CrossMark

ORIGINAL ARTICLE

Genomics of Adaptation to Multiple Concurrent Stresses: Insights from Comparative Transcriptomics of a Cichlid Fish from One of Earth's Most Extreme Environments, the Hypersaline Soda Lake Magadi in Kenya, East Africa

Geraldine D. Kavembe[1,2] · Paolo Franchini[1] · Iker Irisarri[1] ·
Gonzalo Machado-Schiaffino[1] · Axel Meyer[1]

Abstract The Magadi tilapia (*Alcolapia grahami*) is a cichlid fish that inhabits one of the Earth's most extreme aquatic environments, with high pH ($\sim$10), salinity ($\sim$60 % of seawater), high temperatures ($\sim$40 °C), and fluctuating oxygen regimes. The Magadi tilapia evolved several unique behavioral, physiological, and anatomical adaptations, some of which are constituent and thus retained in freshwater conditions. We conducted a transcriptomic analysis on *A. grahami* to study the evolutionary basis of tolerance to multiple stressors. To identify the adaptive regulatory changes associated with stress responses, we massively sequenced gill transcriptomes (RNAseq) from wild and freshwater-acclimated specimens of *A. grahami*. As a control, corresponding transcriptome data from *Oreochromis leucostictus*, a closely related freshwater species, were generated. We found expression differences in a large number of genes with known functions related to osmoregulation, energy metabolism, ion transport, and chemical detoxification. Over-representation of metabolism-related gene ontology terms in wild individuals compared to laboratory-acclimated specimens suggested that freshwater conditions greatly decrease the metabolic requirements of this species. Twenty-five genes with diverse physiological functions related to responses to water stress showed signs of divergent natural selection between the Magadi tilapia and its freshwater relative, which shared a most recent common ancestor only about four million years ago. The complete set of genes responsible for urea excretion was identified in the gill transcriptome of *A. grahami*, making it the only fish species to have a functional ornithine-urea cycle pathway in the gills—a major innovation for increasing nitrogenous waste efficiency.

Keywords *Alcolapia grahami* · RNAseq · Urea cycle · Gene expression · Extremophile

✉ Axel Meyer
axel.meyer@uni-konstanz.de

1 Department of Biology, Chair in Zoology and Evolutionary Biology, University of Konstanz, 78457 Constance, Germany

2 Department of Biology, School of Applied and Pure Sciences, South Eastern Kenya University, P. O. Box 170, 90200 Kitui, Kenya

Introduction

East African cichlid fishes are well-known model systems for investigating a wide range of questions in evolutionary biology owing to their remarkable species-richness, history of explosive adaptive radiations, and astonishing phenotypic diversity (Kornfield and Smith 2000; Seehausen 2006; Kuraku and Meyer 2008). Another key feature of this group is the ability to establish viable populations in challenging habitats characterized by extreme environmental parameters, which gives them an edge over fish lineages confined to more benign habitats (Rijssel and Witte 2012). In fact, some major adaptive radiations of cichlids have been traced to periods of environmental instability characterized by fluctuations in lake levels (Sturmbauer et al. 2001; Danley et al. 2012). Successful colonization of novel niches outside the range of native habitats is a main avenue exploited by most vertebrates to diversify, including cichlids (Lande 2009). This colonization is dependent on the ability of a species to adjust to new

and often stressful environmental conditions (Yoshida et al. 2007). For instance, extreme water level fluctuations of Lake Tanganyika during the late Pleistocene may have played a role in shaping the adaptive evolution and speciation of its current rock-dwelling cichlid fishes (Verheyen et al. 1996).

A remarkable example of East African cichlids' evolutionary success and resilience to persistent and varied environmental stresses is the Magadi tilapia (*Alcolapia grahami*), a small cichlid fish (average adult weight ~3 g) endemic to Lake Magadi, a soda lake in Kenya (Fig. 1a). As in most endorheic soda lakes, Lake Magadi is characterized by extreme environmental conditions: high levels of pH (up to 10), alkalinity, salinity, temperature (up to 40 °C), reactive oxygen species (ROS), and ultraviolet radiation, plus extreme daily fluctuations in oxygen concentration (Coe 1966; Narahara et al. 1996; Johannsson et al. 2014). This unusually stressful environment poses several physiological challenges. High alkalinity, salinity, and temperature can cause degradation of cellular

Fig. 1 a Map showing the geographical locations of the sampling sites—Lakes Magadi and Victoria—that are the focus of this study. Photograph of the two sampling sites: **b** Winam Gulf, Lake Victoria and **c** Fish Spring Lagoon, Lake Magadi

biomolecules such as proteins (Rothschild and Mancinelli 2001). Likewise, high osmotic and ionic gradients brought about by high salinity need to be counteracted to avoid effects on cell structure and function (Gonzalez 2012). High pH often negatively affects nitrogen waste excretion, whereas intense ultraviolet radiation and ROS have detrimental effects on several cellular components including lipid peroxidation, protein carbonylation, and degradation of genetic material (DNA and RNA) (Rothschild and Mancinelli 2001). Surprisingly, despite these stressful conditions, *A. grahami* maintains viable populations in Lake Magadi, which do not show signs of reduced genetic diversity (Seegers et al. 1999; Wilson et al. 2000, 2004; Kavembe et al. 2014).

Organisms in stressful environments deploy an array of adaptive mechanisms to reduce or counteract the adverse effects of their environment. Both phenotypic plasticity or genetic modifications may provide means to respond to stress (Chevin et al. 2010; Yampolsky et al. 2014). Whether both act in an independent or complementary manner is still a widely debated topic (Merilä 2012; Merilä and Hendry 2014). Some studies suggest that populations may initially respond to stressful conditions through behavioral and phenotypic plasticity followed closely by genetic assimilation (canalization), which fixes beneficial traits that have been environmentally induced in the population (Badyaev 2005; Lande 2009). In *A. grahami*, several behavioral and physiological adaptations to its challenging habitat have been described. These include air breathing as a mitigation mechanism against recurring hypoxic conditions, excretion of nitrogen in form of urea (as opposed to ammonia excreted by almost all other teleosts), high metabolic rates, and tolerance to high water alkalinity facilitated by a high pH in blood and tissues (Johansen et al. 1975; Randall et al. 1989; Wood et al. 1989; Pörtner et al. 2010; Johannsson et al. 2014).

A. grahami can be successfully acclimated to freshwater conditions (Wood et al. 2002a, b). In contrast, closely related freshwater species have been shown to die within minutes upon transfer to Magadi water (Wright et al. 1990). As in typical euryhaline species, this unique ability to survive in both saline and freshwater environments is likely modulated by changes in the activity of ion transporters in the gills (Scott et al. 2004). Intriguingly, and in sharp contrast to other known euryhaline species, *A. grahami* has been shown to retain specific adaptations upon acclimation to freshwater conditions, such as obligate urea excretion and drinking of water at rates comparable to marine teleosts (Wood et al. 2002a, b). This suggests that *A. grahami* may possess some species-specific adaptations to cope with its extreme habitat.

The fact that *A. grahami*'s unique tolerance to environmental extremes may have been acquired within a very
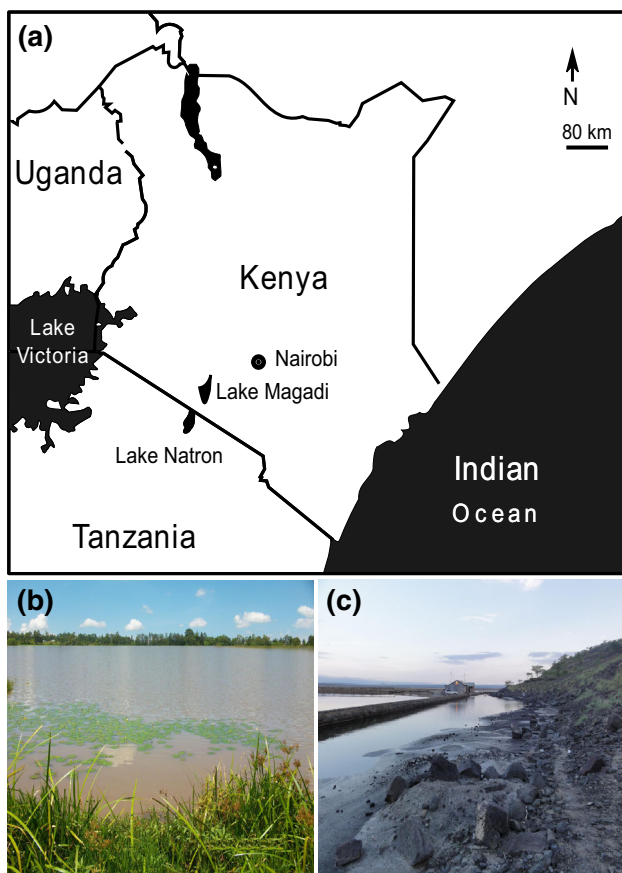
short time is of great evolutionary importance (Pörtner et al. 2010). Geological studies suggest the extreme conditions in Lake Magadi were initiated by a Pleistocene desiccation event (about 13,000 years ago) that resulted in the splitting of a freshwater paleolake Orolonga to form the present-day Lake Magadi, Lake Little Magadi (a small soda lake located North of Magadi), and Lake Natron (a large soda lake located South of Magadi). In a recent study (Kavembe et al. 2014), we inferred the closest extant relative of both A. grahami and its sister species from lake Natron to be Oreochromis variabilis, a freshwater species endemic to Lake Victoria. It has been suggested that Lake Victoria experienced a Pleistocene desiccation, which might imply a recent origin of its native cichlid species flock, composed of mainly haplochromine cichlids (Seehausen 2002). One might then expect an even much younger age for O. variabilis as well, similar to the tilapiine cichlids in Lakes Magadi and Natron.

Much is known about the physiological aspects of A. grahami (Pörtner et al. 2010; Wood et al. 2012), but virtually nothing is known about the molecular changes associated with acclimation to freshwater conditions. Several studies using closely related tilapia and other euryhaline teleost species have provided some insights on the molecular responses to environmental stresses (Boutet et al. 2006; Tine et al. 2008; Xu et al. 2013b; Lam et al. 2014). However, most of these studies focused on molecular responses associated with single environmental variables and could therefore not always capture the complexity of all challenging variables experienced by natural populations (Oleksiak 2011). Additionally, these previous studies were based on candidate gene approaches and microarray technology. These methods require some prior genomic knowledge of the study species, are expensive and limited by the number of genes they can capture, which render them inappropriate for studying non-model species such as A. grahami (Zhu and Zhao 2007; Wang et al. 2009).

With the recent developments in massive parallel sequencing technologies, sequencing of transcriptomes has become rapid and highly cost-effective, allowing the screening of thousands of genes at a real genome scale. A recent study examined the transcriptome-wide differences associated with ionic regulation in the gills of the euryhaline Oreochromis mossambicus (Lam et al. 2014), providing the first genome-scale view on the molecular responses associated with osmoregulation and gill remodeling in a tilapia species. However, like its predecessors, this study focused on a single environmental parameter (salinity), and thus, the genomic basis for responses to multiple extreme environmental variables in tilapia remains largely unexplored. Natural populations of A. grahami provide an excellent opportunity to examine the molecular responses

to multiple co-occurring stresses, both in the wild and following acclimation to freshwater conditions.

Here, we investigate the molecular mechanisms of stress tolerance in natural and freshwater-acclimated populations of A. grahami using a comparative transcriptomics approach based on RNA-seq. We sequenced gill transcriptomes of field-collected and laboratory-acclimated A. grahami in order to identify genes responsible for (i) diverse environmental stress conditions in Lake Magadi and (ii) acclimation to freshwater conditions (Fig. 2). Gills play a central role in the physiology of fishes, including respiration, acid–base balance, nitrogenous waste excretion, and ionoregulation (Evans et al. 2005). Therefore, gills are an ideal tissue to use when examining the transcriptomic responses to water chemistry in fish (Oleksiak 2011; Xu et al. 2013b; Lam et al. 2014). In order to increase the robustness of our analysis (Kelley et al. 2012; Xu et al. 2013b), we compared the transcriptomes of A. grahami with corresponding data from O. leucostictus, a close relative freshwater species from Lake Victoria (Fig. 2). This serves as a baseline in our comparisons, as it reveals the transcriptomic changes that occur during the acclimation process of a closely related obligate freshwater species.

Our study also sought to study the expression of the urea pathway in the gills of A. grahami. Since A. grahami exhibits obligate urea excretion (Wood et al. 2002a) that is retained after acclimation to freshwater conditions, we hypothesized that the genes of the urea pathway would display similar expression levels in both wild and acclimated samples, and that these genes are up-regulated in A. grahami relative to O. leucostictus.

