PHYLOGENETIC DIVERSITY AND FISH: A COMPARISON OF ACTINOPTERYGII PHYLOGENETIC DIVERSITY WITH OTHER VERTEBRATE GROUPS

Word count: 2452 words

Scott, Oenone.

Contact: ojs19@imperial.ac.uk

Supervisors:

James Rosindell (j.rosindell@imperial.org.uk) & Rikki Gumbs (Rikki.Gumbs@zsl.org)

A thesis submitted in partial fulfilment of the requirements for the Computational Methods in Ecology and Evolution Master of Research at Imperial College London

Formatted in the Harvard Style

1 Abstract

2 Introduction

2.1 The need for conservation efforts

Our planet is currently losing species at such a significant rate that this period has been called the "Sixth mass extinction" (Barnosky et al. 2011). This loss of biodiversity is detrimental both to the ecosystems in which these species reside and to humanity, through the loss of ecosystem services such as sources of food, medicine, and employment. Furthermore, it can be argued that biodiversity is worth preserving for its own sake, although quantifying the intrinsic value of biodiverse ecosystems outside of potential anthropogenic uses can be challenging. True extinction events cannot (yet) be reversed, despite the valiant efforts of some, such as **Piotrowska2018** and the Wooly Mammoth. However, through conservation action, species that are extinct in the wild can be reintroduced; vulnerable or threatened species can be protected, and entire ecosystems can be protected. Conservation and reintroduciton of species have been happening across most continents for decades, with successful outcomes realised for a broad variety of taxa - including plants (**Godefroid2011**), freshwater fish (**Cochran-Biederman2015**), and ferrets (**Jachowski2011**). Unfortunately, conservation actions are resource intensive and expensive, and conservationists are unable to protect every threatened or vulnerable species on the planet. This is compounded by a lack of comprehensive, up-to-date data on which species are threatened globally. These limitations, along with other practical challenges such as accessibility, force conservationists to select some species to focus on, meaning some species or entire ecosystems are left without any support.

There are many metrics by which conservationists evaluate species' worthiness of protection. Some focus on a species' "irreplaceabilitiy". Others focus on physiological or genetic 'uniqueness', unusual abilities, or species living in rare habitats (Brooks et al. 2006), and yet prioritise population numbers, roles in ecosystems, or a combination of multiple factors. The problem with some of these approaches is that they can be skewed towards organisms that are charismatic (i.e. the Giant Panda), or populations that are easier to observe (e.g. rhinoceroses) and miss species that are hard to find/ see (i.e. deep-sea organisms), visually unremarkable or belonging to a non-charismatic taxa, but that exciting for other reasons - such as being future source of medicinal chemicals, displaying interesting behaviour, or playing an irreplaceable role within an ecosystem (Taylor 2002). Another way in which conservation can vary is through the breadth of scope. Some species are thought to best benefit from species-wise conservation () whereas other approaches aim to conserve a range of species, and others will try and protect an entire biome (**something about the rainforest**). There are pros and cons to each approach, and as yet there is not a clear framework as to when to use each approach, as there are only a few studies that review and compare such conservation approaches (**find a lot more citations**). There are several clades of fish (such as sharks) that have

TALK ABOUT SPECIES-WISE PROTECTION OF SPECIES AND ECOSYSTEM WIDE PROTECTION

WHY WOULD KNOWING MORE ABOUT SPECIES MEAN THEY ARE BETTER PROTECTED

WHY WOULD FISH BENEFIT FROM SPECIES-WISE PROTECTION

There is significant taxonomic and geographic bias reported in conservation research (Taylor 2002; Darwall et al. 2011; Watson et al. 2017), which highlights the clear need for a method that avoids taxanomic bias, and can be applied widely to identify which species which require conservation prioritisation.

INTRODUCE PHYLOGENETIC DIVERSITY. WHAT IS IS AND WHAT IT MEANS. THEN, WHY WE ARE LOOKING AT IT HERE.

