Using phylogenetics to understand the evolutionary relationships of polyploids in *Hibiscus* section *Furcaria*

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Whitaker Matthew Hoskins
May 2016

Acknowledgments

I would like to thank several people that made this work possible in terms of academic and personal support. First, I want to thank Randy Small for being a flexible, supportive advisor throughout my undergraduate and graduate career. I would also like to thank my committee members Brian O'Meara and Ed Schilling for giving me great ideas on how to expand my project and encouraging my learning process. Additional thanks need to go to several professors that taught me a myriad of skills and tidbits along the way: Karen Hughes, Ken McFarland, Brandon Matheny, Meg Staton, Joe Williams, Phillip Li, and many others. Also, thank you to Justin Hendy, Cassie Dresser, John Reese, Jayne Lampley, and all the graduate students in the EEB department for being great friends and colleagues. Additionally, I would like to thank the Department of Ecology and Evolutionary Biology for providing funding to complete this work. Lastly, thank you to my lovely wife, PJ Hoskins for supporting the decision to come back to school and live in Knoxville once again.

Abstract

Neopolyploids constitute at least 35% of known species of angiosperms, and because polyploidization is a pertinent process in plant diversification and domestication, it is a thriving field of study. *Hibiscus* section *Furcaria* includes several groups of polyploids in addition to ten known diploid species. In previous studies genome groups for *Hibiscus* section *Furcaria* were determined through artificial hybridization experiments and patterns of biogeography were elucidated based on the distribution of diploids and polyploids. For instance, the Australian hexaploids contain 3 genomes (designated G, J, and V) and are thought to have developed from a polyploidization event between an African diploid relative (G) and two unknown donors (J and V). This study seeks to perform phylogenetic analysis using a suite of chloroplast and nuclear regions to determine the maternal genetic relationships between the diploids and the Australian hexaploid lineage, and to reconstruct the origin of this group in order to determine if any surviving diploid donors exist related to the unknown J and V genomes. Four chloroplast regions and two nuclear regions were explored to determine genome relationships. Results showed the Australian hexaploid species form a well-supported clade using chloroplast genes (*ndhC-trnV,ndhF-rpl32R,rpl32F-trnL,rps16-trnK*) and ITS with *Hibiscus sudanensis* as the maternal donor of the G genome group. Possible donors to the J and V genome groups are proposed based on the phylogenies, morphology, and biogeography but more sampling of clones is needed to ensure the identity of possible donors lineages.

keywords: polyploidy, *Hibiscus* section *Furcaria,* phylogenetics, chloroplast DNA, nuclear DNA, genome donor

Table of Contents

Chapter 1: Introduction..... 1

Chapter 2: Methods.....8

Chapter 3: Results.....13

Chapter 4: Discussion.....18

References.....22

Appendices.....27

Appendix A: Tables.....28

Appendix B: Figures.....32

Vita.....45

List of Tables

1. Plant materials used in this study. Voucher specimens at TENN.....29
2. Best fit models for cpDNA and nDNA.....30
3. Percent of variable and parsimony-informative sites from all sequence analyses.....31

List of Figures

1. Species occurrence maps for each study species in *Hibiscus* section *Furcaria*.....33

2. Total species occurrence map for all study species in *Hibiscus* section *Furcaria*.....34

3. Bayesian analysis of 4 cpDNA regions with genome groups, geography, and ploidy level....35

4. Bayesian analysis of GBSSI region.....36

5. Bayesian analysis of ITS region.....37

6. Maximum likelihood analysis of 4 cpDNA regions.....38

7. Maximum likelihood analysis of GBSSI region.....39

8. Maximum likelihood analysis of ITS region.....40

9. Parsimony bootstrap analysis of 4 cpDNA regions.....41

10. Parsimony bootstrap analysis of GBSSI region.....42

11. Parsimony bootstrap analysis of ITS region.....43

12. Probability plot for sampling all genome groups in the Australian hexploids in *Hibiscus* sect. *Furcaria* as a function of the number of clones sampled.....44

Chapter 1: Introduction

What is polyploidy?