Materials and Methods

Samples

Using hand nets, adult fish of A. grahami and O. leucostictus were captured at the Fish Spring Lagoon (1°53′00.52″S; 36°18′15.00″E) and the Winam Gulf (00°14′24″S; 34°34′48″E) within Lake Magadi and Lake Victoria, respectively (Fig. 1a–c, see Table 1 for water parameters). Within 15 min of fish capture, gill tissues of individual fish were excised following MS-222 euthanization and preserved in RNAlater (Ambion, Inc.). The samples were maintained at 4 °C overnight and later frozen at −20 °C. Additionally, free-swimming fry (<1 cm) of both species were collected from breeding pits within the vicinity of the sampling sites. The fry were transported in breathing bags (Kordon® LLC, USA) to the laboratory, acclimated, and raised under normal freshwater conditions (water pH 8, temperature 28 °C, dissolved oxygen 6–8 mg L^{-1}) for a period of 8 months. The fry were fed on Artemia during the
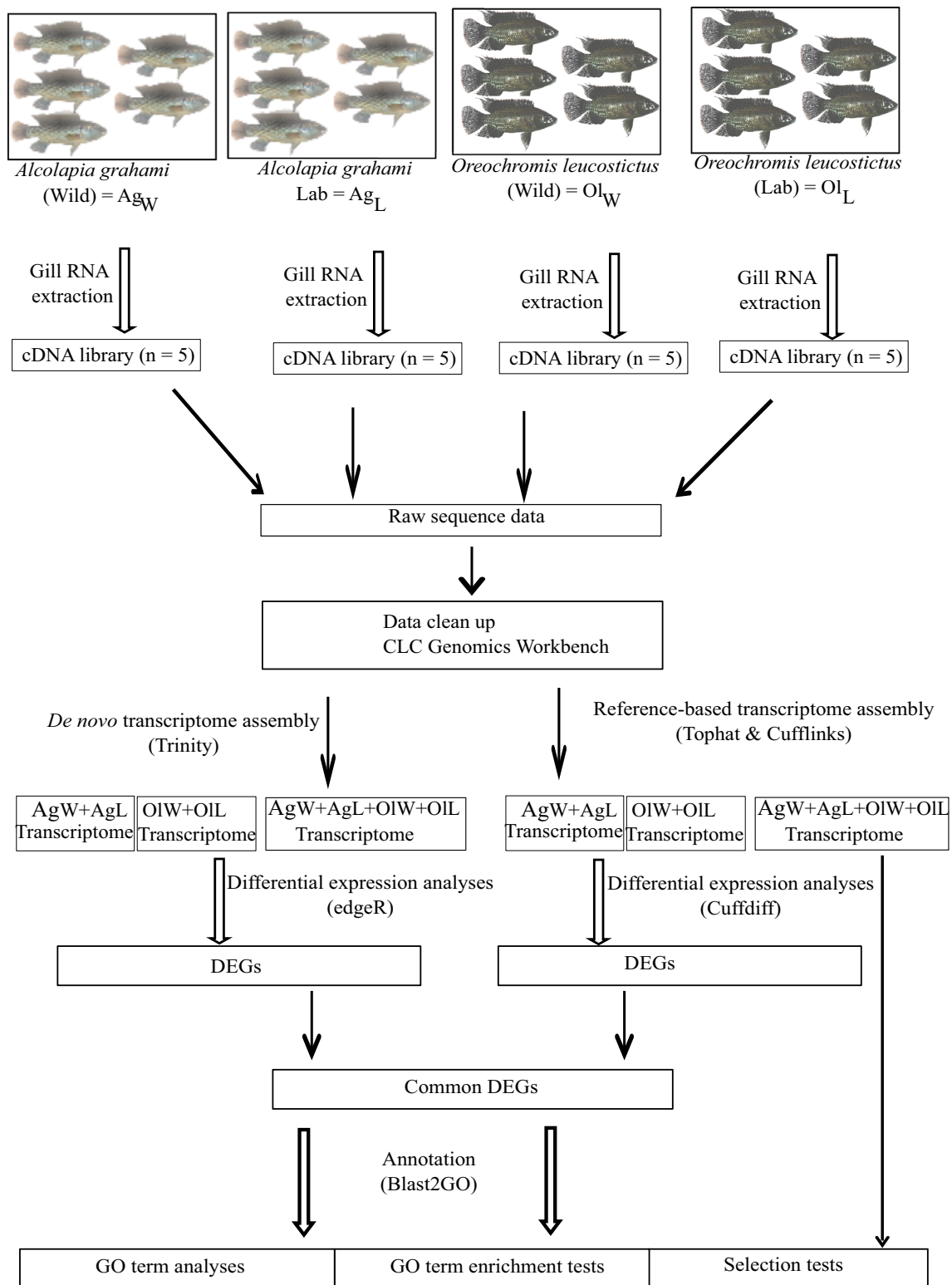
Fig. 2 Scheme representing bioinformatic pipeline used to analyze the gill transcriptomes of wild and laboratory-acclimated individuals of *A. grahami* and *O. leucostictus*

first few weeks and later maintained on flake food. As adults (bright male coloration and body length were used as a proxy to assess maturity), the gills of these freshwater-

acclimated fish were excised and preserved following similar procedures as those used in the field. A total of 20 gills were used in this study, obtaining five biological

Table 1 Physicochemical parameters of Lake Victoria and Lake Magadi (data adapted from Talling and Talling (1965) and Brauner et al. (2012))

Lake	pH	Temp (°C)	Dissolved O_2 (mg L^{-1})	Alkalinity (titrable mM)	Salinity (ppt)	Na^+ (mM)	Ca^+ (mM)	Mg^+ (mM)	Cl^- (mM)	SO_4^{2-} (mM)
Lake Victoria	8.0–8.8	23.5–28.0	4–9	0.758	0.04–0.07	0.42–0.57	0.12–0.36	0.06–0.229	0.11–0.19	0.02
Lake Magadi	7–9	37–40	0.3–3.16	290	21.3	356	0.65	0.04	112	0.8

replicates per experimental group (see below). Sampling was done with research permit NCST/RR1/12/1/MAS/99/4 of the Kenya National Commission for Science, Technology and Innovation (NACOSTI).

RNA Extraction and Massive Parallel Sequencing

Total RNA extraction and cDNA library preparation were performed as described in Henning et al. (2013), with slight modifications. Briefly, for each individual, approximately 50–100 mg of gill tissue was placed in 1 ml Trizol (Invitrogen, Carlsbad, USA) and homogenized using a Qiagen Tissue Lyser (Qiagen, Valencia, USA). The supernatant containing the RNA was separated from the rest of the homogenate by centrifugation, and standard Trizol RNA extraction was performed following manufacturer's instructions. RNA was purified using Qiagen RNeasy Mini Kit as per the manufacturer's instructions. The quality of the RNA was checked on Agarose gels (1.5 %) and with Bioanalyzer 2100 (Agilent technologies, Waldbronn, Germany). All RNA samples had RNA Integrity Number (RIN) scores above 7.5. Approximately, 1 μg of high-quality RNA was used to prepare cDNA libraries using the Illumina TrueSeq mRNA Preparation Kit (Illumina, San Diego, USA) as previously described (Henning et al. 2013). In total, 20 individually barcoded cDNA libraries were constructed representing five individuals and four experimental groups: wild *A. grahami* (Ag$_W$), lab *A. grahami* (Ag$_L$), wild *O. leucostictus* (Ol$_W$), and lab *O. leucostictus* (Ol$_L$) (Fig. 2). Equimolar quantities of all the individually barcoded libraries were pooled and paired-end (2 × 101 bp) sequenced in one lane of the Illumina HiSeq 2000 platform.

Read Processing and Assembly

All paired reads generated above were quality controlled before assembly, read mapping, and downstream analyses (Fig. 2). First, SeqPrep (https://github.com/jstjohn/SeqPrep) was used to remove any remaining adapter sequence and to merge overlapping read pairs. To avoid the generation of chimeric sequences, the minimum overlapping

length to merge reads was set to 15 bp, and a single bp mismatch was allowed only when the overlapping region was ≥50 bp. Paired and long merged reads were quality trimmed using CLC Genomics Workbench v7.0.4 (CLC bio, Aarhus, Denmark). Low-quality reads (CLC parameter 'limit' set to 0.02) and reads shorter than 30 bp were excluded. To control for contamination by external RNA sources, filtered reads were mapped against bacterial, fungal, viral, and protozoan sequences available at NCBI RefSeq (February 2014) using CLC Genomics Workbench and consequently discarded.

Having robust transcriptome assemblies is essential for performing reliable downstream analyses (Chang et al. 2014). Two main approaches are widely used for transcriptome assembly; reference-based and de novo methods. Reference-based approaches are more sensitive and robust to sequencing errors, but depend on the quality of the reference genome and the evolutionary distance to the target organism (Chang et al. 2014). In contrast, de novo methods can reconstruct novel and trans-spliced transcripts and are not affected by misassemblies in the reference genome (Martin and Wang 2011). In order to take advantage of both methods and increase the robustness of the transcriptome assemblies of our non-model species, we used both approaches complementarily.

Reference-based assemblies were generated by aligning the filtered reads to the Nile tilapia genome (Orenil1.0, Ensembl v75) using TopHat v2.0.12 (Trapnell et al. 2009). The resulting alignments were used to derive the transcripts with Cufflinks v2.0.2 (Trapnell et al. 2012). The Nile tilapia is closely related to both species in our study and thus represents an appropriate reference species. De novo assemblies were built with Trinity r2013_11_10 (Grabherr et al. 2011) using default parameters. For both assembly methods, three reference assemblies were generated by combining all reads across samples and biological replicates for (i) *A. grahami,* (ii) *O. leucostictus,* and (iii) a macro-assembly including all reads from the four experimental groups following (Haas et al. 2013), see below.

To test the reliability and completeness of our transcriptomes derived from gills, reconstructed transcripts from reference-based assemblies were subjected to

similarity searches using BLASTx against the NCBI's non-redundant (nr) protein database and the proteomes of four teleost species (tilapia, medaka, stickleback and zebrafish) (Ensembl v. 75). To compare the coverage of transcripts recovered by both reference-based and de novo analyses, all reconstructed transcripts were also subjected to BLASTx searches against the reference proteome of tilapia.

Differential Gene Expression Analyses

Differential gene expression analyses relied on the three reference transcriptomes assembled. This approach avoids the problem of a posteriori matching the "same" transcripts from independent assemblies, and thus it is generally preferred (Haas et al. 2013). To identify differentially expressed genes (DEGs), reads were mapped onto the corresponding reference transcriptomes separately for each individual to account for variability among biological replicates. Differential gene expression was estimated in the following pair-wise comparisons: (i) wild versus lab-acclimated *A. grahami* (Ag_W–Ag_L), (ii) wild versus lab-acclimated *O. leucostictus*, (Ol_W–Ol_L), (iii) wild *A. grahami* versus wild *O. leucostictus*, (Ag_W–Ol_W), and (iv) lab-acclimated *A. grahami* versus lab-acclimated *O. leucostictus* (Ag_L–Ol_L) (Fig. 2). For the reference-based assemblies, the module Cuffdiff was implemented in the Cufflinks package (Trapnell et al. 2012). For the de novo assemblies, gene expression was estimated using RSEM (Li and Dewey 2011) after mapping the reads with Bowtie v. 1.0.0 (Langmead et al. 2009), and differential gene expression was analyzed with edgeR (Robinson et al. 2010) as implemented in the Trinity pipeline. In all cases, a false discovery rate (FDR) threshold of 0.05 was used to correct for multiple comparisons (Benjamini and Hochberg 1995).

To ensure reliable inference of the transcriptomic differences between our experimental groups, only DEGs recovered by both de novo and reference-based methods were considered in downstream analyses. Reciprocal best BLASTn searches (using a stringent *e*-value cut-off of 1*e*-12) were used to identify differentially expressed transcripts that are shared between Cuffdiff (reference) and edgeR (*de novo*) analyses in each of the pair-wise comparisons. These sets of common DEGs were annotated using Blast2GO v.4 (Conesa et al. 2005) by BLASTx similarity searches (*e*-value < 1*e*-6) against the non-redundant database and the proteomes of four teleost species (tilapia, medaka, stickleback and zebrafish). Gene ontology (GO) terms were assigned using an annotation cut-off >55 and GO weight >5. Owing to the large number of DEGs obtained, gene enrichment tests were performed to provide an overview of gene clusters and pathways enriched in different experimental groups. To avoid transcript

redundancies, for each species, the reference-based and the de novo assemblies were merged, and similarity searches were carried out against the Nile tilapia protein dataset (BLASTx, *e*-value < 10*e*-6). GO terms of unique tilapia proteins with BLAST hits were then used as the reference dataset for the enrichment test. GO enrichment analyses were performed with Fisher's exact test, as implemented in Blast2GO to identify significantly over- and under-represented GO terms using the whole assembly as a reference. An additional enrichment test was performed between the assembled transcriptomes of *A. grahami* and those of *O. leucostictus*.

