AN ABSTRACT OF THE THESIS OF

Christopher A. Edwards, for the degree of Honors Baccalaureate of Science in Biology presented on June 5th, 2013. Title: Sequencing the Mitochondrial Genome of the Common Milkweed: *Asclepias syriaca*

Abstract approved: _____

Aaron Liston

Few plant mitochondrial genomes have been published to date because of their structural complexity and difficulties in their physical isolation. Plant mitochondrial genomes are known to have a high level of heteroplasmy and can exist as several different circular or linear strands. Sequencing additional mitochondrial genomes will help to understand their highly variable structures. New sequencing technologies provide the opportunity to greatly facilitate the characterization of plant mitochondrial genomes. The mitochondrial genome of *Asclepias syriaca* was sequenced using high throughput Illumina technology. Reads were assembled using de novo assembly and 10 of the resulting contigs were assembled into a whole circular genome with the aid of scaffolding software and Sanger sequencing. A 687kb genome resulted and was annotated with 42 mitochondrial genes and 16 transfer RNAs. Evidence of other conformational forms of this genome were found and the in vivo structure remains unknown. This study was able to construct the whole mitochondrial genome using only Illumina and Sanger sequence data which usually isn't possible. This may have been due to the lower level of heteroplasmy present in this individual compared to others from the same population.

Keywords: Asclepias syriaca, Mitochondria, Genome, Milkweed, Heteroplasmy

Corresponding e-mail address: chrisedwards080210@yahoo.com

Sequencing the Mitochondrial Genome of the Common Milkweed

Asclepias syriaca

by

Christopher A. Edwards

A PROJECT

Submitted to

Oregon State University

University Honors College

In partial fulfillment of

the requirements

for the degree of

Honors Baccalaureate of Science in Biology (Honors Scholar)

Presented June 5th, 2013

Commencement June 2013

Honors Baccalaureate of Science in Biology, project of Christopher A. Edwards presented on June 5th, 2013

APPROVED:

Mentor, representing Botany and Plant Pathology

Committee Member, representing Botany and Plant Pathology

Committee Member, representing Forestry

Chair, Department of Biology

Dean, University Honors College

I understand that my project will become a part of the permanent collection of Oregon State University, University Honors College. My signature below authorizes the release of my project to any reader upon request.

Christopher A. Edwards, Author

ACKNOWLEDGEMENTS

I would like to thank Aaron Liston for his continuing support throughout the development of this project, he believed in me from the start and never waivered in his support. This thesis would not have been possible without his guidance. I would like to thank him for allowing me the chance to work with him and lend me his expertise. This experience has shaped my entire college experience in a profound way.

I would also like to thank Shannon Straub for her work on the project and for taking time out of her busy schedule to assist me and lend her ideas to the project. The same goes for everyone else from the lab including Kevin Weitemier and Rich Cronn. Thanks to all of you. This wouldn't have been possible without your contributions.

Additionally I want to thank my parents. For their support not only through my college years but through everything in my life that led me to this point. I know these words don't mean much but thank you for all you've done for me, I'll spend a lifetime trying to give it back. Love you guys.

Finally I would like to thank my Grandparents.

Grandpa, you have always believed in me as long as I can remember. You are a role model for the life I want to live. I hope to have as many great stories to tell as you do when I'm your age. I love you both and can't wait to see you at graduation.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

Sequencing the Mitochondrial Genome of the Common Milkweed:

Asclepias syriaca (Apocynaceae)

INTRODUCTION

New technology has allowed for a drastic increase in the ability to quickly sequence the genome of an organism (Shendure, 2008). Most often the nuclear genome of these organisms is being sequenced but the mitochondria and chloroplast organelles also contain their own genomes. Sequencing plant mitochondrial genomes presents unique problems because their existence in vivo is not as simple as nuclear genomes. They have a dynamic structure due to a high level of heteroplasmy which is defined as the "coexistence of divergent mitochondrial genotypes in a cell" (Woloszynska p.657, 2009). This heteroplasmy means that the strands of a mitochondrial genome may exist as several linear or circular stands of varying sizes, and is usually a combination of the two. These separate pieces can also replicate independently of each other (Alverson, 2011). Over the 25 years that mitochondrial genomes have been sequenced their variability has made their structure elusive, and it is still largely unknown. Mapping and sequencing studies tend to find the genome as one large circular structure whereas electrophoretic and microscopic studies typically discover linear and multi-branched molecules (Mower, 2010). This heteroplasmy exists to help maintain genomic stability and may also provide a mechanism for increasing variation in the absence of sexual reproduction (Mare´ chal and Brisson, 2010).

Vast differences also exist between plant and animal mitochondrial genomes. Genome size is measured by the number of nucleotide base pairs, usually reported in kilobases (kb). Animal mitochondrial genomes are typically much smaller than plant mitochondrial genomes. In comparison, the human mitochondrial genome has been reported at 16kb (Anderson, 1981) whereas plant mitochondrial genomes typically range from several hundred kb to as much as 11,000kb in *Silene conica* which has one of the largest known mitochondrial genomes (Sloan, 2010, Table 1). No known correlation exists between the size of the nuclear genome and the mitochondrial genome. The range of mitochondrial genome sizes within seed plants is also highly variable. Even though sizes can vary by an order of magnitude within seed plants they contain mostly similar gene content depending on their lineage. Repetitive regions especially small repeats are often responsible for the large disparity in size of mitochondrial genomes between species (Andre´, 1992). Most of the genome is also non-coding (around 90%) (Kubo and and Mikami 2007).

Range of Mitochondrial Genome Sizes

Species	Type	Size
Human	Animal	16kb
Rice Frog	Animal	17kb
Brassica hirta	Plant	208kb
Watermelon	Plant	379kb
Tulip Tree	Plant	554kb
Spruce	Plant	4,300kb
Silene Conica	Plant	11,000kb

Table 1: The variation of sizes between plant and animal mitochondrial genomes

Seed plant mitochondrial genomes have also been known to uptake foreign DNA and incorporate it into their own genomes, whether this be from intracellular or extracellular horizontal gene transfer (Stern and Lonsdale, 1982). Intracellular gene transfers are common where part of the chloroplast genome is transferred to the mitochondria. Seed plant mitochondrial genomes have been shown to have low

rates of mutation (Wolfe, 1987) meaning that base substitution and other forms of mutation are rare in occurrence. This helps to track the origin of certain regions of the genome. If a piece of the chloroplast has made its way to the mitochondria the low rate of mutation will preserve the evidence of this transfer. Another more complicated characteristic is the high incidence of RNA editing and trans-splicing in the transcription of mitochondrial genes. Highly edited RNA can make it difficult to trace a transcriptome (experimentally sequenced RNA) back to the original DNA sequence.

