

A Latitudinal Gradient of Reference Genomes

Ethan B. Linck and Carlos Daniel Cadena

Accessibility Disclaimer:

For a more accessible version of this document, please submit an accessibility request form through the Montana State University Library website.

ORIGINAL ARTICLE OPEN ACCESS

A Latitudinal Gradient of Reference Genomes

Ethan B. Linck¹  | Carlos Daniel Cadena² ¹Department of Ecology, Montana State University, Bozeman, Montana, USA | ²Departamento de Ciencias Biológicas, Universidad de los Andes, Bogotá, ColombiaCorrespondence: Ethan B. Linck (ethan.linck@montana.edu)

Received: 7 July 2024 | Revised: 17 September 2024 | Accepted: 26 September 2024

Handling Editor: Martin Kardos

Funding: The authors received no specific funding for this work.

Keywords: biodiversity genomics | conservation genetics | macroecology | tetrapods

ABSTRACT

Global inequality rooted in legacies of colonialism and uneven development can lead to systematic biases in scientific knowledge. In ecology and evolutionary biology, findings, funding and research effort are disproportionately concentrated at high latitudes, while biological diversity is concentrated at low latitudes. This discrepancy may have a particular influence in fields like phylogeography, molecular ecology and conservation genetics, where the rise of genomics has increased the cost and technical expertise required to apply state-of-the-art methods. Here, we ask whether a fundamental biogeographic pattern—the latitudinal gradient of species richness in tetrapods—is reflected in the available reference genomes, an important data resource for various applications of molecular tools for biodiversity research and conservation. We also ask whether sequencing approaches differ between the Global South and Global North, reviewing the last 5 years of conservation genetics research in four leading journals. We find that extant reference genomes are scarce relative to species richness at low latitudes and that reduced representation and whole-genome sequencing are disproportionately applied to taxa in the Global North. We conclude with recommendations to close this gap and improve international collaborations in biodiversity genomics.

1 | Introduction

Scientific knowledge reflects both reality and the economic and social conditions that shape the practice of research (Hull 1990; Logino 2019; Stephan 2012). The coincidence of the Scientific Revolution and imperial expansion in Western Europe in the 16th through 18th centuries led to the accumulation and concentration of global capital in institutions of higher learning in the United Kingdom, France, Spain, the Netherlands and Germany, funding foundational investigations in the natural sciences. Industrialisation and the rise of capitalism in the 19th century stoked the boiler of the research endeavour, permitting the continued expansion of universities and birthing a learned aristocratic class with disposable time and income to commit to natural history (Opitz 2004, 2006). This legacy persists to this

day: Research expenditures concentrated in a handful of majority Anglophone countries (May 1997), which in turn produce a disproportionate amount of scholarship (May 1997; King 2004) and often drive citation networks (Pasterkamp et al. 2007; Meneghini, Packer, and Nassi-Calo 2008).

Ecology and evolutionary biology (EEB) face a unique challenge in light of global scientific inequality, as while funding and research effort are disproportionately concentrated at high latitudes (Melles et al. 2019), biological diversity is concentrated at low latitudes (Willig, Kaufman, and Stevens 2003). The consequences of this discrepancy are many, including systematic biases in research effort (Titley, Snaddon, and Turner 2017) and taxonomy (Freeman and Pennell 2021), major gaps in our understanding of natural history and species distributions (Collen

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Author(s). *Molecular Ecology* published by John Wiley & Sons Ltd.

et al. 2008; Feeley and Silman 2011) and a blinkered view of patterns of diversity and diversification across the tree of life (Reddy 2014; Cornwell et al. 2019). Increased research attention on tropical ecosystems has therefore been a stated priority for decades (Collen et al. 2008; Nori, Loyola, and Villalobos 2020), but despite some promising strides towards international collaborations (Perez and Hogan 2018), partners in the Global North frequently set research agendas (Bradley 2008; Asase et al. 2022), leading to accusations of parachute science and persistent gaps in data quantity, quality and type (Stocks et al. 2008; Soares et al. 2023).

EEB disciplines where data collection and analysis are equipment- and cost-intensive are especially likely to show increased discrepancies between the Global North and Global South. In the 1960s and 1970s, assays of molecular genetic markers in non-model organisms laid the groundwork for phylogenetic systematics to shift away from a traditional focus on morphology, increasing statistical power and requiring increasingly expensive tools and specialised skills (Hillis, Moritz, and Mable 1996). Concurrently, the rise of phylogeography brought an evolutionary perspective to population biology, altering views on the fundamental units of management and conservation (Avice 2000). In the late 2000s and early 2010s, reduced representation of high-throughput sequencing using library preparation methods such as RADseq and target capture began to proliferate as an alternative to traditional Sanger sequencing (Ekblom and Galindo 2011; Lemmon and Lemmon 2013), boasting the advantages of rapidly generating much larger datasets and requiring little to no prior information about a focal taxon's genome.

A third revolution is currently underway, as low-coverage whole-genome resequencing (hereafter WGS) begins to supplant reduced representation methods in some fields and taxa (Ellegren 2014; Toews et al. 2016). Unlike RADseq, target capture and related approaches, WGS typically requires a high-coverage reference genome with which to align samples. Because sequencing an individual at a sufficient depth remains costly—and because assembling such large data is computationally intensive—generating high-quality novel reference genomes remains difficult for many labs and, by extension, species.

The problem may be particularly acute in conservation biology where decades of concern about the so-called 'conservation genetics gap' have highlighted a persistent disconnect between the importance managers place on genetic information and its actual use in decision-making (Taylor, Dussex, and van Heezik 2017). Hypotheses to explain the conservation genetics gap include scepticism about its importance, the specialised knowledge required for analysis and interpretation and cost (Hoban et al. 2013), though some recent discussion suggests the term itself obfuscates by lumping the diverse set of obstacles (or 'spaces') between research and implementation (Toomey, Knight, and Barlow 2017; see solutions reviewed by Hogg 2024). At low-resource institutions in the tropics, a broader shift towards WGS approaches may further discourage the development of the field in the very regions where the greatest number of critically endangered species is found (Vamossi and Vamossi 2008; Bertola et al. 2024), especially if technological advances become publication requirements at high-impact and broadly read journals.

