© 2020. Published by The Company of Biologists Ltd | Journal of Experimental Biology (2020) 223, jeb191262. doi:10.1242/jeb.191262

COMMENTARY

# Help, there are 'omics' in my comparative physiology!

**Alex S. Torson[1], Yun-wei Dong[2] and Brent J. Sinclair[1,*]**


## ABSTRACT

'Omics' methods, such as transcriptomics, proteomics, lipidomics or metabolomics, yield simultaneous measurements of many related molecules in a sample. These approaches have opened new opportunities to generate and test hypotheses about the mechanisms underlying biochemical and physiological phenotypes. In this Commentary, we discuss general approaches and considerations for successfully integrating omics into comparative physiology. The choice of omics approach will be guided by the availability of existing resources and the time scale of the process being studied. We discuss the use of whole-organism extracts (common in omics experiments on small invertebrates) because such an approach may mask underlying physiological mechanisms, and we consider the advantages and disadvantages of pooling samples within biological replicates. These methods can bring analytical challenges, so we describe the most easily analyzed omics experimental designs. We address the propensity of omics studies to digress into 'fishing expeditions' and show how omics can be used within the hypothetico-deductive framework. With this Commentary, we hope to provide a roadmap that will help newcomers approach omics in comparative physiology while avoiding some of the potential pitfalls, which include ambiguous experiments, long lists of candidate molecules and vague conclusions.


**KEY WORDS: Transcriptomics, Metabolomics, Genomics, RNA-seq, Hypothesis testing**

## Introduction

A goal of comparative physiology is to understand the diverse mechanisms that allow animals to function and survive in their environments. Comparative physiologists (1) inquire vertically across levels of biological organization, (2) connect processes horizontally among organisms within ecosystems (e.g. host–pathogen interactions), and (3) aim to understand how physiological systems adapt in response to changing environments (see Mykles et al., 2010; Somero et al., 2017). A decade (or so) ago, there was a disconnect between the tools available for the non-model organisms favored by the Krogh principle, and the burden of proof that could be shouldered by transferring those questions to established model organisms (Dow, 2007). Now, high-throughput omics is routine in non-model species, genetic manipulation via RNA interference (RNAi) is almost de rigueur in some taxa (e.g. Howlett et al., 2012; Huang et al., 2018; Maori et al., 2019), CRISPR/Cas9 allows genomic manipulation in non-models (e.g. Gui et al., 2020) and genetic models can be established (relatively) quickly (for discussion, see Matthews and Vosshall, 2020). However, the core use of omics (largely transcriptomics, metabolomics and proteomics; see Glossary) in comparative physiology remains exploratory, identifying candidate molecules and processes – and thus allowing researchers to generate hypotheses – rather than revealing the mechanisms that underlie physiological phenomena. We argue that the latter should drive the questions of 'when' and 'why' to use omics in comparative physiology.

In this Commentary, we draw on our experiences – and especially our own mistakes – to build a road map for incorporating omics into comparative physiology. We do not address the strengths, weaknesses or 'promise' of specific methods or technologies – this is covered and criticized elsewhere (e.g. Dow, 2007; Karahalil, 2016; Madr et al., 2017; Suarez and Moyes, 2012), and rapid changes in technology would render our thoughts obsolete before publication. Nor do we presume to preach to the many comparative physiologists already making sterling use of omics. Rather, we aim to provide an entry point for comparative physiologists who are considering incorporating omics into their studies for the first time. We hope that this Commentary will be useful for designing and interpreting experiments, and for tempering expectations: if there is anything we have learned as a discipline, it is that omics seldom yields clear answers. Because it is our area of focus, we center our examples around invertebrate thermal biology, but the principles presented here should be applicable across comparative physiology. Opinions and recommendations are, of course, our own.

## What are 'omics'?

For the purposes of this Commentary, we define omics as any technique that identifies and quantifies a suite of sub-cellular molecules in a high-throughput fashion (Raghavachari, 2011). Although technology has increased the sophistication of these methods, comparative physiologists have long addressed questions using multivariate and integrative techniques. For example, understanding metabolic flux requires the simultaneous measurement of the activity of many enzymes (Darveau et al., 2005; Driedzic and Hochachka, 1976; Fernández et al., 2011) and understanding changes in membrane dynamics during thermal stress requires the characterization of membrane fluidity, lipid composition and activity of membrane-associated proteins (Biederman et al., 2019). Thus, these multi-molecule approaches have a longer history in the field than their recent categorization as 'omics'.

Omics approaches vary in their functional significance. For example, genomics (see Glossary) can address population-level structure (Kovach et al., 2015) and evolutionary questions (Mock et al., 2017); epigenomics can unveil regulatory mechanisms (Glastad et al., 2019) and transgenerational plasticity (Ho and Burggren, 2010); and microbiomics (usually microbial metagenomics, e.g. Ren et al., 2020) can reveal how microbes influence the phenotypes of their hosts (e.g. Moriano-Gutierrez et al., 2019). However, comparative physiologists tend to focus on functional omics such as transcriptomics, proteomics, lipidomics and metabolomics (Fig. 1), because they most directly connect to

[1]Department of Biology, The University of Western Ontario, London, ON N6A 5B7, Canada. [2]The Key Laboratory of Mariculture, Ministry of Education, Fisheries College, Ocean University of China, Qingdao 266003, PR China.

*Author for correspondence (bsincla7@uwo.ca)

A.S.T., 0000-0002-5142-6162; Y.w.D., 0000-0003-4550-2322; B.J.S., 0000-0002-8191-9910

Journal of Experimental Biology (2020) 223, jeb191262. doi:10.1242/jeb.191262

## Glossary

**Gene co-expression network**
An undirected graphic that shows which genes or gene products have correlated expression profiles. Co-expressed genes are of biological interest because they likely indicate shared regulatory control or functional relationships.

***De novo* transcriptome assembly**
The assembly of short sequences into a transcriptomic sequence without the use of a reference genome.

**Differential expression analysis**
A method of taking RNA sequencing data and conducting a statistical analysis to evaluate quantitative changes in expression between experimental groups.

**Gene ontology (GO)**
A system of classification used to characterize gene and gene-product functions across species under a unified vocabulary. This database is often used to categorize groups of genes with similar function during omics analyses. See http://geneontology.org/.

**Genomics**
The study of the structure and function of all the genetic information encoded in an organism's DNA.

**Genome-guided transcriptome assembly**
A method of transcriptome assembly that relies on existing genomic DNA sequence to align and assemble RNA sequencing reads.

**KEGG (Kyoto Encyclopedia of Genes and Genomes)**
A bioinformatics database linking genomic information to higher-level cellular processes such as metabolism, cell cycling and signal transduction. See https://www.genome.jp/kegg/.

**Metabolomics**
The characterization of a large number of metabolites in a sample. Untargeted metabolomics is the global analysis of all molecules of a similar type (e.g. polyamines) independent of known identities. Targeted metabolomics, in contrast, quantifies a suite of molecules that have already been identified.

**Orthology-based approaches**
The identification and functional characterization of genes and gene products based on sequence similarity and assumed functional conservation among species.

**Phosphoproteomics**
A subcategory of proteomics that identifies changes in the phosphorylation state of proteins.

**Proteomics**
Identification and quantification of the entire set of proteins that is produced or modified by an organism.

**Reference genome**
A template nucleic acid database that contains all or most of the genomic DNA sequence of an organism.

**RNA sequencing (RNA-seq)**
A high-throughput sequencing technique used to identify and quantify all expressed genes (i.e. mRNA) of an organism at a given point in time.

**Transcriptomics**
The study of the complete set of RNA molecules transcribed by an organism using high-throughput methods such as RNA-seq and microarray.

organismal and cellular function. Many of these omics tools are available off-the-shelf from private companies or central platforms, most of which will perform some level of analysis (for a fee, of course). We caution that commercial (usually biomedical) laboratories seldom have the resources to optimize their protocols for non-model systems; as a result, the results can be unsatisfactory or unusable. Identifying a collaborator with a vested interest in the question or organism is typically more effective, especially for metabolomics and proteomics. Nevertheless, generating some level of omics data for almost any question about an organism is now relatively affordable and accessible, whether it is a traditional model system or not.

A typical omics workflow, such as a study using RNA-seq (see Glossary), would start with (1) a simple experimental design that will facilitate pair-wise comparisons in the analysis; (2) extraction of the relevant molecules from a sample; (3) data processing [for RNA-seq, this might include genome-guided transcriptome assembly (see Glossary)]; (4) analysis [for RNA-seq, this might include differential expression analysis (see Glossary) to identify transcript expression that varies in response to a treatment]; and (5) data visualization and interpretation (often with the help of functional databases or pathway-mapping tools). These high-throughput techniques can yield long lists of 'candidate' molecules or pathways, especially when the experiments are intended to be hypothesis generating. For this reason, omics studies have a reputation as 'fishing expeditions', and it can be difficult to publish omics-only studies unaccompanied by additional experiments. How, then, do we effectively use omics in comparative physiology? We argue that a successful omics study in comparative physiology has (1) an experimental design that facilitates inquiry at and across omics-levels, (2) sophisticated interpretation of results that moves past the generation of 'candidates', and (3) an emphasis on hypothesis evaluation and/or explicit hypothesis testing.

