Immune-related transcriptional responses to parasitic infection in a naturally inbred fish: roles of genotype and individual variation

Rebecca Jane Pawluk\textsuperscript{1}, Tamsyn M. Uren Webster\textsuperscript{1}, Joanne Cable\textsuperscript{2}, Carlos Garcia de Leaniz\textsuperscript{1}, Sofia Consuegra\textsuperscript{1*}

\textsuperscript{1}Swansea University, Singleton Park, Swansea, SA2 8PP, Wales, UK
\textsuperscript{2} School of Biosciences, Cardiff University, Cardiff, CF10 3AX, Wales, UK

\textsuperscript{*}Correspondence to s.consuegra@swansea.ac.uk

Keywords: \textit{Kryptolebias marmoratus}, transcriptome, \textit{Argulus folicaeous}, immune response, inbreeding

Data reposition: The datasets generated and/or analysed during the current study have been submitted to the Gene Expression omnibus (GEO) repository (accession number: GSE101084)
Abstract

Parasites are strong drivers of evolutionary change and the genetic variation of both host and parasite populations can co-evolve as a function of parasite virulence and host resistance. The role of transcriptome variation in specific interactions between host and parasite genotypes has been less studied and can be confounded by differences in genetic variation. We employed two naturally inbred lines of a self-fertilising fish to estimate the role of host genotype in the transcriptome response to parasite infection using RNA-seq. In addition, we targeted several differentially expressed immune-related genes to further investigate the relative role of individual variation in the immune response using qPCR, taking advantage of the genomic uniformity of the self-fertilising lines. We found significant differences in gene expression between lines in response to infection both in the transcriptome and in individual gene qPCR analyses. Individual qPCR analyses of gene expression identified significant variance differences between lines for 6 genes but only for 3 genes between infected and control fish. Our results indicate that although the genetic background plays an important role in the transcriptome response to parasites, it cannot fully explain individual differences within genetically homogeneous lines, which can be important for determining the response to parasites.
Introduction

Parasites are one of the strongest drivers of evolutionary change, influencing the abundance, distribution and genetic variability of their hosts (Ebert and Hamilton 1996). The genetic composition of both parasites and hosts can change as a result of co-evolution between parasite virulence and host immune resistance (Hamilton 1980). In this sense, the Red Queen Hypothesis predicts that parasite-mediated selection against hosts with common genotypes should help maintain high genetic variability in the offspring (Morran, et al. 2011), thereby providing a potential explanation for the maintenance of sexual reproduction despite the costs of producing males (Howard and Lively 1994; Morran, et al. 2011). Since the immune response against parasites involves costs and life-history trade-offs (Lochmiller and Deerenberg 2000; Sheldon and Verhulst 1996), hosts need to optimise their defence mechanisms, for example by developing long-term immunity only to certain parasites (e.g. those with high transmission rates and intermediate virulence; (Boots and Bowers 2004). Parasites, in turn, can develop local adaptation to the hosts’ most common genotypes, exhibiting greater fitness in local than in allopatric hosts (Greischar and Koskella 2007). The extent to which parasites can be locally adapted will largely depend on migration rates exhibited by both hosts and parasites (Gandon, et al. 1996), their genetic background (Andrade, et al. 2002), but critically, also on the interactions between host and parasite genomes (Lambrechts, et al. 2006).

Environmental conditions influence the presence and abundance of hosts and their parasites, thereby shaping the composition of parasite communities (Eizaguirre and Lenz 2010; Evans and Neff 2009). However, host resistance seems to be influenced by only a few loci and their epistatic interactions and this may depend more on the genotype combination of hosts and parasites than on the environment (Wilfert and Schmid-Hempel 2008). The host genotype can modulate parasite fitness, infection patterns and virulence by influencing the strength and direction of selection on parasite evolution (De Roode and Altizer 2010). For example, in the
snail *Biomphalaria glabrata*, genetic background can influence responses to infection by the castrating parasite *Schistosoma mansoni* as well as the parasite’s reproductive success (Tavalire, et al. 2016). In Chagas disease, which has a variable spectrum of pathology caused by *Trypanosoma cruzi*, parasite distribution between tissues depends in part on the genetic makeup of the parasite, but largely differs depending on the host’s genotype (Campbell, et al. 2004), specifically on host’s loci involved in the modulation of the infection (e.g. the genes of the Major Histocompatibility Complex or MHC (Andrade, et al. 2002)).

