

Expressed Sequences and Polymorphisms in Rohu Carp (*Labeo rohita*, Hamilton) Revealed by mRNA-seq

Nicholas Robinson · Pramoda K. Sahoo ·
Matthew Baranski · Kanta Das Mahapatra ·
Jatindra N. Saha · Sweta Das · Yashowant Mishra ·
Paramananda Das · Hirak K. Barman ·
Ambekar E. Eknath

Received: 27 October 2011 / Accepted: 23 December 2011 / Published online: 2 February 2012
© Springer Science+Business Media, LLC 2012

Abstract Expressed genes and polymorphisms were identified in lines of rohu *Labeo rohita* selected for resistance or susceptibility to *Aeromonas hydrophila*, an important bacterial pathogen causing aeromoniasis. All animals were grown in a common environment and RNA from ten individuals from each line pooled for Illumina mRNA-seq. De novo transcriptome assembly produced 137,629 contigs with 40× average coverage. Forty-four percent of the assembled sequences were annotated with gene names and ontology terms. Of these, 3,419 were assigned biological process terms related to “stress response” and 1,939 “immune system”. Twenty-six contigs containing

38 single nucleotide polymorphisms (SNPs) were found to map to the *Cyprinus carpio* mitochondrial genome and over 26,000 putative SNPs and 1,700 microsatellite loci were detected. Seventeen percent of the 100 transcripts with coverage data most indicative of higher-fold expression (>5.6 fold) in the resistant line pool showed homology to major histocompatibility (MH), heat shock proteins (HSP) 30, 70 and 90, glycoproteins or serum lectin genes with putative functions affecting immune response. Forty-one percent of these 100 transcripts showed no or low homology to known genes. Of the SNPs identified, 96 showing the highest allele frequency differences between susceptible and

Electronic supplementary material The online version of this article (doi:10.1007/s10126-012-9433-8) contains supplementary material, which is available to authorized users.

N. Robinson · M. Baranski
Nofima,
PO Box 5010, 1432 Ås, Norway

M. Baranski
e-mail: matthew.baranski@nofima.no

P. K. Sahoo · K. D. Mahapatra · J. N. Saha · S. Das · Y. Mishra ·
P. Das · H. K. Barman · A. E. Eknath
Central Institute of Freshwater Aquaculture,
Kausalyaganga,
Bhubaneswar 751 002, India

P. K. Sahoo
e-mail: pksahoo1@hotmail.com

K. D. Mahapatra
e-mail: kdmahapatra@yahoo.co.in

J. N. Saha
e-mail: jnsaha2000@yahoo.co.in

S. Das
e-mail: swetabt@gmail.com

Y. Mishra
e-mail: yashowant@gmail.com

P. Das
e-mail: pdas77@hotmail.com

H. K. Barman
e-mail: hkbarman68@hotmail.com

A. E. Eknath
e-mail: ambekar.eknath@yahoo.co.in

N. Robinson
Flinders University,
Sturt Road, Bedford Park,
Adelaide, SA 5042, Australia

N. Robinson (✉)
31 Hutchinson Avenue, Warrandyte,
Melbourne VIC 3113, Australia
e-mail: nick.robinson@nofima.no

resistant line fish included transcripts with homology to MH class I and galactoside-binding soluble lectin, also with putative functions affecting innate and acquired immune response. A comprehensive sequence resource for *L. rohita*, including annotated microsatellites and SNPs from a mixture of *A. hydrophila*-susceptible and -resistant individuals, was created for subsequent experiments aiming to identify genes associated with *A. hydrophila* resistance.

Keywords *Labeo rohita* · Transcriptome expression · Single-nucleotide polymorphism · Disease resistance · *Aeromonas hydrophila*

Introduction

Next-generation sequencing (NGS), utilising massively parallel sequencing technologies, makes it feasible and efficient to identify marker polymorphisms and differences in gene transcription levels in species for which there is little existing genetic data. Sequencing-based RNA analysis records the numerical frequency of a sequence in the library population, eliminating background signals observed using relative expression profiles generated with microarray hybridization technology, to give a measure of gene expression incorporating both polymorphism and gene expression data in a single step. By taking a more integrative approach, it could also be possible to gain a better understanding of the regulatory polymorphisms, or eQTLs, in *cis* and *trans* elements/factors associated with gene expression differences (Gibson and Weir 2005; Babak et al. 2010), but large datasets showing individual variation in polymorphisms and expression levels are needed because of the high sampling error associated with such data (Robinson et al. 2008).

Carp are the world's most cultured fish. The production of rohu carp (*Labeo rohita*, Hamilton) in 2008 was more than 1.1 million tonnes, mostly in India (Food and Agriculture Organisation of the United Nations, 6 June 2011, http://www.fao.org/fishery/culturedspecies/Labeo_rohita/en#tcNA0112). *L. rohita* is an important source of animal protein in India and Bangladesh and is efficiently grown in earthen pond systems. However, earthen ponds lend themselves to limited possibilities for disease control; many diseases of *L. rohita* are endemic, and mortality and growth loss from disease are likely to have a large effect on total production in India. *Aeromonas hydrophila* is a motile bacteria that becomes more abundant with higher organic loads in ponds and has increased pathogenicity when fish are under stress (reviewed by Jeney and Jeney 1995). Clinical signs of the disease include haemorrhages and ulcers. The disease is difficult to control and treat because little can be done to reduce stressors in the pond environment and there are no effective drugs or vaccines. Disease caused

by *A. hydrophila* affects numerous fish species, and as *A. hydrophila* is ubiquitous and has a cosmopolitan distribution, this disease is a major problem for aquaculture worldwide (reviewed by Harikrishnan and Balasundaram (2005)).

A selective breeding programme with focus on growth rate has been carried out for *L. rohita* since 1992 by the Central Institute of Freshwater Aquaculture in Orissa, India (Gjerde et al. 2005). Significant additive genetic variation in survival to *A. hydrophila* with experimental challenge tests has been found (Das Mahapatra et al. 2008). Rohu that are resistant to *A. hydrophila* have been shown to have higher serum bactericidal activity, ceruloplasmin levels, myeloperoxidase activity and superoxide production in blood phagocytes than susceptible rohu (Sahoo et al. 2004). Selective breeding based on survival analysis to experimental challenge testing is being used to improve resistance (Das Mahapatra et al. 2009), but a major limitation is the inability to directly test highly valuable potential broodstock.

DNA markers linked to quantitative trait loci affecting disease resistance could potentially be used with selective breeding to speed up the genetic improvement of this trait, but no such tests exist for disease resistance in *L. rohita*, and the genomic resources needed to develop them are lacking in this species. Polymorphisms associated with resistance to specific diseases have been detected in other teleosts (e.g. Moen et al. 2009; Heng et al. 2011; Houston et al. 2006). The gene expression response of carp and other teleosts to challenge with specific diseases and stress has been investigated (e.g. Cossins et al. 2006; Gracey et al. 2004; Nayak et al. 2011; Millán et al. 2011), but few studies have compared gene expression in resistant and susceptible fish (Zhang et al. 2011; Aoki et al. 2011). Recent studies have used RNA sequencing (RNA-seq) as an alternative to microarrays to compare the transcriptomics of different genetic types (e.g. Jeukens et al. 2011; Yang et al. 2011; Goetz et al. 2010), to map *cis*-eQTL (Babak et al. 2010) and to study the response of animals to varying environmental or experimental conditions (e.g. Meyer et al. 2011). Measurement of fold expression differences using RNA-seq has been shown to be accurate (high correlation to measurement using qPCR; Wang et al. 2009; Meyer et al. 2011) and to result in a more comprehensive understanding of transcriptional differences (e.g. Jeukens et al. 2011) than other methods such as those using microarrayed expressed sequence. Here, we use mRNA-seq on pools of liver, intestine, muscle, kidney, spleen and brain samples from aeromoniasis-resistant and -susceptible lines of *L. rohita* to develop a large, gene-associated single nucleotide polymorphism and microsatellite resource that can be used to scan the *L. rohita* genome for quantitative trait loci and gene expression differences associated with resistance to *A. hydrophila* and other traits of interest.

