (Winter, 2017) package in R. GO-enrichment analysis was performed using STRING-DB (Szklarczyk et al., 2015) and GORILLA (Eden et al., 2009), in which all orthologue gene symbols were used as a background list. For STRING-DB PPI analysis, we used single lists of gene symbols against human protein references (minimum interaction score: 0.70; text mining disabled).

2.12 | Quantification of linear and nonlinear selection differentials

We estimated linear and nonlinear (i.e., quadratic) selection differentials for each of the 18,717 gene transcripts quantified in 238 individuals based on both corrected and uncorrected survival (binary status, nonsurvivor = 0, survivor = 1; Table S2). Subsequently, estimates of linear or quadratic selection differentials were computed using generalized linear models under the glm function in R. These models used logit-link functions and binomial error distributions for the binary survival response and the predictor of the mean-centred and variance-scaled (mean = 0, *SD* = 1) transcript levels (linear differentials) or added a quadratic scaled transcript-level predictor (nonlinear differentials). To transform the logistic regression model coefficients to selection differentials on the relative fitness scale that is meaningful to microevolutionary studies, we used the method suggested by Janzen and Stern (1998). The *p*-values for each selection differential were estimated using *t* tests that were constructed based on logistic regression coefficients, their standard errors and model residual degrees of freedom. In addition, to calculate the distribution of linear and nonlinear selection differentials when no selection is acting on transcripts, we randomized the survival values (1000 permutations) and estimated selection differentials as described above.

To compare the strength of directional and nondirectional selection, we used a recently developed unified measure, the distributional selection differential (DSD) (Henshaw & Zemel, 2016) for standardized trait values (mean = 0, $SD = 1$). This enabled us to use a single framework to partition total selection (DSD) into selection on the trait mean (dD) and selection on the shape of the trait distribution (dN).

2.13 | Quantification of linear selection gradients

The linear selection gradients for the DE genes were reconstructed from a subset of principal components (Chong et al., 2018), as this approach not only enables the multicollinearity between the predictors to be handled but is also suitable for cases in which the number of predictors exceeds the number of observations. For this purpose, the principle components were calculated from the correlation matrix of the standardized values of 416 DE genes, and the linear selection gradients were subsequently computed for the first 55 PCs (explaining 76% of the variation) with the glm function in R, using the logit-link function and binomial error distribution. The eigenvectors from the original 55 PCs were then matrix multiplied with the resulting linear selection gradients to reconstruct the selection gradients for individual genes (Chong et al., 2018). Similarly, the selection gradients for 416 DE genes were calculated by including FL as a predictor. The standard errors were reconstituted for these gradients as described by Chong et al. (2018). The *t*-statistic was calculated by dividing the gradients by the standard errors, and the *p*-values were estimated from the results using 237 degrees of freedom. The *p*-values were corrected for the FDR using the p.adjust function in R.

3 | RESULTS

3.1 | Parasite load, survival and transcript abundance

Among 18,717 host genes expressed in pelvic fin tissue, 804 covaried with PL quantified in kidney tissue 1 month later (unadjusted $p < .01$, FDR < 0.19; Figure 2c; Table S2). These results indicate that among the top 804 transcripts, ~650 probably represent true positives showing a genuine association between transcript abundance and PL. Consistent with the linear regression results, DAPC (Jombart et al., 2010) identified a host transcriptome signature predictive of PL (Figure 2b). GO analysis revealed that the genes positively correlated with PL represent a nonrandom set of genes showing enrichment for 59 GO terms (GORILLA, Eden et al., 2009; FDR < 0.05; Table S3), with the top three (FDR $< 7.7 \times 10^{-6}$) biological processes of cell division (GO:0051301, $n = 41$), mitotic cell cycle process (GO:1903047, $n = 48$) and cell cycle process (GO:0022402, $n = 57$).

Fish survival was predicted with 84% accuracy based on the transcription profiles for 1,270 genes using RF analysis. Similar to analysis of PL, both DE and DAPC analyses revealed a gene expression signature that covaried with survival ($n = 416$ genes, unadjusted $p < .01$, FDR < 0.45; Figure 2b and c; Table S2). These results suggest that among the top 416 transcripts, ~229 transcripts probably represented true positives showing genuine associations between transcript abundance and survival.

3.2 | Potential links between survival and parasite load

PPI network analysis using STRING-DB indicated that both survival-associated transcripts (PPI enrichment, $p < .001$) and transcripts positively correlated with PL (PPI enrichment, $p < 1.0 \times 10^{-16}$) were highly enriched for genes involved in the mitotic cell cycle (GO:0000278; FDR $= 7.45 \times 10^{-7}$ and 3.87×10^{-24}, respectively; Figure 2d and e). Among the genes associated with both survival and PL are several known oncogenes and tumour suppressors (e.g., *AURKB*, *BTG1*, *UBE2C*, *BIRC5*, *EEF2K*; Figure 2c). At the same time, survival was not dependent on fish size (Welch's two-sample *t* test, $n = 278$, $p = .216$; Wilcoxon's rank sum test, $p = .190$).

To further explore the relationships between PL and survival at the transcriptome level, we carried out WGCNA (Langfelder & Horvath, 2008). The survival-associated genes clustered into seven gene co-expression networks (Figure 3a), which included two modules that correlated with PL (Figure 3b). One particular module consisting of 27 genes (depicted in red in Figure 3b) showed strong enrichment for the mitotic cell cycle (GORILLA, FDR $= 5.9 \times 10^{-5}$; PPI enrichment $p < 1.0 \times 10^{-16}$), similar to the results from individual transcript analysis. Within the red module, the survival-linked genes that showed the highest correlations with PL were *AURKB*, *UBE2C*, *BIRC5* and *CENPH*, which are known key regulators of the mitotic cell cycle (Figure 3c and d).

