

“Diseases are Like the Stars” (data mining exercise on known diseases)
(prepared by Michael Hood, Amherst College – Draft of October 2017)

Goal:

To see whether the number of known parasite species (i.e. Parasite Species Richness) is affected by the extent to which a host is studied.

Introduction:

Often at near the start of the disease ecology undergraduate course, we discuss parasite diversity and how many different diseases a host might encounter. This simple exercise shows how existing inventories of parasites are strongly biased by how much we know about particular host species, where non-agricultural or wild hosts are often poorly studied. We also discuss examples comparing parasite diversity among different types of hosts while controlling for this “diseases are like the stars” effect (e.g. Gibson et al. 2010. *Oikos* 119 (2010): 1919-1928). (The title phrase is from the adage ‘the more you look, the more you’ll see.’)

Method:

1. Use databases on know parasite-host associations to assess Parasite Species Richness.
2. Use ISI Web of Science to assess extent to which a host is studied.
3. Examine for a correlations between these traits in Excel

Part 1 - Plants

1. Plant Species Names were obtained for genera of interest, in particular for plant genera of with species of agricultural importance. Currently accepted names for host are according to ITIS, the Integrated Taxonomic Information System (<http://www.itis.gov/index.html>)
2. Each student assigned about a subset of plant host species in Excel datasheet (Diseases Are Like the Stars 2017.xls).
3. For each host species search in database at the USDA's Systematic Mycology and Microbiology Laboratory (<https://nt.ars-grin.gov/fungaldatabases/index.cfm>)
 - make sure to click on "host name" option, and past the plant name in the top box
 - record the number of “Fungus-Host combinations” reported in pop-up window (given as ___ "records in the Specimen database") in the column "Fungal Parasite Species Richness" of the Excel spreadsheet

Also, for each host species search name in quotes at the ISI's Web of Science (“WoS”) (<http://isiknowledge.com/wos>)

- use the "Topic" option
- record number of citation ("Results:") in column "WoS Citations"

Part 2 - Animals

1. Data on Parasite Species Richness provided by Amy Pedersen (Univ. Edinburgh) for combined known diseases (including protozoa, helminths, and arthropods...) as part of an on-going research project. Animal Species Names put in random order among ~600 in the larger study. (Stephens et al. (2017), *Global Mammal Parasite Database* version 2.0. *Ecology*, 98: 1476. doi:10.1002/ecy.1799) (<http://www.mammalparasites.org>)

- Each student assigned about a subset animal host species in Excel datasheet (Diseases Are Like the Stars 2017.xls).
- As was done with the plants, for each host species search name in quotes at the ISI's Web of Science (<http://isiknowledge.com/wos>)
 - use the "Topic" option
 - record number of citation ("Results:") in column "WoS Citations"

Last Step

Combine and analyze the data. We've recently gone with using a shared GoogleDoc to compile the results. We generally plot parasite number vs citation number per species, separately for plants and animals. Use of log scale axes is helpful to display the data (even though there are entries with zero values).

Typical Results:

