The Company of Biologists

REVIEW

Considerations for the use of transcriptomics in identifying the 'genes that matter' for environmental adaptation

Tyler G. Evans*

ABSTRACT

Transcriptomics has emerged as a powerful approach for exploring physiological responses to the environment. However, like any other experimental approach, transcriptomics has its limitations. Transcriptomics has been criticized as an inappropriate method to identify genes with large impacts on adaptive responses to the environment because: (1) genes with large impacts on fitness are rare; (2) a large change in gene expression does not necessarily equate to a large effect on fitness; and (3) protein activity is most relevant to fitness, and mRNA abundance is an unreliable indicator of protein activity. In this review, these criticisms are re-evaluated in the context of recent systems-level experiments that provide new insight into the relationship between gene expression and fitness during environmental stress. In general, these criticisms remain valid today, and indicate that exclusively using transcriptomics to screen for genes that underlie environmental adaptation will overlook constitutively expressed regulatory genes that play major roles in setting tolerance limits. Standard practices in transcriptomic data analysis pipelines may also be limiting insight by prioritizing highly differentially expressed and conserved genes over those genes that undergo moderate fold-changes and cannot be annotated. While these data certainly do not undermine the continued and widespread use of transcriptomics within environmental physiology, they do highlight the types of research questions for which transcriptomics is best suited and the need for more gene functional analyses. Such information is pertinent at a time when transcriptomics has become increasingly tractable and many researchers may be contemplating integrating transcriptomics into their research programs.

KEY WORDS: Environment, Microarray, Physiology, RNA-seq, Stress, Temperature, Transcriptomics

Introduction

The environment exerts a profound effect on biological systems, and strong spatial and temporal heterogeneity in abiotic variables across the biosphere has shaped the physiologies of life on Earth (Hochachka and Somero, 2002). Consequently, some species are eurytolerant, and able to persist under a broad range of abiotic conditions, while others are stenotolerant, and confined to habitats that experience minimal abiotic change (Somero, 2012). A fundamental goal of environmental physiology is to understand why differences exist among species in their capacity to tolerate environmental change (Hochachka and Somero, 2002; Somero, 2011). Achieving this goal is dependent upon answering a basic question: how do organisms modify environmental tolerance limits? Recently, advances in biotechnology have allowed experiments

Department of Biological Sciences, California State University, East Bay, Hayward, CA 94542, USA.

*Author for correspondence (tyler.evans@csueastbay.edu)

aimed at addressing this question to become increasingly reductionist in nature, with many contemporary studies relating changes in the environment to genome-scale phenomena (Gracey, 2007; Wang et al., 2009a; Merbl and Kirschner, 2011; Pespeni et al., 2013; Welch et al., 2014).

Transcriptomics has emerged as a particularly popular approach for exploring how organisms respond to environmental change (Gracey and Cossins, 2003; Cossins et al., 2006; Ozsolak and Milos, 2011; Gracey, 2007; Wang et al., 2009a; Evans and Hofmann, 2012). Whether in the form of microarrays or more recently RNA sequencing, considerable effort has been directed toward characterizing shifts in mRNA abundance triggered by changes in key environmental variables such as temperature (Logan and Buckley, 2015), salinity (Evans and Somero, 2008), oxygen (Gracey et al., 2011) and pH (Benner et al., 2013; Evans et al., 2013). Within the field of environmental physiology, transcriptomics has been used successfully to address a broad range of questions concerning how or whether organisms can acclimate or adapt to the abiotic conditions associated with life in specific habitats (Evans and Hofmann, 2012). These investigations have demonstrated the complexity of responses to the environment (Chapman et al., 2011; Evans et al., 2011), isolated cellular and physiological processes that are robust or sensitive to environmental change (Logan and Somero, 2010), provided clues as to how organisms cope with life in challenging habitats (Podrabsky and Somero, 2004; Bilyk and Cheng, 2013), helped to predict vulnerabilities or resistance toward climate change (Barshis et al., 2013; Palumbi et al., 2014) and highlighted potentially important genes for future study (Meyer and Manahan, 2010; Whitehead et al., 2013), as well as leading to many other valuable scientific discoveries. However, like any other experimental approach, transcriptomics is associated with a set of limitations, and some have questioned the adequacy of transcriptomics to address particular questions of interest to environmental physiologists (Feder and Walser, 2005; Suarez and Moyes, 2012).

An initial promise was that unbiased screens of the transcriptome would serve to isolate novel or unforeseen genes with large consequences on fitness under different environmental conditions; the so-called 'genes that matter' (Feder and Walser, 2005; Feder and Mitchell-Olds, 2003). In 2005, Feder and Walser concluded that this specific promise had gone largely unfulfilled, offering a pointed description of the major issues facing the use of transcriptomics in finding the genes that matter for environmental adaptation (Feder and Walser, 2005). Their critique focused on three major issues: (1) genes with large impacts on fitness are rare and therefore unlikely to be identified with transcriptomics, (2) the relationship between gene expression and fitness is unreliable and (3) fitness is primarily determined by proteins, and mRNA abundance is a poor proxy for protein abundance. The goal of this review is to re-evaluate the validity of these statements in the context of new experimental evidence, and also to identify new limitations that may have

emerged as a consequence of the increasing popularity of transcriptomics in studies of the environment (Fig. 1). In general, data collected over the last decade indicate that these criticisms remain valid today and that new issues may have also arisen. While this information does not discredit the widespread and continued use of transcriptomics within environmental physiology, it does illustrate the types of research questions that transcriptomic experiments are best suited to address. A wider understanding of potential limitations of transcriptomics is especially relevant given that technological advances have made transcriptomics increasingly accessible, and more and more researchers are likely contemplating integrating this technology into their research programs (Wang et al., 2009a; De Wit et al., 2012; Qian et al., 2014). Information summarized here may also stimulate research aimed at addressing uncertainties about the use of transcriptomics. A major reason for the staying power of transcriptomics is that the scientific community has often responded to criticisms with targeted research that improved how transcriptomic data are acquired and analyzed (e.g. Storey and Tibshirani, 2003; Huang et al., 2008).

Genes with large impacts on fitness are rare

Within the field of environmental physiology, transcriptomics has been touted as a discovery or hypothesis-independent approach (Gracey and Cossins, 2003; Feder and Walser, 2005) – a means to screen thousands of genes simultaneously in order to isolate those that play a major role in determining environmental tolerance (Feder and Mitchell-Olds, 2003). However, the efficacy of transcriptomics to isolate genes consequential to fitness during environmental stress has been disputed (Feder and Walser, 2005). There is no doubt that transcriptomics can be used to identify genes with major impacts on ecologically relevant traits (Gracey, 2007). One prominent example was the identification of calmodulin as a regulator of beak

morphology in Darwin's finches. In this case, microarray analysis revealed that calmodulin was expressed at higher levels in birds with long pointed beaks than in those with thicker, more robust beaks. Subsequent functional analysis demonstrated that overexpression of calmodulin in the frontonasal prominence of chicken embryos caused an elongation of the upper beak that recapitulated the pointed beak morphology of the finch (Abzhanov et al., 2006). Transcriptional profiling of obese mice also identified hepatic stearoyl-CoA desaturase-1 as a key regulator of fat deposition, and this hypothesized function was confirmed in subsequent loss of function experiments (Cohen et al., 2002; Xu et al., 2007). However, Feder and Walser (2005) suggest that the probability of such discoveries is very low, and that the question most relevant to environmental physiology is not whether transcriptomics can facilitate these discoveries, but rather the frequency by which they occur. Addressing this contention directly, even just within the field of environmental physiology, is difficult because it requires both a comprehensive list of transcriptomic studies and an understanding of how often candidate genes from each study were shown to impact fitness in ensuing functional experiments. Alternatively, the probability of transcriptomics isolating genes that impact fitness can be estimated from genome-wide mutagenesis screens that systematically inhibit the expression of a single gene, screen for a phenotype, and then repeat this process for each protein-coding gene across the entire genome (Carpenter and Sabatini, 2004). While deleted genes not displaying an overt phenotype may still influence fitness, these screens nonetheless provide an estimate of the proportion of genes that contribute to overt phenotypic change and are therefore more likely to influence fitness. Feder and Walser (2005) cite a large number of these mutagenesis screens as evidence that the fitness costs associated with loss of function in any one gene are most often negligible, and therefore the likelihood that transcriptomics would isolate any of these rare genes is low (Feder and Walser, 2005). For example, RNA interference was used to systematically eliminate the functions of more than 16,000 genes in the nematode worm *Caenorhabditis elegans*. Remarkably, 89% of single-copy genes show no detectable phenotypic effect (Conant and Wagner, 2004). In the plant model *Arabidopsis thaliana*, 96% of the 25,500 genes screened were considered dispensable (May and Martienssen, 2003). In cultured embryonic cells of the fly *Drosophila melanogaster*, a functional screen of 19,470 genes inhibited by RNA interference (91% of protein-coding genes present in the *Drosophila* genome) reported only 483 genes as essential for growth and viability (Boutros et al., 2004). As summarized by Feder and Walser (2005): 'Thus, most genes are remarkable for not being essential' (Feder and Walser, 2005). Studies published since support the hypothesis that most protein-coding genes are not essential when deleted individually, at least under the conditions of the experiment. An analysis of 4836 deletion mutants in the yeasts *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* revealed that 83% of single-copy orthologs in the two yeasts had conserved dispensability (Kim et al., 2010). Phenotypic screening of a large number of nonsense and essential splice mutations in zebrafish *Danio rerio* revealed that only 6% (74 out of 1216) caused a discernible phenotype during the first 5 days of embryonic development (Kettleborough et al., 2013). In the mouse *Mus musculus*, an analysis of nearly 3900 individually inactivated genes found that approximately half were essential as both singletons and duplicates (Liao and Zhang, 2007).

