

## 30. De novo assembly of a *Setaria viridis* reference genome

Jerry Jenkins,<sup>1</sup> Jeremy Schmutz,<sup>1,2</sup> Malia Gehan,<sup>3</sup> Tom Brutnell,<sup>3</sup> and Todd C. Mockler<sup>3\*</sup>  
([tmockler@danforthcenter.org](mailto:tmockler@danforthcenter.org))<sup>1</sup>

<sup>1</sup>HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, <sup>2</sup>US Department of Energy, Joint-Genome Institute, Walnut Creek, California, and <sup>3</sup>Danforth Plant Science Center, St. Louis, Missouri

<http://sviridis.org/>  
<http://foxmillet.org/>

**Project Goals:** Bioenergy grasses promise to provide a sustainable source of renewable fuels for the US bioenergy economy. To develop bioenergy grasses with the desirable traits needed for large scale production, it will be necessary to develop model systems that are closely related to bioenergy feedstocks, but which are more amenable to genetic analysis. One of the most promising model species is *Setaria viridis*. *S. viridis*, like all major feedstock targets, is a C4 panicoid grass. However, it is much smaller in stature, flowers within 6 weeks of planting, and can be easily transformed with genes of interest. The objectives of this project were to develop a genomic reference sequence for *S. viridis* that will provide insights for engineering and intelligent breeding of improved food, fuel, and fiber crops. Further, the *S. viridis* genome assembly will serve as a useful reference for comparative genomics in the grasses.

The assembly of the *S. viridis* genome was based on 118x of Illumina whole-genome shotgun sequencing data and performed using ALLPATHS. The 1,417 marker map and *Setaria italica* synteny were used to identify a total of 21 misjoins. Scaffolds were then oriented, ordered, and joined together based on the map and synteny, and assembled into 9 chromosomes. When the marker map and synteny were in disagreement the map was given priority. A total of 232 joins were made during this process. The final set of pseudomolecule chromosomes were numbered and oriented to match the *S. italica* chromosomes. A total of 50 gaps in the genome assembly were patched using ~18x coverage of PacBio reads, adding 34,444 bases to the assembly. Patching consisted of assembling the PacBio reads crossing a gap using QUIVER and then integrating the assembled sequence in the genome at the gap. Additionally, 84 homozygous SNPs and 144 homozygous INDELS were corrected using ~70x of independently collected Illumina reads to polish the final assembly. The initial release of the *S. viridis* genome is composed of 12,531 contigs assembled into 9 chromosomes, along with 724 unintegrated scaffolds. The main genome contig N/L50 is 1597/65.5 Kb and the scaffold N/L50 is 4/46.1 Mbp. The main genome scaffolds span a total of 394.9 Mbp.

*This work was supported by the Office of Biological and Environmental Research in the DOE Office of Science (DE-SC0008769).*