As of April 2012 when this project began, 16 mitochondrial genomes of land plants had been published (Alverson, 2010). The sequence of more mitochondrial genomes will help to better understand their complex nature and in vivo existence. The scientific value of more plant mitochondrial genomes is evident in the following statement: "Clearly more data—preferably whole-genome sequences—are necessary to understand the ebb and flow of these remarkable genomes" (Alverson, 2010). This study answers that call presenting a complete and annotated mitochondrial genome of *Asclepias syriaca*.

Why Asclepias?

The genus *Asclepias* contains ca. 115 species of herbaceous perennial plants in North America, named the "milkweeds" because of the milky latex in their stems and leaves which is toxic in most species (Fishbein, 2011). The milkweeds serve as a model system for the evolutionary ecology of plant reproduction, and have also been helpful in studying plant hybridization (Straub, 2011). Pollination in this genus

is also unique because pollen is grouped together into a complex structure called "pollinia" rather than existing as individual grains. *Asclepias* supports a variety of pollinators including honeybees, wasps, beetles, and hummingbirds, though their relationship as a host for monarch butterfly caterpillars is their most important relationship (Morgan, 1997). This relationship has served as a model for studying co-evolution between plants and herbivores. The chloroplast genome has already been sequenced (Straub, 2011) and the nuclear genome is soon to be published. Adding the sequence of the mitochondrial genome will make this collection complete. These compelling reasons provide support for the importance of choosing this species as the focus of this study.

Figure 1: ***Asclepias syriaca*** - Photo by Mark Fishbein

MATERIALS AND METHODS

One individual [*Phippen 5* (OSC)] from a population of *Asclepias syriaca* in Illinois was selected and sufficient DNA was extracted using the CTAB method (standard protocol) and prepared for an 80bp (base pair) paired end high throughput Illumina (GAIIx) sequenced library with an insert size of 450bp. These short Illumina reads were de novo assembled (from scratch/no reference) into longer contigs using Velvet v. 1.0.12 (Zerbino, 2008) that creates a consensus sequence to assemble short reads into longer contigs (continuous strands of DNA, figure 2).

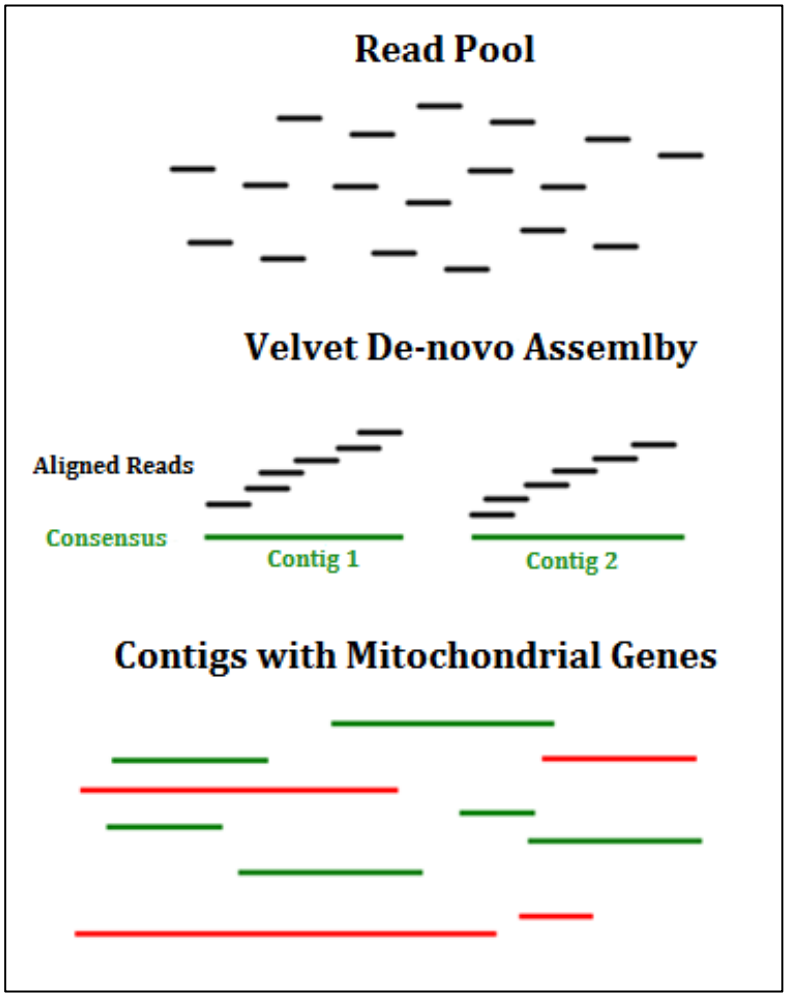

Figure 2: Constructing Contigs From a Read Pool - The process of short read assembly is shown. Small reads are assembled into larger contigs using computational software. In this study only contigs with known mitochondrial genes were selected for further analysis.

The resulting contigs were searched for known mitochondrial genes in tobacco (GenBank:NC_006581) using BLAT (Kent, 2002) and those contigs with matches were used with the program SSPACE v. 1.0 (Boetzer, 2011) for construction into larger scaffolds (series of connected contigs that may have gaps of known lengths) with the aid of a mate pair library constructed by Global Biologics. Primers were designed using Geneious 6.1.4 around resulting gaps in the assembly so that PCR (polymerase chain reactions) could be run with these primers to amplify the sequences with the gaps. PCR was run with standard protocol and products were Sanger sequenced in the CGRB Central Services Lab at Oregon State University. Sanger sequences were inserted into the assembly using Geneious to fill the gaps resulting from the SSPACE model. Additional mitochondrial contigs were then identified by examining the read depth of the velvet contigs pool and added to a new SSPACE assembly. Circos (Krzywinski, 2009) was used next to create a circular representation of the genome. A standard method restriction digest with the enzyme Sph l was also used to confirm the location of a large PCR product that closed the structure into a master circle. Finally mitochondrial genes within the sequence were annotated using the genomes of *Boea hygrometrica* (Zhang 2012) and *Liriodendron tulipifera* (Richardson 2013) as references.