Materials and Methods

Fish

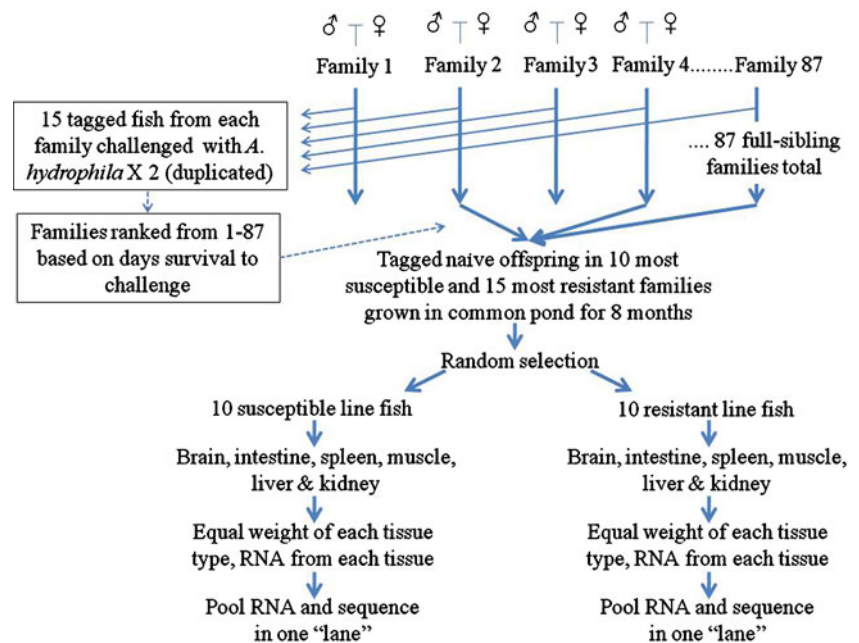
The founding parents of the individuals used for sequencing were 2003 and 2004 year class rohu that had been selectively bred for five generations for higher growth (Das Mahapatra et al. 2008) at the Central Institute of Freshwater Aquaculture, Kausalyaganga, Bhubaneswar, India. Figure 1 shows the strategy used to select fish and tissues for sequencing. Resistant and susceptible families of fish were selected after intra-peritoneal challenge of 87 full sibling families (30 sibs per family, challenged in duplicate batches containing 15 sibs per family) to a virulent strain of *A. hydrophila* (Sahoo et al. 2008). An LD₅₀ dose of 5×10^6 CFU/g body weight was injected intraperitoneally into each fish (Sahoo et al. 2008). Unchallenged individuals from the 15 highest-ranked families (those whose full siblings survived the longest after the disease challenge test) were selected as the first generation of the resistant line, whereas unchallenged individuals from ten lower-ranked families were selected as the founders of the susceptible line. All unchallenged progeny (resistant and susceptible lines) were reared in the same pond for 8 months after they were tagged at 10–15 g weight. Das Mahapatra et al. (2009) tested a sample of these fish and observed a 31.1% response to the first generation of selection (measured as the average difference in days of survival after challenge using the same strain of *A. hydrophila* with which the lines were generated). The unchallenged fish used in subsequent steps were examined for signs of parasitic infection (gross observations and kidney biopsy examination)

and screened for *Aeromonas* infection using two PCR tests (Xia et al. 2004; Kong et al. 2002).

Tissue Collection and Preparation of RNA Pools

Tissue samples were collected from ten resistant (*R*) and ten susceptible (*S*) line fish. Six different tissues (liver, intestine, muscle, kidney, spleen and brain) were obtained. Individual tissue samples were stored in RNAlater (Qiagen) at -30°C for storage until RNA was isolated. RNA pools were prepared in the following way. First, equal quantities (weight) of tissue were used to create 12 tissue mixtures from which RNA was purified, *S*-liver, *R*-liver, *S*-intestine, *R*-intestine, *S*-muscle, *R*-muscle, *S*-kidney, *R*-kidney, *S*-spleen, *R*-spleen, *S*-brain and *R*-brain, where *S* means susceptible line and *R* means resistant line. Next, after RNA extraction of the 12 tissue mixtures, equal quantities of purified RNA were mixed to create a single susceptible line pool and a single resistant line pool (Fig. 1). RNA was extracted (TRIzol® Plus RNA Purification, Invitrogen) according to the manufacturer's instructions. Tissue was ground in liquid nitrogen and 1.5 ml of TRIzol was added. RNA was further purified by chloroform extraction followed by precipitation using isopropanol and one wash with 70% ethanol. RNA was resuspended in 25 μl of RNase-free water. Only good-quality RNA samples with optimal purity ($\text{OD}_{260/280} > 1.8$ and < 2.2 ; $\text{OD}_{260/230} > 1.0$ and < 2.4) and concentration (> 50 and $< 2,500$ ng/ μl), estimated using UV spectrophotometry (Nanodrop) and capillary electrophoresis (Agilent Bioanalyser), were utilised for sequencing. Samples were prepared using the Illumina protocol outlined in "Preparing Samples for Sequencing of mRNA" (Part# 1004898 Rev. A September 2008). mRNA

Fig. 1 Approach used to sample animals and tissues for sequencing



was purified from total RNA using magnetic oligo(dT) beads and fragmented using divalent cations under elevated temperature. cDNA was synthesised from the fragmented mRNA using Superscript II (Invitrogen, USA), followed by second-strand synthesis. cDNA fragment ends were repaired and phosphorylated using Klenow, T4 DNA Polymerase and T4 Polynucleotide Kinase. Next, an ‘A’ base was added to the 3’ end of the blunted fragments, followed by ligation of Illumina adapters via T-A mediated ligation. The ligated products were size-selected by gel purification and then PCR-amplified using Illumina primers. The library size and concentration were determined using capillary electrophoresis (Agilent Bioanalyzer, Agilent Technologies, USA). Fragments of 200 ± 25 base pairs were gel-purified for sequencing after adapter ligation, and a final enrichment of purified cDNA templates was performed using PCR as per the Illumina protocol. No RNA normalisation was performed. Fifty-four cycles of paired-end sequencing was performed on a Genome Analyser II (Illumina, USA) according to the manufacturer’s instructions.

Processing of the Raw Illumina GAI Data Files

Residual adapter and poor quality sequences (quality score threshold of 10) were trimmed from the raw data using the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/commandline.html), and reads shorter than 20 bp were discarded. Fastq files were converted to fasta files and *Trimest* (EMBOSS, <http://emboss.open-bio.org/wiki/Appdoc:Trimest>) used to trim poly A/T tails from the sequence.

Assembly and Annotation

CLC Assembly Cell Software (CLC bio, <http://www.clcbio.com/>) was used to perform a de novo transcriptome assembly on the entire dataset (combined sequence from both lanes) using the default parameters and a minimum contig size of 100. Sequence data were deposited in NCBI’s Gene Expression Omnibus (Edgar et al. 2002) and are accessible through GEO Series accession number GSE27994 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE27994>). Basic Local Alignment Tool applications (BLAST; Altschul

et al. 1997) and BLAST databases were downloaded from the National Centre for Biotechnology Information (NCBI, <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) and installed locally on a UNIX cluster. BLASTX was used to compare nucleotide sequences to the non-redundant protein database with a Karlin–Altschul cutoff E-score of 0.001. The ten best matches below the cutoff threshold were retrieved, and the results were imported into a Blast2GO database for gene ontology (GO) annotation (Götz et al. 2008).

Reads from the resistant and susceptible pools were aligned separately to the de novo assembled contigs using the programme ‘*clc_ref_assemble_short*’. The resulting pool-specific assemblies were used to find SNP allele frequency differences and calculate relative expression differences (based on average contig coverage) as described in “SNP Discovery, Validation and Allele Frequency Estimation” and “Transcript Coverage and Indicative Gene Expression”.

SNP Discovery, Validation and Allele Frequency Estimation

The ‘*find_variations*’ programme in CLC Assembly Cell was used to identify single nucleotide polymorphisms. A range of minimum minor allele frequency (0.05–0.15) and minor allele count (two to four) thresholds were tested. Minimum coverage was set to eight. SNP detection was performed in each lane, and the difference in allele frequency between the lanes was calculated and used to identify possible trait-associated SNPs.