Survey of Published Stress-Response Genes Among DEGs and Urea Pathway Genes

To test for representation of stress-responsive genes in the transcriptomes generated in our study, we selected 25 genes with known physiological functions related to environmental stress (based on published literature) and searched for homologous transcripts among the identified DEG sets. These included genes related to hyper- and hypo-osmotic stress (Kalujnaia et al. 2007; Evans and Somero 2008; Laverty and Skadhauge 2012), heat stress (Fangue et al. 2006; Purohit et al. 2014), hypoxia and oxidative stress (Almeida et al. 2002; Woo et al. 2013) (see Table S1 for the complete list of genes). Additionally, we investigated the expression levels of the entire set of genes of the ornithine-urea cycle (OUC) pathway: *N*-acetylglutamate synthase (NAGS), ornithine carbamoyl transferase (OTC), carbamoyl-phosphate synthase III (CPSIII), argininosuccinate synthase (ASS), argininosuccinate lysase (ASL), and arginase (ARG), and one accessory urea pathway gene (ornithine glutamine synthetase (GS)) (Table 2). For each selected candidate gene, the protein sequence of the Nile tilapia was downloaded from Ensembl v.75, and used as query in TBLASTn searches (*e*-value < 1*e*-3 or 1*e*-5) against the estimated DEG transcripts (in the case of urea genes).

Tests for Signatures of Positive Selection in *Alcolapia grahami* Genes

Selection acting on protein-coding genes was studied by comparing the rates of non-synonymous (dN) and synonymous (dS) substitutions (Yang and Bielawski 2000). We used the yn00 method implemented in PAML v.4.8 (Yang 2007), which calculates pair-wise dN/dS ratios between sequences in an alignment. For these analyses, only the reference-based assemblies were used to avoid redundancy and because in principle they are more sensitive than de novo assemblies. The open reading frame for each transcript was detected using the query coordinates of

Table 2 Summary statistics of the transcriptome assemblies (de novo and reference-based) for *A. grahami* and *O. leucostictus*

	Reference-based			De novo		
	A. grahami	*O. leucostictus*	Macro-assembly	*A. grahami*	*O. leucostictus*	Macro-assembly
Raw reads	153141794	170412638	323554432	153141794	170412638	323554432
Clean reads	137993533	149928898	287922431	137993533	149928898	287922431
No. of transcripts	430896	406591	584655	47037	46212	57366
Number of genes	220040	208040	311359	28316	27957	32525
N50	3980	3981	4227	3482	3012	2482
Total length (bp)	733579864	600401818	717344879	125140067	123666880	158270486

tilapia proteins from the BLASTx output. Orthologous transcripts with a d*S* rate >0.1 were excluded from further analyses to avoid comparing paralogous genes (Bustamante et al. 2005). Both Yang-Nielsen (Yang and Nielsen 2000) and Nei-Gojobori (Nei and Gojobori 1986) methods were used for the d*N*/d*S* computations.

First, orthologous genes in *A. grahami* and *O. leucostictus* were inferred by a reciprocal best BLASTn hit (RBH) procedure against the Nile tilapia proteome (*e*-value < 1*e*-10), and the longest transcripts among those matching the same Nile tilapia protein were selected. To ensure accuracy in the identification of orthologs, only the pairs with the best similarity value with the orthologous tilapia protein were retained for downstream analyses. Positive selection is usually inferred if a gene has a d*N*/d*S* ratio >1 (Yang and Bielawski 2000). As a d*N*/d*S* ratio >0.5 calculated with the yn00 method has been suggested as an appropriate cut-off for genes under positive selection (Swanson et al. 2004), genes with d*N*/d*S* ratios >0.5 were deemed to show moderate signatures of positive selection in the current study.

Results

Transcriptomes of *A. grahami* and *O. leucostictus*

Sequencing of the gill transcriptomes using the Illumina HiSeq 2000 platform resulted in a total of 153,141,794 and 170,412,638 reads in *A. grahami* and *O. leucostictus*, respectively (Table 2). After read processing, approximately, 90 and 88 % clean reads were obtained for *A. grahami* and *O. leucostictus*, respectively (Table 2). One *O. leucostictus* sample rendered appreciably lower mapping values when aligned to the Nile tilapia genome (compared to all other samples, data not shown), and was eliminated from all further analyses. Reference-based assemblies of the clean reads yielded a total of 47,037, 46,212, and 57,366 transcripts with N50 values of 3980,

3981, and 4227 in *A. grahami*, *O. leucostictus*, and the macro-assembly in that order (Table 2). De novo assemblies yielded 430,896, 406,591, and 584,655 transcripts with N50 values of 3482, 3012, and 2482 for *A. grahami*, *O. leucostictus*, and the macro-assembly in that order (Table 2).

BLASTx similarity searches of the transcripts assembled using the Nile tilapia genome (i.e., using the reference-based method) yielded a high number of matches to the five searched databases: the NCBI's nr protein database and the four teleost protein datasets. For instance, BLASTx searches against the teleost proteins using *A. grahami* transcripts as queries recovered 76, 72, 70, and 51 % of the proteins present in the genomes of tilapia, medaka, stickleback, and zebrafish, correspondingly (Table 3). A similar analysis with *O. leucostictus* yielded very similar results (Table 3). In both cases, the number of homologous transcripts recovered consistently decreased with an increase in the phylogenetic distance between our study species and the model species; there was a higher degree of similarity of our transcriptomes to Nile tilapia and least with zebrafish. Similar findings were made in the transcriptome assembly of *Oreochromis mossambicus*, in which a high number of homologous contigs were found in the comparison with its congeneric species *O. niloticus* and were much lower in zebrafish and humans (Lam et al. 2014).

Overview of Gene Expression Patterns

The comparison between acclimated *A. grahami* and *O. leucostictus* (Ag$_L$–Ol$_L$) rendered the highest number of DEGs (5907, for Cuffdiff), while the comparison between the wild and acclimated *O. leucostictus* (Ol$_W$–Ol$_L$) produced the lowest number of DEGs (1485 for Cuffdiff) (Fig. 3a). The other two comparisons yielded intermediate numbers of DEGs (Ag$_W$–Ag$_L$ 3139 and Ag$_W$–Ol$_W$ 2213, for Cuffdiff) (Fig. 3a). This pattern was consistent for both Cuffdiff and edgeR analyses (Fig. 3a). Nevertheless, edgeR analyses (based on de novo assemblies) produced

Table 3 Similarity search (BLASTx) results of reference-based transcripts blasted against the proteomes of model teleost species

Description	Tilapia	Medaka	Zebrafish	Stickleback
Number of non-redundant (nr) protein sequences retrieved from Ensembl	26,763	24,674	43,153	27,576
Transcripts of *Alcolapia grahami* matching	20,328	17,815	22,039	19,314
Transcripts of *Oreochromis leucostictus* matching	20,108	17,574	21,874	18,974

Fig. 3 a Differentially expressed genes (DEGs) from de novo and reference-based methods. **b** Up- and down-regulated common DEGs among the pair-wise comparisons
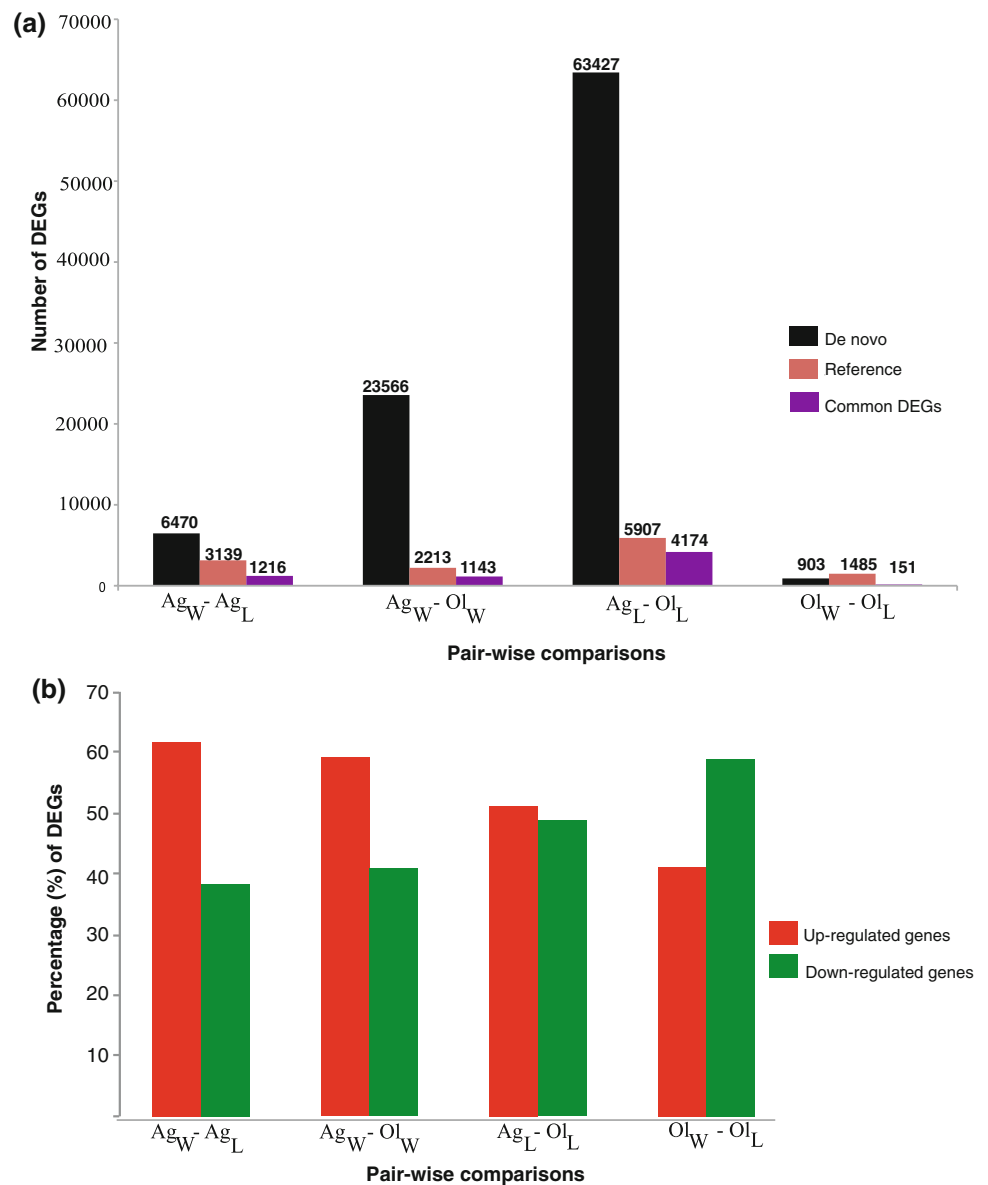
appreciably higher numbers of DEGs than the Cuffdiff analyses (Fig. 3a). Despite differences in the algorithms implemented in Cuffdiff and edgeR (Lu et al. 2013), this discrepancy in DEG numbers is most likely due to the higher number of transcripts assembled de novo (as compared to reference-based assemblies, see "Methods" section). Given that multiple individuals were pooled to assemble the transcriptomes, it is likely that Trinity may have inflated the real number of transcripts due to the presence of among-individual polymorphisms in the data.

This is further reinforced by the fact that both reference-based and de novo assemblies recover a similar total number of proteins (assessed by BLAST searches against the tilapia reference proteome, data not shown). In addition, de novo assemblies could contain trans-spliced transcripts and other transcripts that were misassembled or unassembled in the tilapia reference genome (Chen et al. 2011; Lu et al. 2013). Both Cuffdiff and edgeR analyses recovered a significant number of DEGs in common (Fig. 3a). The comparison between acclimated individuals of both species (Ag$_L$–Ol$_L$) yielded the highest number of common DEGs (4174), of which about half of the genes were up- (52 %) and down-regulated (49 %) (Fig. 3b). The comparison between wild and acclimated samples of O. leucostictus (Ol$_W$–Ol$_L$) yielded the smallest number (151) of DEGs with 41 % of up-regulated and 59 % up- and down-regulated transcripts (Fig. 3b). The pair-wise comparisons between the wild and acclimated A. grahami (Ag$_W$–Ag$_L$) and between wild specimens of the two species (Ag$_W$–Ol$_W$) uncovered intermediate numbers of common DEGs (1216 and 1143, respectively) genes (Fig. 3b).

