

Some Richness Gets Richer v.3.8.25

Wares, JP

Odum School of Ecology, University of Georgia

jpwares@uga.edu

The goal of this data exploration was to understand numerically the extent to which we, as a scientific community, study a tiny subset of most of the diversity that is out there. I began this assessment thinking about trait transitions among species being labile (eg Hart et al 1997, Cunningham 1999, Hart and Podolsky 2005), pathogen transfer being something that needs further study among related organisms (Park et al 2018, Park 2019), and the potential for genomic interactions and transfers. We need more information at recent phylogenetic scales for these comparisons, and they often do not exist in the data that are available for meta/synthetic studies. What can we learn from transitions in traits on the scale of Miocene, Pliocene, or more recently separated populations or species, relative to what we learn when the tips of comparative analyses can only involve contrasts that scale to 100s of millions of years?

In particular, here I focus on the perceived economic or ecosystem effects of species as an element of understanding disparity in study. While species like *Pisaster ochraceus* were included as the first representatives of consumer-driven community control (Paine 1966), in recent years we have seen that even the small, less charismatic members of the same guild play equivalent roles (Gravem and Morgan 2017). So how do our biases drive our broader discovery?

Here I have developed a very simple bibliometric study to ask about the overall incidence of study of a given, randomly selected, taxon from several key marine invertebrate phyla. To what extent does the 'economic' or 'ecosystem' declaration for a species - which itself is part of seeking and being awarded federal funding for most biologists - affect citation patterns, and how do those declarations affect what we know about other key facets of interaction or diversity at finer taxonomic scales?

So, here I query the scientific literature not for meta-analysis but merely *effort* by the scientific community. How does perceived economic or ecosystem "importance" influence the effort that leads to publication, the effort that leads to additional work on pathogens or parasites, or the additional effort required to understand movement and potential local adaptation, or our evidence for climate change response (Sunday et al 2012, Poloczanska et al 2016)? Of course sheer effort will correlate with all of these; there are more publications on *Crassostrea virginica* than dozens of other mollusks combined, and that necessarily invites more work on population genomics and pathogens that affect harvest or the health of consumers. But without the exploration we cannot fully see what we are not studying sufficiently at all.

What segment of diversity is being studied, effectively?

First of all, we ask about what the overall emphasis of diversity is for each of 6 key marine phyla (Annelida, Arthropoda, Cnidaria, Echinodermata, Mollusca, Porifera). To address the relative abundance of studies across taxa, we apply Hill numbers -- here, Simpson's (1949) diversity index, to address the dominance of some species in such data. There are a lot of scientific contributions, as there are more than one paternal contributor in most broods of the barnacle *Chelonibia testudinaria* (Ewers-Saucedo et al 2016). However, the strong skew towards a small number of species is often overlooked in attempts to develop synthetic ideas about "how things work".

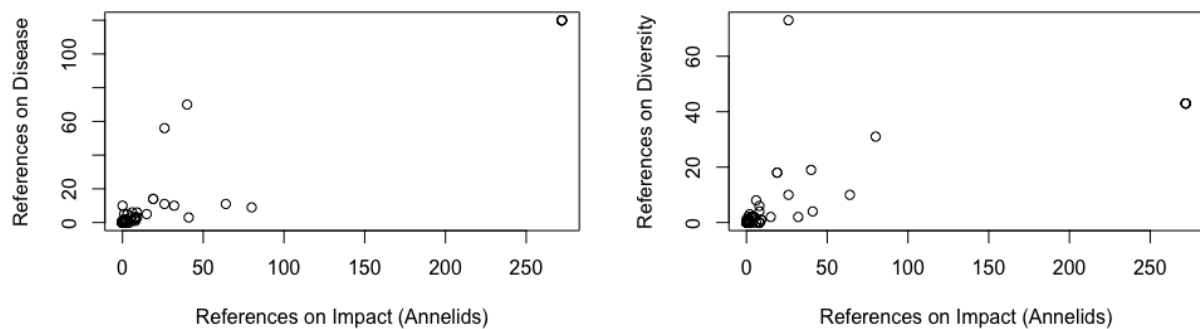
Some of this variation is, of course, driven by availability (e.g. by depth; Dawson et al 2023) or numeric abundance. Nevertheless, what is the citation intensity across the phyla named above? To do this requires 3 key steps. First, I obtained a list of all 'marine' taxa from WoRMS (2024), with metadata for each taxon (no location data, just taxonomic history). From the complete list of species obtained from WoRMS (2024), we have a total of 579156 known marine taxa. From the 6 selected phyla noted above, a random (without replacement) list of 100 taxonomic names was selected in R for each phylum. I tried using the 'acceptedNameUsage' to avoid historical synonymies, however this option tends to result in the majority of the taxonomic entries having only one or a very few observations in the literature. Working instead with 'genus' includes some instances of multiple congeners in our lists, but provides greater numerical signal; when a recognized taxon in WoRMS was missing information for 'genus' (and would for 'species' as well), the script would replace with a common taxon in that phylum with respect to disease or parasite studies (e.g. *Crassostrea* for Mollusca) or else the automated search would fail. The potential for duplicate searches does bias metrics of literature coverage somewhat but not in a way that will obscure the goals of this study.