Here we ask whether a fundamental biogeographic pattern—the latitudinal gradient of species richness in tetrapods—is reflected in reference genomes available through the US National Institute of Health's National Center for Biotechnology Information Genome Browser, an important data resource for various applications of molecular tools for biodiversity research and conservation. We hypothesized that NCBI contributor affiliations combined with inequities in economic development and access to scientific resources would lead the number of species with assembled genomes to be the greatest in the temperate zone, not the tropics. We also ask whether sequencing approaches differ between the Global South and Global North, reviewing the last 5 years of conservation genetics research in four leading English language journals. We conclude by discussing strategies to improve international collaborations in biodiversity genomics and boost the representation of species from the Global South in sequence databases.

2 | Methods

2.1 | Reference Genomes, Species Richness and Latitude

We used the National Center for Biotechnology Information (NCBI) Datasets command-line tools v.16.19.0 (O'Leary et al. 2024) to download taxonomy metadata for the subset of species with an assembled reference genome in the following taxa: Birds (Class: Aves), mammals (Class: Mammalia), squamates (Order: Squamata), amphibians (Class: Amphibia), turtles (Order: Testudines), crocodilians (Order: Crocodylia) and tuataras (Order: Rhynchocephalia). We selected these groups—together comprising extant tetrapods—to provide a snapshot of animal diversity in relatively well-studied clades with different ecologies and evolutionary histories, while restricting the total dataset to a computationally manageable size. From this initial list, we retained species with an exact match to the Global Biodiversity Information Facility's (GBIF) Backbone Taxonomy using rgbif v.3.8.0 (Chamberlain et al. 2024) and downloaded all observations of each backed by georeferenced voucher specimens in natural history museum collections (NHCS), excluding those without coordinates and those flagged for geospatial issues ($n = 3,006,946$). We repeated this process for all species in each higher-level taxon represented in our list of reference genomes (i.e., downloaded metadata for all georeferenced tetrapod specimens on GBIF; $n = 9,303,258$). DOIs for each download are available in the References section (GBIF.org 2024a, 2024b, 2024c, 2024d, 2024e, 2024f, 2024g, 2024h).

Filtering these aggregated datasets to contain only species with 10 or more specimen records, we generated convex hull polygons for each as a coarse approximation of their geographic distribution using the R package sf v.1.0–16 (Pebesma 2018; Pebesma and Bivand 2023). Overlaying these on a shapefile of Earth's landmasses from rnaturalearth v.1.0.1 (Massicotte and South 2024), we calculated species richness as the number of overlapping convex hulls in $2^\circ \times 2^\circ$ grid cells, statistically standardising this value by subtracting the observed mean global species richness and dividing by its standard deviation. We subtracted the number of species with reference genomes from the total species richness to determine the regions with

the largest representation gap in genomic resources, again standardising the difference. To assess the significance and slope of the correlation between species richness and the absolute value (or modulus) of latitude in decimal degrees, we performed simple linear regressions in R v.4.4.0 (R Core Team 2024), analysing species with reference genomes and our full dataset separately.

2.2 | Literature Review

To evaluate how the geography of authorship might impact the sequencing strategy of studies in conservation biology, we performed a restricted Web of Science literature search on 29 June 2024 for English language conservation genetics papers published in the last 5 years in the journals *Conservation Genetics*, *Molecular Ecology*, *Journal of Heredity* and *Conservation Biology*, selected for frequently publishing empirical work on non-model organisms. We used the queries ‘SO=“Conservation Genetics” and ‘SO=“Molecular Ecology” OR “Journal of Heredity” OR “Conservation Biology”) AND (TS=“Conservation Genet*” OR KP=“Conservation Genet*” OR TI=“Conservation Genet*”), excluding reviews, genome announcements, meta-analyses, preprints and studies that were purely simulations. Our criteria aimed to achieve a tractable sample size for a careful study (< 1000 papers) while covering the period in which WGS became commonly used for the conservation genetics of non-model organisms (Fuentes-Pardo and Ruzzante 2017; Hohenlohe, Funk, and Rajara 2021).

We then manually reviewed each study, first assigning the home institution of its first and last author to the Global North, Global South, or both (i.e., joint affiliations) using the 2024 UN Trade and Development Classifications. Because the number of middle authors varied widely across our sample, we assessed their affiliations on a binary basis, indicating only whether a contributor from an institution from the Global South was present outside of the lead and senior positions. Synthesising these data, we assigned papers to mutually exclusive groups based on whether they included one or more Global South authors or only Global North authors. Next, we categorised each study's sequencing approach as reduced representation, WGS, Sanger sequencing, microsatellites or other and described its overall focus using tiered categories based on the discussion by Bertola et al. (2024). These tiers were: (1) Taxonomy/systematics, identification or sexing; (2) phylogeography/population genetic structure, estimating genetic diversity and inferring demographic history and (3) detecting outlier loci, quantifying runs of homozygosity and evaluating adaptive potential. When studies employed more than one sequencing approach or addressed goals belonging to multiple tiers, we assigned them to a single category based on their most data-intensive method or question. To explore geographic patterns in the sequencing effort, we assessed whether each study's taxonomic sampling included (1) at least one species distributed in the Global South and (2) at least one species distributed in the Global North. Because some studies included multiple taxa and some species are broadly distributed or migrate between regions, these categories were not mutually exclusive. To evaluate whether geographic representation in conservation genetics changed over the period

covered by our review, we performed logistic regression using the stats package R v.4.4.0, treating the presence or absence of an author from the Global South as a binary outcome variable and year as the sole independent variable.

3 | Results

3.1 | Sampled Gradients of Species Richness

Our list of tetrapod reference genomes from NCBI included 1159 bird species, 795 mammal species, 123 amphibian species, 39 turtle species, 6 crocodile species and the tuatara (*Sphenodon punctuatus*). This total represented 6.8% of the 30,832 tetrapod species with a georeferenced preserved specimen record on GBIF. Species with a reference genome were associated with 3,048,136 specimens or 40.7% of the 7,478,867 tetrapod specimens with available metadata. Following filtering out species with 10 or fewer observations, we retained 1859 species with a reference genome or 8.6% of the 21,583 tetrapod species meeting the same criteria.