Six groupings of putative SNPs, based on minor allele frequency, coverage, BLAST matches with other contigs (putative repeat sequence) and number of SNPs found in the same contig, were selected for validation testing (Table 1). Single nucleotide polymorphism genotyping was performed using the MassARRAY system from Sequenom (San Diego, CA, USA). PCR and extension primers were designed using the software spectrodesigner v3.0 (Sequenom). All SNP genotyping was performed according to the iPLEX protocol from Sequenom (available at <http://www.sequenom.com>). For allele separation, the Sequenom MassARRAY™ Analyzer (Autoflex mass spectrometer) was used. Genotypes were assigned in real time

Table 1 Selection of putative SNPs for validation testing. Combinations of high and low coverage and high and low minor allele frequency were tested together with the occurrence of putative repetitive sequence (many BLAST hits) and number of SNPs in the same contig

Validation group	Coverage	Minor allele frequency (%)	BLAST hits	Number of SNPs in contig
1	112–426	10–13	1–5	3–13
2	9	22	1	1
3	78–309	42–48	1–2	2–6
4	9–11	44–45	1	1
5	92–476	30–46	7–11	1
6	9–19	33–47	5–11	1

(Tang et al. 1999) by using MassARRAY SpectroTYPER rt v3.4 software (Sequenom) based on the mass peaks present. All results were manually inspected using the MassARRAY TyperAnalyzer v3.3 software (Sequenom). Several SNPs in contigs annotated as MHC I and other SNPs that were found to differ in allele frequency between the pools of resistant and susceptible line animals were also tested for validation. Overall, 54 SNPs were tested across 96 individual rohu samples.

Transcript Coverage and Indicative Gene Expression

We performed separate assemblies for the two pools using the de novo assembly of all sequences (“Assembly and Annotation”) as a reference. The programme ‘assembly_info’ in CLC Assembly Cell was used to calculate the average sequence coverage for each contig in resistant and susceptible pools. Coverage data were compared as an indicator of differential transcript expression between the resistant and susceptible pools. Quantile-based rank normalisation was used to correct for the differences between flow cells or pools (Bullard et al. 2010). Data were visually represented in Bland-Altman MA scale plots where $M = \log_2 R - \log_2 S$ and $A = 0.5 * (\log_2 R + \log_2 S)$ and where R and S were the average coverage depths for each contig in the resistant line pool and the susceptible line pool, respectively, using the Bioconductor package in the statistical R programming language (Gentleman et al. 2005). Spearman’s rank correlation was tested between sequence length and average gene expression between lanes (A), between sequence length and total gene expression across lanes and between sequence length and differential gene expression between lanes (M).

Quantitative Real-Time PCR Validation of Gene Expression

Total RNA was isolated from the skin, muscle, liver, kidney, spleen, gill, eye and brain of 30 resistant and 15 susceptible second-generation selected line rohu using a GeneJET RNA purification kit (Fermentas Life Sciences, as per the manufacturer’s instructions). The purity and quantity of RNA samples was checked using Nanodrop ND 1000 (Thermo Scientific, as per the manufacturer’s instructions). After DNase I (Fermentas Life Sciences) treatment, 1 µg total RNA was reverse-transcribed to single-strand cDNA using M-MLV reverse transcriptase (Sigma, USA) according to the manufacturer’s instructions. qRT-PCR was carried out on LightCycler 480 (Roche, Germany) using LightCycler 480 SYBR Green I Master reaction mix (Roche, Germany) according to the manufacturer’s instructions using a specific primer pair designed to amplify contig 88601 with homology to part of the major histocompatibility class I antigen (MH I, forward 5′ → 3′ primer sequence GAATGGAGACAAACGCAAAGA, reverse 5′ → 3′ primer

sequence TCCAGCATAGGCAAGATGAG). A stable housekeeping gene, β-actin (5′ → 3′ forward primer TTGGCAATGAGAGGTTTCAGGT and reverse primer TTGGCATAACAGGTCCTTACGG), was used as control. Pure cDNAs were diluted tenfold and 2 µl of diluted cDNA product was used as a PCR template. The qRT-PCR programme consisted of pre-denaturation at 95°C for 10 min and 45 cycles of amplification at 95°C for 10 s, 58°C for 10 s, and 72°C for 20 s. Each assay for resistant samples was performed simultaneously with susceptible samples in the same plate. All samples were run in duplicate and n -fold differential expression was calculated using the comparative C_t method (Livak and Schmittgen 2001) where n -fold differential expression = $2^{-\Delta\Delta C_t}$ and where ΔC_t was the mean C_t value (number of PCR cycles necessary to achieve a given level of fluorescence where exponential PCR amplification occurs) for MH I less the mean C_t value for β-actin over the duplicate measurements. For the purpose of our analysis, n -fold ≥ 2.0 was defined as increased mRNA expression, n -fold > 0.5 and < 2.0 as no differential expression and n -fold ≤ 0.5 as decreased mRNA expression.

qRT-PCR specificity was verified by melting curve analysis at a temperature of 95°C for 10 s, 65°C for 1 min and 95°C for 1 min. Percent amplification efficiency (E) was calculated using twofold serial dilutions of cDNA for each primer pair. Efficiency was defined as $E = -1 + 10^{(-1/\text{slope})}$ where slope is the slope of the standard curve for the serial dilutions. “No-template” controls were included on each run (no amplification of these controls was observed).

Mitochondrial Sequence and Polymorphisms

The complete common carp *Cyprinus carpio* mitochondrial genome (Accession X61010.1, Chang et al. 1994) from GenBank (NCBI) was used as a reference sequence for the de novo assembled rohu contigs using CLC Assembly Cell (CLC bio, <http://www.clcbio.com/>). SNPs mapping to the *L. rohita* mitochondrial genome were identified and mapped.

Microsatellite Loci

Contig sequences were searched using PrimerPro (<http://webdocs.cs.ualberta.ca/~yifeng/primerpro/>) which is a pipeline for microsatellite identification integrating the microsatellite identification tool MISA, primer design software Primer3 and NCBI BLAST software. First contig sequences were trimmed of distal stretches containing ambiguous bases (two in a 5′ 40-base-pair window stretch). Mono-, di-, tri-, tetra-, penta- and hexanucleotide sequences repeated more than 12, 8, 6, 6, 5 and 5 times, respectively, were captured as microsatellites with potential application for mapping and genome scans. Mononucleotide repeats of A or T, which often

occurred at the 3' or 5' end of the contigs, were removed from the list. The criteria used for primer design included a GC clamp of one base and an optimum melting temperature of 55°C (minimum 50°C and maximum 60°C).

Results

Raw Sequence Data Processing, Assembly and Annotation

Trimming of residual adaptor and poor-quality end sequence resulted in 36,352,981 reads with an average read length of 45 bp (Table 2). De novo assembly of the complete dataset resulted in 137,629 contigs with an average contig length of 207 bp (range 100–6,387 bp). Excluding contigs smaller than 200 bp resulted in a total of 40,596 contigs, with an average length of 380 bp. Average coverage across both lanes of data was approximately 40× (around 20× per lane). BLAST searches provided annotation for 47.4% of the contigs; most homology (> 90%) was found to be with genes in the zebra fish (*Danio rerio*) genome.

Gene ontology annotation showed that the expressed genes were involved in a diverse range of biological processes (ESM 1). Out of a total of 44,292 contigs with GO-terms, 3,419 (7.7%) were annotated with the biological process “stress response” while 1,939 (4.4%) were annotated with “immune system processes” (Table 3). Many of the contigs were implicated in broad metabolic, cellular and biological regulation processes (22,191, 30,356 and 16,547 contigs, respectively; ESM 1).

Differential Gene Expression Between Resistant and Susceptible Lines

M-values across all contigs had a median of 0.002 and interquartile range of 1.61 (Fig. 2). mRNA-seq coverage data indicated that up to ninefold up-regulation was occurring (contig 21365 with weak homology to a hypothetical protein in *D. rerio*), while one gene (contig 111051 with strong homology with an E-value of 1.3E-112 to a hypothetical protein described in *D. rerio*) showed more than tenfold down-regulation in the resistant line (12.7-fold differential expression). Of the 100 contigs with the highest indicative differential expression in the resistant line animals, six showed homology to putative heat shock protein genes (HSP 30, 90 and 70), seven to putative zona pellucida glycoprotein, one to serum lectin isoform 1 precursor and three to putative major histocompatibility (MH) class I antigen or heavy chain genes (ESM 2). The ten contigs with the highest indicative differential expression showed homology to a hypothetical protein in zebrafish, HSP 30 (three occurrences), choriogenin h, zona pellucida glycoprotein, serum lectin isoform 1 precursor and three contigs with no or low homology to known genes.

