

Individual Development Plan 2021

Will Dumm

Career Goals

Long-Term and 10-year Goals

In the long term, my goal is to be productive in a flexible, self-directed but collaborative, location-agnostic position contributing as a programmer to an interesting project in science or applied math. This job will be:

- senior-level
- flexible, in the sense that I have some freedom to pursue projects that interest me
- creative, not purely an application of technical skills
- related to a goal or cause that is important to me

Getting there will be mostly a journey of self-cultivation, rather than credentialing and experience building. My current position offers most of these characteristics already, but I can learn to take advantage of them better. For example, I can use my freedom in my position to be more creative about how I approach my projects. I think this long-term goal is achievable within the next ten years.

This Year's Goals

Within the next year, my goal is to identify and execute a set of milestones for promotion in my position. These milestones will include:

- Becoming a proficient c++ programmer
- Building familiarity with relevant data formats and structures, such as the intricacies of variant calling format and the Usher Mutation Annotated Tree
- Completion, at least in the scope of the history-DAG and parsimony, of the Parsimony Plateau paper as a coherent, well-rounded, and useful document
- Implementation of useful applications of the history-DAG, including

- a deep parsimony search using Usher utilities, applied to SARS-CoV-2 data. It will be interesting to focus on how much better we can do than the ‘accepted’ ancestral tree for SARS-CoV-2 constructed by Usher.
- a useful application to GCTree inference, hopefully optimizing inference via the DAG structure.
- Progress in understanding and implementing the ideas described in the ‘conditional damara distribution’ grant application.

Development of Project-Specific Knowledge

Project Description

My projects for the foreseeable future revolve around the description of a compact object (the history-DAG) capable of describing an ensemble of labeled trees, organized by associations of internal labels with leaf label partitions. In the near-term, this will involve exploring the properties of the history-DAG constructed on weighted trees, especially maximum parsimony trees.

In the medium-term, I will focus on developing applications for this object. For example, properties of the history-DAG could be useful in expediting a search for maximally parsimonious trees. The history-DAG will likely also be useful in compactly expressing a large number of trees, both in the Bayesian setting of posterior distributions with large support, and perhaps also as a way of characterizing the space of all max-parsimony trees on fixed taxa.

I will also explore ways to use the history-DAG to improve inference in the package GCTree. Immediate benefits may be realized by using the history-DAG to efficiently utilize a larger number of maximally parsimonious trees in the existing inference pipeline. It may also be possible to streamline GCTree inference algorithms by making use of the structure of the history-DAG.

As a related project, I will seek to extend the group’s work on the subsplit-DAG to allow use of the history-DAG or another new structure to constrain the search space in the variational inference problem described in the ‘conditional damara distribution’ grant application.

New Skills and Knowledge Gained

- Biology background, including basic immunology and phylogenetics
- Mathematical phylogenetic background, including parsimony, basic understanding of applications of MCMC and variational inference to Bayesian phylogenetics
- Re-familiarization with Python, DAG-related algorithms, shell, and cluster use
- Basic understanding of many of the group’s projects

Skills and Knowledge Needed

The following skills will be needed to tackle my projects:

- Familiarization with c++, including basic knowledge, productive problem-specific idioms, and toolchain proficiency
- More familiarity with variational inference applied to phylogenetics
- More familiarity with the branching process model and how it's used in GCTree
- Detailed understanding of the Usher MATUtils and MATOptimize code-base
- Continued progress in understanding the history-DAG and its implications

Career Skills

Communication Skills

So far, opportunities to practice and improve communications skills have been limited to writing and informal/impromptu verbal communication of project ideas and questions. I could improve my general readiness to explain project ideas without preparation. One way I may do this is by verbalizing each day a new idea that I've had during that day, even if just to myself. This will also help me prepare for more formal exercises in verbal communication, such as group meeting presentations.

The most immediate and accessible improvements in writing skills will be automatic, as I become more familiar with the notation and vocabulary that is most common in phylogenetics, and that works best with my project. I also intend to mindfully implement Erick's suggestions for writing from the group wiki.

Other Opportunities for Improvement

I expect to improve my time management in this position:

- Keeping sight of immediate goals for each day or subproject, and staying on track
- Knowing better when to shift attention when progress on a task slows
- Having a more complete array of useful tasks in mind, and better recognizing directions with high potential for progress

I can address the first two items by pausing to check in with my goals and progress throughout the day. The third can be helped by continuing to write down any ideas or questions that come up while working on other tasks.

Opportunities for Contacts and Collaboration

I am looking forward to communicating with authors of Usher about my project, when I have ideas that are ready to share, and questions that require their help. Erick also mentioned communicating with Mike Steel regarding parsimony on the history-DAG.