RESULTS

The de novo assembly produced 1,483 contigs of at least 150bp. Of these contigs seven large contigs and one small contig were found to have mitochondrial genes and were selected for use in the construction of scaffolds using SSPACE. Four scaffolds resulted containing a total of four internal gaps (figure 3). Primers were designed to sequence these internal gaps in order to create a continuous sequence. Additional primers were designed to sequence beyond the ends of the scaffolds to explore how these scaffolds might connect to each other, the goal being to connect them all into one continuous master circle.

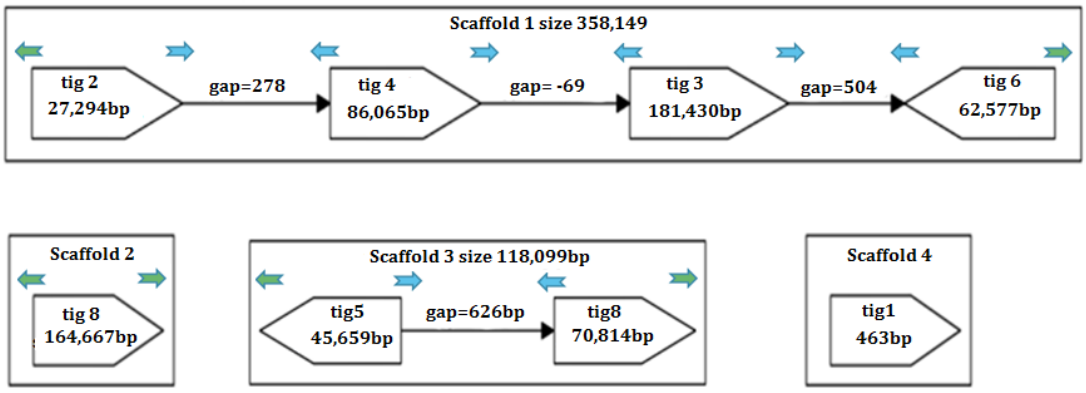

Figure 3: SSPACE Scaffolds - The four resulting scaffolds of the SSPACE assembly ranging from 463 to 358,149bp, with four internal gaps of varying estimated sizes. Blue arrows represent primers designed to sequence internal gaps and green arrows represent primers designed to sequence beyond the scaffolds to explore how they may connect to one another.

All primers designed to sequence internal gaps were successful in producing Sanger sequences that could be inserted into the assembly spanning the gaps. The

small (463bp) scaffold 4 had twice the sequence depth of the other contigs, and was hypothesized to be repeated twice within this genome. Sanger sequencing revealed both repeats. One repeat of this scaffold was found to be located within the gap on scaffold 3, and the other repeat was found in the first gap of scaffold 1 (figure 4). Insertion of the Sanger sequences was straight forward except for the gap between tig 4 and tig 3, this gap was estimated to be negative and as revealed by Sanger sequencing the initial scaffold assembly, had more sequence here than actually existed. A deletion of 70bp from the assembly was required to close the gap.

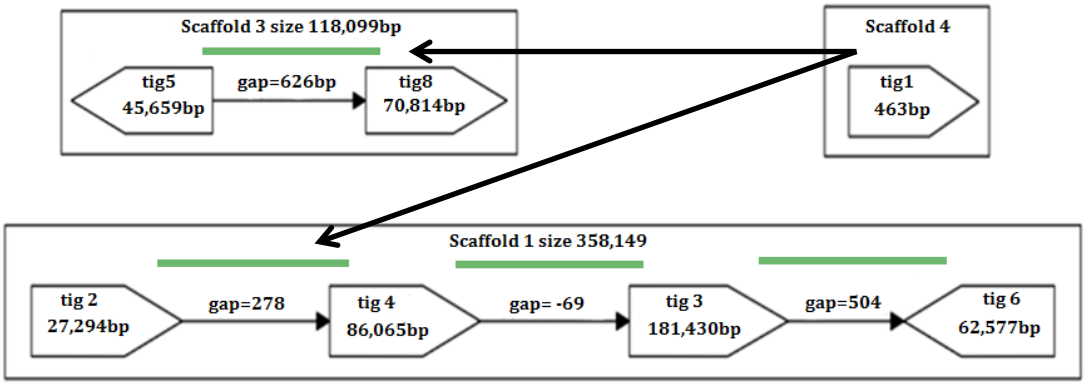

Figure 4: Inserting The First Round of Sanger Sequences - The results of Sanger sequencing from the primers is shown. Green lines indicate the Sanger sequences and where they were inserted into the assembly to close the gaps. Scaffold 4 was found to be within the gaps on scaffold 1 and 3, the black arrows point to where it is located.

Primers sequencing off the ends of the scaffolds revealed that scaffold 1 connects in a circular structure (figure 5 and 6). This structure is not a representative master circle because it contains only one of the scaffolds.

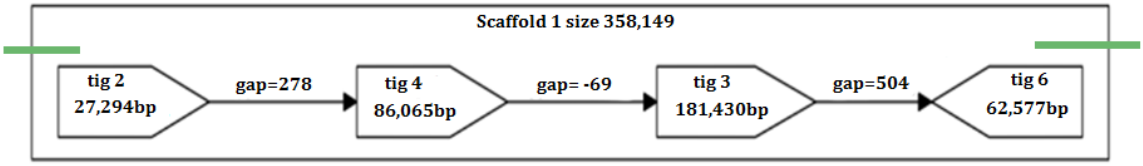

Figure 5: Exploring the Structure of Scaffold 1 - The Sanger sequences off the end of scaffold 1 (green lines) show overlap, connecting scaffold 1 in a circular structure.

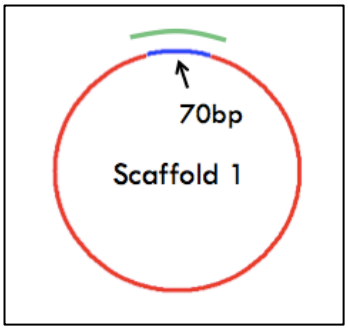

Figure 6: Circular Structure of Scaffold 1 - Shows the circular conformation of scaffold 1, and the 70bp revealed by Sanger sequencing that were required to make this connection. (Green line represents Sanger sequence)

None of the other primers designed to sequence off the ends of the scaffolds were successful in linking any scaffolds together. This was due in part to not having identified all of the mitochondrial contigs in the pool of contigs. Further examination of the contig pool found two large contigs (9 and 10) that matched the same read depth as the other mitochondrial contigs (figure 7).

