

## JGI CSP Letter of Intent 2008

CSP\_LOI\_793583

**PI-**Todd P. Michael, Waksman Institute for Microbiology, Rutgers, The State University of New Jersey

**Co-PIs-**Randall A. Kerstetter, Joachim Messing, Jorg Schwender, John Shanklin, Elias Landolt, Klaus Appenroth, Tokitaka Oyama

### Proposal Title:

Genome sequencing of the duckweed *Spirodela polyrhiza*: a biofuels, bioremediation and carbon cycling crop

### 1. Sequencing Description

The **Lemnaceae**, commonly known as **Duckweeds**, are the **smallest, fastest** growing and **simplest** of flowering plants, representing a high-impact **biofuel feedstock** that is ripe for exploitation. There are forty species representing five Genera in this tiny aquatic monocot family, ***Spirodela*, *Lemna*, *Landoltia*, *Wolffia*** and ***Wolffiella***. The individual plants range in size from 1.5 cm long (***Spirodela polyrhiza***) to less than one millimeter (***Wolffia globosa***). Some of the current uses of Lemnaceae are a testimony to its commercial and scientific utility: basic research **model system**, **toxicity** testing organism, **biotech** protein factories, wastewater **remediation**, high protein **animal feed**, and **carbon cycling** (Stomp 2005). John Cross's website, Charms of Duckweed (<http://www.mobot.org/jwccross/duckweed/duckweed.htm>) is an excellent resource that provides additional information concerning both basic and applied research on duckweeds.

We propose whole genome shotgun sequencing of the Greater Duckweed, ***Spirodela polyrhiza* (L.) Schleiden**. We have chosen *S. polyrhiza* to represent the Lemnaceae because it has the **smallest genome** in the Family. The haploid genome size of *S. polyrhiza* 7003 has been estimated by flow cytometry to be similar to *Arabidopsis thaliana* at **~150 MB** using *Brachypodium distachyon* (300 MB), *Carica papaya* (345 MB) and *Physcomitrella patens* (480 MB) as internal controls. We estimate that **~1,200 MB** of Sanger whole genome shotgun sequencing will be required for 8X coverage. To facilitate assembly and gene prediction, we also propose the sequencing of a modest number of ESTs (~100,000), which is the equivalent of an additional **~100 MB** of sequencing assuming 700-900 bp per EST read. We will supply high quality genomic DNA for the construction of BAC libraries and normalized poly-A+ RNA or cDNA libraries for EST sequencing.

Since JGI has **next generation sequencing** capability, we would be very interested in increasing coverage for both the **genome** and **transcriptome (EST)** to potentially uncover unique genomic features traditionally resistant to Sanger sequencing. Recently the **complex** and **heterozygous** *Vitis vinifera* (cultivated grape) genome was successfully assembled using a mixture of Sanger-based whole genome shotgun sequencing (6.5x) and 454 sequence-by-synthesis (4.2x) (Velasco et al. 2007). In this scenario, we propose 4-6x coverage by traditional sequencing for a total of ~800 MB of high quality Sanger sequence. Then, using a combination of next generation sequencing, such as 454 (250 bp), Illumina (40 bp) and/or SOLiD mate-pair (25 bp ~ 25bp), we propose that additional 20-50x coverage be generated to identify unique genome features. Based on our estimates, four 454 runs would provide 3x coverage (assuming 120 MB, 200-300 bp read lengths per run). Two Illumina runs would provide 20x coverage (1500 MB, 32 bp reads), and one SOLiD mate-pair run would provide at

least 20x coverage (3000 MB, 50 bp mate-pair per run). For the SOLiD run, we would use two sizes of mate-pair libraries (one per slide; 1 kb and 10 kb) to increase power through repeat regions. In combination, these technologies will allow us to resolve any residual heterozygosity, repetitive regions, and other unique genome structure or sequences that are resistant to traditional Sanger sequencing.

Stomp A-M (2005) The duckweeds: A valuable plant for biomanufacturing.

*Biotechnology Annual Review* **11**: 69-99.

Velasco R, Zharkikh A, Troggio M (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS ONE* **2**(12): e1326.