The minimum tetrapod species richness in both datasets was one. The maximum species richness calculated from the reference genome dataset was 705, occurring in a grid cell centred on south Florida, USA (Figure 1C). The maximum species richness in the full dataset was 2698, occurring in a grid cell centred on the western Amazon basin and east slope of the Andes in Ecuador near the Peruvian border (Figure 1C); this was also where the greatest gap between sequenced species and total species richness was observed (Figure 1A).

Species richness was negatively related to the absolute value latitude in both regressions, albeit with a much steeper slope when data from all tetrapod species were included (reference genomes only, $\beta = -4.892$, adjusted $R^2 = 0.6439$; $p < 0.001$; full data, $\beta = -18.65$, adjusted $R^2 = 0.7856$ $p < 0.001$) (Figure 1B). Because data visualisation indicated there might be a distinct breakpoint in the relationship at midlatitudes—potentially reflecting a transition from the influence of sampling effort effects to true biogeographic signal—we fit an additional piecewise linear regression model with the R package segmented v.2.1–0. This model identified a breakpoint at 39.819, fitting a segment with a slope of $\beta = -0.312$ before it and a segment with a slope of $\beta = -7.3634$ after it ($p = 0.0408$; adjusted $R^2 = 0.7184$).

3.2 | Literature Review

After excluding papers that did not meet our stated criteria, we reviewed 394 empirical conservation genetics articles published between 1 January 2020 and 29 June 2024. This list included 345 papers from the journal *Conservation Genetics*, 11 papers from *Journal of Heredity*, 35 articles from *Molecular Ecology* and 3 articles from *Conservation Biology*. Of these, 62 included a first or senior author from the Global South; 40 more included Global South scientists as middle authors. Nearly 87% ($n = 342$) of papers had a first or senior author from the Global North. Ninety-eight included sampling from a focal taxon or focal taxa in the Global South, with 277 sampling a taxon or taxa from the Global North. From 2020 to 2024, the odds of a paper including

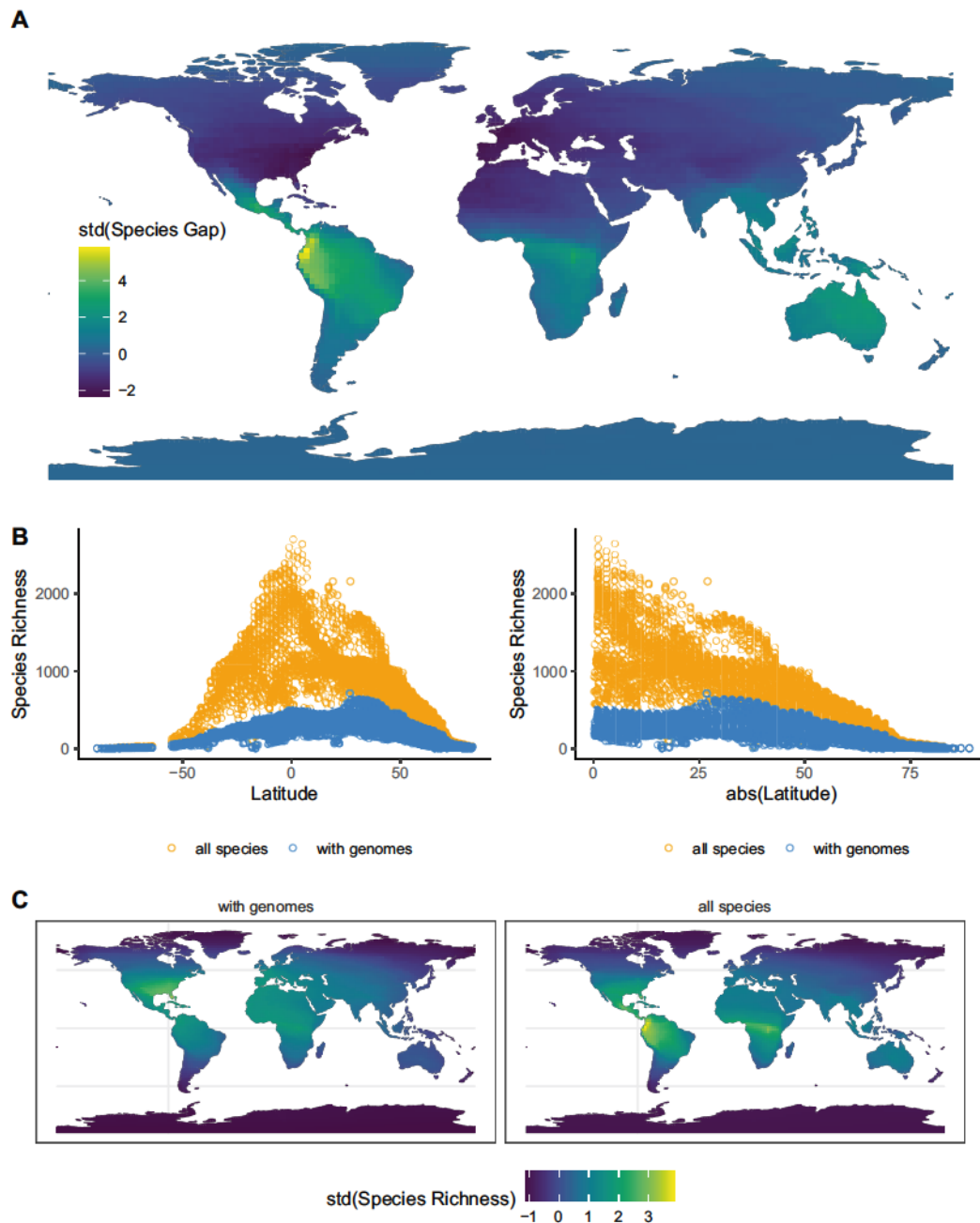


FIGURE 1 | (A) Tropical species are underrepresented among available reference genomes. Colours reflect the standardised difference between total species richness and the number of species with an assembled reference genome on the NCBI Genome Browser. Negative values indicate grid cells falling below calculated global mean species richness. (B) Richness among species with reference genomes does not reflect the global patterns of biodiversity. Gold circles represent total local species richness in 2° by 2° grid cells, while blue circles represent richness of species with assembled reference genomes. (C) Global patterns of species richness calculated from species with reference genomes and all species with NHC specimen records on GBIF, respectively.

an author from the Global South did not significantly increase (log-odds, 1.1096; 95% CI, -0.5864 to 0.2667; $p=0.210$). Microsatellites were the most used sequencing strategy among Global South authors and in studies of Global South taxa, while reduced representation genome sequencing approaches were most common in the Global North. For authors and taxa in both the Global South and the Global North, Tier 2 studies (phylogeography/population genetics structure, estimating genetic diversity or inferring demographic history) were the most common. Further details are provided in Tables 1 and 2.

