Marine Genomics 35 (2017) 63–68

Contents lists available at ScienceDirect

Marine Genomics

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

Method paper

De novo transcriptome assembly and annotation for the desert rainbowfish (*Melanotaenia splendida tatei*) with comparison with candidate genes for future climates

Katie Gates [a], Jonathan Sandoval-Castillo [a], Louis Bernatchez [b], Luciano B. Beheregaray [a,*]

[a] Molecular Ecology Lab, Flinders University, Adelaide 5001, South Australia, Australia
[b] Institut de Biologie Intégrative et des Systèmes, Université Laval, Québec, QC G1V 0A6, Canada

ARTICLE INFO

Article history:
Received 30 January 2017
Accepted 15 May 2017
Available online 22 May 2017

Keywords:
RNA-seq
Climate change
Gene expression
Ecological genomics
Thermal tolerance
Teleost

ABSTRACT

Transcriptomics *via* RNA-seq has rapidly emerged as a powerful tool for ecological and evolutionary studies, enabling genome-scale studies of adaptation *via* regulation of global gene expression. Here we present a *de novo* transcriptome for the desert rainbowfish (*Melanotaenia splendida tatei*) based on individuals sampled in the Lake Eyre Basin, Australia's arid zone river system. Recently developed methods in RNA-seq and bioinformatics were used for sequencing, assembling and annotating a high-quality liver transcriptome suitable for studies of ecology and adaptation in desert rainbowfish. Transcript annotation in UniprotKB using BLASTx assigned unique protein matches to ~47% of 116,092 Trinity genes, while BLASTp assigned unique protein matches to ~35% of 62,792 predicted protein-coding regions. A full Trinotate annotation report is provided for predicted genes and their corresponding transcripts. Annotations were compared with previously identified genes for transcriptional regulation and heritable plasticity in future climates in the subtropical rainbowfish (*M. duboulayi*), finding ~57% of these candidate genes present in the desert rainbowfish transcriptome. We discuss the utility of transcriptomics methods for ecological studies of adaptation, while emphasising a range of methodological considerations for dealing with transcriptome datasets. This newly assembled transcriptome is expected to help elucidate mechanisms for adaptation to high temperatures and a variable climate in desert aquatic fauna.

1. Introduction

Genetic research on local adaptation and population divergence has so far mostly focussed on sequence variation rather than differential gene expression (Lenz, 2015). However, it is now recognised that the regulation of gene expression may represent an important adaptive mechanism to changing environments (Merila, 2015). Studies of gene expression are gaining traction in attempts to predict responses of species to anthropogenically-induced environmental changes such as global warming (Smith et al., 2013; McCairns et al., 2016). Transcriptomics methods (*e.g.* RNA-seq) offer a promising perspective for such studies, as recently indicated in work with a range of species (Krabbenhoft and Turner, 2016), including Australian rainbowfishes (Smith et al., 2013; McCairns et al., 2016).

The desert rainbowfish, *Melanotaenia splendida tatei*, is a small-bodied freshwater fish endemic to Australia's arid zone system, the Lake Eyre Basin (McGuigan et al., 2000). They are a member of the '*australis*' species complex of Australian rainbowfishes (family *Melanotaeniidae*) (Unmack et al., 2013), which represents an emerging system for studying adaptation and evolutionary resilience to climate change (Smith et al., 2013; McCairns et al., 2016). Australian rainbowfishes are both abundant and diverse, and despite relatively recent radiation, occupy many bioregions across the country and a wide range of aquatic environments (Unmack et al., 2013). Interestingly, species ranges in Australia correspond closely to major climatic regions (McGuigan et al., 2000; Unmack et al., 2013), suggesting the possibility of recent climatic-driven diversification.

Initial studies of adaptation in rainbowfishes provided evidence of heritable morphological divergence in relation to contrasting flow regimes (McGuigan et al., 2003; McGuigan et al., 2005), a key attribute in the context of climate change. More recently, experimental studies of climate-change using the subtropical rainbowfish *M. duboulayi* provided evidence of extensive plastic transcriptional responses to predicted future temperatures (Smith et al., 2013; McCairns et al., 2016), as well as pedigree-based support for a heritable basis to these responses (McCairns et al., 2016). These findings make Australian rainbowfishes a valuable system for studying local adaptation and testing genome-level responses to climate change. Transcriptomes now exist for the subtropical *M. duboulayi* (Smith et al., 2013) and the temperate *M. fluviatilis* rainbowfishes (Sandoval-Castillo et al., in preparation; Beheregaray et al., unpublished). Studies incorporating multiple climatic ecotypes will allow exploration of adaptation and evolutionary resilience to future climates across biogeographic provinces and climatic regions.