Table 2 Summary of Illumina GAI sequencing performed on cDNA pools of resistant and susceptible individuals of *L. rohita* and summary of contig assembly statistics

	Resistant (<i>R</i>)	Susceptible (<i>S</i>)	<i>R</i> + <i>S</i>
Before trimming			
Number of reads	22,517,180	20,671,468	43,188,648
Number of bases	1,215,927,720	1,116,259,272	2,332,186,992
Average read length	54	54	54
After trimming			
Number of reads	18,766,707	17,586,274	36,352,981
Number of bases	867,076,189	798,525,599	1,665,601,788
Average read length	46.20	45.41	45.82
After assembly			
Number of contigs			137,629
Total sites			28,513,006
Average contig length			207
Longest contig			6,387
No. of assembled reads	12,971,267	12,137,730	25,108,999
Average coverage	21	19	40

Resistant and susceptible ‘phenotypes’ were sequenced in independent lanes of an eight-lane flow cell. Column *R* + *S* represents the combined dataset, corresponding to both lanes. Results are shown before and after adaptor, poly-A trimming and quality filtering. De novo transcriptome assembly was performed using the combined dataset. Results of the ‘mapping assemblies’ for the detection of SNPs are also shown

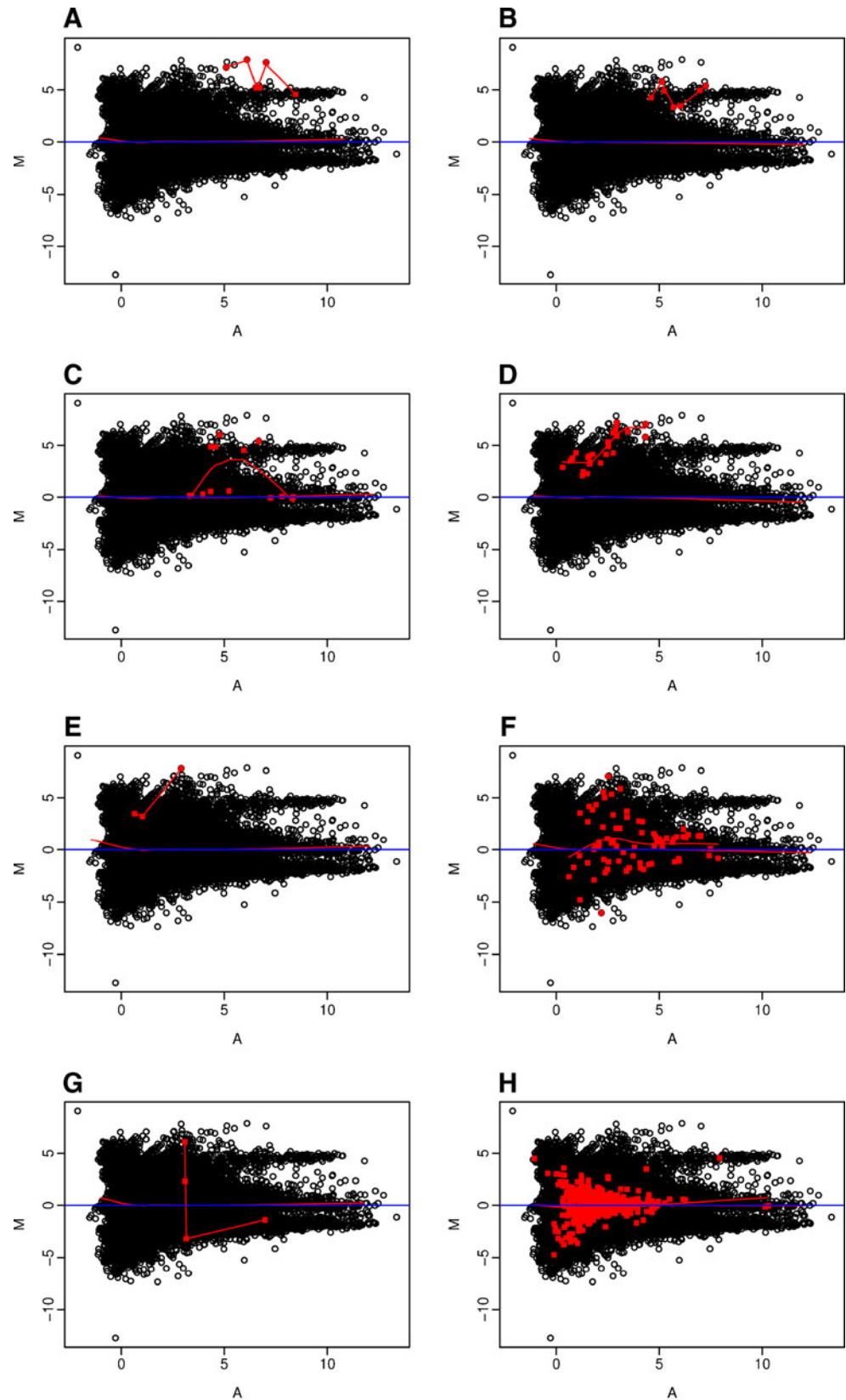
Contigs with homology to heat shock proteins 30, 90 and 70 showed indications of up-expression in the resistant line animals sampled (many more than fivefold; Fig. 2a–c). All the six contigs that showed homology to HSP30 exhibited more than fourfold up-regulation in the resistant line (median 6.17, inter quartile range 2.29; Fig. 2a); seven contigs showing homology to HSP 70 were up-regulated by more than threefold (median 4.89, inter quartile range 1.31; Fig. 2b), while of the 14 contigs with homology to HSP 90 (median 0.426, inter quartile range 4.69; Fig. 2c) five were indicative of more than 4.5-fold up-expression in the resistant line fish sampled and the remaining nine showed very little difference in expression between resistant and susceptible line fish sampled.

All 26 contigs with homology to zona pellucida glycoprotein showed indications of two- to sevenfold higher

Table 3 Examples of gene ontology annotation categories relevant to disease resistance for *L. rohita* contigs (a complete list of biological process categories is shown in ESM 1)

GO ID	GO term	Number of contigs
GO:0002376	Immune system process	1,939
GO:0022610	Biological adhesion	1,803
GO:0055114	Oxidation reduction	1,740
GO:0006950	Response to stress	3,419
GO:0006508	Proteolysis	2,962

Fig. 2 Bland–Altman MA plots of quantile normalised \log_2 transformed coverage data. Overall plot for the 137,629 assembled contigs, with the contigs highlighted showing homology to HSP30 (**a**), HSP70 (**b**) and HSP90 (**c**), zona pellucida glycoprotein (**d**), choriogenin (**e**), MH class I antigen (**f**), MH class I heavy chain (**g**) and ubiquitin (**h**)



expression in the resistant line fish sampled (Fig. 2d; median 4.13, interquartile range 2.45). The highest homology for

contig 11721 was for zona pellucida glycoprotein 2 (ZP2) in common carp (*C. carpio*, E-value 2.9E-18,

alignment length 57 bp). All three choriogenin contigs (contigs 89859, 75375 and 125023) showed indications of 3.5-fold higher expression in the resistant fish sampled on average than the susceptible fish sampled (interquartile range 2.3; Fig. 2e). One contig (89859 with homology to choriogenin h, and also with zona pellucida protein 2 in *Carassius auratus* and *C. carpio*, respectively) was indicated as more than sevenfold more highly expressed, while the other two were more than threefold.

Only one rohu contig (contig 90383) was found to show homology to the serum lectin isoform 1 precursor gene (E-value 4.5E-14), and mRNA-seq indicated that this contig was more than sevenfold up-regulated in the resistant line fish sampled.

mRNA-seq indicated that some contigs showing homology to MH class I antigens and heavy chain genes were more than fivefold down or up-regulated in the resistant line individuals (Fig. 2f, g, respectively). For the 76 contigs with homology to the MH class I antigen, mRNA-seq indicated that several contigs were more than fivefold up- or down-regulated (interquartile range 2.70; Fig. 2f). Of the four contigs showing homology to the MH class I heavy chain, mRNA-seq indicated that two were up-regulated by six- and fourfold while two were down-regulated by one- and threefold with an interquartile range of 5.08 (Fig. 2g).

An analysis of 1,263 contigs showing homology to genes involved in the T-cell receptor pathway (contigs identified using the annotation search terms ubiquitin, tyrosine kinase, phosphoprotein phosphatase, interferon, granulocyte, tumour necrosis factor, cyclin-dependant kinase 4 and interleukin) revealed little or no indications of differential expression from mRNA-seq coverage data (with median log expression ratios ranging from 0.0872 for ubiquitin to -0.403 for interleukin, interquartile ranges 1.0 to 1.40; Fig. 2h, ubiquitin).

qRT-PCR amplified MH I with 89% efficiency while β actin was amplified with 87% efficiency. All dissociation (melt) curves showed one single peak, which indicated that both qPCR products were likely to correspond to single amplicons. qRT-PCR revealed an average of more than fivefold higher expression of the MH I gene in intestine, gill and skin tissues in second-generation selected resistant line compared to second-generation selected susceptible line fish (Table 4). MH I was also more highly expressed in the brain of resistant line fish (more than fourfold), whereas no differential expression was detected in the other tissues (liver, kidney, spleen and muscle).