Several hypotheses have been formulated to explain the relatively low number of essential genes encoded in eukaryotic genomes. One of these explanations, the contingent function hypothesis, states that
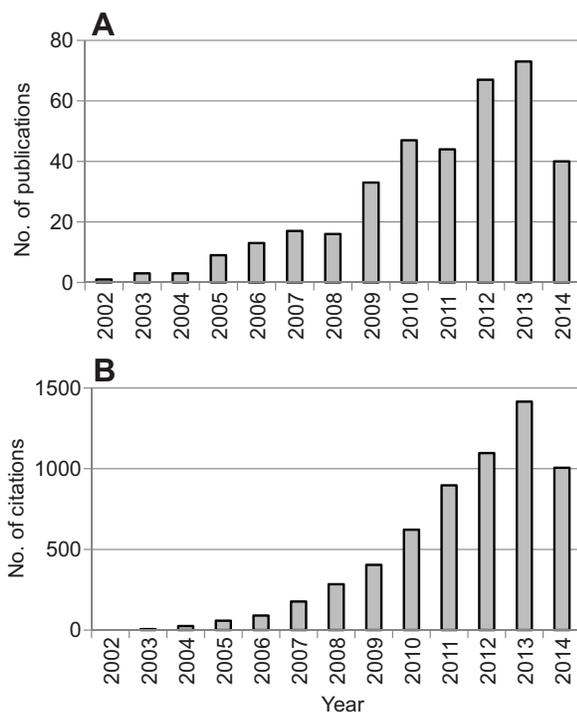

Fig. 1. Increasing popularity of transcriptomics in environmental science. The number of publications (A) and number of citations (B) retrieved from the Web of Science database using search terms 'transcriptomics' and 'environment'. Data apply to a search performed on 17 September 2014.

genes have conditional phenotypes that manifest only under certain cellular states (Thatcher et al., 1998). This hypothesis has obvious relevance to transcriptomics and environmental physiology because it attests that genes deemed dispensable during non-stress conditions may nonetheless have major fitness consequences when environmental conditions deviate from optima. Importantly, the ability of transcriptomics to isolate genes that strongly influence environmental tolerance would increase if a greater number of genes were shown to impact fitness during environmental stress than during non-stress (i.e. control) conditions. But do more genes have contingent functions during adaptive responses to the environment than under non-stress conditions? Experiments aimed at addressing this question are relatively few, and are presently confined to simple laboratory model organisms. However, results suggest that in these species, the probability of any one gene significantly influencing fitness, even under shifting environmental conditions, is low. In the yeast *S. cerevisiae*, exposure of 4783 single gene deletion strains to acute heat stress (increasing temperature from 30 to 50°C) yielded only 55 deletions (1.2% of genes assayed) that differed significantly in heat sensitivity relative to control populations (Gibney et al., 2013). Similar trends also seem to underlie acquired thermotolerance in yeast, where exposure to an initial mild stress confers resistance toward a subsequent more severe stress. Pre-treatment of the same collection of yeast deletion strains to 37°C prior to increasing temperature to 50°C, identified only 10 genes (0.2% of genes assayed) that were able to modify acquired thermotolerance. These 10 strains were largely a subset of the 55 genes shown to influence tolerance toward acute heat stress, suggesting that only a small number of genes underlie both innate and acquired thermotolerance in yeast (Gibney et al., 2013). Additional genome-wide analyses in *S. cerevisiae* support this conclusion. A separate screen of 4786 viable haploid deletions identified a similarly small number of genes (*N*=38) capable of altering sensitivity to heat stress (Mir et al., 2009). Studies in yeast also suggest relatively small numbers of genes are capable of modifying sensitivity toward other abiotic factors. A screen of 4828 yeast deletion strains identified only 95 genes (2% of genes assayed) that alter sensitivity toward ethanol, 42 genes (0.9% of genes assayed) capable of modifying osmotic stress tolerance, and 30 genes (0.6% of genes assayed) that change tolerance toward oxidative stress (Auesukaree et al., 2009). At least in yeast, only a small subset of protein-coding genes appear capable of modifying tolerance toward abiotic factors.

Yeast are comparatively simple, single-celled eukaryotes, and the trends described above may not be conserved in more complex eukaryotic species. *Caenorhabditis elegans* is the only metazoan where comparable high-throughput screens for genes involved in thermotolerance have been performed. In a genome-wide RNA interference screen for regulators of the heat shock response in *C. elegans*, seven genes were identified as required for the induction of a heat shock response, along with 52 genes that act as negative regulators and whose knockdown leads to constitutive activation of the heat shock response. The heat shock response is a defining characteristic of the molecular reaction to heat stress, is highly conserved across taxa and is an important biochemical indicator of thermal tolerance limits (Tomanek, 2010). These data suggest only 59 genes comprise the heat shock regulatory network in *C. elegans* (Guisbert et al., 2013). Similar genome-wide screens have been employed to identify genetic modifiers of proteostasis; that is, genes capable of increasing the cellular capacity for protein folding, preventing proteins from aggregating, or suppressing protein toxicity. Unfolded (i.e. denatured) proteins are a hallmark of heat stress and genes capable of increasing or decreasing the cellular capacity for protein folding are likely to influence fitness during exposure to elevated temperatures. In *C. elegans*, a genome-wide RNA interference screen identified 88 genes that modified protein aggregation (Silva et al., 2011), attesting that a very small proportion of the approximately 20,000 protein-coding genes in the *C. elegans* genome (Hillier et al., 2005) are actually capable of modifying this key aspect of the response to thermal stress.

Systematic, genome-wide gene functional analyses in yeast and *C. elegans* indicate that a small number of individual genes are considered essential under non-stress conditions, and that few individual genes appear capable of modifying thermotolerance or aspects of the heat shock response in these two model organisms. The contingent function hypothesis is impossible to reject because it is not feasible to experimentally test all functional contexts. However, under the specific conditions used in these studies, a relatively small number of genes were shown to have temperature-dependent fitness effects. Consequently, the probability of transcriptomics isolating these genes with major impacts on fitness would not increase because of contingent functions under the conditions tested. Most studies investigating transcriptomic responses to acute environmental stress report significant differential expression in hundreds to thousands of transcripts, far greater than the number of essential or conditionally important genes suggested for yeast or *C. elegans*. Data from these single-gene deletion studies suggest that lists of significantly differentially expressed genes from a typical transcriptomic experiment may be biased toward genes that: (1) are truly dispensable, (2) are redundant in function ('marginal benefit' hypothesis; Thatcher et al., 1998) and/or (3) make small contributions to fitness (Feder and Walser, 2005). The concepts of gene functional redundancy and small individual gene contributions to fitness have important implications for environmental adaptation because they imply that environmental tolerance will depend on how many genes are modified, in addition to which ones. For example, loss of function of a single chaperone protein may not influence thermotolerance if a paralog or functionally related protein is able to compensate for its absence. There is evidence to suggest that multi-copy genes show loss of function phenotypes less often than single-copy genes (Chen et al., 2012). Both heat shock protein 70 (Hsp70) and heat shock factor (HSF) null *Drosophila* mutants maintain some degree of thermotolerance, suggesting functional compensation from other chaperone proteins (Gong and Golic, 2006). Thermotolerance in *Drosophila* also increases linearly with Hsp70 copy number (Bettencourt et al., 2008). While these studies reveal important trends in environmental adaptation, and emphasize the need for genome-wide functional screens in a wider range of organisms, they are of course not without caveats. Outcomes of functional screens are influenced by the metric of fitness used and by the severity of environmental change administered. Studies that only screen for genes that modify the upper thermal maxima disregard genes that influence fitness through sub-lethal changes in growth, reproduction or any other ecologically relevant trait. Similarly, genes necessary for coping with mild abiotic stress may not be the same as those required for severe stress (Logan and Somero, 2010).

Differential expression may not equal importance

The effectiveness of transcriptomics to isolate genes that underlie environmental adaptation is dependent upon genes differentially expressed during environmental stress also being genes with large contributions to fitness under those conditions. While in many cases this assumption may be valid, at the very least transcriptomics will

overlook the potentially important functions of genes that are constitutively expressed and do not change in abundance in response to environmental perturbation. Feder and Walser (2005) cite changes in the activity of constitutively expressed proteins as a major factor contributing to the lack of correlation between mRNA abundance and fitness (Giaever et al., 2002; Warringer et al., 2003; Feder and Walser, 2005). Assessing the relative contributions of constitutive and differentially expressed genes to fitness during environmental stress requires knowledge of all the genes differentially expressed in response to a specific change in the environment, and knowledge of how both differentially expressed and constitutively expressed genes impact fitness under the same environmental conditions. At present, this information is only available for *S. cerevisiae*; however, experiments in this simple eukaryote are providing novel insights into the relationship between environmentally regulated gene expression and fitness. Surprisingly, genes important for protection against acute environmental stress in yeast have very little overlap with genes strongly up- or down-regulated in response to that stress (Fig. 2). Alternatively, genes that are constitutively expressed, and whose expression is relatively unchanged during environmental stress, are more likely to influence fitness. A survey of 2954 yeast single-gene deletions for both a significant effect on survival and a significant change in gene expression during heat stress showed that large changes in expression do not correlate with a strong survival response for either induced (Fig. 2A) or repressed genes (Fig. 2B). This relationship between gene expression and fitness during heat stress in yeast even applies to heat shock proteins, whose rapid and abundant up-regulation is considered a hallmark of heat stress. Of the 29 heat shock proteins induced or repressed in response to heat stress in yeast, only two, jjj2 (a member of the DnaJ family of chaperones) and hsp104, significantly affected survival (Fig. 2C; Gibney et al., 2013). Lack of concordance between gene expression and functional significance is actually a well-established outcome in studies of environmental tolerance in yeast. Scans of the promoter sequences of genes that confer sensitivity and/or resistance toward

