Transcriptome Responses of *Spirodela polyrhiza*

13

Paul Fourounjian

Abstract

In order to analyze the transcriptome of any species, RNA-seq has become the gold standard and evolved into a variety of library preparations and sequencing platforms to study more than mRNA abundance. This chapter reviews the transcriptional studies of *Spirodela polyrhiza*, the best-characterized member of the Lemnaceae family in a genomic sense. To date, there have been three studies of its transcriptome. The first two analyzed ribosomal RNA depleted total RNA of fronds and fronds developing into turions after exposure to abscisic acid. The first study analyzed 154 down-regulated genes involved in growth and 208 upregulated genes involved in starch, anthocyanin production, and seed development. The second study found 66 sites where chloroplast mRNAs were edited to create a functional protein, supporting the hypothesis that mRNA editing was evolved once, and the conservation of editing sites was phylogenetically correlated. The third study, also performed in the 7498 ecotypes, was sequencing of the uncapped polyadenylated transcripts. While the main aim was to observe miRNA induced cleavage, differences in the post-transcriptional regulation or abundance of degraded transcripts across the eight sequencing conditions can be observed. Taken together, these studies cover mRNA expression, post-transcriptional editing, and finally degradation.

Scientists have been interested in gene expression ever since discovering the central dogma of biology and have developed a number of methods over the decades to measure RNA quantity. After reverse transcription was discovered in 1970, Northern blot and Sanger sequencing followed in 1977, qPCR came out in the late 1980s, and then in 2005, the Roche 454 sequencing platform applied shotgun genome sequencing technology to massively parallel RNA-sequencing and quantification (Cieślik and Chinnaiyan 2017). The data from these Roche and Illumina sequencers were typically thousands to millions of 50–200 nt reads that need to be mapped to the genome and reassembled to determine splicing patterns and gene expression as fragments per kilobase per million (FPKM). Joining the next-generation sequencers are the high throughput, long-read sequencers like PacBio and Oxford Nanopore systems available in 2011 which often produce 20–200 kB reads that can easily span entire mRNA and long non-coding RNA transcripts (1–2 kb), thereby eliminating the reassembly steps to more precisely map the transcriptome with its splicing patterns and alternative polyadenylation sites. Unfortunately, the present challenge with these

P. Fourounjian (✉)
Waksman Institute of Microbiology, Rutgers University, Piscataway 08854, USA
e-mail: pjf99@scarletmail.rutgers.edu

X. H. Cao et al. (eds.), *The Duckweed Genomes*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-11045-1_13

reads is their indels and mismatches which can be corrected via deeper PacBio sequencing of the same read, alignment with the reference genome, and alignment with short-read sequencing, all of which can be combined (An et al. 2018a). With these technologies, it is now possible to accurately characterize and measure the transcriptome of virtually any species with a reference genome or through *de novo* assembly.

When looking at the RNA-seq studies of the duckweed family, we see that most of the gene expression analyses have been done in the recently sequenced *Lemna minor*, or in *Landoltia punctata* and *Lemna aequinoctialis* with *de novo* transcriptomes assembled from small reads (An et al. 2018b). These species have genome sizes ranging from 379 to 650 Mb and de novo transcriptomes of 74,797 and 72,105 unique contigs, while *Lemna minor* has 22,382 annotated genes, and *Spirodela* has 19,623 and 18,507 in strains 7498 and 9509, respectively, suggesting a wide variety of possible mRNAs from the roughly 20,000 genes found in Lemnaceae genomes (Tao et al. 2013; Wang et al. 2014a, 2016; Van Hoeck et al. 2015; Michael et al. 2017; Yu et al. 2017). In addition to the transcriptomic analysis of ABA-induced turion formation, there have been a couple of other RNA-seq experiments within *Spirodela polyrhiza* that measure aspects other than gene expression to understand the full complexity of these plant transcriptomes. A reinvestigation of turion development RNA-seq identified the chloroplast genes that undergo mRNA editing and how this relates to the rest of the monocots. Another study was the sequencing of the degraded RNAs in *Spirodela* within eight conditions as a measure of miRNA induced cleavage, which could also show a rough measure of expression. These three studies provide researchers a chance to witness mRNA expression, editing, and degradation.

