TOWARD RESOLUTION OF GREEN PLANT PHYLOGENY

Pls: D.F. Mandal, J. Boore, M. Donoghue, R. Olmstead, K. Renzaglia, A. Smith, P. Wolf, B. Mishler, & C.J. O'Kelly

Addresses for Pls and our project home page are on the back of the mini-poster handouts below.

The tree of life is inherently fractal. Look closely at one lineage of a phylogeny and it dissolves into many separate lineages, and so on down to a very fine scale. A great body of phylogenetic research has provided numerous tools applicable at particular, usually fairly constrained, scales. These tools have left many phylogenetic questions unanswered. We think these questions will remain unanswered until it is possible to do analyses across multiple scales.

The green plant lineage represents the most suitable system for this research because it is one of the oldest and most diverse branches of the tree of life and it contains prototypical examples of the known phylogenetic problems. Investigations on it may draw on a tradition of interdisciplinary collaborative research, facilitated by the Green Plant Phylogeny Research Coordination Group (GPPRCG, "Deep Green").

Our overall objective is to resolve the primary pattern of evolutionary diversification among green plants, and establish a model for doing so that will be applicable to other groups of organisms with long evolutionary histories. We believe that a solid backbone based on genomic and ultramolecular data for selected exemplar taxa will enable the integration of previous and ongoing studies of many more taxa into a comprehensive picture of green plant phylogeny.

To achieve this objective, we will:

- use molecular biology and genomics to a) sequence the genomes of chloroplasts and mitochondrial genomes, b) develop a Bacterial Artificial Chromosome (BAC) nuclear genome libraries from 1C 100Mb for each major lineage of green plants, c) determine the evolutionary relationships of the chloroplast genes using sequences from these libraries, and d) groundtruth the resulting phylogenies with existing nuclear gene data;

- produce a comprehensive set of comparable morphological and ultrastructural data for those major lineages of green plants;

- incorporate inferences from across the phylogenetic hierarchy in green plants using methods designed to permit scaling across studies.

![Figure 1](image)

**Figure 1.** Outline phylogeny of the green plants, indicating the currently understood phylogenetic position for 51 candidate exemplar lineages (represented by generic names) and the higher-level taxa to which they are thought to belong. All branches are subject to further testing, but the best-supported clades are indicated with thick lines, branches with some support are indicated with thin lines, and areas of the tree that remain controversial are indicated by labelled ovals.

1. Base of green plants: outgroup relationships and basal branching among prasinophytes. 2. Relationships among major lineages of Chlorophyta (some groups are more firmly established, e.g., Ulvales/Ulotrichales, Chlorophyceae/ Trebouxiophyceae). 3. Relationships among the chlorophycean algae and their placement in the Chlorophyta. 4. Base of the land plants. 5. Base of the ferns (moniliforms).

**MOLECULAR BIOLOGY & GENOMICS**

- Spherical chloroplasts?
- Gal purify org DNA?
- Identify clones
- FAC5s-RCA
- BIG DNA?
- Small insert limited library
- Large insert BAC library
- Shotgun
- Sequencing & gap fill
- Assemble & annotate
- Assemble trees

**SCALING ACROSS STUDIES**

The blue light photoreceptor cryptochrome (CRY) was chosen to investigate nuclear gene phylogeny as blue light penetrates both aquatic and xeric environments. In seed plants, a clade including all flowering plant CRY2 sequences was found sister to a monophyletic gymnosperm clade (CRY2). Performing plant CRY1 sequences formed a monophyletic clade sister to the seed plant CRY2 clade. A second gymnosperm CRY clade (CRY17) was recovered and was found unresolved with one of two CRY clades and the clade containing the remaining seed plant sequences. Thus far, four unique clades were recovered for Marsilea, consistent with the diversity of cryptochrome genes previously reported in Atlantaria.

The "prasinophyte" genus Phaeocystis and its relatives contain what may be the most ancient green plants (Group 1). Phaeocystis species are found in oceanic waters worldwide. Swimming cells (Panel A) form in large, oxygen-rich cysts ("phytocysts"); Panel B, Phycomastigina congregate within cysts; Panel C, Alveolates are present in the swimming cell surface; similar scales are found in the most ancestral members of all algal lineages. This clade contains the prasinophyte, a prasinophyte with derived morphological and ultrastructural characters relative to Phaeocystis, conserve many eubacterial elements not found in other plants.

**SUMMARY OF OUR PROGRESS**

- We estimate we have >25 genomes in various stages of FACS-RCA, and 18 genomes from BACs in progress.
- A single clone was recovered for Synechocystis and for Anoarcheae suggesting that diversity in trichophytes in the nuclear gene, CRY, originated after the divergence of trichophytes from their phototrophic relatives.
- Among autotrophic algae, there are some domains in which gene sequences evolve more rapidly than the chlorophycean algae into which these sequences are not found. This phenomenon obscures relationships; for example, both types of domains are found among algae at the base of the eukaryotes, leading to the impression that this clade is not monophyletic.
- We have established a framework for managing, aligning, and analyzing our genomic data. Further genomes will be fed into our database as they are completed. Firm phylogenetic conclusions will be drawn after the larger set of genomes is completed.
- Outreach activities have included a special symposium for student presentations at the 2004 meeting of the Botanical Society of America (BSA), a workshop for teachers at the 2002 meeting of BSA, and the development of teaching materials on cryptophyte morphology and phylogeny for use at the secondary school level.