* Corresponding author at: School of Biological Sciences, Flinders University, Adelaide, SA 5001, Australia.
E-mail address: luciano.beheregaray@flinders.edu.au (L.B. Beheregaray).

http://dx.doi.org/10.1016/j.margen.2017.05.008
1874-7787/

Here, we used RNA-seq technology to sequence, assemble and annotate a *de novo* liver transcriptome for the desert rainbowfish *M. splendida tatei*. We chose liver because previous research has linked this tissue type to temperature variation, and has provided a suite of genes shown to be under climate-mediated selection in the closely related subtropical *M. duboulayi* (Smith et al., 2013; McCairns et al., 2016). We compared these genes with annotations for the *M. splendida tatei* transcriptome to determine presence of genes that may be functionally relevant to heat stress adaptation. To the best of our knowledge, a transcriptome has not yet been published for a desert-adapted freshwater fish. This resource will enable comparisons of transcriptomic features and responses to environmental changes across a range of desert aquatic faunas.

2. Methods

2.1. Sample collection

Adult male desert rainbowfish *(M. splendida tatei*) were collected from Algebuckina Waterhole (N − 27.900, E 135.815) using fyke nets in September 2014 (spring) (Table 1). The habitat is a permanent non-flowing waterhole situated on the Neales River in the arid far north of South Australia. The waterhole is sparsely vegetated with mostly native riparian vegetation and, despite disturbance from human activity such as cattle grazing, provides important refuge for native fish and other wildlife during drought periods (Cockayne et al., 2013). The location is approximately 60 km from Oodnadatta, which holds the record for the highest temperature in Australia (50.7 °C) and frequently experiences summer mean maximums of above 35 °C (BOM, 2017). The sampled males were of similar size (6–7 cm) to control for sex- and age-related effects on transcriptional variation. Fish were euthanized by overdose of anaesthetic AQUI-S, and livers were extracted and immediately preserved in RNAlater™. Samples were stored in −80 °C until RNA extraction.

2.2. Wet-lab and in silico procedures

An overview of the wet-lab and *in silico* procedures for all methods detailed hereafter is found in the schematic diagram (Fig. 1).

2.3. RNA extraction, library preparation and sequencing

Total RNA was extracted from preserved liver samples of six individuals using MagMAX™-96 Total RNA Isolation Kit (Ambion) following manufacturer's protocol. Only male fish were used, in accordance with previous rainbowfish transcriptomic work (Smith et al., 2013; McCairns et al., 2016). Quantity and quality was assessed using a Bioanalyzer 2100 (Agilent Technologies). Total RNA was converted into complimentary DNA libraries for sequencing, following TruSeq RNA™ Sample Preparation Low Throughput Protocol (Illumina). Unique indexing adaptors (Illumina MID tags 2, 4–7, 12–16, 18, 19) were ligated to each sample, and fragment size distribution and concentration of libraries were verified using a Bioanalyzer 2100 (Agilent Technologies). RNA concentration of individual samples were normalised by dilution to approximately 15 ng/μL for a total volume of 50 μL per individual. Normalised libraries were pooled in a group of 12 (including six samples from another project) for sequencing within a single batch in one Illumina HiSeq2500 lane. Sequencing was performed at the McGill University Génome Québec Innovation Centre in Canada, to produce paired-end, 100 base reads.

2.3.1. Quality trimming

Sequence data were sorted by individual and trimmed of indexing adaptors at the sequencing facility. Quality trimming was performed using Trimmomatic (Bolger et al., 2014) (Sliding Window: 5:20; leading/trailing threshold: 20; Head Crop: 12). Reads less than 45 bp were dropped from the dataset.

2.3.2. Transcriptome assembly and evaluation

Trimmed, strand-specific reads were assembled *de novo* using Trinity 2.1.0 (Haas et al., 2013) with default parameters and in-silico normalisation. The success of the completed transcriptome assembly was evaluated using read content statistics (% raw reads present), contig length distribution (N50), annotation-based metrics (% full length transcripts) and Benchmarking Universal Single-Copy Orthologs (BUSCO) software (Simao et al., 2015). Where two or more transcripts showed 80% or higher similarity or blasted to the same gene, all but the longest transcript were removed to generate a non-redundant set of transcripts (unigenes).

Table 1
MixS information for transcriptome assembly of *Melanotaenia splendida tatei*.