3.3 | Linear selection differentials and gradients

To quantify the strength and form of contemporary natural selection on transcript abundance, we first estimated standardized linear (s) and quadratic selection differentials (λ) for 18,717 transcripts. Selection differentials quantify selection (both direct and indirect) on a trait in terms of the effects of trait values on relative fitness in units of standard deviations of the trait, allowing direct comparisons among traits, populations and species (Lande & Arnold, 1983). We compared their magnitudes to a large data set of phenotypic selection estimates based on a variety of traits and taxa (1,834 published estimates of s) (Siepielski et al., 2017). The vast majority of s values, which measure the change in a population's mean trait value before and after selection, were small (median($|s|$) = 0.047; 95% values of s between −0.132 and 0.129, $\widehat{S} \pm SE = -0.0011 \pm 0.0005$, $t_{18716} = -2.14$, $p = .033$), whereas the coregulated gene

associated with survival showed larger values of s (Figure 4a). Similar results were obtained for the linear differential estimates calculated using both uncorrected and corrected survival information (Figure S4).

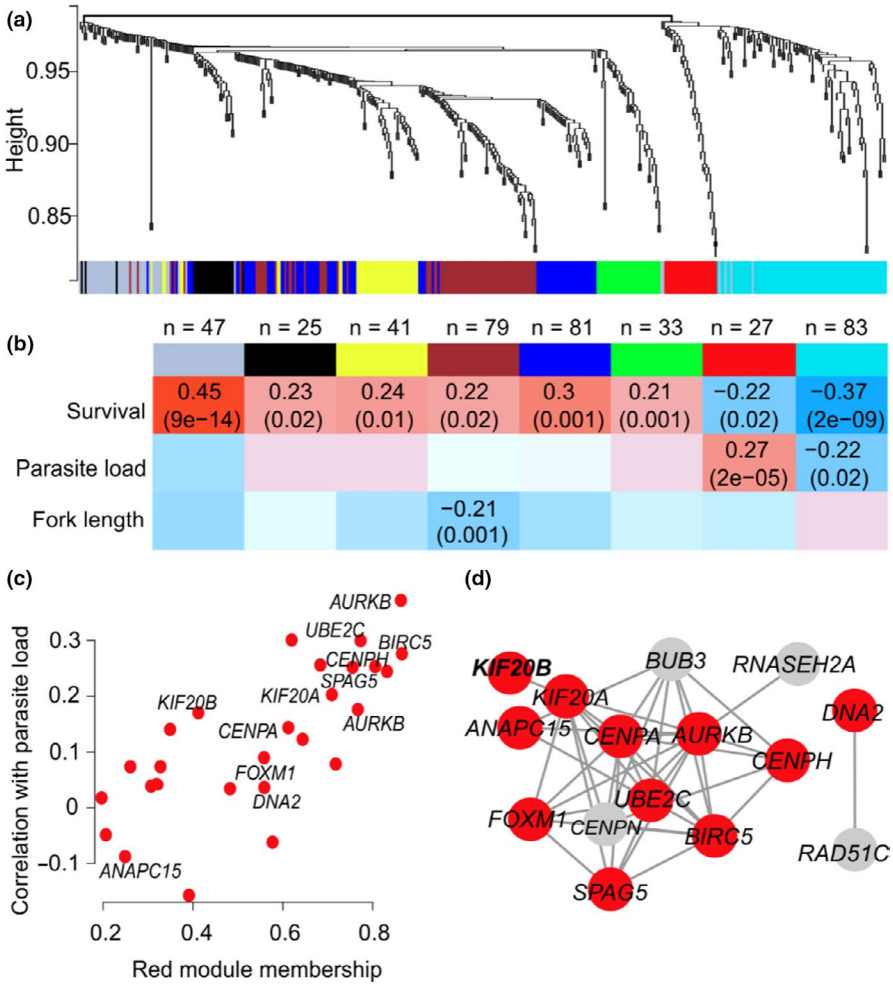

FIGURE 3 Weighted gene co-expression network analysis (WGCNA) of survival genes and their relationship with parasite load. (a) Gene dendrogram with the corresponding seven modules. Each colour represents a module with highly connected genes. (b) Relationships of module eigengenes and survival, parasite load (PL) and fork length (FL). The numbers in the table represent the Pearson correlation coefficients between the corresponding module eigengene and trait, with the *p*-values in parentheses. (c) Module membership of the red module genes and the corresponding Pearson correlation coefficients with parasite load. (d) Protein–protein network of the red module genes involved in the mitotic cell cycle (GO:0000278) shown as red circles

Next, we measured the linear selection gradients (β) for each of the 416 transcripts that covaried with survival after removing the effect of indirect selection in a multiple regression framework using principal component scores (Chong et al., 2018). Among the reconstituted linear selection gradients for the 416 DE genes, a total of 67 estimates of β remained significant (unadjusted $p < .01$, FDR < 0.05). Similar to the differentials, genes showing significant linear gradients were enriched for regulation of the cell cycle (GO:0051726, FDR = 0.031, $n = 9$) and comprised known key regulators of mitotic cell cycle, including *CENPH*, *CENPN* and *KIF20A*. Interestingly, body size, which had no direct effect on survival, showed a significant selection gradient ($\beta = 0.473$; FDR = 6.0×10^{-4}). However, the reconstituted selection gradients for 416 DE genes either controlled or nor controlled for body size (Table S2) were highly correlated ($r^2 = 0.950$) indicating that fish size has little effect on these estimates.

3.4 | Nonlinear selection

Direct comparison of the strength of linear and nonlinear selection using distributional selection differentials (Henshaw & Zemel, 2016) revealed that the linear component of selection was generally stronger than the nonlinear component, which represents selection on the shape of the trait distribution (mean dD = 0.053, dN = 0.031; signed test, $p = 9.4 \times 10^{-206}$; Figure 4b). Nevertheless, for 7273 (40.8%) genes, the nonlinear differentials were higher than the linear selection differentials (Figure S5). Furthermore, permutations indicated that while a small proportion of transcripts were affected by directional selection, the data set was highly enriched for transcripts influenced by disruptive selection, reflecting the elevated survival associated with extreme transcript abundance (Figure 5); the distribution of λ was shifted strongly towards the right tail (Figure 4b, 95% values of λ between −0.137 and 0.279), and its mean differed from zero ($\hat{\lambda} \pm SE = 0.058 \pm 0.001$, $t_{18,716} = 78$, $p = 2.2 \times 10^{-16}$; compared to 658 phenotypic λ estimates [Siepielski et al., 2017], two-sample Wilcoxon test $p = 2.2 \times 10^{-16}$). Similar results were obtained for the quadratic selection differentials calculated for both corrected and uncorrected survival information (Figure S4).

