Geographical bias in physiological data limits predictions of global change impacts

Craig R. White\textsuperscript{1,2} | Dustin J. Marshall\textsuperscript{1,2} | Steven L. Chown\textsuperscript{1} | Susana Clusella-Trullas\textsuperscript{3} | Steven J. Portugal\textsuperscript{4} | Craig E. Franklin\textsuperscript{5} | Frank Seebacher\textsuperscript{6}

\textsuperscript{1}School of Biological Sciences, Monash University, Melbourne, Australia; \textsuperscript{2}Centre for Geometric Biology, Monash University, Melbourne, Australia; \textsuperscript{3}Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, Stellenbosch, South Africa; \textsuperscript{4}Department of Biological Sciences, Royal Holloway, University of London, Egham, UK; \textsuperscript{5}School of Biological Sciences, The University of Queensland, Brisbane, Australia and \textsuperscript{6}School of Life and Environmental Sciences A08, University of Sydney, Sydney, Australia

\textbf{Abstract}

1. Climate affects all aspects of biology. Physiological traits play a key role in mediating these effects, because they define the fundamental niche of each organism.

2. Climate change is likely to shift environmental conditions away from physiological optima. The consequences for species are significant: they must alter their physiology through plasticity or adaptation, move, or decline to extinction. The ability to understand and predict such organismal responses to global change is, however, only as good as the geographical coverage of the data on which these predictions are based.

3. Geographical biases in the state of physiological knowledge have been identified, but it has not been determined if these geographical biases are likely to limit our capacity to predict the outcomes of global change. We show that current knowledge of physiological traits is representative of only a limited range of the climates in which terrestrial animals will be required to operate, because data for animals from only a limited range of global climates have been incorporated in existing compilations.

4. We conclude that geographical bias in existing datasets limits our capacity to predict organismal responses in the vast areas of the planet that are unstudied, and that this geographical bias is a much greater source of uncertainty than the difference between the current climate and the projected future climate. Addressing this bias is urgent to understand where impacts will be most profound, and where the need for intervention is most pressing.

\textbf{Keywords}
climate, forecasting, functional traits, interactive effects, meta-analysis, precipitation, rainfall, temperature
1 | INTRODUCTION

Physiological traits mediate the responses of species to climate change (Hoffmann & Sgrò, 2011; Seebacher et al., 2015). Physiological information therefore has considerable predictive power for forecasting how climate change will affect species (Diamond, Nichols et al., 2012; Elith et al., 2010; Humphries et al., 2002; Kearney & Porter, 2009; Mathewson et al., 2017). However, data accumulated over past decades exhibit a profound spatial bias: more data are available for species from North America, Western Europe, South Africa and Australia than for species from other regions of the world (Addo-Bediako et al., 2001; Chown et al., 2002; Clusella-Trullas et al., 2011; Dillon et al., 2010; Seebacher et al., 2015; Figure 1).

The extent to which this spatial bias compromises understanding of global responses to climate change will depend on (a) the extent to which physiological traits differ among species, and particularly whether the sampled species represent species in unsampled environments, and (b) the extent to which the studied environments are representative of unstudied and future environments.

Physiological performance of organisms varies among species living in different environments, because climate is a potent driver of trait evolution (Amarasekare & Johnson, 2017; Beaman et al., 2016; Chevin et al., 2010). Species from cold environments have relatively high metabolic rates (Addo-Bediako et al., 2002; Anderson & Jetz, 2005; Lovegrove, 2003; White, Blackburn, Martin et al., 2007) and are more tolerant of cold (Addo-Bediako et al., 2000; Clusella-Trullas et al., 2011; Kellermann et al., 2012) than species from warm environments; species from dry environments have lower rates of water loss (Addo-Bediako et al., 2001; Portugal et al., 2014) and improved desiccation resistance (Kellermann et al., 2009, 2012) compared to species from more moist environments; and species from stable environments exhibit greater physiological plasticity than those from more variable environments (Seebacher et al., 2015), although this differs among traits (Gunderson & Stillman, 2015). Species from understudied regions will therefore be functionally distinct from those from better studied regions, if the regions differ in climate (Amarasekare & Johnson, 2017; Chevin et al., 2010; Marshall & Burgess, 2015).