4 | Discussion

Reference genomes available through the US National Institute of Health's National Center for Biotechnology Genome Browser fail to reflect the overwhelming concentration of tetrapod species richness in the tropics and are strongly biased towards species at midlatitudes in the Northern Hemisphere (Figure 1). This pattern is almost certainly a result of global inequalities in economic development and its resulting effects on research productivity (May 1997; King 2004). Its consequences will likely

TABLE 1 | Summary of regional authorship affiliation, sequencing strategy and sampled focal species range for empirical conservation genetics papers from 2019 to 2024 in four leading journals.

Sequencing strategy	Global South author	Only Global North authors	Global South taxon	Global North taxon
Sanger	22 (0.0558)	25 (0.0634)	26 (0.0660)	26 (0.0659)
Microsatellites	43 (0.1066)	105 (0.2690)	45 (0.1142)	108 (0.2741)
Reduced representation	29 (0.0736)	129 (0.3299)	31 (0.0787)	131 (0.3325)
WGS	5 (0.0013)	24 (0.0609)	11 (0.0152)	18 (0.0279)
Other	4 (0.0101)	8 (0.0203)	4 (0.0102)	8 (0.0203)

Note: Integers indicate the total number of studies in each category, while numbers in parentheses refer to its proportion out of all reviewed articles ($n = 394$). Regional authorship affiliation is mutually exclusive, while the distribution of focal taxa is not, as studies could involve multiple taxa, broadly distributed taxa or long-distance migrants. Papers were assigned to a sequencing strategy based on the most data-intensive approach they employed (i.e., a study applying both Sanger sequencing and microsatellites would be assigned to the 'Microsatellites' category).

TABLE 2 | Summary of regional authorship affiliation, study goals and sampled focal species range among reviewed papers.

Study goals	Global South author	Only global North authors	Global South taxon	Global North taxon
1. Taxonomy/systematics, identification or sexing	9 (0.0228)	19 (0.0635)	11 (0.0279)	17 (0.04315)
2. Phylogeography/population genetics structure, estimating genetic diversity and inferring demographic history	90 (0.2284)	241 (0.6192)	102 (0.2589)	248 (0.6294)
3. Detecting outlier loci, quantifying runs of homozygosity and evaluating adaptive potential	4 (0.0106)	31 (0.0786)	4 (0.0102)	31 (0.0787)

Note: Study goals refer to broad tiers of research questions with increasing data requirements. Papers were assigned to each on the basis of the most data-intensive analysis they employed (i.e., a paper inferring population genetics structure and identifying loci under selection would be assigned to tier 3). As in Table 1, 'Global South author' and 'Only Global North authors' are mutually exclusive categories, while 'Global South taxon' and 'Global North taxon' are not.

include increasing an already profound methodological gap in sequencing approaches between molecular ecologists in the Global North and the Global South (Table 1).

Our analysis contrasts patterns inferred from both a traditional source of biodiversity datavouchered specimens in NHCs—and a contemporary genomic resource archive, the National Institute of Health's NCBI Genome Browser. Our description of the latitudinal gradient of total tetrapod species richness is broadly similar to other recent macroecological studies of the phenomenon (Roll et al. 2017; Quintero et al. 2023), with global hotspots concentrated in northwest South America and Central Africa (Figure 1C) and an approximately monotonic decline in latitudinal species richness maxima from the equator to the poles. In contrast, the latitudinal gradient of richness of tetrapod species with reference genomes is flattened, showing only moderate declines at high latitudes and a midlatitude peak in species richness in the Northern Hemisphere (Figure 1B).

This difference is especially notable because we made no effort to correct for disparities in historical specimen collection across the latitude, with the consequence that our 'true' species richness gradient significantly underestimates biodiversity in the tropics. Across the longitude, our analysis appears to underestimate diversity in East Asia, Indonesia and Oceania (Quintero et al. 2023), likely due to both the coarse grain of our study and the region's

greater distance from the large NHCs in Europe and North America that are the backbone of curated GBIF data. Geopolitical factors complicate a simple interpretation of these results: Use of GBIF and NCBI by scientists outside the Anglosphere may be limited by data sovereignty laws or other impacts of economic and military rivalries. Regardless, species with publicly available reference genomes as of July 2024 are more reflective of socioeconomic conditions than biogeographic reality.

If both natural history collections and contemporary bioinformatics resources reflect historical inequalities in development and scientific capacity, why do data from the former better approximate the latitudinal gradient of species richness? Part of the answer lies in their different goals: While the mission of many NHCs is to explicitly catalogue and archive regional or global biodiversity, the NCBI Genome Browser is typically used as a repository for open data publication requirements and is less often an end unto itself. Another possibility is that researchers looking to obtain genetic material from the tropics face unique obstacles to their work. As the birth and golden age of NHCs coincided with the heyday of Western colonialism (De Vos 2007; Quintero Toro 2012), scientific collection was less restricted by concerns of either sovereignty or Indigenous land tenure, let alone local or international regulations on the transfer of biological samples (Khan and Tyagi 2021; Sherkow et al. 2022). Relatedly, one early reader of this manuscript

pointed to the added challenge of obtaining high-quality tissue samples in hot, humid conditions.

We do not believe access to sources of DNA is a major obstacle: Science is rich with examples of researchers overcoming far greater hurdles to their work, and extant frozen tissue collections at institutions in both the Global South and Global North include thousands of samples awaiting sequencing. Instead, we suggest that evolving norms and persistent obstacles of cost have led many relatively well-resourced scientists in the Global North to prioritise generating reference genomes for local taxa. In spite of rapid declines in the per base-pair cost of whole-genome sequencing (Lou et al. 2021), high-coverage sequencing remains a significant expense: While averages are hard to come by in this increasingly privatised sector, one of us recently paid ~\$13,000 US (September 2023) for high-depth long-read and HiC sequencing of a North American passerine bird from an industry contractor. Even in North America, this figure likely pushes small, single PI labs to prioritise investing in generating resources for species at the center of their research program, or otherwise likely to provide long-term utility—which are often those in their own backyards.

