Diversity Analysis Exercise

Jack Sepkoski's Database

Some of J.J. Sepkoski's career work was in cataloging the diversity of marine life through the Phanerozoic by spending years pouring over scientific journals tabulating the occurrence of fossil genera. All told, 20 years of this work amounted to study of well over 30,000 genera. An impressive feat! After his death in early 1999 his colleagues saw to it that his most recent version of the database made it into press with the *Bulletins of American Paleontology* (Figure 1).

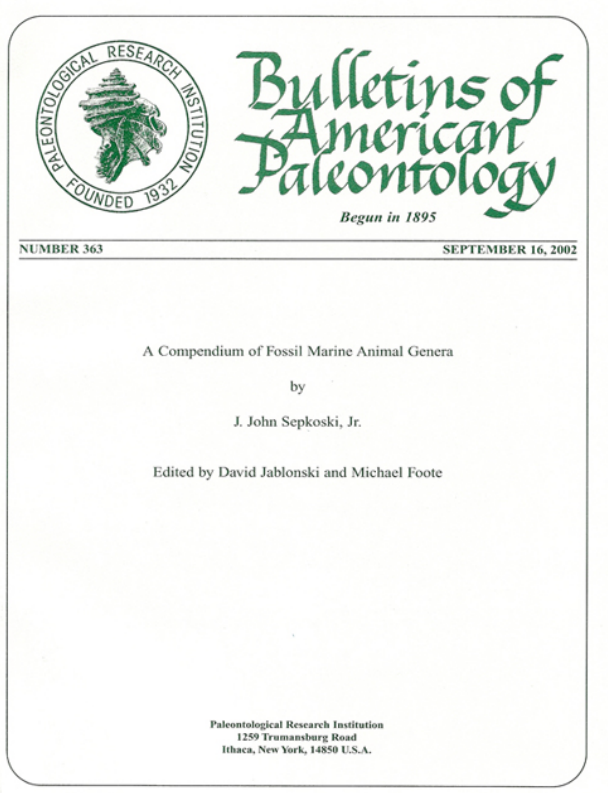

Figure 1. Front cover of the Sepkoski "Compendium" published in the *Bulletins of American Paleontology* in Sept. of 2002. Jablonski and Foote editors.

What Sepkoski did was quite simple really... He tallied FADs and LADs for marine organisms as reported in the paleontological literature and, for some groups, he gathered occurrence data within the taxon's duration as well. What Sepkoski was able to do then was to construct diversity curves for many different marine taxa individually AND all marine taxa taken collectively. Since his data amount to mostly taxon durations and are binned at the *Stage* level (sub-divisions of Periods) we can manipulate the data almost any way we want! With just some simple mathematical conversions we can produce diversity curves using the standing, boundary crossers, or range-through methods for example. We can even calculate extinction rates and origination rates through time.

Accessing the Data

Shanan Peters at the University of Wisconsin - Madison has done a great service to us and the paleontological community as a whole by providing a searchable and easy-to-use version of Sepkoski's (2002) database. We owe him a giant "*Thanks!*" since we'll be accessing the troves of Jack's data through his website. Point your browser to the following website to get started: http://strata.geology.wisc.edu/jack/

After reading Shanan's note and beginning your search (by clicking *"Begin your search*!") you'll see a dizzying array of Phyla and Classes with check boxes next to them. You'll also see a "drop-down" menu at the top of the page that will spit back groupings of data ("*All Genera*", "*Invertebrates Only*", etc.). The numbers in parentheses after the blue ALL-CAPS taxonomic groups indicate how many genera are in the Sepkoski database for that group. Notice that there is only one nematode worm and very few (n=3) comb-jellies of the phylum Ctenophora. Why do you suppose this is? _____
_____.

Clicking on any blue ALL CAPS word will return a list of all fossil genera within that Phylum/Class complete with their FAD and LAD in the stage binning names that Jack decided on. It is worth noting here that it is almost always the case that the paper that describes the first appearance of some genus is not the same paper that announces the last appearance of that genus. Just imagine how many paleontological papers Sepkoski must have looked through to produce this database! FOP stands for First Occurrence Period and FOS stands for First Occurrence Stage. Each of these is equivalent to "FAD". Likewise, LOP and LOS stand for Last Occurrence Period and Last Occurrence Stage respectively. Find the genus *Limulus* in the Order Xiphosurida among the phylum Arthropoda. This is the horseshoe crab (Figure 2). How long has this genus been around in "morphological stasis"?
_____

R = Recent
Q = Quaternary
T = Tertiary
K = Cretaceous
J = Jurassic
Tr = Triassic
P = Permian
C = Carboniferous
D = Devonian
S = Silurian
O = Ordovician
Cm = Cambrian
V = Vendian (latest Precambrian)

Table 1. Period abbreviations in Sepkoski's database.