Visualization of the common DEGs in a Venn diagram highlights that several DEGs (in the order of tens or hundreds, depending on the sets) were shared among pair-wise comparisons (Fig. 4a). The highest number (504) of DEGs was shared between the wild (Ag$_W$–Ol$_W$) and laboratory-acclimated (Ag$_L$–Ol$_L$) among-species comparisons, as well as between the latter and A. grahami species comparison (Ag$_W$–Ag$_L$) (Fig. 4a).

Survey of Published Candidate Genes, with Special Emphasis on the Ornithine-Urea Cycle

Screening of the DEGs using previously published stress-related candidate genes and from the urea cycle pathway recovered a significant number of homologous sequences (Tables S1, S2). For the 25 candidate genes queried, TBLASTn consistently found the same genes in the corresponding sets of DEGs from both Cuffdiff and edgeR pipelines at both 1e-5 (Table S1) and 1e-3 (results not shown) e-value cut-off values. Indeed, varying numbers of candidate genes were recovered among the pair-wise comparisons, consistent with the pattern we observed in the distribution of DEGs among our pair-wise comparisons. Except for the comparison between the A. grahami samples (Ag$_W$ vs. Ag$_L$), more candidate genes were consistently recovered from DEGs derived from the de novo methods (edgeR) (Tables S1, S2). However, for some candidate genes such as prolactin and some urea pathway genes, we could not find any homologous transcripts among the lists of differentially expressed genes (Tables S1, S2).

In our interest to study the expression patterns of the urea gene pathway (Fig. S1), we attempted to confirm the presence of all its constituent genes in the individual species transcriptomes by searching all the genes in the urea pathway in the assembled transcriptomes. We found the entire set of genes of the urea pathway (Table S2). Consistent with previous work suggesting the existence of a functional OUC cycle in A. grahami, all the OUC enzymes and the lead-in ammonia-trapping enzyme glutamine synthetase (GS) were expressed in both the wild and acclimated A. grahami (Tables S2–S4). Curiously, CPSIII, the rate-limiting enzyme of the OUC pathway was not found in the reference-derived (Table S2) assembly but it was found on the de novo based assemblies (Table S2). This emphasizes the need to consider both assembly methods in gene expression studies. Although most of the OUC genes were also present in O. leucostictus, the CPSIII gene that catalyzes the first committed step in the OUC pathway was not found (Table S2). Importantly, except for arginase, no significant expression changes were found in the OUC genes between wild and laboratory samples of A. grahami. In contrast, a significant up-regulation of the OUC genes was found between A. grahami and O. leucostictus in both wild and laboratory conditions (Tables S3, S4).

Gene Expression Changes Associated with Adaptation to Multiple Stresses in Wild A. grahami

We sought to identify DEGs in wild A. grahami relative to acclimated specimens (Ag$_L$) and wild O. leucostictus (Ol$_W$), as this might represent potential candidate genes for adaptation to stressful conditions in Lake Magadi. To do this, we searched for the overlap between up- and down-regulated annotated DEGs in wild A. grahami (Ag$_W$), first relative to laboratory-acclimated A. grahami (Ag$_L$) and then relative to wild O. leucostictus (Ol$_W$). A total of 317 and 247 annotated DEGs were up-regulated in Ag$_W$ relative to the Ag$_L$ and Ol$_W$, respectively, out of which 63 were commonly up-regulated in Ag$_W$ (Fig. 4b, Table S5). Generally, most of the up-regulated DEGs are associated with energy metabolism (e.g., transaldolase-like isoform × 1 and hexokinase-2-like isoform × 1) and responses to stress, e.g., pH (carbonic anhydrases), immunity (immunoglobulin superfamily member 1-like), and oxidative stress (superoxide dismutase-partial) (Table S5). Among the annotated down-regulated DEGs, 675 and 309 genes were down-regulated in Ag$_W$ compared to the Ag$_L$ and Ol$_W$, respectively (Fig. 4c, Table S5). Out of these, 47 common DEGs were down-regulated in Ag$_W$ compared to either Ag$_L$ or Ol$_W$, including genes involved in osmoregulation (e.g., mitogen-activated protein kinase 10-like, claudin-10a-like isoform x1n) and oxidation activities (e.g., mpv17-like) (Table S5).

Fig. 4 a Venn diagram
showing the distribution of
differentially expressed genes
(p value <0.05, FDR <0.05)
coincidentally estimated by both
reference-based and de novo
analyses in four statistical
comparisons of wild and
laboratory-acclimated *A.
grahami* and *O. leucostictus.*
The numbers of differentially
expressed genes for the four
main experimental groups are
shown in brackets. **b, c**
Common genes that are
differentially expressed in wild
A. grahami relative to
laboratory *A. grahami* and *O.
leucostictus.*
Diagram constructed with
VENNY ([http://bioinfogp.cnb.
csic.es/tools/venny/](http://bioinfogp.cnb.csic.es/tools/venny/))
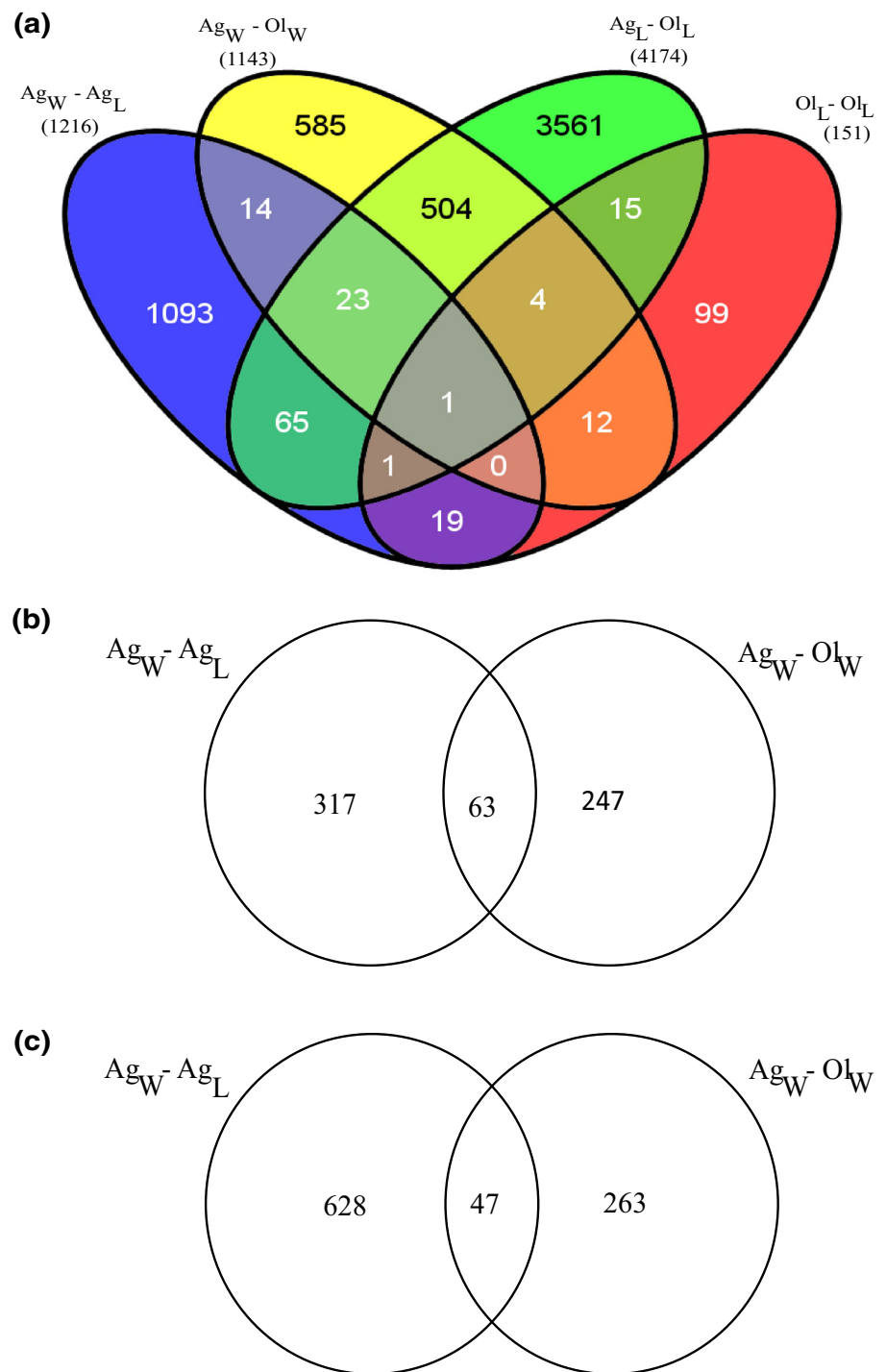
Functional Classification of Differentially Expressed Genes

Gene enrichment tests were used to identify significantly enriched gene categories among our experimental groups. In the *A. grahami* group comparison, relative to the laboratory samples (Ag_L), the gills of wild (Ag_W) samples were enriched in GO terms related to transmembrane activity (e.g., GO: 1901677 phosphate transmembrane transporter activity and GO: 0008509 anion transmembrane activity), as well as in metabolism-related processes (e.g., GO:0009226 nucleotide-sugar biosynthetic process and GDP mannose metabolic process) (Fig. 5a, Table S6). Among the down-regulated genes in wild *A. grahami* relative to their laboratory counterparts, GO enrichment yielded terms linked to cell nuclear processes (e.g.,
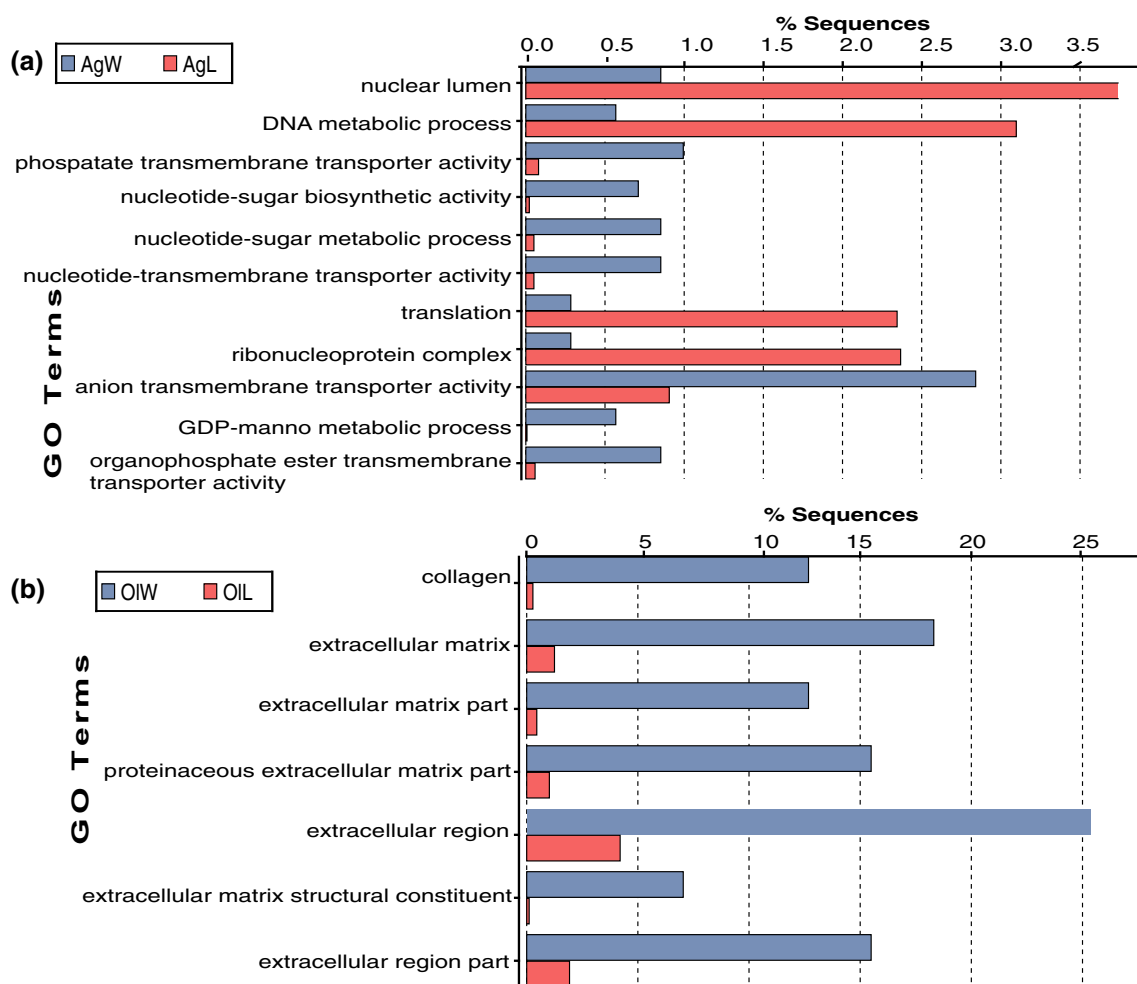
Fig. 5 Enriched GO terms of differentially expressed genes (DEGs) between **a** wild and lab-acclimated *A. grahami*; **b** wild and lab-acclimated *O. leucostictus* samples and between species **c** following laboratory acclimation and **d** in wild conditions

GO:0031981 nuclear lumen, GO:0006259 DNA metabolic process and GO:0006412 translation) (Fig. 5a, Table S6). In contrast, very few GO terms were assigned to the comparison between *O. leucostictus* groups suggesting only minimal transcriptomic changes occur in the species upon acclimation to laboratory conditions (Fig. 5b, Table S7). Nevertheless, a marked over-representation of GO terms related to extracellular matrix was observed in wild *O. leucostictus* relative to laboratory samples of the species (Fig. 5b, Table S7).