INTRODUCE EDGE METRIC, AND THEN TOUCH ON IT'S APPLICATION AND ONE-TWO SUCCESS STORIES

The 'EDGE' metric seeks to solve this problem of biased prioritisation by calculating prioritisation lists of species using the evolutionary history represented by species - it's "Evolutionarily distinctness" (ED) score combined with its level of global endangerment (GE). A species' ED score is calculated using a phylogenetic tree, and the score represents the numbers of millions of years of evolutionary history that is within that species (Isaac et al. 2007). Species that have speciated recently, or that have many recent 'cousin' species will have lower ED scores than species that have no close living relatives, and that speciated a long time ago.

The GE score of a species is calculated using the IUCN's 'Red List' assessment for that species (Isaac et al. 2007), in which species are categorised according to how stable their populations are and how likely they are to become extinct (IUCN 1975). These categories are shown in Figure 1. This metric was chosen for this study as it provides a mechanism to comprehensively

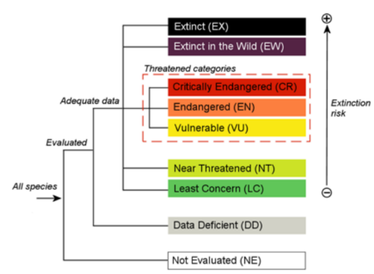

Figure 1: IUCN Red List categories (IUCN 1975)

assess species within family groups (or larger), and this study aims to produce results that can be used to meaningfully influence the manner in which these species are considered.

2.2 Why fish?

The group considered in this study are the *Actinopterygii*, commonly known as the ray-finned fish. Actinopterygii are the largest class of vertebrate, with over 33,000 constituent species, with more species being described regularly (Froese and Pauly n.d.). For context, mammals are thought to have closer to 6,400 extant species (Mammalogists 2020); reptiles have 9,500 species (Pincheira-Donoso et al. 2013) ; and birds have around 18,000 (Barrowclough et al. 2016). Unfortunately, the Actinopterygii are also proportionally the most understudied class of the vertebrates. for example, just over 50% of Actinopterygii are assessed in the IUCN Red List, compared with the 62-90% of other classes, as laid out in Table 2.2.

Class	# of Species	# Assessed	% Assessed
Actinopterygii	33,000	17,955	54.4
Mammalia	6,500	5850	90.0
Aves	18,000	11,147	62.0
Reptilia	9,400	7892	84.0

Table 2.2: IUCN Red List assessments for vertebrate classes (Mammalogists 2020; Pincheira-Donoso et al. 2013) (Barrowclough et al. 2016; Froese and Pauly n.d.)

The IUCN is not the only research body in which Actinopterygii are poorly represented. A quick search in WebofScience also demonstrates the lack of research into this class. When searching for topics that include the class names of the groups mentioned in Table 2.2, the results are as follows: Actinopterygii -2055; Mammalia - 8905; Reptilia - 6,030; Aves - 13,116. This bias is acutely present in conservation science - a recent study by Watson et al. 2017 revealed that, of the vertebrate groups, fish had the highest imbalance between number of published studies and number of described species (16% and 50% respectively).

Despite the lack of studies into Actinopterygii, many species of fish are acknowledged to be very important as a source of food and employment - through fishing, tourism and entertainment. Fish consumption has been rising steadily for the past 6 decades, as highlighted clearly in Figure 2.

Many people, especially in Europe and Northern America, are increasing the amount of fish in their diet - some to improve their health (Rimm 2006)), and others to lessen their impact on the environment (Mary and Shoup 2018). Younger generations are particularly aware of the impact their food choices have on the planet (Mary and Shoup 2018), so it could be assumed that this trend of eating more fish is unlikely to change. This increased demand for fish (and fish-related products) has lead to an increase in aquaculture (farming fish) (Belton, Bush, and Little 2018). Wild-caught fish number are in decline due to stock exhaustion and populating collapse rather than a decrease in interest in wild caught fish (Tremblay-Boyer et al. 2011; Belton, Bush, and Little 2018).

