Genome Improvement of DOE JGI Flagship Plant Genomes

David Sims[1]* (dsims@hudsonalpha.org), Jerry Jenkins[1], Wellington Muchero[2], Gerald A. Tuskan[2, 3], Jim Bristow[3], Kerrie Barry[3], Daniel S. Rokhsar[3], Jeremy Schmutz[1,3], and Jane Grimwood[1,3], and **Richard M. Myers**[1]

[1] DOE JGI HudsonAlpha Genome Sequencing Center, Huntsville, AL
[2] BiosciencesDivision, Oak Ridge National Laboratory, Oak Ridge, TN, 37831, USA
[3] DOE Joint Genome Institute, Walnut Creek, CA, USA

Project Goals:

As part of the DOE JGI Plant Program, we are working to experimentally and computationally improve the genomic sequences of key DOE plant species which we have designated the Plant Flagships. These genomic sequences and related data support mission aims in producing cellulosic feedstocks and understanding carbon cycling and carbon sequestration.

Through discussion with the JGI Plant Genome Advisory Committee and DOE, a set of JGI plant genomes that are the most important to DOE mission and plant science have been designated as JGI Plant Flagship Genomes. This selected set of critical species allows us to focus our computational and experimental efforts to move beyond sequence to function and to provide the most direct benefit to mission science. The current Plant Flagships are:

- Poplar – the DOE tree, the basis for cellulosic research at ORNL
- Sorghum – widely planted grass crop for biomass, cellulose, and sugar
- *Brachypodium* – small grass model organism
- *Chlamydomonas* – the most studied algal species, model algal organism
- Soybean – the source of biodiesel and the number two US economic crop
- Foxtail millet – a grass model, recently evolutionary diverged from switchgrass
- *Physcomitrella* – moss model organism, basic comparator for land plants

Although the flagship designation was initially intended for genomes that had already been draft sequenced, we also consider several additional genomes as proto-flagships because of their importance to DOE biofuel mission; initial sequencing efforts are on-going along with the creation of genomic resources:

- Panicum virgatum (switchgrass) a candidate biofuel feedstock that grows on marginal soil and is being used by all of the BioEnergy centers as a model crop species
- Miscanthus – a perennial grass species that produces large amounts of cellulosic material with low agricultural inputs, one of the top feedstock candidates

- Panicum hallii (Hall's panicgrass) – a small, evolutionary nearby diploid relative of switchgrass that may serve as laboratory model organism for switchgrass research

For genome improvement, the objective is to completely (or as near as possible) resolve each base pair within a localized region and then using mapping resources, position and incorporate the sequence into chromosome scale pseudomolecule assemblies. In some cases, we have needed to construct the mapping resources ourselves to push forward a chromosome scale release. For base pair level improvement, which we carry out for the entire gene space of an organism or directed regions, we collect experimental data in the form of primer walks with a variety of chemistries and templates, transposon mediated sequencing and shatter libraries from clones (the latter two techniques being essentially subprojects of the subprojects). We have ongoing efforts to explore practical uses of next-generation sequencing technologies for improving the extremely large and polyploid plant genomes. We release new versions of the Plant Flagship genomes periodically as we complete significant iterative improvements. The sequence and subsequent annotation is made public through the phytozome web site at www.phytozome.net.

We have recently completed major updates to the sequences of *Chlamydomonas*, *Physcomitrella* and poplar, all of which are currently in the annotation phase. We are proceeding with a directed improvement project of sorghum gene space and a complete improvement project for *Brachypodium.* With these improved sequences we seek to 1) provide near complete information for genomes of direct DOE mission importance, 2) support BRC and others efforts for cellulosic biofuel development and plant customization efforts, 3) foster communities to develop scientific research programs working with DOE organisms and attract new scientists to solve DOE problems by providing tool sets to work with these organisms, and 4) build a solid foundation for diversity, phenotype, resequencing, transcriptomic studies, and functional studies in these organisms.