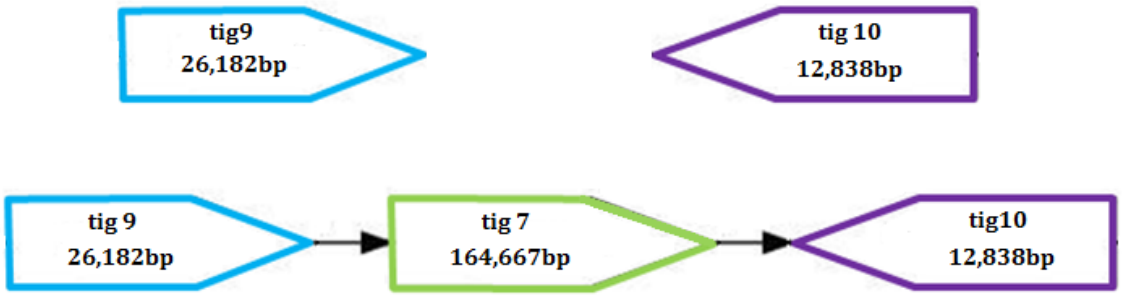

Figure 7: Two New Contigs and Their Connection to Contig 7 - Contigs 9 and 10 were identified with BLAT and used in a second version of the scaffold assembly. The lower assembly shows that the new contigs matched on either end of contig 7, which was not linked to any other contigs in the initial assembly.

BLAT searches of these newly identified contigs revealed exact matches at or near the ends of contigs in our assembly, specifically contig 7, which was initially its own scaffold in the original assembly. Clearly these contigs belonged in the assembly but were not included in the initial assembly because they did not contain any mitochondrial genes. Including these new contigs in the SSPACE assembly resulted in a new organization of the contigs (figure 8).

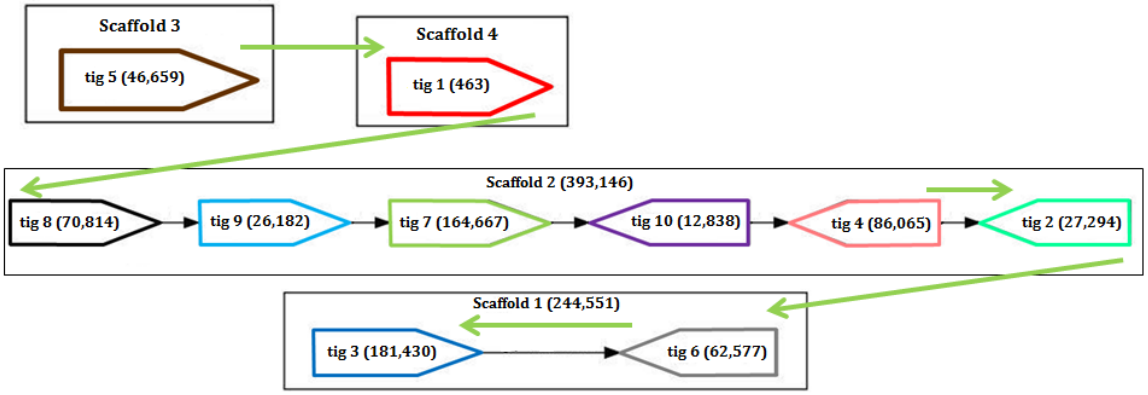

Figure 8: The Orientation of All Contigs - The second scaffold assembly that contains the two additional contigs. Green arrows indicate the PCR evidence for the confirmation of this structure from the original primers. A circular structure can be drawn linking scaffold 3 to 4, 4 to 2, and 2 to 1. Scaffold 3 also has PCR evidence linking it back to the start of the circle at scaffold 3 (not shown).

The new SSPACE model also resulted in four scaffolds but with a larger average size. Adding the PCR and Sanger sequence from the original SSPACE assembly allowed for the confirmation for the locations of many contigs in this new structure. More importantly PCR evidence showed that all of the scaffolds could be connected together into one master circular structure, which is the goal of most mitochondrial sequencing studies. The output of Circos creates a clearer picture of

the genome (figure 9).