In the spirit of the traditional mission of NHCs, the past decade has seen several interrelated, international initiatives to increase taxonomic diversity in high-quality reference genomes (O'Brien, Haussler, and Ryder 2014; Koepfli et al. 2015; Cheng et al. 2018; Rhie et al. 2021). Of these initiatives, the Earth BioGenome Project (EBP)—a 'moonshot' endeavour to sequence all Eukarya within 10 years—is by design the most ambitious, bringing its smaller-scale predecessors together under a common set of goals and standards (Lewin et al. 2018; Lewin et al. 2022). These collaborations have had a profound impact on biodiversity genomics: As of March 2021, EBP estimated at least one species from 15.6% of all Eukaryotic families that had been sequenced, marking notable progress towards its Phase I goal of complete family-level representation within 3 years. Though this focus on phylogenetic diversity has introduced its own bias by undersampling lineages with high diversification rates, the initiative has nonetheless helped to close what was surely an even larger gap in the empirical patterns of species richness and species richness represented by NCBI. In some cases, EBP or the initiatives under its umbrella have also provided researchers with early access to draft assemblies of nonmodel organisms (Linck, Freeman, and Dumbacher 2020) or with support and resources to produce new assemblies (Cadena et al. 2024), a scenario that suggests the resource gap may be slightly less dire in practice than that reported here.

Yet, in a world of limited time, finite resources and incompletely described biodiversity, sequencing at scale is not immune to its own biases. For example, a recent publication introducing an attempt to generate reference genomes for all vertebrates (Rhie et al. 2021) included 127 authors affiliated with 102 institutions. Of these, only 13 are in the Global South: 4 in China, 3 in Korea, 2 in Malaysia, 1 in Singapore, 1 in Qatar and 1 in Colombia. No authors or institutional affiliations from Africa, Indonesia or Oceania outside of Australia and New Zealand were included. To date, institutions in the Global North have generated most animal (Hotaling, Kelley, and Frandsen 2021) and plant (Marks et al. 2021) genomes available on NCBI's GenBank, a

situation mirrored by an analysis of squamates alone by Pinto et al. (2023). Though understandable considering the current distribution of scientists and resources—though not of species—the imbalance seems likely to perpetuate representational biases in the near-term.

Current efforts by international collaborations like the Amphibian Genomics Consortium to increase geographic representation and to offer support and opportunities to researchers from developing countries and underrepresented groups (Kosch et al. 2024) are steps in the right direction in the path to make the field of biodiversity genomics more equitable. We are particularly enthusiastic about EBP's model of pairing targeted sequencing efforts with capacity building workshops (see Sharaf et al. 2024 for a recent example from 11 African countries). In line with the Convention on Biological Diversity's Nagoya Protocol (Secretariat of the Convention on Biodiversity 2011), another critical dimension of the conversation about equitable generation of encyclopedias of reference genomes is the need for researchers to build strong partnerships with Indigenous people and other local communities, allowing them to participate in and benefit from the different phases and products of sequencing projects (Ambler et al. 2021; Colella et al. 2023; Mc Cartney et al. 2023).

If tropical species are underrepresented on NCBI, we would expect that they are infrequently studied using whole-genome resequencing (and other sequencing strategies dependent on a reference genome). Our review of conservation genetics papers published in *Molecular Ecology*, *Journal of Heredity*, *Conservation Genetics* and *Conservation Biology* over the last 5 years suggests this is indeed the case. While the Global South/Global North binary and measures of human development more generally are only imperfectly correlated with latitude, their association nonetheless indicates that WGS is only rarely applied in conservation genetics studies in the tropics ($n = 6$) and almost never by a leading researcher with a primary affiliation to a research institution in the region ($n = 1$) (Table 1). Correspondingly, study goals that are highly data-intensive—such as outlier detection or quantifying runs of homozygosity—are much more commonly pursued in papers written by authors from the Global North (Table 2).

Overall, microsatellites remain the most common molecular approach for Global South scientists and for Global South focal taxa, for scientists and focal taxa in the Global North, reduced representation approaches dominate. We believe this is reflective of the expense and limited availability of high-throughput sequencing in the Global South, regardless of whether reads are assembled de novo or aligned to a reference. Lastly, we point out that across all categories, the vast majority of scientists (87%) and species (77%) in our sample originate in the Global North. Though our choice of North American or European-based journals precludes generalisation, the primacy of the English language in global science (Montgomery 2013) leads us to interpret this as indicating the field of conservation genetics—let alone conservation genomics—is in its infancy in the tropics. We thus doubt the patterns described here would change appreciably if we were to include publications in other languages. Setting appropriate goals, targets and indicators to effectively conserve and monitor global genetic diversity will require this situation to be remedied (Hoban et al. 2021) (Table 2).

We highlight discrepancies between available reference genomes and global biogeographic patterns to encourage increased, equitable collaboration between scientists in the Global North and Global South (see Sawchuk et al. 2024 for an African-based perspective on equity and inclusion in population genetics and ancient DNA studies). In light of this, we make three simple recommendations (see also Bertola et al. 2024). First, we encourage scientists from resource-rich institutions to consider allocating effort and funds towards generating reference genomes that serve the needs of managers and researchers in the Global South. Second, we support the continued development of multinational sequencing projects but ask funders and senior personnel to increasingly consider prioritisation, inclusion and capacity building in areas of the world with rich biodiversity and limited resources to study it using genomic tools. Third, we ask journals to consider the issues of access and cost in editorial guidelines and decisions: While high-throughput sequencing is increasingly expected by editorial boards and reviewers at high-impact journals, it is not essential to address a variety of research questions (Bertola et al. 2024) and remains out of reach for most scientists residing where most of the world's species occur. If ecology, evolution and conservation aim to accurately catalogue and effectively protect life on Earth, remedying inequalities in genomic resources should be a major priority.