SNP Detection and Validation

Using a minimum minor allele count of three and a minimum minor allele frequency of 0.05, 33,730 SNPs and indels were detected. In total, 70% of the SNPs tested for validation were polymorphic (Table 5). SNPs with low minor allele frequency

Table 4 MH I (contig 88601) expression revealed by qRT-PCR in different tissues of susceptible and resistant lines of rohu

Tissue	$\Delta C_t \pm$ standard error		$\Delta\Delta C_t$	$2^{-\Delta\Delta C_t}$ /fold change
	Susceptible line	Resistant line		
Liver	14.4±2.1	14.0±1.5	-0.5	1.4
Kidney	11.8±1.6	10.8±1.3	-0.9	1.9
Intestine	15.1±2.7	12.7±2.1	-2.4	5.3
Spleen	14.6±2.3	18.1±4.2	3.5	0.1
Gill	12.1±1.8	9.3±1.4	-2.9	7.3
Brain	15.4±2.6	13.1±1.3	-2.2	4.7
Muscle	10.2±2.6	10.8±2.5	0.6	0.7
Skin	11.6±3.3	9.2±2.6	-2.4	5.1

were those that were most frequently monomorphic (e.g. group 1 and 2 SNPs with 43% and 13% polymorphic loci, respectively; Table 5). The incidence of monomorphism was also lower with higher coverage so that group 3 with high coverage and high minor allele frequency SNPs showed the highest level of polymorphism (100%; Table 5).

SNP Allele Frequencies in the Susceptible and Resistant Line Fish Sampled

Fixed allele differences between the individuals sampled for the resistant and susceptible line pools were detected at 96 SNP loci (allele frequency differences of 1.0), while 104 SNP loci contained allele frequency differences >0.95 between the pools (ESM 3). Of the SNPs, those showing large allele frequency differences between pools were at

Table 5 Percentage of SNP loci validated as polymorphic (the validation group criteria are shown in Table 1)

Validation group	Number of SNP loci tested	% polymorphic	Contig numbers
1	7	43	110199, 110696, 111522, 111795, 111877, 111919, 132709
2	8	13	67015, 68065, 59391, 59407, 63199, 96623, 115991, 83318
3	6	100	53381, 53451, 53504, 54210, 54608(1), 54608(2)
4	7	86	54150, 57982, 58395, 63206, 65328, 67476, 70886
5	8	88	110212, 110463, 132633, 52671, 54095, 55955, 87872, 87988
6	7	86	54666, 92876, 58602, 60942, 57930, 73092, 110141

position 59 in contig 11876, 139 in contig 89274, 35 in contig 89352 and positions 13 and 22 in contig 99704, all of which showed strong homology to the MH class I antigen (ESM 3). All susceptible animals included in the pool contained one MH class I antigen haplotype, while all resistant line animals contained an alternative haplotype for these five SNPs. Six SNPs in contigs with homology to the galactoside-binding soluble lectin 9 gene also showed fixed allele frequency differences between animals sampled from the susceptible and resistant line pools (allele frequency differences of 1.0 for positions 73, 207, 90 and 95 in contig 88103 and positions 90 and 124 in contig 115438; ESM 3).

L. rohita Mitochondrial DNA Transcriptome

Twenty-six contigs were found to map to the *C. carpio* mitochondrial DNA sequence in GenBank (GenBank X61010.1). Thirty-eight SNPs were detected in these 26 contigs (ESM 4; Fig. 3). Coverage of the *L. rohita* mitochondrial transcriptome was patchy, but it was clear that all 15 protein coding genes were expressed in the tissues studied. Where SNPs were found to occur, 374× average coverage was found within resistant and susceptible lanes (minimum of 2× for NADH1 contig_110148 and maximum of 3,719× for NADH2 contig_111443; ESM 4).

Microsatellite Loci

A total of 1713 microsatellite containing contigs were discovered (1,784 microsatellites, of which 45% were mono-, 35% di-, 18% tri-, 1% tetra-, 1% penta- and 0.5% hexa-nucleotide repeats; ESM 5). Seven hundred eighty-eight were either A or T mononucleotide repeats which were excluded from analysis. Seventy contigs contained more than one microsatellite and 56 microsatellites consisted of compound formations. Of the microsatellite containing contigs, primers for PCR amplification could be designed for 961 microsatellite loci (ESM 6). Some of the microsatellites detected were found to occur in contigs with homology to genes with putative health- and stress-related functions, such as contig 55086 (containing a TG motif repeated eight times) with homology to a small heat shock protein and contig 110428 (containing a TG motif repeated nine times) with homology to MHC I antigen.

Discussion

Polymorphic Loci in *L. rohita*

This study has developed a large resource of single nucleotide polymorphism and microsatellite markers that will be used for mapping-based approaches to identify

QTL associated with disease resistance of rohu to *A. hydrophila*. Resistant and susceptible line animals were sampled for the detection of SNPs and microsatellites to boost the likelihood of finding markers that segregate in association with disease resistance loci in rohu families. The stringency for SNP detection that was applied yielded a high proportion of valid SNP loci. One hundred percent of high coverage and high minor allele frequency putative SNPs tested were found to be polymorphic. Detection of polymorphic loci was performed using samples taken from two sub-populations derived from the same rohu population that will be scanned for associations with disease resistance. Microsatellites identified in this study could be used to supplement SNP data. In the process of scanning the genome, we will also be able to produce a linkage map of all informative SNPs and microsatellites, around half of which have been annotated with gene names and gene ontology terms.

Mitochondrial DNA

L. rohita mitochondrial contigs and SNPs were mapped with reference to the whole common carp *C. carpio* mitochondrial genome sequence. Coverage of all the mitochondrial protein coding genes was detected, but many gaps were found to exist. Gaps could be due to a relatively low expression (and consequently poor coverage) of particular mitochondrial genes in the tissues sampled or because of differences in the organisation or the mitochondrial genome sequence between the two species. Sequence coverage was highly variable, ranging from double for contig 110148 in NADH1 to 3,719× for contig 111443 in NADH2. Contigs that were found to match the *C. carpio* mitochondrial genome did so with high identity. The top ten matches showed greater than 85% identity (ranging from 85.5% for contig 11443 of 715 base pairs length to 93.5% for contig 110148 of 1,925-base-pair length) and few gaps (0–0.8%). Four percent genetic divergence between *L. rohita* and *C. carpio* 12 s-mitochondrial DNA sequences has been detected before (with 0.53%, 9.26% and 7.41% genetic divergence detected between *L. rohita* and *L. calbasu* and between *Cirrhinus mrigala* and *Ctenopharyngodon idella*, respectively) (Carreras-Carbonell et al. 2008). Many tRNA genes were also sequenced (tRNA-Val, -Leu, -Ile, -Gln, -Met, -Ala, -Asn, -Cys, -Tyr and -Arg; Fig. 3).

Differences Detected Between the Susceptible and Resistant Line Animals Sampled

The results indicate possible differences in gene expression and allele frequency between the sampled resistant and susceptible individuals at many loci that can be implicated as having a role in disease resistance.