General features of classification	
Classification	Eukaryota; Animalia; Chordata; Vertebrata; Actinopterygii; Atheriniformes; Melanotaenioidei; Melanotaeniidae; *Melanotaenia splendida tatei*
Investigation type	Eukaryote transcriptome
Project name	Comparative Evolutionary Genomics of Australian Rainbowfishes (L.B. Beheregaray & L. Bernatchez)
Environment	
Latitude, longitude	N −27.900, E 135.815
Geographical location	Algebuckina Waterhole, South Australia, Australia
Collection date	2014-09
Biome	Desert
Feature	Waterhole
Material	Freshwater
Sequencing	
Sequencing method	Illumina HiSeq 2500 paired-end
Estimated size	171 Mb
Organ or tissue source	Liver tissue
Assembly	
Method	*De novo* assembly
Program	Trinity 2.1.0
Finishing strategy	High quality transcriptome assembly
Data accessibility	
Database name	NCBI
Project name	PRJNA369450
Sample name	SRR5223656,SRR5223655,SRR5223654,SRR5223653,SRR5223652,SRR5223651

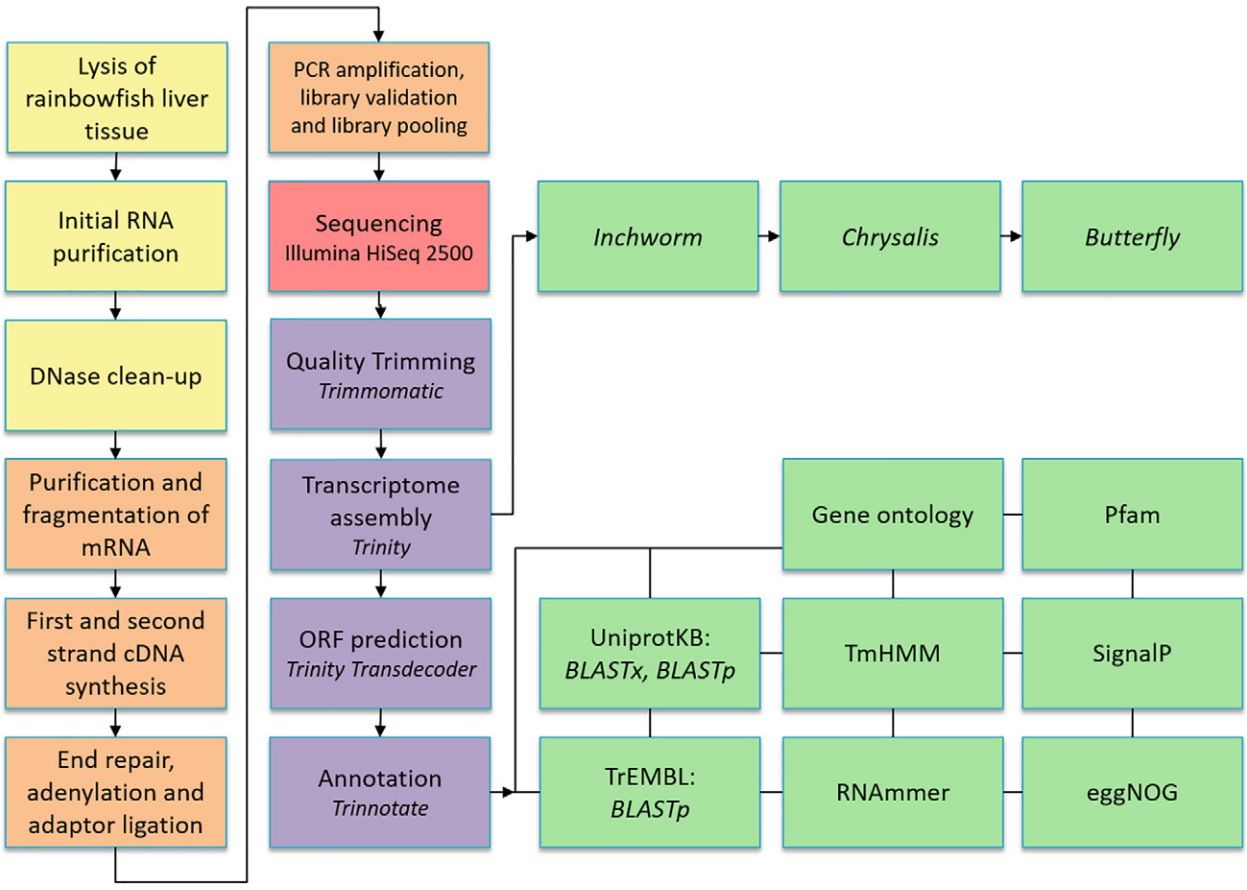

Fig. 1. Workflow of RNA extraction, sequencing and bioinformatics procedures implemented in this study (yellow = RNA extraction; orange = cDNA library preparation; red = sequencing; purple = major bioinformatics procedures; green = subsets of bioinformatics procedures). Names in italics correspond to software.

2.3.3. Functional annotation

The presence of open reading frames (ORFs) of a minimum length of 100 peptides was analysed using TransDecoder scripts. The transcriptome was functionally annotated using the Trinotate package within the Trinity suite (Grabherr et al., 2011). BLASTx (for transcript) and BlastP (for ORFs) (Altschul et al., 1990) were used to query contigs against the UniprotKB and TrEMBL protein databases (using default E-value cut-off) to identify homology to known proteins.