heat stress for specific transcription factor binding elements revealed only 8% of the genes shown to regulate thermotolerance are able to bind HSF1, indicating that heat shock proteins and other genes regulated by HSF1 are again not the most critical for modifying thermotolerance. Similarly, the promoter regions of only 6% of the genes that modify heat sensitivity in yeast contain binding sites for the general stress transcription factors Msn2p and Msn4p (Jarolim et al., 2013). Similar principles apply to the gene networks that underlie acquired stress resistance in yeast. Berry et al. (2011) exposed the entire non-essential yeast deletion collection to a mild stressor (elevated temperature, increased NaCl or increased dithiothreitol) and then tested each deletion strain for subsequent resistance toward severe oxidative stress (H_2O_2 exposure). The majority of genes whose expression increased following each mild stress pretreatment played no role in subsequent H_2O_2 tolerance (Berry et al., 2011). In their review of the role of gene expression in response to environmental stress in yeast, de Nadal et al. (2011) state that: 'There is generally a low overlap between those genes that are transcriptionally induced in response to stress and those genes that seem to be essential for adaptation' (de Nadal et al., 2011). Lopez-Maury et al. (2008) offer a similar sentiment, concluding that many of the genes regulated during stress in yeast do not seem to have any direct functional relevance to the specific perturbation (López-Maury et al., 2008). Finally, Giaever and Nislow (2014) state that there is little correlation between the genes required for fitness in a condition and those genes whose transcription is up-regulated in that condition (Giaever and Nislow, 2014). Previously described functional redundancy and small individual contributions to fitness are probable reasons why genes responding transcriptionally to environmental change do not individually exert a large impact on fitness in yeast. While the relevance of these data to more complex eukaryotes remains a major caveat, using transcriptomics to identify genes that underlie environmental adaptation in yeast will surely disregard the majority of functionally important genes.
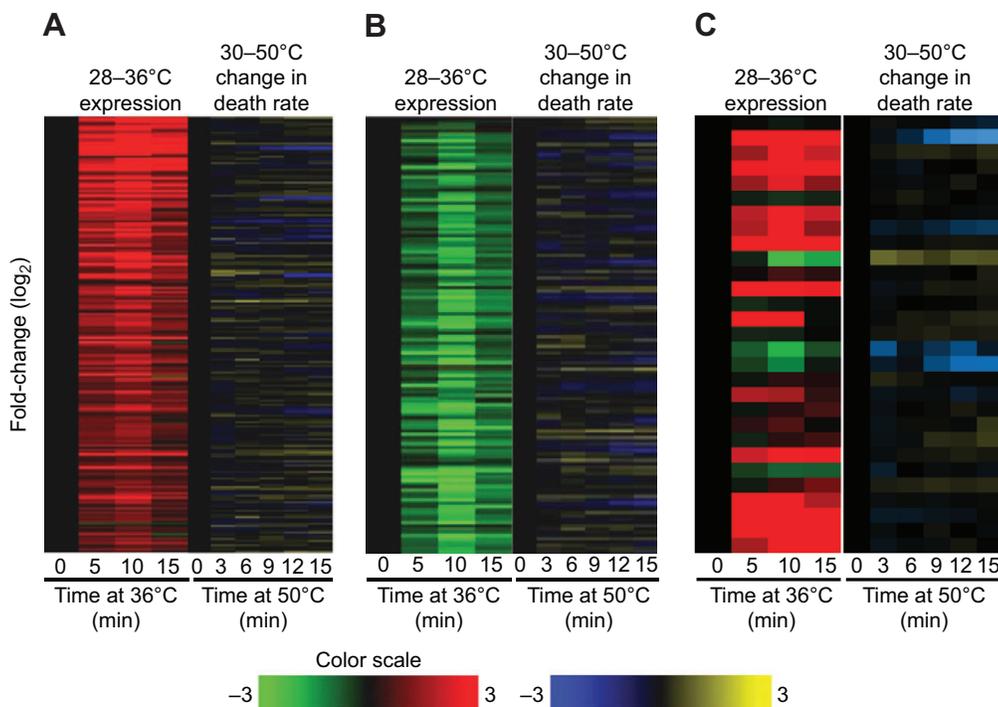

Fig. 2. Lack of correlation between heat-induced changes in gene expression and genes that modify heat sensitivity in yeast *Saccharomyces cerevisiae*. Hierarchical clustered heatmaps showing that genes strongly up-regulated (A) or down-regulated (B) in response to heat shock (temperature increase from 26 to 36°C) are not the same genes that modify sensitivity toward heat stress. This relationship also applies to heat shock proteins (*N*=29), with the exception of hsp104 and jjj2 (C). Each row represents the log₂ expression or effect on death rate of a single gene and each column represents expression of that particular gene at a given time (0–15 min). Bright red coloration indicates strong up-regulation of expression and bright green indicates strong down-regulation. Bright yellow indicates that deletion of that gene decreased death rate during heat stress, while bright blue indicates that deletion of that gene increased death rate during heat stress. Adapted from Gibney et al. (2013).

Individual genes that do influence sensitivity toward the environment in yeast act as so-called 'hub' genes. Hub genes interact with far more than the average number of protein partners, and therefore are much more likely to display phenotypes when knocked-out compared with genes whose proteins have no or few interacting partners (Jeong et al., 2001; Han et al., 2004; Feder and Walser, 2005). Gene functional annotation shows that the subset of genes capable of modifying thermotolerance in yeast, almost all of which do not respond transcriptionally to heat stress, contain a significantly higher proportion of genes involved in cell signaling and chromatin regulation (Gibney et al., 2013; Jarolim et al., 2013), and therefore may act as hub genes by influencing the expression and/or activity of a large number of downstream genes. For example, several genes identified as necessary for heat shock survival in *S. cerevisiae* are either kinases that covalently attach phosphate groups to proteins (e.g. serine/threonine/tyrosine kinase; beta subunit of casein kinase 2), or phosphatases that remove phosphate groups (e.g. trehalose 6-phosphate phosphatase; putative tyrosine phosphatase similar to Oca1), highlighting the importance of these post-translational modifications in determining thermotolerance (Gibney et al., 2013). Hub genes are less likely to be detected through transcriptomic screens because these genes tend to exhibit stable expression, instead influencing fitness through post-translational changes in protein activity (Batada et al., 2006; Batada and Hurst, 2007; Lu et al., 2007; Lehner, 2008; MacNeil and Walhout, 2011). In contrast, transcriptomics is best-suited to interrogate the downstream effectors of biological responses because these genes tend to have high expression variability (Gracey, 2007). Effector genes also tend to exhibit low connectivity to other genes within environmentally regulated gene networks (Batada and Hurst, 2007; Lu et al., 2007; Lehner, 2008; MacNeil and Walhout, 2011), which helps to explain their comparatively small effects on fitness as single-gene deletions. Gene networks underlying adaptive responses to arsenic in yeast follow this genetic architecture. Genes that conferred the most sensitivity to arsenic were hub genes upstream of the arsenic detoxification pathways, while expression profiling identified downstream genes that protect against toxicity, but which share redundant functions and therefore have no apparent phenotypic effect when deleted individually (Haugen et al., 2004; Gracey, 2007).

Systems-level experiments using yeast deletion strains emphasize that genes exerting a large influence on environmental tolerance are few in number, constitutively expressed, do not alter abundance significantly during environmental change and are hub genes involved in regulatory events. Transcriptomics is not the most appropriate experimental method to identify these types of genes. Alternatively, a detailed understanding of the interactome or the 'connectedness' of genes within environmentally regulated gene networks may more reliably predict genes with large effects on environmental tolerance. Analyses of the interactome suggest that environmentally regulated gene networks share common design principles (i.e. topologies) (Chalancon et al., 2012). Integration of gene expression data with gene network topology has shown that the set of interacting genes activated by stress are distinguished by an 'autocratic' topology, where a few master regulators control the expression of entire cascades of genes with little combinatorial regulation. Differences in the expression or activity of these few key regulatory hubs are hypothesized to contribute to phenotypic variability at the population level (Chalancon et al., 2012). Gene network analysis of the response to environmental stress in the Zhikong scallop *Chlamys farreri*, the killifish *Fundulus heteroclitus* and diverse lineages of bacteria support this general network topology. In the scallop, temperature increase transcriptionally activates six modules of co-expressed genes, with each module targeting specific cellular pathways important to coping with environmental change, including protein folding, apoptosis and metabolism. Each module contains between 107 and 1640 genes, and is regulated by very few, highly connected hub genes (Fu et al., 2014). Populations of killifish that differ in their capacity to acclimate to changes in environmental salinity possess distinct gene networks. In populations able to acclimate to a wider range of environmental salinity, natural selection has targeted upstream regulators of transcription to more precisely control the expression of the downstream genes that enable phenotypic plasticity in response to salinity change. Consequently, gene networks in killifish populations that exhibit the broadest range of salinity tolerance have fewer transcriptional regulators and a more streamlined topology (Shaw et al., 2014). A meta-analysis of gene networks governing responses to the environment in bacteria that differ in their ability to tolerate a suite of abiotic factors also shows that most of the variation within environmentally regulated gene networks (in both gene sequence and gene content) occurs within upstream sensing and signal-transduction elements rather than downstream effector genes (Singh et al., 2008). Dysfunctional regulatory elements may also be associated with a limited capacity to respond to environmental change. Antarctic notothenioid fish are extremely stenothermal, a trait that has been mechanistically linked to the inability to up-regulate heat shock proteins during temperature increase (Hofmann et al., 2000). Interestingly, several species of notothenioid actually produce large amounts of heat shock protein mRNA constitutively (Buckley et al., 2004; Place et al., 2004; Place and Hofmann, 2005; Buckley and Somero, 2009). In fact, the gill and liver transcriptomes of field-acclimated cryopelagic *Pagothenia borchgrevinki* are enriched for molecular chaperone transcripts relative to the eurythermal zebrafish *D. rerio* (Bilyk and Cheng, 2013). These data suggest that dysfunctional regulatory elements controlling the heat-induced expression of molecular chaperones is at least partly responsible for the extreme stenothermal lifestyle of these fish (Buckley et al., 2004).