Author Contributions

E.B.L. and C.D.C. jointly designed the study and conducted the literature review. E.B.L. analyzed data and wrote the manuscript with contributions from C.D.C.

Acknowledgements

We thank Marty Kardos for the invitation to participate in this special issue. Carolyn Hogg and an anonymous reviewer provided helpful feedback on the study. E.B.L. expresses appreciation for his mobile phone's Wi-Fi hotspot, which was essential for submitting the initial draft of this manuscript from the field. Ongoing genomic work by C.D.C. is supported by the Facultad de Ciencias at Universidad de Los Andes through a Programa de Investigación (Convocatoria 2024–2026).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Processed GBIF datasets, NCBI metadata and a digital notebook containing code to perform these analyses and generate Figure 1 are available at https://github.com/elinck/lat_grad_genome and from Data Dryad (<https://doi.org/10.5061/dryad.2v6wwpzhx>).

References

- Ambler, J., P. K. Dearden, P. Wilcox, M. Hudson, and N. Tiffin. 2021. "Including Digital Sequence Data in the Nagoya Protocol Can Promote Data Sharing." *Trends in Biotechnology* 39, no. 2: 116–125.
- Asase, A., T. I. Mzumara-Gawa, J. O. Owino, A. T. Peterson, and E. Saupe. 2022. "Replacing "Parachute Science" With "Global Science" in Ecology and Conservation Biology." *Conservation Science and Practice* 4, no. 5: e517.
- Avice, J. C. 2000. *Phylogeography: The History and Formation of Species*. Cambridge, MA: Harvard University Press.

- Bertola, L. D., A. Brüniche-Olsen, F. Kershaw, I. R. M. Russo, A. J. MacDonald, and P. Sunnucks. 2024. "A Pragmatic Approach for Integrating Molecular Tools Into Biodiversity Conservation." *Conservation Science and Practice* 6, no. 1: e13053.
- Bradley, M. 2008. "On the Agenda: North–South Research Partnerships and Agenda-Setting Processes." *Development in Practice* 18, no. 6: 673–685.
- Cadena, C. D., L. Pabón, C. DoNascimento, L. Abueg, T. Tilley, and B. O-Toole. 2024. "A Reference Genome for the Andean Cavefish *Trichomycterus rosablanca* (Siluriformes, Trichomycteridae): Building Genomic Resources to Study Evolution in Cave Environments." *Journal of Heredity* 115, no. 3: 311–316.
- Chamberlain, S., V. Barve, D. Mcglinn, et al. 2024. "rgbif: Interface to the Global Biodiversity Information Facility API. R package version 3.8.0." <https://CRAN.R-project.org/package=rgbif>.
- Cheng, S., M. Melkonian, S. A. Smith, S. Brockington, J. M. Archibald, and P. M. Delaux. 2018. "10KP: A Phylodiverse Genome Sequencing Plan." *GigaScience* 7, no. 3: giy013.
- Colella, J. P., L. Silvestri, G. Súzan, M. Weksler, J. A. Cook, and E. P. Lessa. 2023. "Engaging With the Nagoya Protocol on Access and Benefit-Sharing: Recommendations for Noncommercial Biodiversity Researchers." *Journal of Mammalogy* 104, no. 3: 430–443.
- Collen, B., M. Ram, T. Zamin, and L. McRae. 2008. "The Tropical Biodiversity Data Gap: Addressing Disparity in Global Monitoring." *Tropical Conservation Science* 1, no. 2: 75–88.
- Cornwell, W. K., W. D. Pearse, R. L. Dalrymple, and A. E. Zanne. 2019. "What We (Don't) Know About Global Plant Diversity." *Ecography* 42, no. 11: 1819–1831.
- De Vos, P. S. 2007. "Natural History and the Pursuit of Empire in Eighteenth-Century Spain." *Eighteenth-Century Studies* 40, no. 2: 209–239.
- Eklblom, R., and J. Galindo. 2011. "Applications of Next Generation Sequencing in Molecular Ecology of Non-Model Organisms." *Heredity* 107, no. 1: 1–15.
- Ellegren, H. 2014. "Genome Sequencing and Population Genomics in Non-Model Organisms." *Trends in Ecology & Evolution* 29, no. 1: 51–63.
- Feeley, K. J., and M. R. Silman. 2011. "The Data Void in Modeling Current and Future Distributions of Tropical Species." *Global Change Biology* 17, no. 1: 626–630.
- Freeman, B. G., and M. W. Pennell. 2021. "The Latitudinal Taxonomy Gradient." *Trends in Ecology & Evolution* 36, no. 9: 778–786.
- Fuentes-Pardo, A. P., and D. E. Ruzzante. 2017. "Whole-Genome Sequencing Approaches for Conservation Biology: Advantages, Limitations and Practical Recommendations." *Molecular Ecology* 26, no. 20: 5369–5406.
- GBIF.org. 2024a. "GBIF Occurrence Download." <https://doi.org/10.15468/dl.p4r6hf>. Accessed June 12, 2024 from R via rgbif <https://github.com/ropensci/rgbif>.
- GBIF.org. 2024b. "GBIF Occurrence Download." <https://doi.org/10.15468/dl.njrdnn>. Accessed June 20, 2024 from R via rgbif <https://github.com/ropensci/rgbif>.
- GBIF.org. 2024c. "GBIF Occurrence Download." <https://doi.org/10.15468/dl.pw99uq>. Accessed June 20, 2024 from R via rgbif <https://github.com/ropensci/rgbif>.
- GBIF.org. 2024d. "GBIF Occurrence Download." <https://doi.org/10.15468/dl.jbg6xy>. Accessed June 21, 2024 from R via rgbif <https://github.com/ropensci/rgbif>.
- GBIF.org. 2024e. "GBIF Occurrence Download." <https://doi.org/10.15468/dl.5pkrsd>. Accessed June 21, 2024 from R via rgbif <https://github.com/ropensci/rgbif>.