## Designing experiments with omics in mind

Elegant experimental design remains a cornerstone of comparative physiology. Omics is now readily affordable, such that cost is increasingly unlikely to constrain experimental design. However, there are practical and design-related decisions that must be made before embarking on an omics experiment in comparative physiology. Here, we break down some of these important decisions.

### Availability of existing omics resources

Some non-model systems already have omics tools available. The existence of reference genomes (see Glossary), transcriptomes or metabolite databases diminishes some of the challenges that come with starting from scratch. However, although a high-quality reference genome or transcriptome will simplify RNA-seq workflow, published genomes vary widely in their completeness and quality (for discussions see: Hanschen and Starkenburg, 2020; Mardis et al., 2002; Seppey et al., 2019). If a reference genome is unavailable, *de novo* transcriptome assembly (see Glossary) is necessary, but relatively straightforward (e.g. Duan et al., 2015; Poelchau et al., 2011; Toxopeus et al., 2018). However, there are drawbacks associated with incomplete (i.e. draft-level) reference genomes and *de novo*-assembled reference transcriptomes: it is much more difficult to detect the involvement of regulatory elements such as promoters or alternative splicing events (Conesa et al., 2016), and some epigenomics techniques may be of little value without a high-quality reference genome (Cazaly et al., 2019).

Other omics techniques, such as metabolomics and proteomics, also benefit from established resources. Accurate identification of metabolites and proteins (as opposed to metabolite or protein 'features', which may simply be peaks in a spectrogram) is critical for informed analysis and interpretation of results (Calvete, 2014; Pinu et al., 2019). Model systems benefit from sequenced genomes and transcriptomes (which facilitate protein identification; e.g. Modahl et al., 2018), and established metabolite libraries and databases (e.g. MetaboLights, a cross-species database which covers metabolite structure and reference spectra; Haug et al., 2020). For non-models, metabolite libraries may need to be developed in-house and protein identification must rely on orthology-based approaches (see Glossary), which can limit conclusions. For example, MacMillan et al. (2016) were limited

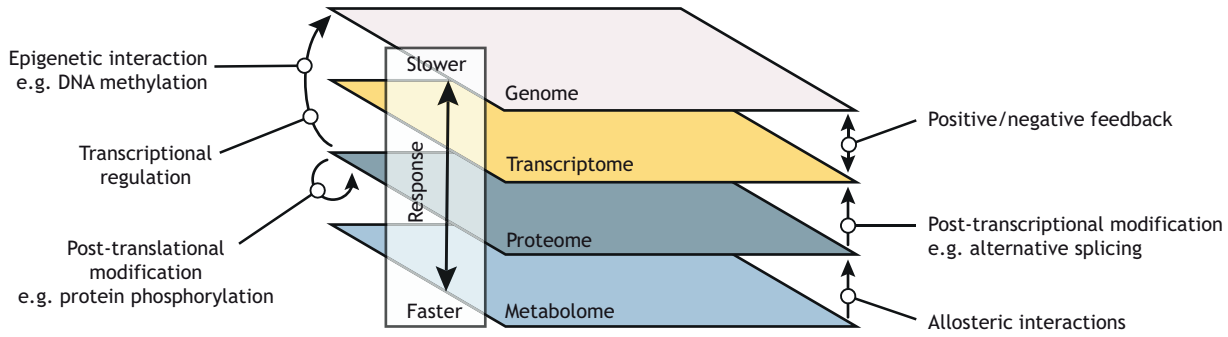


**Fig. 1. Systems-level integration of biological levels of organization.** White circles and arrows represent examples of interaction between levels and the direction of regulation, respectively. The response time box is relative to the four main omics levels, with variation at the metabolomic level capable of driving fast physiological changes and changes at the genomic level only capable of driving generational-scale changes in phenotype.

to 34 identified metabolites in *D. melanogaster* because of a small reference panel of metabolites associated with their untargeted metabolomics, whereas Hariharan et al. (2014) identified over 1000 metabolites in the same species with a larger reference library. Although a small reference panel decreases inferential power in untargeted metabolomics studies, if the hypothesis being tested predicts specific patterns of response, then identifying all the metabolites is less important than identifying large-scale patterns and the key drivers of those patterns; alternatively, a small reference panel of highly relevant metabolites will suffice. Characterizing the features in proteomics often relies on matching peptide fragments to a reference genome (Yang et al., 2019). In the absence of high-quality reference information, this process depends on orthologous relationships with existing resources, which rapidly becomes less precise as the references are less closely related to the study species (Armengaud, 2016).

## Time scales of biological processes influence the choice of omics

Metabolic and gene regulatory networks drive cellular-level responses to stimuli on distinct time scales (Shamir et al., 2016). RNA-seq is widely used by comparative physiologists because it is affordable and sensitive, and turn-key analyses are available. However, biological processes occur over a range of timescales. In eukaryotes, transcriptomic or proteomic responses (requiring ~5–10 min or tens of minutes, respectively) are too slow to regulate, for example, metabolic flux, which requires control via allosteric interaction and post-translational modification (Milo and Phillips, 2015). In ectotherms, these processes will also be temperature dependent. We suggest that omics choice should be governed by the timescale of the process: biological processes that take minutes might require, for example, phosphoproteomics (see Glossary) or metabolomics, those on the order of tens of minutes to hours are well-suited to transcriptomics and proteomics, generational-scale phenomena might be best addressed in the epigenome, and longer-term (evolutionary) processes might require input from the genome (Fig. 1).

As an example in insects, a brief (minutes) exposure to mild low temperatures increases tolerance to more extreme cold exposures, a process called rapid cold-hardening (RCH; Teets et al., 2020). However, transcriptomic signatures of RCH have been elusive: for example, Zhang et al. (2011) identified only 20 genes upregulated after a RCH-like treatment in *D. melanogaster*. In retrospect, this is unsurprising. Rapid cold-hardening can be induced in as little as 10 min, a timeframe that is unlikely to be influenced heavily by transcriptional regulation – in fact, RCH does not even require protein synthesis (Misener et al., 2001). However, such rapid changes fall within the timeframe captured by metabolomics and phosphoproteomics, which revealed gluconeogenesis, cryoprotectant synthesis and phosphorylation of cytoskeletal and stress response proteins during RCH in *Sarcophaga bullata* (Teets and Denlinger, 2016; Teets et al., 2012). By contrast, low-temperature acclimation over 6 days also improves *D. melanogaster* cold tolerance, but these physiological changes are accompanied by differential expression across one-third of the transcriptome (MacMillan et al., 2016). Thus, transcriptomic-level inquiry, which is a common first step into omics studies, may not be appropriate for all physiological processes.

## Choice of sample material could restrict inference

Animals integrate and partition physiological processes among tissues and cells. Tissue-level resolution in omics is especially important when tissues within a system respond differently to a given stimulus. In insects, for example, water balance is governed by the balance of excretion by the Malpighian tubules and absorption by the hindgut (Nation, 2015). Increased expression or activity of sodium pumps would thus increase active transport in the hindgut but decrease transport across Malpighian tubules (Des Marteaux et al., 2018a), but the antagonistic responses would be masked in a sample that includes both tissues.

While vertebrate biologists have almost always used specific tissues (e.g. the liver, brain or gills; Akashi et al., 2016; Windisch et al., 2014; Zhang et al., 2015), invertebrate comparative physiologists often homogenize entire animals (e.g. Deng et al., 2018; Des Marteaux et al., 2019; Robert et al., 2016; Torson et al., 2017; Zhang et al., 2011). This whole-animal approach is forgivable for very small animals, fast physiological processes (i.e. those that proceed more quickly than the time needed to dissect out tissues) or large experiments, and has been used to successfully generate testable hypotheses (e.g. Gleason and Burton, 2015; MacMillan et al., 2016; Meyer et al., 2011; Poupardin et al., 2015; Teets et al., 2012; Torson et al., 2015). However, we argue that, at best, using whole animals limits the capacity to draw inferences about sub-organismal processes, and at worst, may obscure some processes entirely. When whole-animal sampling is unavoidable, additional care can be inserted into data processing. For example, bacterial, fungal and plant contamination included with the gut tissue can be identified through matches to databases and tools such as DeconSeq (Sangiovanni et al., 2019; Schmieder and Edwards, 2011). It is possible to extend this argument further to include among-cell variation within a tissue (e.g. there are distinct cell types in the Malpighian tubules, Halberg et al., 2015), and techniques are available for single-cell omics (e.g. scRNA-seq; Kolodziejczyk

Journal of Experimental Biology (2020) 223, jeb191262. doi:10.1242/jeb.191262

et al., 2015), should these be necessary to answer a specific question. However, such single-cell approaches are technically challenging and likely impractical in many non-model systems. As a result, one needs to choose an appropriate tissue or cell type when designing omics experiments. These decisions require some working knowledge of the physiological system in question and careful logistical considerations related to optimizing sample preparations. Even if the fingerprint of the hypothesized response is observable in whole-organism homogenates, a tissue-specific approach will increase signal and reduce noise. To keep results repeatable, the sample collection also needs to be repeatable. For both single-cell and tissue-specific studies, this might require a formal detailed protocol in the methods, or even stand-alone methods descriptions (e.g. Torson et al., 2020).