Trinotate was used to automate protein domain identification within the PFAM database (HMMR), protein signal peptide and transmembrane domain prediction (signalP and tmHMM), and leveraging gene ontology (GO) annotation databases. Functional annotation data were uploaded to a Trinotate SQLite database (Supplementary Table S1), with GO assignments extracted for each gene feature. Functional annotation results were contrasted with genes previously identified in the closely related *M. duboulayi* as candidates for temperature stress response in both short-term (Smith et al., 2013) and long-term (McCairns et al., 2016) climate change scenarios.

3. Results

3.1. Transcriptome sequencing and assembly

The half-lane of Illumina HiSeq sequencing produced approximately 299 million reads for *Melanotaenia splendida tatei* (Table 2), with reads per individual ranging from 22 to 30 million. After trimming and quality control, a total of 241,339,863 reads were assembled *de novo*.

Of total aligned reads, 74% were properly paired, with a total of 165,976 contigs ('Trinity transcripts'), 116,092 genes ('Trinity genes') and 67,938 unigenes. Transcriptome contig Nx statistics revealed an N50 of 2113, median contig length of 463, and average contig length of 1032. Assembly completeness assessment using BUSCO found a high percentage of genes in common (complete plus fragmented) with the eukaryotic (90%), the Metazoan (95%) and vertebrate (75%) gene datasets (Table 3).

3.2. Functional annotation of assembled transcriptome

A search of assembled Trinity transcripts against the UniprotKB protein database matched 77,291 (67%) of the total 116,092 isolated Trinity genes (Table 4). Of these, 53,275 (46% of Trinity genes) were determined to represent unique proteins. Transdecoder-predicted protein coding regions (ORFs) aligned in UniprotKB and TrEMBL resulted in 22,286 and 29,989 unique protein hits, respectively. A total of 96,202 GO terms were assigned to the assembled transcripts (Supplementary

Table 2
Summary of transcriptome assembly statistics for *Melanotaenia splendida tatei*.

Raw reads	299,308,290
Total aligned reads	241,339,863
Total assembled bases	171,263,802
Number of contigs ('Trinity transcripts')	165,976
Number of genes ('Trinity genes')	116,092
Number of Unigenes	67,938
Contig N50 (bp)	2113
Maximum contig length (bp)	17,870
Minimum contig length (bp)	201
Median contig length (bp)	463
Average contig length (bp)	1032
Open reading frames	62,792

Table 3
Assessment of transcriptome assembly completeness for *M. splendida tatei* as it compares to Benchmarking Universal Single-Copy Orthologs (BUSCO) for eukaryotic, metazoan and vertebrate gene sets.

	Eukaryote	Metazoan	Vertebrate
% complete	85	86	67
% duplicated	17	16	10
% fragmented	5	9	8
% missing	10	5	25
No. of genes assessed	429	843	3023

Table S1). Approximately 70% of ontologies corresponded to biological processes, 21% to molecular functions and 9% to cellular component.

Comparison of functional annotation results found 23 genes in common with 50 candidate genes identified for temperature stress response in *M. duboulayi* in a short-term climate change experiment (Smith et al., 2013) and eight in common with 12 candidates identified in a long-term climate change experiment (McCairns et al., 2016) (Table 5). The most highly represented biological functions among these annotations included metabolic processes (31%) and transcription regulation (25%).

4. Discussion

4.1. Evaluation of assembly and annotation

We used RNA-seq to assemble a *de novo* transcriptome for the desert rainbowfish *Melanotaenia splendida tatei* from wild individuals sampled in Australia's arid zone. Sequencing compared favourably to the previous assembly of the *M. duboulayi* transcriptome by Smith et al. (2013), with the 299 million raw reads (~150 m paired-end) equating to more than twice the number of sequence reads for the equivalent lane space. As with many aspects of RNA-seq, ideal sequencing depth is dependent on the biological question. Because coverage across the transcriptome varies among transcripts, the ideal number of reads in any study is dependent on the least abundant RNA transcripts of interest, and a greater number of reads can enable the discovery of those that are more lowly-expressed (Sims et al., 2014). This will depend on the organism and tissue type, as some samples may be dominated by very highly expressed genes, which can take up a large percentage of the allotted reads (Conesa et al., 2016). However, recent studies indicate that *de novo* transcriptome sequencing for a majority of animal phyla can be performed to a high standard with a total number of paired-end reads (all replicates combined) as low as 20 million, given a length of 75–100 bp (Francis et al., 2013). These requirements were easily satisfied by the sequencing depth obtained for *M. splendida tatei*.