Given that species traits vary with climate, the extent to which current knowledge of functional trait diversity will be useful to predict global responses to climate change will depend on the extent to which studied environments are representative of the full range of current global environments, and the range of environments that organisms will be exposed to in the future. Here we quantitatively evaluate the concordance between the environments for which we have some knowledge of functional traits (‘studied environments’) and global current and projected future environments by combining georeferenced trait data from published studies with climate data from recent (1970–2000) and projected future (2041–2060) climatologies. We compare studied environments to randomly chosen current and projected future environments to assess how well our current knowledge is representative of the range of current global environments and future environments.

2 | MATERIALS AND METHODS

We gathered a total of 5,329 spatially explicit measurements of the physiological traits likely to dictate the responses of terrestrial animals to climate change: critical upper and lower thermal limits, supercooling temperature, optimal temperatures for performance, rates of metabolism, patterns of gas exchange and acclimation capacity (Table 1). The physiological traits we chose as means to identify studied environments define the range of temperatures over which animals can survive (critical upper and lower thermal limits, Figure 1).
supercooling temperature), the rates at which they utilise energy and the demands they place on their environment (metabolic rate, which exhibits associations with survival and fitness that vary among environmental and ecological contexts: Boratyński & Koteja, 2009; Boyce et al., 2020; Pettersen et al., 2016, 2020), their susceptibility to desiccation stress (gas exchange patterns: Schimpf et al., 2012; White, Blackburn, Terblanche, et al., 2007), the temperatures at which functional performance (development, growth, and locomotion) is maximised, and their capacity to compensate physiologically for changes in the thermal environmental (acclimation capacity). The dataset includes records for 2,637 species, including insects, arachnids, fish, amphibians, reptiles, birds and mammals (Table 1). We drew exclusively upon published compilations with broad spatial or taxonomic coverage (Anderson & Jetz, 2005; Buckley et al., 2018; Bushuev et al., 2018; Clussella-Trullas et al., 2011; Diamond, Nichols, et al., 2012; Hoffmann et al., 2013; Irlich et al., 2009; Londoño et al., 2015; Naya et al., 2018; Portugal et al., 2014; Rohr et al., 2018; Seebacher et al., 2015; Sørensen et al., 2018; White, Blackburn, Martín, et al., 2007; White, Blackburn, Terblanche, et al., 2007; White et al., 2019; Wiersma et al., 2007). Spatial bias in existing data is already well documented (Addo-Bediako et al., 2001; Chown et al., 2002; Clussella-Trullas et al., 2011; Dillon et al., 2010; Seebacher et al., 2015), and our aim was to quantify the extent to which this known bias limits our capacity to forecast responses to climate change.

For each trait record, we first extracted associated data for mean annual temperature, total annual precipitation, the standard deviation of annual temperature and the coefficient of variation of annual precipitation were log_{10}(x + 1) transformed, and the global range of values for each climate variable was scaled to fall between 0 and 1. Next, we assessed the distribution of potential future climates, by generating 5,329 randomly selected locations (i.e. the same size as the empirical dataset) around the globe, using the dismo v1.3–3 (Hijmans et al., 2013) package in R v4.0.3 (R Core Team, 2020), and extracted climate data for these locations. This process of generating 5,329 randomly selected locations and associated climate data was repeated 1,000 times to yield 1,000 ‘sets’ of randomly selected locations and climates.

We also extracted climate projections (CMIP Phase 6) of mean annual temperature, total annual precipitation, the standard deviation of annual temperature and the coefficient of variation of annual rainfall for the year 2050 (average for 2041–2060) at 10-arc-min resolution based on eight global climate models (BCC-CSM2-MR, CanESM5, CNRM-CM6-1, CNRM-ESM2-1, IPSL-CM6A-LR, MIROC-ES2L, MIROC6 and MRI-ESM2-0) under the Shared Socio-Economic Pathway 2-4.5 scenario (Eyring et al., 2016). Data for annual precipitation were log_{10}(x + 1) transformed, and all future climate projections were scaled using the data ranges for the 1970–2000 climatologies. We then extracted 125 sets of 5,329 random locations for each of the eight models (1,000 sets in total). For each of the 1,000 sets, we report the mean value, the range and the 95% confidence interval estimated using the 2.5th and 97.5th percentiles.