### More replicates are better, but pooling can reduce costs without losing (too much) signal

Balancing the difficulty of obtaining samples against the expected among-individual variation in a response defines the acceptable sample size for any experiment in comparative physiology (McNab, 2003). Three biological replicates are currently acceptable for comparative physiology omics experiments (Fig. 2A), but this is really an analytical minimum rather than an example of best practice for either statistical or inferential power. The cost of omics has been falling (e.g. genome sequencing; Wetterstrand, 2019), allowing larger sample sizes. This allows easier identification of outliers and, therefore, increased accuracy (Fig. 2B). When performing omics studies, one must consider the relative importance of among-individual versus among-treatment variation. In studies that focus on differences between treatments, small variations in the age or prior experience of individuals (especially if they were field-collected) is noise. By contrast, if differences among individuals are paramount (e.g. where there are precise phenotypic measurements for each individual), then the variation among individuals is signal. The best way to improve the signal-to-noise ratio and statistical power is to increase replication (Fig. 2C). However, pooling (Fig. 2D) can mask inter-individual variation, increasing the signal-to-noise ratio when comparing treatments at the cost of decreased statistical power and the capacity to detect individual differences. Thus, if individual variation is not a primary research objective then pooling is both cost-effective and improves the capacity to detect a treatment effect.

Although omics is getting cheaper, cost is still important, especially when the cost is per sample. For example, metabolomics usually requires sequential sample processing through chromatography/mass-spectrometry instruments. In these cases, the cost of replication scales linearly with sample size. In other technologies, such as RNA-seq, analyses are parallel, and many replicates can be sequenced on a single ‘lane’ or ‘cell’ (at the expense of number of reads per sample, i.e. sequencing depth). In

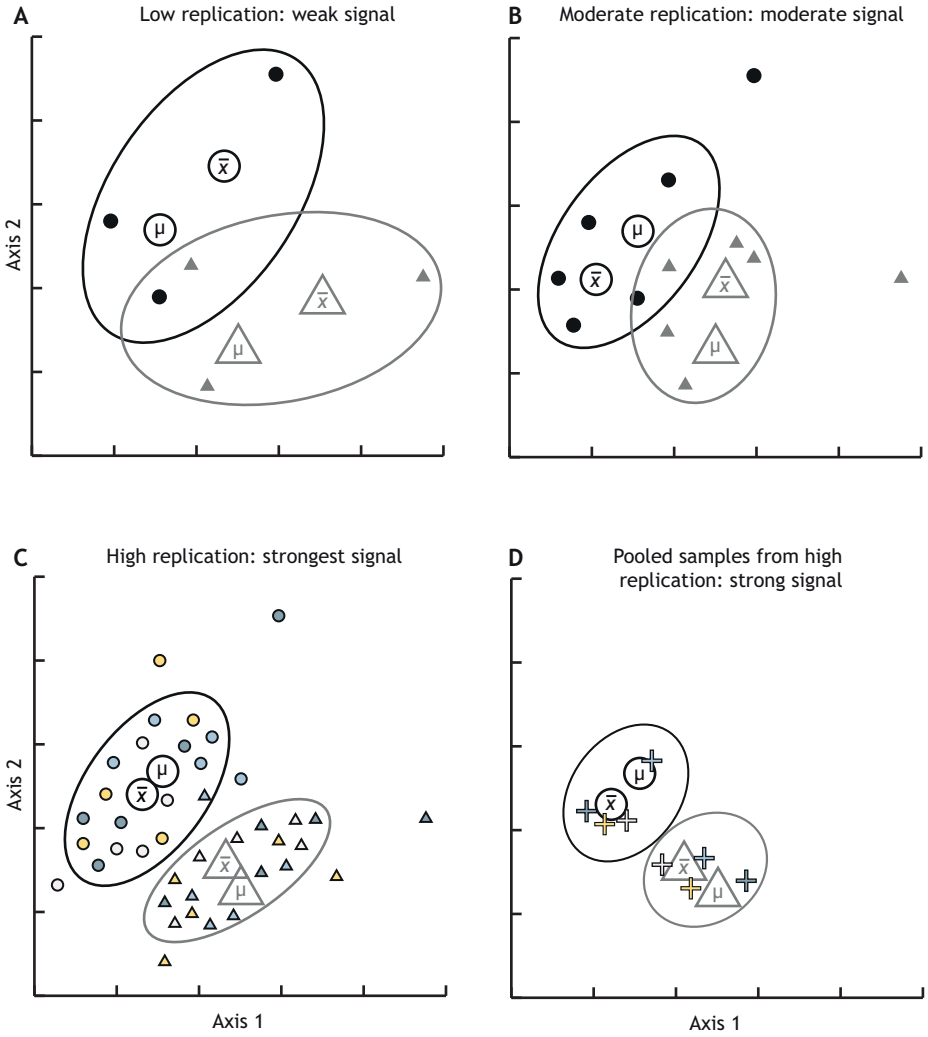


**Fig. 2. Hypothetical principal component analyses of omics profiles.** Increasing replication and pooling can improve signal to noise in omics profiles. (A) Low biological replication (e.g. $n$=3) will support most omics statistical analyses, but significant biological variation in the two groups (black and grey) could yield an inaccurate estimate ($\bar{x}$) of the true mean (μ). (B) With moderate replication, estimates are more accurate, and outliers can be identified and addressed. (C) High replication gives the most accurate estimate of mean and variance to detect the true differences among groups. (D) Pooling individuals from high-replication sampling (represented by shapes of similar color in C is an effective cost-saving strategy that gives an accurate estimate of the signal but masks underlying individual variation. Average values of each pool in C are represented by colored + symbols.

Journal of Experimental Biology (2020) 223, jeb191262. doi:10.1242/jeb.191262

this case, there is still a per-replicate cost for sample preparation, but adding more samples to a sequencing run bears no additional cost: the ability to detect differentially expressed genes plateaus after about 10 million reads (Liu et al., 2014), but, of course, this plateau increases if one is interested in isoform-level resolution. Thus, more biological replicates at a cost of sequencing depth returns the greatest statistical power per unit cost. If there is a per-replicate cost to acquiring omics data, then pooling individuals within biological replicates can increase the signal-to-noise ratio with a small sample size, as discussed above.

## Simple experiments are still best

Analyzing omics output from complex experiments is not trivial. Most omics analysis pipelines such as the Tuxedo Suite for RNA-seq (Trapnell et al., 2013, 2012) and MetaboAnalyst for metabolomics (Chong et al., 2019) are optimized for pairwise comparisons. Experimental designs comparing two treatments or multiple treatments back to a common control (Fig. 3A) are easily analyzed under this framework. Pairwise tools are also appropriate for more complex designs that compare controls and treatments among tissues or populations (Fig. 3B; e.g. Des Marteaux et al., 2017). However, a typical 'simple' comparative physiology experiment might be a 2×2 design examining the interaction of two treatments (Fig. 3C). Ordinarily, we would analyze such data using some version of a two-way ANOVA. Pairwise comparisons allow the dissection of main effects, but not the interactions (which may, in fact, be the purpose of the experiment). Such designs can be analyzed (for RNA-seq, in any case) using generalized linear models, which are more complex than the turn-key analyses provided in many packages (McCarthy et al., 2012). Pairwise analyses are also ill-suited to time series data (Fig. 3D), since dividing a time series analysis into discrete pair-wise comparisons fails to account for temporal autocorrelation (Spies et al., 2019). However, there are tools for clustering temporal patterns of gene expression (e.g. maSigPro; Nueda et al., 2014) that can identify sets of genes that respond similarly over time (e.g. Toxopeus et al., 2018). The tools to address these complex designs are still in the early stages of their development and are challenging for those without great familiarity with omics analyses. Thus, there are currently no well-established standards and the most-common

