GEOMETRIC MODELS OF BIOLOGICAL PHENOMENA

organized by
Louis Billera, Susan Holmes, and Karen Vogtmann

Workshop Summary

The scientific goals of this conference were to initiate and foster communication between four disparate groups of people all interested in the study of phylogenetic trees. The four main target groups were biologists, probabilists/statisticians, geometers and combinatorialists. To help attain these goals, we chose representatives from each of the four groups to present a "tutorial" the first two mornings, aimed specifically at the group farthest from their specialty; these speakers were chosen especially for their ability as communicators as well as for their prominence in the field, and we were in general quite pleased with the results. A relatively small number of additional speakers presented more traditional lectures on their work; we found that these speakers, too, perhaps influenced by the "tutorial" lectures, made an effort to reach all members of the audience. Finally, we scheduled a discussion session each of the first four days, to focus on issues in the four fields related to the study of phylogenetic trees. These discussions were quite lively and animated, and generated and articulated a number of interesting problems for further investigation. Excellent notes on these discussions were taken by Francis Su, and are now posted on AIM's web page. In particular, they give a very good exposition of the questions raised, and of the future directions.

Su's notes on the final discussion on Thursday contain discussion of many of the issues raised at the conference, and the following is a summary of these notes.

The organizers were encouraged by the degree of interaction and amount of discussion which took place during the conference. Several groups seemed to find interesting problems to work on, and gained significant insight into the perspective of other groups.

Diaconis cited an example of residuals for discrete data analysis, as something worth following up, as well as the discussion in the geometry session about what are natural measures on tree space. He noted that the workshop was valuable because he couldn't think of another setting in which all these ideas from different arenas could all come together.

Other problems and issues that were mentioned in the final discussion included:

The idea of concatenating sequences as an "averaging" process Billera noted that in the discussion about averaging, he was struck with the idea that concatenation was an average in sequence space, an idea that he hadn't appreciated until this workshop. He noted that the conference had touched on many ways to go from sequences to trees, and one of the big questions was whether these methods were coherent with concatenation-averaging.

Penny said that it is worrisome that there are many different ways of concatenating, and it is not clear how to handle an increase in the number of parameters.

How can one determine likelihood contours in tree space?

There was discussion about confidence sets and likelihood contours, and how these might intersect transition points where the tree space branches in several directions. For instance, if a maximum likelihood process produces a tree with very short edge, then it would be near a transition. There was also discussion of whether the notion of curvature on tree space would be helpful for determining likelihood contours.

Is there a notion of a “multivariate metric” on the space of trees?

Susan Holmes said that in the data that arises for trees, you get *multivariate* distances, i.e. distances with several values for one tree, and asked whether there is anything similar in geometry. Diaconis mentioned that there are some distances that take values in partially ordered sets, and ordered abelian groups.

Are there alternate ways to represent phylogenetic relationships (i.e. not trees)?

Epstein asked if are there other objects besides trees that would be helpful for understanding phylogeny, and whether making a space of such objects would be useful. St John mentioned *networks*, which are graphs with several non-tree portions, but coarsely tree-like.

What does a random walks on tree space look like?

Vert said that one can imagine a random walk in the space of trees; one would start at a tree, generate data, then estimate a new tree (by maximum likelihood, e.g), etc. He suggested that it might be interesting to study this walk.

Bi-colored trees.

In response to Snel’s talk, Shareshian asked whether it would be useful to study spaces of bi-colored trees, in order to model orthology and paralogy.

During the session, Diaconis suggested that, as a follow up to this conference, people can now get together in small groups and communicate and start to work on a problem, whereas this wasn’t true before the beginning of the week. AIM could facilitate such follow-up meetings.

There was some discussion about holding a follow-up conference to this one, since all the participants now have a solid foundation on which to try to work on some joint problems.