GO analysis indicated that genes shaped by disruptive selection ($\lambda > 0.2$, $n = 1{,}652$) showed enrichment of many molecular processes (GORILLA, 51 GO terms, FDR < 0.05; Table S5), including multi-organism process (GO:0051704, FDR = 0.02, $n = 158$), regulation of cell death (GO:0010941, FDR = 0.045, $n = 187$), iron ion homeostasis (GO:0055072, FDR = 0.035, $n = 18$), vesicle-mediated transport (GO:0016192, FDR = 0.003, $n = 208$) and neutrophil activation (GO:0042119, FDR = 0.041, $n = 73$). The transcripts affected by

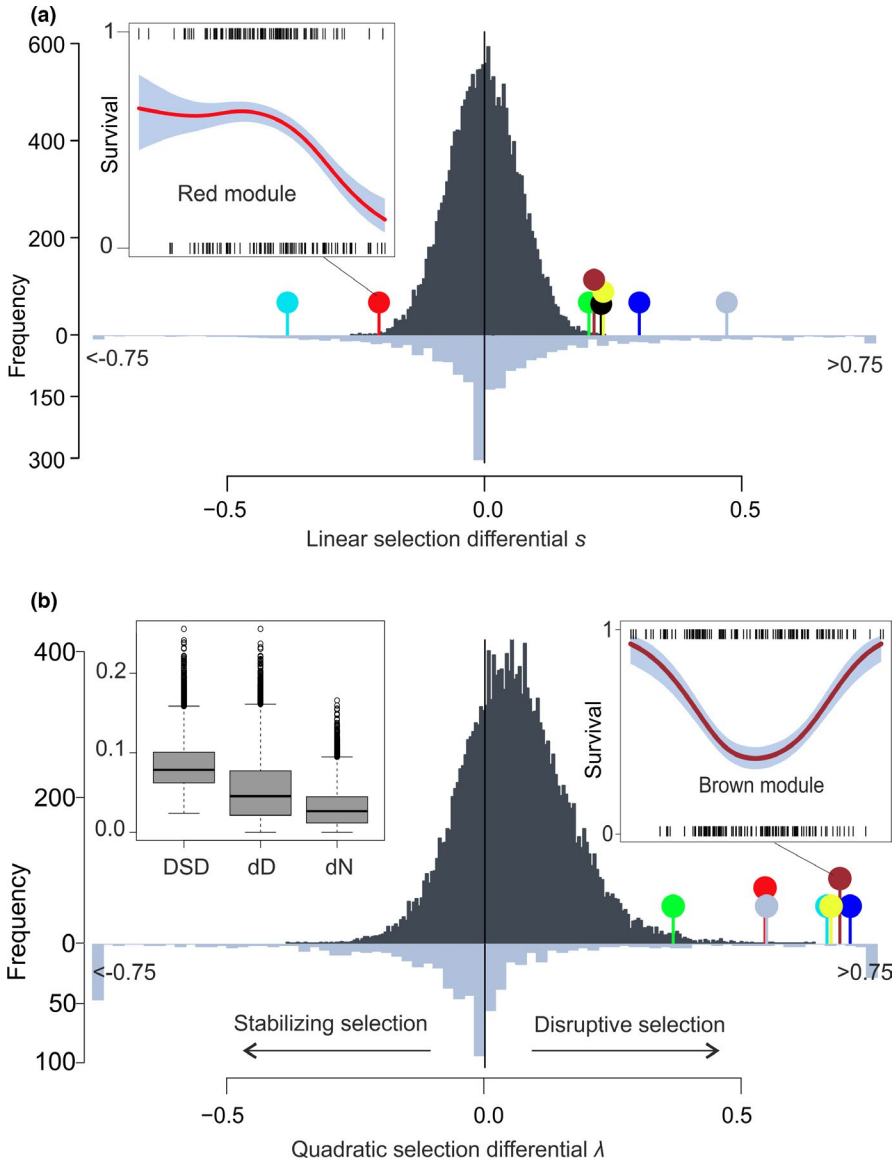

FIGURE 4 Strength of survival selection on 18,717 transcripts and published phenotypic traits. (a) Linear selection differentials *s*. (b) Quadratic selection differentials λ. Differentials for transcripts and published phenotypic traits (Siepielski et al., 2017) are shown as dark and light grey histograms, respectively. Negative and positive λ values reflect stabilizing and disruptive selection, respectively. Estimates <−0.75 were assigned a value of −0.75, and estimates >0.75 were assigned a value of 0.75. Selection differentials for the WGCNA gene modules are shown as coloured pins. The inserted figures illustrate the relationships between survival and module eigengenes as cubic spline (Schluter, 1988) functions (95% CI in grey) for the red and brown modules; short insert lines reflect individual data. The inserted boxplot illustrates total selection as measured by the distributional selection differential (DSD; Henshaw & Zemel, 2016), which is broken down into components representing selection on the trait mean (dD = |*s*|) and selection on the shape of the trait distribution (dN). The line across the box represents the median; the box edges represent the upper and lower quartiles; the whiskers extend to a maximum of 1.5× interquartile range beyond the box; and the points represent outliers

disruptive selection ($\lambda > 0.2$) were clustered into six co-expressed gene modules that all showed higher variance among survivors compared to nonsurvivors, a hallmark of disruptive selection favouring extreme trait values (Levene's test, FDR $< 2.0 \times 10^{-4}$, Figure 4b). The constructed co-expressed gene modules showed further enrichment for more specific GO terms, such as the cellular response to cytokine stimulus (brown module, GO:0071345, FDR = 0.015, $n = 23$) and antigen processing and presentation of peptide antigen via MHC class II (brown module, GO:0002495, FDR = 0.049, $n = 10$).

4 | DISCUSSION

There has long been interest in understanding the relative roles of drift and selection shaping gene expression variation within and between species (Dunn et al., 2013; Romero et al., 2012). The common approach to this complex question encompasses phylogenetic or comparative analyses that aim to indirectly identify patterns of expression, which do not fit neutral expectations over evolutionarily long time periods. However, these approaches describe the response to selection (*R*) and not the strength of selection (*S*) when expressed in the context of the breeder's equation ($R = Sh^2$), where h^2 is narrow-sense heritability. By combining 3′ RNA-sequencing, genetic mark–recapture and selection analysis, we adopted an alternative approach as in Groen et al. (2020) to directly quantify the intensity and form of contemporary natural selection on transcript abundance. As a result, we were able to characterize the transcriptomic targets and potential molecular pathways involved in the process of contemporary parasite-driven selection.