## 2. Sequencing Justification

The doubling time of the fastest growing duckweeds in optimal growth conditions is **less than 30 hours**, nearly twice as fast as other “fast”-growing flowering plants and more than **double** that of conventional crops. This represents an unprecedented opportunity for biomass production. Theoretical yield estimates for pond-cultivated wild duckweeds are 4 metric tons per hectare per day (fresh weight) or 80 metric tons per hectare per year (dry weight) with reported yields ranging from 0.5 to 1.5 metric tons/hectare/day fresh weight or 13 to 38 metric tons/hectare/year dry weight (Journey 1993). This “realized” yield with unselected and unimproved duckweeds rivals the “goal” of 33.6 metric tons/hectare/year that might one day be achieved through genetic engineering a 300% increase in grass biomass production (Hamilton 2006).

Many species are currently developed for industrial uses. For instance, the Environmental Protection Agency uses *Lemna minor* and *Lemna gibba* for ecotoxicological bioassays and for water quality testing, and other species are being developed as well. The plants readily grow on agricultural and municipal **wastewater**, an abundant, infinitely renewable, low-cost substrate that is not typically used for agricultural applications in developed countries. The plants are perennials with **worldwide distribution**; growing anywhere there is fresh water and sunlight. These tiny plants require little mechanical support or **vascular tissue** – the smallest members of the family completely lack differentiated xylem and phloem. As a result of expending little energy on supportive structures, the composition of vegetatively propagating fronds resembles maturing leaves with high levels of protein and carbohydrate and negligible lignin. *Spirodela*, *Lemna*, and *Wolffia* species form specialized over-wintering fronds, called **turions** that accumulate high levels of starch (40 to 70%). Their high starch content increases density causing them to sink to the bottom of the water column where they are more likely to survive freezing conditions. The change in density that accompanies starch accumulation would providing an ideal system for continual harvest. Current efforts focus on harnessing this developmental switch to allow for continual growth and harvesting of high-starch duckweed biomass.

Before the days of Arabidopsis, duckweeds and more specifically *Lemna* was an important **model system** for plant biology. Since *Lemna* is small, morphologically reduced (although with root and leaf-like structure), fast-growing, easily cultivated under aseptic conditions, **transformable**, crossable, and particularly suited to **biochemical studies** (direct contact with media), it is an ideal system for biological research. Much of what we know about photoperiodic flowering responses comes from fundamental research conducted on *Lemna* by the preeminent plant biologist Dr. William Hillman at the Brookhaven National Laboratories. Genome sequence will change the way basic science can progress in the duckweeds.

Journey WK, Skillicorn P, & Spira W. (1993) Duckweed Aquaculture – A New Aquatic Farming System for Developing Countries. *The World Bank*.  
Hamilton R (2006) Biotechnology for biofuels production in *A High Growth Strategy for Ethanol*. Monsma DW & Riggs JA eds. ASPEN INSTITUTE: Washington DC pp 55-60.

### 3. Sequencing Utilization

We have chosen *S. polyrhiza* 7003, originally collected in Louisiana, for sequencing due to the small size of its genome. However, we anticipate that *S. polyrhiza* (the basal most taxon) will form the foundation for future genome projects and tools in the other Lemnaceae family members. We are developing a duckweed genome portal, **LemnaBase (<http://lemna.rutgers.edu>)**, where genome data will be shared with community. Our goal is to make sequences available as soon as practical to stimulate research. Considering the broad use of duckweeds, genome sequence will facilitate research at every level from genomics, systematics, genetics, and biochemistry to applied industrial research. To date there are only a few higher plant genomes that are completely sequenced, *Arabidopsis thaliana*, *Populus trichocarpa*, *Oryza sativa*, *Vitis vinifera* and multiple draft sequences or genomes on the way. The *S. polyrhiza* genome would be the first monocot that is not a grass, thus providing an important link in the evolutionary relationships between higher plants.