Genotype-by-genotype interaction between hosts and parasites can result in increasing host genetic diversity, e.g. allelic diversity in immune-related genes (Carius, et al. 2001). The highly polymorphic MHC genes are probably the most studied in host-parasite genotype interactions (Bernatchez and Landry 2003). However, the MHC-based adaptive immune response can be a relatively slow response to infection (Acevedo-Whitehouse and Cunningham 2006), and it does not fully explain host resistance (Rauch, et al. 2006b). Growing evidence suggests that the innate immune response plays central role in the interactions between host and parasite genotypes, as it has a tight relationship with acquired immunity (Magnadóttir 2006) and provides a more immediate form of defence against pathogens, which could act within hours of infection (Rauch, et al. 2006a).

Specific interactions between host and parasite genotypes also manifest as variation in gene expression when, for example, different parasite genotypes elicit variable responses in the host transcriptome (Barribeau, et al. 2014). Differences in immune gene expression have been observed in response to infection by different parasite genotypes in several organisms, including bumble bees (Barribeau, et al. 2014) and sticklebacks (Haase, et al. 2014). Host specific gene expression responses to parasite infection can be more challenging to identify, particularly in vertebrates where individual genetic variability confounds interpretation. Laboratory inbred mice lines, along with other model organisms, are commonly used to help
understand the role of host genetic background in response to infection (Srivastava, et al. 2009) and the influence of allele-specific variation on transcript abundance (Keane, et al. 2011). However, the extent to which results from model organisms can explain processes occurring in natural populations is debatable, and it is recognised that evolutionary mutant models (organisms with particular traits of interest that have developed over generations of natural, instead of artificial, selection) can provide additional insights into the genetic factors and gene-by-environment interactions that affect the immune response (Albertson, et al. 2009). With this in mind, we took advantage of the self-fertilising and naturally inbred mangrove killifish (Kryptolebias marmoratus) to examine the relative roles of genotype versus individual variation in the immune-related transcriptome response of the host to parasite infection.
Methods

Study species and experimental design

We used two different *K. marmoratus* selfing lines (R and DAN) originating from Belize that have undergone at least 30 generations of self-fertilisation (Ellison, et al. 2013). Fish were reared in individual aquaria (12 x 8 x 8.5 cm) containing 750 ml of brackish water (14 ppt salinity, constituted from dechlorinated water and marine filtered water) under controlled conditions (12:12 hours light:dark photoperiod, 24°C). Eighty 8-month-old *K. marmoratus* were selected from the two lines (40 DAN and 40 R). Twenty fish from each line were size-matched and were individually infected with a single *Argulus foliaceus* (an ectoparasitic freshwater louse) following (Stewart, et al. 2017), whereas the other twenty from each line were unexposed to the parasite to serve as controls. The culture of *A. foliaceus* originated from carp (*Cyprinus carpio*) caught in a still water fishery in North Lincolnshire, July 2014, and thereafter was maintained on *Gasterosteus aculeatus* (three-spined sticklebacks) at Cardiff University as detailed in Stewart et al. (2017). *A. foliaceus* is a generalist parasite that tends to spend variable periods away from the host, often resting on the substrate and can be monitored visually without anaesthetising the fish. The attachment of the parasites was facilitated by manually placing a single *Argulus* on each fish’s skin, allowing for suction by the parasite to the body surface and afterwards the presence of the parasite attached to the fish was visually monitored every 2 hours. Fish size was not significantly different between lines (DAN mean size=14.8 mm, SD=0.027; R mean size=14.3mm, SD=0.012; Mann-Whitney U=287.000; P=0.984) and *Argulus* successfully attached to all exposed fish, with attachment times varying between 18 and 48 hours. Attachment times (grouped by 6 h intervals) did not differ between lines (Kruskal-Wallis chi-squared = 1.52 df = 4, P = 0.822). After 48 hours approximately 50% of the fish were still infected and the experiment was terminated, all fish were humanely...
euthanized, stored in RNA later and frozen at -80°C prior to RNA extraction and library preparation.