Historically, polyploids were defined as organisms containing multiple chromosomes compared to an ancestral state, and ploidy level was determined by chromosome number (Lutz 1907, Husband 2013). Modern studies emphasize whole genome duplication as the inherent process leading to polyploidy (Husband 2013). Whole genome duplications can be determined through the comparison of the number of genes and genome size in terms of megabases (Mb) in related genomes. Yet, this criterion expands the definition to include chromosome numbers that do not always show strict doubling and may be the result of a restructured genome and/or loss of paralogs. Research on polyploidy is increasing as molecular technology reaches a point where whole genomes can be sequenced more efficiently and less expensively. Polyploids can be classified as neopolyploids from a recent doubling event, or paleopolyploids from an ancient duplication that subsequently underwent diploidization, but there is no specific temporal distinction (Wolfe 2001, Adams & Wendel 2005). Diploidization is an evolutionary process that converts quadrivalent chromosome pairing in tetraploids to bivalent chromosome pairing and these species function genetically like diploids while maintaining the same overall chromosome number (Wolfe 2001). Jiao et al. (2011) highlight the importance of paleopolyploidy in the development of diverse plant groups by using expressed-sequence-tag (EST) sequences to estimate ancient duplication events. The authors find one polyploidization event prior to the diversification of seed plants and one prior to the diversification of angiosperms, thus suggesting a correlation exists between polyploidy and cladogenesis.

Hybridization is another fundamental distinction in the origin of polyploids. Autopolyploids represent whole genome duplications occurring within a species and perhaps just

one individual, while allopolyploids result from hybridization between different species. However, it may be difficult to categorize a species as autopolyploid or allopolyploid due to recurring introgression or paralog loss over time (Mason-Gamer 2013). Therefore the distinction between autopolyploids and allopolyploids is not always clear within certain taxa (Stebbins 1970, Ramsey & Schemske 1998, Tate et al. 2005).

Two main mechanisms can form polyploids- somatic doubling and gametic non-reduction. Somatic doubling is a result of the failure of mitosis and cytokinesis in a zygote or meristematic shoot. Gametic non-reduction is a result of a failure of one of the meiotic divisions to halve the chromosome number. Gametic non-reduction is expected to form more viable hybrids by using a 'triploid bridge' to allow polyploids to form offspring and populate an area. The triploid bridge involves the backcrossing of triploids with diploids or a self-fertilizing triploid to form fertile tetraploid individuals (Ramsey & Schemske 1998). An interesting component to the development of polyploidy in plants is the conspicuous absence of a centrosome for the microtubule-organizing center (Marshall 2009). Andreuzza & Siddiqi (2008) provide a perspective piece on d-Erfurth et. al (2008) that explains how a mutant *AtPS1* gene specific to plants can produce a parallel spindle in *Arabidopsis* which leads to gametic non-reduction. The creation of non-reduced gametes leads to polyploidization in plants and may explain why plants develop polyploid individuals at a higher rate than other kingdoms of life. Cellular mechanisms like these, along with phylogenetic history and ecology, are key to the understanding and utilization of polyploid plants in biotechnology. Resurgence occurring in polyploid research is built around trying to understand ecological implications, genetic complexity, and possible innovations in agricultural production (Glennon et al. 2014, Soltis et al. 2010, Sattler et al. 2016).

Importance of Polyploidy in Plants

As human population continues to increase crop production must keep pace in order to supply adequate food for future generations (Hooke et al. 2012). This means the knowledge of how to maximize sustainable production of plants is paramount and this is applied to the increasingly inter-related fields of biotechnology and agriculture. Many of our existing crop plants are polyploid (Renny-Byfield & Wendel 2014, Blanc & Wolfe 2004) which has implications for understanding how crops and plants in general respond to selection.

On an organismal level polyploids may have certain advantages over diploids due to their multiple gene copies. One advantage is the ability of polyploids to mask recessive alleles that are harmful (Comai 2005). Also, polyploid plants can self-fertilize which allows polyploids to overcome the reproductive barrier of minority cytotype exclusion. Minority cytotype exclusion is the process by which most gametes of polyploids (2x) are lost due to their incompatibility with the more abundant (1x) gametes in their environment (Levin 1975). Apomixis, or the formation of embryos asexually, is another way polyploids may overcome reproductive barriers and continue to proliferate. It has even been suggested these apomictic species are competitive but reach a dead-end in terms of their evolutionary history (Stebbins 1970). Lastly, polyploidization can lead to new functions or sub-functionalization of homeologs for adaptation (Adams & Wendel 2005). Disadvantages also exist but mostly apply to cellular machinery and epigenetic consequences (Comai 2005). These advantages stem from the complex molecular consequences of polyploidization. Unfortunately, these complexities can also make it difficult to study polyploid plants both *in situ* and *ex situ.*

