

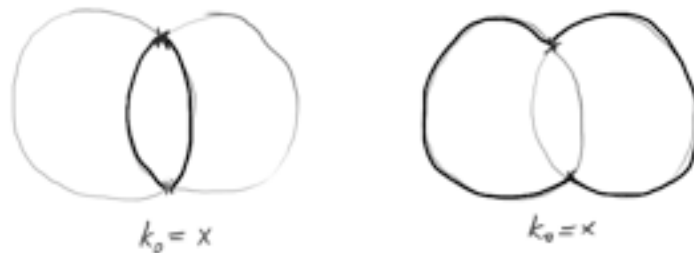
## The Week's News for Andy

### Progress

*Pre-Cenozoic radiolaria, diatom diversity: projects on hold*  
*Diatom FIB, diatom sonication: projects dropped*

*Diatom morphospace, valve outline approach [dormant since start of 2011]*

- Forged ahead in partnership with Sébastien in December on trying to create a morphospace of valve view outline shapes using a piecewise parametrization by sides. Sébastien ran into a number of problems, many of which exceeded my mathematical understanding, some of which were just within my grasp.



- Amongst the problems I understood: the same parameter values can describe different shapes. In the simplest case with only one parameter per side (i.e. one constant curvature) there are almost always two solutions (see above); the morphospace probably becomes more degenerate the more parameters there are.
- Further problems were more difficult for me to grasp fully: There's a problem with the scaling of lengths. Curvature depends on scale (scale up a circle and its curvature decreases, even though its shape stays the same), so the question is—what do you normalize the length scale to? My idea was to use the half-distance between the two vertices of the two-vertex case as the unit distance. This seemed intuitively appealing because the case of a circle would then have a curvature of 1. But here's the rub. If we then have a set of parameters (let's say, in a simple case, an arc of some given curvature  $K$  joining the two vertices), the curvature itself is expressed as the derivative of the tangent angle of the arc with arc length— $d\theta/ds$ . But what is the length of the arc,  $s$ ? Well, to find that you have to integrate over the curvature function... So we're stuck, because that in itself requires knowing  $s$ . Basically, if we have a parameter  $K$  for curvature, and we start out at a point and want to head towards the other point,  $K$  will tell us how curved our path of travel is. But the units of  $K$  are in terms of the total distance of the travelled path itself—i.e., for every 100th part of the total distance, turn  $x$  degrees to the left. But we don't know what that total distance is. We also don't know which direction you need to start heading out in to ensure that you will end up at the next vertex.
- Sébastien proceeded by defining the curvature of segments by using an *angle* (which is scale-invariant) rather than a curvature, similar to his bubble/coleochaete morphospace. This is easy to understand for the 1-parameter case, but I was unable to understand the extension of this approach for higher “modes” (2, 3, 4, etc. parameters).
- Met Sébastien December 13th for the last time. He showed me an implementation of this new approach that seemed to do an OK job, though I was unable to follow.

- My task was to work on the other part of the problem—taking images of diatoms, extracting the outline, and finding the set of parameters that best fits the observed shape. My attempts here were spectacularly unsuccessful. I was unable to find a straightforward description of *how* outlines are extracted from raster (pixel) images—the descriptions were all very mathematically sophisticated and involved complex and case-specific smoothing functions to make up for the fact that the outline is always described in very discretized steps of one of four (or eight) cardinal directions (adjacent pixels). My discussions with Sébastien and Jacques about this were also rather dispiriting, because the process of finding the appropriate parameters will not simply be a matter of tinkering with the parameters until the shapes look right (which was basically what I had in mind). It would actually involve writing an algorithm that compares the outline generated by a given parameter set with the observed outline, calculates some error function, and then iterates through parameter space until an optimum fit / minimum error is found.
- Sébastien was gone on vacation at the start of the year, so I moved on to the character-based morphospace, and haven't returned to this project—it has essentially been dormant since the start of 2011.