**RNA extraction, library construction and sequencing**

For transcriptomic analyses, we selected fish that had remained infected for 48 hours. Total RNA was extracted from 20 whole individuals (5 R controls, 5 DAN controls, 5 R infected individuals, 5 DAN infected individuals) using the Bioline Isolate II RNA mini kit (Bioline, London, United Kingdom) according to the manufacturer’s instructions. The concentration and quality of RNA in each sample was determined using a NanoDrop 2000 Spectrophotometer (NanoDrop Technologies, U.S.A.) and a Qubit™ fluorometer. Library construction was completed using the Illumina TruSeq kit v2 according to the manufacturer’s instructions (Illumina, San Diego, CA) using approximately 500-1000 ng of RNA. The concentration and quality of the libraries were determined using the Qubit (Invitrogen) and Bioanalyser 2100 (Agilent Technologies). All samples were sequenced (126 bp paired end) on an Illumina Hiseq2500 platform (Illumina, San Diego, CA, USA). Samples were pooled for library construction when RNA concentrations were <500 ng resulting in 4 pool groups of 3 fish (control DAN, control R, infected DAN, infected R), 1 pool of 2 fish (infected R) and 6 fish individually sequenced (2 of each control DAN, control R, infected DAN).

**Transcriptome assembly and annotation**

Raw sequences were processed using Trimmomatic, version 0.33 (Bolger, et al. 2014), to ensure Illumina adapter sequences were removed and poor quality 3’ ends were trimmed using a sliding window (Q > 20). *De novo* transcriptome assembly was conducted with reads from all combined samples using Trinity version r2013-02-25 (Grabherr, et al. 2011; Haas, et al. 2013), encompassing an initial *in silico* normalisation with an optimised K value of 40. The
final transcriptome assembly was annotated using Blastx (Altschul, et al. 1990) against Ensembl peptide databases for zebrafish, medaka, stickleback, mice and humans using an e-value cutoff < 1e-5. Most of the annotations were from zebrafish. Additional annotation of the remaining unidentified sequences was conducted using Blastn against NCBI refseq databases in order to obtain as many annotated sequences as possible.

Differential gene expression analysis and functional analysis

In addition to the samples that were sequenced as pools due to low RNA concentration, three sequence pools were made from the individually sequenced samples (2 control fish DAN, 2 infected fish DAN and 2 control fish R) to ensure a balanced number of pooled samples per treatment (i.e. 2 pools for each one of the DAN and R infected and control groups, each pool consisting of 2 and 3 individuals respectively). As the number of reads were similar between pairs of individuals to be pooled (Table S1), we randomly down-sampled the sequence reads of the largest library of each pair to the size of the smallest one, using the functions cut (to downsize) and cat (to pool). Bowtie2 v.2.0.2 (Langmead and Salzberg 2012) was used to align reads from all pools against the final transcriptome assembly, using the –K 1 parameter to report a single best hit for each read. Following alignment, read counts for each transcript were generated using idxstats in Samtools v.1.2, (Li, et al. 2009).

The EdgeR package (Robinson, et al. 2010) was used to calculate significant differences in gene expression between infected and non-infected groups of fish from the two different selfing lines (DAN and R) using treatment and line as factors, via a quasi-likelihood negative binomial generalized log-linear model (glmQLFTest). Only transcripts with >4 reads were considered and prior to the analysis tagwise dispersion was used to moderate the degree of over-dispersion amongst transcripts using the recommended prior.df, of 10. Transcripts with
FDR < 0.05 and P < 0.001 were considered to be differentially expressed. In addition, a Gene ontology (GO) enrichment analysis was completed for differentially expressed transcripts, using the Database for Annotation, Visualization and Integrated Discovery, version 6.8 (DAVID) (Dennis, et al. 2003).

