

Detecting Convergence of Bayesian Searches in Computational Phylogeny

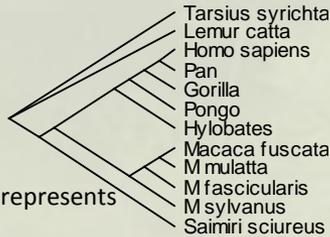
William Dawson, Grant Brammer, Dr. Tiffani Williams

Department of Computer Science and Engineering

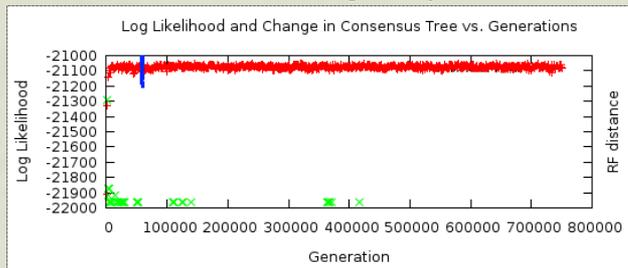
Objective

The Issue

- Computational Phylogeny is Concerned with Building Trees of Life.
- But what particular tree best represents the data?
- Requires a heuristic search – keep trying various trees trying to find the best one.

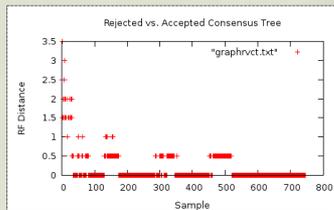


Objective: develop stopping criteria for Bayesian searches. Current criteria can't get the job done.



Results

Data sets were run collecting data both for metrics and tracking changes in the consensus tree for comparison.



Taxa Set	Last Change	Breadth Complete
Anolis Lizards	522000 gens	523408 gens
Beetles	530000 gens	747714 gens

Generated with 1000 anchor points.

Methodology

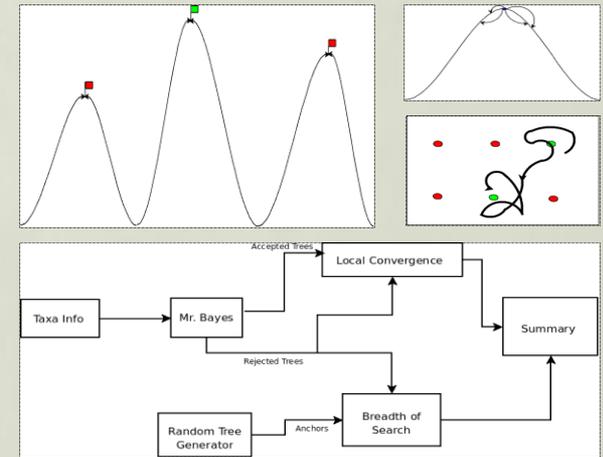
Think about searching for the highest mountain.

We need to know when we're at the top of one mountain.

> Can use the consensus tree of the rejected proposals because of way we walk through space.

And we want to make sure we've looked at all the mountains.

> Set up anchor points and check off which we have gone near.



Discussion

The opposite of convergence – is being stuck in one region of space. Meaning when the chain gets unstuck a consensus tree change will happen. Thus we check convergence by looking for a lack of “stuckness”.

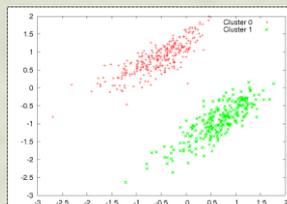
Solve a recognized problem - Many times researchers will run the program multiple times because they are concerned they've gotten stuck in certain regions. Checking the breadth quantifies this need by reporting the percentage of tree space covered.

Challenges with checking breadth of search – tree space is vast, it would be useless to check the breadth by looking at the percentage of trees searched. Instead we need to look at the variety of trees searched, and anchor points are the means for this.

Exploit Search Technique - The Markov Chain Monte Carlo tree space wandering technique ensures that when we're near an anchor point, we're there for a good amount of time.

Made possible by new fast algorithms – HashCS, HashRF.

Future Work - focus on optimizing the current metrics, exploring tree distance metrics, and using metrics to create smarter searches.



Conclusion

The metrics were shown to be both predictive and conservative, demonstrating practical use for detecting convergence.