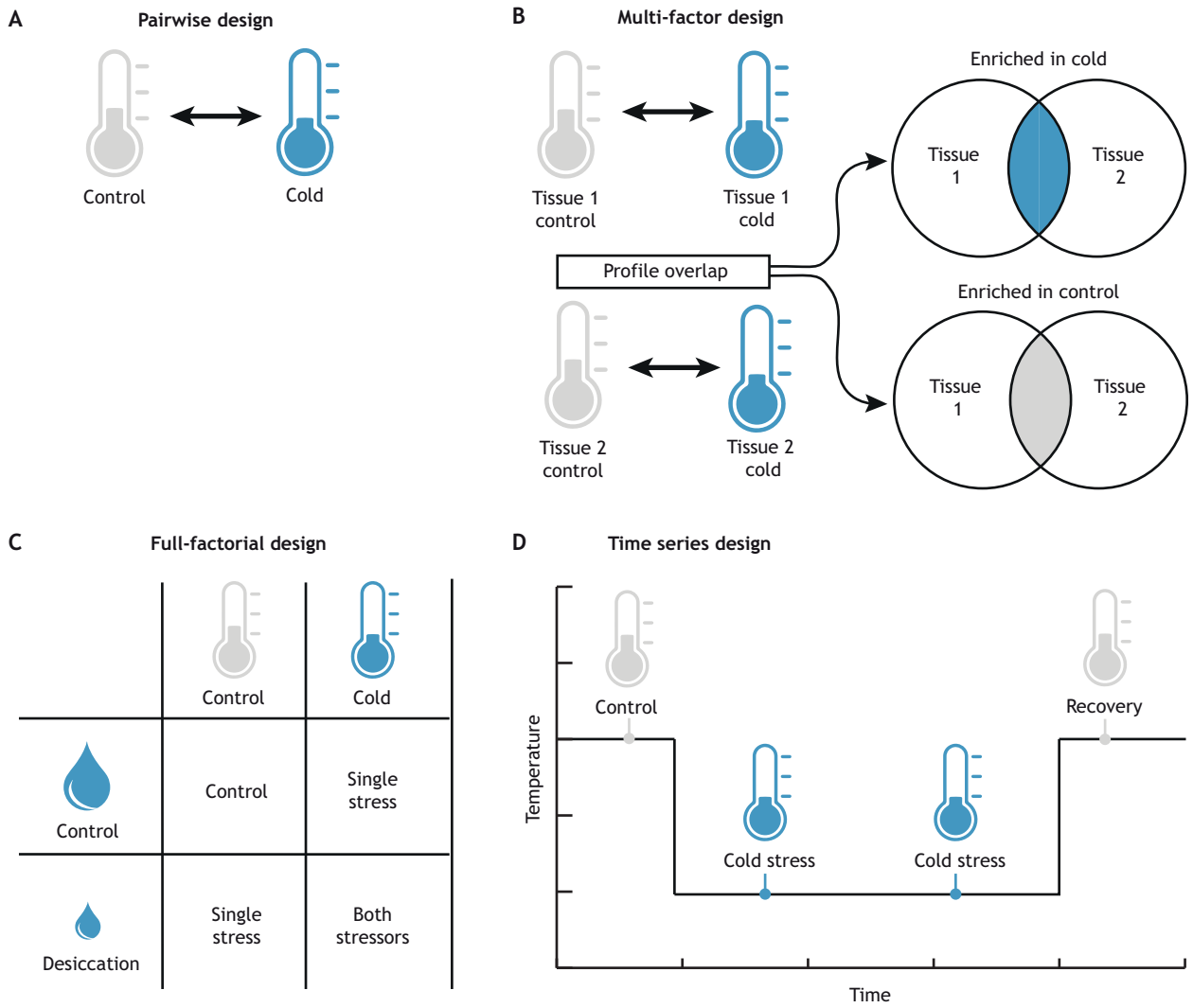


**Fig. 3. Designing experiments with omics in mind.** (A) Experiments consisting of simple pairwise comparisons are easily analyzed using common analytical pipelines. (B) Multi-factor experiments, such as those measuring responses to cold stress in multiple tissues can be analyzed in a pairwise fashion (black arrows) and interactions between tissues can be addressed via overlap of differential expression profiles between tissues (represented as Venn diagrams). Analyzing (C) full-factorial designs and (D) time-series data is more challenging (see main text for further discussion).

methods for complex analyses are often out-performed by pairwise differential expression methods (summarized by Spies et al., 2019). Nevertheless, the proverbial statistician's cry still holds true in the face of omics: 'plan your data analysis before you do the experiment!'

## Interpreting omics: moving beyond the fishing expedition

A key advantage of omics is that when few or no hypotheses exist, omics should allow the generation of testable hypotheses by identifying candidate molecules and pathways underlying physiological processes. However, this approach is often derided as a 'fishing expedition' (Ning and Lo, 2010). It may be argued that such omics approaches fall into the 'hypothesis formulation' stage of the classic scientific method (Doughty and Kerkhoven, 2020). Nevertheless, because large amounts of omics data can be acquired cheaply and easily, and the analysis can be challenging, an unkind criticism might be that the quality of the hypotheses generated has not generally been high. If this is indeed the case, how do we improve the quality of omics as a hypothesis-generating tool?

Most omics analyses generate long lists of candidate molecules as their output. In some cases, these candidates are themselves direct evidence of a molecule's involvement in a physiological process. For example, high concentrations of hemolymph glycerol in emerald ash borer (*Agrilus planipennis*) prepupae during winter strongly suggest that this metabolite is a cryoprotectant (Crosthwaite et al., 2011). Differential expression of genes associated with a transcriptionally regulated process (e.g. caspase-8 in apoptosis; Li et al., 2014) could directly imply that pathway's involvement in a phenotype. Alas, we are rarely so lucky in omics analyses. Often the evidence is indirect and requires either more-sophisticated omics-level analyses or additional experiments.

When faced with an overwhelming degree of differential expression from RNA-seq or proteomics experiments, many authors use orthology-based approaches to categorize their candidates into functional classes (using, for example, gene ontology, GO; see Glossary) or biological pathways (e.g. using the Kyoto Encyclopedia of Genes and Genomes, KEGG; see Glossary), and test for enrichment of these functions against a reference. For example, Zhang et al. (2020) evaluated differential expression of proteins using a metabolic pathway analysis to show that fatty acid degradation increased during diapause in rice water weevil, *Lissorhoptrus oryzophilus*. Because some of the tools used to categorize candidates overlap, there can be integration among omics datasets: for example, pathway outputs from metabolomics and transcriptomics can be compared (e.g. Košt'ál et al., 2016). However, there are caveats to these analyses: (1) not all genes/transcripts are annotated meaningfully, and in most datasets a significant number of genes and transcripts have no annotation or known function, which introduces bias into the processes that will (or can) be identified; (2) gene products could serve purposes that have not yet been functionally annotated (e.g. moonlighting proteins; Jeffery, 2014), which could lead to inaccurate conclusions about the involvement of a process in a phenotype; (3) gene ontology and pathway analyses rely heavily on the assumption of homologous relationships in gene-product function across taxa, which could allow novel functions or processes to be overlooked; and (4) taxonomic coverage is limited in most databases and heavily biased towards biomedical models, which means omics results from the more weird and wonderful Krogh models may miss key processes or functions (Gaudet et al., 2017).

One way to avoid the bias inherent in these functional analyses is to use clustering and gene co-expression network analyses (see Glossary) to identify associated groups of molecules or processes without *a priori* expectations about their function or pathway (Horvath, 2011; Scaria et al., 2016; Wu et al., 2003). For example, Riddell et al. (2019) used a co-expression network analysis to identify functional modules of gene expression and show that plasticity in the desiccation tolerance of a montane salamander correlates with suites of genes involved in blood vessel development. Thus, clustering and co-expression analyses can contribute directly to the generation of testable hypotheses about novel gene function, which can then be followed up with functional experiments. The implementation and interpretation of the results generated from these more advanced analytical tools requires a robust bioinformatics skill set.

## Omics in a hypothetico-deductive framework

Designing and testing elegant, efficient and discriminatory hypotheses is the epitome of experimental science (Platt, 1964). Although the emergence of omics led to claims that this moved science to a post-hypothesis world (e.g. Ning and Lo, 2010), most omics still require follow-up 'small science' hypothesis-driven studies (cf. Alberts, 2012) to draw conclusions. What, then, is the role of omics in hypothesis-driven comparative physiology, and what opportunities for testing hypotheses do omics technologies afford?

Increasing the sophistication of omics-level analyses will increase the quality of the resulting hypotheses, but if an omics study generates hypotheses, reviewers and editors will probably expect the authors to test those hypotheses. We note that the hypothesis testing need not take place at the same level of organization – indeed, to do so would be overwhelming if there are tens or hundreds of potentially interacting candidate molecules, and there are limited (or no) genetic tools for the organism in question. However, omics-derived hypotheses can be readily tested using the tools with which physiologists are comfortable (Table 1). For example, transcriptomics suggested that cold-acclimated crickets remodel their hindgut cytoskeleton (Des Marteaux et al., 2017), which was then tested using microscopy (Des Marteaux et al., 2018b). Similarly, Whitehead et al. (2012) used transcriptomics to identify aryl-hydrocarbon receptor pathways upregulated after crude oil exposure in killifish and evaluated this response with immunohistochemical measurements of gill cytochrome p450 A1 protein. Thus, in these situations, omics are contributing directly to the integrative goals of the comparative physiology program, but the omics are used to generate, rather than test, the hypotheses (Table 1).