**RT-qPCR analysis**

To further assess the role of individual variation versus genetic background in the gene expression in response to infection, we selected a group of immune-related genes from the transcriptome analyses which were analysed using real time quantitative PCR (RT-qPCR) in 48 fish (24 DAN and 24 R, including 9 of the fish sequenced for RNA-seq, of which there were 9 controls and 15 infected fish from each line). Four immune-related genes were selected among those differentially expressed (MHC I-uka, MHC II-dab, CD4-1 and CXCL 11.8; FDR<0.05 and P<0.0001) and five more based on P<0.05 (LECT2, C7) or fold change >2 (AHSA1B, FGG and IRGFl) (Supplementary material 1, Tables S2-S3). 18S rRNA and EF1a were used as reference genes following previous work (Olsvik, et al. 2005; Rhee, et al. 2010; Small, et al. 2008). Specific primers for immune targets were designed using NCBI primer-BLAST (Ye, et al. 2012), followed by Beacon designer (ver. 2.1, PREMIER Biosoft) to check for the absence of secondary structure. For this analysis RNA from 48 fish (15 infected fish from each line and 9 control fish from each line) was extracted as described earlier.. The concentration and purity of RNA in each sample was determined using a Qubit™ fluorometer and a NanoDrop 2000 Spectrophotometer. Total RNA (2 µg) was first treated with DNase (Promega), then reverse transcribed with GoScript™ Reverse Transcriptase (Promega, Medison, WI) using 10 µM random hexamer primers (MWG-Biotech).

All primer optimization and amplification reactions were completed using 5µl SYBR green Supermix (Biorad), 3.5µl water, 0.25 µl forward primer (10 uM), 0.25 µl reverse primer
(10 uM) and 1 µl cDNA per sample. A CFX96 Touch™ Real-Time PCR Detection System (Biorad) was used to run samples using the following protocol: 95°C for 10 min; 40 cycles of 95°C for 10 s, 60°C for 45 s; 95°C for 1 min; 55°C for 1 min; 80 cycles starting at 55°C for 10 s with a melting curve program of 55–95°C and a heating rate of 0.5°C every 10 s. The annealing temperature was adjusted accordingly to optimise primer efficiency for each target gene (Supplementary material 1, Table S4). PCR efficiencies (E=10^{1/slope}) for each primer pair were derived from standard curves (mean quantification cycle (Cq) vs. log cDNA dilution) using a two or ten-fold dilution series with pooled cDNA. All optimised primer pairs had efficiency values between 89.5-119.2, and standard curve R^2 values > 0.95 (Table S4). Melt curve analysis confirmed the specific amplification of a single PCR products in each case. Following primer optimisation all samples (diluted at 1:2) were run in triplicate for each gene in accordance with the sample maximization method (Hellemans, et al. 2007). Samples displaying non-specific amplification (contaminant melt peaks) or high variation in Cq values between technical replicates (standard deviation >1.0) were removed from the analysis.

BestKeeper (Pfaffl, et al. 2004) was used to estimate the stability of the reference genes. Mean Cq values were extracted for all samples and the relative expression of each gene was calculated using the comparative 2^{-ΔΔCt} method including gene specific efficiency correction (Pfaffl 2001), and normalising to the geometric mean for reference genes (Hellemans, et al. 2007; Vandesompele, et al. 2002).

Data analysis

Differences in individual gene expression between groups (fish from different lines and treatment) were initially assessed with a generalised linear model using the function glm in R version 3.4.0 and a gaussian link function. The full model included infection status (yes/no) and genetic background (line) as fixed factors plus their interactions. Model selection was
carried out based on AIC (Akaike Information Criterion) and log-likelihood (LR) ratio analyses. Generalised Linear Mixed-effects models (GLMM) of the combined gene expression for the target genes were fitted using the \textit{lmer} function in the R package \textit{lme4} (Bates, et al. 2014) using individuals as random effects. Comparisons between models with and without random factors were carried by AIC comparisons with respect to the GLMM fitted by Maximum Likelihood. Variances between groups were compared using the Fligner-Killeen test of homogeneity of variances (Conover, et al. 1981). All analyses were run on R version 3.4.0 (R Core Team 2014).