The uncertain relationship between mRNA and protein

Since its inception, transcriptomics has been criticized for the lack of correspondence between mRNA and protein abundance (Greenbaum et al., 2003; Tian et al., 2004). Fitness is primarily the result of proteins, and should mRNA abundance serve as an insufficient approximation of protein abundance, the ability of transcriptomics to resolve biologically meaningful information is greatly reduced (Feder and Walser, 2005). This potential discrepancy has spurred a substantial amount of research aimed at assessing how changes in transcript level relate to subsequent shifts in protein abundance. In summarizing results from a large number of studies aimed at quantifying the relationship between mRNA and protein abundance, Feder and Walser (2005) state: 'The probability of predicting whether a particular protein's concentration increases or decreases under stress would seem to be greater for a flip of a coin (50%) than for transcriptomics (typically <50%)' (Feder and Walser, 2005). Next-generation nucleic acid sequencing and improved proteomic capabilities provide new and more powerful opportunities to re-evaluate the validity of this statement. Presently, measurements of the absolute concentration of mRNAs and proteins are available from various organisms, including bacteria, yeast, worm, fly and human cells (Vogel and Marcotte, 2012). These recent data mostly corroborate the conclusion reached by Feder and Walser (2005) that on average, transcription regulation is only half

the story of protein abundance (Plotkin, 2010). In general, in both bacteria and eukaryotes, the cellular concentrations of proteins correlate with the abundance of their corresponding mRNAs. However, the correlation is not strong: approximately 40% of the variation in protein concentration can be explained by knowing mRNA abundance (de Sousa Abreu et al., 2009; Maier et al., 2009; Schwanhäusser et al., 2011; Vogel and Marcotte, 2012). In human cells, only 27% of protein concentration can be explained by mRNA abundance, while the remaining 73% of variation is explained by other known and unknown factors (Fig. 3; Vogel and Marcotte, 2012). These data suggest a key role for processes downstream of transcription that are missed when using transcriptomic profiling.

Cellular concentrations of mRNA and protein are ultimately the net result of transcription, mRNA degradation, translation and protein degradation. As a result, changes in mRNA stability can achieve the same cellular outcome as changes in transcript abundance. A constitutively expressed mRNA with a long half-life may produce the same amount of protein as an abundant mRNA with a short half-life, and vice versa. However, only the change in transcript abundance will be detected via transcriptomics. Interestingly, natural selection appears to have modified mRNA and protein half-lives according to protein function (Schwanhäusser et al., 2011). In mammalian cells, genes with mRNAs and proteins that have long half-lives are functionally enriched for constitutive cellular processes like translation (e.g. ribosomal proteins) and metabolism (e.g. glycolytic and citric acid cycle enzymes). In contrast, the subset of mammalian genes whose mRNAs and proteins have short half-lives are significantly enriched in transcription factors, signaling genes, chromatin-modifying enzymes and genes with cell cycle-specific functions (Schwanhäusser et al., 2011). Ontologies of genes whose mRNAs and proteins have short half-lives overlap with ontologies of genes that are apparently most important for environmental adaptation in model organisms. Recall that genes required for heat tolerance in *S. cerevisiae* were also enriched for genes involved in cell signaling and chromatin modification (Gibney et al., 2013). The transient nature of regulatory mRNAs makes these transcripts more difficult to identify using transcriptomics.
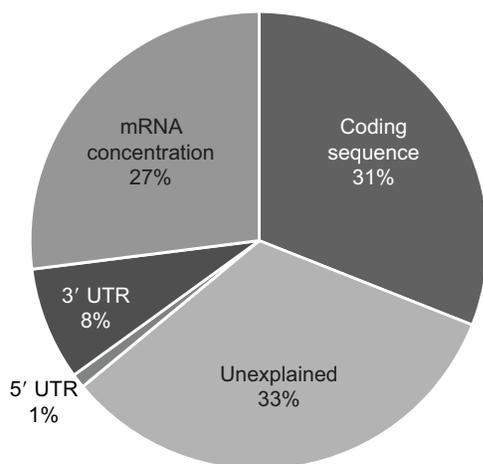

Fig. 3. Percentage of variance in protein abundance explained by various factors in a human medulloblastoma cell line. As shown, 27% of variance in protein abundance can be explained by mRNA abundance. The remaining 73% of variance is explained by other factors and is indicative of extensive post-transcriptional and post-translational regulation. UTR, untranslated region. Adapted from Vogel and Marcotte (2012).

Repeatedly identifying the same genes

Previous sections of this review have described biological phenomenon that place limits on the capacity for transcriptomics to isolate genes with large effects on environmental tolerance. However, technical issues also constrain the value of data derived from transcriptomic experiments. Aspects of transcriptomic experimental design and data analysis have been questioned in the past. Early transcriptomic studies were hampered by inappropriate statistical analyses, a problem that was resolved through the development of statistical methods specific to transcriptomic datasets (Storey and Tibshirani, 2003; Hackstadt and Hess, 2009; Rau et al., 2013). Analyzing long lists of differentially expressed genes also posed a challenge for early transcriptomic experiments (Gracey, 2007). The creation of data analysis tools that relate changes in gene expression to larger-scale cellular and physiological outcomes has greatly reduced this burden (Thomas et al., 2003; Huang et al., 2008). Yet, as transcriptomic studies accumulate in the scientific literature (Fig. 1), new technical biases arise that may also require attention. One bias pertinent to the field of environmental physiology is the potential for transcriptomic studies to repeatedly isolate the same set of environmentally regulated genes. While this result is in itself quite meaningful, providing empirical evidence for an important unifying theme in biology, this tendency limits insight into genes responsible for differences in environmental tolerance between species.

The cellular stress response is an important unifying theme in biology. The concept of a cellular stress response posits that virtually all life on Earth responds to acute environmental change by inducing a homologous set of proteins. This response represents a reaction to macromolecular damage, and because many different types of environmental stress cause similar types of cellular damage (e.g. protein unfolding), the same set of genes are up-regulated in response to shifts in different abiotic variables. These environmentally regulated proteins, which include molecular chaperones, DNA damage and repair proteins, proteolytic proteins and certain metabolic enzymes, act to prevent and repair macromolecular damage and stabilize basic aspects of cell function. Proteins contributing to the cellular stress response are among the most highly conserved across the three superkingdoms of life, a trend that asserts their fundamental importance (Kültz, 2003, 2005). Standard procedures in the analysis of transcriptomic data are likely biasing the resulting set of significantly differentially expressed genes toward those involved in the cellular stress response. The transcriptomic data analysis pipeline often prioritizes highly conserved and strongly up- or down-regulated mRNAs at the expense of genes that cannot be annotated or exhibit muted fold-changes. Imposing a fold-change cut-off is common practice, but introduces bias (Dalman et al., 2012), and is rooted in the potentially flawed assumption that a large change in mRNA abundance equates to a large impact on fitness (Gibney et al., 2013). Furthermore, only genes that can be annotated are typically subject to more detailed analyses, and genes comprising the cellular stress response are easily annotated because they are widely conserved and well studied. Similar analytical approaches have biased proteomic datasets toward the repeated identification of cellular stress response proteins. Meta-analyses of the comparative proteomics literature have generated a list of repeatedly detected proteins that are consistently differentially expressed independent of target species, *in vivo* or *in vitro* conditions, tissue or organ assayed, or experimental objective (Petrak et al., 2008; Mariman, 2009; Wang et al., 2009b). These proteins account for 23.2% of those considered differentially expressed, but represent only 4.9% of unique proteins

across the species assayed. In a striking example of this phenomenon, 83% of the studies included in a proteomics meta-analysis reported a change in a member of the Hsp70 family. Gene ontology classification and tests for functional enrichment of the 44 most frequently detected proteins yield striking similarity to ontologies comprising the cellular stress response (Fig. 4; Wang et al., 2009b). While similar meta-analyses are not available for transcriptomic datasets, biases in the proteomic literature are the result of the same data-filtering practices frequently implemented in transcriptomics that emphasize highly expressed and easily annotated genes. That diverse taxa respond to acute environmental stress in such a conserved manner is certainly informative, but provides limited insight as to why species differ so widely in their tolerance toward environmental change. In fact, the ability to turn the cellular stress response on and off, or modify the abundance of specific components of the response, is more likely to contribute to differences in environmental tolerance (Healy et al., 2010; Tomanek, 2010), and this possibility again emphasizes the importance of regulatory genes that are less likely to alter mRNA abundance in response to environmental change.