Issues with the study of polyploids

Polyploids are organisms with homeologs, a term for paralogous chromosomes developing from a genome duplication event (Wolfe 2001). The effect of polyploidization on the phylogenetics of plant populations is increasingly studied as genetic technology enables the assessment of these inherently difficult study systems. An example of these challenges includes possible introgression among a polyploid and diploid that may generate polyphyletic clades and incongruent gene trees (Mason-Gamer 2013). Additionally, the extinction of a diploid progenitor of polyploid species may lead to no matching in a particular homeolog at all (Mason-Gamer 2013).

In addition to the genetic conundrums that polyploids present, there is a challenge to ecologists because few universal ecological differences have been demonstrated between diploids and polyploids (Soltis et al. 2010). Even when the duplication event is intraspecific (autopolyploid) the ecological consequences are not necessarily predictable. In other words, even species developing from autopolyploidization show little predictable ecological effects between ploidy levels. Potential issues to explain the ecology of polyploids versus their diploid counterparts include: climatic tolerance, secondary chemistry, invasiveness, flowering phenology, and mating system (Hahn et al. 2012, Hull-Sanders et al. 2009, Pandit et al. 2014, Robertson et al. 2011, Segraves & Thompson 1999).

Hibiscus Taxonomy

The traditional recognition of *Hibiscus* L. (Malvaceae) includes 250-350+ species (Pfeil 2005) however this group is not monophyletic and contains species within several tribes of Hibisceae (Pfeil & Crisp 2005). *Hibiscus* section *Furcaria* is a group of ca. 200 species that

occurs around the globe and is prevalent in tropical and sub-tropical zones, especially in the Southern Hemisphere. The section name *Furcaria* comes from the bifurcating bracteoles that emerge below the calyx, sometimes referred to as an epicalyx. The distinguishing characters of this section are the thickened midribs and marginal ribs of the calyx and presence or absence of nectaries along these ribs. This group also includes two important crop species, *H. sabdariffa* or roselle, used in making tea and jellies, and *H. cannabinus*, or kenaf, used in fiber and paper production.

Several mating studies by Margaret Menzel and colleagues determined species within *Furcaria* belong to distinct "genome groups" (Menzel & Martin 1971, Menzel & Wilson 1963, Menzel & Martin 1980, Wilson 1993) and these genome groups are designated by letters (A, B, C, D, E, G, H, J, P, R, V, X, Y). *Hibiscus* section *Furcaria* contains only 10 known diploids. Several tetraploid lineages also exist (AB, AX, AY, BY, GH, GP) as well as species with higher levels of polyploidy (BG _ _, CDEG, CDEGR, GJV). These studies used the compatibility of chromosome pairing in pollen mother cells (PMCs) of hybrid individuals to determine the gene affinity index (GAI) for each pair of species crossed. GAI ranges from 0 to 1 and is determined based on the number of chromosomes that aligned during meiosis, higher scores means more compatibility. In these studies evidence was found for a group of hexaploid Australian species that were mostly incompatible with all other species in section *Furcaria* except those of the G genome group.

The Australian hexaploid group, designated by the letters GJV, has been the focus of recent taxonomic update. Craven (2003) revised the taxonomy of the Australian species and produced an identification key and distribution maps for 23 species within Northern and Western Australia but in total 29 species of *Furcaria* are present on the continent. These species are

shrubs and small trees found within the coastlands of tropical and subtropical regions of northern Australia (Kimberley, Northern Territory, and Queensland). Fig. 1 and 2 show the known ranges of the 4 Australian hexaploids used in this study. Some species are even found on other islands such as Indonesia/Papua New Guinea.