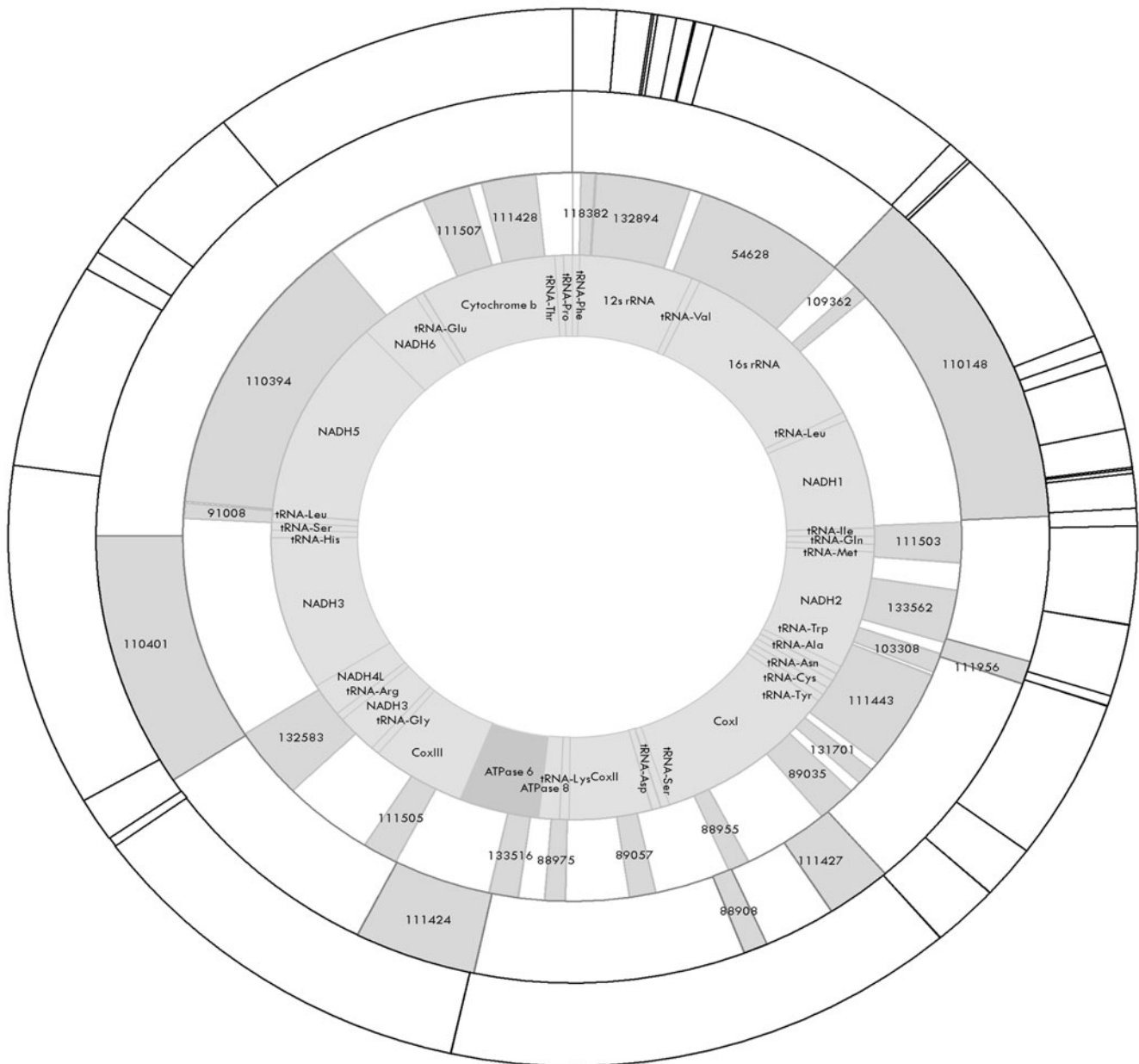


Fig. 3 Map of the *C. carpio* mitochondrial genome (Chang et al. 1994) showing the position of *L. rohita* contigs and SNPs. The inner ring shows the location of mitochondrial genes and t-RNA. The positions of the assembled contigs are represented by shaded patches

in the two concentric rings out from the inner ring (labelled contig ID numbers) and SNP positions are marked by lines in the outer ring. A line at 0° slices through all the rings and marks the start position for the *C. carpio* mitochondrial DNA sequence

Six of the 100 contigs showing highest indications of differential up-expression in sampled resistant line fish showed homology to heat shock proteins (HSP30, HSP70 and HSP90). Extracellular and membrane-bound heat-shock proteins are thought to play important immunostimulatory functions and bind antigens and present antigens to the immune system (e.g. review of HSP70, Radons and Multhoff 2005). They also act as intra-cellular chaperones (helping to stabilise proteins) and are thought to play a general

role in providing a degree of stress tolerance. In Sydney rock oysters, small HSPs showed twofold higher transcript expression in animals selected for disease resistance compared to wild-caught oysters (Green et al. 2009). The ability of HSPs to stimulate innate and acquired antigen-specific immunity has made them candidates for vaccine development (Sudheesh et al. 2007; Beachy et al. 2007; Walker et al. 2007; Segal et al. 2006; Keeble and Walker 2009). Of the 100 contigs showing highest indications

of differential up-expression in sampled resistant line fish, seven showed homology to putative zona pellucida glycoprotein and three to choriogenin (I, h and h-beta, with high homology also to zona pellucida glycoprotein 2). High molecular weight glycoprotein polymers called mucins are found on the outer body surfaces and intestine of fish, and higher levels of production of these mucins could offer better protection against bacterial infection as mucin forms a highly viscous gel that protects the epithelium from chemical, physical and microbial disturbances. Mucin genes show little sequence conservation across species (Roussel and Delmotte 2004) but can be predicted from the common amino acid composition of mucin domains (Lang et al. 2004). Common carp respond quickly to increased bacterial loads in the water by increasing the amount and total glycosylation of high molecular weight glycoproteins in the skin (van der Marel et al. 2010; Schroers et al. 2009). The contigs with homology to zona pellucida glycoprotein 2 that showed the highest differential expression were those that showed the highest expression in the resistant pool ($r_s=0.92$), while transcript expression was generally relatively lower and less variable in the susceptible pool (average log2 coverage of -0.16 versus 4.34 , with variance 0.47 versus 3.13 , in susceptible and resistant pools, respectively).

Sevenfold higher indicative expression of the serum lectin isoform 1 precursor gene, and the occurrence of alternative isoforms of galactoside-binding soluble lectin 9 in resistant and susceptible line fish, is also an interesting finding as serum lectins are found in mucus and have been shown to agglutinate to and alter the viability and pathogenicity of Gram-negative bacteria including *A. hydrophila* (Fock et al. 2001; Dutta et al. 2005; Young et al. 2007; Merino-Contreras et al. 2001).

Different MH class I antigen haplotypes were found to be fixed in the susceptible and resistant pools of individuals (five SNP loci) and some contigs with homology to MH class I antigens were indicated to be either more than fivefold down- or up-regulated in resistant line individuals from mRNA-seq. Quantitative real-time PCR confirmed more than fivefold up-regulation of one these MH class I haplotypes (contig 88601) in the skin, gill and intestine of resistant line fish. MH class I molecules could also play an important role in producing anti-bacterial immune responses in fish (Nikolich-Zugich et al. 2004). T cells require major histocompatibility antigen presenting proteins on the cell surface to recognise and present pathogen-derived peptides in order to mount an immune response against the detected pathogen. As natural targets in candidate gene studies for disease resistance, three MH class I alleles have been found to be associated with improved resistance and four MH class II alleles were associated with increased susceptibility to *Aeromonas salmonicida* in Atlantic salmon (Grimholt et al. 2003). Selective pressure to recognise a rapidly evolving and diverse array of pathogen

peptides is thought to maintain the high polymorphism and heterozygosity observed in the major histocompatibility antigen-presenting proteins on the cell surface. Research in salmonids has found that MH class I exons are “shuffled” and that this is at least one way that new polymorphisms are generated (Shum et al. 2001; Grimholt et al. 2002).

From evidence in the available literature, it is likely that the MH loci, glycoproteins, heat shock proteins, and lectins all play a role in keeping fish, and other organisms in a state of readiness for disease. The results of this study indicate that selection has potentially resulted in allele frequency and basal gene expression differences at some of these loci. Besides being involved in severe fin, skin and tail rot, *A. hydrophila* causes haemorrhagic septicaemia, deep ulcers affecting musculature with cutaneous lesions and intestinal erosion (Sahoo et al. 1988). In most instances, the portal of entry for this bacterial pathogen remains eroded skin and intestine. Thus, it is expected that the fish with up-regulation of stress-related genes at the site of entry would probably be more competent to handle this pathogen.