Between species, the GO terms that were over-represented in acclimated *A. grahami* compared to corresponding samples of *O. leucostictus* included terms related to stress (e.g., GO:0055114 oxidation reduction process, GO:0001666 response to hypoxia and GO:0050764 regulation of phagocytosis), calcium metabolism (e.g., GO:0005544 calcium-dependent phospholipid binding, or transmembrane activity (GO: 1901677 phosphate trans-membrane transporter activity) (Fig. 5c, Table S8), among

others. Under-represented GO terms in *A. grahami* for this comparison included nuclear-related process (e.g., GO:0006396 RNA processing, GO:0003676 nucleic acid binding) (Fig. 5c, Table S8). Similarly, compared to wild *O. leucostictus*, the wild samples of *A. grahami* revealed over-representation of terms related to stress (e.g., GO:0009617 response to bacterium and GO:0009410 response to xenobiotic stimulus) (Fig. 5d, Table S9). Other terms that were over-represented in *A. grahami* compared to *O. leucostictus* under natural conditions (Ag_W–Ol_W) included terms related with several metabolic processes and catalytic activities (Fig. 5d, Table S9).

Genes Showing Signatures of Positive Selection

Out of a total of 22,746 and 22,699 different transcripts obtained by the reference-based analysis in *A. grahami* and *O. leucostictus,* respectively, the application of stringent screening criteria resulted in 14,618 putative orthologous

Fig. 5 continued

(c) ☐ AgL ☐ OIL
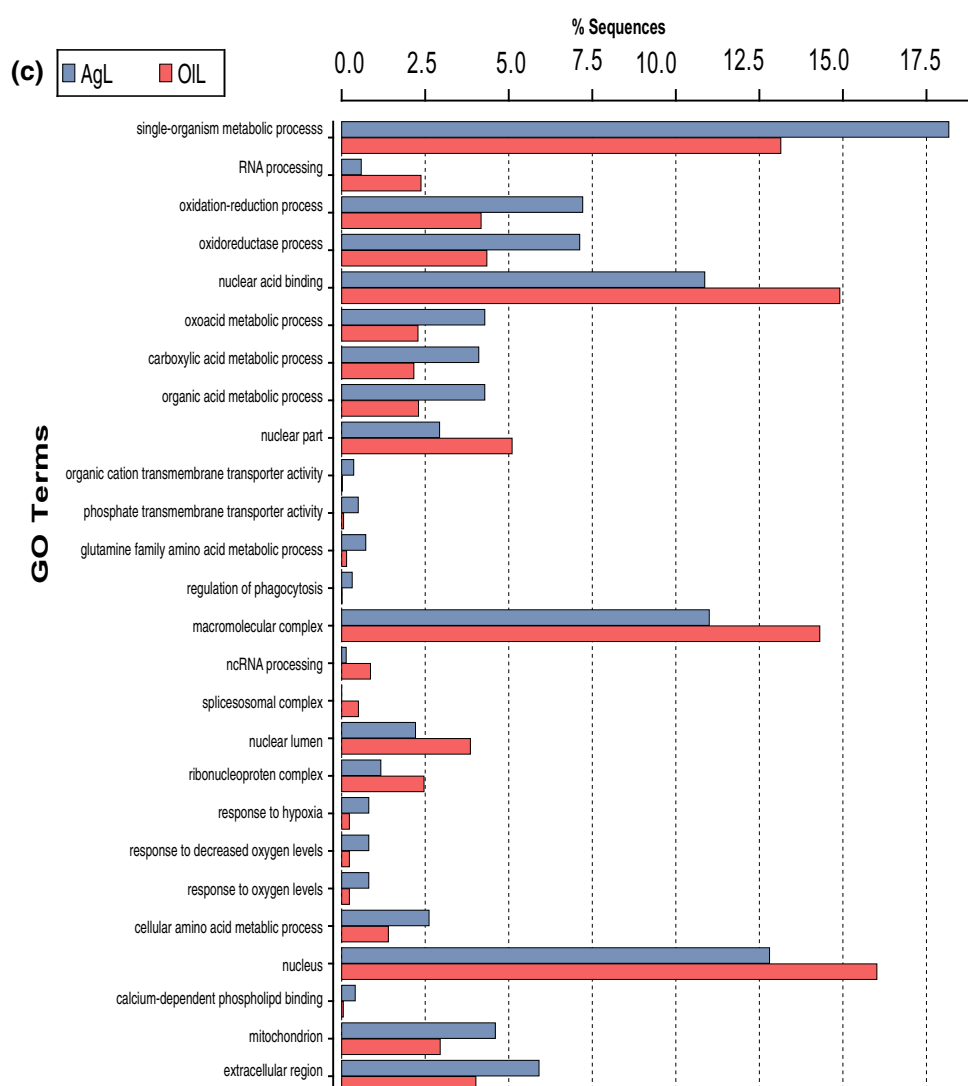
genes for the two species. Despite this high number of genes, only 25 genes (representing 0.17 % of the orthologous genes) showed signatures of positive selection using the yn00 method. 10 genes (40 %) had dN/dS > 1 suggestive of signs of strong positive selection and 15 (60 %) genes had 0.5 < dN/dS < 1 representing signatures of moderate positive selection (Table 4). The Nei–Gojobori method recovered seven with dN/dS ratios above one (dN/dS > 1) and a further 15 genes with dN/dS ratios above 0.5 (0.5 < dN/dS < 1) (Table 4). Annotation of the transcripts showing signatures of positive selection resulted in genes orthologous to annotated genes in the nr database, most of which have physiological functions related to stress (Table 4). For instance, among the genes showing strong positive signatures of selection (dN/dS > 1) were genes with functions related to chemical detoxification (*CYP3A40*), immune system (AEBP1), cellular transport (PIGS), and neural development (SYDE2) (Table 4). Other

genes showing signs of positive selection have functions related to sensory perception, tumor suppression, and cellular detoxification. However, GO enrichment tests of the genes did not yield any enriched GO terms at FDR < 0.05.

Discussion

De Novo and Reference-Based Approaches for Differential Gene Expression Analysis

To date, most published differential gene expression analyses have relied solely on data derived from either reference-based or de novo analyses (Lu et al. 2013), despite the potential drawbacks of either methods. If used complementarily, both methods can significantly improve the robustness of differential expression analysis. The

Fig. 5 continued
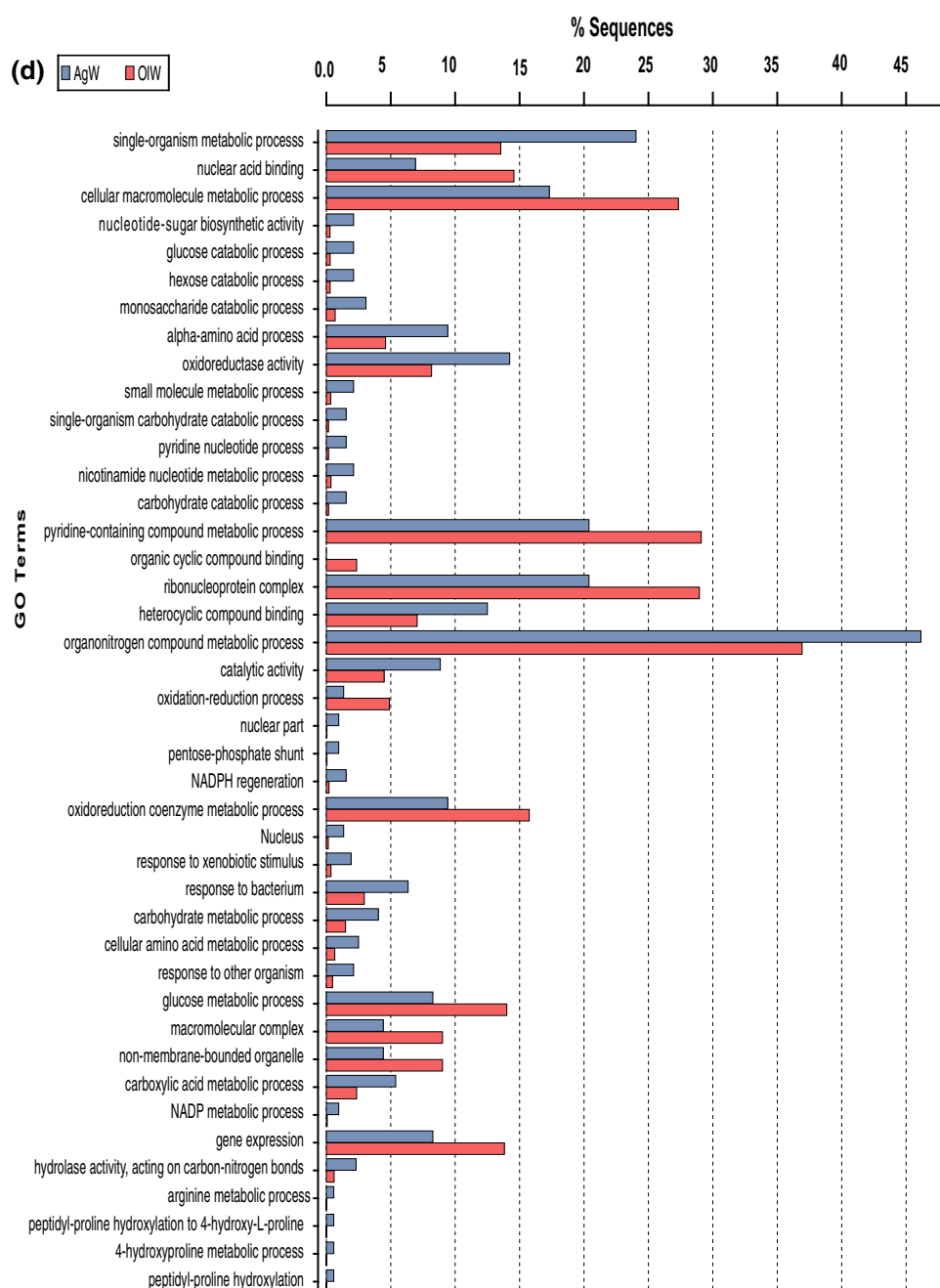
present study exploited this approach by examining gene expression patterns on datasets derived from both reference-based and de novo assemblies. More specifically, we used transcriptome data derived from Trinity which have been shown to recover the maximum number of genes and Tophat-Cufflinks that yields full-length transcripts (Jain et al. 2013), thus increasing the robustness of our analyses. Despite the differences in the algorithms employed by the two methods, significantly large sets of common DEGs were recovered by both approaches. Moreover, the relative expression levels of genes between conditions were consistent in both analysis pipelines. By interrogating only the common DEGs, our complementary approach allowed us to select a more robust set of DEGs that represent possible candidate genes associated with adaptation to stress in both hypersaline-adapted and freshwater-acclimated *A. grahami*.