One unique aspect of the duckweed lifecycle is the formation of turions. Their role as an asexual organ of perennation makes them analogous to both seeds, which sexually give rise to an entire organism, and tubers or buds, which asexually survive winter and other unfavorable conditions. In 2014, an RNA-seq study was performed on *Spirodela* during the development of turions, making it the first genome wide-study of gene expression in a plant tuber, which was followed by a study of potatoes in 2015 (Wang et al. 2014b). It was previously discovered that 3 days of exposure to 10 μM abscisic acid, ABA, induced irreversible turion development and an increase of two enzymes involved in starch and cell wall production. This study, therefore, compared four biological replicates of *Spirodela* fronds with and without 3 days of exposure to ABA by sequencing 15–41 million 75 bp reads of ribosomal, rRNA, depleted total RNA on a SOLiD 5500 instrument. They were able to map reads to the nuclear, mitochondrial, and chloroplast genomes, with 28–39% of reads deriving from the organelles. Reads were aligned with bowtie and tophat, normalized and compared in cufflinks, and annotated for GO term enrichment through blast2go and GOseq. The results showed 154 genes down-regulated during turion development, meaning that they were minimally four-fold less abundant, with roughly half being 0.2–0.1 the expression compared to control. These results had a false discovery rate, FDR, less than 0.01 thanks to the eight biological replicates. The 154 down-regulated genes were largely involved in carbon fixation, protein synthesis, DNA replication, and growth in general since turions no longer grow. For the 208 upregulated genes, the GO term enrichment showed that many of these upregulated genes functioned in starch and anthocyanin production, hormone response and signal transduction, cell wall synthesis, and seed dehydration. There were 13 genes in cell wall and anthocyanin production that were specific to turion induction. Similar to a desiccating seed of a terrestrial plant developing turions upregulated five and expressed two previously silent genes of the late embryogenesis abundant protein family. These LEA family proteins protect other proteins and confer resistance to dehydration, salinity, and cold stress. This transcriptomic study was properly timed to observe not only the structural changes of turion development, but the signaling pathway. They noticed upregulation of seven ABA-responsive, three ethylene-responsive, and two heat shock

responsive transcription factors. There were also ABA transcription factor binding sites in 30 of the upregulated genes, while 119 had a bind site for ethylene-responsive transcription factors. This pathway matches the ABA or environment triggered, calcium-dependent signal pathway observed in maturing seeds, reinforcing the similarity of turions and seeds on a molecular, invisible level.

The same authors performed a second investigation of the rRNA depleted RNA-seq experiment in fronds and developing turions (Wang et al. 2015). Since 26% of the total RNA sequenced mapped to the chloroplast genome, they had 1000-fold coverage of most genes after stringent filtering. The PPR proteins are a massive family, characterized by the 35 amino acid pentatricopeptide repeat motif that specifically binds the 4th and 34th residues in the pfam model to an RNA base, creating a pattern of these motifs that bind to a specific RNA sequence (Barkan et al. 2012; Manna 2015). While these PPR proteins are found in prokaryotes and eukaryotes acting in splicing, processing, editing, stability, and translation of RNAs, this study focused on the DYW-type PPRs that correct certain missense mutations in the plastid genome by editing the mRNA from a cytosine to uracil residue, thereby creating a functional mRNA and protein product. Mapping the RNA-seq reads and detecting C to U SNPs with SAMtools revealed 66 sites of RNA editing with an average efficiency of 76% and a range of 6–100%. Comparison to developing turions showed very similar gene expression with no differentially expressed genes. There were, however, six over and five under edited sites (>two-fold difference, p value <0.05) in seven genes during turion development compared to fronds. So while expression was constant, 1/6th of the sites were differentially edited, thereby altering the functional protein abundance of seven genes. These differences in editing efficiency even varied as much as 8–100% at multiple sites within the same gene due to the sequence-specific nature of PPR protein editing. A phylogenetic analysis with the Mega6 program revealed the 66 editing sites in *Spirodela* had an 81% overlap with the 75 in coconut. There was a 42 and 38% overlap with the 35 and 26 sites observed in the more evolutionarily distant rice and maize. This correlation confirms the hypothesis of a single origin of RNA editing PPR proteins in the early land plants like ferns, that have hundreds of edited sites that were gradually reduced and differentiated over time to ~80 in the basal monocots and 25–40 in the more recent angiosperm species.

Another indirect study of *Spirodela* mRNA expression was the degradome experiment found in Fourounjian et al. (2019). The primary purpose of this experiment was to confirm the cleavage activity of miRNAs on target mRNAs and observe regulatory differences between the biological triplicates of the control, 0 °C, 37 °C, ABA, kinetin, copper, nitrate, and sucrose conditions by sequencing 28–63 million uncapped mRNAs per library. While this degradome sequencing is not a perfect correlate to mRNA expression, the normalized read count of each gene (not kilobase normalized), its expression pattern can be viewed in this program hosted by the Myers laboratory of the Danforth Center https://mpss.danforthcenter.org/~private/dbs/index.php?SITE=messing_SPIRODELA_PARE. These patterns can even reveal unannotated exons, since all reads were polyadenylated. Finally, the miRNA cleavage study revealed that 15 genes, mainly well-conserved transcription factors were expressed and cleaved in four or more conditions, while 71% of the results were condition-specific targets, many of which had more structural and metabolic functions. Of these conditional specific changes, sucrose created the largest difference, followed by copper and heat exposure. This large transcriptomic and metabolic change of sucrose addition suggests that laboratory experiments modeling duckweeds in outdoor applications should avoid this often added media component.