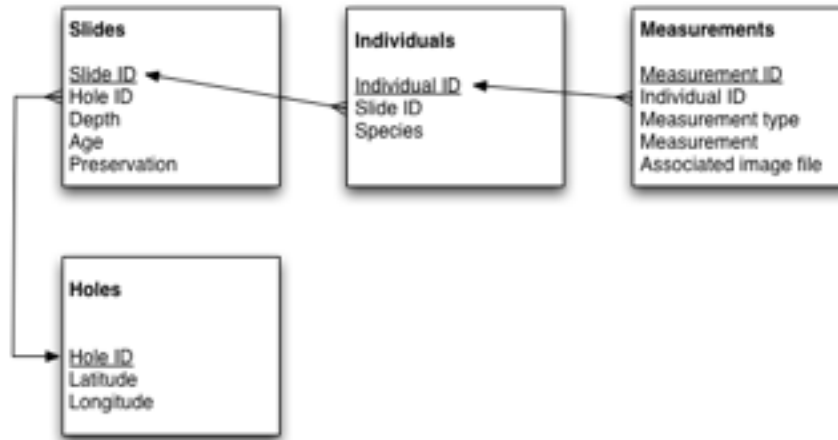
#### *Diatom morphospace, discrete character approach*

- Gave this project my full attention starting January 3rd. Finished assembling the list of characters on January 20th.
- I attempted to avoid artificially introducing taxonomic structure to the morphospace by paying particular attention to formulating my morphological characters in as general a way as possible, to avoid taxonomically biased terminology. In many cases this required going counter to established practice in how diatoms are described. For many aspects of diatom morphology, the same shapes or structures are given different names in the literature, or are grouped in different categories depending on what taxonomic groups they are in—even if the shapes or structures are clearly homologous.
 

*(As an example, the pores and their arrangement on diatom frustules of the two major groups (centrics vs. pennates) are described by two essentially non-overlapping vocabularies, although the structures are obviously comparable and almost certainly homologous (although, for a morphospace, homology should not be a guiding feature, in my opinion). So instead of constructing separate categories for striation (for pennates) and areolation (for centrics), and applying the associated established sets of terminologies for describing them, I have referred to these openings generically as "pores" and compiled a set of characters that applies as much as possible to both centric and pennate diatoms.)*
- While these choices have led to a long list, and will make the morphospace characters more challenging to read for diatom specialists (who will be accustomed to seeing structures described in a particular way), I hope it will provide a more taxonomically agnostic view of shape, and an honest depiction of how morphology is explored through time.
- Sent this out to Wiebe Kooistra, who referred me on to Heroen Verbruggen and David Mann. Mann did not reply upon first or second attempts; Heroen replied with positive comments and good questions, which I was able to answer (at least to my own satisfaction). His only substantive advice was to show the list to “show this to people that are working with diatoms if you have not already done so”.
- Not sure whom to contact, if anyone. Talked to Mike Foote while he was here, who suggested blazing ahead without heed to the authorities, who might be blinded by taxonomic thinking—which he strongly felt I was right to shield myself from.
- This project is essentially ready for “churning”—filling out the data matrix with character codings.

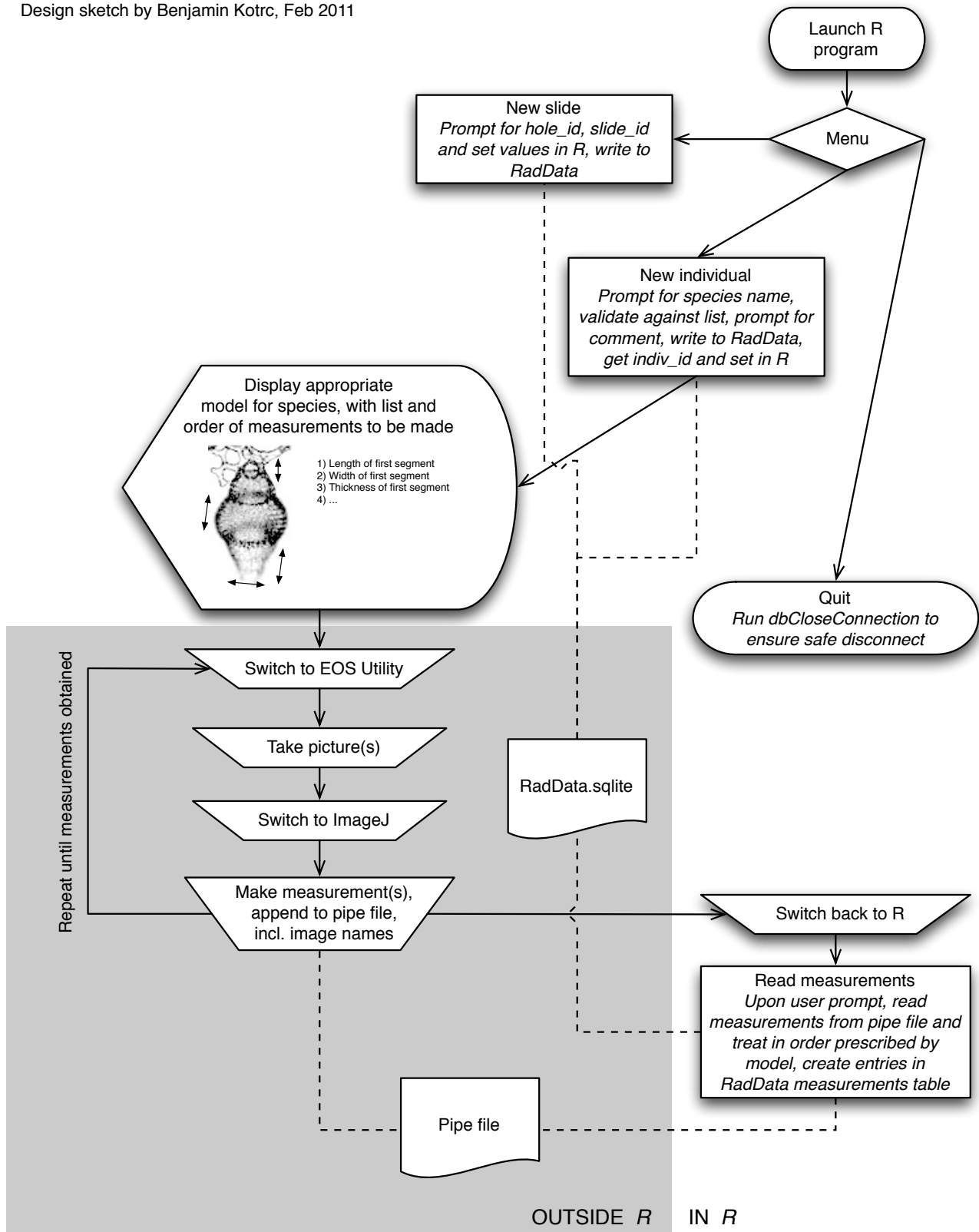
### *Radiolarian silicification, by-lineage*