4.1 | The strength of natural selection on transcript abundance

Based on uni- and multivariate regression analysis, we identified a small number of transcripts potentially affected by selection. This is

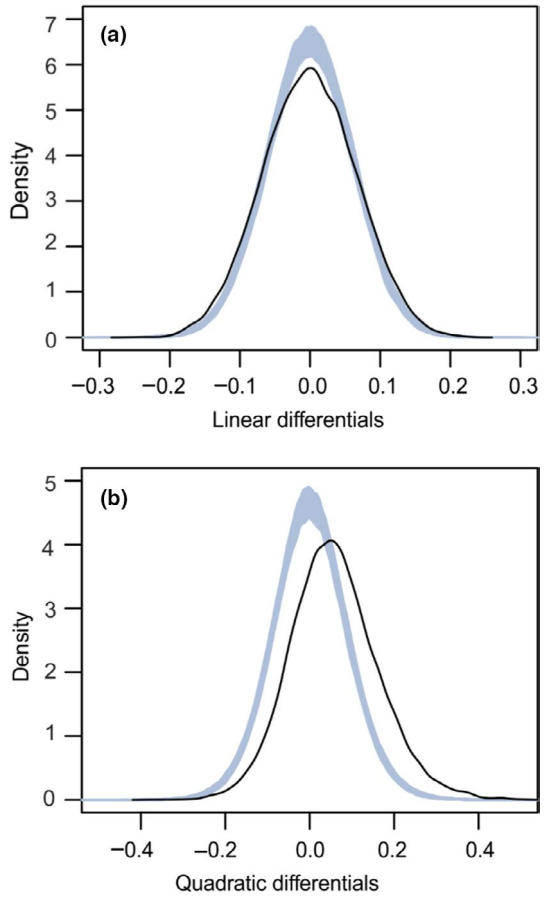

FIGURE 5 The distribution of linear and quadratic selection differentials. (a) Linear selection differentials. (b) Quadratic selection differentials. The black line corresponds to observed data, and grey lines represent 1000 randomizations (no selection)

consistent with recent work in rice, suggesting that directional selection is generally weak at microevolutionary times, and the strength of selection depends on environmental conditions (Groen et al., 2020). Nevertheless, the estimated selection differentials measuring both indirect and direct selection on traits ranged widely from −0.26 to 0.23, implying that if heritable variation is present and constraints are absent, selection can exert evolutionary changes in transcript abundances at evolutionarily short timescales (Campbell-Staton et al., 2017; Donihue et al., 2020; Kingsolver et al., 2001; Kingsolver & Pfennig, 2007). When we compared our transcriptomic data with published phenotypic traits (n = 1834) in terms of the strength of linear selection (which includes both indirect and direct selection components), similar frequency distributions were observed, with the large majority of estimated differentials being close to zero. However, the phenotypic traits possessed longer "tails" for selection differentials than the transcripts, suggesting either rare but very strong linear selection on some phenotypic traits and/or potential bias due to small sample sizes. On the other hand, selection differentials quantify selection considering each trait separately and measure the total selection acting on the trait (both direct and indirect). Thus, when traits are highly correlated, as is likely among the many transcripts, it becomes impractical to distinguish separate influences of individual transcript abundances on relative fitness. To overcome this limitation, we subsequently calculated selection gradients (β) for 416 transcripts that covaried with survival to quantify the strength of direct selection on individual genes after removing indirect selection from other correlated transcripts. Altogether, we identified 67 significant β estimates ranging from −0.47 to −0.15 and from 0.15 to 0.63, indicating that direct selection on transcript abundances has the potential to cause substantial evolutionary changes at relatively short timescales. However, further studies are clearly needed to shed light on how environmental conditions driven by climatic fluctuations influence the strength and form of selection on transcript abundances (Groen et al., 2020). Given that gene expression variation has a strong environmental component, we expect that the patterns of selection often vary considerably among years and populations, and changes in the direction of selection are frequent, as observed for phenotypic traits (Campbell-Staton et al., 2017; Donihue et al., 2020; Siepielski et al., 2009, 2017).

4.2 | The form of natural selection on transcript abundance

Direct comparison of the strength of linear and nonlinear selection using the distributional selection differentials (Henshaw & Zemel, 2016) revealed that for 40% of transcripts, the nonlinear differentials were higher than the linear selection differentials. Furthermore, when nonlinear selection was partitioned into stabilizing and disruptive components, our data set was highly enriched for transcripts showing signatures indicative of disruptive selection. This is unexpected because disruptive selection is thought to be rare in nature (Kingsolver et al., 2001; Kingsolver & Pfennig, 2007). Moreover, this finding contrasts with the expectation that stabilizing selection is more common than disruptive selection if most populations are well adapted to their current environment (Kingsolver et al., 2001). On the other hand, it has been suggested that disruptive selection may be more widespread than previously thought, reflecting density-dependent or frequency-dependent competition for resources (Kingsolver & Pfennig, 2007). Thus, our results corroborate with phenotypic selection estimates, and also suggest that host transcript abundance may be influenced by disruptive selection in response to parasite infection. For example, it may be more beneficial for a host to either invoke a strong immune response (i.e., highly resistant hosts with the lowest PL and lowest disease expression as measured by kidney hyperplasia) or tolerate the damage from a high PL than to partially control the parasite load (i.e., hosts suffering from damage by both parasites and immunopathology). The functional categorization of genes and gene modules under disruptive selection supported this hypothesis, because they were highly enriched for biological processes related to host immune defence, host–pathogen interactions, cellular repair and maintenance. These inferred functions presumably reflect the complex nature of host–parasite interactions, as the transcripts shaped by disruptive selection were involved in a wide range of molecular processes.