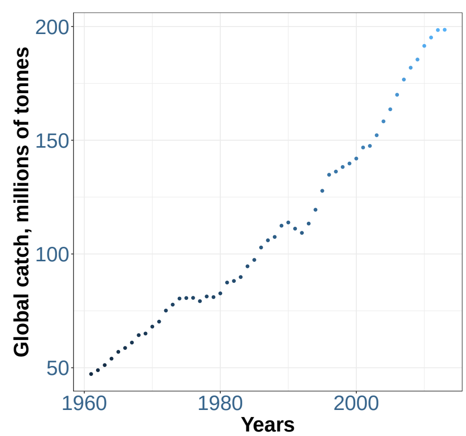

Figure 2: Global catch of freshwater, demersal, pelagic and marine fish, (Department and Aquaculturer 2020)

Marine ecosystems in particular are under increasing pressure from human action. Globally, governments and scientific bodies have agreed that ecosystems need to be protected in order to protect the future of marine resources (Muraki Gottlieb et al. 2018). Marine Protected Areas (MPAs) are one such form of protection - however the definition of these vary significantly from one location to another (Marinesque, Kaplan, and Rodwell 2012). In general, an MPA's success is evaluated by its effectiveness at conserving 'keystone' species, or their value to local fishing activities (Marinesque, Kaplan, and Rodwell 2012). As discussed above, by considering only 'keystone' species, considerable ecosystem damage may occur un-measured, and many species may go un-protected.

Given that many species of fish exist in both the marine and freshwater realms, and many more migrate around the globe, clear conservation prioritisation for fish will be crucial to ensure we do not fail to protect vulnerable and non-charismatic species.

2.3 X, Y and Z

subsection: why these fish in particular

[now to sum everything up]

It is clear that the Actinopterygii group is both under-studied and under-protected, despite their importance to human populations across the globe. This study is a first step in directly addressing this imbalance. Using the EDGE metric, we create conservation prioritisation lists for [X, Y and Z] and, along with geospatial analysis of their current habitats and threats, highlight methods through which vulnerable species could be protected. This study also provides a roadmap that will be useful for the creation of reports regarding other family groups within the Actiopterygii class.

3 Methods

3.1 Taxanomic Matching

Data Collection

Before performing the EDGE2 analysis, I needed to create the most up-to-date species list of each taxanomic group, source a phylogenetic tree set, and then adjust the species in the tree according to the latest species list.

Of the vertebrate groups that I calculated EDGE for, Birds, Squamates and Amphibians had already gone through the process of taxanomic matching and 'tree cleaning'. Chondrichythes, Actinopterygii, Mammals and Agnatha all required this process.

The first step was to establish a taxanomic authority. For Mammals we used the data from 'VertLife'(), which pulls taxaonomic data from the ASM Mammal Diversity Database. For Condrichythes, Actinopterygii and Agnatha, we used FishBase as the taxanomic authority (Froese and Pauly n.d.) . This means that we used the species data from this group, and adjusted species names in our other data sources to match the authority.

The second data set we require was a phylogenetic tree/ set of trees. Each of these data came from a specialised site. For the Actinopterygii class we used the set of 100 trees from **Chang2019**. For the Chondrichthyes we used the trees from INSERT SHARK TREE SOURCE. For Mammals we used the tree set from HERE. For the Agnatha [insert what actually happened when I do this piece of work].

The third data set we used was the IUCN Red List, as the probability of extincion (pext) value is calulcated based on the Red List "Extinction" Category (IUCN 1975).

To create the 'master species' list the first step was to strip out any sub-speciies from the master species list - this was done by removing any data entries with three componants i.e. [example]). We then compared the list of species in the phylogenetic trees and in the

Red List data with the taxanomic authority. For any species that were in the trees or in the red list that weren't in the taxanomic authority, we manually checked to see if these species were either spelling mistakes or synonyms of valid species. If the species was an incorrect spelling of a valid species, then the name was corrected. If the species was a synonym of a valid species, then the database was checked to see if the correct species was in the database. If it was not, then the name was changed to the valid species name. If the valid species was also present in the database, then the species was discarded. If the species was neither a spelling mistake nor a synonym, then it was also discarded. [Create a diagram for this?]

At the end of the taxanomic matching, we had a data frame for each taxanomic group containing the taxanomic species name, genus and family, the tree species, and the red list category (n.b. the tree species and red list category were entered as NA if the valid species weren't found in the tree or evaluated by the red list).