Overlooking taxonomically restricted genes

Differentially expressed genes that cannot be annotated are typically given limited attention when analyzing transcriptomic datasets. With no homology from which to base conclusions, speculating on the function or importance of these genes without exceeding the bounds of the data is challenging. With each new microarray or RNA sequencing dataset, sequenced genome or expressed sequence tag collection, the number of these taxonomically restricted genes (also referred to as 'orphan' or 'unknown' genes) proliferates (Khalturin et al., 2009). Clearly, this is a major obstacle to elucidating the genetic basis for any trait, not just for understanding environmental adaptation (Oh et al., 2012). Comparative genome analyses indicate that every taxonomic group studied so far contains 10–20% of genes that lack recognizable homologs in other species and therefore cannot be annotated (Khalturin et al., 2009). Even in model organisms with long histories of research use, the number of genes with uncharacterized functions is substantial. Analyses from

Fig. 4. Gene ontology classification of proteins generally detected in comparative proteomics. Each of the 44 generally detected proteins is assigned to only one functional category. The percentage of these proteins in each functional category is shown. Adapted from Wang et al. (2009b). Stars indicate that this functional category is also part of the minimal stress proteome defined by Kültz (2005).

the 12 *Drosophila* genomes project revealed that at least 15% of genes from other *Drosophila* congeners do not show any significant sequence similarity to *D. melanogaster* gene models (Drosophila 12 Genomes Consortium, 2007). Similarly, 14% of *S. pombe* protein-coding genes could not be found in *S. cerevisiae* (Wood et al., 2002) and 11% of genes in *C. elegans* do not have any significant homology to genes in the *Caenorhabditis briggsae* genome (Stein et al., 2003). Over 36% of the minimal set of *Daphnia pulex* genes have no detectable homology to genes in other species (Colbourne et al., 2011). Comparative analyses of 122 bacterial genomes suggest that the number of taxonomically restricted genes in bacteria is actually increasing (Wilson et al., 2005). Some have hypothesized that taxonomically restricted genes comprise spurious, non-functional open reading frames (Clamp et al., 2007), but increasing evidence refutes this possibility and most of these genes appear functional (Gibson et al., 2013). Others predicted that the number of unknown genes would decrease quickly as molecular databases incorporated sequences from more and more organisms, yet this prediction has not been validated either (Casari et al., 1996). While the number of sequences deposited in GenBank is increasing at an exponential rate, the proportion of genes that show no similarity to previously sequenced genes remains at 10–20%, depending on the cut-off threshold used in Basic Local Alignment Search Tool (BLAST) protein similarity searches (Khalturin et al., 2009). Poor sequence assembly or quality is not a major contributor to the high proportion of unknown genes in eukaryotic genomes (Gibson et al., 2013). Despite the ubiquity of taxonomically restricted genes, and the potential for these genes to contribute to the unique traits of different species, they have been granted comparatively little attention in transcriptomic responses to the environment and from the scientific community in general.

A reasonable assumption is that taxonomically restricted genes are responsible for the specific adaptations of organisms to their ecological setting and contribute to species-specific responses to the environment (Khalturin et al., 2009; Voolstra et al., 2011). The function of taxonomically restricted genes will eventually need to be investigated if environmental physiologists are to comprehensively identify genes that are important for modifying tolerance limits. In a clear illustration of this premise, 23 of the 55 genes shown to influence heat sensitivity in *S. cerevisiae* annotate as having unknown functions (Gibney et al., 2013). Similarly, 17%, 10% and 6% of the genes significantly affecting tolerance toward ethanol, H_2O_2 and NaCl in *S. cerevisiae* were also classified as unknown genes, respectively (Auesukaree et al., 2009). Among the genes that modify the capacity for protein folding in *C. elegans*, 19% are of unknown function. Similar trends are also apparent in plants. In the Brassicaceae, of which the plant model *A. thaliana* is a member, family-specific genes are more likely to alter expression in response to abiotic stresses than more conserved genes (Oh et al., 2012).

Improving experimental designs

Future transcriptomic studies are likely to benefit from improved experimental designs and reduced costs that offer a means to increase the probability of identifying genes with adaptive significance. Comparative approaches that relate transcriptional responses to the environment between one or more taxons offer considerable improvement over single-species analyses (Whitehead, 2012). Comparative experiments allow conserved responses to environmental change to be separated from responses that are exclusive to particular species or populations (Stillman and Tagmount, 2009; Meyer and Manahan, 2010; Healy et al., 2010;
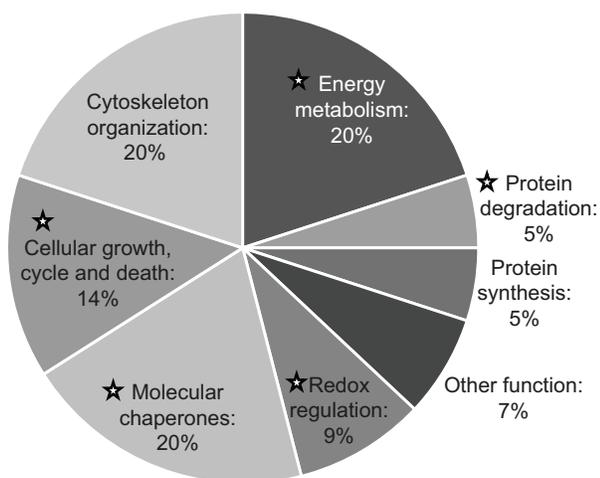
Lockwood et al., 2010; Lockwood and Somero, 2011; Barshis et al., 2013). Genes responding identically to environmental change across species are more likely to be under strong evolutionary constraint, such as those involved in the conserved cellular stress response, and less likely to contribute to species-specific differences in environmental tolerance. Genes exhibiting expression patterns unique to particular taxa or phenotypes are more likely to have evolved to enable novel life histories (Whitehead, 2012; Whitehead et al., 2011). The comparative approach can be further improved by integrating phylogeny (Whitehead, 2012) or genetic background (Meyer and Manahan, 2010) into downstream analyses. For example, if three or more species or populations are being compared and a robust phylogeny is available, one may test whether the transcriptomic response across species is consistent with neutral expectations or rejects neutral expectations. These more sophisticated designs can better isolate variation in gene expression that is adaptive from variation that arose from neutral evolutionary processes (for a complete description and examples, see Whitehead, 2012). Of course implementing these more informative approaches is reliant upon having genetic or phylogenetic and genomic resources; however, prior knowledge of the value of these data may persuade researchers to pursue both pieces of information.

Integrating gene network analyses

A major obstacle presently limiting the scientific value of transcriptomics in studies of environmental adaptation is the inability to develop unbiased criteria that can identify genes with disproportionate effects on fitness from lists of hundreds or thousands of candidate genes. Using transcriptomic data to construct environmentally regulated gene networks represents a promising approach to overcoming this obstacle. Studies in yeast demonstrate that connectivity is one metric that can be used to more reliably identify genes that influence fitness. Construction of a genome-scale genetic interaction map for *S. cerevisiae* reveals a strong correlation between gene connectedness and fitness: single mutants with overt fitness defects tend to exhibit an increased number of genetic interactions (r=0.73). These hub genes showed a high degree of pleiotropy, suggesting that strong influences on fitness are a consequence of genes participating in multiple cellular pathways. Consistent with the prediction that genes with prominent effects on fitness are rare, only a very small number of *S. cerevisiae* genes serve as network hubs (<1%; Costanzo et al., 2010). Network analyses also provide a means to assess connectivity in genes that do not undergo expression changes in response to environmental perturbation. Gene networks are assembled using databases of known genetic interactions, and, consequently, differential expression is not a prerequisite for inclusion in network modelling. For example, Costanzo et al. (2010) used only 30% of the *S. cerevisiae* genome as query genes, but recovered genetic interactions for approximately 75% of the genome. Nonetheless, incorporating gene expression data does increase the accuracy of resulting networks because genes exhibiting similar expression profiles can be connected into modules of co-expressed genes (Fu et al., 2014). Knowledge of network topology can also facilitate predictive power. Yeast gene networks assembled using transcriptomic data accurately predict transcriptional responses to conditions not considered during network construction (Danziger et al., 2014).

Tools for gene network construction are numerous and accessible (Lee and Tzou, 2009), yet are under-utilized in comparison with other data analysis strategies such as gene functional annotation and gene set enrichment analysis (Thomas et al., 2003; Subramanian

et al., 2005; Huang et al., 2008). Forays into network analyses within the context of environmental adaptation in non-model systems have proven informative and should increase in the future. Application of artificial neural network models to transcriptomic profiles of wild Eastern oysters *Crassostrea virginica* provided a means to isolate genes responding to specific environmental variables in highly dynamic estuarine environments (Chapman et al., 2011). A similar approach identified genes responding to changes in anthropogenic land use in *C. virginica* (Chapman et al., 2009), and genes underlying egg quality in striped bass *Morone saxatilis* (Chapman et al., 2014). In European flounder, network models integrating transcriptional, metabolic and phenotypic information were able to predict chemical exposure in the wild and identify new genes participating in chemical response pathways (Williams et al., 2011).