I then searched the Web of Science using the Clarivate Starter API (developer.clarivate.com/) and code `rwosstarter` by Casajus (2023), evaluating by Topic (which searches title, abstract, keyword plus, and author keywords). Searches were first for the random set of 100 genera from each phylum. Searches for each sample of taxa (by phylum) included all publications per taxon, those including potential *impact* ("economic" or "ecosystem"), those including potential *disease* ("parasite" or "pathogen"), and those including exploration with *genomic* approaches ("intraspecific" or "genomic"). These are clearly not exhaustive approaches to search but guide towards larger patterns.

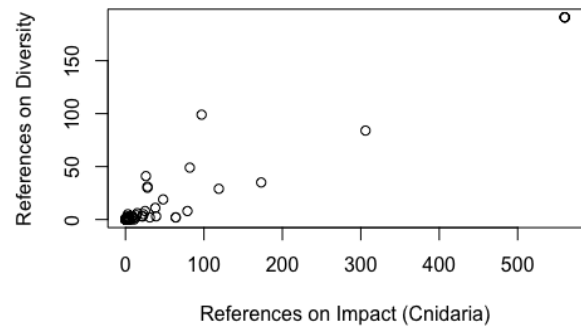
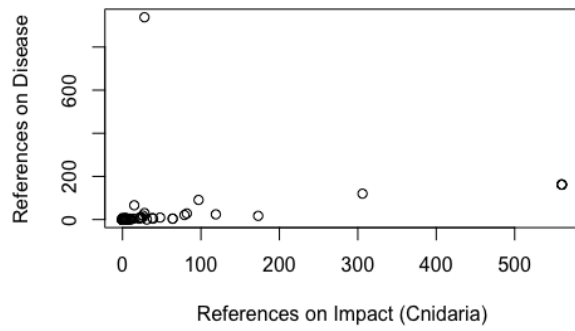
For example, the Phylum Mollusca includes over 164,000 entries in WoRMS. By randomly sampling 100 genera from this phylum (this random sample happens to include *Crassostrea*), we find a really strong selection for which taxa get the most observation. Using the count of published papers about each genus, the inverse Simpson's index (10.2) tells us how skewed this distribution is, indicating that about 10.7% of genera from our small subsample dominate publishing effort. Similar results were found for the other five phyla (ranging from 10.1-14.9%, with only notable exception being Porifera where inverse Simpson's index divided by number of taxa was 36.7%; however, sponges are also studied an order of magnitude less frequently by these numbers).

Below are plots of the number of references for each phylum assessed by “impact” relative to pathogen/parasite or within-species diversity, loosely assessed (see *Appendix A* for data).

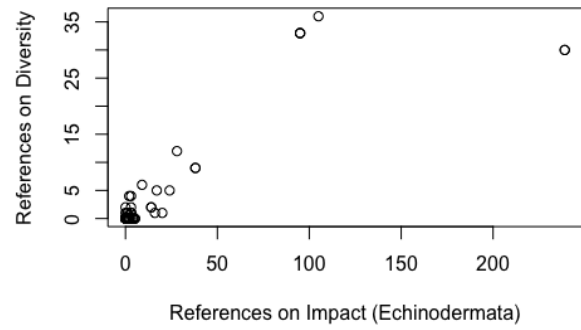
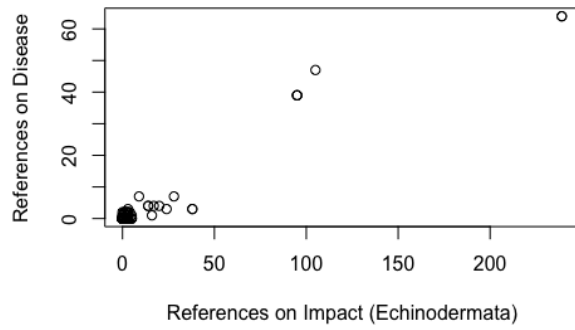
Annelids