One approach to the use of omics in hypothesis-driven comparative physiology is to use omics data to evaluate *a priori* hypotheses. For example, MacMillan et al. (2016) articulated six hypotheses associated with the mechanisms underlying insect cold acclimation, and used a combination of metabolomics and transcriptomics to support four of those hypotheses. One of the other hypotheses concerned phospholipid properties (a potential mechanism that is likely to be neither transcriptionally regulated nor identifiable using a metabolomics approach that focused on polar metabolites). The remaining hypothesis was on the regulation of apoptosis: this process is transcriptionally regulated (Franzetti et al., 2012) and should produce a metabolomic signature, so a lack of evidence in the omics datasets suggests that this hypothesis is not well supported. We refer to this approach as 'hypothesis evaluation', rather than testing, because an omics approach will rarely provide a direct and definitive test of a physiological hypothesis. To be most effective, we advocate the systematic compilation of hypotheses to

**Table 1. Examples of omics-derived hypotheses and follow-up experiments crossing biological levels of organization**

| Species | Initial omics approach | Omics-derived hypothesis | Follow-up level (technique) | Reference(s) |
|---|---|---|---|---|
| Fall field cricket (*Gryllus pennsylvanicus*) | RNA-seq | Crickets remodel the cytoskeleton during low temperature acclimation | Cellular (microscopy) | Des Marteaux et al. (2017, 2018b) |
| Cabbage looper (*Trichoplusia ni*) | RNA-seq | Dietary ion loading leads to changes in paracellular permeability in the distal ileac plexus of lepidopteran Malpighian tubules | Protein (immunohistochemistry); gene expression (qPCR); cellular (microscopy, pharmacology) | Kolosov et al. (2019a,b) |
| Alfalfa leafcutting bee (*Megachile rotundata*) | RNA-seq | Brief warm exposures elicit a beneficial oxidative stress response in chill-injured insects | Protein (enzyme activity); cellular (oxidative damage) | Torson et al. (2015, 2019) |
| Mangrove killifish (*Kryptolebias marmoratus*) | Proteomics | Structural proteins for bone mineralization are responsible for weight-responsiveness of the musculoskeletal system | Tissue (histology, bone density measurements, mechanical tests) | Turko et al. (2017) |
| Gulf killifish (*Fundulus grandis*) | RNA-seq | Toxic levels of crude oil reduce reproductive success via interference with oestrogen signaling pathways | Organismal (egg production, fertilization success) | Whitehead et al. (2012) |
| Rice water weevil (*Lissorhoptrus oryzophilus*) | Proteomics | Glycerol accumulation is driven by fatty acid breakdown during diapause | Metabolomic (glycerol and glycogen measurements); gene expression (qPCR) | Zhang et al. (2020) |
| Blue mussels (*Mytilis galloprovincialis*; *Mytilis trossulus*) | Proteomics | Sirtuin-dependent deacylation regulates heat stress response in intertidal mussels | Protein (chemical inhibition of sirtuin) | Vasquez et al. (2017, 2020) |

yield *a priori* expectations about candidate molecules and processes. This approach can then facilitate the evaluation of multiple (and possibly competing) hypotheses based on predicted omics-level responses under the strong-inference paradigm (cf. Platt, 1964). Analysis and interpretation can then be two-pronged: (1) hypothesis evaluation and (2) hypothesis generation based on the dataset. Explicit hypothesis evaluation can be applied to data analysis and interpretation after data collection, making it possible to reconcile otherwise overwhelming 'fishing expedition' data sets with a more recognizable hypothetico-deductive approach.

Can omics be used to directly test a hypothesis? The short answer to this question is 'yes', but designing and interpreting that hypothesis is more challenging. Omics methods should lend themselves well to testing hypotheses that predict patterns of responses. For example, Sinclair et al. (2013) hypothesized that insects have similar cellular responses to cold and dehydration stress (i.e. that they display cross tolerance). This hypothesis yields clear predictions of shared patterns of omics responses to the two stressors, which can be readily evaluated (Fig. 4). Similarly, Harada and Burton (2020) used a transcriptomic approach to test the

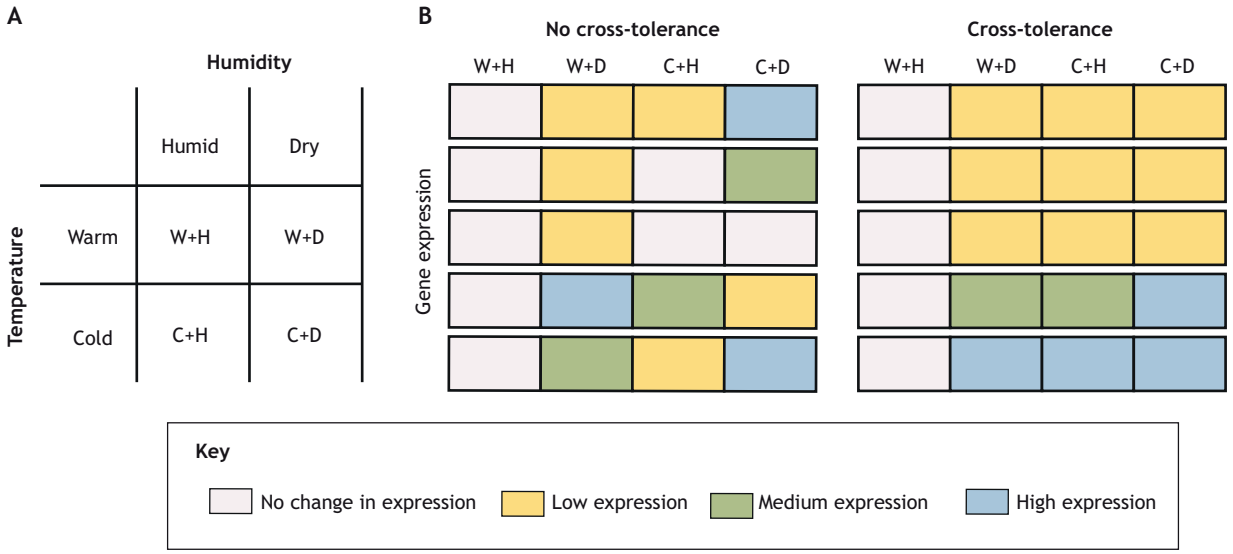


**Fig. 4. A hypothetical experimental design to investigate cross tolerance of cold and dehydration stress in insects.** This experiment uses a full-factorial design (A) to test the hypothesis that the physiological responses of an insect to cold stress and dehydration share similar regulatory mechanisms. This hypothesis is easily tested using an omics approach by evaluating the degree to which molecular profiles (e.g. RNA-seq) are similar or dissimilar among the single-stress and combined-stress exposures. (B) Low levels of overlap among, for example, gene expression profiles (which could be evaluated using heat maps) would not support the hypothesis and would indicate divergent stress response mechanisms to cold stress and dehydration. Expression levels, relative to the warm temperature, high humidity treatment, from low to high are represented as a gradient from yellow to blue, with grey representing no change in expression. A high degree of overlap in profiles would support the hypothesis that cross tolerance exists between these stressors. W, warm; C, cold; H, humid; D, dry.

hypothesis that the heat shock response in the intertidal copepod *Tigriopus californicus* is regulated by heat shock transcription factor-1 (HSF-1). After exposure to a heat shock, HSF-1 trimers bind to highly conserved sequences in the genome (heat shock elements; HSEs) to promote the expression of genes involved in the heat shock response (Vihervaara and Sistonen, 2014). Thus, this hypothesis predicts that knocking down HSF-1 prior to heat stress should result in a decreased expression of genes with HSEs in their promoter regions and should ultimately hinder the heat shock response. Knocking down HSF-1 using RNAi resulted in decreased survival after heat stress but did not alter the expression of any of the approximately 400 genes in *C. californicus* containing at least one HSE in their promoter region, not supporting the original hypothesis.

## Conclusions

Omics techniques are now readily accessible for comparative physiology investigations in non-model species, giving researchers new tools to study complex physiological problems. The physiological process under investigation will dictate the most appropriate omics method, the tissue sampling and replication approach, and the experimental design. However, analysis of omics experiments can be challenging, so begin with simple experimental designs. A strength of omics is the ability to generate new hypotheses, but these studies need not be fishing expeditions. The hypotheses generated by omics can be tested (often at another level of biological organization) using more conventional physiological tools, but omics data themselves can also be used to evaluate competing hypotheses. We recommend compiling these hypotheses *a priori* and developing clear predictions to allow them to be evaluated. With some care, omics can even be used to directly test hypotheses about physiological processes. Ultimately, we hope that this Commentary saves those who are new to omics from falling victim to their many pitfalls and inspires established comparative physiologists to integrate these tools into their research.

### Acknowledgements

We are grateful to members of the Sinclair lab and three anonymous referees for their critical feedback on an earlier draft of the paper. Thanks to Andrew Whitehead, Nick Teets, Dennis Kolosov, Richelle Tanner, Andy Turko and Eric Riddell for suggestions of studies that integrate omics into hypothesis testing.

### Competing interests

The authors declare no competing or financial interests.