We propose to use duckweed as a model for the partitioning of carbon in this simplest of higher plants. Duckweed can accumulate 40-70% by dry weight of starch in its dormant vegetative stages (turions). We propose to identify the mechanisms by which carbon storage occurs and will investigate ways to affect the balance of carbon storage between starch and more energy dense forms such as lipids because oil contains 8 times more energy per volume than starch. The genome will be a critical starting point for this work as it will be used to construct a genome scale metabolic model. Metabolic flux analysis with the use of stable isotope labeling will demonstrate which pathways predominate under defined physiological conditions. This technique is ideal for use with duckweed because of its small size and aquatic habit, whereas such studies are not possible on larger terrestrial flowering plants. The ability to transform duckweed in conjunction with its short generation time will allow us to efficiently test hypotheses regarding carbon partitioning and lay the foundation for rational metabolic engineering. The aqueous habit will also allow us to test the physiological consequences of growth under a variety of conditions and environmental stimuli such as low or high nitrogen or phosphorus under uniform conditions, something that is relatively difficult to achieve with higher plants. The knowledge gained from the proposed studies on carbon partitioning in duckweed will have several applications. First, it will enable us to evaluate duckweed directly as a source for renewable carbon. Second, the principles learned from duckweed will be a useful resource for those working on terrestrial biofuel crops.

### 4. Genomics Community Interest

A PubMed search for duckweed reveals at least a thousand hits, representing only a fraction of the research on duckweeds over the past century; there were 47 hits just in the last year (2007). Most of the research focuses on using duckweeds for toxicity testing, growth rate research, expression of foreign proteins and phytoremediation. The community using duckweeds for research is a broad international community and we predict that the *S. polyrhiza* genome sequence will stimulate even more basic research using duckweeds.

A complete genome sequence for a duckweed species should also prove of critical importance in elucidating molecular mechanisms of both environmental phytotoxicity and phytoremediation. Duckweeds assimilate contaminants from the aquatic phases of the environment in which they reside. Rapid growth provides a simple, sensitive, and easily quantified measure of the impact of toxins and environmental contaminants on plants. As a result, several duckweed species are currently employed for ecotoxicity testing. Duckweeds are frequently found growing in waters contaminated with municipal, agricultural, industrial, and mineralogical wastes. Wild species have been found to sequester or degrade a wide range of environmental contaminants – from heavy metals such as lead, cadmium, and arsenate to halogenated organic compounds (Barber et al., 1995). A complete genome sequence would provide a rational basis for generating improved genotypes that could be employed in aquatic phytoremediation strategies and lay a groundwork for understanding the biological and molecular mechanisms that lead to toxicity or that are employed during sequestration or remediation of toxic molecules.

Barber, J.T., H.A. Sharma, H.E. Ensley, M.A. Polito, and D.A. Thomas. 1995. Detoxification of phenol by the aquatic angiosperm, *Lemna gibba*. *Chemosphere*. 31(6):3567-3574.

## 5. DOE Relevance

Sequencing the *S. polyrhiza* genome represents a perfect match to the three-fold mission of the DOE-JGI and the Department of Energy's Office of Biological and Environmental Research – alternative energy, bioremediation, and global carbon cycling. *S. polyrhiza* and other members of the family Lemnaceae have enormous potential as biofuel feedstocks. These plants produce biomass faster than any other flowering plant, do not compete for land with food or other agricultural commodities, can be cultured with low agronomic inputs, have high energy content in the form of easily fermentable starch (up to 40 to 70% of biomass), have negligible lignin content, and favorable processing characteristics (small particle size, easily harvested and transported). Members of the this plant family are currently employed for waste remediation and distributed by the Environmental Protection Agency for environmental monitoring of water quality and ecotoxicological bioassays. Propagated on agricultural and municipal wastewater, *Spirodela* and related species efficiently extract excess nitrogen and phosphate pollutants, reduce algal growth (by shading), coliform bacterial counts, suspended solids, evaporation, biological oxygen demand, and mosquito larvae while maintaining pH, concentrating heavy metals, sequestering or degrading halogenated organic and phenolic compounds, and encouraging the growth of other aquatic animals such as frogs and fowl. Primitive aquatic plants (*Azolla* species with rapid growth characteristics similar to *Spirodela*) have been implicated as the primary source of carbon sequestration that drove global climate change during the Early Eocene (Moran *et al* 2006 *Nature* 441:601-5). Although the *S. polyrhiza* genome sequence wouldn't be expected to provide direct insights into the historic carbon cycling that gave us our current low CO<sub>2</sub> climate, it would unlock the remarkable potential of a rapidly growing aquatic plant for carbon sequestration, carbon cycling, and biofuel production.