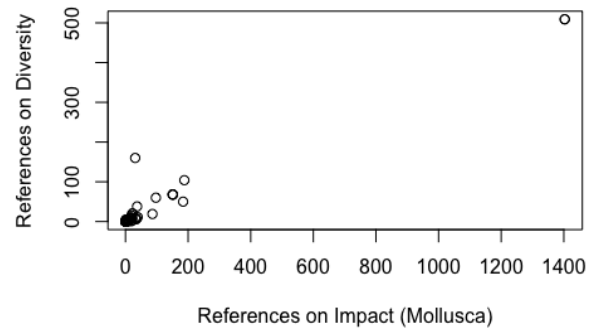
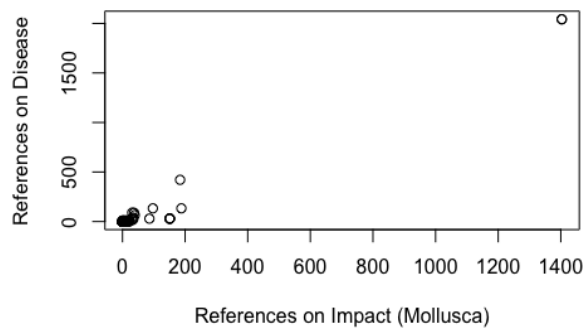
Cnidaria



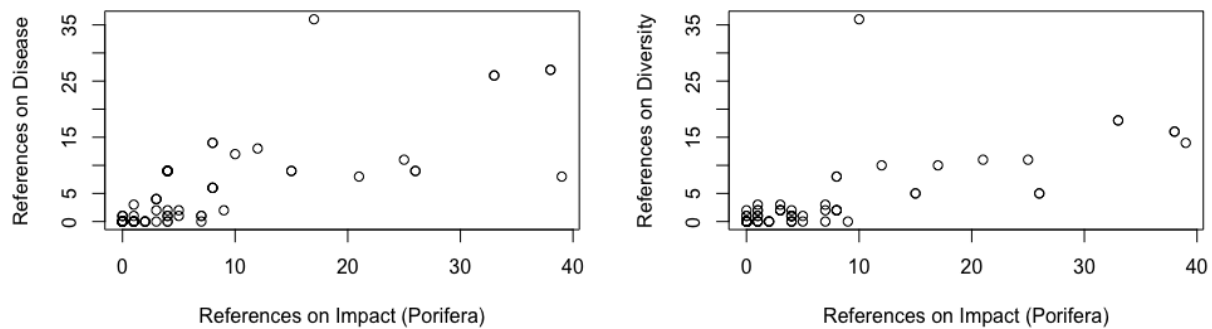
Echinoderms



Mollusks



Sponges



The fact that our research efforts are this skewed is not a surprise, of course. Some species are easier to find in abundance, others have well-documented effects of their 'impact' on human society in terms of economic or ecosystem factors, and of course as more information is constructed around a species we have new insights to build upon (see Dietrich et al 2020).

Still, this leads to predictable information gaps that span huge phylogenetic distances, which can be of vital importance for understanding patterns of transmission when disease emerges in a taxon of interest. In all 6 phyla examined here, the correlation between what is known about an organism's 'impact' generally has a strong correlation on how much that organism is evaluated for parasites or pathogens (Annelida 0.963; Arthropoda 0.922; Cnidaria 0.356¹; Echinodermata 0.966; Mollusca 0.988; Porifera 0.795) or within-species genomic diversity (Annelida 0.811; Arthropoda 0.989; Cnidaria 0.975; Echinodermata 0.870; Mollusca 0.976; Porifera 0.740). Again, this doesn't indicate a pattern of causation, but identifies that our knowledge is unevenly distributed across the tree of life. Many traits - including those that permit pathogens or parasites, or that influence patterns of abundance or diversity within a taxon - can be highly labile (Guo and Wares 2017, Pappalardo et al 2020), yet our focus on a select group of species often means that any attempt to generalize may be using data points that are easily 100s of millions of years isolated from one another (e.g. the 2 most abundantly-studied molluscan genera in a random sample were *Crassostrea* and *Conus*, separated 536 mya, Kumar et al 2022 -- though of course each of these taxa themselves include some fascinating intrageneric comparisons to be considered).