Aligned reads constituted approximately 80% of the original number, falling into an expected range of around 70–90% (Conesa et al., 2016), with a majority properly paired. Nx statistics indicated that half of all assembled bases were found in contigs of at least 2113 bp. Again, this compared well to the *M. duboulayi* transcriptome assembly (Smith et al., 2013), which received an N50 value of 1856. Although Nx values can be inflated by high numbers of isoforms in long transcripts or by increasing user-defined minimum contig length, it remains an important metric in that values much lower than benchmark can signal poor quality assembly (Haas et al., 2013; Rana et al., 2016). The 67,938 predicted unigenes also compared well to other recent fish transcriptome assemblies, including those using more than one tissue type. For example, a full body transcriptome of the Chilean pencil catfish (*Trichomycterus areolatus*) produced 64,889 non-redundant transcripts (Schulze et al., 2016), while a combined gill, liver, kidney and muscle transcriptome of the striped catfish (*Pangasianodon hypophthalmus*) produced 66,451 non-redundant transcripts (Nguyen Thanh et al., 2014).

Evaluation using BUSCO (Simao et al., 2015) compares gene content with orthologous genes found in more than 90% of species for selected clades. Assembled genes from the *M. splendida tatei* transcriptome performed well next to other assemblies (Hara et al., 2015; Seemann et al., 2015). For example, a recent quality assessment of twenty vertebrate genomes found the number of missing data to range from 3.4% (panda) to 17% (pig) against the BUSCO vertebrate gene set (Seemann et al., 2015). When considering that these figures represent complete genomes, as opposed to the snapshot offered by a transcriptome of single-tissue expression at one point in time, the *M. splendida tatei* transcriptome performs remarkably well with only 25% missing data. This high representation speaks well for the choice of liver when only a single tissue can be analysed due to financial or other reasons. It can also be noted that missing genes can reflect biological novelty in the study species (Simao et al., 2015). For example, a simulation using the Japanese lamprey (*Lethernteron japonicum*) (Hara et al., 2015) received only 21% completeness against the vertebrate gene set, *versus* a 70% completeness score for the metazoan gene set, which could be the result of an overrepresentation of gene sets from more intensively studied lineages.

A reference-free functional annotation was achieved for *M. splendida tatei* using a homology search in the protein databases UniprotKB and TrEMBL. The resulting percentage of full-length transcript coverage found approximately 46% of genes matching unique proteins in the UniprotKB database. A survey of recent *de novo* transcriptome assemblies of non-model fishes revealed percentages of alignment matches ranging from approximately 30%–65%, although it was not consistently stated whether these matches represented unique proteins, or included proteins matching to multiple isoforms or gene regions (Ji et al., 2012; Smith et al., 2013; Salisbury et al., 2015; Rana et al., 2016).

Table 4
Annotation summary of BLAST searches of UniprotKB and TrEMBL protein databases for the *Melanotaenia splendida tatei* transcriptome. Percentages indicate proportion of genes receiving matches for BLASTx results and proportion of ORFs receiving matches for BLASTp searches, respectively.

Database	Total BLAST matches		No. of unique proteins	
	BLASTx	BLASTp	BLASTx	BLASTp
UniprotKB	77,291 (67%)	23,334 (37%)	53,275 (46%)	22,286 (35%)
TrEMBL	n/a	30,332 (48%)	n/a	29,989 (47%)

4.2. Implications for climatic adaptation research

This study provides a high quality annotated transcriptome which will be a valuable complement to existing studies of Australian rainbowfishes. Previous laboratory studies of differential expression using the transcriptome of the subtropical ecotype *M. duboulayi* have so far discovered predictable plastic responses to future climates through experimental manipulation of temperature conditions, and provided a suite of candidate genes for temperature tolerance (Smith et al., 2013). Additional research has also provided evidence for heritable plasticity in the expression of these candidate genes and showed that their molecular pathways co-occur with genes inferred to be under climate-mediated selection in wild *M. duboulayi* populations (McCairns et al., 2016). We found that approximately half of the previously identified candidate genes were present in the desert rainbowfish transcriptome, with the most highly represented functional annotations relating to metabolic processes and transcription regulation.

The dependence of ectotherms' metabolism on temperature has long been established (Nwewll and Northcroft, 1967; Hawkins, 1995), so it is unsurprising to find that genes relating to metabolic regulation are among the most highly regulated in temperature stress scenarios. For example, NR1D4B, a candidate for climate change response in in *M. duboulayi* (Smith et al., 2013; McCairns et al., 2016) and present in the desert rainbowfish transcriptome, is involved in regulatory pathways governing metabolism (Desvergne et al., 2006) and has previously been found to upregulate in response to cold stress in larval zebrafish (*Danio rerio*) (Long et al., 2012). Other genes, such as the protein coding

Table 5
Annotations of putative genes in *de novo* transcriptome assembly of *M. splendida tatei* found to be in common with previously identified candidate genes for temperature stress response in *M. duboulayi* (Smith et al., 2013; McCairns et al., 2016). Gene ontology abbreviations: P = biological process, F = molecular function, C = cellular component.