Expanding functional analyses

Even in cases where robust experimental design and sophisticated data analysis tools can be implemented, the actual contribution of a particular gene to whole-organism tolerance limits will remain uncertain until functional assessment can be performed. As stated by Feder and Walser (2005): 'As even the most avid advocates of transcriptomics admit, transcriptomics should not be an end in itself, but a means for revealing candidate mRNA species whose abundance and phenotypes as proteins require subsequent study' (Feder and Walser, 2005). In the context of this statement, an effective use of transcriptomics involves coupling gene expression profiling with functional analyses such as targeted gene knock-down or over-expression. Presently, comprehensive gene functional information is restricted to a few, simple model systems under a small number of environmental conditions. While these studies are providing valuable information concerning the relationship between gene expression and fitness, a major uncertainty is whether these same trends will hold true across a wider range of species. Transcriptomics offers a sensitive means to identify environmentally regulated genes; however, the expanded use of transcriptomics has meant the identification of environmentally regulated genes (including many 'unknown' genes) has outpaced subsequent functional characterization. The inability to transition from experimental approaches aimed largely at identifying genes involved in environmental adaptation toward those that actually evaluate the functional importance of genes in setting tolerance limits is impeding understanding of how organisms adapt to the environment. The need for gene functional analyses does seem palpable within the scientific community (Kettleborough et al., 2013), and much-needed progress is being made to expand gene functional analyses beyond model organisms to target specific genes of interest in a wider range of species (e.g. CRISPRi) (Larson et al., 2013; Qi et al., 2013) and manipulate the expression of those genes at specific times (e.g. condition-specific strategies) (Skarnes et al., 2011). A more informed perspective of the value of transcriptomics to environmental physiology will emerge as gene functional analyses become more prevalent.

Summary

Transcriptomics has made an overwhelmingly positive contribution to the field of environmental physiology, and should continue to do so as genomic approaches become more tractable and affordable. However, some limitations of transcriptomics have become apparent along with these considerable advances. The ability of transcriptomics to identify genes with major effects on fitness during environmental stress has been questioned previously (Feder

and Walser, 2005), and evidence for this uncertainty remains today. Systems-level experiments in simple model organisms imply that few individual genes are capable of modifying organism tolerance limits, and often the expression of these genes is not altered by the environment. Accumulating evidence indicates that regulatory molecules whose expression does not change during environmental stress, and that are consequently overlooked in transcriptomic screens, play major roles in environmental adaptation. Transcriptomics has also been criticized for the inconsistent relationship between mRNA abundance and fitness (Feder and Walser, 2005), and new research indicates not only that an increase or decrease in mRNA may not equate to a corresponding change in protein abundance or activity but also that regulatory proteins more likely to influence environmental sensitivity have mRNAs with short half-lives. These limitations, combined with technical issues such as the tendency to focus on genes that are highly conserved and undergo large fold-changes, indicate that transcriptomics is not a suitable method to comprehensively identify genes involved in adaptive responses to the environment. As transcriptomics infiltrates new areas of research and more investigators consider integrating this technology into their research programs, knowledge of these biases, in addition to the substantial benefits offered by transcriptomics, is critical to determining the types of experimental questions that transcriptomics can be directed toward, improving experimental designs and accurately interpreting transcriptomic data.

References

Abzhanov, A., Kuo, W. P., Hartmann, C., Grant, B. R., Grant, P. R. and Tabin, C. J. (2006). The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature* **442**, 563-567.

Auesukaree, C., Damnernsawad, A., Kruatrachue, M., Pokethitiyook, P., Boonchird, C., Kaneko, Y. and Harashima, S. (2009). Genome-wide identification of genes involved in tolerance to various environmental stresses in *Saccharomyces cerevisiae*. *J. Appl. Genet.* **50**, 301-310.

Barshis, D. J., Ladner, J. T., Oliver, T. A., Seneca, F. O., Traylor-Knowles, N. and Palumbi, S. R. (2013). Genomic basis for coral resilience to climate change. *Proc. Natl. Acad. Sci. USA* **110**, 1387-1392.

Batada, N. N. and Hurst, L. D. (2007). Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat. Genet.* **39**, 945-949.

Batada, N. N., Hurst, L. D. and Tyers, M. (2006). Evolutionary and physiological importance of hub proteins. *PLoS Comput. Biol.* **2**, e88.

Benner, I., Diner, R. E., Lefebvre, S. C., Li, D., Komada, T., Carpenter, E. J. and Stillman, J. H. (2013). *Emiliania huxleyi* increases calcification but not expression of calcification-related genes in long-term exposure to elevated temperature and pCO$_2$. *Philos. Trans. R. Soc. B. Biol. Sci.* **368**, 20130049.

Berry, D. B., Guan, Q., Hose, J., Haroon, S., Gebbia, M., Heisler, L. E., Nislow, C., Giaever, G. and Gasch, A. P. (2011). Multiple means to the same end: the genetic basis of acquired stress resistance in yeast. *PLoS Genet.* **7**, e1002353.

Bettencourt, B. R., Hogan, C. C., Nimali, M. and Drohan, B. W. (2008). Inducible and constitutive heat shock gene expression responds to modification of Hsp70 copy number in *Drosophila melanogaster* but does not compensate for loss of thermotolerance in Hsp70 null flies. *BMC Biol.* **6**, 5.

Bilyk, K. T. and Cheng, C.-H. C. (2013). Model of gene expression in extreme cold - reference transcriptome for the high-Antarctic cryopelagic notothenioid fish *Pagothenia borchgrevinki*. *BMC Genomics* **14**, 634.

Boutros, M., Kiger, A. A., Armknecht, S., Kerr, K., Hild, M., Koch, B., Haas, S. A., Heidelberg Fly Array Consortium, Paro, R. and Perrimon, N. (2004). Genome-wide RNAi analysis of growth and Viability in *Drosophila* cells. *Science* **303**, 832-835.

Buckley, B. A. and Somero, G. N. (2009). cDNA microarray analysis reveals the capacity of the cold-adapted Antarctic fish *Trematomus bernacchii* to alter gene expression in response to heat stress. *Polar Biol.* **32**, 403-415.

Buckley, B. A., Place, S. P. and Hofmann, G. E. (2004). Regulation of heat shock genes in isolated hepatocytes from an Antarctic fish, *Trematomus bernacchii*. *J. Exp. Biol.* **207**, 3649-3656.

Carpenter, A. E. and Sabatini, D. M. (2004). Systematic genome-wide screens of gene function. *Nat. Rev. Genet.* **5**, 11-22.

Casari, G., de Daruvar, A., Sander, C. and Schneider, R. (1996). Bioinformatics and the discovery of gene function. *Trends Genet.* **12**, 244-245.

Chalancon, G., Ravarani, C. N. J., Balaji, S., Martinez-Arias, A., Aravind, L., Jothi, R. and Babu, M. M. (2012). Interplay between gene expression noise and regulatory network architecture. *Trends Genet.* **28**, 221-232.

Chapman, R. W., Mancia, A., Beal, M., Veloso, A., Rathburn, C., Blair, A., Sanger, D., Holland, A. F., Warr, G. W. and Didonato, G. U. Y. (2009). A transcriptomic analysis of land-use impacts on the oyster, *Crassostrea virginica*, in the South Atlantic bight. *Mol. Ecol.* **18**, 2415-2425.

Chapman, R. W., Mancia, A., Beal, M., Veloso, A., Rathburn, C., Blair, A., Holland, A. F., Warr, G. W., Didinato, G., Sokolova, I. M. et al. (2011). The transcriptomic responses of the eastern oyster, *Crassostrea virginica*, to environmental conditions. *Mol. Ecol.* **20**, 1431-1449.

Chapman, R. W., Reading, B. J. and Sullivan, C. V. (2014). Ovary transcriptome profiling via artificial intelligence reveals a transcriptomic fingerprint predicting egg quality in Striped Bass, *Morone saxatilis*. *PLoS ONE* **9**, e96818.

Chen, W.-H., Trachana, K., Lercher, M. J. and Bork, P. (2012). Younger genes are less likely to be essential than older genes, and duplicates are less likely to be essential than singletons of the same Age. *Mol. Biol. Evol.* **29**, 1703-1706.

Clamp, M., Fry, B., Kamal, M., Xie, X., Cuff, J., Lin, M. F., Kellis, M., Lindblad-Toh, K. and Lander, E. S. (2007). Distinguishing protein-coding and noncoding genes in the human genome. *Proc. Natl. Acad. Sci. USA* **104**, 19428-19433.

Cohen, P., Miyazaki, M., Socci, N. D., Hagge-Greenberg, A., Liedtke, W., Soukas, A. A., Sharma, R., Hudgins, L. C., Ntambi, J. M. and Friedman, J. M. (2002). Role for stearoyl-CoA desaturase-1 in leptin-mediated weight loss. *Science* **297**, 240-243.

Colbourne, J. K., Pfrender, M. E., Gilbert, D., Thomas, W. K., Tucker, A., Oakley, T. H., Tokishita, S., Aerts, A., Arnold, G. J., Basu, M. K. et al. (2011). The ecoresponsive genome of *Daphnia pulex*. *Science* **331**, 555-561.

Conant, G. C. and Wagner, A. (2004). Duplicate genes and robustness to transient gene knock-downs in *Caenorhabditis elegans*. *Proc. R. Soc. Lond. B Biol. Sci.* **271**, 89-96.

Cossins, A., Fraser, J., Hughes, M. and Gracey, A. (2006). Post-genomic approaches to understanding the mechanisms of environmentally induced phenotypic plasticity. *J. Exp. Biol.* **209**, 2328-2336.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., Ding, H., Koh, J. L. Y., Toufighi, K., Mostafavi, S. et al. (2010). The genetic landscape of a cell. *Science* **327**, 425-431.

