

COMMENTARY

Best practices for building and curating databases for comparative analyses

Lisa E. Schwanz^{1,*}, Alex Gunderson², Maider Iglesias-Carrasco³, Michele A. Johnson⁴, Jacinta D. Kong⁵, Julia Riley⁶ and Nicholas C. Wu⁷

ABSTRACT

Comparative analyses have a long history of macro-ecological and -evolutionary approaches to understand structure, function, mechanism and constraint. As the pace of science accelerates, there is ever-increasing access to diverse types of data and open access databases that are enabling and inspiring new research. Whether conducting a species-level trait-based analysis or a formal meta-analysis of study effect sizes, comparative approaches share a common reliance on reliable, carefully curated databases. Unlike many scientific endeavors, building a database is a process that many researchers undertake infrequently and in which we are not formally trained. This Commentary provides an introduction to building databases for comparative analyses and highlights challenges and solutions that the authors of this Commentary have faced in their own experiences. We focus on four major tips: (1) carefully strategizing the literature search; (2) structuring databases for multiple use; (3) establishing version control within (and beyond) your study; and (4) the importance of making databases accessible. We highlight how one's approach to these tasks often depends on the goal of the study and the nature of the data. Finally, we assert that the curation of single-question databases has several disadvantages: it limits the possibility of using databases for multiple purposes and decreases efficiency due to independent researchers repeatedly sifting through large volumes of raw information. We argue that curating databases that are broader than one research question can provide a large return on investment, and that research fields could increase efficiency if community curation of databases was established.

KEY WORDS: Biomechanics, Morphology, Open science, Phylogenetics, Physiology, Systematic review

Introduction

Comparative studies have long been inspired by scientific questions seeking patterns or insights that cannot be answered with a single species (Schmidt-Nielsen, 1972; Schmidt-Nielsen, 1975; Somero, 2000; Seebacher et al., 2015; Geange et al., 2021). These approaches may address common mechanistic processes, reveal


constraints on traits, highlight trade-offs in resource allocation and functional design, and generate new hypotheses (e.g. Brown et al., 2004; Vogel, 2008; Pörtner et al., 2017). As such, they can provide major advances in biology, particularly in the fields of physiology and biomechanics. Comparative studies typically rely on high-quality data sourced from many independent empirical studies (Davidson et al., 2011; Muñoz and Price, 2019). Compiling these data into a database can be an arduous process with many pitfalls. However, careful consideration of the challenges and trade-offs can lead to a useful and effective database with enduring benefits to the research field (Whitlock, 2011; White et al., 2013).

Many decisions in database curation depend on the form of comparative research being undertaken (Fig. 1). First, analyses of comparative trait data (hereafter 'trait-based analyses') examine specific traits collected at a species or population level, typically aimed at questions of macro-ecological or -evolutionary interest (e.g. physiological scaling relationships; Francis et al., 2018; White et al., 2019). Second, meta-analyses estimate the overall strength of evidence for a particular hypothesized effect by examining associations within many independent studies that have each addressed the hypothesis (e.g. effect of the environment on physiology and phenotype; Noble et al., 2018a; Gunderson and Stillman, 2015; Iglesias-Carrasco et al., 2020; Wu and Seebacher, 2020). Finally, qualitative comparative reviews of a research question provide synthesis of a topic and data, but do not present analyses of data (e.g. Vitousek et al., 2018a, Bodensteiner et al., 2021). The approach used influences whether the most relevant data are observational or experimental, and whether databases must be compiled *de novo* from primary studies or involve extracting data from taxonomic compendia or other accessible databases (Fig. 1). Importantly, the form of comparative study has profound implications for database-building that begin immediately during the searching and filtering process, through decisions regarding the structure of a database, and the duration of curation.

Despite the eccentricity of each database, all databases share some major challenges that create trade-offs for researchers as well as the research field. We highlight three over-arching and inter-related challenges that span multiple stages of database curation and that will motivate our tips. (1) Curating a database requires large effort. The time commitment required usually demands that a team of researchers collaborate on the project. (2) Much effort goes into discarding information. Resources need to be screened for data that are relevant and meet quality thresholds. Thus, only a small proportion of information screened will be retained for the database. Over time, multiple researchers repeatedly screen the same information leading to repeated effort within the field. (3) A database may only have immediate utility for a single study. All databases should have certain qualities about them: transparency, (re)usability and reproducibility (Borries et al., 2016; Wilkinson et al., 2016). However, trade-offs between effort and utility typically

¹Evolution and Ecology Research Centre, and the School of Biological, Earth, and Environmental Sciences, University of New South Wales, Sydney, NSW 2035, Australia. ²School of Science and Engineering, Tulane University, New Orleans, LA 70118, USA. ³Ecology and Evolution of Sexual Interactions group, Doñana Biological Station-CSIC, Sevilla 41001, Spain. ⁴Department of Biology, Trinity University, San Antonio, TX 78212, USA. ⁵School of Natural Sciences, Trinity College Dublin, Dublin 2, Ireland. ⁶Department of Biology, Mount Allison University, Sackville, New Brunswick, E4L 1E4, Canada. ⁷School of Life and Environmental Sciences, The University of Sydney, NSW 2006, Australia.

*Author for correspondence (l.schwanz@unsw.edu.au)

 L.E.S., 0000-0001-5864-7112; A.G., 0000-0002-0120-4246; M.I.-C., 0000-0003-0349-7967; M.A.J., 0000-0002-9278-5639; J.D.K., 0000-0002-1085-8612; J.R., 0000-0001-7691-6910; N.C.W., 0000-0002-7130-1279