Of 50 candidate genes for temperature response (Smith et al., 2013)		
Gene name	Protein	Primary gene Ontology
HMGCS1	Hydroxymethylglutaryl-CoA synthase, cytoplasmic	P: isoprenoid biosynthetic process
UBE2	Ubiquitin-conjugating enzyme E2	P: endosome transport
MAO	Amine oxidase [flavin-containing]	P: catecholamine metabolic process
TKT	Transketolase	F: transketolase activity
UBE2V2	Ubiquitin-conjugating enzyme E2 variant 2	F: acid-amino acid ligase activity
CCDC47	Coiled-coil domain-containing protein 47	P: embryonic development
UQCRC2	Cytochrome b-c1 complex subunit 2, mitochondrial	F: metalloendopeptidase activity
NFE2L1	Nuclear factor erythroid 2-related factor 1	P: heme biosynthetic process
GST	Glutathione *S*-transferase	F: glutathione transferase activity
PER3	Period circadian protein homolog 3	C: cytoplasm
ALAS2	5-Aminolevulinate synthase, erythroid-specific, mitochondrial	P: response to hypoxia
MPP1	55 kDa erythrocyte membrane protein	C: intracellular non-membrane-bounded organelle
GAA	Lysosomal alpha-glucosidase	F: carbohydrate binding
ABLIM1	Actin-binding LIM protein 1	P: axon guidance
UGT2A2	UDP-glucuronosyltransferase 2A2	F: transferase activity, transferring hexosyl groups
GFOD1	Glucose-fructose oxidoreductase domain-containing protein 1	C: extracellular region
DYRK1B	Dual specificity tyrosine-phosphorylation-regulated kinase 1B	P: protein amino acid autophosphorylation
SPAG9	C-Jun-amino-terminal kinase-interacting protein	F: protein binding
SETD3	Histone-lysine *N*-methyltransferase setd3	P: peptidyl-lysine monomethylation
PPARAB	Peroxisome proliferator-activated receptor alpha	P: transcription regulation
CYP1A1	Cytochrome P450 1A	C: endoplasmic reticulum membrane
NR1D2	Nuclear receptor subfamily 1 group D member 2	P: steroid hormone mediated signaling pathway
VTG2	Vitellogenin-2	F: lipid transporter activity
Of 12 candidate genes for temperature response showing heritability of transcriptional variation (McCairns et al., 2016)		
Gene name	Protein	Primary gene Ontology
CYP1A	Cytochrome P450 1A1	C: endoplasmic reticulum membrane
HMGCS1	Hydroxymethylglutaryl-CoA synthase, cytoplasmic	P: isoprenoid biosynthetic process
HSP90AA1.2	Heat shock protein 90-alpha 2	P: response to heat
NR1D4B	Nuclear receptor subfamily 1 group D member 4b	P: transcription regulation
PPARAB	Peroxisome proliferator activated receptor alpha	P: transcription regulation
TMX2B	Thioredoxin-related transmembrane protein 2-B	P: cell redox homeostasis
UGT2A4	UDP-glucuronosyltransferase	P: metabolic process
UQCRC2B	Ubiquinol-cytochrome *c* reductase core protein II	P: aerobic respiration

HSP90A (heat shock protein), have been found to play important roles in cellular signal transduction networks (Young et al., 2001) and have been highly up-regulated in response to heat stress in zebrafish (*Danio rerio*) and tambaqui (*Colossoma macropomum*) (Long et al., 2012; Prado-Lima and Val, 2016). It is expected that the candidate genes identified in the subtropical *M. duboulayi* play an important role in this species' thermal tolerance. The same is likely to apply for many of these genes to the desert rainbowfish, particularly given the link established with temperature stress in the transcriptional response of other teleost species. While the link between transcriptional variation and fitness is far from fully established, it is currently understood that transcriptional variation can respond to selection and therefore assist in adaptation to environmental variables, such as temperature (Leder et al., 2015; McCairns et al., 2016). These genes may therefore present important targets of selection in both rainbowfish species in a warming environment.

Importantly however, further experimental studies are required to determine whether transcriptional responses to temperature stress conditions are similar in the desert rainbowfish, and whether these candidate genes play homologous roles among different species. For example, the same genes can show contrasting regulation patterns between closely related species under identical experimental conditions (Liu et al., 2012), and it has been found that transcription can sometimes differ even where translation is highly conserved between species (Wang et al., 2015). Given the adaptation of *M. splendida tatei* to the extremes of an arid environment, the desert species may possess an even more extensive suite of genes and isoforms capable of regulation under thermal stress. For example, a comparison of responses to cold stress between tomato species (Liu et al., 2012) found that a greater variety of GO terms were enriched among upregulated genes in cold-tolerant wild species than in more cold-sensitive domesticated species. Investigating these key transcriptional differences between ecotypes is worthwhile, as it may help to elucidate both genetic and related physiological characteristics which increase the likelihood of survival and species persistence under a more variable climatic regime.