When you check a box (or choose a "*Fauna*" from the drop down menu) and then click "*Submit*" at the bottom, the website spits back a different set of data that is the stage-binned, age-tied tallies of diversity. These stages are within the Periods that you know. Refer to your book's geologic time scale for help in figuring out what these stages are... The various columns after "Date" are described in the legend at the bottom of the data tabulation.

Figure 2. The common horseshoe crab (*Limulus*). Truly a "living fossil"!

For any of these "data dumps" you may copy all the cells in the web view by highlighting them all and then paste them into MS EXCEL for additional massaging and analysis. Practice this transfer.

Working with the Data

After you are savvy with the web interface that Peters has set up for Sepkoski's data, I want you to conduct a few analyses and reflect on them. Do all this work on your own over the next few days. All of this will be due Monday 3/9 in class.

1. Pick a class, any class (so long as it has >100 genera representing it in the database). Figure out what that class is. Check its box in Peters' website, spit out the data and construct a series of genus diversity curves and other plots for that group through time. Write up a one page (double-spaced) report about interesting patterns that you find. Try plotting all sorts of data either directly from the data dump or after some manipulation. Include at least three plots in your report and reference them in the text you write. One of them should be the standing diversity plot and the other 2+ are anything you find interesting.
2. Pick two phyla, any two phyla (so long as they each have >1000 genera representing them in the database). Bring the diversity data over to EXCEL for both and go about writing up a 1.5 page (double spaced) report that describes the similarities and differences between these phyla. Reference your numbered figures within. Include standing diversity plots for both (drawn on one graph) as well as other comparison data.
3. Select "*All Genera*" from the Choose Fauna drop down menu and bring those data over to EXCEL. Discuss (in one page, double spaced) these data as plotted in a standing diversity curve and in two other ways as well.

Tips for Producing Diversity/Origination/Extinction Diagrams

- We want our various plots/curves to be constructed with time getting toward the Recent from left to right across the screen/paper. EXCEL does this in the opposite fashion for whatever reason. A quick fix is to plot your data with young to the left and then *ctrl*-click or right click the X-axis and go to *Format Axis...* Within that window (*Scale*) there will be an option for you to reverse the order of the axis. This will fix that...
- To make it so that Age in Ma or Stage is the scale across the bottom, simply go to the Source Data by right clicking and then set Category X axis labels as the numerical ages provided in the data dump.
- If you go to View, Toolbars, Drawing... this will give you the ability to do some additional annotation on the figures you create. You can draw arrows to particular features or bracket zones, etc. Use these tools to really show what you have found.
- BE SURE YOU LABEL YOUR AXES IN A DETAILED WAY. TITLE THE GRAPH USING A SENTENCE LENGTH COLLECTION OF WORDS, NOT JUST "DIVERSITY", FOR EXAMPLE.

Example Grading Rubric:

	A - Excellent	BC - Good/Average	DF - Poor
Rationale for analysis (in written product)	Makes clear case for why a particular analysis was done --- cites previous knowledge from other courses or earlier in this course --- strong statement of expectation/hypothesis	Makes clear case for why particular analysis was done --- more than "I've always like snails so..." --- weak expectation statement	Weak or non-existent reason for running a particular analysis --- No statement of expectations
Data Usage/Display (in MS Excel figures)	Appropriate data for rationale --- sophisticated calculations and data manipulation --- plotted in such a way as to address hypothesis --- plots formatted correctly and relate to one another	Appropriate data for rationale --- only simple calculations and data manipulation --- plotted clearly but unclear how addresses statements in the rationale --- plots formatted with minor errors	Inappropriate data for stated rationale --- no calculations or data manipulation --- plotted in weird and inappropriate way --- plots formatted poorly with major errors
Analysis/ Interpretation (in written product)	Well written w/ no grammatical errors --- addresses what is learned from exercise in a detailed way --- discusses all plotted data --- poses additional hypotheses	Well written w/ few grammatical errors --- addresses rationale --- discusses all plotted data --- fails to pose additional hypotheses or lines of research	Poorly written w/ many grammatical errors --- detached from thoughts in rationale --- fails to pose additional hypotheses or lines of research --- fails to discuss all plotted data

Questions? email Dave Sunderlin - Lafayettte College - sunderld@lafayette.edu