Results

Transcriptome comparison between treatments

We assembled a \textit{de novo} transcriptome of \textit{Kryptolebias marmoratus} from 294 million Illumina RNA-sequencing reads derived from the 20 fish from two different inbred lines (DAN and R) subject to two treatments (infection with \textit{Argulus foliaceus} and non-infected control). The final assembly consisted of 291,771 transcripts with an average length of 1,136 bp and a N50 of 2,575 bp. Annotation against Ensemble databases resulted in a total of 67,822 annotated transcripts. Edge R analysis identified 276 differentially expressed genes between groups (FDR < 0.05; P<0.001); (Supplementary material 1, Table S3). Most of the differences were due to line (205 genes were DE when only differences between lines were considered). MDS grouping and a heat map indicated greater differentiation between lines than between treatments (Figure 1 and Supplementary material 2, Figure S1). The differentially expressed genes included several immune-related genes (\textit{C7, MHC I-uka, MHC II-dab, CD4-1}) as well as genes involved in response to bacteria (\textit{LECT2}) and inflammation (\textit{CXCL11.8}).

Functional analysis of treatment group differences
Functional analysis of differentially expressed gene revealed 23 enriched GO terms ($P<0.1$) (Supplementary material, Figure S2) and 6 annotation clusters, thirteen of which were involved in immune response (Level: GO ALL). In addition to genes directly involved in immune response, many genes involved in membrane transfer and GTPase activity were highly enriched. Functional classification confirmed that the major enriched gene families included the Major Histocompatibility Complex and Immunoglobulin families (Supplementary material 1, Tables S5-S6).

**Individual gene analyses**

The stability of the reference genes was estimated by their standard deviation (SD) and the correlation with the BestKeeper index. SD was 1.51 for $18S$ rRNA and 1.50 for $EF1a$, both were highly correlated (Pearson $r=0.944$, $P=0.001$) and displayed high correlation with the BestKeeper index ($18S$ rRNA: $r=0.986$ $P=0.001$; $EF1a$: $r=0.986$ $P=0.001$). Samples were tested for their expression stability using the intrinsic variance of expression (InVar) implemented in BestKeeper and five of them (2 DAN and 3 R) were excluded from the rest of the analyses based on their high (>3) overexpression values, as recommended in (Pfaffl, et al. 2004). Of the nine immune-related genes analysed individually using RT-qPCR ($MHC$ I-uka, $MHC$ II-dab, $FGG$, $IRGF1$, $C7$, $CXCL11.8$, $CD4-1$, $LECT2$ and $AHSA1B$), five displayed significant differences either in expression between lines ($MHC$ II-dab $P=0.013$; $IRFG1$ $P=0.037$), infection status ($FGG$ $P=0.034$), or both ($CD4-1$ line $P=0.048$, infection $P=0.011$; $CXCL11.8$, line $P=0.0025$, infection $P=0.016$) (Figure 2). There were no significant interactions between the genetic background (line) and infection status in any case. The role of the individual immune response versus the genetic background (line) was estimated by analysing the expression of all the nine target genes and of the five differentially expressed (DE) genes in
relation to infection and line. Comparison among four models including and excluding interactions between line and infection, two of them including individual (ID) as random factor, indicated that the model which included infection, line and ID (but no interactions) provided the best fit to the data \( \text{lmer(Gexpress~Line+Infect+(1|ID))} \) when all the 9 candidate genes were considered \( (\text{AIC}=1045.6; \text{Chisq}=3.8410 \text{ df}=1 \text{ P}=0.05) \), although considering individual (ID) as random factor did not improve the fit over the most basic model \( \text{lmer(Gexpress~Line+Infect}) \) when only the 5 DE genes were considered (Supplementary material 2, Figure S5).

Six of nine genes \( (\text{IRGF1, CD4-1, MHC II-dab, AHSA1b, CXCL11.8 and LECT2}) \) displayed significant differences in variance between lines (Supplementary material 2, Figure S3). Of these, \textit{MHC II-dab} and \textit{CXCL11.8} additionally displayed significant differences in variance between treatments (infected versus control). In addition, \textit{FGG} displayed significant differences only between treatments (Table 1). For all nine genes, the variance was higher within infected than control groups.

**Discussion**

Evidence of the role of genotype-specific resistance and infectivity in maintaining the dynamics of the host-parasite arms race has been accumulating for both plants and animals over the last decades (Carius, et al. 2001; Rodenburg, et al. 2017; Srivastava, et al. 2009). However, the importance of transcriptome variation in response to specific interactions between host and parasite genotypes has only received attention recently (Barribeau, et al. 2014). Specific transcriptome responses to parasite infection can be difficult to identify in natural populations where there is high individual genetic variability. By using naturally inbred (selfed) fish reared under controlled conditions, we have been able to estimate significant
differences in transcript expression between genotypes, as well as variance differences in gene expression which were significant between lines and treatments.