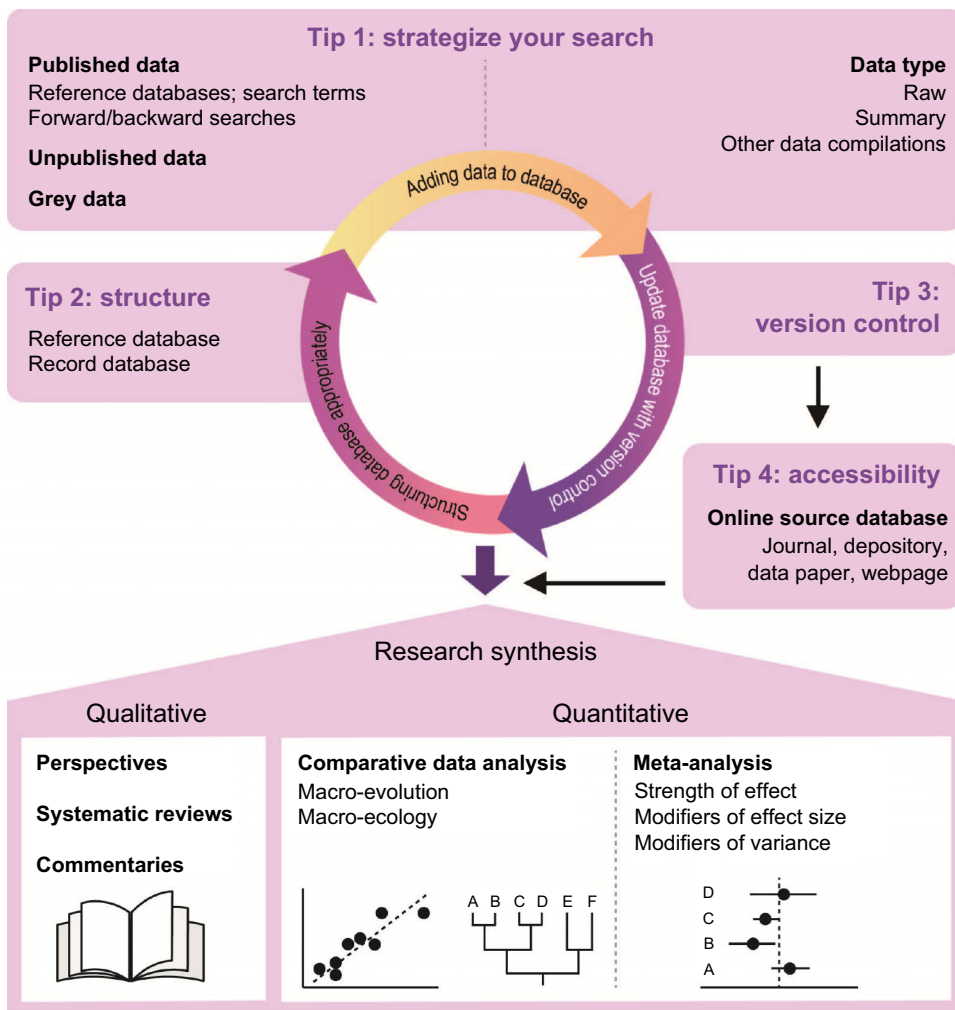


Fig. 1. Three types of comparative research syntheses and a relational schematic of the database curation required. We provide tips for initial search strategies, decisions regarding database structure, and the necessity of version control and curating accessible databases. The type of synthesis, particularly when quantitative, impacts strategies of searching and structuring databases, and creates trade-offs between multi-use and curation effort.

lead to single-use structures in databases as a matter of efficiency for a specific research question.

Most researchers undertaking comparative analyses, including ourselves, are biologists with little training in database curation. The effort and challenges of compiling a quantitative database may partly explain why qualitative reviews (literature and systematic reviews) remain vastly more common than quantitative reviews across physiology and biomechanics journals (Fig. 2A). Meta-analyses, in particular, remain under-utilized in comparative physiology and biomechanics. Although meta-analysis studies appear to be more commonly published in broader-audience journals rather than discipline journals, they remain modest in number, especially if other types of reviews are similarly enriched in the broad-audience journals [Fig. 2A; Noble et al., 2022 (this issue)]. In this Commentary, we aim to lower the barrier to all types of quantitative reviews by providing tips for building a database that are based on our experiences and perspectives (Fig. 1). By navigating the trade-offs associated with database creation, researchers can efficiently curate a database that is effective for the current research question, supports immediate additional studies, and inspires new inquiry by other researchers.

Tip 1: strategize your search

Comparative analyses often begin with exploratory perusals through known published papers or datasets. This exploration phase can help refine the research question(s), optimize search terms, and

strategize data extraction (Stewart et al., 2013). However, it is best to step back after exploration and develop a search strategy before beginning the formal search (Côté et al., 2013; Foo et al., 2021 preprint). This refinement is of critical importance to consider how best to match your search with your identified inclusion criteria and the analytical approach. The quality of comparative science can be impacted early in the searching stage.

Comparative analyses aim to include all relevant, quality data, and thus begin with a methodical search. By this stage, the search terms, exclusion and inclusion criteria, the type of data (see Fig. 1) and the key variables to be extracted from the original sources should be clearly defined (Côté et al., 2013; Forero et al., 2019). To reduce the chances of overlooking relevant data, it is usually recommended to use several bibliographical databases, such as Google Scholar, Web of Science, PubMed or Scopus (Falagas et al., 2008; Forero et al., 2019). In addition, it is advisable to search forwards (for papers that cite the original study) and backwards (for previous papers that the original study cited) on influential reviews on the topic. It is also useful to target unpublished data and grey literature – as far as the quality of the data remain similar to those in published studies – in an effort to make the dataset comprehensive (Côté et al., 2013).

The research question and type of comparative analysis will have specific demands on the search and data screening protocols. For example, trait-based analyses often focus on broad questions, so they frequently include data compiled from other reviews, datasets