4.3 | Functional annotation of putative targets of selection

Variation in transcript abundance, similar to that in morphological or physiological traits, is expected to be shaped by selection through whole-animal performance, which can be defined as how well an organism accomplishes certain ecologically relevant tasks (Arnold, 1983). Therefore, it is pertinent to ask what performance traits are "visible" to selection in the studied host–parasite system. The functional categorization of genes and correlated gene modules provides some clues to this question, as both survival- and PL-associated transcripts shaped by linear selection were highly enriched for genes involved in the mitotic cell cycle. First, it is unlikely that the genes associated with survival reflect variation for general stress response of the host caused by *Tetracapsuloides bryosalmonae* infection. This is because most stress factors lead to an arrest of mitosis (Burgess et al., 2014; Kassahn et al., 2009; Martín-Hernández et al., 2017), whereas we detected that PL associated with up-regulation, rather than down-regulation (what may be expected for arrest of mitosis), of the key mitotic cell cycle host genes (*AURKB*, *UBE2C*, *BIRC5*; Figure 3). Second, the observed associations between fin tissue transcriptome, PL and survival may reflect the host response to parasite entry because *T. bryosalmonae* enters its salmonid host through surface tissues, which may include fins (Longshaw et al., 2002). However, we currently lack experimental evidence that *T. bryosalmonae* entry causes upregulation of cell-cycle activity in fin or/and other mucosal tissues of the host. Third, the coupling of the transcription of mitotic cell cycle genes in fin, PL and survival may reflect the severe physiological impact of PKD on the host at the whole organismal level. For example, previous studies in salmonids have demonstrated that one of the main PKD symptoms is tumour-like proliferation of the lymphoid renal tissue, where the kidney may increase in size to over ten times its normal volume (Figure 1; Bettge et al., 2009; Clifton-Hadley et al., 1987; Hedrick et al., 1993). Similarly, PKD causes enlargement of the spleen, and several studies suggest that PKD in salmonids is a systemic disease that affects multiple organs and tissues (Bettge et al., 2009; Bruneaux et al., 2016; Clifton-Hadley et al., 1987; Hedrick et al., 1993; Longshaw et al., 2002; Okamura et al., 2011). In teleost fishes, pelvic fins consist of epidermis, bony rays, ligaments, nerve fibres, connective tissue cells and blood vessels. The observed associations between cell-cycle-related host genes, PL and survival may, therefore, reflect the importance of blood homeostasis and sustaining normal kidney function. However, analysis of multiple tissues, including renal, blood and fin transcriptomes, during the progression of PKD is clearly needed to further dissect the molecular mechanisms of the host response, as we currently lack comprehensive understanding of the inflammatory, mitotic and immune processes across organs (Chevrier, 2019). Regardless of the specific physiological mechanism, this work adds to the increasing body of work showing that parasitism can have an effect on the host's cellular machinery (Guo et al., 2016; Kassahn et al., 2009; Martín-Hernández et al., 2017).

Two limitations in this study may be addressed in future research. First, despite the high electrofishing effort, low dispersal and relatively high recapture probabilities, we probably did not recover all survivors. We therefore carried out functional and selection analysis based on both initial recapture information and by treating 13 putatively uncaptured individuals as survivors, as suggested by their transcript profiles that matched recaptured individuals. However, given that the main findings (e.g., distribution of the linear and nonlinear selection coefficients, GO enrichment patterns) remained very similar irrespective of the classification, imperfect classification of small number of survivors probably had only a small effect on the main conclusions. Second, even though it was not possible to analyse the primary target tissue of the parasite (kidney), requiring lethal sampling and thereby preventing survival estimates, the systemic nature of PKD conceivably enabled us to acquire biologically meaningful information from fin biopsies with a minimal expected effect on fish survival (Gjerde & Refstie, 1988). Similarly, transcript abundances in fin tissue have recently been associated with ageing-related mortality in fish, demonstrating the usefulness of fin tissue for linking gene expression and whole-organism performance (Baumgart et al., 2016). More generally, because of their major role in pathogen defence, mucosal surface tissues have been widely used to study innate and adaptive immune responses in teleost fishes (Gomez et al., 2013).

In summary, our work demonstrates the power and challenges of integrating nonlethal sampling and transcriptomics with classical ecological methods to dissect complex high-order organismal traits, such as survival in the wild, into functionally interpretable molecular processes. As such, our study provides a novel perspective for studying contemporary selection at the suborganismal level and is readily applicable to other species and systems, where nonlethal sampling of blood, mucosal and other tissues is feasible. We anticipate that the approach described here will enable critical information on the molecular mechanisms and targets of natural selection to be obtained in real time, as wild populations increasingly contend with novel selective pressures (Hendry et al., 2008), including those imposed by global warming (Hoffmann et al., 2011).

ACKNOWLEDGEMENTS

We thank M. F. Oleksiak, J. M. Henshaw, T. Aykanat, V. Kisand, C. Primmer, T. Tenson, R. J. Pawluk and A. Krasnov for commenting on earlier drafts of the manuscript; K. Haugjärv for help during sample collection; K. Salminen from the Center of Evolutionary Applications, University of Turku, for RNA extractions and library preparations; and Põlula Fish Rearing Centre (RMK) for their support during fieldwork. Bioinformatic analyses used resources at the CSC—IT Center for Science, Finland. This work was supported by the Academy of Finland (grant 266321), the Estonian Ministry of Education and Research (institutional research funding project IUT8-2), the Estonian Research Council grant (PRG852), Ella & Georg Ehrnrooth foundation, University of Turku Foundation, Archimedes Foundation Scholarship in smart specialization growth areas, and a German Research Foundation Research Fellowship (DE 2405/1-1).

AUTHOR CONTRIBUTIONS

A.V. conceived the study. A.V., F.A., P.V.D., S.K. and L.P. collected the samples. P.V.D. measured haematocrit and kidney swollenness. I.N. carried out microsatellite analysis and parasite quantification. M.O. carried out mark–recapture analysis. F.A. carried out bioinformatic and transcriptomic analyses. F.A. and P.V.D. estimated selection differentials. All authors participated in interpretation of the results. A.V., F.A. and P.V.D. drafted the manuscript, and all others commented. Competing interests: the authors declare no competing interests.

DATA AVAILABILITY STATEMENT

Data supporting the findings of this work are available in the Supporting Information. The sequence data have been deposited in the NCBI BioProject database under accession no. PRJNA517427 and are publicly available. The R scripts and input files for each analysis are available from the Dryad Digital Repository (https://doi.org/10.5061/dryad.612jm641t).