There is growing evidence that phenotypic traits often have a polygenic basis and that their molecular signal of adaptation is difficult to detect *via* commonly used methods such as population genome scans or genome-wide association studies (Bernatchez, 2016; Wellenreuther and Hansson, 2016). In these instances, even highly heritable trait variation can be difficult to attribute to a genetic basis (Bernatchez, 2016). It is recognised that adaptation to thermal stress is highly likely to be polygenic in nature due to the wide-ranging effects of temperature on different biochemical reactions and physiological responses (Hawkins, 1995; Schulte, 2015). Therefore, approaches such as RNA-seq, which allow a transcriptome-wide perspective of functional responses to the environment provide a valuable complement to DNA-based studies of adaptation (Romero et al., 2012).

5. Conclusion

Understanding the effects of temperature variation on biodiversity is important for informing management practices in a changing climate. Analysis of gene expression in organisms adapted to extreme environmental conditions, particularly when compared to that of their mild climate cousins, will help to elucidate processes and pathways enabling adaptation to future climates. To this end, we have assembled and annotated the first transcriptome for a desert freshwater fish, *M. splendida tatei*. The desert rainbowfish transcriptome has received even better

evaluation metrics for a majority of quality assessments than that of the previously assembled *M. duboulayi* transcriptome, and should therefore provide an excellent basis for future comparative studies. The presence of a majority of the candidate genes in the desert rainbowfish transcriptome reflects the close relatedness of this species to the subtropical *M. duboulayi*, and supports the use of these species for comparisons of ecotypes across Australian biomes.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.margen.2017.05.008.

Acknowledgements

We thank Catherine Attard, Chris Brauer, and David Schmarr for assistance with fieldwork. This study was funded by the Australian Research Council (ARC grants DP150102903 and FT130101068). Animal ethical approval was received from Flinders University (E429/16).