Objectives

The following research seeks to add to the body of knowledge on the Australian hexaploid species of section *Furcaria* in terms of their phylogenetic relatedness to other species within *Furcaria* and add to the growing literature on polyploid phylogenetics. The research objectives start in a broad sense on polyploids and narrow in on the study group- *Hibiscus* section *Furcaria.*

1. Determine the phylogenetic relationships of diploid species to polyploid species using chloroplast and nuclear DNA.
2. Infer the possible diploid donors for the J and V genomes found in the Australian hexaploids.
3. Discuss the known ranges of the possible diploid donors and the biogeography of hybridization events that allowed for the allohexaploids to develop in Australia.

Predictions

Since chloroplasts are typically inherited maternally (Greiner et al. 2015) the cpDNA phylogeny will only show maternal inheritance relationships. The G genome species are known to be compatible with the Australian hexaploids, represented as GJV, (Menzel & Martin 1980) so

we expect the cpDNA phylogeny to show close relationships with any species that contains a G genome. Species used for this study in the G genome group are *Hibiscus sudanensis,* a diploid representing the G genome*, Hibiscus rostellatus,* a tetraploid representing the GH genomes, and *Hibsicus furcellatus,* a tetraploid representing the GP genomes. Nuclear data was also included in this study to address the final objective of the origin of the J and V genome groups. Three possible outcomes for the origin of the J and V genome groups could occur. (1) Cloned copies of the J and V group may be located within a clade of extant species in section *Furcaria*. For example, a genome copy, perhaps J, aligns with the A group while the other copy V aligns with Y. (2) Cloned copies of the J and V groups may be located in the clade with other G species as well because they align closely with another group, either the H or P, that is contained within a tetraploid. This could mean the J, V, or both are modified H and P genomes, which in turn could be modified from the G group. (3) Cloned copies of the J and V groups will not follow a phylogenetic pattern because there are no extant progenitors of these groups, these progenitors have not been found and sequenced, or the clones for the Australian hexaploid species were not adequately sampled.

Chapter 2: Methods

Samples

Seeds of 28 species were obtained from the USDA National Seed Storage Laboratory in Dry Branch, GA or the Southern Regional Plant Introduction Station at the University of Georgia Griffin Campus. Table 1 shows all samples used in this study, several are from previous work done by Randy Small on the diploid and tetraploid species of this group. These species were selected to fully sample the potential diploid genome groups and sample some tetraploid species of interest. Eleven species of the Australian hexaploids in section *Furcaria* were selected to represent the Australian hexaploid group.

Table 2 shows the eleven Australian species, 2 of which are varieties, attempted for growth and sampling in this study. These seeds were soaked in a mix of 10% bleach treatment for an hour then rinsed to prevent fungal contamination. Seeds were then nicked and soaked in water for several days. Eight of the twelve species germinated and were planted (see Table 2) in soil saturated with water and ZeroTol (algaecide/fungicide) in order to prevent colonization of foreign contaminants. Seedlings were watered every five days with low concentration liquid fertilizer (Dyna-Gro), and ZeroTol and grown in controlled light conditions. Three plant species within *Hibiscus* section *Furcaria* were successfully grown from seed (*H. marenitensis, H. meraukensis, H. splendens*) and one previously extracted (*H. arnhemensis*). For each successfully grown specimen, a voucher specimen was placed in the University of Tennessee Herbarium. Collection numbers and voucher information are available in Table 2. The intent of this study is not to fully sample the Australian group but use representatives to determine the phylogenetic relationships of the Australian group to other genome groups.

Molecular methods

Total genomic DNA was isolated using standard protocols with the DNeasy Plant Mini Kit (Qiagen) by extracting from young leaves of lab grown plants. PCR amplifications followed published reaction conditions (Small 2004). Seven cpDNA regions from Shaw et al. (2005) were screened. Four regions were used (*ndhC-trnV,ndhF-rpl32R,rpl32F-trnL,rps16-trnK*) that showed the greatest variability and thus potential phylogenetic utility. The nuclear ribosomal internal transcribed spacer (ITS) region was amplified using ITS4 and ITS5 primers (White et. al 1990). Reaction conditions were adapted from White et al. (1990). GBSSI (granule-bound starch synthase I), a protein-coding, single copy nuclear gene was amplified using primers 2B-9R* (5'-*CCAATGAACCCAATCAAGGGAG*-3') and 2-X2F (5'-*TGACNGTGTCTCCTCGCTATGA*-3'). The GBSSI reaction conditions were 95°C for 3 minutes, 40 cycles of 95°C for 30 seconds, 66°C for 30 seconds, 72°C for 2 minutes. DNA sequencing was performed on column purified (QIAquick PCR Purification Kit) or ExoSAP-IT (USB) cleaned PCR products using the ABI Prism Big Dye Terminator Cycle Sequencing kits v. 3.1 (Applied Biosystems), and read on an ABI 3100 automated sequencer at The University of Tennessee Molecular Biology Resource Facility, Knoxville.