- GBIF.org. 2024f. "GBIF Occurrence Download." <https://doi.org/10.15468/dl.jd85b2>. Accessed July 5, 2024 from R via rgbif <https://github.com/ropensci/rgbif>.
- GBIF.org. 2024g. "GBIF Occurrence Download." <https://doi.org/10.15468/dl.59eyey>. Accessed July 5, 2024 from R via rgbif <https://github.com/ropensci/rgbif>.
- GBIF.org. 2024h. "GBIF Occurrence Download." <https://doi.org/10.15468/dl.vybgce>. Accessed May 5, 2024 from R via rgbif <https://github.com/ropensci/rgbif>.
- Hillis, D. M., C. Moritz, and B. K. Mable, eds. 1996. *Molecular Systematics*. 2nd ed. Sinauer, Sunderland, MA: Sinauer Associates, Inc.
- Hoban, S., M. W. Bruford, W. C. Funk, P. Galbusera, M. P. Griffith, and C. E. Grueber. 2021. "Global Commitments to Conserving and Monitoring Genetic Diversity Are Now Necessary and Feasible." *Bioscience* 71, no. 9: 964–976.
- Hoban, S. M., H. C. Haufler, S. Pérez-Espona, J. W. Arntzen, G. Bertorelle, and J. Bryja. 2013. "Bringing Genetic Diversity to the Forefront of Conservation Policy and Management." *Conservation Genetics Resources* 5: 593–598.
- Hogg, C. J. 2024. "Translating Genomic Advances Into Biodiversity Conservation." *Nature Reviews Genetics* 25, no. 5: 362–373.
- Hohenlohe, P. A., W. C. Funk, and O. P. Rajora. 2021. "Population Genomics for Wildlife Conservation and Management." *Molecular Ecology* 30, no. 1: 62–82.
- Hotelling, S., J. L. Kelley, and P. B. Frandsen. 2021. "Toward a Genome Sequence for Every Animal: Where Are We Now?" *Proceedings of the National Academy of Sciences* 118, no. 52: e2109019118.
- Hull, D. L. 1990. *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*. Chicago, IL: University of Chicago Press.
- Khan, A., and A. Tyagi. 2021. "Considerations for Initiating a Wildlife Genomics Research Project in South and South-East Asia." *Journal of the Indian Institute of Science* 101, no. 2: 243–256.
- King, D. A. 2004. "The Scientific Impact of Nations." *Nature* 430, no. 6997: 311–316.
- Koepfli, K. P., B. Paten, Genome 10K Community of Scientists, and S. J. O'Brien. 2015. "The Genome 10K Project: A Way Forward." *Annual Review of Animal Biosciences* 3, no. 1: 57–111.
- Kosch, T. A., M. Torres-Sanchez, H. C. Liedtke, K. Summers, M. H. Yun, and A. J. Crawford. 2024. "The Amphibian Genomics Consortium: Advancing Genomic and Genetic Resources for Amphibian Research and Conservation." *bioRxiv*. <https://doi.org/10.1101/2024.06.27.601086>.
- Lemmon, E. M., and A. R. Lemmon. 2013. "High-Throughput Genomic Data in Systematics and Phylogenetics." *Annual Review of Ecology, Evolution, and Systematics* 44: 99–121.
- Lewin, H. A., S. Richards, E. Lieberman Aiden, M. L. Allende, J. M. Archibald, and M. Bálint. 2022. "The Earth BioGenome Project 2020: Starting the Clock." *Proceedings of the National Academy of Sciences* 119, no. 4: e2115635118.
- Lewin, H. A., G. E. Robinson, W. J. Kress, W. J. Baker, J. Coddington, and K. A. Crandall. 2018. "Earth BioGenome Project: Sequencing Life for the Future of Life." *Proceedings of the National Academy of Sciences* 115, no. 17: 4325–4333.
- Linck, E., B. G. Freeman, and J. P. Dumbacher. 2020. "Speciation and Gene Flow Across an Elevational Gradient in New Guinea Kingfishers." *Journal of Evolutionary Biology* 33, no. 11: 1643–1652.
- Logino, H. 2019. "The Social Dimensions of Scientific Knowledge." In *The Stanford Encyclopedia of Philosophy (Summer 2019 Edition)*, edited by E. N. Zalta. Metaphysics Research Lab: Stanford University. <https://plato.stanford.edu/archives/sum2019/entries/scientific-knowledge-social/>.
- Lou, R. N., A. Jacobs, A. P. Wilder, and N. O. Therkildsen. 2021. "A Beginner's Guide to Low-Coverage Whole Genome Sequencing for Population Genomics." *Molecular Ecology* 30, no. 23: 5966–5993.
- Marks, R. A., S. Hotelling, P. B. Frandsen, and R. VanBuren. 2021. "Representation and Participation Across 20 Years of Plant Genome Sequencing." *Nature Plants* 7, no. 12: 1571–1578.
- Massicotte, P., and A. South. 2024. "rnatuarearth: World Map Data from Natural Earth. R Package Version 1.0.1.9000." <https://github.com/ropensci/rnatuarearth>, <https://docs.ropensci.org/rnatuarearth/>, <https://docs.ropensci.org/rnatuarearth/>.
- May, R. M. 1997. "The Scientific Wealth of Nations." *Science* 275, no. 5301: 793–796.
- Mc Cartney, A. M., M. A. Head, K. S. Tsosie, B. Sterner, J. R. Glass, and S. Paez. 2023. "Indigenous Peoples and Local Communities as Partners in the Sequencing of Global Eukaryotic Biodiversity." *npj Biodiversity* 2, no. 1: 8.
- Melles, S. J., C. Scarpone, A. Julien, J. Robertson, J. B. Levieva, and C. Carrier. 2019. "Diversity of Practitioners Publishing in Five Leading International Journals of Applied Ecology and Conservation Biology, 1987–2015 Relative to Global Biodiversity Hotspots." *Ecoscience* 26, no. 4: 323–340.
- Meneghini, R., A. L. Packer, and L. Nassi-Calo. 2008. "Articles by Latin American Authors in Prestigious Journals Have Fewer Citations." *PLoS One* 3, no. 11: e3804.
- Montgomery, S. L. 2013. *Does Science Need a Global Language?: English and the Future of Research*. Chicago, IL: University of Chicago Press.
- Nori, J., R. Loyola, and F. Villalobos. 2020. "Priority Areas for Conservation of and Research Focused on Terrestrial Vertebrates." *Conservation Biology* 34, no. 5: 1281–1291.
- OBrien, S. J., D. Haussler, and O. Ryder. 2014. "The Birds of Genome10K." *GigaScience* 3, no. 1: 32. <https://doi.org/10.1186/2047-217X-3-32>.
- O'Leary, N. A., E. Cox, J. B. Holmes, W. R. Anderson, R. Falk, and V. Hem. 2024. "Exploring and Retrieving Sequence and Metadata for Species Across the Tree of Life With NCBI Datasets." *Scientific Data* 11, no. 1: 732.
- Opitz, D. L. 2004. *Aristocrats and Professionals: Country-House Science in Late-Victorian Britain*. University of Minnesota. Doctoral dissertation, Ann Arbor, MI: ProQuest Dissertations and Theses Global.
- Opitz, D. L. 2006. "This House is a Temple of Research: Country-House Centres for Late-Victorian Science." In *Repositioning Victorian Sciences: Shifting Centres in Nineteenth-Century Thinking*, edited by D. Clifford, E. Wadge, A. Warwick, and M. Willis. London UK: Anthem Press.
- Pasterkamp, G., J. Rotmans, D. de Kleijn, and C. Borst. 2007. "Citation Frequency: A Biased Measure of Research Impact Significantly Influenced by the Geographical Origin of Research Articles." *Scientometrics* 70, no. 1: 153–165.
- Pebesma, E. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data." *R Journal* 10, no. 1: 439–446. <https://doi.org/10.32614/RJ-2018-009>.
- Pebesma, E., and R. Bivand. 2023. *Spatial Data Science: With applications in R*. Boca Raton, FL: Chapman and Hall/CRC. <https://doi.org/10.1201/9780429459016>.
- Perez, T. M., and J. A. Hogan. 2018. "The Changing Nature of Collaboration in Tropical Ecology and Conservation." *Biotropica* 50, no. 4: 563–567.
- Pinto, B. J., T. Gamble, C. H. Smith, and M. A. Wilson. 2023. "A Lizard is Never Late: Squamate Genomics as a Recent Catalyst for Understanding Sex Chromosome and Microchromosome Evolution." *Journal of Heredity* 114, no. 5: 445–458.
- Quintero, I., M. J. Landis, W. Jetz, and H. Morlon. 2023. "The Build-Up of the Present-Day Tropical Diversity of Tetrapods." *Proceedings of the National Academy of Sciences* 120, no. 20: e2220672120.