¹ Note this low correlation is driven in part by *Myxobolus*, itself an unusual parasite, highlighting another issue with my search terms - some taxa *are* parasites, some will *have* parasites, both will be caught in this search I'm afraid.

As an example of what we miss with large phylogenetic distances, even very similar-seeming organisms may merit additional attention into why -- or why not -- they share a pathogen, parasite, or other syndrome. The mussel *Geukensia*, for instance, is more than 300 mya diverged from the mussel *Mytilus* despite being in the same family Mytilidae (Audino et al 2020). They are both even more diverged from the soft-shell clam *Mya arenaria* (Family Myidae). The unusual form of disease known as bivalve transmissible neoplasia (transmissible tumors) is found in *Mya* (Giersch et al 2022) and *Mytilus* (Yonemitsu et al 2019; Skazina et al 2021,2023), yet I don't think there is any work on whether or not these are found in the ribbed mussel *Geukensia*. A hook for the study of less-studied organisms is insight into how their context interacts with their traits in ways that we cannot get just by “model” organisms alone (Bertile et al 2023). Look, it's just an opinion paper with a little bit of data sprinkled on top! Yet the fact that nobody tends to eat *Geukensia* likely is the difference in effort to know more about them.

I guess I'm arguing for the fact, in two collaborations I'm currently part of, that when we focus on model systems will the extrapolation be so great? One of these collaborations exhibits few signals for life history traits affecting genomic diversity, but would it be different in more closely related forms, like developmentally polymorphic polychaetes (Zakas et al 2013)? Would we gain more if we knew how specific *as well as* transferrable the insights were from key organisms? The more labile a trait is, the more resolution we need in a phylogeny to truly understand it (Blomberg et al 2003). Since we are not talking about relatively stable morphological developments (e.g. “chitinous exoskeleton” or “segmentation”) but traits that are likely complexly driven by both evolutionary history (i.e. cospeciation) as well as environmental context (host transfer or factors affecting within-species diversity, both in a constantly changing environment), this seems to be important (Poulin et al 2011, Hensen et al 2023).

This is a call to early career scientists as well as those of us with the privilege to take on new projects with our resources. Every species you can find in reasonable abundance has a pathogen or parasite of interest; they all exhibit intraspecific diversity; and many have yet to be fully assessed for what they do for our natural ecosystems. Yet, by concentrating our effort on a subset of distantly-related diversity we don't fully understand how these simple factors are interrelated. There is still so much to be done as biologists. Evaluating more recent (intrafamilial, intrageneric) potential for trait-based transitions in these respects is likely key to understanding how impact, interactions, and diversity are driven. This is only going to happen by bypassing our biases towards the well-known species and towards those that need to become well-known.

REFERENCES

Audino et al 2020 DOI:[10.1093/zoolinnean/zlaa011](https://doi.org/10.1093/zoolinnean/zlaa011)

Bertile, F., Matallana-Surget, S., Tholey, A. *et al.* Diversifying the concept of model organisms in the age of -omics. *Commun Biol* 6, 1062 (2023).
<https://doi.org/10.1038/s42003-023-05458-x>

Blomberg SP, Garland T Jr, Ives AR. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*. 2003 Apr;57(4):717-45. doi: 10.1111/j.0014-3820.2003.tb00285.x. PMID: 12778543.

Casajus N (2023) rwsstarter: An R Client to the Web of Science Starter API. R package version 0.1. <https://github.com/FRBCesab/rwsstarter>

Collins, T.M., K. Frazer, A.R. Palmer, G.J. Vermeij, W.M. Brown. 1996. Evolutionary history of northern hemisphere *Nucella* (Gastropoda, Muricidae): Molecular, morphological, ecological, and paleontological evidence. *Evolution* 50:2287-2304.

Cunningham, CW. 1999. Some limitations of ancestral character-state reconstruction when testing evolutionary hypotheses. *Systematic Biology* 48: 665-674.

Dawson et al 2023 doi: 10.1086/727969

Dietrich et al 2020 How to choose your research organism
<https://doi.org/10.1016/j.shpsc.2019.101227>

Ewers-Saucedo C, Hope NB, Wares JP. The unexpected mating system of the androdioecious barnacle *Chelonibia testudinaria* (Linnaeus 1758). *Mol Ecol*. 2016 May;25(9):2081-92. doi: 10.1111/mec.13593.