Cloning

Since multiple copies of each nuclear region were expected in the Australian hexaploid species (Menzel & Martin 1980) based on hybridization experiments, and multiple bands were shown through gel electrophoresis, the PCR products were cloned prior to sequencing. Additionally, tetraploids and some heterozygous diploid species were cloned in previous work for the ITS region. All cloning used the vector pGEM-T (Promega) system as described in Small

et al. (1998). Cloning plates using *E. coli* were grown overnight in a Luria broth and ampicillin mixture and plasmid DNA was purified as described in Small et al. (1998). 10 colonies were picked per plate. The insert DNA was PCR amplified and sequenced to obtain clean nuclear sequence copies. An analysis of the probability of detecting all gene copies in sampling but this was done post-cloning (see Fig. 12).

Alignment and phylogenetic analyses

Pre-alignment sequence editing (base correction using chromatograms, forward and reverse sequence overlap) was performed in Sequencher v. 5.1. Sequence alignment was performed using Muscle v. 3.7. Post-alignment sequence editing was performed in MacClade v. 4.08 and Gblocks (Castresana 2000). An unaligned NBRF file was uploaded to the Gblocks server and all 3 options for less stringent selection were selected; then the resulting file was downloaded, converted to a fasta file, and used in subsequent analyses. Indels in the cpDNA dataset were coded as binary characters at the end of alignments but no indels were present in the GBSSI or ITS regions. The four cpDNA regions were edited separately, concatenated, then partitioned into their 4 regions. The GBSSI region was edited as a whole then partitioned into an exon region and an intron region (sequences represent a small portion of exon 2, all of exon 3-9 and introns 2,3, 5-8). The ITS region was edited whole and not partitioned.

Phylogenetic model testing in MEGA v. 6 (Tamura et al. 2013) was performed using discrete characters and lowest BIC scores were used for the maximum likelihood and Bayesian analyses (see Supplementary Table 1). Each of the four chloroplast regions showed T92 or T92+G as the best model. The exons of GBSSI were best modeled by K2+G and the introns were best modeled by T92 but GTR+G+I was employed due to low phylogenetic signal in

preliminary tree formation. The ITS region showed K2+G as the best model but GTR+G+I was employed due to low phylogenetic signal in preliminary tree formation. The Tamura 3-parameter model (T92) adds a G-C content bias for the sequence data to the Kimura 2-parameter model (Tamura 1992). The Kimura 2-parameter model uses equal base frequencies, one transition rate, and one transversion rate. The Generalized Time Reversible (GTR) model uses 6 substitution rates and is a more neutral model (Yang 1994). The gamma distribution represents rate variation among sites (Kimura 1980). The invariable sites model (I) allows for a proportion of sites to be invariable.

Phylogenetic analyses were conducted using parsimony (PAUP*4.0, Swofford 2003), maximum likelihood (MEGA v. 6, Tamura et al. 2013), and Bayesian methods (Mr. Bayes v. 3.2.6, Ronquist & Huelsenbeck 2003, or under the CIPRES Gateway, Miller et al. 2010). Parsimony analyses were conducted using a "fast"-stepwise stepwise addition with 10,000 replicates and retaining groups with greater than 50% frequency. Maximum likelihood parameters were set to the appropriate substitution rate in MEGA v. 6 and 200 bootstrap replicates were used in order to maintain adequate sampling (Pattengale 2009). A uniform rate was used for variation among sites and the nearest-neighbor-interchange (NNI) heuristic was employed with a very strong branch swap filter.