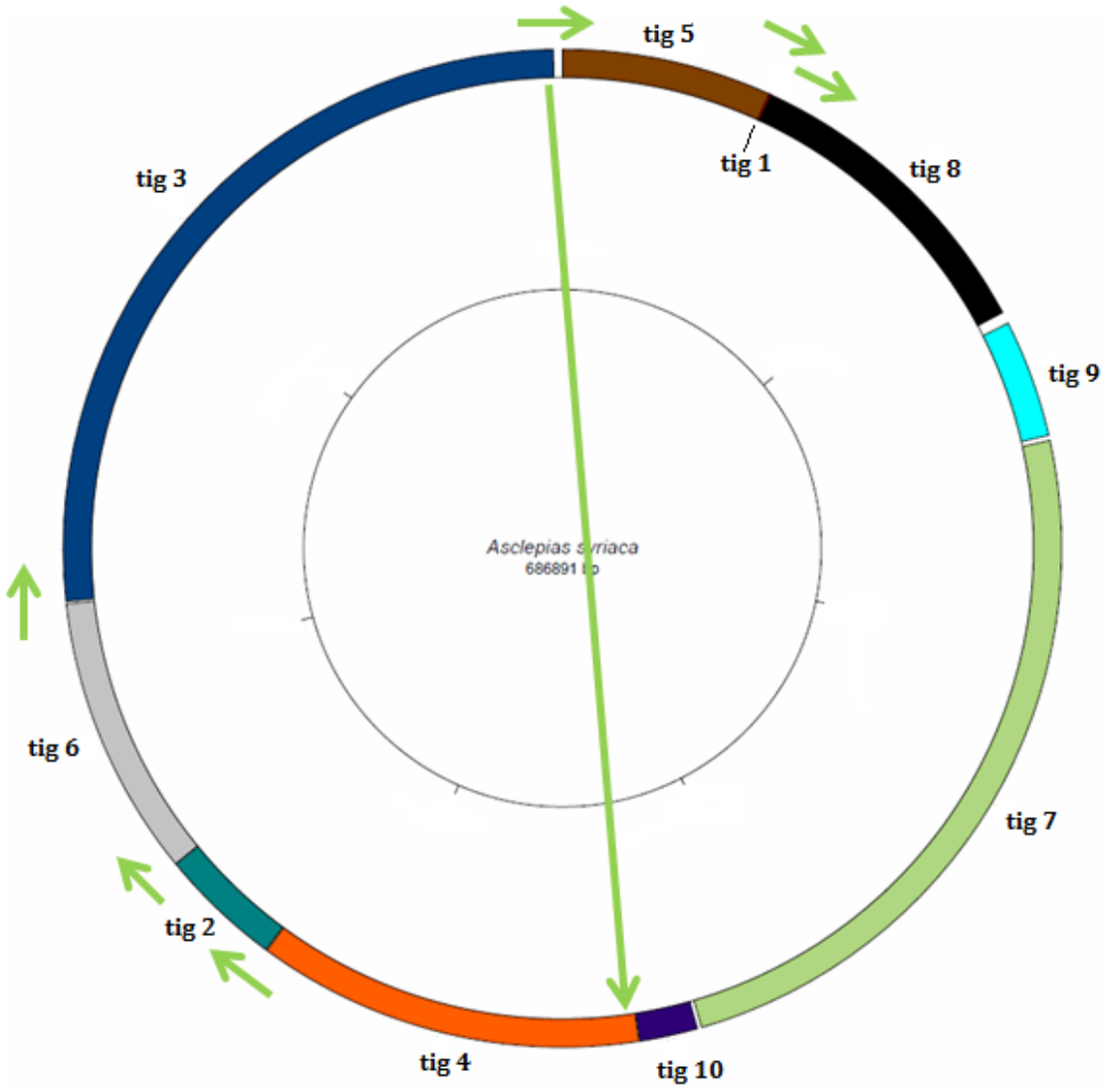

Figure 9: The Circular Structure of The Genome - The second SSPACE scaffolds are represented as a master circle. Green arrows indicate gaps that have PCR sequence across them confirming their place in the assembly. The long arrow across the center shows the PCR that creates a circle out of scaffold 1 from the original assembly.

This model creates four new gaps around the newly inserted contigs. More primers were designed around these gaps to sequence them. Primers in these

locations were successful in obtaining sequence within these gaps. Only the gap between contig 3 and 5 at the top of the model produced a Sanger sequence that was unable to be inserted into the assembly. It appears to be a large repeat that is also present at position 558kb of this genome. A restriction digest on the PCR product provides evidence that this sequence is in fact connecting these two contigs but its nature appears to be too complex to be successfully Sanger sequenced. The final constructed genome was 687kb containing 10 contigs joined by 10 Sanger sequences.

Features of the Genome

Several other features of the genome were revealed during sequencing and assembly. Primer 7R was designed on the top of contig 9 (figure 10). This primer when run in combination with any other primer always produced the same result. When run by itself, it also produced this same product meaning it is functioning as both a forward and a reverse primer. This phenomenon indicates that there is likely to be an inverted repeat within the 148kb region of the genome.

Sequencing the gap between contig 8 and 9 revealed a region very similar to the chloroplast genome of *Asclepias*. Further exploration of this region found a 10kb section that was highly similar to the plastome (figure 10). Shannon Straub (et al) is currently working on a manuscript soon to be published that further investigates this finding.

Many repeated regions were found within this genome, the largest of which is the region at the top of the genome between contigs 3 and 5. This section contains

a large repeat of the genome from the 327kb and 558kb regions. This repeat created such difficulty in trying to Sanger sequence the final PCR product that it was deemed unreasonable to sequence.