The parasite we used (an argulid) induces a consistent innate response with the addition of an adaptive response approximately 7-10 days’ post-infection (Stewart, et al. 2017). We found a number of differentially expressed genes, with higher levels of differentiation between selfing lines than between treatments (infected and controls fish). In addition, we examined several immune-related genes that displayed differential expression between lines and/or infection status, selected based on FDR< 0.05 and P<0.05. In addition, as we had no a-priori assumption of the number of DE genes that we would find in relation to infection and/or line differences, we used a threshold of fold change>2, which is often considered as a significant change in the expression of immune related genes in fish in response to infection (e.g. El Aamri, et al. 2015; Polinski, et al. 2014; Purcell, et al. 2004), as the limit below which we were not further investigating a particular gene. Five of the nine genes selected on this basis were found differentially expressed in the qPCR analyses either between lines, treatments or both. Of them, CXCL11.8, a chemokine involved in regulating cell trafficking of leukocytes that can be critical for the recruitment of immune cells to the sites of infection, has a dual role in immune response and normal physiological conditions (Alejo and Tafalla 2011). It was targeted for qPCR analysis based on its differential expression between lines (upregulated in DAN) and the individual analyses indicated that differential expression between control and infected groups and between lines, as well as higher level of individual variation among DAN than R individuals, supporting the results from the transcriptome. MHC class II-dab was also differentially expressed between lines (upregulated in R) based on EdgeR analyses, and this was confirmed on the individual qPCR analysis, which also indicated differences in variance between line and treatment groups. MHC II molecules are expressed on antigen presenting cells which have direct functional relevance to teleost immune responses (Chen, et al. 2006;
Hofmann, et al. 2017; Janeway, et al. 1997) and have been identified in various teleost species (Grimholt, et al. 2000; Hofmann, et al. 2017; Pang, et al. 2013) with very high allelic diversity maintained by natural and sexual selection (Aguilar and Garza 2007; Consuegra and Garcia de Leaniz 2008), which could result in allele specific differences in gene expression. MH C II expression in K. marmoratus was consistent with that observed in Atlantic salmon when infected with lice from the genus Lepeophtheirus (Fast, et al. 2006). CD4 binds to MH C II molecules on the surface of dendritic cells which are important for antigen presentation (Leahy 1995; Yoon, et al. 2015). In contrast, and despite its high degree of polymorphism and the differential expression observed in the transcriptome analysis, MH C class I gene expression and variance did not differ between infected R and DAN individuals (although R fish tended to have higher expression of this gene, in similar to that observed at the transcriptome level). MH C I is responsible for the presentation of intracellularly derived antigens to the TCR/CD8 complex of cells and, as for MH C II, its high polymorphism is likely to be maintained by natural and, potentially, sexual selection (Ellison, et al. 2012; Ellison, et al. 2013). As for MH C II-dab, we found that CD4-1 displayed differential expression between lines and between control and infected groups based on transcriptome analyses (more expressed in DAN and infected individuals). Individual qPCR confirmed differences in expression between lines and between infected and control fish, albeit in the opposite direction (overexpressed in R). The difference between analyses could be the result of individual variation, as potentially indicated by the significant differences in variance between lines.

FGG (one of the genes encoding the peptide chains of fibrinogen) is important for the formation of clots which can be relevant for the response to external parasites (Vo, et al. 2013). Although not in the original list of DE genes, we targeted it based on having fold change >2 and in the individual analysis we found it differentially expressed between control and infected groups in both lines. Individual variation was much higher in infected individuals than in
control specimens but did not differ significantly between lines. In contrast, IRGF1 (an immunity-related GTPase) displayed high variation in amplification between lines, both at the transcriptome and qPCR levels. IRGF1 only amplified in 22% of R control specimens and 12% of R infected specimens, with very low levels of expression in all cases and displayed significantly larger individual variance in infected R individuals than DAN.