Gill Transcriptomes to Study Abiotic Stress in Fish

The reliability of differential expression analyses depends largely on the sequencing depth and the number and

Table 4 Genes showing signs of positive selection in *A. grahami*

Gene ID	Uniprot ID	Description	dN/dS-yn	dN/dS	Putative function
AEBP1	Q8IUX7	Adipocyte enhancer-binding protein 1-like	2.167	1.5	Immune response
EPB41L3	Q9Y2J2	Band -like protein 3-like isoform x7	0.739	0.6	Apoptosis/tumor suppression
BOC	Q9BWV1	Brother of cdo-like	0.564	0.628	Myogenesis
cyp3a40	Q98T91	Cytochrome p450 3a40-like	2.606	2.147	Oxidative metabolism of Environmental chemicals and endogenous compounds
HUWE1	Q7Z6Z7	e3 ubiquitin-protein ligase huwe1-like	0.5	0.5	Apoptosis
WWP1	Q9H0M0	e3 ubiquitin-protein ligase nedd4-like isoform x2	0.857	1.087	Component of the pain pathway
EHD3	Q9NZN3	eh domain-containing protein 3	1	<0.5	Endocytic transport
SLC1A3		Excitatory amino acid transporter 1-like	0.974	0.731	Neural transmission
FRRS1	P43003	Ferric-chelate reductase 1-like	1.231	0.974	Oxidation–reduction process (Fe^{3+} to Fe^{2+})
PIGS	Q6ZNA5	gpi transamidase component pig	1.05	0.96	Cellular transport
GIMAP1	Q96S52	gtpase imap family member 7-like	0.873	0.629	Immune response
HLA-DRA	Q8NHV1	hla class ii histocompatibility dr alpha chain-like	1.5	1.787	Auto immune response
METTL9	P01903	Methyltransferase-like protein 9-like	0.674	0.581	Immune response
NF1	P11881	Neurofibromatosis type 1	0.789	0.75	Tumor suppressor activity
NUCB2	Q9H1A3	Nucleobindin 2a	0.967	0.75	Calcium homeostasis, Anorexigenic activity
PICALM	P21359	Phosphatidylinositol-binding clathrin assembly isoform x5	1.2	0.917	Involved in clathrin-mediated endocytosis
IQGAP3	P80303	ras gtpase-activating-like protein iqgap3-like	0.556	0.574	Cytoskeleton organization
RAB30	Q13492	ras-related protein rab-30-like isoform x2	0.734	0.77	Intracellular membrane trafficking.
SYDE2	Q86VI3	rho gtpase-activating protein syde2-like	1.313	1.111	Involved in neuronal development
SLC25A47	Q923S9	Solute carrier family 25 member 47-a-like	1.228	1.28	Formation and maintenance of tight junctions
SULT2B1	Q5VT97	Sulfotransferase family cytosolic 2b member 1-like	0.526	<0.5	Transmembrane transport
TJP2	Q6Q0C1	Tight junction protein zo-2-like isoform x5	1.1	0.5	Development
YTHDF2	O00204	yth domain family	0.707	1	Intracellular
Tol2	–	tol2 transposon gene	0.943	0.759	Gene transfer vector

diversity of transcripts that can be reconstructed. This is in turn dependent on the choice of the target tissue, number of biological replicates, and number of conditions (treatments), among other factors (Strickler et al. 2012; Romero et al. 2012; Sims et al. 2014). Until recently, it was thought that in order to capture the highest number of transcripts, it is desirable to sample various tissues across developmental stages, sexes, and treatments, which can then be pooled and sequenced for gene expression studies (Strickler et al. 2012). This approach is not only costly, but importantly, it increases the number of factors (variance components) to consider when analyzing data derived from such studies. Several recent studies, however, have shown that fairly complete transcriptomes can also be reconstructed from single tissues such as gill (Kelley et al. 2012) and testis (Soumillon et al. 2013). Kelley et al. (2012) speculated that the gill may contain exceptionally high numbers of

transcripts, making it a good tissue for transcriptome assembly. Indeed, in the current study, we were able to capture up to 76 % of the genes present in the reference proteome of tilapia using single adult fish gill tissues (Table 3).

Gene Expression Changes in *A. grahami*

Overall, our study recovered a high number of differentially expressed genes associated with various physiological adaptations in *A. grahami* in comparison to *O. leucostictus* in both laboratory and wild conditions. The up- and down-regulation of genes might reflect short-term adaptations through plasticity and/or long-term genetic adaptations. When comparing gene expression levels among species, differences may result both from treatments (environment) and from species-specific differences. Both

factors may in turn have compounding effects on gene expression (Verne et al. 2011; Romero et al. 2012).We would thus expect high numbers of DEGs in the comparisons between species both in the wild (Ag$_W$–Ol$_W$) and laboratory (Ag$_L$–Ol$_L$) conditions, and many shared DEGs between these two comparisons. However, a high number of DEGs were exclusive to the comparison between species under laboratory conditions (4174 DEGs in Ag$_L$–Ol$_L$ vs. 1143 DEGs in Ag$_W$–Ol$_W$, (Figs. 3a, 4a) and only about 500 DEGs were shared between the comparisons of the species in the wild and laboratory conditions (Fig. 4a).

A possible explanation for the high number of DEGs between the two lab-acclimated samples could be a result of constitutive expression of stress genes in *A. grahami*. In such a case, genes that would typically be expressed under Magadi-like water conditions would remain expressed even following acclimation to freshwater conditions. Since such a scenario does not necessarily mean that freshwater inducible genes are not expressed, a combination of genes derived from both effects may lead to increased numbers of DEGs. This phenomenon is best understood in the context of genetic assimilation, in which genes and consequently traits that are unconditionally adaptive become genetically fixed (canalized) such that they may continue to be expressed long after the stressor has been removed (Stearns 1994; Badyaev 2005; Lande 2009). Several studies suggest that constitutive up-regulation (transcriptome uploading) of stress-responsive genes may actually represent a major phenomenon in animals evolved under chronic environmental stress (Latta et al. 2012; Purohit et al. 2014). For instance, *Trematomus bernacchii* (family Nototheniidae) and several other Antarctic fishes have been shown to have lost the ability to induce heat shock proteins but constitutively express heat shock proteins (Buckley et al. 2004; Shin et al. 2014). In their case, this adaptation has been suggested as an evolutionary innovation to deal with physiological challenges associated with chronic cold exposure such as denaturation of proteins (Shin et al. 2014). Constitutive expression of heat shock proteins has also been suggested in the hot spring-adapted *Channa striatus* (family Channidae) that are exposed to high temperatures of 36 °C (Purohit et al. 2014). In *A. grahami,* expression of urea genes in freshwater conditions (see section on urea genes) could represent a classical example of constitutive expression of genes in response to chronic stress, in this case high alkalinity and pH, which make it impossible for these soda lake-adapted cichlids to excrete ammonia, as other teleosts do.

Adaptive Gene Expression Differences Between Wild and Laboratory Conditions of *A. grahami*

Within species, a high number of DEGs were found between *A. grahami* in wild (hypesaline) versus laboratory (freshwater) conditions (Ag$_W$–Ag$_L$, 1,216 DEGs) as compared to a similar comparison in *O. leucostictus* (Ol$_W$–Ol$_L$, 151 DEGs) (Figs. 3a, 4a). This is consistent with the contrasting differences in water conditions between the native habitat of *A. grahami* and those in the laboratory. This also suggests that *A. grahami* can modulate its gene expression according and most appropriately to the necessary environmental conditions. In the wild, natural populations of *A. grahami* are exposed to hyper-osmotic stress, while under laboratory (freshwater) conditions, they are exposed to extreme hypo-osmotic stress compared to their natural environment (Wood et al. 2002a). Thus, substantial physiological differences are expected in *A. grahami* between the two extremely different water conditions. In the case of *O. leucostictus*, the laboratory water conditions are more similar to its native freshwater habitat and hence the species undergoes minimal physiological adjustments, as revealed by the small number of DEGs.

In concordance with previous reports that *A. grahami* has one of the highest metabolic rates owing to its active life style, several genes and GO terms related to metabolism were recovered in various pair-wise comparisons involving wild *A. grahami* (Fig. 5a, d). We particularly noted an over-representation of metabolic-related terms in the comparison between wild versus acclimated *A. grahami,* suggesting that acclimation to laboratory conditions greatly decreases the metabolic requirements of this species. Reduced metabolic rates in *A. grahami* following acclimation to freshwater conditions have previously been measured directly, and are thought to result from the removal of the high cost of acid–base regulation in Lake Magadi water (Wood et al. 2002a). Similar results were reported in the closely related cichlid *Sarotherodon melanotheron*, in which fish acclimated to hypersaline water showed an up-regulation of genes related to metabolism when compared to specimens acclimated to freshwater (Tine et al. 2008).

Further screening of the DEGs revealed several genes of the claudin superfamily that were differentially expressed between wild and acclimated *A. grahami* as well as between *A. grahami* and *O. leucostictus,* in both wild and laboratory conditions (Table S5). Up-regulation of claudin genes in gills has been suggested in other species following the transfer of fish from saline water to freshwater and has also been shown to be more highly expressed in freshwater than in saline water fish (Czesny et al. 2012; Lam et al. 2014) suggesting their role in osmoregulation. Consistent with this pattern, several of the isoforms of claudin were up-regulated in laboratory-acclimated specimens of *A. grahami* compared to *O. leucostictus*, and were down-regulated in wild *A. grahami*. Several members of the mitogen-activated protein kinase (MAPK) pathway, another important pathway induced in response to osmotic stress (Kültz and Burg 1998; Fiol and Kültz 2007),

displayed significant expression differences in *A. grahami* and *O. leucostictus* both in the wild and lab conditions (Table S5). Other highly expressed genes in wild *A. grahami* compared to laboratory samples of the species and wild *O. leucostictus* include sodium potassium-transporting ATPase (Na^+/K^+-atpase) and angiotensin-related genes (Table S5), both of which have also been implicated in response to ionic and/or osmotic stress (Boutet et al. 2006). We also found an up-regulation of genes related to carbonic anhydrase activity. This enzyme has been implicated in ion regulation, carbon dioxide (CO_2) excretion, and acid–base balance in the gills of euryhaline fish (Gilmour and Perry 2009; Gilmour 2012). Indeed, in all comparisons involving wild *A. grahami*, several isoforms of carbonic anhydrase were found, emphasizing the importance of this enzyme in the metabolism of this species under its native challenging environmental conditions in the wild (Table S5).

In general, there was over-representation of GO terms related to body detoxification (e.g., oxidase-reductase activity), especially between the comparison of the wild samples of *A. grahami* and *O. leucostictus* (Fig. 5d). Further examination of the DEGs revealed a number of genes belonging to the cytochrome P450 gene family and other genes involved in ROS detoxification (e.g., glutathione reductases and superoxide dismutase) among all the pairwise comparisons considered. These genes are involved in various metabolic activities including oxidative, peroxidative, and reductive metabolism of endogenous and exogenous substrates (Danielson 2002). Lake Magadi is a closed lake in which intensive ultraviolet radiation, high salinity, high pH, and high alkalinity, together with daytime hyperoxia synergize to produce high concentrations of ROS (Johannsson et al. 2014).

In relation to hypoxia stresses, *A. grahami* are exposed to changing partial oxygen pressures, which are brought about by photosynthetic and respiratory activity of cyanobacteria that increase and decrease the levels of dissolved oxygen during day and night, respectively (Narahara et al. 1996). We thus expected to find several genes and enriched terms related to hypoxia in wild *A. grahami*. Although several isoforms of hypoxia-inducible genes (e.g., cytochrome c oxidase, hypoxia-inducible factor 1, and hemoglobin subunit beta-a) were up-regulated in wild *A. grahami* compared to other groups, no GO terms related to hypoxia were enriched in this group. This may as well be attributed to various alternative defense mechanisms present in wild *A. grahami* that guard the fish against gaseous assault including air breathing facilitated by the presence of an air sac (Maina et al. 1996a), a thin water-blood barrier in the gill epithelium (Maina et al. 1996a, b) and a modification in hemoglobin in order to increase oxygen absorption especially during anoxic conditions (Lykkeboe et al. 1975). Increased

hemoglobin-oxygen affinity and improved oxygen transport have also been suggested to increase hypoxia tolerance in Lake Victoria cichlids exposed to prolonged hypoxia stress (Rutjes et al. 2007).