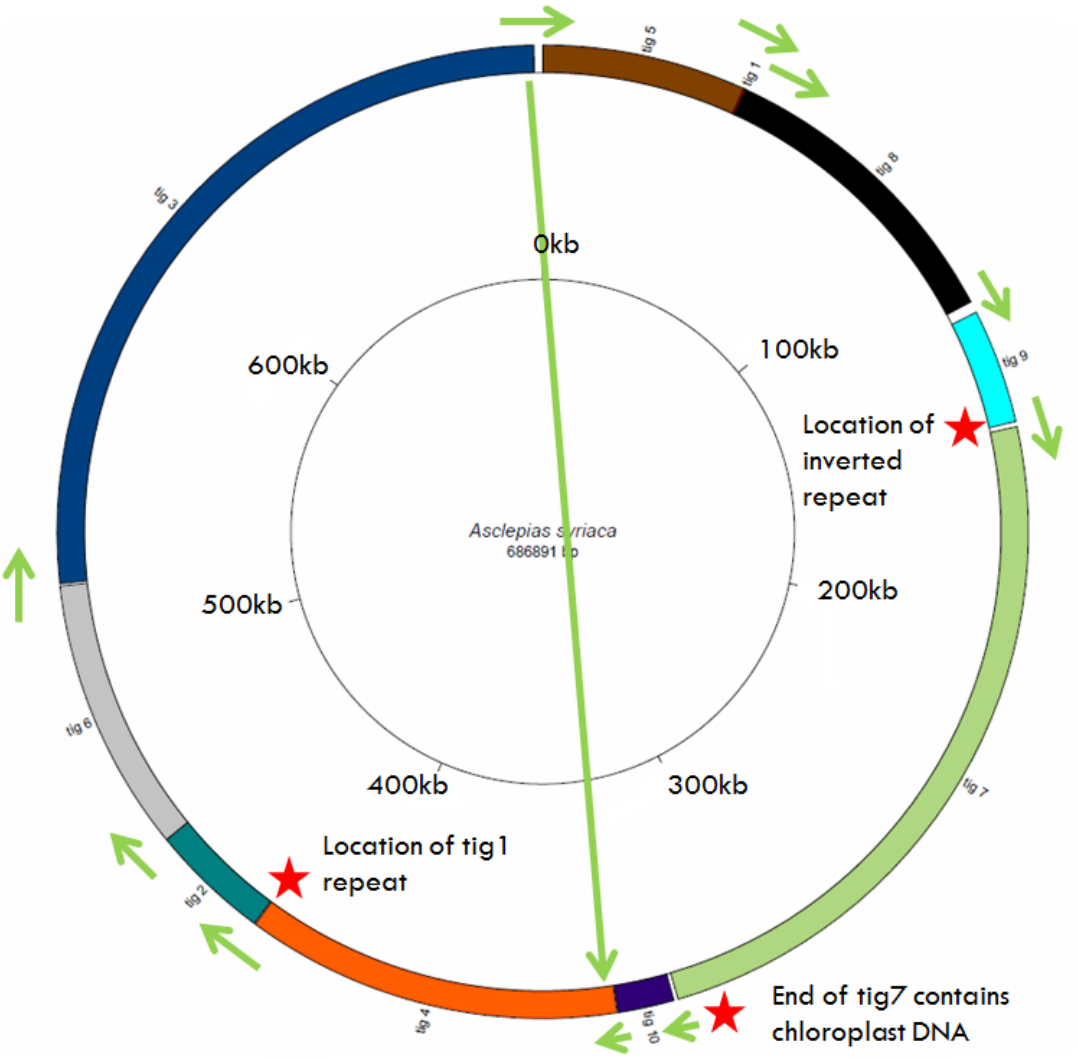

Figure 10: Features of the Genome - The complete circular structure of the mitochondrial genome including all 11 Sanger sequences and the locations of the inverted repeat, the chloroplast region, and the repeat of contig 1.

Gene Annotation

The published mitochondrial genomes of *Boea hygrometrica* (Zhang, 2012) and *Liriodendron tulipifera* (Richardson, 2013) were used as references to find the genes contained within the assembled mitochondrial genome. Forty-two genes were found within the genome some of which may be pseudogenes (figures 11 & 12).

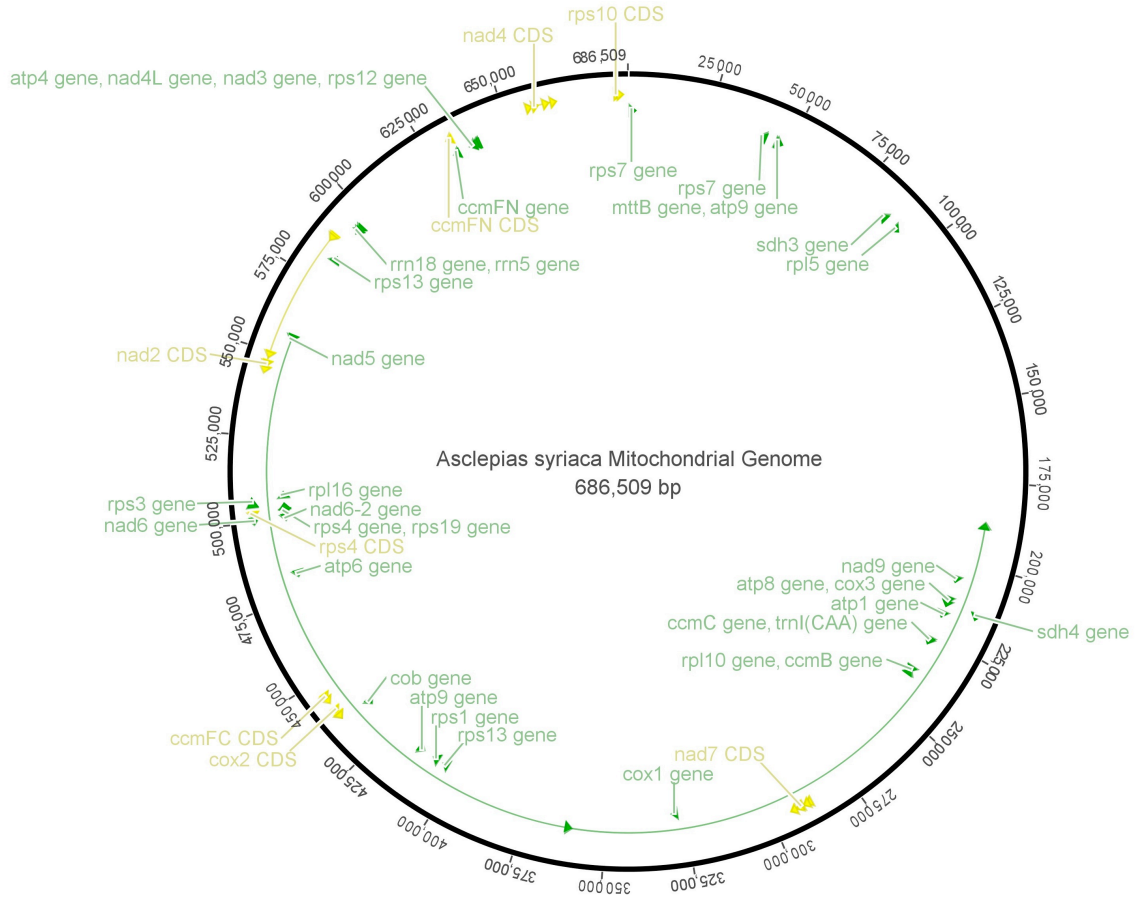

Figure 11: Gene Annotations - The annotated genome of *Asclepias syriaca* showing the location of the genes.