Dalman, M. R., Deeter, A., Nimishakavi, G. and Duan, Z.-H. (2012). Fold change and p-value cutoffs significantly alter microarray interpretations. *BMC Bioinformatics* **13**, S11.

Danziger, S. A., Ratushny, A. V., Smith, J. J., Saleem, R. A., Wan, Y., Arens, C. E., Armstrong, A. M., Sitko, K., Chen, W.-M., Chiang, J.-H. et al. (2014). Molecular mechanisms of system responses to novel stimuli are predictable from public data. *Nucleic Acids Res.* **42**, 1442-1460.

de Nadal, E., Ammerer, G. and Posas, F. (2011). Controlling gene expression in response to stress. *Nat. Rev. Genet.* **12**, 833-845.

de Sousa Abreu, R., Penalva, L. O., Marcotte, E. M. and Vogel, C. (2009). Global signatures of protein and mRNA expression levels. *Mol. Biosyst.* **5**, 1512-1526.

De Wit, P., Pespeni, M. H., Ladner, J. T., Barshis, D. J., Seneca, F., Jaris, H., Therkildsen, N. O., Morikawa, M. and Palumbi, S. R. (2012). The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Mol. Ecol. Resour.* **12**, 1058-1067.

***Drosophila* 12 Genomes Consortium** (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203-218.

Evans, T. G. and Hofmann, G. E. (2012). Defining the limits of physiological plasticity: how gene expression can assess and predict the consequences of ocean change. *Philos. Trans. R. Soc. B. Biol. Sci.* **367**, 1733-1745.

Evans, T. G. and Somero, G. N. (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.

Evans, T. G., Hammill, E. D. D., Kaukinen, K., Schulze, A. D., Patterson, D. A., English, K. K., Curtis, J. M. R. and Miller, K. M. (2011). Transcriptomics of environmental acclimatization and survival in wild adult Pacific sockeye salmon (*Oncorhynchus nerka*) during spawning migration. *Mol. Ecol.* **20**, 4472-4489.

Evans, T. G., Chan, F., Menge, B. A. and Hofmann, G. E. (2013). Transcriptomic responses to ocean acidification in larval sea urchins from a naturally variable pH environment. *Mol. Ecol.* **22**, 1609-1625.

Feder, M. E. and Mitchell-Olds, T. (2003). Evolutionary and ecological functional genomics. *Nat. Rev. Genet.* **4**, 649-655.

Feder, M. E. and Walser, J.-C. (2005). The biological limitations of transcriptomics in elucidating stress and stress responses. *J. Evol. Biol.* **18**, 901-910.

Fu, X., Sun, Y., Wang, J., Xing, Q., Zou, J., Li, R., Wang, Z., Wang, S., Hu, X., Zhang, L. et al. (2014). Sequencing-based gene network analysis provides a core set of gene resource for understanding thermal adaptation in Zhikong scallop *Chlamys farreri. Mol. Ecol. Resour.* **14**, 184-198.

Giaever, G. and Nislow, C. (2014). The yeast deletion collection: a decade of functional genomics. *Genetics* **197**, 451-465.

Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B. et al. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387-391.

Gibney, P. A., Lu, C., Caudy, A. A., Hess, D. C. and Botstein, D. (2013). Yeast metabolic and signaling genes are required for heat-shock survival and have little overlap with the heat-induced genes. *Proc. Natl. Acad. Sci. USA* **110**, E4393-E4402.

Gibson, A. K., Smith, Z., Fuqua, C., Clay, K. and Colbourne, J. K. (2013). Why so many unknown genes? Partitioning orphans from a representative transcriptome of the lone star tick *Amblyomma americanum. BMC Genomics* **14**, 135.

Gong, W. J. and Golic, K. G. (2006). Loss of Hsp70 in *Drosophila* is pleiotropic, with effects on thermotolerance, recovery from heat shock and neurodegeneration. *Genetics* **172**, 275-286.

Gracey, A. Y. (2007). Interpreting physiological responses to environmental change through gene expression profiling. *J. Exp. Biol.* **210**, 1584-1592.

Gracey, A. Y. and Cossins, A. R. (2003). Application of microarray technology in environmental and comparative physiology. *Annu. Rev. Physiol.* **65**, 231-259.

Gracey, A. Y., Lee, T.-H., Higashi, R. M. and Fan, T. (2011). Hypoxia-induced mobilization of stored triglycerides in the euryoxic goby *Gillichthys mirabilis. J. Exp. Biol.* **214**, 3005-3012.

Greenbaum, D., Colangelo, C., Williams, K. and Gerstein, M. (2003). Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol.* **4**, 117.

Guisbert, E., Czyz, D. M., Richter, K., McMullen, P. D. and Morimoto, R. I. (2013). Identification of a tissue-selective heat shock response regulatory network. *PLoS Genet.* **9**, e1003466.

Hackstadt, A. J. and Hess, A. M. (2009). Filtering for increased power for microarray data analysis. *BMC Bioinformatics* **10**, 11.

Han, J.-D. J., Bertin, N., Hao, T., Goldberg, D. S., Berriz, G. F., Zhang, L. V., Dupuy, D., Walhout, A. J. M., Cusick, M. E., Roth, F. P. et al. (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **430**, 88-93.

Haugen, A. C., Kelley, R., Collins, J. B., Tucker, C. J., Deng, C., Afshari, C. A., Brown, M. J., Ideker, T. and Van Houten, B. (2004). Integrating phenotypic and expression profiles to map arsenic-response networks. *Genome Biol.* **5**, R95.

Healy, T. M., Tymchuk, W. E., Osborne, E. J. and Schulte, P. M. (2010). Heat shock response of killifish (*Fundulus heteroclitus*): candidate gene and heterologous microarray approaches. *Physiol. Genomics* **41**, 171-184.

Hillier, L. W., Coulson, A., Murray, J. I., Bao, Z., Sulston, J. E. and Waterston, R. H. (2005). Genomics in *C. elegans*: so many genes, such a little worm. *Genome Res.* **15**, 1651-1660.

Hochachka, P. W. and Somero, G. N. (2002). The diving response and its evolution. In *Biochemical Adaptation. Mechanism and Process in Physiological Evolution*, pp. 158-185. Oxford, UK: Oxford University Press.

Hofmann, G. E., Buckley, B. A., Airaksinen, S., Keen, J. E. and Somero, G. N. (2000). Heat-shock protein expression is absent in the Antarctic fish *Trematomus bernacchii* (family Nototheniidae). *J. Exp. Biol.* **203**, 2331-2339.

Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44-57.

Jarolim, S., Ayer, A., Pillay, B., Gee, A. C., Phrakaysone, A., Perrone, G. G., Breitenbach, M. and Dawes, I. W. (2013). *Saccharomyces cerevisiae* genes involved in survival of heat shock. *G3* **3**, 2321-2333.

Jeong, H., Mason, S. P., Barabasi, A.-L. and Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature* **411**, 41-42.

Kettleborough, R. N. W., Busch-Nentwich, E. M., Harvey, S. A., Dooley, C. M., de Bruijn, E., van Eeden, F., Sealy, I., White, R. J., Herd, C., Nijman, I. J. et al. (2013). A systematic genome-wide analysis of zebrafish protein-coding gene function. *Nature* **496**, 494-497.

Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R. and Bosch, T. C. G. (2009). More than just orphans: are taxonomically-restricted genes important in evolution? *Trends Genet.* **25**, 404-413.

Kim, D.-U., Hayles, J., Kim, D., Wood, V., Park, H.-O., Won, M., Yoo, H.-S., Duhig, T., Nam, M., Palmer, G. et al. (2010). Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe. Nat Biotech.* **28**, 617-623.

Kültz, D. (2003). Evolution of the cellular stress proteome: from monophyletic origin to ubiquitous function. *J. Exp. Biol.* **206**, 3119-3124.

Kültz, D. (2005). Molecular and evolutionary basis of the cellular stress response. *Annu. Rev. Physiol.* **67**, 225-257.

Larson, M. H., Gilbert, L. A., Wang, X., Lim, W. A., Weissman, J. S. and Qi, L. S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8**, 2180-2196.

Lee, W.-P. and Tzou, W.-S. (2009). Computational methods for discovering gene networks from expression data. *Brief. Bioinform.* **10**, 408-423.

Lehner, B. (2008). Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Mol. Syst. Biol.* **4**, 170.

Liao, B.-Y. and Zhang, J. (2007). Mouse duplicate genes are as essential as singletons. *Trends Genet.* **23**, 378-381.

Lockwood, B. L. and Somero, G. N. (2011). Transcriptomic responses to salinity stress in invasive and native blue mussels (genus *Mytilus*). *Mol. Ecol.* **20**, 517-529.

Lockwood, B. L., Sanders, J. G. and Somero, G. N. (2010). Transcriptomic responses to heat stress in invasive and native blue mussels (genus *Mytilus*): molecular correlates of invasive success. *J. Exp. Biol.* **213**, 3548-3558.

Logan, C. A. and Buckley, B. A. (2015). Transcriptomic responses to environmental temperature in eurythermal and stenothermal fishes. *J. Exp. Biol.* **218**, 1915-1924.

Logan, C. A. and Somero, G. N. (2010). Transcriptional responses to thermal acclimation in the eurythermal fish *Gillichthys mirabilis* (Cooper 1864). *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **299**, R843-R852.

López-Maury, L., Marguerat, S. and Bähler, J. (2008). Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat. Rev. Genet.* **9**, 583-593.