Lastly, several molecular modifications are employed in fish to guard the cell against long-term harmful effects of thermal stress, among them the recruitment of molecular chaperones, especially of the heat shock family. Surprisingly, compared to previous studies (Xu et al. 2013a; Narum et al. 2013), only a few heat shock proteins (HSP70, HSP30, and HSC70) were found up-regulated in *A. grahami* in our study, again suggesting either constitutive expression of heat shock proteins and/or the existence of alternative adaptive mechanisms in *A. grahami* for dealing with chronic temperature stress. For instance, the hemoglobin proteins in *A. grahami* have been shown to exhibit a high degree of thermostability (Franklin et al. 1994), which ensures a continuous supply of oxygen to cells despite the high temperatures. Such alternative physiological modifications in *A. grahami* could provide important complementary and probably more reliable long-term adaptive strategies against chronic temperature stress compared to the short-lived molecular chaperones (e.g., HSPs).

Novel Expression of Ornithine-Urea Cycle Pathway Genes in the Gills of *A. grahami*

A key finding in the present study is the recovery of all the genes of the OUC pathway in the gill transcriptome of *A. grahami*. To the best of our knowledge, this is the first report of a functional OUC pathway in the gill tissue of any teleost fish. By expressing the complete complement of the OUC pathway genes, the gill becomes another unusual tissue known to express urea pathway genes in *A. grahami,* following an earlier equally surprising finding of the complete OUC enzyme repertoire in the muscles (Randall et al. 1989). The liver is the usual site for urea enzyme activity in terrestrial ureotelic vertebrates and some teleosts (Wright and Wood 2009). Although it is generally agreed that OUC pathway genes may actually be present in the early life stages of most teleost species, most adult teleost fish are ammoniotelic (Wright and Fyhn 2001). In fact, the expression of OUC pathway in adults has only been described in a few teleost species and *A. grahami* is so far the only documented obligate ureotelic teleost. We suggest that the presence of a functional OUC pathway in the gills, together with the ammonia-trapping lead-in enzyme glutamine synthetase (GS) in *A. grahami*, could offer the first step in piecing together the puzzle on how *A. grahami* continues to excrete urea in freshwater conditions (Wood et al. 1989, 1994, 2002a; Wright et al. 1990) in which ammonia excretion should be favored,

given its relatively high levels in its bloodstream (Wood et al. 1989). With a functional OUC in the gills, any ammonia diffusing through the gill epithelium would be trapped and converted into urea before being expelled into the surrounding water.

Since CPSIII represents the first committed step in the OUC pathway, its absence in *O. leucostictus* led us to suggest that the species does not have a functional OUC in its gills. Given that we detected the expression of all other OUC genes in the species, it may however be useful to screen other teleost species for mRNA and/or enzymatic evidence of OUC expression, especially those that are facultative ureotelic as adults.

Discordance Between Number of DEGs and Number of Genes Under Positive Selection

Compared to the high number of differentially expressed genes recovered in our study, transcriptome-wide screening identified only 25 genes as having experienced positive selection. The relatively low number of positively selected genes versus the high number of DEGs in our study suggests that most adaptive responses to multiple stresses in *A. grahami* may be driven by variation in gene expression as opposed to genetic modifications. Nevertheless, most of the positively selected genes in our study are associated with physiological functions related to stress (Table 4). *AEBP1* (dN/dS 2.167) plays a key role in apoptosis and cell survival by facilitating enhanced macrophage inflammatory responsiveness (Majdalawieh et al. 2007). Under extreme environmental stress, the immune system of fish is greatly compromised, and thus, genes with functions related to immune response such as AEBP1 are expected to be under strong positive selection in stressful environments. Members of the CYP family (including CYP3A40 with the highest dN/dS of 2.606 in our study) have been associated with detoxification processes such as in the European eel (Pujolar et al. 2012). Owing to high levels of ROS and other free radicals associated with numerous stresses in Lake Magadi (Johannsson et al. 2014), genes that promote their destruction would be selectively propagated in the genome.

Interestingly, the transcripts showing signatures of positive selection were not among the list of differentially expressed transcripts. This suggests that the evolution of gene expression and sequence divergence between species may involve different genes with related functions. A recent study of killifish species (family Fundulidae) adapted to contrasting osmotic (freshwater versus saline water) gradients found that adaptation to osmotic stress involved different sets of genes, some showing divergence in gene expression and others in coding sequences (Kozak

et al. 2014). The authors suggested that as populations adapt to complex physiological challenges, gene expression and coding sequences may evolve independently but converge in function to ensure that an optimal phenotype for the respective environments is achieved (Kozak et al. 2014). However, the relationship between gene expression and protein sequence evolution remains largely debated. Some studies support decoupling of gene expression from sequence evolution as is suggested by our study, whereas several other studies have suggested a positive correlation between the two evolutionary processes in several species including bacteria, yeast (Kim and Yi 2007), *Drosophila* (Lemos et al. 2005), ants (Hunt et al. 2013), and mammals (Khaitovich et al. 2005; Warnefors and Kaessmann 2013). Several factors have been suggested that may influence the evolution of protein sequences and gene expression including organismic attributes (e.g., protein–protein interactions), mutation rates, and the strength of selection, which in turn determine if the two processes are coupled or decoupled.

Conclusion

Our study represents the first genome-wide profile of expressed genes in *A. grahami* from natural and laboratory-acclimated fish. Our results provide new important insights into the molecular mechanisms underpinning unique adaptations to multiple abiotic stressor environments such as the Magadi soda lake. The differentially expressed genes, as well as the genes showing signatures of positive selection provide potential candidates for further studies investigating the molecular aspects of stress tolerance in *A. grahami* and other extremophile species.

References

Almeida JA, Diniz YS, Marques SFG et al (2002) The use of the oxidative stress responses as biomarkers in Nile tilapia (Oreochromis niloticus) exposed to in vivo cadmium contamination. Environ Int 27:673–679

Badyaev AV (2005) Stress-induced variation in evolution: from behavioural plasticity to genetic assimilation. Proc R Soc B 272:877–886. doi:10.1098/rspb.2004.3045

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B 57:289–300

Boutet I, Long Ky CL, Bonhomme F (2006) A transcriptomic approach of salinity response in the euryhaline teleost, Dicentrarchus labrax. Gene 379:40–50. doi:10.1016/j.gene.2006.04.011

Brauner CJ, Gonzalez RJ, Wilson JM (2012) Extreme environments: hypersaline, alkaline, and ion-poor waters. Fish Physiol 32:435–476. doi:10.1016/B978-0-12-396951-4.00009-8

Buckley BA, Place SP, Hofmann GE (2004) Regulation of heat shock genes in isolated hepatocytes from an Antarctic fish, Trematomus bernacchii. J Exp Biol 207:3649–3656. doi:10.1242/jeb.01219

Bustamante CD, Fledel-Alon A, Williamson S et al (2005) Natural selection on protein-coding genes in the human genome. Nature 437:1153–1157. doi:10.1038/nature04240

Chang Z, Wang Z, Li G (2014) The impacts of read length and transcriptome complexity for de novo assembly: a simulation study. PLoS One 9:e94825. doi:10.1371/journal.pone.0094825

Chen G, Li R, Shi L et al (2011) Revealing the missing expressed genes beyond the human reference genome by RNA-Seq. BMC Genom 12:590. doi:10.1186/1471-2164-12-590

Chevin L-M, Lande R, Mace GM (2010) Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. PLoS Biol 8:e1000357. doi:10.1371/journal.pbio.1000357

Coe MJ (1966) The biology of Tilapia grahami Boulenger in Lake Magadi, Kenya. Acta Trop XXIII:146–175

Conesa A, Götz S, García-Gómez JM et al (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676. doi:10.1093/bioinformatics/bti610

Czesny S, Epifanio J, Michalak P (2012) Genetic divergence between freshwater and marine morphs of alewife (Alosa pseudoharengus): a "next-generation" sequencing analysis. PLoS One 7:e31803. doi:10.1371/journal.pone.0031803

Danielson PB (2002) The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans. Curr Drug Metab 3:561–597

Danley PD, Husemann M, Ding B et al (2012) The impact of the geologic history and paleoclimate on the diversification of East african cichlids. Int J Evol Biol. doi:10.1155/2012/574851

Evans TG, Somero GN (2008) A microarray-based transcriptomic time-course of hyper- and hypo-osmotic stress signaling events in the euryhaline fish Gillichthys mirabilis: osmosensors to effectors. J Exp Biol 211:3636–3649. doi:10.1242/jeb.022160

Evans DH, Piermarini PM, Choe KP (2005) The multifunctional fish gill: dominant site of gas exchange, osmoregulation, acid-base regulation, and excretion of nitrogenous waste. Physiol Rev 85:97–177. doi:10.1152/physrev.00050.2003

Fangue NA, Hofmeister M, Schulte PM (2006) Intraspecific variation in thermal tolerance and heat shock protein gene expression in common killifish, Fundulus heteroclitus. J Exp Biol 209:2859–2872. doi:10.1242/jeb.02260

Fiol DF, Kültz D (2007) Osmotic stress sensing and signaling in fishes. Fed Eur Biochem Soc 274:5790–5798. doi:10.1111/j.1742-4658.2007.06099.x

Franklin CE, Crockford T, Johnston IA, Kamunde C (1994) The thermostability of haemoglobins from the hot-spring fish, Oreochromis alcalicus grahami: Comparisons with antarctic and temperate species. J Therm Biol 19:277–280. doi:10.1016/0306-4565(94)90051-5

Gilmour KM (2012) New insights into the many functions of carbonic anhydrase in fish gills. Respir Physiol Neurobiol 184:223–230

Gilmour KM, Perry SF (2009) Carbonic anhydrase and acid-base regulation in fish. J Exp Biol 212:1647–1661. doi:10.1242/jeb.029181

Gonzalez RJ (2012) The physiology of hyper-salinity tolerance in teleost fish: a review. J Comp Physiol B 182:321–329. doi:10.1007/s00360-011-0624-9

Grabherr MG, Haas BJ, Yassour M et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652. doi:10.1038/nbt.1883

Haas BJ, Papanicolaou A, Yassour M et al (2013) De novo transcript sequence reconstruction from RNA-SEQ: reference generation and analysis with Trinity platform for reference generation and analysis. Nat Protoc 8:1494–1512. doi:10.1038/nprot.2013.084.De

Henning F, Jones JC, Franchini P, Meyer A (2013) Transcriptomics of morphological color change in polychromatic Midas cichlids. BMC Genom 14:171. doi:10.1186/1471-2164-14-171

Hunt BG, Ometto L, Keller L, Goodisman MA (2013) Evolution at two levels in fire ants: the relationship between patterns of gene expression and protein sequence evolution. Mol Biol Evol 30:263–271. doi:10.1093/molbev/mss234

Jain P, Krishnan NM, Panda B (2013) Augmenting transcriptome assembly by combining de novo and genome-guided tools. PeerJ 1:e133. doi:10.7717/peerj.133

Johannsson OE, Bergman HL, Wood CM et al (2014) Air breathing in Magadi tilapia Alcolapia grahami, under normoxic and hyperoxic conditions, and the association with sunlight and reactive oxygen species. J Fish Biol 84:844–863. doi:10.1111/jfb.12289

Johansen K, Maloiy GM, Lykkeboe G (1975) A fish in extreme alkalinity. Respir Physiol 24:159–162. doi:10.1016/0034-5687(75)90110-3

Kalujnaia S, McWilliam IS, Zaguinaiko VA et al (2007) Salinity adaptation and gene profiling analysis in the European eel (Anguilla anguilla) using microarray technology. Gen Comp Endocrinol 152:274–280. doi:10.1016/j.ygcen.2006.12.025

Kavembe GD, Machado-Schiaffino G, Meyer A (2014) Pronounced genetic differentiation of small, isolated and fragmented tilapia populations inhabiting the Magadi Soda Lake in Kenya. Hydrobiologia 739:55–71. doi:10.1007/s10750-013-1648-9

Kelley JL, Passow CN, Plath M et al (2012) Genomic resources for a model in adaptation and speciation research: characterization of the Poecilia mexicana transcriptome. BMC Genom 13:652. doi:10.1186/1471-2164-13-652

Khaitovich P, Hellmann I, Enard W et al (2005) Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Sci 309:1850–1854. doi:10.1126/science.1108296

Kim SH, Yi SV (2007) Understanding relationship between sequence and functional evolution in yeast proteins. Genetica 131:151–156. doi:10.1007/s10709-006-9125-2

Kornfield I, Smith PF (2000) Africa cichlid fishes: model systems for evolutionary biology. Annu Rev Ecol Syst 31:163–196