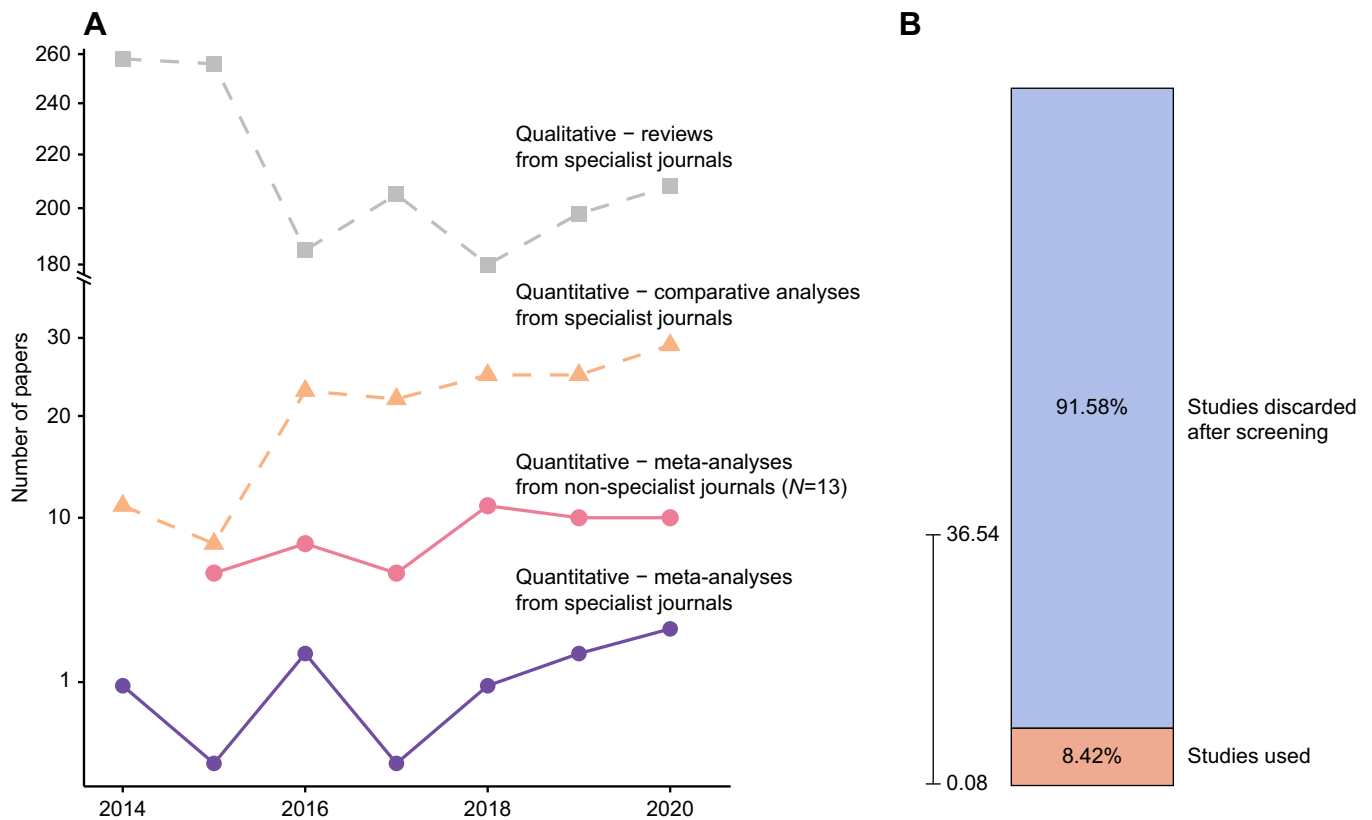


Fig. 2. Recent abundance and search effort associated with quantitative syntheses. (A) Annual frequency of qualitative reviews, trait-based analyses, and meta-analyses between January 2014 and July 2021 found in a survey of all issues within six major journals focused on physiology and biomechanics (*Conservation Physiology*, *Integrative and Comparative Biology*, *Journal of Comparative Physiology A*, *Journal of Comparative Physiology B*, *Journal of Experimental Biology* and *Journal of Experimental Botany*). Description of the data collected is provided on Github (<https://github.com/nicholaswunz/JEB-commentary-database>). For comparison, the number of meta-analyses in the field published in 13 non-specialist journals found via systematic search in Noble et al. (2022; this issue) is also plotted (continuous pink line). (B) Based on 29 meta-analyses [out of 62 from Noble et al. (2022; this issue)] that clearly reported sample sizes, the percentage of primary studies included (orange) and discarded (blue) after the first (title and abstract) and second (full-text) screenings. The black line represents the minimum and maximum of the percentage of studies included across all 29 meta-analyses. For more information on the systematic search of meta-analyses (e.g. search terms, inclusion criteria, PRISMA), see Noble et al. (2022; this issue). Data were obtained with permission from D.W.A. Noble.

or taxonomic compendia (particularly data on body size, taxonomy, life-history traits), as well as target trait data (e.g. hormone levels, range size, bite force). In contrast, meta-analyses typically have a focused question, for which the data can only be sourced from a limited number of primary studies that have examined the focal question. Because of this difference, meta-analyses and trait-based analyses often differ in the rigidity of the data search and collection. While trait-based analyses can be flexible and opportunistic in data collection, meta-analyses usually require rigid search and screening protocols to avoid bias, as well as a rigorous assessment of the data quality (e.g. sample size) and uncertainty.

In meta-analysis, the screening process has two steps: first, an initial screening of the title and abstract of the list of studies that matched the searching criteria; and then, a second full-text screening of the reduced list of the potential studies with available data, followed by data extraction (Côté et al., 2013). Fortunately, there are several tools that facilitate this screening process, including Rayyan (Ouzzani et al., 2017), Abstrackr (Wallace et al., 2012), Covidence (www.covidence.org) or the R package *revtools* (Westgate, 2019), which usually provide a more visual and summarized view of the studies to explore. There are also tools to help with data extraction such as the R package *metaDigitise* (Pick et al., 2019) or the program DataThief (www.datathief.org). These steps have to follow the best practices for transparency and repeatability in a systematic

review (Forero et al., 2019; O’Dea et al., 2021a; Salameh et al., 2020), such as providing a PRISMA diagram (Moher et al., 2009) alongside formal, clearly explained search strategies and terms, and clear exclusion/inclusion criteria of studies.

Data heterogeneity is an important, and sometimes unanticipated, factor to consider during a search. In particular, data quality thresholds must be established that include the types of sources used and whether or not to rely on data with varying accuracy (see Borries et al., 2016; Gerstner et al., 2017). In addition, a database requires complete descriptions of the variables, units and detailed definitions about how variables were measured to ensure the collected data are comparable. This is extremely important when the traits measured vary depending on taxon, sex or environmental factors (Johnson et al., 2018), or indeed, when traits have multiple definitions or assaying protocols across research domains or taxonomic groups (e.g. critical temperatures; Bates and Morley, 2020). In this regard, an advantage of meta-analyses is that differences in units, methods or measurements between the studies are not a barrier [see Noble et al., 2022 (this issue)], because unitless effect sizes are being analysed. However, for trait-based analyses, reduced comparability between raw data or missing information in the original studies (e.g. studies of mass-corrected traits that do not present mass) might be problematic. In this sense, tools like the R packages *Rphylopars* (Goolsby et al., 2017) and *mice* (van Buuren and Groothuis-

Oudshoorn, 2011) can allow, in some scenarios, the interpolation and imputation of missing data. Similarly, several R packages can handle uncertain or variable taxonomy for phylogenetic analyses (e.g. *Taxa*; Foster et al., 2018). The multiple and often complicated design elements of original studies means that researchers need to trade off narrowing the search and data extraction criteria to fit their specific question, against the inclusion of additional data that might be used to answer similar questions. A good balance would be to carefully decide the inclusion criteria for your current study, but keep track of excluded or accessory data that could be used for expanding into future studies (see Tip 2). For example, metadata referring to taxonomic reference databases, such as NCBI Taxonomic Database or Open Tree of Life (Hinchliff et al., 2015), may aid in updating your database if taxonomy changes.