We next estimated four-dimensional hypervolumes (Blonder et al., 2014, 2018) for the empirical dataset of 5,329 data locations (‘studied environments’), and for the 1,000 randomly generated datasets of 5,329 random locations (‘random environments’), using the hypervolume function of the R package hypervolume v 2.0.12. Hypervolumes are characterised by n axes that constitute an n-dimensional Euclidean space, from which a geometrical shape can be defined and used to describe the size, position and geometry of the system (Blonder et al., 2014, 2018). The hypervolume concept was introduced by Hutchinson (1957) to describe the range of environments that permit a species to exist, with axes corresponding to environmental factors that affect organismal performance (Holt, 2009). Here we estimate hypervolumes that describe the range of environments occupied by species that have been studied to date, as well as for the environments available to species at present and in the future, with axes corresponding to measures of the mean and

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Summary of the traits for which data were compiled, and the distribution across major taxonomic groups. Note that the total number of records is larger than the total number of species, because data are available for more than one trait for some species</th>
</tr>
</thead>
<tbody>
<tr>
<td>acc</td>
<td>CTmax</td>
</tr>
<tr>
<td>Actinopterygii</td>
<td>137</td>
</tr>
<tr>
<td>Amphibia</td>
<td>32</td>
</tr>
<tr>
<td>Arachnida</td>
<td>24</td>
</tr>
<tr>
<td>Aves</td>
<td>0</td>
</tr>
<tr>
<td>Insecta</td>
<td>4</td>
</tr>
<tr>
<td>Mammalia</td>
<td>0</td>
</tr>
<tr>
<td>Reptilia</td>
<td>34</td>
</tr>
<tr>
<td>Total</td>
<td>231</td>
</tr>
</tbody>
</table>

Abbreviations: acc, acclimation capacity (Seebacher et al., 2015); CTmax, critical thermal maximum (°C); CTmin, critical thermal minimum temperature (°C); DGC, Discontinuous gas exchange characteristics (White et al., 2007b), which are related to fitness (Schimpf et al., 2012); FMR, Field metabolic rate (kJ/day); MMR, Metabolic rate at maximum temperature (kJ/day); T.LL, Lower lethal temperature (°C); T.UL, Upper lethal temperature (°C); Topt_d, Optimum temperature for development (°C); Topt_g, Optimum temperature for growth (°C); Topt_l, Optimum temperature for locomotion (°C); Tsc, Supercooling point (°C) and an indication of freezing tolerance (FT) or freeze avoidance (FA).
variability of temperature and precipitation. A hypervolume can also be described as a probability distribution over these axes, which can be used to create a geographical suitability map based on a set of input climatologies and the probability density function within the hypervolume (Blonder et al., 2018).

We then quantified the discordance between the current climate at studied environments, and the current or future climates at random environments. Discordance was determined by calculating the proportion of the hypervolume of studied environments that falls outside of the hypervolume of the random environments, and vice versa, using the `hypervolume_overlap_statistics` function. A null expectation for discordance between hypervolumes was generated by comparing 1,000 pairs of random four-dimensional hypervolumes, each generated using 5,329 random locations. On average, discordance between randomly generated hypervolumes within a pair was 6.8% (range: 3.6%–15.4%; 95% CI: 4.3%, 11.8%) of the total hypervolume.

Finally, we calculated a geographical suitability map by projecting the four-dimensional hypervolume for current environments onto the 1970–2000 climatology, and onto each of the eight future climate projections, to visualise the extent to which our existing knowledgebase is applicable to the current and future global range of environments. This was achieved using the `hypervolume_project` function.

**FIGURE 2** Geographical suitability maps depicting the congruence between the climate at studied locations and at (a) current and (b) future global environments. Suitability scores (probability densities) near zero indicate locations with climates that are poorly represented by the climates at studied locations; higher suitability scores indicated locations with climates that are well represented by studied locations. (a) Estimates of the four-dimensional hypervolume describing the current climates for studied environments projected onto current (1970–2000) global climatologies. (b) Suitability score estimates of the four-dimensional hypervolume describing the current climates for studied environments projected onto future climatologies (CMIP Phase 6) for the year 2050 (average for 2041–2060) under the Shared Socio-Economic Pathway 2–4.5 scenario, averaged over eight global climate models (BCC-CSM2-MR, CanESM5, CNRM-CM6-1, CNRM-ESM2-1, IPSL-CM6A-LR, MIROC-ES2L, MIROC6 and MRI-ESM2-0). Four-dimensional hypervolumes were constructed using data for mean annual temperature, total annual precipitation, the standard deviation of annual temperature and the coefficient of variation of annual rainfall.
3 | RESULTS

Studied environments exhibit a profound spatial bias (Figure 1). A mean of 3.2% (range: 2.0%–4.8%, 95% CI: 2.4%, 4.2%) of the hypervolume describing the studied environments fell outside of the hypervolume describing the randomly selected current environments. Based on comparison of 95% CIs, this is significantly less than expected by chance, presumably because of the inclusion of a small number of studies that each intensively sample a small number of sites (e.g. Bushuev et al., 2018; Londoño et al., 2015). A mean of 85.9% (range: 84.9%–86.7%, 95% CI: 85.2%, 86.7%) of the hypervolume describing randomly selected current environments fell outside of the hypervolume describing the studied environments, which is significantly more than expected by chance.