- Quintero Toro, C. 2012. *Birds of Empire, Birds of Nation: A History of Science, Economy, and Conservation in United States-Colombia Relations*. Bogotá, DC: Ediciones Uniandes-Universidad de los Andes.
- R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org>.
- Reddy, S. 2014. "What's Missing From Avian Global Diversification Analyses?" *Molecular Phylogenetics and Evolution* 77: 159–165.
- Rhie, A., S. A. McCarthy, O. Fedrigo, J. Damas, G. Formenti, and S. Koren. 2021. "Towards Complete and Error-Free Genome Assemblies of all Vertebrate Species." *Nature* 592, no. 7856: 737–746.
- Roll, U., A. Feldman, M. Novosolov, A. Allison, A. M. Bauer, and R. Bernard. 2017. "The Global Distribution of Tetrapods Reveals a Need for Targeted Reptile Conservation." *Nature Ecology & Evolution* 1, no. 11: 1677–1682.
- Sawchuk, E. A., K. A. Sirak, F. K. Manthi, E. K. Ndiema, C. A. Ogola, and M. E. Prendergast. 2024. "Charting a Landmark-Driven Path Forward for Population Genetics and Ancient DNA Research in Africa." *American Journal of Human Genetics* 111, no. 7: 1243–1251.
- Secretariat of the Convention on Biodiversity. 2011. "Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising From Their Utilization to the Convention on Biological Diversity: Text and Annex." Montreal, Canada, p. 15. <https://doi.org/10.25607/OBP-789>.
- Sharaf, A., L. T. Nesengani, I. Hayah, et al. 2024. "Establishing African Genomics and Bioinformatics Programs Through Annual Regional Workshops." *Nature Genetics* 56: 1–10.
- Sherkow, J. S., K. B. Barker, I. Braverman, et al. 2022. "Ethical, Legal, and Social Issues in the Earth BioGenome Project." *Proceedings of the National Academy of Sciences* 119, no. 4: e2115859119.
- Soares, L., K. L. Cockle, E. Ruelas Inzunza, J. T. Ibarra, C. I. Miño, and S. Zuluaga. 2023. "Neotropical Ornithology: Reckoning With Historical Assumptions, Removing Systemic Barriers, and Reimagining the Future." *Ornithological Applications* 125, no. 1: duac046.
- Stephan, P. 2012. *How Economics Shapes Science*. Cambridge, MA: Harvard University Press.
- Stocks, G., L. Seales, F. Paniagua, E. Maeher, and E. M. Bruna. 2008. "The Geographical and Institutional Distribution of Ecological Research in the Tropics." *Biotropica* 40, no. 4: 397–404.
- Taylor, H. R., N. Dussex, and Y. van Heezik. 2017. "Bridging the Conservation Genetics Gap by Identifying Barriers to Implementation for Conservation Practitioners." *Global Ecology and Conservation* 10: 231–242.
- Titley, M. A., J. L. Snaddon, and E. C. Turner. 2017. "Scientific Research on Animal Biodiversity is Systematically Biased Towards Vertebrates and Temperate Regions." *PLoS One* 12, no. 12: e0189577.
- Toews, D. P., L. Campagna, S. A. Taylor, C. N. Balakrishnan, D. T. Baldassarre, and P. E. Deane-Coe. 2016. "Genomic Approaches to Understanding Population Divergence and Speciation in Birds." *Auk* 133, no. 1: 13–30.
- Toomey, A. H., A. T. Knight, and J. Barlow. 2017. "Navigating the Space Between Research and Implementation in Conservation." *Conservation Letters* 10, no. 5: 619–625.
- Vamosi, J. C., and S. M. Vamosi. 2008. "Extinction Risk Escalates in the Tropics." *PLoS One* 3, no. 12: e3886.
- Willig, M. R., D. M. Kaufman, and R. D. Stevens. 2003. "Latitudinal Gradients of Biodiversity: Pattern, Process, Scale, and Synthesis." *Annual Review of Ecology, Evolution, and Systematics* 34, no. 1: 273–309.