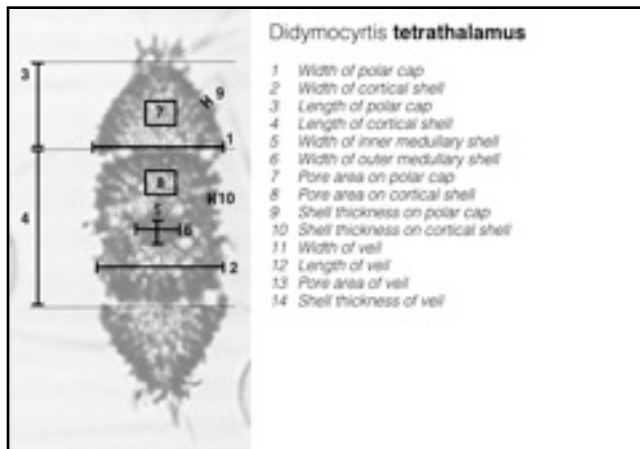
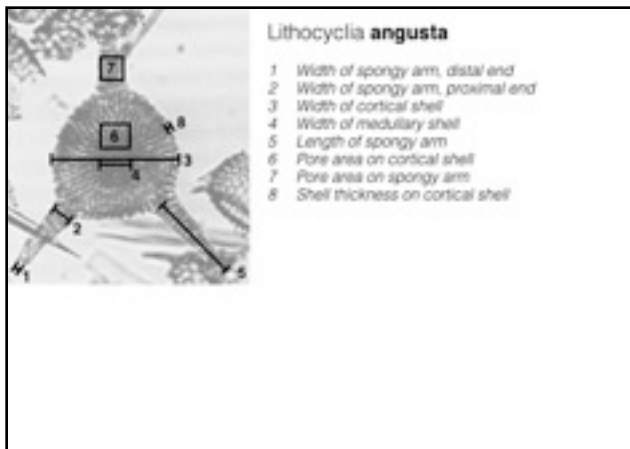
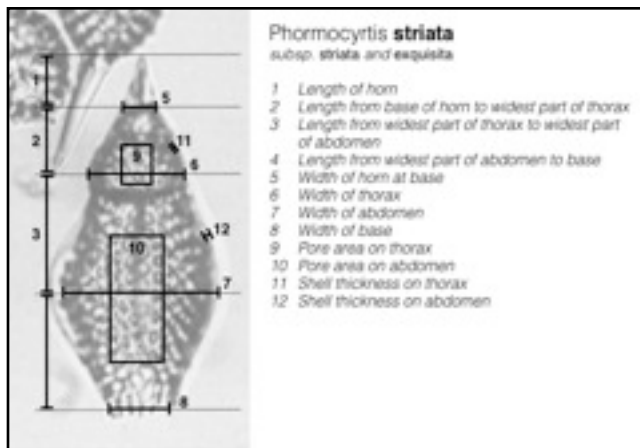
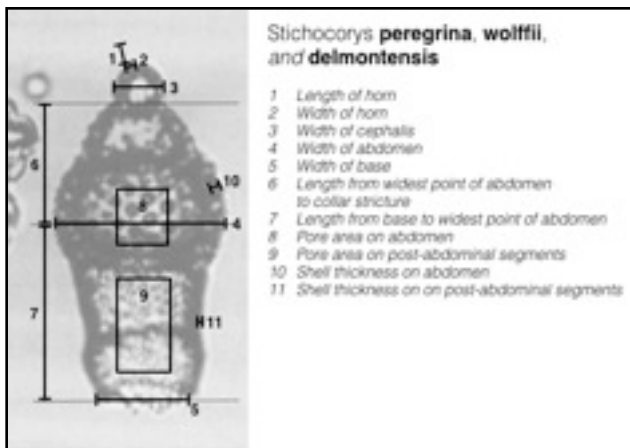
- Restarted this project January 21st. Followed Dave Lazarus' advice—and what I learned from my negative experiences with unwieldy Excel spreadsheets—and decided to create a relational database to hold the morphometric measurements I will make.