A final key element of efficient searching is to recruit collaborators to assist with the large time and effort required in collecting data from the original sources. In meta-analyses, for example, around 92% of studies initially identified during a search do not match the criteria needed to answer the research question [Fig. 2B; from Noble et al., 2022 (this issue)]. Most biologists building databases are unlikely to have a dedicated data curator; the work will be done by students, post-doctoral researchers and more senior scientists with a specific project in mind, and future curation will only occur with new projects. Thus, it is incredibly valuable to have several collaborators to divide the effort. It is critical that contributions from collaborators are consistent and comparable. Inconsistencies in data entry among individuals can be minimized by agreeing upon inclusion and exclusion criteria during the planning phase and training on a random subset of initial studies to assess agreement. During data compilation, random checks for consistency and quality can be conducted in combination with version control to correct arising issues (see Tip 3).

Tip 2: structure your databases for multiple uses

At minimum, a database requires two layers: (1) the data, and (2) associated reference information for the data. Each of these layers needs a metadata file to explain the data columns. The data layer may include trait values (typically raw or summary data) and their associated error metrics, as well as important covariates, such as location, taxonomic rank and reference database identifiers, and units (Fig. 3). The associated reference information could include (in addition to the bibliographic reference for each data point) definitions of trait measurements or relevant information on methodology. Creating a reference database that is rich in information about each study adds value to the database beyond its original purpose. For example, a reference dataset that includes all papers that were full-text screened and coding for exclusion/inclusion, as well as noting additional relevant variables not used in the initial study, will provide future users with a starting point if they wish to broaden the inclusion criteria. Having detailed and clear metadata overcomes problems associated with vague or inconsistent trait measurements in the data by unambiguously defining the variable for the user (which may be your future self).

Once you have decided which variables to include in your database, there are two ways of structuring the database ('effect size' format and 'stacked' format; Fig. 3) that depend on its intended purpose. Because a meta-analysis is focused on examining the strength of effect associated with a specific hypothesis, the statistical analyses often require a dataset where each row is a meta-analytic effect size arising from a quantitative correlation (e.g. an *r*-value) or a paired comparison of treatments [e.g. standardized mean difference (Hedges' *g*) or odds ratio]. Thus, each row of effect sizes effectively represents two or more data points (e.g. control mean versus treatment mean; Curtis et al., 2013). Trait-based analyses have a longer, 'stacked' format, with one datum per row representing a single observation (e.g. mean trait value for a group). Both types of databases have additional columns for relevant

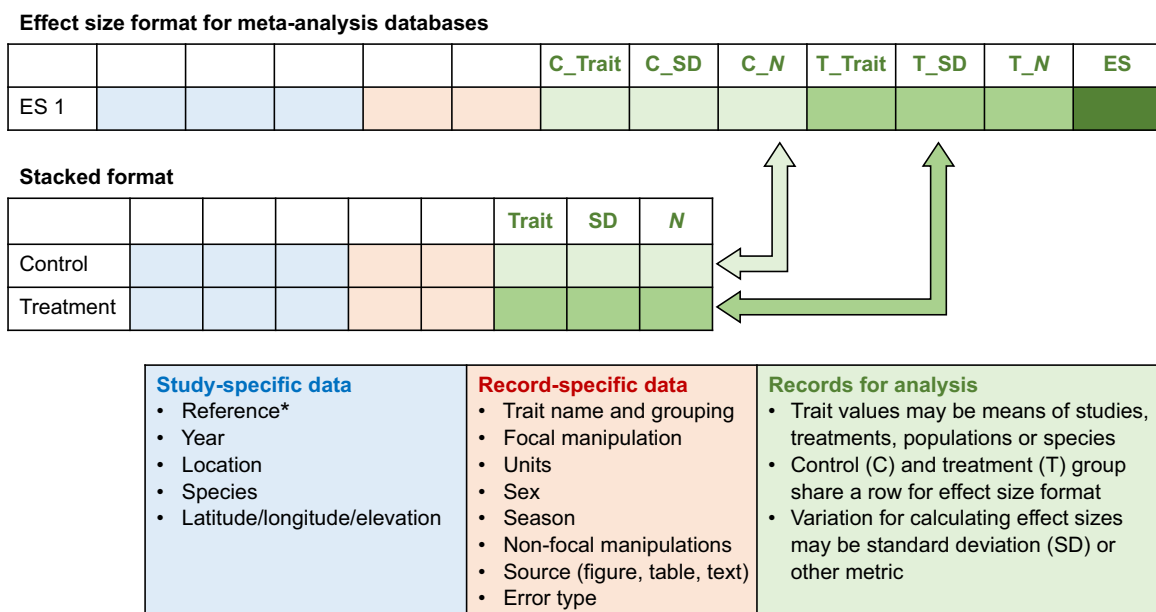


Fig. 3. Relationship between an 'effect size' format for meta-analysis and a 'stacked' format for general comparative databases. Study-specific data (blue) and record-specific data (red) may be shared between both formats. Columns refer to trait values for different groups within the study (C: control; T: treatment). A pairwise comparison in an 'effect size' format translates to rows in a 'stacked' format. Calculated effect sizes (ES) are unique to meta-analyses. In our diagram, *N* is sample size and *SD* is standard deviation. *Note that databases may have multiple records per row (e.g. mass, lifespan, metabolic rate) compiled from multiple sources; columns for each record and bibliographic reference will be included.

covariates, moderators or grouping categories (e.g. ‘male’ or ‘female’).

These two general structures conflict between best practice for specific immediate use and best practice for future multi-use. Thus, trade-offs arise when deciding between database structure and function. This trade-off is particularly strong for meta-analyses. Effect sizes are highly specific for the intended question and an indication of the associated trait data is required if the same data points are applied to other questions. An ‘effect size’ format can be expanded into a ‘stacked’ format to facilitate use in other comparative analyses as long as the trait data are provided in the database, but this process is arduous and error-prone. Furthermore, some moderators or covariates of interest to the meta-analysis may be contrast specific and lost when the data are stacked, while other important covariates that were treatment specific may not be documented in the ‘effect size’ format. We argue that creating an initial ‘stacked’ format database, even for meta-analyses, has little cost and large benefit because of the greater applicability and ease of use in a wide range of comparative questions. The stacked database can be the primary focus for version updates (see Tip 3) and accessibility for other researchers (see Tip 4), whereas ‘effect size’ format versions can be linked directly to the research paper they support.