ORCID

Freed Ahmad https://orcid.org/0000-0002-8994-4723
Paul V. Debes https://orcid.org/0000-0003-4491-9564
Anti Vasemägi https://orcid.org/0000-0002-2184-5534

REFERENCES

Arnold, S. J. (1983). Morphology, performance and fitness. *American Zoologist*, *23*(2), 347–361. https://doi.org/10.1093/icb/23.2.347

Baumgart, M., Priebe, S., Groth, M., Hartmann, N., Menzel, U., Pandolfini, L., Koch, P., Felder, M., Ristow, M., Englert, C., Guthke, R., Platzer, M., & Cellerino, A. (2016). Longitudinal RNA-seq analysis of vertebrate aging identifies mitochondrial complex I as a small-molecule-sensitive modifier of lifespan. *Cell Systems*, *2*(2), 122–132. https://doi.org/10.1016/J.CELS.2016.01.014

Bettge, K., Segner, H., Burki, R., Schmidt-Posthaus, H., & Wahli, T. (2009). Proliferative kidney disease (PKD) of rainbow trout: Temperature- and time-related changes of T*etracapsuloides bryosalmonae* DNA in the kidney. *Parasitology*, *136*(06), 615. https://doi.org/10.1017/s0031182009005800

Bishop, S. C., Doeschl-Wilson, A., & Woolliams, J. A. (2012). Uses and implications of field disease data for livestock genomic and genetics studies. *Frontiers in Genetics*, *3*, 1–5. https://doi.org/10.3389/fgene.2012.00114

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bruneaux, M., Visse, M., Gross, R., Pukk, L., Saks, L., & Vasemägi, A. (2016). Parasite infection and decreased thermal tolerance: Impact of proliferative kidney disease on a wild salmonid fish in the context of climate change. *Functional Ecology*, *31*(1), 216–226. https://doi.org/10.1111/1365-2435.12701

Burgess, A., Rasouli, M., & Rogers, S. (2014). Stressing mitosis to death. *Frontiers in Oncology*, *4*, 140. https://doi.org/10.3389/fonc.2014.00140

Butler, D. G., Cullis, B. R., Gilmour, A. R., & Gogel, B. J. (2009). *Mixed models for S language environments ASReml-R reference manual*.

Campbell-Staton, S. C., Cheviron, Z. A., Rochette, N., Catchen, J., Losos, J. B., & Edwards, S. V. (2017). Winter storms drive rapid phenotypic, regulatory, and genomic shifts in the green anole lizard. *Science (American Association for the Advancement of Science)*, *357*(6350), 495–498. https://doi.org/10.1126/science.aam5512

Chaussabel, D., Quinn, C., Shen, J., Patel, P., Glaser, C., Baldwin, N., Stichweh, D., Blankenship, D., Li, L., Munagala, I., Bennett, L., Allantaz, F., Mejias, A., Ardura, M., Kaizer, E., Monnet, L., Allman, W., Randall, H., Johnson, D., … Pascual, V. (2008). A modular analysis framework for blood genomics studies: Application to systemic lupus erythematosus. *Immunity*, *29*(1), 150–164. https://doi.org/10.1016/J.IMMUNI.2008.05.012

Chevrier, N. (2019). Decoding the body language of immunity: Tackling the immune system at the organism level. *Current Opinion in Systems Biology*, *18*, 19–26. https://doi.org/10.1016/j.coisb.2019.10.010

Chong, V. K., Fung, H. F., & Stinchcombe, J. R. (2018). A note on measuring natural selection on principal component scores. *Evolution Letters*, *2*(4), 272–280. https://doi.org/10.1002/evl3.63

Clifton-Hadley, R., Bucke, D., & Richards, R. H. (1987). A study of the sequential clinical and pathological changes during proliferative kidney disease in rainbow trout, *Salmo gairdneri* Richardson. *Journal of Fish Diseases*, *10*(5), 335–352. https://doi.org/10.1111/j.1365-2761.1987.tb01081.x

Cobb, J. P., Mindrinos, M. N., Miller-Graziano, C., Calvano, S. E., Baker, H. V., Xiao, W., … Inflammation and Host Response to Injury Large-Scale Collaborative, Research Program. (2005). Application of genome-wide expression analysis to human health and disease. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(13), 4801–4806. https://doi.org/10.1073/pnas.0409768102

Dash, M., & Vasemägi, A. (2014). Proliferative kidney disease (PKD) agent *Tetracapsuloides bryosalmonae* in brown trout populations in Estonia. *Diseases of Aquatic Organisms*, *109*(2), 139–148. https://doi.org/10.3354/dao02731

Debes, P. V., Gross, R., & Vasemägi, A. (2017). Quantitative genetic variation in, and environmental effects on, pathogen resistance and temperature-dependent disease severity in a wild trout. *The American Naturalist*, *190*(2), 244–265. https://doi.org/10.1086/692536

Doeschl-Wilson, A., Bishop, S. C., Kyriazakis, I., & Villanueva, B. (2012). Novel methods for quantifying individual host response to infectious pathogens for genetic analyses. *Frontiers in Genetics*, *3*, 1–9. https://doi.org/10.3389/fgene.2012.00266

Donihue, C. M., Kowaleski, A. M., Losos, J. B., Algar, A. C., Baeckens, S., Buchkowski, R. W., Herrel, A. (2020). Hurricane effects on neotropical lizards span geographic and phylogenetic scales. *Proceedings of the National Academy of Sciences*, *117*(19), 10429–10434. https://doi.org/10.1073/pnas.2000801117

Dunn, C. W., Luo, X., & Wu, Z. (2013). Phylogenetic analysis of gene expression. *Integrative and Comparative Biology*, *53*(5), 847–856. https://doi.org/10.1093/icb/ict068

Eden, E., Navon, R., Steinfeld, I., Lipson, D., & Yakhini, Z. (2009). GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, *10*(1), 1–7. https://doi.org/10.1186/1471-2105-10-48

Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A. S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G. B., Gunnarsdottir, S., Mouy, M., Steinthorsdottir, V., Eiriksdottir, G. H., Bjornsdottir, G., Reynisdottir, I., Gudbjartsson, D., Helgadottir, A., Jonasdottir, A., Jonasdottir, A., … Stefansson, K. (2008). Genetics of gene expression and its effect on disease. *Nature*, *452*(7186), 423–428. https://doi.org/10.1038/nature06758