A mean of 4.8% (range: 3.6%–6.4%, 95% CI: 4.0%, 5.7%) of the hypervolume describing the studied environments fell outside of the hypervolume describing the randomly selected future environments, which is not significantly different from that expected by chance. A mean of 85.0% (range: 84.4%–85.6%, 95% CI: 84.4%, 85.6%) of the hypervolume describing randomly selected future environments fell outside of the hypervolume describing the studied environments, which is significantly more than expected by chance.

Projections of the studied environment hypervolume onto current (Figure 2a) and projected future (Figure 2b) climatologies revealed strikingly similar patterns. In both cases, the geographical suitability maps exhibit regions with near zero suitability score (probability density) estimates, indicating that large regions of the surface have climates that are not encapsulated by studied environments.

4 | DISCUSSION

Our analysis confirms that the studied environments exhibit a profound spatial bias (Figure 1), and that the environments from which physiological data were obtained represent just a subset of the range of predicted future climates. These findings are significant because current knowledge of physiological traits is used to predict the impact of climate change (Diamond, Sorger, et al., 2012; Duarte et al., 2012; Janion-Scheepers et al., 2018; Pinsky et al., 2019), but this knowledge is derived from species living in environments that are not representative of many current and predicted future environments (Figure 2).

Many of the regions that are not represented by the studied environments are at high latitude (e.g. Antarctica and Greenland), but there are also significant data-poor expanses elsewhere, including the Saharan and Sahel regions of Africa, parts of the Andes, the northmost regions of north America and much of western Europe and south-east Asia (Figure 2). A surprising point of distinction between the maps of geographical (Figure 1) and environmental (Figure 2) coverage is that although the tropics are in general rather poorly sampled (Figure 1), environmental coverage through this region is nearly complete (Figure 2). Projections of the climate data from studied environments onto current (Figure 2a) and future (Figure 2b) climates are strikingly similar, suggesting that the major source of uncertainty is not derived from the difference between current and future climates, but instead arises from our poor understanding of animals from many geographical regions of the world (Figures 1 and 2). The responses of animals from unstudied environments to future conditions are difficult to predict, because they may have trait values, including capacities for trait plasticity, that differ in systematic but currently unknown ways from the species that have so far been studied.

In recent years it has become clear that the effects of environmental stressors are typically interactive, such that the response to multiple stressors can be greater or lesser than the additive combination of the individual effects (Alton & Franklin, 2017; Crain et al., 2008). Yet the vast majority of assessments of climate change risk consider only a single stressor (Kauhanto et al., 2016; O’Brien et al., 2019; Rosenblatt & Schmitz, 2016), and most broad-scale studies of the effect of climate on functional traits test only for additive effects, although there are exceptions (Naya et al., 2018; White, Blackburn, Terblanche, et al., 2007). Extrapolation from our current knowledge base to predict the outcomes of climate change will be challenging because predictions are required for animals that live in environments with combinations of environmental conditions that have not yet been studied.

The recognition of geopolitical biases in epidemiology led to the formulation of specific strategies to redress this bias (Jones et al., 2008), and recent work has proposed strategies to overcome spatial biases in conservation research (Reboredo Segovia et al., 2020). We suggest that a similar effort is necessary to understand where the impacts of climate change will be most profound, and where conservation efforts will be most necessary. Expansion of data compilations to include the non-Anglophone literature would be a useful first step, followed by efforts to measure species from regions with climates that are not represented in the existing data (Figure 2).

ACKNOWLEDGEMENTS

This research was funded by the Australian Research Council (FT130101493, DP170101046, DP170101114, DP180103925) and the National Research Foundation of South Africa. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modelling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. We thank the anonymous reviewers who provided constructive feedback on previous versions of the manuscript, and whose thoughtful suggestions greatly improved the work.

AUTHORS’ CONTRIBUTIONS

C.R.W., F.S. and C.E.F. conceived the study; All authors contributed to the methodology; C.R.W. analysed the data and led the writing of the study. All authors contributed critically to the drafts and gave final approval for publication.

DATA AVAILABILITY STATEMENT

All data were collected from published databases cited in the main text.
REFERENCES

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.