Giersch RM, Hart SFM, Reddy SG, Yonemitsu MA, Orellana Rosales MJ, Korn M, Geleta BM, Countway PD, Fernández Robledo JA, Metzger MJ. Survival and Detection of Bivalve Transmissible Neoplasia from the Soft-Shell Clam *Mya arenaria* (MarBTN) in Seawater. *Pathogens*. 2022 Feb 23;11(3):283. doi: 10.3390/pathogens11030283

Hart, MW, M Byrne, MJ Smith <https://doi.org/10.1111/j.1558-5646.1997.tb05108.x>

Hart, MW and RD Podolsky 2005. Mitochondrial DNA phylogeny and rates of larval evolution in *Macrophiothrix* brittlestars. *Mol. Phyl. Evol.* 34: 438-447.

Hensen et al 2023 <https://doi.org/10.1016/j.ympev.2023.107938>

Gravem, S.A. and S.G. Morgan. 2017. Shifts in intertidal zonation and refuge use by prey after mass mortalities of two predators. *Ecology* 98:1006-1015.

Guo B, Wares JP. 2017. Large-scale gene flow in the barnacle *Jehlius cirratus* and contrasts with other broadly-distributed taxa along the Chilean coast. *PeerJ* 5:e2971 <https://doi.org/10.7717/peerj.2971>

Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB (2022) TimeTree 5: An Expanded Resource for Species Divergence Times. *Mol Biol Evol* doi.org/10.1093/molbev/msac174

Pappalardo, P. et al 2018 DOI: 10.1111/geb.13008

Park, A. W., M. J. Farrell, J. P. Schmidt, S. Huang, T. A. Dallas, P. Pappalardo, J. M. Drake, et al. 2018. "Characterizing the Phylogenetic Specialism-Generalism Spectrum of Mammal Parasites." *Proceedings of the Royal Society B: Biological Sciences* 285 (1874). <https://doi.org/10.1098/rspb.2017.2613>.

Park, A. W. 2019. "Phylogenetic Aggregation Increases Zoonotic Potential of Mammalian Viruses." *Biology Letters* 15 (12). <https://doi.org/10.1098/rsbl.2019.0668>.

Poulin R, Krasnov BR, Mouillot D, Thieltges DW. The comparative ecology and biogeography of parasites. *Philos Trans R Soc Lond B Biol Sci*. 2011 Aug 27;366(1576):2379-90. doi: 10.1098/rstb.2011.0048. PMID: 21768153; PMCID: PMC3130428.

Skazina M, Odintsova N, Maiorova M, Ivanova A, Väinölä R, Strelkov P. First description of a widespread *Mytilus trossulus*-derived bivalve transmissible cancer lineage in *M. trossulus* itself. *Sci Rep*. 2021 Mar 11;11(1):5809. doi: 10.1038/s41598-021-85098-5

Maria Skazina, Nelly Odintsova, Mariia Maiorova, Lidia Frolova, Irina Dolganova, Kira Regel, Petr Strelkov, Two lineages of bivalve transmissible neoplasia affect the blue mussel *Mytilus trossulus* Gould in the subarctic Sea of Okhotsk, *Current Zoology*, Volume 69, Issue 1, February 2023, Pages 91–102, <https://doi.org/10.1093/cz/zoac012>

Paine, Robert T. "Food Web Complexity and Species Diversity." *The American Naturalist* 100, no. 910 (1966): 65–75. <http://www.jstor.org/stable/2459379>.

Poloczanska et al 2016 <https://doi.org/10.3389/fmars.2016.00062>

Simpson, E.H. (1949) Measurement of Diversity. *Nature*, 163, 688

Sunday, JM, AE Bates, and NK Dulvy. 2012. Thermal tolerance and the global redistribution of animals. *Nature Climate Change* 2: 686-690.

WoRMS Editorial Board (2024). World Register of Marine Species. Available from <https://www.marinespecies.org> at VLIZ. Accessed 2024-09-03. doi:10.14284/170

Marisa A Yonemitsu Rachael M Giersch Maria Polo-Prieto Maurine Hammel Alexis Simon Florencia Cremona Fernando T Avilés Nicolás Merino-Véliz Erika AV Burioli Annette F Muttray James Sherry Carol Reinisch Susan A Baldwin Stephen P Goff Maryline Houssin Gloria Arriagada Nuria Vázquez Nicolas Bierne Michael J Metzger (2019) A single clonal lineage of transmissible cancer identified in two marine mussel species in South America and Europe *eLife* 8:e47788.

Zakas, C. and J.P. Wares. 2012. Consequences of a poecilogonous life history for genetic structure in coastal populations of the polychaete *Streblospio benedicti*. *Mol. Ecol.* 21: 5447-5460.