Fraser, H. B. (2013). Gene expression drives local adaptation in humans. *Genome Research*, *23*(7), 1089–1096. https://doi.org/10.1101/gr.152710.112

Fraser, H. B., Moses, A. M., & Schadt, E. E. (2010). Evidence for widespread adaptive evolution of gene expression in budding yeast. *Proceedings of the National Academy of Sciences*, *107*(7), 2977–2982. https://doi.org/10.1073/pnas.0912245107

Gilad, Y., Oshlack, A., & Rifkin, S. A. (2006). Natural selection on gene expression. *Trends in Genetics*, *22*(8), 456–461. https://doi.org/10.1016/j.tig.2006.06.002

Gjerde, B., & Refstie, T. (1988). The effect of fin-clipping on growth rate, survival and sexual maturity of rainbow trout. *Aquaculture*, *73*(1–4), 383–389. https://doi.org/10.1016/0044-8486(88)90071-3

Gomez, D., Sunyer, J. O., & Salinas, I. (2013). The mucosal immune system of fish: The evolution of tolerating commensals while fighting pathogens. *Fish & Shellfish Immunology*, *35*(6), 1729–1739. https://doi.org/10.1016/J.FSI.2013.09.032

Graham, A. L., Shuker, D. M., Pollitt, L. C., Auld, S. K. J. R., Wilson, A. J., & Little, T. J. (2010). Fitness consequences of immune responses: Strengthening the empirical framework for ecoimmunology. *Functional Ecology*, *25*(1), 5–17. https://doi.org/10.1111/j.1365-2435.2010.01777.x

Groen, S. C., Ćalić, I., Joly-Lopez, Z., Platts, A. E., Choi, J. Y., Natividad, M., Dorph, K., Mauck, W. M., Bracken, B., Cabral, C. L. U., Kumar, A., Torres, R. O., Satija, R., Vergara, G., Henry, A., Franks, S. J., & Purugganan, M. D. (2020). The strength and pattern of natural selection on gene expression in rice. *Nature (London)*, *578*(7796), 572–576. https://doi.org/10.1038/s41586-020-1997-2

Guo, Z., González, J. F., Hernandez, J. N., McNeilly, T. N., Corripio-Miyar, Y., Frew, D., Morrison, T., Yu, P., & Li, R. W. (2016). Possible mechanisms of host resistance to *Haemonchus contortus* infection in sheep breeds native to the canary islands. *Scientific Reports*, *6*(1), 26200. https://doi.org/10.1038/srep26200

Hari, R. E., Livingstone, D. M., Siber, R., Burkhardt-Holm, P., & Guttinger, H. (2006). Consequences of climatic change for water temperature and brown trout populations in alpine rivers and streams. *Global Change Biology*, *12*(1), 10–26. https://doi.org/10.1111/j.1365-2486.2005.001051.x

Hedrick, R. P., MacConnell, E., & de Kinkelin, P. (1993). Proliferative kidney disease of salmonid fish. *Annual Review of Fish Diseases*, *3*, 277–290. https://doi.org/10.1016/0959-8030(93)90039-e

Hendry, A. P., Farrugia, T. J., & Kinnison, M. T. (2008). Human influences on rates of phenotypic change in wild animal populations. *Molecular Ecology*, *17*(1), 20–29. https://doi.org/10.1111/j.1365-294x.2007.03428.x

Henshaw, J. M., & Zemel, Y. (2016). A unified measure of linear and nonlinear selection on quantitative traits. *Methods in Ecology and Evolution*, *8*(5), 604–614. https://doi.org/10.1111/2041-210x.12685

Hoffmann, A. A., & Sgrò, C. M. (2011). Climate change and evolutionary adaptation. *Nature*, *470*(7335), 479–485. https://doi.org/10.1038/nature09670

Husak, J. F. (2016). Measuring selection on physiology in the wild and manipulating phenotypes (in Terrestrial Nonhuman Vertebrates). *Comprehensive physiology*, *6*(1), 63–85). https://doi.org/10.1002/cphy.c140061

Janzen, F. J., & Stern, H. S. (1998). Logistic regression for empirical studies of multivariate selection. *Evolution*, *52*(6), 1564. https://doi.org/10.2307/2411330

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 94. https://doi.org/10.1186/1471-2156-11-94

Jones, O. R., & Wang, J. (2010). COLONY: A program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, *10*(3), 551–555. https://doi.org/10.1111/j.1755-0998.2009.02787.x

Kassahn, K. S., Crozier, R. H., Pörtner, H. O., & Caley, M. J. (2009). Animal performance and stress: Responses and tolerance limits at different levels of biological organisation. *Biological Reviews*, *84*(2), 277–292. https://doi.org/10.1111/j.1469-185X.2008.00073.x

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. https://doi.org/10.1038/nmeth.3317

Kingsolver, J. G., Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hill, C. E., Hoang, A., Gibert, P., & Beerli, P. (2001). The strength of phenotypic selection in natural populations. *The American Naturalist*, *157*(3), 245–261. https://doi.org/10.1086/319193

Kingsolver, J. G., & Pfennig, D. W. (2007). Patterns and power of phenotypic selection in nature. *BioScience*, *57*(7), 561–572. https://doi.org/10.1641/B570706

Lande, R., & Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution*, *37*(6), 1210. https://doi.org/10.2307/2408842

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 1–13. https://doi.org/10.1186/1471-2105-9-559

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 Biology*, *15*(2), R29. https://doi.org/10.1186/gb-2014-15-2-r29

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., & Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Computational Biology*, *9*(8), e1003118. https://doi.org/10.1371/journal.pcbi.1003118

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., & Storey, J. D. (2012). The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, *28*(6), 882–883. https://doi.org/10.1093/bioinformatics/bts034

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Computer Applications in the Biosciences*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., Hvidsten, T. R., 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., … Davidson, W. S. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature*, *533*(7602), 200–205. https://doi.org/10.1038/nature17164