We then updated the phylogenetic trees. Using a base set of 1000 trees, we removed all of the species in the tree that were not in the taxonomic database (the 'invalid' species). We then checked to see if there were any family groups that were missing from the tree, and imputed a single species from each family group. We repeated this with the genera. Then, we imputed the rest of the missing species. As the imputation process is somewhat random, the result was 1000 unique trees for each taxanomic group.

3.2 EDGE, EDGE2 and PD calculations

Once we had all of the updated phylogenetic trees and IUCN data we looked at EDGE2 scores and phylogenetic diversity loss across the whole taxanomic tree for each of our clades. The EDGE2 metric is an updated way version of the EDGE metric, originally designed by **Steel2007**.

How to calculate EDGE scores:

How to calculate EDGE2 scores:

How to calculate PD loss:

For each of the vertebrate groups we: -Calculated EDGE2 Scores using 7 pext values* (including clade-specific where possible) – Did a spearman analysis of the top 100 EDGE species in each pext forecast to see if the pext forcast affects the EDGE ranking - Calculated EDGE scores for each vertegrate group – did a spearman anaylsis of the top 100 EDGE species with the top 100 EDGE2 species for each forcast -Calculated percentage PD loss for each clade using the 7* pext values (*including clade-specific wehre possible)

— FOR JUST FISH OR FOR ALL VERT GROUPS? - Scrambled pext (for all 6 pext forcasts) and calculcated percentage PD loss. – Do some kind of analysis to see if there is a difference between the PD loss in the switched up data and the correct data set.

3.3 Clade selection

We picked three clades for the in-depth threat analysis. To pick the most suitable clades we first screened each family group for size, Red List coverage and presence of vulnerable species. We then created super-family clusters of families that: had ¿50% Red List Coverage and at least 1 vulnerable species, ¿100 species, and were monophyletic. This left us with 21 large clusters (is this for the results section?) [have table with all of the groups in??]. We then looked into each of these clades to find if there were any biological meaning to these clusters of family groups. This left us with XXXX groups. We then picked 3 final groups for [x y z reasons].

3.4 Clade threat analysis

For the selected clades we: - Produced an EDGE list, using the EDGE 2 metric. - Threat analysis - Range analysis - Overlap with MPAs? Especially of important areas? - Trends in threat? Increasing/ decreasing/ static?

3.5 Packages and technical specifications

rfishbase rredlist phylbase capter phytools data.table geiger pez

4 Results

4.1 Updated phylogenetic trees

SHOULD I TALK ABOUT HOW MANY SPECIES I NEEDED TO IMPUTE for each taxonomic group?

4.2 PD loss

Insert: image with percentage PD loss across 6 vertebrate groups with all 6 pext forecasts

Insert: Tree showing the evolutionary relationship of the 6 vertebrate groups.

4.3 EDGE1 vs EDGE 2

Insert: eeither multi-panel of the spearman analyses of the EDGE 2 vs EDGE 1 scoers for all vertebrate groups, or just pick one/ two (dpeending on results) and descript.

4.4 EDGE 2 - variation by pext

Insert: show image from spearman analysis comparing top 100 across all of the EDGE 2 groups.

4.5 PD at threat: random or not

Insert: barplot showing range of PD loss with real data vs scrambled, plus an analysis to see if there is a statistical difference.

4.6 THREAT analysis

Something about threats.

5 Discussion

References

Barnosky, Anthony D. et al. (2011). "Has the Earth's sixth mass extinction already arrived?" In: *Nature* 471.7336, pp. 51–57. ISSN: 00280836. DOI: 10.1038/nature09678.

Barrowclough, George F. et al. (2016). "How many kinds of birds are there and why does it matter?" In: *PLoS ONE* 11.11, pp. 1–15. ISSN: 19326203. DOI: 10.1371/journal.pone.0166307.

Belton, Ben, Simon R. Bush, and David C. Little (2018). "Not just for the wealthy: Rethinking farmed fish consumption in the Global South". In: *Global Food Security* 16.October 2017, pp. 85–92. ISSN: 22119124. DOI: 10.1016/j.gfs.2017.10.005. URL: https://doi.org/10.1016/j.gfs.2017.10.005.