For any comparative analysis, extracting a comprehensive set of information from original sources may improve the re-usability of the database, the applicability for other questions, and the ease of combining the database with other relevant databases (e.g. phylogenies or global climate databases). However, this benefit occurs at the expense of the time and effort to obtain these data. It may be that these additional variables are not critical for the initial analysis, but are commonly reported covariates in certain fields, such as latitude or longitude, species identity and body mass. When possible, adding these common covariates as reported in the original sources is preferable to appending values from an independent comparative dataset, because a measure reported in such a dataset may differ from the mean of the population for which the focal data are compiled. Arguably the foresight to include common covariates represents a significant saving in time compared with repeatedly extracting these data, enabling multiple research studies from the single database effort. For example, during the creation of HormoneBase (hormonebase.org; Vitousek et al., 2018b), in addition to the focal hormone data, collaborators compiled a wide range of geographical, body size and life history data (Johnson et al., 2018) that have led, to date, to 10 publications, including seven trait-based analyses (e.g. Vitousek et al., 2019; Injaian et al., 2020; Husak et al., 2021).

Tip 3: version control your database

During the curation of a database (within a single study or across multiple studies), it will change and evolve as you use it – through finding errors in a previous version, updating recent literature and data, and expanding or changing the purpose of your database. Changes in the database can escalate quickly, especially when multiple collaborators are involved. This process can lead to general chaos: uncertainty of which is the most up-to-date version of the database, one person doing something incorrect and no means of correcting it without substantial effort, and uncertainty regarding the steps you took in creating your database. Ideally, the group would be able to document each of their changes individually, including what they were and who made them, particularly because this would facilitate reproducibility and transparency (Ram, 2013; Shaw et al., 2016; Lowndes et al., 2017; Powers and Hampton,

Box 1. Checklist for making a database

1. Survey the literature to establish the scope of your research question(s), and determine the ideal statistical approach to address the question
2. Determine your strategy for a systematic search of the primary (and possibly grey) literature
3. Determine the list of variables (and their standardized units) to be compiled from each study, including any covariates and non-target traits that may have long-term utility (e.g. latitude and longitude of study location, body mass) and any necessary reference information
4. Establish a strategy for version control of your database
5. Compile the data in your database and use the database to test your question(s) of interest
6. Make your database accessible (i.e. on a webpage or repository) to maximize use to other researchers
7. Publish a description of the database in a data paper to maximize its visibility to other researchers

2019). One straightforward means of accomplishing this is version control, also known as a revision or source control (Ruparella, 2010).

Even if version control seems unfamiliar, you have probably been using a very simplified form of it to keep track of versions of draft manuscripts or datasets already! The most basic and technologically simple type of version control would be to keep your database (e.g. a .csv file) saved as versions with different, sequential names, paired with a meta-document detailing the major differences between the versions. However, there are also many elegant and easy-to-use systems for more complex and collaborative means of version control [i.e. Apache™; Subversion® (Pilato et al., 2008); Concurrent Versions Systems (Grune, 1986); Git (Somasundaram, 2013); Perforce; and as reviewed in Zolkifli et al., 2018]. The use of these systems requires technological know-how, but can be a worthwhile investment in the long run. The technological challenges can be navigated by having a designated person within your collaboration lead their use, as well as employing graphical user interfaces, like Sourcetree (www.sourcetreeapp.com) or GitKraken (www.gitkraken.com). Using version control, files are saved as they are edited and you can jump to any instance in time. That means if you make a mistake in the database and notice it in the short- or long-term, then you can simply revert to the version of the database before that mistake was made! Two aspects of version control that are also particularly helpful are: (1) providing comments on what was modified and how, and (2) tracking who made each change. These features facilitate understanding how your database evolves over time and determine with whom to follow-up to discuss any changes that were made. Indeed, version control is critical for the facilitation of a ‘living database’ that is updated and used for multiple studies over time, as it is a clear method for tracking discrete database updates. A ‘living database’ maintains its scientific value over time by providing up-to-date data, but obviously generates a cost associated with performing the updated searches, data extraction and version control. We assert that version-controlled updates should be prioritized over maintaining a continuously updated ‘dynamic database’ because researchers can directly refer to the version that they base their study on. It is important to report which version of your database is used in a publication and, if your database is publicly available on a website, to keep all versions of your database accessible in case someone wants to reproduce an older study. Overall, version control is an efficient way of ensuring reproducibility and transparency in database construction, and as an offshoot will streamline the construction process, especially for projects with many collaborators.

Tip 4: make your database accessible

There are several ways to ensure that a comparative database will be of maximum use to other researchers, either for their own projects or to replicate your results. First, it is important to consider which data to share. Most comparative analyses focus on the mean and variability of response variables, and authors often share only those values. However, in many cases it is beneficial to share the raw data from which those summary statistics were calculated.

Next, it is important to ensure that the database is stored in a format from which it is easy to extract information. For example, data in a table format saved as a comma-separated values (.csv) or text (.txt) file can be opened by many programs such as R and Python for downstream analysis. Avoid unusable formats such as word documents and PDFs, as these essentially force others to re-enter your data. Most published papers today provide useable data, but it is still relatively common for data to be shared in an unusable format. For example, when surveying the meta-analyses in comparative physiology identified by Noble et al. (2022; this issue), only 70% (37) of the 53 papers that provided data did so in a useable format and 30% (16) did so with an unusable format. Similarly, around 64% of data archives in ecology and evolution are not usable (Roche et al., 2015).

Where and how data are hosted is also important. Data can be hosted on a researcher's website (e.g. Reptile Development Database, www.repdevo.com), as a supplementary document hosted by the journal (e.g. Gunderson and Stillman, 2015), or on a data repository site such as Dryad (www.datadryad.org), Figshare (www.figshare.com; e.g. Vitousek et al., 2018b; hormonebase.org), the Open Science Framework (www.osf.io; e.g. Merklung et al., 2018) or Github (www.github.com). Ideally, the data are associated with a doi number so that they have a permanent location that can be reliably accessed; doi numbers are most easily assigned by hosting on journal or data repository websites. If the dataset is large in size and scope, particularly if larger than any single comparative analysis performed using it, writing a 'data paper' maximizes the profile, accessibility and citability of the dataset. A data paper would be in addition to any papers written using the dataset to test the original, specific biological hypotheses and predictions. Data papers describe the features of the dataset, such as what information is included, how it was compiled, and how it can be used (e.g. Noble et al., 2018b; Vitousek et al., 2018b). Data papers often also include tables and graphics that summarize key features of the data. Furthermore, data papers can be used to inform other researchers about the existence of your dataset, the location where it is stored, as well how it may be best utilized for future research. Ideally, promoting your database in this way will maximize the utility of your database for the advancement of the field.