Lohman, B. K., Weber, J. N., & Bolnick, D. I. (2016). Evaluation of TagSeq, a reliable low-cost alternative for RNAseq. *Molecular Ecology Resources*, *16*(6), 1315–1321. https://doi.org/10.1111/1755-0998.12529

Longshaw, M., Le Deuff, R., Harris, A. F., & Feist, S. W. (2002). Development of proliferative kidney disease in rainbow trout, *oncorhynchus mykiss* (walbaum), following short-term exposure to *Tetracapsula bryosalmonae* infected bryozoans. *Journal of Fish Diseases*, *25*(8), 443–449. https://doi.org/10.1046/j.1365-2761.2002.00353.x

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/s13059-014-0550-8

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

Martín-Hernández, R., Higes, M., Sagastume, S., Juarranz, Á., Dias-Almeida, J., Budge, G. E., Meana, A., & Boonham, N. (2017). Microsporidia infection impacts the host cell's cycle and reduces host cell apoptosis. *PLoS One*, *12*(2), e0170183. https://doi.org/10.1371/journal.pone.0170183

Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance*. Belknap Press of Harvard University Press.

Meyer, E., Aglyamova, G. V., & Matz, M. V. (2011). Profiling gene expression responses of coral larvae (*Acropora millepora*) to elevated temperature and settlement inducers using a novel RNA-seq procedure. *Molecular Ecology*, 3599–3616. https://doi.org/10.1111/j.1365-294x.2011.05205.x

Miller, K. M., Li, S., Kaukinen, K. H., Ginther, N., Hammill, E., Curtis, J. M. R., Patterson, D. A., Sierocinski, T., Donnison, L., Pavlidis, P., Hinch, S. G., Hruska, K. A., Cooke, S. J., English, K. K., & Farrell, A. P.

(2011). Genomic signatures predict migration and spawning failure in wild Canadian salmon. *Science*, *331*(6014), 214–217. https://doi.org/10.1126/science.1196901

Mo, T. A., & Jørgensen, A. (2016). A survey of the distribution of the PKD-parasite *Tetracapsuloides bryosalmonae* (cnidaria: Myxozoa: Malacosporea) in salmonids in norwegian rivers - additional information gleaned from formerly collected fish. *Journal of Fish Diseases*, *40*(5), 621–627. https://doi.org/10.1111/jfd.12542

Moll, P., Ante, M., Seitz, A., & Reda, T. (2014). QuantSeq 3′ mRNA sequencing for RNA quantification. *Nature Methods*, *11*(12), i–iii. https://doi.org/10.1038/nmeth.f.376

Ogle, D. H. (2017). *FSA: Fisheries stock analysis*. R package 0.8.17.

Okamura, B., Hartikainen, H., Schmidt-Posthaus, H., & Wahli, T. (2011). Life cycle complexity, environmental change and the emerging status of salmonid proliferative kidney disease. *Freshwater Biology*, *56*(4), 735–753. https://doi.org/10.1111/j.1365-2427.2010.02465.x

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3), 290–295. https://doi.org/10.1038/nbt.3122

R Core Team. (2017). *R: A language and environment for statistical computing. R foundation for statistical computing*. Vienna.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47. https://doi.org/10.1093/nar/gkv007

Romero, I. G., Ruvinsky, I., & Gilad, Y. (2012). Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics*, *13*(7), 505–516. https://doi.org/10.1038/nrg3229

Schluter, D. (1988). Estimating the form of natural selection on a quantitative trait. *Evolution*, *42*(5), 849. https://doi.org/10.2307/2408904

Siepielski, A. M., DiBattista, J. D., & Carlson, S. M. (2009). It's about time: The temporal dynamics of phenotypic selection in the wild. *Ecology Letters*, *12*(11), 1261–1276. https://doi.org/10.1111/j.1461-0248.2009.01381.x

Siepielski, A. M., Morrissey, M. B., Buoro, M., Carlson, S. M., Caruso, C. M., Clegg, S. M., Coulson, T., DiBattista, J., Gotanda, K. M., Francis, C. D., Hereford, J., Kingsolver, J. G., Augustine, K. E., Kruuk, L. E. B., Martin, R. A., Sheldon, B. C., Sletvold, N., Svensson, E. I., Wade, M. J., & MacColl, A. D. C. (2017). Precipitation drives global variation in natural selection. *Science*, *355*(6328), 959–962. https://doi.org/10.1126/science.aag2773

Skovgaard, A., & Buchmann, K. (2012). *Tetracapsuloides bryosalmonae* and PKD in juvenile wild salmonids in Denmark. *Diseases of Aquatic Organisms*, *101*(1), 33–42. https://doi.org/10.3354/dao02502

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., & von Mering, C. (2015). STRING v10: Protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, *43*(D1), D447–D452. https://doi.org/10.1093/nar/gku1003

Tops, S., Lockwood, W., & Okamura, B. (2006). Temperature-driven proliferation of *Tetracapsuloides bryosalmonae* in bryozoan hosts portends salmonid declines. *Diseases of Aquatic Organisms*, *70*(3), 227–236. https://doi.org/10.3354/dao070227

Vasemägi, A., Nousiainen, I., Saura, A., Vähä, J. P., Valjus, J., & Huusko, A. (2017). First record of proliferative kidney disease agent *Tetracapsuloides bryosalmonae* in wild brown trout and European grayling in Finland. *Diseases of Aquatic Organisms*, *125*(1), 73–78. https://doi.org/10.3354/dao03126

Winter, D. J. (2017). Rentrez: An R package for the NCBI eUtils API. *The R Journal*, *9*(2), 520. https://doi.org/10.32614/rj-2017-058

Wright, M. N., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, *77*(1), 1–17. https://doi.org/10.18637/jss.v077.i01

Zhao, S., & Fernald, R. D. (2005). Comprehensive algorithm for quantitative real-time polymerase chain reaction. *Journal of Computational Biology*, *12*(8), 1047–1064. https://doi.org/10.1089/cmb.2005.12.1047

Zippin, C. (1958). The removal method of population estimation. *The Journal of Wildlife Management*, *22*(1), 82. https://doi.org/10.2307/3797301

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Ahmad F, Debes PV, Nousiainen I, et al. The strength and form of natural selection on transcript abundance in the wild. *Mol Ecol*. 2021;30:2724–2737. https://doi.org/10.1111/mec.15743