

## Gene phylogeny reconstruction based on whole genome shotgun sequence data.

Tomoaki Nishiyama<sup>1,2\*</sup>, Takako Tanahashi<sup>3,4</sup>, and Mitsuyasu Hasebe<sup>2,3,4</sup>

<sup>1</sup>Advanced Science Research Center, Kanazawa University, 920-0934 Kanazawa, Japan

<sup>2</sup>ERATO Hasebe Reprogramming Evolution Project, Japan Science and Technology Agency, Okazaki, Aichi 444-8585, Japan

<sup>3</sup>National Institute for Basic Biology, 444-8585 Okazaki, Japan

<sup>4</sup>School of Life Science, The Graduate University for Advanced Studies, Okazaki, Aichi 444-8585, Japan

\* e-mail: tomoakin@kenroku.kanazawa-u.ac.jp

Since whole genome shotgun (WGS) sequence data of *Physcomitrella patens* and *Selaginella moellendorffii*, which are pivotal taxa to understand land plant evolution, became available, we started a project to perform phylogenetic analysis on every gene that is important for the development in angiosperms. For the reconstruction, related sequences are automatically collected from *P. patens* and *S. moellendorffii* WGS sequences, *Cyanidioschyzon merolae* and *Chlamydomonas reinhardtii* predicted proteins, and the NCBI nr dataset. The collected sequences are aligned and the alignment is converted to a nexus file with hint information for which region to use for the analysis. The file can be edited with MacClade to remove excess genes and reselect for the region to use for down stream phylogenetic analysis. Based on the edited nexus file, a Neighbor-Joining tree with bootstrap test and a maximum likelihood tree are obtained. The taxonomy data are extracted from the NCBI taxonomy database. The trees are rerooted based on the distances of branches or taxonomic distributions and scalable vector graphics files with coloration according to the taxonomy are generated. The accession no., the organism name and the definition are written as the OTU label. The automation of sequence collection and tree drawing allowed researchers to concentrate on the judgment of alignment. Several hundred genes have been analyzed with this system and some example will be shown.