Brooks, T. M. et al. (2006). “Global biodiversity conservation priorities”. In: *Science* 313.5783, pp. 58–61. ISSN: 00368075. DOI: 10.1126/science.1127609.

Darwall, William R.T. et al. (2011). “Implications of bias in conservation research and investment for freshwater species”. In: *Conservation Letters* 4.6, pp. 474–482. ISSN: 1755263X. DOI: 10.1111/j.1755-263X.2011.00202.x.

Department, Fisheries and Aquaculturer (2020). *Fisheries and Aquaculture Department.* URL: http://www.fao.org/fishery/ (visited on 03/23/2020).

Froese, R and D Pauly (n.d.). *FishBase.* URL: www.fishbase.org,%20version%20(12/2019)..

Isaac, Nick J.B. et al. (2007). “Mammals on the EDGE: Conservation priorities based on threat and phylogeny”. In: *PLoS ONE* 2.3. ISSN: 19326203. DOI: 10.1371/journal.pone.0000296.

IUCN (1975). *IUCN RED LIST CATEGORIES AND CRITERIA*. Tech. rep., 59P.

Mammalogists, American Society of (2020). *Mammal Diversity Database.* URL: www.mammaldiversity.org. (visited on 2020).

Marinesque, Sophie, David M. Kaplan, and Lynda D. Rodwell (2012). “Global implementation of marine protected areas: Is the developing world being left behind?” In: *Marine Policy* 36.3, pp. 727–737. ISSN: 0308597X. DOI: 10.1016/j.marpol.2011.10.010. URL: http://dx.doi.org/10.1016/j.marpol.2011.10.010.

Mary, By and Ellen Shoup (Oct. 2018). “Cargill : Americans are eating more fish than they did five years ago , prioritizing nutrition and sustainability”. In: *Food Nativator USA.* URL: https://www.foodnavigator-usa.com/Article/2018/10/02/Cargill-Americans-are-eating-more-fish-than-they-did-five-years-ago-prioritizing-nutrition-and-sustainability.

Muraki Gottlieb, Hiroko et al. (2018). “Area Based Management Tools , Including Marine Protected Areas in Areas Beyond National Jurisdictionn, 9 – 11 October 2018 at IUCN Headquarters, Gland, Switzerland. Gland, Switzerland: IUCN.” In: October, pp. 1–16. URL: https://www.iucn.org/sites/dev/files/content/documents/iucn%7B%5C_%7Doct2018%7B%5C_%7Dabmt%7B%5C_%7Din%7B%5C_%7Dabnj%7B%5C_%7Dworkshop%7B%5C_%7Dreport%7B%5C_%7Dcorr.pdf.

Pincheira-Donoso, Daniel et al. (2013). “Global Taxonomic Diversity of Living Reptiles”. In: *PLoS ONE* 8.3, pp. 1–10. ISSN: 19326203. DOI: 10.1371/journal.pone.0059741.

Rimm, Eric B (2006). “Fish Intake, Contaminants, and Human Health”. In: 296.15, pp. 1885–1900.

Taylor, Raymond L. (2002). “Taxanomic Bias in Conservation Research”. In: *Science* 297.10, pp. 191–192. ISSN: 19457197. DOI: 10.1210/jcem-10-10-1361. URL: https://www.jstor.org/stable/3077032.

Tremblay-Boyer, Laura et al. (2011). “Modelling the effects of fishing on the biomass of the world’s oceans from 1950 to 2006”. In: *Marine Ecology Progress Series* 442, pp. 169–185. ISSN: 01718630. DOI: 10.3354/meps09375.

Watson, James E.M. et al. (2017). “Changing trends and persisting biases in three decades of conservation science”. In: *Global Ecology and Conservation* 10, pp. 32–42. ISSN: 23519894. DOI: 10.1016/j.gecco.2017.01.008. URL: http://dx.doi.org/10.1016/j.gecco.2017.01.008.

6 Appendix: Supplementary materials