Lu, X., Jain, V. V., Finn, P. W. and Perkins, D. L. (2007). Hubs in biological interaction networks exhibit low changes in expression in experimental asthma. *Mol. Syst. Biol.* **3**, 98.

MacNeil, L. T. and Walhout, A. J. M. (2011). Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Res.* **21**, 645-657.

Maier, T., Güell, M. and Serrano, L. (2009). Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* **583**, 3966-3973.

Mariman, E. C. M. (2009). 2DE-proteomics meta-data indicate the existence of distinct cellular stress-responsive mechanisms. *Expert Rev. Proteomics* **6**, 337-339.

May, B. P. and Martienssen, R. A. (2003). Transposon mutagenesis in the study of plant development. *CRC Crit. Rev. Plant Sci.* **22**, 1-35.

Merbl, Y. and Kirschner, M. W. (2011). Protein microarrays for genome-wide posttranslational modification analysis. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **3**, 347-356.

Meyer, E. and Manahan, D. T. (2010). Gene expression profiling of genetically determined growth variation in bivalve larvae (*Crassostrea gigas*). *J. Exp. Biol.* **213**, 749-758.

Mir, S. S., Fiedler, D. and Cashikar, A. G. (2009). Ssd1 is required for thermotolerance and Hsp104-mediated protein disaggregation in *Saccharomyces cerevisiae. Mol. Cell. Biol.* **29**, 187-200.

Oh, D.-H., Dassanayake, M., Bohnert, H. J. and Cheeseman, J. M. (2012). Life at the extreme: lessons from the genome. *Genome Biol.* **13**, 241.

Ozsolak, F. and Milos, P. M. (2011). RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* **12**, 87-98.

Palumbi, S. R., Barshis, D. J., Traylor-Knowles, N. and Bay, R. A. (2014). Mechanisms of reef coral resistance to future climate change. *Science* **344**, 895-898.

Pespeni, M. H., Sanford, E., Gaylord, B., Hill, T. M., Hosfelt, J. D., Jaris, H. K., LaVigne, M., Lenz, E. A., Russell, A. D., Young, M. K. et al. (2013). Evolutionary change during experimental ocean acidification. *Proc. Natl. Acad. Sci. USA* **110**, 6937-6942.

Petrak, J., Ivanek, R., Toman, O., Cmejla, R., Cmejlova, J., Vyoral, D., Zivny, J. and Vulpe, C. D. (2008). Déjà vu in proteomics. A hit parade of repeatedly identified differentially expressed proteins. *Proteomics* **8**, 1744-1749.

Place, S. P. and Hofmann, G. E. (2005). Comparison of Hsc70 orthologs from polar and temperate notothenioid fishes: differences in prevention of aggregation and refolding of denatured proteins. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **288**, R1195-R1202.

Place, S. P., Zippay, M. L. and Hofmann, G. E. (2004). Constitutive roles for inducible genes: evidence for the alteration in expression of the inducible hsp70 gene in Antarctic notothenioid fishes. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **287**, R429-R436.

Plotkin, J. B. (2010). Transcriptional regulation is only half the story. *Mol. Syst. Biol.* **6**, 406.

Podrabsky, J. E. and Somero, G. N. (2004). Changes in gene expression associated with acclimation to constant temperatures and fluctuating daily temperatures in an annual killifish *Austrofundulus limnaeus. J. Exp. Biol.* **207**, 2237-2254.

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P. and Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183.

Qian, X., Ba, Y., Zhuang, Q. and Zhong, G. (2014). RNA-Seq technology and its application in fish transcriptomics. *OMICS* **18**, 98-110.

Rau, A., Gallopin, M., Celeux, G. and Jaffrézic, F. (2013). Data-based filtering for replicated high-throughput transcriptome sequencing experiments. *Bioinformatics* **29**, 2146-2152.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature* **473**, 337-342.

Shaw, J. R., Hampton, T. H., King, B. L., Whitehead, A., Galvez, F., Gross, R. H., Keith, N., Notch, E., Jung, D., Glaholt, S. P. et al. (2014). Natural selection canalizes expression variation of environmentally induced plasticity-enabling genes. *Mol. Biol. Evol.* **31**, 3002-3015.

Silva, M. C., Fox, S., Beam, M., Thakkar, H., Amaral, M. D. and Morimoto, R. I. (2011). A genetic screening strategy identifies novel regulators of the proteostasis network. *PLoS Genet.* **7**, e1002438.

Singh, A. H., Wolf, D. M., Wang, P. and Arkin, A. P. (2008). Modularity of stress response evolution. *Proc. Natl. Acad. Sci. USA* **105**, 7500-7505.

Skarnes, W. C., Rosen, B., West, A. P., Koutsourakis, M., Bushell, W., Iyer, V., Mujica, A. O., Thomas, M., Harrow, J., Cox, T. et al. (2011). A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337-342.

Somero, G. N. (2011). Comparative physiology: a "crystal ball" for predicting consequences of global change. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **301**, R1-R14.

Somero, G. N. (2012). The physiology of global change: linking patterns to mechanisms. *Annu. Rev. Mar. Sci.* **4**, 39-61.

Stein, L. D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M. R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A. et al. (2003). The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol.* **1**, e45.

Stillman, J. H. and Tagmount, A. (2009). Seasonal and latitudinal acclimatization of cardiac transcriptome responses to thermal stress in porcelain crabs, *Petrolisthes cinctipes*. *Mol. Ecol.* **18**, 4206-4226.

Storey, J. D. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100**, 9440-9445.

Suarez, R. K. and Moyes, C. D. (2012). Metabolism in the age of 'omes'. *J. Exp. Biol.* **215**, 2351-2357.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545-15550.

Thatcher, J. W., Shaw, J. M. and Dickinson, W. J. (1998). Marginal fitness contributions of nonessential genes in yeast. *Proc. Natl. Acad. Sci. USA* **95**, 253-257.

Thomas, P. D., Kejariwal, A., Campbell, M. J., Mi, H., Diemer, K., Guo, N., Ladunga, I., Ulitsky-Lazareva, B., Muruganujan, A., Rabkin, S. et al. (2003). PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res.* **31**, 334-341.

Tian, Q., Stepaniants, S. B., Mao, M., Weng, L., Feetham, M. C., Doyle, M. J., Yi, E. C., Dai, H., Thorsson, V., Eng, J. et al. (2004). Integrated genomic and proteomic analyses of gene expression in mammalian cells. *Mol. Cell. Proteomics* **3**, 960-969.

Tomanek, L. (2010). Variation in the heat shock response and its implication for predicting the effect of global climate change on species' biogeographical distribution ranges and metabolic costs. *J. Exp. Biol.* **213**, 971-979.

Vogel, C. and Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227-232.

Voolstra, C. R., Sunagawa, S., Matz, M. V., Bayer, T., Aranda, M., Buschiazzo, E., DeSalvo, M. K., Lindquist, E., Szmant, A. M., Coffroth, M. A. et al. (2011). Rapid evolution of coral proteins responsible for interaction with the environment. *PLoS ONE* **6**, e20392.

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

Wang, P., Bouwman, F. G. and Mariman, E. C. M. (2009b). Generally detected proteins in comparative proteomics – a matter of cellular stress response? *Proteomics* **9**, 2955-2966.

Warringer, J., Ericson, E., Fernandez, L., Nerman, O. and Blomberg, A. (2003). High-resolution yeast phenomics resolves different physiological features in the saline response. *Proc. Natl. Acad. Sci. USA* **100**, 15724-15729.

Welch, A. J., Bedoya-Reina, O. C., Carretero-Paulet, L., Miller, W., Rode, K. D. and Lindqvist, C. (2014). Polar bears exhibit genome-wide signatures of bioenergetic adaptation to life in the arctic environment. *Genome Biol. Evol.* **6**, 433-450.

Whitehead, A. (2012). Comparative genomics in ecological physiology: toward a more nuanced understanding of acclimation and adaptation. *J. Exp. Biol.* **215**, 884-891.

Whitehead, A., Roach, J. L., Zhang, S. and Galvez, F. (2011). Genomic mechanisms of evolved physiological plasticity in killifish distributed along an environmental salinity gradient. *Proc. Natl. Acad. Sci. USA* **108**, 6193-6198.

Whitehead, A., Zhang, S., Roach, J. L. and Galvez, F. (2013). Common functional targets of adaptive micro- and macro-evolutionary divergence in killifish. *Mol. Ecol.* **22**, 3780-3796.

Williams, T. D., Turan, N., Diab, A. M., Wu, H., Mackenzie, C., Bartie, K. L., Hrydziuszko, O., Lyons, B. P., Stentiford, G. D., Herbert, J. M. et al. (2011). Towards a system level understanding of non-model organisms sampled from the environment: a network biology approach. *PLoS Comput. Biol.* **7**, e1002126.

Wilson, G. A., Bertrand, N., Patel, Y., Hughes, J. B., Feil, E. J. and Field, D. (2005). Orphans as taxonomically restricted and ecologically important genes. *Microbiology* **151**, 2499-2501.

Wood, V. Gwilliam, R. Rajandream, M.-A. Lyne, M. Lyne, R. Stewart, A. Sgouros, J. Peat, N. Hayles, J. Baker, S. et al. (2002). The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**, 871-880.

Xu, H., Wilcox, D., Nguyen, P., Voorbach, M., Smith, H., Brodjian, S., Suhar, T., Reilly, R. M., Jacobson, P. B., Collins, C. A. et al. (2007). Hepatic knockdown of stearoyl-CoA desaturase 1 via RNA interference in obese mice decreases lipid content and changes fatty acid composition. *Front. Biosci.* **12**, 3781-3794.