Kozak GM, Brennan RS, Berdan EL et al (2014) Functional and population genomic divergence within and between two species of killifish adapted to different osmotic niches. Evolution 68:63–80. doi:10.1111/evo.12265

Kültz D, Burg M (1998) Review evolution of osmotic stress signaling via map kinase cascades. J Exp Biol 201: 3015–3021

Kuraku S, Meyer A (2008) Genomic analysis of cichlid fish "natural mutants". Curr Opin Genet Dev 18:551–558. doi:10.1016/j.gde.2008.11.002

Lam SH, Lui EY, Li Z et al (2014) Differential transcriptomic analyses revealed genes and signaling pathways involved in iono-osmoregulation and cellular remodeling in the gills of euryhaline Mozambique tilapia, Oreochromis mossambicus. BMC Genomics 15:921

Lande R (2009) Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. J Evol Biol 22:1435–1446

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. doi:10.1186/gb-2009-10-3-r25

Latta LC, Weider LJ, Colbourne JK, Pfrender ME (2012) The evolution of salinity tolerance in Daphnia: a functional genomics approach. Ecol Lett 15:794–802. doi:10.1111/j.1461-0248.2012.01799.x

Laverty G, Skadhauge E (2012) Adaptation of teleosts to very high salinity. Comp Biochem Physiol Part A 163:1–6. doi:10.1016/j.cbpa.2012.05.203

Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL (2005) Evolution of proteins and gene expression levels are coupled in Drosophila and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions. Mol Biol Evol 22:1345–1354. doi:10.1093/molbev/msi122

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform 12:323. doi:10.1186/1471-2105-12-323

Lu B, Zeng Z, Shi T (2013) Comparative study of de novo assembly and genome-guided assembly strategies for transcriptome reconstruction based on RNA-Seq. Sci China Life Sci 56:143–155. doi:10.1007/s11427-013-4442-z

Lykkeboe G, Johansen K, Maloiy G (1975) Functional properties of hemoglobins in the teleost Tilapia grahami. J Comp Physiol 104:1–11

Maina J, Wood CM, Narahara A et al (1996a) Morphology of the swim (air) bladder of a cichlid teleost: Oreochromis alcalicus grahami (Trewavas, 1983), a fish adapted to a hyperosmotic, alkaline, and hypoxic environment: a brief outline of the structure and function of the swimbladder. In: Datta Munshi JS, Dutta HM (eds) Fish morphology: horizons of new research. Science Publishers, Lebanon, pp 179–192

Maina JN, Kisia SM, Wood CM et al (1996b) A comparative allometric study of the morphometry of the gills of an alkalinity adapted cichlid fish, Oreochromis alcalicus grahami, of Lake Magadi, Kenya. Int J Salt Lake Res 5:131–156

Majdalawieh A, Zhang L, Ro H (2007) Adipocyte enhancer-binding protein-1 promotes macrophage inflammatory responsiveness by up-regulating NF-κB via IκBα negative regulation. Mol Biol Cell 18:930–942. doi:10.1091/mbc.E06

Martin JA, Wang Z (2011) Next-generation transcriptome assembly. Nat Rev Genet 12:671–682. doi:10.1038/nrg3068

Merilä J (2012) Evolution in response to climate change: in pursuit of the missing evidence. BioEssays 34:811–818. doi:10.1002/bies.201200054

Merilä J, Hendry AP (2014) Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. Evol Appl 7:1–14

Narahara A, Bergman HL, Laurent P et al (1996) Respiratory physiology of the Lake Magadi Tilapia (Oreochromis alcalicus grahami), a fish adapted to a hot, alkaline, and frequently hypoxic environment. Physiol Zool 69:1114–1136

Narum SR, Campbell NR, Meyer KA et al (2013) Thermal adaptation and acclimation of ectotherms from differing aquatic climates. Mol Ecol. doi:10.1111/mec.12240

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Oleksiak MF (2011) Genomic approaches with natural fish populations. J Fish Biol 76:1067–1093. doi:10.1111/j.1095-8649.2010.02563.x.Genomic

Pörtner HO, Schulte PM, Wood CM, Schiemer F (2010) Niche dimensions in fishes: an integrative view. Physiol Biochem Zool 83:808–826. doi:10.1086/655977

Pujolar JM, Marino IAM, Milan M et al (2012) Surviving in a toxic world: transcriptomics and gene expression profiling in response to environmental pollution in the critically endangered European eel. BMC Genom 13:507. doi:10.1186/1471-2164-13-507

Purohit GK, Mahanty A, Suar M et al (2014) Investigating hsp gene expression in liver of Channa striatus under heat stress for understanding the upper thermal acclimation. Biomed Res Int 2014:1–10. doi:10.1155/2014/381719

Randall DJ, Wood CM, Perry SF et al (1989) Urea excretion as a strategy for survival in a fish living in a very alkaline environment. Nature 337:165–166

Rijssel JC, Witte F (2012) Adaptive responses in resurgent Lake Victoria cichlids over the past 30 years. Evol Ecol 27:253–267. doi:10.1007/s10682-012-9596-9

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140. doi:10.1093/bioinformatics/btp616

Romero IG, Ruvinsky I, Gilad Y (2012) Comparative studies of gene expression and the evolution of gene regulation. Nat Rev Genet 13:505–516. doi:10.1038/nrg3229

Rothschild LJ, Mancinelli RL (2001) Life in extreme environments. Nature 409:1092–1101. doi:10.1038/35059215

Rutjes HA, Nieveen MC, Weber RE et al (2007) Multiple strategies of Lake Victoria cichlids to cope with lifelong hypoxia include hemoglobin switching. Am J Physiol 293:1376–1383. doi:10.1152/ajpregu.00536.2006

Scott GR, Richards JG, Forbush B et al (2004) Changes in gene expression in gills of the euryhaline killifish Fundulus heteroclitus after abrupt salinity transfer. Am J Physiol 287:300–309. doi:10.1152/ajpcell.00054.2004

Seegers L, Sonnenberg R, Yamamoto R (1999) Molecular analysis of the Alcolapia flock from lakes Natron and Magadi, Tanzania and Kenya (Teleostei: Cichlidae), and implications for their systematics and evolution. Ichthyol Explor Freshwaters 10:175–199

Seehausen O (2002) Patterns in fish radiation are compatible with Pleistocene desiccation of Lake Victoria and 14,600 year history for its cichlid species flock. Proc R Soc London B 269:491–497. doi:10.1098/rspb.2001.1906

Seehausen O (2006) African cichlid fish: a model system in adaptive radiation research. Proc Biol Sci B 273:1987–1998. doi:10.1098/rspb.2006.3539

Shin SC, Ahn DH, Kim SJ et al (2014) The genome sequence of the Antarctic bullhead notothen reveals evolutionary adaptations to a cold environment. Genome Biol 15:468. doi:10.1186/s13059-014-0468-1

Sims D, Sudbery I, Ilott NE et al (2014) Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet Genet 15:121–132. doi:10.1038/nrg3642

Soumillon M, Necsulea A, Weier M et al (2013) Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. Cell Rep 3:2179–2190. doi:10.1016/j.celrep.2013.05.031

Stearns SC (1994) The evolutionary links between fixed and variable traits. Acta Palaeontol Polon 38:215–232

Strickler SR, Bombarely A, Mueller LA (2012) Designing a transcriptome next-generation sequencing project for a non-model plant species. Am J Bot 99:257–266. doi:10.3732/ajb.1100292

Sturmbauer C, Baric S, Salzburger W et al (2001) Lake level fluctuations synchronize genetic divergences of cichlid fishes in African lakes. Mol Biol Evol 18:144–154

Swanson WJ, Wong A, Wolfner MF, Aquadro CF (2004) Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. Genetics 168:1457–1465. doi:10.1534/genetics.104.030478

Talling JF, Talling IB (1965) The chemical composition of African Lake waters. Int Rev der gesamten Hydrobiol und Hydrogr 50:421–463

Tine M, de Lorgeril J, D'Cotta H et al (2008) Transcriptional responses of the black-chinned tilapia *Sarotherodon melanotheron* to salinity extremes. Mar Genom 1:37–46. doi:10.1016/j.margen.2008.06.001

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111. doi:10.1093/bioinformatics/btp120

Trapnell C, Roberts A, Goff L et al (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562–578. doi:10.1038/nprot.2012.016

Verheyen E, Rüber L, Snoeks J, Meyer A (1996) Mitochondrial phylogeography of rock-dwelling cichlid fishes reveals evolutionary influence of historical lake level fluctuations of Lake Tanganyika, Africa. Philos Trans R Soc Lond B Biol Sci 351:797–805

Verne S, Jaquish B, White R et al (2011) Global transcriptome analysis of constitutive resistance to the white pine weevil in spruce. Genome Biol Evol 3:851–867. doi:10.1093/gbe/evr069

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63

Warnefors M, Kaessmann H (2013) Evolution of the correlation between expression divergence and protein divergence in mammals. Genome Biol Evol 5(7):1324–1335. doi:10.1093/gbe/evt093

Wilson PJ, Wood CM, Maina JN, White BN (2000) Genetic structure of Lake Magadi tilapia populations. J Fish Biol 56:590–603. doi:10.1006/jfbi.1999.1184

Wilson PJ, Wood CM, Walsh PJ et al (2004) Discordance between genetic structure and morphological, ecological, and physiological adaptation in Lake Magadi tilapia. Physiol Biochem Zool 77:537–555. doi:10.1086/422054

Woo S, Denis V, Won H et al (2013) Expressions of oxidative stress-related genes and antioxidant enzyme activities in *Mytilus galloprovincialis* (Bivalvia, Mollusca) exposed to hypoxia. Zool Stud 52:15. doi:10.1186/1810-522X-52-15

Wood CM, Perry SF, Wright PA et al (1989) Ammonia and urea dynamics in the Lake Magadi tilapia, a ureotelic teleost fish adapted to an extremely alkaline environment. Respir Physiol 77:1–20

Wood C, Bergman H, Laurent P et al (1994) Urea production, acid-base regulation and their interactions in the Lake Magadi Tilapia, a unique teleost adapted to a highly alkaline environment. J Exp Biol 189:13–36

Wood CM, Wilson P, Bergman HL et al (2002a) Obligatory urea production and the cost of living in the Magadi tilapia revealed by acclimation to reduced salinity and alkalinity. Physiol Biochem Zool 75:111–122. doi:10.1086/340626

Wood CM, Wilson P, Bergman HL et al (2002b) Ionoregulatory strategies and the role of urea in the Magadi tilapia (*Alcolapia grahami*). Can J Zool 80:503–515. doi:10.1139/Z02-019

Wood CM, Bergman HL, Bianchini A et al (2012) Transepithelial potential in the Magadi tilapia, a fish living in extreme alkalinity. J Comp Physiol B 182:247–258

Wright PA, Fyhn JH (2001) Ontogeny of nitrogen metabolism and excretion. In: Anderson PA, Wright PA (eds) Fish physiology, 20th edn. Academic Press, Orlando, pp 149–200

Wright PA, Wood CM (2009) A new paradigm for ammonia excretion in aquatic animals: role of Rhesus (Rh) glycoproteins. J Exp Biol 212:2303–2312

Wright PA, Perry SF, Randall DJ et al (1990) The effects of reducing water pH and total CO_2 on a teleost fish adapted to an extremely alkaline environment. J Exp Biol 151:361–369

Xu J, Ji P, Wang B et al (2013a) Transcriptome sequencing and analysis of wild Amur Ide (*Leuciscus waleckii*) inhabiting an extreme alkaline-saline lake reveals insights into stress adaptation. PLoS One 8:e59703. doi:10.1371/journal.pone.0059703

Xu J, Li Q, Xu L et al (2013b) Gene expression changes leading extreme alkaline tolerance in Amur ide (*Leuciscus waleckii*) inhabiting soda lake. BMC Genom 14:682. doi:10.1186/1471-2164-14-682

Yampolsky LY, Schaer TMM, Ebert D (2014) Adaptive phenotypic plasticity and local adaptation for temperature tolerance in freshwater zooplankton. Proc R Soc B 281:20132744. doi:10.1098/rspb.2013.2744

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591. doi:10.1093/molbev/msm088

Yang Z, Bielawski J (2000) Statistical methods for detecting molecular adaptation. Trends Ecol Evol 15:496–503

Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17:32–43

Yoshida T, Goka K, Ishihama F et al (2007) Biological invasion as a natural experiment of the evolutionary processes: introduction of the special feature. Ecol Res 22:849–854. doi:10.1007/s11284-007-0435-3

Zhu M, Zhao S (2007) Candidate gene identification approach: progress and challenges. Intern J Biol Sci 3:420–427