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
Bernatchez, L., 2016. On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. J. Fish Biol. 89, 2519–2556.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.
BOM, 2017. Daily maximum temperature: Oodnadatta Airport, Climate Data online, Australian Government Bureau of Meteorology. viewed 01/03/2017. http://www.bom.gov.au/climate/data/.
Cockayne, B., Schmarr, D., Duguid, A., Mathwin, R., 2013. Lake Eyre Basin Rivers Assessment (LEBRA) 2012 Monitoring Report. Report to LEBRA Oversight Group.
Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., Mortazavi, A., 2016. A survey of best practices for RNA-seq data analysis. Genome Biol. 17, 13.
Desvergne, B., Michalik, L., Wahli, W., 2006. Transcriptional regulation of metabolism. Physiol. Rev. 86, 465–514.
Francis, W.R., Christianson, L.M., Kiko, R., Powers, M.L., Shaner, N.C., Haddock, S.H.D., 2013. A comparison across non-model animals suggests an optimal sequencing depth for de novo transcriptome assembly. BMC Genomics 14, 167.
Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652.
Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512.
Hara, Y., Tatsumi, K., Yoshida, M., Kajikawa, E., Kiyonari, H., Kuraku, S., 2015. Optimizing and benchmarking de novo transcriptome sequencing: from library preparation to assembly evaluation. BMC Genomics 16, 977.
Hawkins, A.J.S., 1995. Effects of temperature change on ectotherm metabolism and evolution: metabolic and physiological interrelations underlying the superiority of multilocus heterozygotes in heterogeneous environments. J. Therm. Biol. 20, 23–33.
Ji, P., Liu, G., Xu, J., Wang, X., Li, J., Zhao, Z., Zhang, X., Zhang, Y., Xu, P., Sun, X., 2012. Characterization of common carp transcriptome: sequencing, de novo assembly, annotation and comparative genomics. PLoS One 7, e35152.
Krabbenhoft, T.J., Turner, T.F., 2016. Analytical Considerations for Comparative Transcriptomics of Wild Organisms. bioRxiv.
Leder, E.H., McCairns, R.S., Leinonen, T., Cano, J.M., Viitaniemi, H.M., Nikinmaa, M., Primmer, C.R., Merilä, J., 2015. The evolution and adaptive potential of transcriptional variation in sticklebacks—signatures of selection and widespread heritability. Mol. Biol. Evol. 32, 674–689.
Lenz, T.L., 2015. Transcription in space - environmental vs. genetic effects on differential immune gene expression. Mol. Ecol. 24, 4583–4585.
Liu, H., Ouyang, B., Zhang, J., Wang, T., Li, H., Zhang, Y., Yu, C., Ye, Z., 2012. Differential modulation of photosynthesis, signaling, and transcriptional regulation between tolerant and sensitive tomato genotypes under cold stress. PLoS One 7, e50785.
Long, Y., Li, L., Li, Q., He, X., Cui, Z., 2012. Transcriptomic characterization of temperature stress responses in larval zebrafish. PLoS One 7, e37209.
McCairns, R.J.S., Smith, S., Sasaki, M., Bernatchez, L., Beheregaray, L.B., 2016. The adaptive potential of subtropical rainbowfish in the face of climate change: heritability and heritable plasticity for the expression of candidate genes. Evol. Appl. 9, 531–545.
McGuigan, K., Zhu, D., Allen, G.R., Moritz, C., 2000. Phylogenetic relationships and historical biogeography of melanotaeniid fishes in Australia and New Guinea. Mar. Freshw. Res. 51, 713–723.
McGuigan, K., Franklin, C.E., Moritz, C., Blows, M.W., 2003. Adaptation of rainbow fish to lake and stream habitats. Evolution 57, 104–118.
McGuigan, K., Chenoweth, S., Xa, F., Blows, M., Xa, W., C, M., Associate Editor: James, 2005. Phenotypic divergence along lines of genetic variance. Am. Nat. 165, 32–43.
Merila, J., 2015. Evolutionary biology: perplexing effects of phenotypic plasticity. Nature 525, 326–327.
Nguyen Thanh, H., Zhao, L., Liu, Q., 2014. De novo transcriptome sequencing analysis and comparison of differentially expressed genes (DEGs) in *Macrobrachium rosenbergii* in China. PLoS One 9, e109656.
Nwewll, R.C., Northcroft, H.R., 1967. A re-interpretation of the effect of temperature on the metabolism of certain marine invertebrates. J. Zool. 151, 277–298.
Prado-Lima, M., Val, A.L., 2016. Transcriptomic characterization of tambaqui (*Colossoma macropomum*, Cuvier, 1818) exposed to three climate change scenarios. PLoS One 11, e0152366.
Rana, S.B., Zadlock, F.J., Zhang, Z., Murphy, W.R., Bentivegna, C.S., 2016. Comparison of de novo transcriptome assemblers and k-mer strategies using the killifish, *Fundulus heteroclitus*. PLoS One 11, e0153104.
Romero, I.G., Ruvinsky, I., Gilad, Y., 2012. Comparative studies of gene expression and the evolution of gene regulation. Nat. Rev. Genet. 13, 505–516.
Salisbury, J.P., Sîrbulescu, R.F., Moran, B.M., Auclair, J.R., Zupanc, G.K.H., Agar, J.N., 2015. The central nervous system transcriptome of the weakly electric brown ghost knifefish (*Apteronotus leptorhynchus*): de novo assembly, annotation, and proteomics validation. BMC Genomics 16, 166.
Schulte, P.M., 2015. The effects of temperature on aerobic metabolism: towards a mechanistic understanding of the responses of ectotherms to a changing environment. J. Exp. Biol. 218, 1856.
Schulze, T.T., Ali, J.M., Bartlett, M.L., McFarland, M.M., Clement, E.J., Won, H.I., Sanford, A.G., Monzingo, E.B., Martens, M.C., Hemsley, R.M., Kumar, S., Gouin, N., Kolok, A.S., Davis, P.H., 2016. De novo assembly and analysis of the Chilean pencil catfish *Trichomycterus areolatus* transcriptome. J. Genomics 4, 29–41.
Seemann, S.E., Anthon, C., Palasca, O., Gorodkin, J., 2015. Quality assessment of domesticated animal genome assemblies. Bioinf. Biol. Insights 9, 49–58.
Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.
Sims, D., Sudbery, I., Ilott, N.E., Heger, A., Ponting, C.P., 2014. Sequencing depth and coverage: key considerations in genomic analyses. Nat. Rev. Genet. 15, 121–132.
Smith, S., Bernatchez, L., Beheregaray, L.B., 2013. RNA-seq analysis reveals extensive transcriptional plasticity to temperature stress in a freshwater fish species. BMC Genomics 14.
Unmack, P.J., Allen, G.R., Johnson, J.B., 2013. Phylogeny and biogeography of rainbowfishes (Melanotaeniidae) from Australia and New Guinea. Mol. Phylogenet. Evol. 67, 15–27.
Wang, Z., Sun, X., Zhao, Y., Guo, X., Jiang, H., Li, H., Gu, Z., 2015. Evolution of gene regulation during transcription and translation. Genome Biol. Evol. 7, 1155–1167.
Wellenreuther, M., Hansson, B., 2016. Detecting polygenic evolution: problems, pitfalls, and promises. Trends Genet. 32, 155–164.
Young, J.C., Moarefi, I., Hartl, F.U., 2001. Hsp90: a specialized but essential protein-folding tool. J. Cell Biol. 154, 267–274.