C7 and LECT2 play a role in the complement activation in fish (C7; (Guo, et al. 2016) and in the immune regulation response to bacterial infection (LECT2; (Chen, et al. 2009; Lin, et al. 2007; Lu, et al. 2013)) and were listed as DE with P<0.05 in the transcriptome analysis. Both had higher number of counts in the transcriptome of DAN fish compared to R and displayed higher variance in infected R individuals when compared with DAN in the individual analyses, although the differences were not significant. No differences were observed either in the expression or variance of AHSA1B, involved in the regulation of cell growth and apoptosis (Shao et al. 2016), which had been selected on the basis of a fold change >2.

In general the results of the transcriptome and the individual analyses displayed a good agreement, with 5 of the 9 genes identified as DE in the transcriptome being confirmed at the individual level (all but one in the same direction). Comparing control and exposed fish was critical to be able to interpret the potential genetic or individually-linked differences in gene expression in response to infection. Our transcriptome results indicate that the differences in gene expression were larger between selfing lines than between infected and control fish, and this pattern was supported by the targeted approach looking at selected immune-related genes, suggesting that there could be a genotype-related pattern of gene expression, similar to the one described in bumble bees (Barribeau, et al. 2014). Natural variation in gene expression has also been observed in populations of contrasting Fundulus species, where 18% of 900 genes displayed significant differences in expression among wild individuals within populations (Oleksiak, et al. 2002). Individual patterns of gene expression in humans also display
differences among individuals that can be as large as those comparing humans and chimpanzees (Enard, et al. 2002). Individual variation in gene expression in human blood seems to be variable among genes, but crucially several of the genes identified as having high intrinsic variation are immune-related genes with high polymorphism (e.g. MHC II genes) (Enard, et al. 2002), suggesting that genotype variation can be (at least in part) responsible for the differences in gene expression. Microsatellite analyses had previously indicated that fish from the R line were genetically identical and homozygous at 28 of 29 microsatellites, whereas fish from the line DAN formed three different homozygous genotypic groups separated by variation at only one microsatellite locus (Ellison, et al. 2013). Both DAN and R selfing lines displayed significant differences in the degree of variance in gene expression for the 9 target genes, and the variance was consistently larger among infected than among control individuals, suggesting an individual component of the immune response, despite the genetic homogeneity of the individuals from each line. Variation in the regulatory regions of the DNA that affect gene expression (e.g. transcriptional regulatory sequences) result in individual variation in expression patterns (Handunnetthi, et al. 2010), which can be targets of selection and play an important role in adaptation. This could explain the differences in variance observed between lines, despite their high homozygosity and highly inbred condition. Yet, it does not fully explain the individual variation in gene expression observed within each experimental line, where individuals were genetically homogeneous, offspring the same age from the same parent and reared in a common environment, and we cannot rule out that other mechanisms, such as gene expression stochasticity, could be involved. The model that best predicted the gene expression patterns included individual as random factor, albeit only when all the nine target genes were included, suggesting that not only the genotype, but also intrinsic factors could be involved in the immune-related gene response to infection. Stochasticity in gene expression has been observed for example in isogenic bacteria subject to identical environmental
conditions, resulting in phenotypic differences (Thattai et al. 2004). As for stochasticity among cells from a single organism (Kaern et al. 2005), in whole individuals this could represent a mechanism which provides flexibility to survive in fluctuating environments.

Whether regulated by genotype variation or not, our study suggests that individual differences in gene expression can also be important for determining the response to parasites. Given that the response to parasitic infection cannot normally be explained by genotype alone, we suggest that the naturally inbred mangrove killifish is an ideal model species to further investigate transcriptomic responses to vertebrate infection and their regulation mechanisms.

**Declarations**

*Ethics approval and consent to participate*

All the experiments in this study have been conducted following Home Office regulations, approved by both Swansea and Cardiff University Ethics Committees and under Home Office licence number PPL 302357.

*Competing interests*

The authors declare that they have no competing interests.

*Funding*

This work was supported by the Natural Environment Research Council through a NERC Industrial CASE studentship reference NE/L00948X. SC, JC, CGL and TUW were partially supported by the Welsh Government and Higher Education Funding Council for Wales (HEFCW) through the Sêr Cymru National Research Network for Low Carbon Energy and Environment (NRN-LCEE).