### Funding

This work was supported by a Genome Canada (LSARP 10106) grant for the Biosurveillance of Alien Forest Enemies (BioSAFE) project and the National Natural Science Foundation of China (41776135).

### References

**Akashi, H. D., Diaz, A. C., Shigenobu, S., Makino, T. and Kawata, M.** (2016). Differentially expressed genes associated with adaptation to different thermal environments in three sympatric Cuban *Anolis* lizards. *Mol. Ecol.* **25**, 2273-2285. doi:10.1111/mec.13625

**Alberts, B.** (2012). The end of 'small science'? *Science* **337**, 1583. doi:10.1126/science.1230529

**Armengaud, J.** (2016). Next-generation proteomics faces new challenges in environmental biotechnology. *Curr. Opin. Biotechnol.* **38**, 174-182. doi:10.1016/j.copbio.2016.02.025

**Biederman, A. M., Kuhn, D. E., O'Brien, K. M. and Crockett, E. L.** (2019). Physical, chemical, and functional properties of neuronal membranes vary between species of antarctic notothenioids differing in thermal tolerance. *J. Comp. Physiol. B* **189**, 213-222. doi:10.1007/s00360-019-01207-x

**Calvete, J. J.** (2014). Challenges and prospects of proteomics of non-model organisms. *J. Proteomics* **105**, 1-4. doi:10.1016/j.jprot.2014.04.034

**Cazaly, E., Saad, J., Wang, W., Heckman, C., Ollikainen, M. and Tang, J.** (2019). Making sense of the epigenome using data integration approaches. *Front. Pharmacol.* **10**, 126. doi:10.3389/fphar.2019.00126

**Chong, J., Wishart, D. S. and Xia, J.** (2019). Using metaboanalyst 4.0 for comprehensive and integrative metabolomics data analysis. *Curr. Protoc. Bioinformatics* **68**, e86. doi:10.1002/cpbi.86

**Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L. and Zhang, X.** (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17**, 13. doi:10.1186/s13059-016-0881-8

**Crosthwaite, J. C., Sobek, S., Lyons, D. B., Bernards, M. A. and Sinclair, B. J.** (2011). The overwintering physiology of the emerald ash borer, *Agrilus planipennis* Fairmaire (Coleoptera: Buprestidae). *J. Insect Physiol.* **57**, 166-173. doi:10.1016/j.jinsphys.2010.11.003

**Darveau, C.-A., Hochachka, P. W., Roubik, D. W. and Suarez, R. K.** (2005). Allometric scaling of flight energetics in orchid bees: Evolution of flux capacities and flux rates. *J. Exp. Biol.* **208**, 3593-3602. doi:10.1242/jeb.01777

**Deng, Y., Li, F., Rieske, L. K., Sun, L.-L. and Sun, S.-H.** (2018). Transcriptome sequencing for identification of diapause-associated genes in fall webworm, *Hyphantria cunea* Drury. *Gene* **668**, 229-236. doi:10.1016/j.gene.2018.05.023

**Des Marteaux, L. E., McKinnon, A. H., Udaka, H., Toxopeus, J. and Sinclair, B. J.** (2017). Effects of cold-acclimation on gene expression in fall field cricket (*Gryllus pennsylvanicus*) ionoregulatory tissues. *BMC Genomics* **18**, 357. doi:10.1186/s12864-017-3711-9

**Des Marteaux, L. E., Khazraeenia, S., Yerushalmi, G. Y., Donini, A., Li, N. G. and Sinclair, B. J.** (2018a). The effect of cold acclimation on active ion transport in cricket ionoregulatory tissues. *Comp. Biochem. Physiol. A* **216**, 28-33. doi:10.1016/j.cbpa.2017.11.005

**Des Marteaux, L. E., Stinziano, J. R. and Sinclair, B. J.** (2018b). Effects of cold acclimation on rectal macromorphology, ultrastructure, and cytoskeletal stability in *Gryllus pennsylvanicus* crickets. *J. Insect Physiol.* **104**, 15-24. doi:10.1016/j.jinsphys.2017.11.004

**Des Marteaux, L. E., Hula, P. and Košťál, V.** (2019). Transcriptional analysis of insect extreme freeze tolerance. *Proc. R. Soc. B* **286**, 20192019. doi:10.1098/rspb.2019.2019

**Doughty, T. and Kerkhoven, E.** (2020). Extracting novel hypotheses and findings from RNA-seq data. *FEMS Yeast Res.* **20**, foaa007. doi:10.1093/femsyr/foaa007

**Dow, J. A. T.** (2007). Integrative physiology, functional genomics and the phenotype gap: a guide for comparative physiologists. *J. Exp. Biol.* **210**, 1632-1640. doi:10.1242/jeb.002691

**Driedzic, W. R. and Hochachka, P. W.** (1976). Control of energy-metabolism in fish white muscle. *Am. J. Physiol.* **230**, 579-582. doi:10.1152/ajplegacy.1976.230.3.579

**Duan, J., Ladd, T., Doucet, D., Cusson, M., vanFrankenhuyzen, K., Mittapalli, O., Krell, P. J. and Quan, G. X.** (2015). Transcriptome analysis of the emerald ash borer (EAB), *Agrilus planipennis*: *De novo* assembly, functional annotation and comparative analysis. *PLoS ONE* **10**, e0134824. doi:10.1371/journal.pone.0134824

**Fernández, M. J., Bozinovic, F. and Suarez, R. K.** (2011). Enzymatic flux capacities in hummingbird flight muscles: a "one size fits all" hypothesis. *Can. J. Zool.* **89**, 985-991. doi:10.1139/z11-074

**Franzetti, E., Huang, Z.-J., Shi, Y.-X., Xie, K., Deng, X.-J., Li, J.-P., Li, Q.-R., Yang, W.-Y., Zeng, W.-N. and Casartelli, M.** (2012). Autophagy precedes apoptosis during the remodeling of silkworm larval midgut. *Apoptosis* **17**, 305-324. doi:10.1007/s10495-011-0675-0

**Gaudet, P., Škunca, N., Hu, J. C. and Dessimoz, C.** (2017). Primer on the gene ontology. In *The Gene Ontology Handbook, Methods in Molecular Biology*, vol. 1446 (ed. C. Dessimoz and P. Gaudet), pp. 189-203. New York, NY: Humana Press.

**Glastad, K. M., Hunt, B. G. and Goodisman, M. A. D.** (2019). Epigenetics in insects: Genome regulation and the generation of phenotypic diversity. *Annu. Rev. Entomol.* **64**, 185-203. doi:10.1146/annurev-ento-011118-111914

**Gleason, L. U. and Burton, R. S.** (2015). RNA–seq reveals regional differences in transcriptome response to heat stress in the marine snail *Chlorostoma funebralis*. *Mol. Ecol.* **24**, 610-627. doi:10.1111/mec.13047

**Gui, S., Taning, C. N. T., Wei, D. and Smagghe, G.** (2020). First report on CRISPR/Cas9-targeted mutagenesis in the Colorado potato beetle, *Leptinotarsa decemlineata*. *J. Insect Physiol.* **121**, 104013. doi:10.1016/j.jinsphys.2020.104013

**Halberg, K. A., Terhzaz, S., Cabrero, P., Davies, S. A. and Dow, J. A. T.** (2015). Tracing the evolutionary origins of insect renal function. *Nat. Commun.* **6**, 6800. doi:10.1038/ncomms7800

**Hanschen, E. R. and Starkenburg, S. R.** (2020). The state of algal genome quality and diversity. *Algal Res.* **50**, 101968. doi:10.1016/j.algal.2020.101968

**Harada, A. E. and Burton, R. S.** (2020). Consequences of HSF knockdown on gene expression during the heat shock response in *Tigriopus californicus*. *J. Exp. Biol.* **223**, jeb208611. doi:10.1242/jeb.208611

**Hariharan, R., Hoffman, J. M., Thomas, A. S., Soltow, Q. A., Jones, D. P. and Promislow, D. E. L.** (2014). Invariance and plasticity in the *Drosophila melanogaster* metabolomic network in response to temperature. *BMC Syst. Biol.* **8**, 139. doi:10.1186/s12918-014-0139-6

**Haug, K., Cochrane, K., Nainala, V. C., Williams, M., Chang, J., Jayaseelan, K. V. and O'Donovan, C.** (2020). MetaboLights: a resource evolving in response to the needs of its scientific community. *Nucleic Acids Res.* **48**, D440-D444. doi:10.1093/nar/gkz1019

**Ho, D. H. and Burggren, W. W.** (2010). Epigenetics and transgenerational transfer: a physiological perspective. *J. Exp. Biol.* **213**, 3-16. doi:10.1242/jeb.019752

**Horvath, S.** (2011). *Weighted Network Analysis: Applications in Genomics and Systems Biology*. New York: Springer-Verlag.