Potential Approach to Find Genes Associated with Disease Resistance

This study highlights the potential usefulness and pitfalls of RNA-seq as a first-pass approach for identifying genes whose differential expression and polymorphism are associated with disease resistance. The study was carried out at a time when the costs of RNA-seq were rapidly decreasing and when the technology for RNA-seq, and ability for analysis of such data, was rapidly improving. The assembly of RNA sequence data will be affected by the occurrence of different isoforms, paralogs and splice variants, and such occurrences are difficult to investigate with short reads in the absence of a reference genome sequence. We expect that a full genome reference sequence for this, or a closely related species, will soon be available. Along with a full genome reference sequence, the mRNA-seq data presented here will be a useful resource for discriminating isoforms, paralogs and splice variants. The data presented in this study point to several important considerations that need to be made when designing experiments and sequencing strategies for detection of such effects:

1. Large data sets, containing expression and allelic values for individuals, are needed to give confidence for statistical analysis. This study used expression data for over 100,000 contigs in 20 animals (two pools), and there would therefore be a high sampling error. That is, by randomly selecting and sequencing transcripts from each pool, there would be many transcripts that appear to be more represented in one pool than the other because of sampling bias (not because of differential

transcript expression associated with disease resistance between resistant and susceptible lines).

2. Pooling does not enable normalisation of the average coverage depth gene expression data within pools, only between the resistant and susceptible pools. Therefore, the overall MA plot obtained in this study shows distinct lines of data that could be due to different quantities of RNA from individuals within pools, and the transcripts that appear to be most differentially expressed could be derived from those animals whose RNA was most under- or over-represented in each pool (unequal representation bias).
3. We would expect that family-based selection for disease resistance is resulting in allele frequency changes at different loci across the population (depending on which causative mutations are segregating in each family and the size of their effect on the trait). Under such a scenario, we would not expect gene expression to be affected by selection at these different loci in the same way in every individual. In future studies, we recommend individual sample barcoding so that individual expression values for particular tissues can be detected by RNA-seq and so that normalisation can be used to reduce unequal representation bias within pools. Replication (e.g. pooled samples from additional susceptible or resistant line animals across more lanes of the sequencer) would also add power to the analysis, but both replication and tagging cost more to implement.
4. Differential transcript expression could be due to environmental factors and not due to differences in a line's ability to resist disease. Therefore, it is important to rear all animals in a common "low stress" environment. In our study, after all the animals were tagged at 10–15 g weight, they were reared in the same pond (common environment) for 8 months before sampling so that differential environmental effects on gene expression were minimised.

Other studies have shown that transcript abundance changes have the potential to evolve much faster than changes at the DNA sequence level (Denver et al. 2005). This is believed to be because mutations in many genes combine to affect transcriptional differences. Differential transcript expression could be controlled by genetic differences at a number of loci in the divergent lines. Differences in the modulation of gene expression over a number of loci could be strongly affecting disease resistance. The potential for selection based directly on gene expression profile associations (e.g. a breeding value based on a number of detected transcriptional differences such as that proposed by Robinson et al. (2008)) or its use as a supplement with marker-assisted selection or traditional selective breeding warrants testing.

Summary

In conclusion, the mRNA-seq approach used in this study has resulted in a large resource of new gene-associated SNP and microsatellite markers. A number of the transcripts that showed indications of differential expression or showed allele frequency differences between the animals sampled in resistant and susceptible lines were homologous to genes that have been implicated in resistance to bacterial diseases based on previous studies. An indication of differential expression from mRNA-seq, for one contig with homology to a major histocompatibility gene, was confirmed using quantitative real-time PCR on an independent set of samples. Sampling error is likely to be high for the mRNA-seq approach used in this study, and we have considered future sampling designs that would need to be used with RNA-seq in order to give greater confidence about differences that are associated with greater protection to bacterial disease. Novel genes that were found to share little homology with sequences annotated at NCBI will be interesting to investigate further. The study highlights the need to develop powerful experimental designs in order to fully exploit the potential of NGS technologies so that useful discoveries can be made and applied to animal breeding.

Acknowledgements This work was supported by the Norwegian Research Council (project number 183833), Department of Biotechnology of the Government of India, the Central Institute of Freshwater Aquaculture and Nofima. The authors would like to thank Steve Carson who provided computer help and advice, Thomas Moen for his assistance with analysis scripts, and the staff at Genotypic Technology Pty Ltd in Bangalore India (especially Anand Damodaran) who prepared the RNA and organised for the sequencing of samples. The authors are thankful to Dr. S Ayyappan, Director General, Indian Council of Agricultural Research, for his guidance and constant support.

References

- Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Aoki T, Hirono I, Kondo H, Hikima J, Jung TS (2011) Microarray technology is an effective tool for identifying genes related to the aquacultural improvement of Japanese flounder, *Paralichthys olivaceus*. *Comp Biochem Physiol Part D Genomics Proteomics* 6:39–43
- Babak T, Garrett-Engle P, Armour C, Raymond C, Keller M, Chen R, Rohl C, Johnson J, Attie A, Fraser H, Schadt E (2010) Genetic validation of whole-transcriptome sequencing for mapping expression affected by *cis*-regulatory variation. *BMC Genomics* 11:473
- Beachy SH, Kisalus AJ, Repasky EA, Subjeck JR, Wang XY, Kazim AL (2007) Engineering secretable forms of chaperones for immune modulation and vaccine development. *Methods* 43:184–193