Conclusions

The collective experience of the authors in creating databases for use in a variety of quantitative comparative studies has revealed to us the potential for more efficient database creation. It is striking that ~90% of papers examined during a targeted literature search are discarded as being inappropriate for the immediate research question; this represents a massive loss of effort that may be repeated by multiple research teams considering related questions. Instead, we argue that the creators of new comparative databases may maximize the future utility of these databases with the inclusion of common, standardized covariates and non-target traits, even if they are not of direct interest to the question that drives the initial effort. In addition, we appeal for greater uptake of open science principles that facilitate synthesis and database curation, including

reducing barriers to accessing data, implementing standardized methodologies and documenting reproducible analyses (O'Dea et al., 2021b). The effort and pay-off associated with compiling a useful comparative database should compel our community to work towards the creation of databases with more collaborative, multi-use potential. To this end, participating in community curation of data (e.g. an 'open synthesis community'; Nakagawa et al., 2020) and initiatives that aggregate datasets (e.g. Open Traits Network; Gallagher et al., 2020), will increase efficiency and reduce redundancy in scientific synthesis, as well as (ideally) establish common workflows or approaches and formalize metadata across databases.

Comparative approaches offer the exciting opportunity to address biological questions at increasingly large taxonomic, geographic, temporal and conceptual scales. As theoretical and statistical tools are being continually developed and refined, the concurrently increasing availability of comparative databases provide the means with which macro-ecological and -evolutionary hypotheses may be tested (Beck et al., 2012). The fields of physiology and biomechanics, along with many other areas of biology, are thus poised to take full advantage of the power of comparative analyses. Yet the first step in many comparative studies – the compilation of a database that allows a team of investigators to examine a specific research question – remains a daunting task for many of us. Our hope is that this Commentary demystifies this process, and provides a resource for others who wish to pursue it (Box 1).

Acknowledgements

We would like to thank Craig Franklin for the invitation to contribute this Commentary to the Special Issue, and two anonymous reviewers for their constructive comments. We are grateful to Dan Noble, Patrice Pottier and Sammy Burke for extracting information for our use during their screening of meta-analyses in the field, and for sharing the data with us.

Competing interests

The authors declare no competing or financial interests.

Funding

While preparing the Commentary, J.R. was supported by a Natural Sciences Engineering Research Council of Canada (NSERC) postdoctoral fellowship, and M.I.-C. was supported by a postdoctoral fellowship from the Andalusian Government.

References

- Bates, A. E. and Morley, S. A. (2020). Interpreting empirical estimates of experimentally derived physiological and biological thermal limits in ectotherms. *Canad. J. Zool.* **98**, 237–244. doi:10.1139/cjz-2018-0276
- Beck, J., Ballesteros-Mejia, L., Buchmann, C. M., Dengler, J., Fritz, S. A., Gruber, B., Hof, C., Jansen, F., Knapp, S. and Kreft, H. (2012). What's on the horizon for macroecology? *Ecography* **35**, 673–683. doi:10.1111/j.1600-0587.2012.07364.x
- Bodensteiner, B. L., Agudelo-Cantero, G. A., Arietta, A. Z. A., Gunderson, A. R., Muñoz, M. M., Refsnider, J. M. and Gangloff, E. J. (2021). Thermal adaptation revisited: how conserved are thermal traits of reptiles and amphibians? *J. Exp. Zool. A* **335**, 173–194. doi:10.1002/jez.2414
- Borries, C., Sandel, A. A., Koenig, A., Fernandez-Duque, E., Kamilar, J. M., Amoroso, C. R., Barton, R. A., Bray, J., Di Fiore, A., Gilby, I. C. et al. (2016). Transparency, usability, and reproducibility: guiding principles for improving comparative databases using primates as examples. *Evol. Anthropol.* **25**, 232–238. doi:10.1002/evan.21502
- Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M. and West, G. B. (2004). Toward a metabolic theory of ecology. *Ecology* **85**, 1771–1789. doi:10.1890/03-9000
- Côté, I. M., Curtis, P. S., Rothstein, H. R. and Stewart, G. B. (2013). Gathering data: searching literature and selection criteria. In *Handbook of Meta-analysis in Ecology and Evolution* (ed. J. Koricheva, J. Gurevitch and K. Mengersen), pp. 37–51. Princeton: Princeton University Press.
- Curtis, P. S., Mengersen, K., Lajeunesse, M. J., Rothstein, H. R. and Stewart, G. B. (2013). Extraction and critical appraisal of data. In *Handbook of Meta-analysis in Ecology and Evolution* (ed. J. Koricheva, J. Gurevitch and K. Mengersen), pp. 52–60. Princeton: Princeton University Press.