Name	Minimum	Maximum	Length	# Intervals	Direction	Transferred Similarity
atp1 gene	217,457	218,980	1,524	1	forward	91.83%
atp4 gene	637,899	638,483	585	1	reverse	86.57%
atp6 gene	481,849	482,655	807	1	reverse	94.92%
atp8 gene	212,860	213,336	477	1	forward	89.51%
atp9 gene	47,204	47,428	225	1	forward	94.67%
atp9 gene	412,837	413,061	225	1	reverse	94.67%
ccmB gene	238,703	239,323	621	1	reverse	96.62%
ccmC gene	226,798	227,520	723	1	forward	93.64%
ccmFc CDS	443,100	445,344	2,245	2	reverse	77.87%
ccmFn gene	631,866	633,600	1,735	1	reverse	72.09%
cob gene	434,752	435,957	1,206	1	reverse	94.61%
cox1 gene	326,721	329,253	2,533	1	forward	95.13%
cox2 CDS	437,982	440,220	2,239	3	reverse	90.49%
cox3 gene	214,002	214,799	798	1	forward	97.24%
mttB gene	46,381	47,147	767	1	forward	86.98%
nad1 gene	-	-	-	5	-	-
nad2 CDS	544,995	588,608	43,614	5	reverse	-
nad3 gene	639,447	639,803	357	1	reverse	98.04%
nad4 CDS	657,545	665,181	7,637	4	forward	96.84%
nad4L gene	638,636	638,938	303	1	reverse	96.37%
nad5 gene	187,701	557,306	369,606	5	reverse	-
nad6 gene	500,348	500,965	618	1	reverse	97.57%
nad7 CDS	288,624	293,523	4,900	4	forward	98.39%
nad9 gene	205,836	206,408	573	1	forward	97.38%
rpl10 gene	237,946	238,434	489	1	forward	91.21%
rpl16 gene	507,355	507,909	555	1	forward	96.22%
rpl5 gene	91,099	91,641	543	1	forward	73.98%
rps1 gene	407,518	408,111	594	1	forward	85.55%
rps10 CDS	683,368	684,508	1,141	2	forward	90.93%
rps12 gene	639,021	639,398	378	1	reverse	94.44%
rps13 gene	583,235	583,586	352	1	forward	80.97%
rps13 gene	403,965	404,315	351	1	forward	96.30%
rps19 gene	503,924	504,202	279	1	forward	92.28%
rps3 gene	504,380	507,500	3,121	1	forward	71.10%
rps4 gene	501,544	502,614	1,071	1	reverse	81.06%
rps7 gene	986	1,432	447	1	forward	92.44%
rps7 gene	41,171	41,639	469	1	reverse	68.34%
rrn18 gene	593,652	595,469	1,818	2	reverse	88.14%
rrn5 gene	595,606	595,722	117	1	forward	99.15%
sdh3 gene	85,448	85,777	330	1	reverse	70.91%
sdh4 gene	214,727	215,128	402	1	forward	90.12%

Table 2: Gene Names and Locations – The gene names, locations, and lengths in relation to the genome are shown. The transferred similarity indicates how similar the gene in *Asclepias* is to that of *Boea* or *Liriodendron*.

tRNA Annotations and Locations

Sixteen unique tRNAs were found in 23 different locations (figure 12 & table 3).

Figure 12: tRNA Locations – The locations of the 23 tRNAs located within the *Asclepias syriaca* mitochondrial genome.

Name	Minimum	Maximum	Length	Direction
trnC(GCA) tRNA	552,234	552,304	71	reverse
trnD(GTC) tRNA	300,336	300,409	74	reverse
trnE(TTC) tRNA	450,371	450,442	72	forward
trnF(GAA) tRNA	452,838	452,911	74	forward
trnF(GAA) tRNA	485,225	485,297	73	reverse
trnfM(CAT) tRNA	628,117	628,190	74	forward
trnG(GCC) tRNA	375,541	375,613	73	forward
trnH(GTG) tRNA	448,303	448,378	76	reverse
trnI(CAA) gene	227,541	227,614	74	forward
trnI(CAT) tRNA	422,996	423,076	81	forward
trnK(TTT) tRNA	202,832	202,904	73	reverse
trnM(CAT) tRNA	541,790	541,862	73	forward
trnN(GTT) tRNA	550,869	550,940	72	reverse
trnN(GTT) tRNA	433,327	433,392	66	forward
trnP(TGG) tRNA	453,148	453,220	73	forward
trnP(TGG) tRNA	480,538	480,606	69	reverse
trnQ(TTG) tRNA	377,399	377,470	72	forward
trnS(GCT) tRNA	452,451	452,538	88	forward
trnS(TGA) tRNA	363,892	363,978	87	reverse
trnS(TGA) tRNA	299,671	299,757	87	reverse
trnW(CCA) tRNA	480,364	480,437	74	reverse
trnY(GTA) tRNA	550,083	550,165	83	reverse

Table 3: tRNA Locations – The location of the tRNAs are shown along with their length and direction.

Heteroplasmy

In several instances it was discovered that a different individual from the same population as Phippen 5 was expressing a different structure of the mitochondrial genome. In PCR reactions with primers designed to sequence gaps other individuals often expressed multiple products (figure 13).

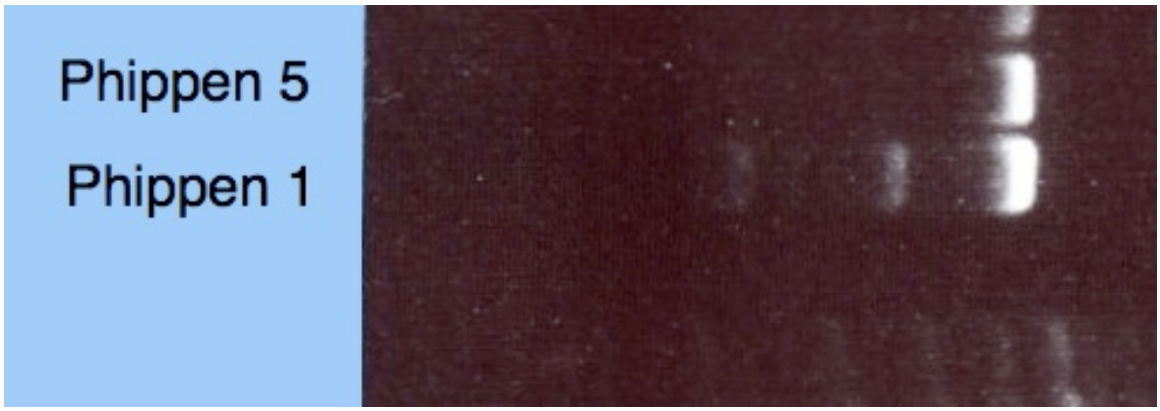

Figure 13: Heteroplasmy – In a PCR reaction using the same primers, Phippen 5 expressed only one strong product whereas Phippen 1 expressed the same product but also displays two other small products. This is likely a result of heteroplasmy.

DISCUSSION

The result of this study was a 687kb mitochondrial genome presented as a master circle. This assembly included ten conitgs connected by ten Sanger sequences. The genome contained forty-two genes and twenty-three tRNAs. The entire project relied only on Illumina and a small amount of Sanger sequencing to construct the genome.

Initial assembly of the *A. syriaca* mitochondrial genome contained 4 gaps. Filling the gaps was required to confirm the arrangement of the contigs within the scaffolds and to create a complete sequence. De novo assemblies of short read sequences tend to break at locations that are highly repeated, making it difficult to design primers and sequence those regions.