- Spent until around January 31st learning the basics of relational database design, SQL (a programming language for relational databases), and its implementation in/interaction through R, called RSQLite. In parallel, designed the schema of the database (above), which consists of separate tables that hold data describing entities of a particular kind, and are related (hence the “relational” aspect of the database) and thereby linked to other entities of different kinds, whose data are stored in separate tables.
- By February 7th, had the SQL database up and running in the form shown in the schema above.
- The next step was to design a measurement protocol and, as a part of that, an interface with which I will interact with this database. In its raw form, inserting information into a SQL database is clumsy (requires typing instructions and data in at the command line, which besides being slow introduces a lot of room for error), so I decided to design a more elegant solution to greatly speed up data collection (see design sketch overleaf).
- Spent February 9th–11th figuring out how to get images to display in R, so that I can display a graphical instruction set outlining the order of measurements to be made for each of the different species I will encounter. Since each species will have a different geometric model for estimating silica use associated with it, each species will require a different set of measurements to be made—this will make it clear which measurements need to be made for the specimen at hand.
- Spent February 11th–13th collecting the taxonomic literature describing each of the 21 species in the 5 lineages I am initially going to study, from the material on loan from the Smithsonian MRC library.
- Spent February 14th–17th coming up with geometric models for each of the 21 species, a prerequisite for the rest of the interface development since I need to know which measurements need to be made for each of the species.
- Spent February 18th–23rd turning these models into the graphical prompts for the interface. A couple of examples are shown two pages ahead.

**RadData** relational database interface for collecting Radiolarian morphometric data  
 Design sketch by Benjamin Kotrc, Feb 2011





- Spent February 24th–March 1st writing macros for ImageJ, the program in which the actual measurements on digital images from the microscope camera will be made. The macros handle setting the magnification scale for the images, keeping track of which measurement number you’re on, calculating area proportions, and getting the measurement data from ImageJ to R.
- Completed the first working implementation of the interface—with all bits from ImageJ to R to SQL in place—on March 7th. A little over six weeks, all told. A big time investment, but hopefully worthwhile in the long run, considering how much data acquisition can be accelerated.

### Other updates/questions

- PlanktonTech know I can’t make it to their meeting, but they asked for a short (no more than ½ a page) general-audience report on progress in our research program. What to include?
- Kati and I are married (but we’re keeping it a secret until the wedding in June—shhh!), so the “adjustment of status” petition incl. ~4,3 kg of paperwork have been submitted to DHS.
- Heroen Verbruggen will be around for NEAS meeting at MBL (which is April 15-17), we may get him to come and visit, and perhaps give a talk at lab meeting (depending on his timing, Friday April 8 or 22, or otherwise on another day of the week)?

**Short-term goals & strategy**

- Finish a few remaining details on the radiolarian measurement database interface—a function to allow making regular back-ups of the database file, a macro to include the magnification for each image in the image file name,
- Although it's ready to go, I'm planning to put the radiolarian project on hold (parking on a downslope!) and return to the discrete character diatom morphospace, to attempt to finish that project as quickly as possible.