**Howlett, N., Dauber, K. L., Shukla, A., Morton, B., Glendinning, J. I., Brent, E., Gleason, C., Islam, F., Izquierdo, D., Sanghavi, S. et al.** (2012). Identification of chemosensory receptor genes in *Manduca sexta* and knockdown by RNA interference. *BMC Genomics* **13**, 211. doi:10.1186/1471-2164-13-211

**Huang, X. Y., Xu, Y., Zhang, Y., Liu, J., Luo, S., Wu, J. M.** (2018). Spatzle4 gene of silkworm, *Bombyx mori*: Identification, immune response, and the effect of RNA interference on the antimicrobial peptides' expression in the integument. *Saudi J. Biol. Sci.* **25**, 1817-1825. doi:10.1016/j.sjbs.2018.07.004

**Jeffery, C. J.** (2014). An introduction to protein moonlighting. *Biochem. Soc. Trans.* **42**, 1679-1683. doi:10.1042/BST20140226

**Karahalil, B.** (2016). Overview of systems biology and omics technologies. *Curr. Med. Chem.* **23**, 4221-4230. doi:10.2174/0929867323666160926150617

**Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. and Teichmann, S. A.** (2015). The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58**, 610-620. doi:10.1016/j.molcel.2015.04.005

**Kolosov, D., Donly, C., MacMillan, H. and O'Donnell, M. J.** (2019a). Transcriptomic analysis of the Malpighian tubules of *Trichoplusia ni*: Clues to mechanisms for switching from ion secretion to ion reabsorption in the distal ileac plexus. *J. Insect Physiol.* **112**, 73-89. doi:10.1016/j.jinsphys.2018.12.005

**Kolosov, D., Jonusaite, S., Donini, A., Kelly, S. P. and O'Donnell, M. J.** (2019b). Septate junction in the distal ileac plexus of larval lepidopteran *Trichoplusia ni*: alterations in paracellular permeability during ion transport reversal. *J. Exp. Biol.* **222**, jeb204750. doi:10.1242/jeb.204750

**Košťál, V., Korbelová, J., Stetina, T., Poupardin, R., Colinet, H., Zahradníčková, H., Opekarová, I., Moos, M. and Šimek, P.** (2016). Physiological basis for low-temperature survival and storage of quiescent larvae of the fruit fly *Drosophila melanogaster*. *Sci. Rep.* **6**, 32346. doi:10.1038/srep32346

**Kovach, R. P., Muhlfeld, C. C., Boyer, M. C., Lowe, W. H., Allendorf, F. W. and Luikart, G.** (2015). Dispersal and selection mediate hybridization between a native and invasive species. *Proc. R. Soc. B* **282**, 20142454. doi:10.1098/rspb.2014.2454

**Li, M., Pang, Z., Xiao, W., Liu, X., Zhang, Y., Yu, D., Yang, M., Yang, Y., Hu, J. and Luo, K.** (2014). A transcriptome analysis suggests apoptosis-related signaling pathways in hemocytes of *Spodoptera litura* after parasitization by *Microplitis bicoloratus*. *PLoS ONE* **9**, e110967. doi:10.1371/journal.pone.0110967

**Liu, Y. W., Zhou, J. and White, K. P.** (2014). RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics* **30**, 301-304. doi:10.1093/bioinformatics/btt688

**MacMillan, H. A., Knee, J. M., Dennis, A. B., Udaka, H., Marshall, K. E., Merritt, T. J. S. and Sinclair, B. J.** (2016). Cold acclimation wholly reorganizes the *Drosophila melanogaster* transcriptome and metabolome. *Sci. Rep.* **6**, 28999. doi:10.1038/srep28999

**Madr, A., Glatz, Z. and Crha, I.** (2017). 'Omics' techniques: Genomics, transcriptomics, proteomics and metabolomics in embryo developmental capacity assessment. *Chemicke Listy* **111**, 551-558.

**Maori, E., Garbian, Y., Kunik, V., Mozes-Koch, R., Malka, O., Kalev, H., Sabath, N., Sela, I. and Shafir, S.** (2019). A transmissible RNA pathway in honey bees. *Cell Reports* **27**, 1949-1959.e6. doi:10.1016/j.celrep.2019.04.073

**Mardis, E., McPherson, J., Martienssen, R., Wilson, R. K. and McCombie, W. R.** (2002). What is finished, and why does it matter. *Genome Res.* **12**, 669-671. doi:10.1101/gr.032102

**Matthews, B. J. and Vosshall, L. B.** (2020). How to turn an organism into a model organism in 10 'easy' steps. *J. Exp. Biol.* **223**, jeb218198. doi:10.1242/jeb.218198

**McCarthy, D. J., Chen, Y. and Smyth, G. K.** (2012). Differential expression analysis of multifactor RNA-seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288-4297. doi:10.1093/nar/gks042

**McNab, B.** (2003). Sample size and the estimation of physiological parameters in the field. *Funct. Ecol.* **17**, 82-86. doi:10.1046/j.1365-2435.2003.00706.x

**Meyer, E., Aglyamova, G. and Matz, M. V.** (2011). Profiling gene expression responses of coral larvae (*Acropora millepora*) to elevated temperature and settlement inducers using a novel RNA–seq procedure. *Mol. Ecol.* **20**, 3599-3616. doi:10.1111/j.1365-294X.2011.05205.x

**Milo, R. and Phillips, R.** (2015). *Cell Biology by the Numbers*. Garland Science.

**Misener, S. R., Chen, C.-P. and Walker, V. K.** (2001). Cold tolerance and proline metabolic gene expression in *Drosophila melanogaster*. *J. Insect Physiol.* **47**, 393-400. doi:10.1016/S0022-1910(00)00141-4

**Mock, T., Otillar, R. P., Strauss, J., McMullan, M., Paajanen, P., Schmutz, J., Salamov, A., Sanges, R., Toseland, A., Ward, B. J. et al.** (2017). Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature* **541**, 536-540. doi:10.1038/nature20803

**Modahl, C. M., Frietze, S. and Mackessy, S. P.** (2018). Transcriptome-facilitated proteomic characterization of rear-fanged snake venoms reveal abundant metalloproteinases with enhanced activity. *J. Proteomics* **187**, 223-234. doi:10.1016/j.jprot.2018.08.004

**Moriano-Gutierrez, S., Koch, E. J., Bussan, H., Romano, K., Belcaid, M., Rey, F. E., Ruby, E. G. and McFall-Ngai, M. J.** (2019). Critical symbiont signals drive both local and systemic changes in diel and developmental host gene expression. *Proc. Natl Acad. Sci. USA* **116**, 7990-7999. doi:10.1073/pnas.1819897116

**Mykles, D. L., Ghalambor, C. K., Stillman, J. H. and Tomanek, L.** (2010). Grand challenges in comparative physiology: Integration across disciplines and across levels of biological organization. *Integr. Comp. Biol.* **50**, 6-16. doi:10.1093/icb/icq015

**Nation, J. L.** (2015). *Insect Physiology and Biochemistry*, 3rd edn. Boca Raton, FL: CRC Press.

**Ning, M. M. and Lo, E. H.** (2010). Opportunities and challenges in omics. *Transl. Stroke Res.* **1**, 233-237. doi:10.1007/s12975-010-0048-y

**Nueda, M. J., Tarazona, S. and Conesa, A.** (2014). Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. *Bioinformatics* **30**, 2598-2602.

**Pinu, F. R., Goldansaz, S. A. and Jaine, J.** (2019). Translational metabolomics: current challenges and future opportunities. *Metabolites* **9**, 108. doi:10.3390/metabo9060108

**Platt, J. R.** (1964). Strong inference - certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science* **146**, 347-353. doi:10.1126/science.146.3642.347

**Poelchau, M. F., Reynolds, J. A., Denlinger, D. L., Elsik, C. G. and Armbruster, P. A.** (2011). A *de novo* transcriptome of the Asian tiger mosquito, *Aedes albopictus*, to identify candidate transcripts for diapause preparation. *BMC Genomics* **12**, 619. doi:10.1186/1471-2164-12-619

**Poupardin, R., Schottner, K., Korbelova, J., Provaznik, J., Dolezel, D., Pavlinic, D., Benes, V. and Košťál, V.** (2015). Early transcriptional events linked to induction of diapause revealed by RNA-seq in larvae of drosophilid fly, *Chymomyza costata*. *BMC Genomics* **16**, 720. doi:10.1186/s12864-015-1907-4

**Raghavachari, N.** (2011). Overview of omics. In *OMICS: Biomedical Perspectives and Applications* (ed. D. Barh, K. Blum and M. A. Madigan), pp. 622. Boca Raton: CRC Press.