Including the two additional contigs in the assembly allowed for all the pieces of the mitochondrial to be connected in one master circle. Doing so ensures that no pieces of the assembly are missing. It is unlikely that the actual structure of the genome is circular in vivo, but it is the most concise form to report. PCR evidence connecting the original scaffold 1 as its own circle shows the presence of different forms of the genome, this lends evidence to the fact that the actual genome probably exists as several circular or linear structures. The inability to connect the two ends of the circle together with Sanger sequence may also suggest that the genome exists as a linear structure in the cell.

Phippen 5, the individual used in this study was a fortuitous candidate for sequencing because it showed less heteroplasmy than 4 other individuals from the

same population. Primers designed to sequence gaps on Phippen 5 often produced multiple products when tested with other individuals from the same population. A lower level of heteroplasmy in Phippen 5 likely helped to reduce complications that could have been encoutnered from sequencing different conformations of the same genome.

The large region of chloroplast DNA on contigs 8 and 9 is likely a region that was transferred to the mitochondrial genome from the plastome. Low mutation rates help to keep this region similar enough that it can still be identified as chloroplast DNA. The transfer also appears to be fairly recent on an evolutionary time scale.

Gene content is similar to closely related species like *Boea hygrometrica*, but distantly related species like *Liriodendron tulipifera* can also have similar gene content. Both of these genomes were equally similar to *Asclepias* even though they are very distant evolutionarily. Depending on the lineage mitochondrial genomes can be conserved across a wide range of taxa. The genes families of ATP, NAD, RPL, and RPS represent the majority of the genes present. Finding many genes from the ATP family is not surprising since the main function of the mitochondria is to synthesize ATP. The NAD genes are related to this function as they code for parts of the electron transport chain like NADH dehydrogenase. The RPL genes along with the RPS genes help to construct the many ribosomes present in mitochondria.

Closing Remarks

Other mitochondrial studies often rely on several forms of data collection to create a complete mitochondria genome. This could include different types of hybridization, cloning, or plastid studies in addition to the initial shotgun sequence. This project was able to construct the genome using only Illumina and Sanger sequence data without requiring other types of analyses. The arrangement of the contigs also has a high level of reputable evidence for their conformation in the final structure. While the in vivo structure may not be known, hopefully this study will provide useful data to others working to reveal the mysteries of the structure of mitochondrial genomes.

Further steps of this project will include annotating the repeats within the genome to better understand them. The mitochondrial sequence is also soon to be published and will be available from Gen Bank in the near future.

REFERENCES

Alverson, Andrew J., et al. "Insights into the evolution of mitochondrial genome size from complete sequences of Citrullus lanatus and Cucurbita pepo (Cucurbitaceae)." *Molecular biology and evolution* 27.6 (2010): 1436-1448.

Alverson, Andrew J., et al. "Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber." *The Plant Cell Online* 23.7 (2011): 2499-2513.

Anderson, Sharon, et al. "Sequence and organization of the human mitochondrial genome." *Nature* 290 (1981): 457-465.

Andre, C., A. Levy, and V. Walbot. "Small repeated sequences and the structure of plant mitochondrial genomes." *Trends in genetics* 8.4 (1992): 128-132.

Boetzer, Marten, et al. "Scaffolding pre-assembled contigs using SSPACE." *Bioinformatics* 27.4 (2011): 578-579.

Fishbein, Mark, et al. "Phylogenetic relationships of Asclepias (Apocynaceae) inferred from non-coding chloroplast DNA sequences." *Systematic Botany* 36.4 (2011): 1008-1023.

Geneious version R6 created by Biomatters

Krzywinski, Martin, et al. "Circos: an information aesthetic for comparative genomics." *Genome research* 19.9 (2009): 1639-1645.

Kubo, Tomohiko, and Tetsuo Mikami. "Organization and variation of angiosperm mitochondrial genome." *Physiologia Plantarum* 129.1 (2007): 6-13.

Maréchal, Alexandre, and Normand Brisson. "Recombination and the maintenance of plant organelle genome stability." *New Phytologist* 186.2 (2010): 299-317.

Morgan, M. T., and D. J. Schoen. "Selection on reproductive characters: floral morphology in Asclepias syriaca." *Heredity* 79.4 (1997): 433-441.

Mower, Jeffrey P., et al. "Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a monkeyflower (Mimulus guttatus) lineage with cryptic CMS." *Genome biology and evolution* 4.5 (2012): 670-686.

Nystedt, Björn, et al. "The Norway spruce genome sequence and conifer genome evolution." *Nature* (2013). 1-6.

Palmer, Jeffrey D., and Laura A. Herbo. "Unicircular structure of the Brassica hirta mitochondrial genome." Current genetics 11.6-7 (1987): 565-570.

Richardson, Aaron O., et al. "The “fossilized” mitochondrial genome of Liriodendron tulipifera: Ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate." *BMC Biology* 11 (2013): 11-29.

Rodríguez-Moreno, Luis, et al. "Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest reported mitochondrial genome in plants contains a significant amount of DNA having a nuclear origin." *BMC genomics* 12.1 (2011): 424-438.

Shendure, Jay, and Hanlee Ji. "Next-generation DNA sequencing." *Nature biotechnology* 26.10 (2008): 1135-1145.

Sloan, Daniel B., et al. "Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm Silene latifolia." BMC evolutionary biology 10.1 (2010): 274.

Stern, David B., and David M. Lonsdale. "Mitochondrial and chloroplast genomes of maize have a 12-kilobase DNA sequence in common." *Nature* 299.5885 (1982): 698.

Straub, Shannon, et al. "Building a model: developing genomic resources for common milkweed (Asclepias syriaca) with low coverage genome sequencing." *BMC genomics* 12.1 (2011): 211.

Wolfe, Kenneth H., Wen-Hsiung Li, and Paul M. Sharp. "Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs." *Proceedings of the National Academy of Sciences* 84.24 (1987): 9054-9058.

Woloszynska, Magdalena. "Heteroplasmy and stoichiometric complexity of plant mitochondrial genomes—though this be madness, yet there's method in't." *Journal of experimental botany* 61.3 (2010): 657-671.

Zerbino, Daniel R., and Ewan Birney. "Velvet: algorithms for de novo short read assembly using de Bruijn graphs." Genome research 18.5 (2008): 821-829.

Zhang, Tongwu, et al. "The complete chloroplast and mitochondrial genome sequences of Boea hygrometrica: insights into the evolution of plant organellar genomes." *PloS one* 7.1 (2012): e30531.