- Bullard JH, Purdom E, Hansen KD, Dudoit S (2010) Evaluation of statistical methods for normalisation and differential expression in mRNA-Seq experiments. *BMC Bioinforma* 11:94
- Carreras-Carbonell J, Macpherson E, Pascual M (2008) Utility of pairwise mtDNA genetic distances for predicting cross-species microsatellite amplification and polymorphism success in fishes. *Conserv Genet* 9:181–190
- Chang YS, Huang FL, Lo TB (1994) The complete nucleotide sequence and gene organization of carp (*Cyprinus carpio*) mitochondrial genome. *J Mol Evol* 38:138–155
- Cossins A, Fraser J, Hughes M, Gracey A (2006) Post-genomic approaches to understanding the mechanisms of environmentally induced phenotypic plasticity. *J Exp Biol* 209:2328–2336
- Das Mahapatra K, Gjerde B, Sahoo PK, Saha JN, Barat A, Sahoo M, Mohanty BR, Odegard J, Rye M, Salte R (2008) Genetic variations in survival of rohu carp (*Labeo rohita*, Hamilton) after *Aeromonas hydrophila* infection in challenge tests. *Aquaculture* 279:29–34
- Das Mahapatra K, Sahoo PK, Saha JN, Mohanty B, Sahoo M, Rye M, Salte R (2009) Response to selection for resistant to *Aeromonas hydrophila* in a susceptible line of rohu carp *Labeo rohita*. *World Aquaculture Society, Asia Pacific Aquaculture, Kuala Lumpur, Malaysia*
- Denver DR, Morris K, Strelman JT, Kim SK, Lynch M, Thomas WK (2005) The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nature Genet* 37:544–548
- Dutta S, Sinha B, Bhattacharya B, Chatterjee B, Mazumder S (2005) Characterization of a galactose binding serum lectin from the Indian catfish, *Clarias batrachus*: possible involvement of fish lectins in differential recognition of pathogens. *Comp Biochem Physiol C-Toxicol Pharmacol* 141:76–84
- Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30:207–210
- Fock WL, Chen CL, Lam TJ, Sin YM (2001) Roles of an endogenous serum lectin in the immune protection of blue gourami, *Trichogaster trichopterus* (Pallus) against *Aeromonas hydrophila*. *Fish Shellfish Immunol* 11:101–113
- Gentleman R, Carey V, Huber W, Irizarry R, Dudoit S (2005) *Bioinformatics and computational biology solutions using R and bioconductor*. Springer, New York
- Gibson G, Weir B (2005) The quantitative genetics of transcription. *Trends Genet* 21:616–623
- Gjerde B, Mahapatra KD, Jana RK, Saha JN, Rye M (2005) Selective breeding in Indian aquaculture. *INFOFISH International INFOFISH, Kuala Lumpur, Malaysia* 1: 13–15
- Goetz F, Rosauer D, Sitar S, Goetz G, Simchick C, Roberts S, Johnson R, Murphy C, Bronte CR, Mackenzie S (2010) A genetic basis for the phenotypic differentiation between siscowet and lean lake trout (*Salvelinus namaycush*). *Mol Ecol* 19:176–196
- Götz S, García-Gómez JM, Terol J, Williams TD, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 36:3420–3435
- Gracey AY, Fraser EJ, Li WZ, Fang YX, Taylor RR, Rogers J, Brass A, Cossins AR (2004) Coping with cold: an integrative, multitissue analysis of the transcriptome of a poikilothermic vertebrate. *Proc Natl Acad Sci U S A* 101:16970–16975
- Green TJ, Dixon TJ, Devic E, Adlard RD, Barnes AC (2009) Differential expression of genes encoding anti-oxidant enzymes in Sydney rock oysters, *Saccostrea glomerata* (Gould) selected for disease resistance. *Fish Shellfish Immunol* 26:799–810
- Grimholt U, Drablos F, Jorgensen SM, Hoyheim B, Stet RJM (2002) The major histocompatibility class I locus in Atlantic salmon (*Salmo salar* L.): polymorphism, linkage analysis and protein modelling. *Immunogenetics* 54:570–581
- Grimholt U, Larsen S, Nordmo R, Midtlyng P, Kjoeglum S, Storset A, Saeb S, Stet RJM (2003) MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. *Immunogenetics* 55:210–219
- Harikrishnan R, Balasundaram C (2005) Modern trends in *Aeromonas hydrophila* disease management with fish. *Rev Fish Sci* 13:281–320
- Heng JF, Su JG, Huang T, Dong J, Chen LJ (2011) The polymorphism and haplotype of TLR3 gene in grass carp (*Ctenopharyngodon idella*) and their associations with susceptibility/resistance to grass carp reovirus. *Fish Shellfish Immunol* 30:45–50
- Houston RD, Guy DR, Hamilton A, Ralph J, Spreckley N, Taggart JB, McAndrew BJ, Haley CS, Bishop SC (2006) Detection of major QTL affecting resistance to infectious pancreatic necrosis (IPN) in a commercial Atlantic salmon population. *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Minas Gerais, Brazil, 13–18 August, 2006. Instituto Prociencia, Minas Gerais, Brazil*
- Jeney Z, Jeney G (1995) Recent achievements in studies on diseases of common carp (*Cyprinus carpio* L.). *Aquaculture* 129:397–420
- Jeukens J, Renaut S, St-Cyr J, Nolte AW, Bernatchez L (2011) The transcriptomics of sympatric dwarf and normal lake whitefish (*Coregonus clupeaformis* spp., Salmonidae) divergence as revealed by next-generation sequencing. *Mol Ecol* 19:5389–5403
- Keeble JR, Walker KB (2009) Therapeutic vaccine comprising Mycobacterium HSP70. *Expert Opin Ther Patents* 19:95–99
- Kong RYC, Lee SKY, Law TWF, Law SHW, Wu RSS (2002) Rapid detection of six types of bacterial pathogens in marine waters by multiplex PCR. *Water Res* 36:2802–2812
- Lang T, Alexandersson M, Hansson GC, Samuelsson T (2004) Bioinformatic identification of polymerizing and transmembrane mucins in the puffer fish *Fugu rubripes*. *Glycobiology* 14:521–527
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2^(-ΔΔC_T) method. *Methods* 25:402–408
- Merino-Contreras ML, Guzman-Murillo MA, Ruiz-Bustos E, Romero MJ, Cadena-Roa MA, Ascencio F (2001) Mucosal immune response of spotted sand bass *Paralabrax maculatofasciatus* (Steindachner, 1868) orally immunised with an extracellular lectin of *Aeromonas veronii*. *Fish Shellfish Immunol* 11:115–126
- Meyer E, Aglyamova GV, Matz MV (2011) Profiling gene expression responses of coral larvae (*Acropora millepora*) to elevated temperature and settlement inducers using a novel RNA-Seq procedure. *Mol Ecol* 20:3599–3616
- Millán A, Gómez-Tato A, Pardo BG, Fernández C, Bouza C, Vera M, Alvarez-Dios JA, Cabaleiro S, Lamas J, Lemos ML, Martínez P (2011) Gene expression profiles of the spleen, liver, and head kidney in turbot (*Scophthalmus maximus*) along the infection process with *Aeromonas salmonicida* using an immune-enriched oligo-microarray. *Mar Biotechnol* 13:1099–1114
- Moen T, Baranski M, Sonesson A, Kjoeglum S (2009) Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): population-level associations between markers and trait. *BMC Genomics* 10:368
- Nayak SP, Mohanty BR, Mishra J, Rauta PR, Das A, Eknath AE, Sahoo PK (2011) Ontogeny and tissue-specific expression of innate immune related genes in rohu, *Labeo rohita* (Hamilton). *Fish Shellfish Immunol* 30:1197–1201
- Nikolich-Žugich J, Fremont DH, Miley MJ, Messaoudi I (2004) The role of mhc polymorphism in anti-microbial resistance. *Microbes Infect* 6:501–512

- Radons J, Multhoff G (2005) Immunostimulatory functions of membrane-bound and exported heat shock protein 70. *Exerc Immunol Rev* 11:17–33
- Robinson N, Goddard M, Hayes B (2008) Use of gene expression data for predicting continuous phenotypes for animal production and breeding. *Animal* 2:1413–1420
- Roussel P, Delmotte P (2004) The diversity of epithelial secreted mucins. *Curr Org Chem* 8:413–437
- Sahoo PK, Mukherjee SC, Sahoo SK (1988) *Aeromonas hydrophila* versus *Edwardsiella tarda*: a pathoanatomical study in *Clarias batrachus*. *J Aqua* 6:57–66
- Sahoo PK, Meher PK, Das Mahapatra K, Saha JN, Jana RK, Reddy PVK (2004) Immune responses in different fullsib families of Indian major carp, *Labeo rohita*, exhibiting differential resistance to *Aeromonas hydrophila* infection. *Aquaculture* 238:115–125
- Sahoo PK, Das Mahapatra K, Saha JN, Barat A, Sahoo M, Mohanty BR, Gjerde B, Odegard J, Rye M, Salte R (2008) Family association between immune parameters and resistance to *Aeromonas hydrophila* infection in the Indian major carp, *Labeo rohita*. *Fish Shellfish Immunol* 25:163–169
- Schroers V, van der Marel M, Neuhaus H, Steinhagen D (2009) Changes of intestinal mucus glycoproteins after peroral application of *Aeromonas hydrophila* to common carp (*Cyprinus carpio*). *Aquaculture* 288:184–189
- Segal BH, Wang XY, Dennis CG, Youn R, Repasky EA, Manjili MH, Subjeck JR (2006) Heat shock proteins as vaccine adjuvants in infections and cancer. *Drug Discov Today* 11: 534–540
- Shum BP, Guethlein L, Flodin LR, Adkison MA, Hedrick RP, Nehring RB, Stet RJM, Secombes C, Parham P (2001) Modes of salmonid MHC class I and II evolution differ from the primate paradigm. *J Immunol* 166:3297–3308
- Sudheesh PS, LaFrentz BR, Call DR, Siems WF, LaPatra SE, Wiens GD, Cain KD (2007) Identification of potential vaccine target antigens by immunoproteomic analysis of a virulent and a non-virulent strain of the fish pathogen *Flavobacterium psychrophilum*. *Dis Aquat Org* 74:37–47
- Tang K, Fu DJ, Julien D, Braun A, Cantor CR, Koster H (1999) Chip-based genotyping by mass spectrometry. *Proc Natl Acad Sci U S A* 96:10016–10020
- van der Marel M, Caspari N, Neuhaus H, Meyer W, Enns ML, Steinhagen D (2010) Changes in skin mucus of common carp, *Cyprinus carpio* L., after exposure to water with a high bacterial load. *J Fish Dis* 33:431–439
- Walker KB, Keeble J, Colaco C (2007) Mycobacterial heat shock proteins as vaccines—a model of facilitated antigen presentation. *Curr Mol Med* 7:339–350
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10:57–63
- Xia C, Ma Z-H, Habibur Rahman M, Wu Z-G (2004) PCR cloning and identification of the β -haemolysin gene of *Aeromonas hydrophila* from freshwater fishes in China. *Aquaculture* 229:45–53
- Yang SS, Tu ZJ, Cheung F, Xu WW, Lamb JF, Jung HJ, Vance CP, Gronwald JW (2011) Using RNA-Seq for gene identification, polymorphism detection and transcript profiling in two alfalfa genotypes with divergent cell wall composition in stems. *BMC Genomics* 12:199
- Young KM, Russell S, Smith M, Huber P, Ostland VE, Brooks AS, Hayes MA, Lumsden JS (2007) Bacterial-binding activity and plasma concentration of ladderlectin in rainbow trout (*Oncorhynchus mykiss*). *Fish Shellfish Immunol* 23:305–315
- Zhang ZB, Niu CJ, Storset A, Bogwald J, Dalmo RA (2011) Comparison of *Aeromonas salmonicida* resistant and susceptible salmon families: a high immune response is beneficial for the survival against *Aeromonas salmonicida* challenge. *Fish Shellfish Immunol* 31:1–9