- Davidson, A. M., Jennions, M. and Nicotra, A. B. (2011). Do invasive species show higher phenotypic plasticity than native species and, if so, is it adaptive? A meta-analysis. *Ecol. Lett.* **14**, 419–431. doi:10.1111/j.1461-0248.2011.01596.x
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A. and Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J.* **22**, 338–342. doi:10.1096/fj.07-9492LSF
- Foo, Y. Z., O'Dea, R. E., Koricheva, J., Nakagawa, S. and Lagisz, M. (2021). A practical guide to question formation, systematic searching and study screening for literature reviews in ecology and evolution. *Methods Ecol. Evol.* **12**, 1705–1720. doi:10.32942/osf.io/6v54p
- Forero, D. A., Lopez-Leon, S., Gonzalez-Giraldo, Y. and Bagos, P. G. (2019). Ten simple rules for carrying out and writing meta-analyses. *PLoS Comput. Biol.* **15**, e1006922. doi:10.1371/journal.pcbi.1006922
- Foster, Z. S. L., Chamberlain, S. and Grünwald, N. J. (2018). Taxa: an R package implementing data standards and methods for taxonomic data. *F1000Res.* **7**, 272. doi:10.12688/f1000research.14013.2
- Francis, C. D., Donald, J. W., Downs, C. J., Fuxjager, M. J., Goymann, W., Hau, M., Husak, J. F., Johnson, M. A., Kircher, B. K., Knapp, R. et al. (2018). Metabolic scaling of stress hormones in vertebrates. *Integr. Comp. Biol.* **58**, 729–738. doi:10.1093/icb/icy063
- Gallagher, R. V., Falster, D. S., Maitner, B. S., Salguero-Gómez, R., Vandvik, V., Pearse, W. D., Schneider, F. D., Kattge, J., Poelen, J. H., Madin, J. S. et al. (2020). Open Science principles for accelerating trait-based science across the Tree of Life. *Nat. Ecol. Evol.* **4**, 294–303. doi:10.1038/s41559-020-1109-6
- Geange, S. R., Arnold, P. A., Catling, A. A., Coast, O., Cook, A. M., Gowland, K. M., Leigh, A., Notarnicola, R. F., Posch, B. C. and Venn, S. E. (2021). The thermal tolerance of photosynthetic tissues: a global systematic review and agenda for future research. *New Phytol.* **229**, 2497–2513. doi:10.1111/nph.17052
- Gerstner, K., Moreno-Mateos, D., Gurevitch, J., Beckmann, M., Kambach, S., Jones, H. P. and Seppelt, R. (2017). Will your paper be used in a meta-analysis? Make the reach of your research broader and longer lasting. *Method. Ecol. Evol.* **8**, 777–784. doi:10.1111/2041-210X.12758
- Goolsby, E. W., Bruggeman, J. and Ané, C. (2017). Rphylopar: fast multivariate phylogenetic comparative methods for missing data and within-species variation. *Methods Ecol. Evol.* **8**, 22–27. doi:10.1111/2041-210X.12612
- Grune, D. (1986). *Concurrent Versions Systems, A Method for Independent Cooperation*. VU Amsterdam. Subfaculteit Wiskunde en Informatica.
- Gunderson, A. R. and Stillman, J. H. (2015). Plasticity in thermal tolerance has limited potential to buffer ectotherms from global warming. *Proc. R. Soc. B* **282**, 20150401. doi:10.1098/rspb.2015.0401
- Hinchliff, C. E., Smith, S. A., Allman, J. F., Burleigh, J. G., Chaudhary, R., Coghill, L. M., Crandall, K. A., Deng, J., Drew, B. T., Gazis, R. et al. (2015). Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proc. Natl. Acad. Sci. USA* **112**, 12764–12769. doi:10.1073/pnas.1423041112
- Husak, J. F., Fuxjager, M. J., Johnson, M. A., Vitousek, M. N., Donald, J. W., Francis, C. J., Goymann, W., Hau, M., Kircher, B. K., Knapp, R. et al. (2021). Life history and environment predict variation in testosterone across vertebrates. *Evolution* **75**, 1003–1010. doi:10.1111/evo.14216
- Iglesias-Carrasco, M., Aich, U., Jennions, M. D. and Head, M. L. (2020). Stress in the city: meta-analysis indicates no overall evidence for stress in urban vertebrates. *Proc. R. Soc. B* **287**, 20201754. doi:10.1098/rspb.2020.1754
- Injaian, A. S., Francis, C. D., Ouyang, J. Q., Dominoni, D. M., Donald, J. W., Fuxjager, M. J., Goymann, W., Hau, M., Husak, J. F., Johnson, M. A. et al. (2020). Baseline and stress-induced corticosterone levels across birds and reptiles do not reflect urbanization levels. *Conserv. Physiol.* **8**, e01110. doi:10.1093/conphys/coz110
- Johnson, M. A., Francis, C. D., Miller, E. T., Downs, C. J. and Vitousek, M. N. (2018). Detecting bias in large-scale comparative analyses: methods for expanding the scope of hypothesis-testing with HormoneBase. *Integr. Comp. Biol.* **58**, 720–728. doi:10.1093/icb/icy045
- Lowndes, J. S. S., Best, B. D., Scarborough, C., Afflerbach, J. C., Frazier, M. R., O'Hara, C. C., Jiang, N. and Halpern, B. S. (2017). Our path to better science in less time using open data science tools. *Nature Ecol. Evol.* **1**, 0160. doi:10.1038/s41559-017-0160
- Merklings, T., Nakagawa, S., Lagisz, M. and Schwanz, L. E. (2018). Maternal testosterone and offspring sex-ratio in birds and mammals: a meta-analysis. *Ecol. Biol.* **45**, 96–104. doi:10.1007/s11692-017-9432-9
- Moher, D., Liberati, A., Tetzlaff, J. and Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* **339**, b2535. doi:10.1136/bmj.b2535
- Muñoz, M. M. and Price, S. A. (2019). The future is bright for evolutionary morphology and biomechanics in the era of big data. *Integr. Comp. Biol.* **59**, 599–603. doi:10.1093/icb/icz121
- Nakagawa, S., Dunn, A., Lagisz, M., Bannach-Brown, A., Grames, E. M., Sánchez-Tójar, A., O'Dea, R. E., Noble, D. W. A., Westgate, M. J., Arnold, P. A. et al. (2020). A new ecosystem for evidence synthesis. *Nature Ecol. Evol.* **4**, 498–501. doi:10.1038/s41559-020-1153-2
- Noble, D. W. A., Stenhouse, V. and Schwanz, L. E. (2018a). Developmental temperatures and phenotypic plasticity in reptiles: a systematic review and meta-analysis. *Biol. Rev.* **93**, 72–97. doi:10.1111/brv.12333
- Noble, D. W. A., Stenhouse, V., Riley, J. E., Warner, D. W., While, G. M., Du, W.-G., and Uller, T. and Schwanz, L. E. (2018b). A comprehensive database of thermal developmental plasticity in reptiles. *Sci. Data* **5**, 17. doi:10.1038/sdata.2018.138
- Noble, D. W. A., Pottier, P., Lagisz, M., Burke, S., Drobnik, S. M., O'Dea, R. E. and Nakagawa, S. (2022). Meta-analytic approaches and effect sizes to account for 'nuisance heterogeneity' in comparative physiology. *J. Exp. Biol.* **225**, jeb243225. doi:10.1242/jeb.243225
- O'Dea, R. E., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W. A., Parker, T. H., Gurevitch, J., Page, M. J., Stewart, G., Moher, D. et al. (2021a). Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension. *Biol. Rev.* **96**, 1695–1722. doi:10.1111/brv.12721
- O'Dea, R. E., Parker, T. H., Chee, Y. E., Culina, A., Drobnik, S. M., Duncan, D. H., Fidler, F., Gould, E., Ihle, M., Kelly, C. D. et al. (2021b). Towards open, reliable, and transparent ecology and evolutionary biology. *BMC Biol.* **19**, 68. doi:10.1186/s12915-021-01006-3
- Ouzzani, M., Hammady, H., Fedorowicz, Z. and Elmagarmid, A. (2017). Rayyan – a web and mobile app for systematic reviews. *Syst. Rev.* **5**, 210. doi:10.1186/s13643-016-0384-4
- Pick, J. L., Nakagawa, S. and Noble, D. W. A. (2019). Reproducible, flexible and high-throughput data extraction from primary literature: the metaDigitise R package. *Methods Ecol. Evol.* **10**, 426–431. doi:10.1111/2041-210X.13118
- Pilato, C. M., Collins-Sussman, B. and Fitzpatrick, B. W. (2008). *Version Control with Subversion: Next Generation Open Source Version Control*. O'Reilly Media, Inc..
- Pörtner, H.-O., Bock, C. and Mark, F. C. (2017). Oxygen- and capacity-limited thermal tolerance: bridging ecology and physiology. *J. Exp. Biol.* **220**, 2685–2696. doi:10.1242/jeb.134585
- Powers, S. M. and Hampton, S. E. (2019). Open science, reproducibility, and transparency in ecology. *Ecol. Appl.* **29**, e01822. doi:10.1002/eap.1822
- Ram, K. (2013). Git can facilitate greater reproducibility and increased transparency in science. *Source Code Biol. Med.* **8**, 7. doi:10.1186/1751-0473-8-7
- Roche, D. G., Kruuk, L. E. B., Lanfear, R. and Binning, S. A. (2015). Public data archiving in ecology and evolution: how well are we doing? *PLoS Biol.* **13**, e1002295. doi:10.1371/journal.pbio.1002295
- Ruparelia, N. B. (2010). The history of version control. *ACM SIGSOFT Software Engineering Notes* **35**, 5–9. doi:10.1145/1668862.1668876
- Salameh, J.-P., Bossuyt, P. M., McGrath, T. A., Thoms, B. D., Hyde, C. J., Macaskill, P., Deeks, J. J., Leeflang, M., Korevaar, D. A., Whiting, P. et al. (2020). Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. *BMJ* **370**, m2632. doi:10.1136/bmj.m2632
- Schmidt-Nielsen, K. (1972). Locomotion: energy cost of swimming, flying, and running. *Science* **177**, 222–228. doi:10.1126/science.177.4045.222
- Schmidt-Nielsen, K. (1975). Scaling in biology: the consequences of size. *J. Exp. Zool.* **194**, 287–307. doi:10.1002/jez.1401940120
- Seebacher, F., White, C. R. and Franklin, C. E. (2015). Physiological plasticity increases resilience of ectothermic animals to climate change. *Nat. Clim. Change* **5**, 61–66. doi:10.1038/nclimate2457
- Shaw, R. G., Moore, A. J., Noor, M. and Ritchie, M. G. (2016). Transparency and reproducibility in evolutionary research. *Ecol. Evol.* **6**, 4605. doi:10.1002/ece3.2291
- Somasundaram, R. (2013). *Git: Version Control for Everyone*. Packt Publishing Ltd.
- Somero, G. N. (2000). Unity in diversity: a perspective on the methods, contributions, and future of comparative physiology. *Annu. Rev. Physiol.* **62**, 927–937. doi:10.1146/annurev.physiol.62.1.927
- Stewart, G. B., Côté, I. M., Rothstein, H. R. and Curtis, P. S. (2013). Extraction and critical appraisal of data. In *Handbook of Meta-analysis in Ecology and Evolution* (ed. J. Koricheva, J. Gurevitch and K. Mengersen), pp. 52–60. Princeton: Princeton University Press.
- van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: multivariate imputation by chained equations in R. *J. Stat. Software* **45**, 1–67. doi:10.18637/jss.v045.i03
- Vitousek, M. N., Johnson, M. A. and Husak, J. F. (2018a). Illuminating endocrine evolution: the power and potential of large-scale comparative analyses. *Integr. Comp. Biol.* **58**, 712–719. doi:10.1093/icb/icy098
- Vitousek, M. N., Johnson, M. A., Donald, J. W., Francis, C. D., Fuxjager, M. J., Goymann, W., Hau, M., Husak, J. F., Kircher, B. K., Knapp, R. et al. (2018b). HormoneBase, a population-level database of steroid hormone levels across vertebrates. *Sci. Data* **5**, 180097. doi:10.1038/sdata.2018.97
- Vitousek, M. N., Johnson, M. A., Downs, C. J., Miller, E. T., Martin, L. B., Francis, C. D., Donald, J. W., Fuxjager, M. J., Goymann, W., Hau, M. et al. (2019). Macroevolutionary patterning in glucocorticoids suggests different