The assembly of the *Spirodela* genome for strains 7498 and 9509 provided not only a scaffold for easy and accurate mapping of RNA-seq data, but a context for the gene expression. This is both in a physical sense for the chromatin

modeling and DNA methylation studies (Cao et al. 2016; Michael et al. 2017), and in a physiological sense where the studies of turion development, for example, can be linked to the results (Kuehdorf et al. 2014). As it stands the research community can observe the Spirodela transcriptome in two or eight conditions as mRNA expression, editing, and degradation. It is expected that the transcriptional research will expand to include more stimuli exposure, tissue specific, life cycle, and microbe interaction experiments. These *Spirodela* genomes and transcriptome studies facilitate research across the family by providing a reference for the other genomes or transcriptomes. Even in cases of de novo assemblies and isoform sequencing of any other related species the *Spirodela* genome will provide annotated and characterized gene models.

References

An D, Cao H, Li C et al (2018a) Isoform sequencing and state-of-art applications for unravelling complexity of plant transcriptomes. Genes (Basel) 9:43. https://doi.org/10.3390/genes9010043

An D, Li C, Zhou Y et al (2018b) Genomes and transcriptomes of duckweeds. Front Chem 6. https://doi.org/10.3389/fchem.2018.00230

Barkan A, Rojas M, Fujii S et al (2012) A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. PLoS Genet 8:e1002910. https://doi.org/10.1371/journal.pgen.1002910

Cao HX, Vu GTH, Wang W et al (2016) The map-based genome sequence of *Spirodela polyrhiza* aligned with its chromosomes, a reference for karyotype evolution. New Phytol 209:354–363. https://doi.org/10.1111/nph.13592

Cieślik M, Chinnaiyan AM (2017) Cancer transcriptome profiling at the juncture of clinical translation. Nat Rev Genet 19:93–109. https://doi.org/10.1038/nrg.2017.96

Fourounjian P, Tang J, Tanyolac B, Feng Y, Gelfand B, Kakrana A, Tu M, Wakim C, Meyers BC, Ma J, Messing J (2019) Post-transcriptional adaptation of the aquatic plant under stress and hormonal stimuli. Plant J

Kuehdorf K, Jetschke G, Ballani L, Appenroth K-J (2014) The clonal dependence of turion formation in the duckweed *Spirodela polyrhiza*—an ecogeographical approach. Physiol Plant 150:46–54. https://doi.org/10.1111/ppl.12065

Manna S (2015) An overview of pentatricopeptide repeat proteins and their applications. Biochimie 113:93–99. https://doi.org/10.1016/J.BIOCHI.2015.04.004

Michael TP, Bryant D, Gutierrez R et al (2017) Comprehensive definition of genome features in *Spirodela polyrhiza* by high-depth physical mapping and short-read DNA sequencing strategies. Plant J 89:617–635. https://doi.org/10.1111/tpj.13400

Tao X, Fang Y, Xiao Y et al (2013) Comparative transcriptome analysis to investigate the high starch accumulation of duckweed (*Landoltia punctata*) under nutrient starvation. Biotechnol Biofuels 6:72. https://doi.org/10.1186/1754-6834-6-72

Van Hoeck A, Horemans N, Monsieurs P et al (2015) The first draft genome of the aquatic model plant Lemna minor opens the route for future stress physiology research and biotechnological applications. Biotechnol Biofuels 8:1–13. https://doi.org/10.1186/s13068-015-0381-1

Wang W, Haberer G, Gundlach H et al (2014a) The *Spirodela polyrhiza* genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. Nat Commun 5. https://doi.org/10.1038/ncomms4311

Wang W, Wu Y, Messing J (2014b) RNA-Seq transcriptome analysis of Spirodela dormancy without reproduction. BMC Genom 15:60. https://doi.org/10.1186/1471-2164-15-60

Wang W, Zhang W, Wu Y et al (2015) RNA editing in chloroplasts of *Spirodela polyrhiza*, an aquatic monocotelydonous species. PLoS ONE 10:e0140285. https://doi.org/10.1371/journal.pone.0140285

Wang W, Li R, Zhu Q et al (2016) Transcriptomic and physiological analysis of common duckweed Lemna minor responses to NH4+ toxicity. BMC Plant Biol 16:92. https://doi.org/10.1186/s12870-016-0774-8

Yu C, Zhao X, Qi G et al (2017) Integrated analysis of transcriptome and metabolites reveals an essential role of metabolic flux in starch accumulation under nitrogen starvation in duckweed. Biotechnol Biofuels 10:167. https://doi.org/10.1186/s13068-017-0851-8