**Ren, C. C., Sylvia, K. E., Munley, K. M., Deyoe, J. E., Henderson, S. G., Vu, M. P. and Demas, G. E.** (2020). Photoperiod modulates the gut microbiome and aggressive behavior in Siberian hamsters. *J. Exp. Biol.* **223**, jeb212548. doi:10.1242/jeb.212548

**Riddell, E. A., Roback, E. Y., Wells, C. E., Zamudio, K. R. and Sears, M. W.** (2019). Thermal cues drive plasticity of desiccation resistance in montane salamanders with implications for climate change. *Nat. Commun.* **10**, 4091. doi:10.1038/s41467-019-11990-4

**Robert, J. A., Bonnett, T., Pitt, C., Spooner, L. J., Fraser, J., Yuen, M. M. S., Keeling, C. I., Bohlmann, J. and Huber, D. P. W.** (2016). Gene expression analysis of overwintering mountain pine beetle larvae suggests multiple systems involved in overwintering stress, cold hardiness, and preparation for spring development. *PeerJ* **4**, e2109. doi:10.7717/peerj.2109

**Sangiovanni, M., Granata, I., Thind, A. S. and Guarracino, M. R.** (2019). From trash to treasure: detecting unexpected contamination in unmapped NGS data. *BMC Bioinformatics* **20**, 168. doi:10.1186/s12859-019-2684-x

**Scaria, T., Stephen, G. and Mathew, J.** (2016). Gene expression data analysis using fuzzy c-means clustering technique. *Intl. J. Computer Appl.* **135**, 33-36. doi:10.5120/ijca2016908470

**Schmieder, R. and Edwards, R.** (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS ONE* **6**, e17288. doi:10.1371/journal.pone.0017288

**Seppey, M., Manni, M. and Zdobnov, E. M.** (2019). BUSCO: assessing genome assembly and annotation completeness. In *Gene Prediction*, pp. 227-245: Springer.

**Shamir, M., Bar-On, Y., Phillips, R. and Milo, R.** (2016). SnapShot: timescales in cell biology. *Cell* **164**, 1302-1302.e1. doi:10.1016/j.cell.2016.02.058

**Sinclair, B. J., Ferguson, L. V., Salehipour-shirazi, G. and MacMillan, H. A.** (2013). Cross-tolerance and cross-talk in the cold: Relating low temperatures to desiccation and immune stress in insects. *Integr. Comp. Biol.* **53**, 545-556. doi:10.1093/icb/ict004

**Somero, G. N., Lockwood, B. L. and Tomanek, L.** (2017). *Biochemical Adaptation: Response to Environmental Challenges, from Life's Origins to the Anthropocene*. Sunderland, MA: Sinauer.

**Spies, D., Renz, P. F., Beyer, T. A. and Ciaudo, C.** (2019). Comparative analysis of differential gene expression tools for RNA sequencing time course data. *Brief. Bioinform.* **20**, 288-298. doi:10.1093/bib/bbx115

**Suarez, R. K. and Moyes, C. D.** (2012). Metabolism in the age of 'omes'. *J. Exp. Biol.* **215**, 2351-2357. doi:10.1242/jeb.059725

**Teets, N. M. and Denlinger, D. L.** (2016). Quantitative phosphoproteomics reveals signaling mechanisms associated with rapid cold hardening in a chill-tolerant fly. *J. Proteome Res.* **15**, 2855-2862. doi:10.1021/acs.jproteome.6b00427

**Teets, N. M., Gantz, J. D. and Kawarasaki, Y.** (2020). Rapid cold hardening: ecological relevance, physiological mechanisms and new perspectives. *J. Exp. Biol.* **223**, jeb203448. doi:10.1242/jeb.203448

**Teets, N. M., Peyton, J. T., Ragland, G. J., Colinet, H., Renault, D., Hahn, D. A. and Denlinger, D. L.** (2012). Combined transcriptomic and metabolomic approach uncovers molecular mechanisms of cold tolerance in a temperate flesh fly. *Physiol. Genomics* **44**, 764-777. doi:10.1152/physiolgenomics.00042.2012

**Torson, A. S., Yocum, G. D., Rinehart, J. P., Kemp, W. P. and Bowsher, J. H.** (2015). Transcriptional responses to fluctuating thermal regimes underpinning

differences in survival in the solitary bee *Megachile rotundata*. *J. Exp. Biol.* **218**, 1060-1068. doi:10.1242/jeb.113829

**Torson, A. S., Yocum, G. D., Rinehart, J. P., Nash, S. A., Kvidera, K. M. and Bowsher, J. H.** (2017). Physiological responses to fluctuating temperatures are characterized by distinct transcriptional profiles in a solitary bee. *J. Exp. Biol.* **220**, 3372-3380. doi:10.1242/jeb.156695

**Torson, A. S., Yocum, G. D., Rinehart, J. P., Nash, S. A. and Bowsher, J. H.** (2019). Fluctuating thermal regimes prevent chill injury but do not change patterns of oxidative stress in the alfalfa leafcutting bee, *Megachile rotundata*. *J. Insect Physiol.* **118**, 103935. doi:10.1016/j.jinsphys.2019.103935

**Torson, A. S., des Marteaux, L. E., Bowman, S., Lei Zhang, M., Ong, K., Doucet, D., Sinclair, B. J. and Roe, A. D.** (2020). Dissection of *Anoplophora glabripennis* (Coleoptera: Cerambycidae) larval tissues for physiological and molecular studies. *Can. Entomol.* **152**, 399-409. doi:10.4039/tce.2020.22

**Toxopeus, J., Des Marteaux, L. E. and Sinclair, B. J.** (2018). How crickets become freeze tolerant: the transcriptomic underpinnings of acclimation in *Gryllus veletis*. *Compar. Biochem. Physiol. D* **29**, 55-66. doi:10.1016/j.cbd.2018.10.007

**Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L. and Pachter, L.** (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

**Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L. and Pachter, L.** (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46-53. doi:10.1038/nbt.2450

**Turko, A. J., Kültz, D., Fudge, D., Croll, R. P., Smith, F. M., Stoyek, M. R. and Wright, P. A.** (2017). Skeletal stiffening in an amphibious fish out of water is a response to increased body weight. *J. Exp. Biol.* **220**, 3621-3631. doi:10.1242/jeb.161638

**Vasquez, M. C., Beam, M., Blackwell, S., Zuzow, M. J. and Tomanek, L.** (2017). Sirtuins regulate proteomic responses near thermal tolerance limits in the blue mussels Mytilus galloprovincialis and Mytilus trossulus. *J. Exp. Biol.* **220**, 4515-4534. doi:10.1242/jeb.160325

**Vasquez, M. C., Martinez, D. A. and Tomanek, L.** (2020). Multiple stressor responses are regulated by sirtuins in *Mytilus* congeners. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **246**, 110719. doi:10.1016/j.cbpa.2020.110719

**Vihervaara, A. and Sistonen, L.** (2014). HSF1 at a glance. *J. Cell Sci.* **127**, 261-266. doi:10.1242/jcs.132605

**Wetterstrand, K.** (2019). DNA sequencing costs: data from the NHGRI large-scale genome sequencing program. www.genome.gov/sequencingcostsdata.

**Whitehead, A., Dubansky, B., Bodinier, C., Garcia, T. I., Miles, S., Pilley, C., Raghunathan, V., Roach, J. L., Walker, N. and Walter, R. B.** (2012). Genomic and physiological footprint of the Deepwater Horizon oil spill on resident marsh fishes. *Proc. Natl Acad. Sci. USA* **109**, 20298-20302. doi:10.1073/pnas.1109545108

**Windisch, H. S., Frickenhaus, S., John, U., Knust, R., Pörtner, H. O. and Lucassen, M.** (2014). Stress response or beneficial temperature acclimation: transcriptomic signatures in Antarctic fish (*Pachycara brachycephalum*). *Mol. Ecol.* **23**, 3469-3482. doi:10.1111/mec.12822

**Wu, F.-X., Zhang, W.-J. and Kusalik, A. J.** (2003). A genetic K-means clustering algorithm applied to gene expression data. In Conference of the Canadian Society for Computational Studies of Intelligence, pp. 520-526: Springer.

**Yang, L., Cao, Y., Zhao, J., Fang, Y., Liu, N. and Zhang, Y.** (2019). Multidimensional proteomics identifies declines in protein homeostasis and mitochondria as early signals for normal aging and age-associated disease in *Drosophila*. *Mol. Cell. Proteomics* **18**, 2078-2088. doi:10.1074/mcp.RA119.001621

**Zhang, J., Marshall, K. E., Westwood, J. T., Clark, M. S. and Sinclair, B. J.** (2011). Divergent transcriptomic responses to repeated and single cold exposures in *Drosophila melanogaster*. *J. Exp. Biol.* **214**, 4021-4029. doi:10.1242/jeb.059535

**Zhang, Y., Sun, J., Mu, H., Li, J., Zhang, Y., Xu, F., Xiang, Z., Qian, P.-Y., Qiu, J.-W. and Yu, Z.** (2015). Proteomic basis of stress responses in the gills of the pacific oyster *Crassostrea gigas*. *J. Proteome Res.* **14**, 304-317. doi:10.1021/pr500940s

**Zhang, X., Yang, S., Zhang, X., Wang, S., Zhang, J. and Xi, J.** (2020). TMT-Based quantitative proteomic profiling of overwintering *Lissorhoptrus oryzophilus*. *Front. Physiol.* **10**, 1623. doi:10.3389/fphys.2019.01623