- selective pressures shape baseline and stress-induced levels. *Am. Nat.* **193**, 866-880. doi:10.1086/703112
- Vogel, S.** (2008). Modes and scaling in aquatic locomotion. *Integr. Comp. Biol.* **48**, 702-712. doi:10.1093/icb/icn014
- Wallace, B. C., Small, K., Brodley, C. E., Lau, J. and Trikalinos, A.** (2012). Deploying an interactive machine learning system in an evidence-based practice center: abstrackr. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*. New York: ACM; 2012, 819-824. doi:10.1145/2110363.2110464
- Westgate, M. J.** (2019). revtools: an R package to support article screening for evidence synthesis. *Res. Syn. Meth.* **10**, 606-614. doi:10.1002/jrsm.1374
- White, E. P., Baldrige, E., Brym, Z. T., Locey, K. J., McGlenn, D. J. and Supp, S. R.** (2013). Nine simple ways to make it easier to (re)use your data. *Ideas Ecol. Evol.* **6**, 1-10. doi:10.4033/iee.2013.6b.6.f
- White, C. R., Marshall, D. J., Alton, L. A., Arnold, P. A., Beaman, J. E., Bywater, C. L., Condon, C., Crispin, T. S., Janetzki, A., Pirtle, E. et al.** (2019). The origin and maintenance of metabolic allometry in animals. *Nature Ecol. Evol.* **3**, 598-603. doi:10.1038/s41559-019-0839-9
- Whitlock, M. C.** (2011). Data archiving in ecology and evolution: best practices. *Trends Ecol. Evol.* **26**, 61-65. doi:10.1016/j.tree.2010.11.006
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. et al.** (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018. doi:10.1038/sdata.2016.18
- Wu, N. C. and Seebacher, F.** (2020). Effect of the plastic pollutant bisphenol A on the biology of aquatic organisms: a meta-analysis. *Glob. Chang. Biol.* **26**, 3821-3833. doi:10.1111/gcb.15127
- Zolkifli, N. N., Ngah, A. and Deraman, A.** (2018). Version control system: a review. *Procedia Computer Science* **135**, 408-415. doi:10.1016/j.procs.2018.08.191