*Acknowledgements*
We are grateful to Alex Stewart for helping with parasite rearing, data collection and immunology insights and to Chloe Robinson for helping with the experimental infection and data collection.
Figure legends

**Figure 1.** Multidimensional scaling (MDS) plot of transcriptome analyses of four treatment groups of *Kryptolebias marmoratus* (R control: Light green, R infected: dark green, DAN control: light blue, DAN infected: dark blue).

**Figure 2.** Relative differences in expression (calculated using the comparative 2-ΔΔCt method and analysed using a Generalised Linear Mixed-effects model) of 9 target immune-related genes. Four groups were analysed using RT-qPCR: two lines- R (green) and DAN (blue), as well as two treatments- infected (dark) and control (light).
References


Chen J, Yang HY, Shi YH, Li MY 2009. An interaction between leukocyte cell-derived chemotaxin 2 and transferrin of ayu, Plecoglossus altivelis. Fish Shellfish Immunol 26: 536-542. doi: http://dx.doi.org/10.1016/j.fsi.2009.02.010


Table 1. Variance (Var) in gene expression compared between the two lines of *Kryptolebias marmoratus* (R and DAN) and treatments (infection and control) for each one of 9 immune-related target genes, analysed using RT-qPCR. Significant differences in italics (P<0.05). The target genes chosen serve important immune functions including antigen presentation (*MHC I*-uka, *MHC II*-dab, *CD4*-1), membrane regulation and attack (*C7* and *IRGF1*), clot formation (*FGG*), immune regulation and cell trafficking (*LECT2* and *CXCL11.8*) and the ATPase activity of heat shock proteins (*AHSAb1*). Variances between groups were compared using the Fligner-Killeen test of homogeneity of variances (Conover, et al. 1981).

<table>
<thead>
<tr>
<th></th>
<th>MHC I</th>
<th>IRGF1</th>
<th>CD4-1</th>
<th>FGG</th>
<th>C7</th>
<th>MHC II</th>
<th>AHSAb1</th>
<th>CXCL11.8</th>
<th>LECT2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var line DAN</td>
<td>0.9531</td>
<td>0.1729</td>
<td>0.0003</td>
<td>2.6146</td>
<td>0.0158</td>
<td>0.0019</td>
<td>0.0005</td>
<td>0.0337</td>
<td>0.0007</td>
</tr>
<tr>
<td>Var line R</td>
<td>0.7725</td>
<td>0.0000</td>
<td>0.0012</td>
<td>3.9938</td>
<td>0.0914</td>
<td>0.0132</td>
<td>0.0016</td>
<td>0.0023</td>
<td>0.0000</td>
</tr>
<tr>
<td>P (between lines)</td>
<td>0.072</td>
<td>3.32e-07</td>
<td>0.012</td>
<td>0.627</td>
<td>0.057</td>
<td>8.61e-05</td>
<td>0.038</td>
<td>0.018</td>
<td>0.008</td>
</tr>
<tr>
<td>Var Control</td>
<td>0.2302</td>
<td>0.0418</td>
<td>0.0002</td>
<td>0.4798</td>
<td>0.0052</td>
<td>0.0059</td>
<td>0.0002</td>
<td>0.0003</td>
<td>0.0001</td>
</tr>
<tr>
<td>Var Infection</td>
<td>1.2466</td>
<td>0.1306</td>
<td>0.0010</td>
<td>4.2703</td>
<td>0.0755</td>
<td>0.0095</td>
<td>0.0014</td>
<td>0.0279</td>
<td>0.0006</td>
</tr>
<tr>
<td>P (between treatments)</td>
<td>0.096</td>
<td>0.109</td>
<td>0.154</td>
<td>0.011</td>
<td>0.102</td>
<td>0.038</td>
<td>0.769</td>
<td>0.024</td>
<td>0.447</td>
</tr>
</tbody>
</table>
Figure 1.
Figure 2.

MHC II dab

CD4

IRGF1

LECT2

C7

CXCL11

MHC I uka

AHSA1B

FGG

* * *

** **

Normalised transcript concentration

Normalised transcript concentration