Biological Design of *Lemnaceae* Aquatic Plants for Biodiesel Production

Evan Ernst*,[1,2] James Birchler,[3] Eric Lam,[4] Jorg Schwender,[5] John Shanklin,[5] and **Robert A. Martienssen (martiens@cshl.edu)**[1,2]

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; [2]Howard Hughes Medical Institute, Cold Spring Harbor, NY; [3]University of Missouri, Columbia, MO; [4]Rutgers University, New Brunswick, NJ; and [5]Brookhaven National Laboratory, Brookhaven, NY

Project Goals:

1. **Leveraging our transformation methods, we will develop a comprehensive toolset for genetic manipulation of *Lemnaceae*. We will establish CRISPR/Cas9 genome editing to complement our artificial miRNA silencing methods. We will construct artificial chromosomes in *Lemna minor* to potentiate whole pathway engineering.**
2. **Resting and over-wintering fronds have higher starch content than corn kernels, but the energy density of oil is more than twice that of starch. We will use regulatory network and metabolic flux modeling to re-engineer the carbon allocation pathways to optimize triacylglyceride (TAG).**
3. **We will use comparative genomics of multiple *Lemnaceae* genome sequences, an extensive living collection of global accessions, and systems network analysis to characterize gene expression networks underpinning developmental and environmental responses to maximize bioenergy products while preserving rapid biomass accumulation. Nutrient deprivation and CO_2 irrigation will be used to enhance yield.**

Lemnaceae species (commonly called duckweeds) are the world's smallest aquatic flowering plants. Under optimal conditions, their rapid clonal growth rate can double the number of fronds in 30 hours and produce 64 grams of biomass per gram starting weight in a week, which is far beyond that of terrestrial crops such as corn (2.3 g/g /week), and unencumbered by secondary products such as lignin. *Lemnaceae* offer an attractive alternative to algae as biofuel feedstocks because of their robust growth in open ponds and the relative ease of harvesting dry material. Convenient metabolic labeling in culture makes Lemna a good system for pathway modeling and engineering, as nutrients are taken up from liquid growth media, and non-responsive stomata can utilize very high levels of atmospheric CO_2. Our goal is to divert a substantial portion of accumulated carbon from starch to oil metabolism in *Lemnaceae*, using resting fronds as the storage tissue.

Recent efforts in the Martienssen and Lam labs have produced three new reference quality *Lemnaceae* genome assemblies complete with chromosome structures, methylomes, small RNA transcriptomes, and structural variant analysis for these novel genomes. We performed single molecule long-read genome sequencing of diploid *L. gibba*, diploid *W. australiana*, and allotetraploid *L. minor* clones using Oxford Nanopore technology followed by Hi-C. Comparisons of the resulting chromosome-scale assemblies reveal that the 21 chromosomes of diploid *L. gibba* are highly colinear with each of the subgenomes of allotetraploid *L. minor*,

while the *W. australiana* has 20 chromosomes with significant architectural differences with the 20 chromosomes of the giant duckweed *S. polyrhiza*. *S. polyrhiza*, *L. gibba*, and the two subgenomes of *L. minor* all encode around 18,000 genes – significantly fewer than terrestrial monocots such as rice and Brachypodium, and comparable to the unicellular alga *Chlamydomonas reinhardtii*. *W. australiana* has undergone an even more striking reduction to only 14,000 genes. Whole methylome sequencing has shown that *Spirodela polyrhiza* has among the lowest cytosine methylation levels in plants and lacks CpG gene body methylation. Coincident with reduced methylation, *S. polyrhiza* has very little retrotransposon sequence, while *L. gibba* and *L. minor* retain a similar retrotransposon content to other monocot genomes, and *W. australiana* is intermediate. Small RNA sequencing has revealed dramatic differences between the three genera consistent with known pathways of RNA directed DNA methylation. We have analyzed orthologous gene content across the *Lemnaceae*, 11 other monocots and 9 non-monocots, revealing variations that likely account for some of these differences, as well as for reduced morphology, clonal reproduction, and aquatic growth habit.

The Birchler Lab has completed the design and construction of a transgene stacking system with alternating transformation vectors that enable iterative recombination into a specific genomic site specified by a previously integrated landing pad. This novel design is compatible with consecutive transformation of *Lemnaceae* undergoing strictly clonal propagation. In addition, antibodies against Centromeric histone H3 have been raised for the four sequenced species mentioned above, and efforts are underway to visualize centromere organization and validate predicted centromeric repeat sequences.

Key experiments in the Lam Lab have confirmed that natural genetic variation in *S. polyrhiza* leads to more or less turion production under the same induction by phosphate limitation. Further, methods for RNA extraction from high starch (>75%) turions have been optimized and applied to high throughput RNA-sequencing experiments of two genotypes at the extremes of turion yield along with their corresponding vegetative fronds. These experiments have identified turion specific genes associated with dormancy and starch biosynthesis along with others encoding putative transcription factors that may be involved in the developmental transition.

Critically, we have already developed engineered *L. minor* exhibiting a significant increase in oil content, building on the successful engineering of sugarcane to achieve 2-5% leaf TAG in the Shanklin Lab under ARPA-E support. Engineered lines include stable overexpressers of WRINKLED, DGAT and PDAT1, all exhibiting marked increases in TAG content. As expected, efforts to eliminate futile cycling and FA cytotoxicity in our first generation transgenics have resulted in dramatic increases in oil content. In our most recent multigene overexpression lines including OLE1, we now observed TAG accumulation up to 6% of dry weight in *L. minor* with no apparent growth rate defect.

Funding for this project is provided by the DOE Office of Biological & Environmental Research (DE-SC0018244).