Mr. Bayes parameters were set to the appropriate modeled rates of likelihood. Model partitioning was used for the Bayesian analyses based on differences in appropriate substitution models for the 4 cpDNA regions and the exon and intron region of GBSSI. Indels were coded as 0 (absence) and 1 (presence) at the end of the concatenated cpDNA sequence. The MKv model was used to analyze this portion of sequence (Lewis 2001). The Bayesian analyses were conducted in two independent runs with four chains each (3 heated, 1 cold, temp=0.2) and a

sample was taken every 1,000 generations with an initial 1,000,000 generations used. The first 25% of generations were used as burn-in. The ITS analysis required an additional 4,000,000 runs due to high standard deviation of split frequencies. Runs reached stationarity after burn-in and continued until the standard deviation of the split frequencies was lower than 0.01 and a visual examination of likelihood scores showed equal mixing of the two independent runs.

Chapter 3: Results

cpDNA analysis

Each cpDNA region amplified from 1000-1400 bp with a concatenated sequence of 5333 bp, and contained 6.5-9.5% variable sites with 2.1-3.6% parsimony-informative sites. A basic local alignment search in GenBank showed the chloroplast sequences were highly similar to sequences from Shaw et al. (2007), with the highest match equal to 97% for *H. cannabinus.* All three phylogenetic analyses show similar patterns of evolution within section *Furcaria* but the Bayesian analyses will be discussed due to stronger clade support using posterior probability distributions. (See Figures 6-8 for maximum likelihood analyses and Figures 9-11 for parsimony analyses.)

Overall, the genome groups discovered through previous hybridization studies (Menzel & Martin 1971, Menzel & Wilson 1963, Menzel & Martin 1980) form well-supported clades (Fig. 3). The B genome group forms a clade with *H. acetosella* and *H. radiatus* (tetraploids) and *H. surattensis*- (diploid). The two specimens used for *H. surattensis* show an unusual relationship because they are not sister to one another but this relationship has low support (0.55). The Y genome forms a clade with *H. mechowii* (diploid) and *H. sabdariffa* (tetraploid). The G genome forms a clade with the Australian hexaploid group (*H. arnhemensis, H. marenitensis, H. meraukensis,* and 2 *H. splendens*), *H. sudanensis* (diploid), *H. rostellatus* (GH tetraploid), and *H. furcellatus* (GP tetraploid). This confirms the G genome lineage as the maternal donor for the Australian hexaploids. *Hibiscus splendens* specimen 4-3BP shows a long branch from the other *H. splendens* specimen 1423. This is due to a tandem repeat of a single insertion of thymine throughout each cpDNA sequence. The A genome with all A diploid species (*H. asper, H. cannabinus, H.* greenwayi) also forms a clade. The X genome forms a clade with two diploid

species (*H. hiernianus, H. mastersianus*) and two tetraploid species (*H. meeusei, H. nigricaulis*). 1 outgroup was used for this phylogeny- *Urena procumbens.* Additionally, there is split between the B, Y species and the A, G, X species towards the base of the tree. However, the A genome exists in tetraploid species of the B and Y genomes as well (Fig.3). *H. costatus* shows as a sister species to the B and Y genome group but with low support and a long branch. This is consistent with previous findings on *H. costatus* in an unknown genome group.

GBSSI analysis

The exon regions of GBSSI amplified a total of approximately 800 bp, and contained 13.9% variable sites with 5.1% parsimony-informative sites. The intron regions of GBSSI amplified a total of approximately 546 bp, and contained 26.7%, variable sites with 9.7% parsimony-informative sites. A basic local alignment search in GenBank showed the GBSSI sequences were much different than most of the sequences on GenBank and the highest match was equal to 78% for *H. moschuetos* subsp. *moschuetos*.

The Bayesian analysis of the GBSSI region (Fig. 4) shows support for the grouping of species within section *Furcaria* but not as strongly as the cpDNA analysis. A clade of 8 GJV copies, including three of the Australian hexaploid species sampled, forms with *Hibiscus sudanensis*, confirming the G genome as a donor lineage to the Australian hexaploids. A clade with low support (0.7 posterior probability) forms with *H. mechowii* of the Y group, *H. surattensis* of the B group, and *4 H.* splendens clones from the GJV group. Lastly, a separate clade of 8 GJV copies, involving all 4 of the Australian hexaploid species sampled forms near the base of the tree, indicating uncertainty in the genome group it may represent. 1 outgroup was used for this phylogeny- *H. macrophyllus.* Interestingly A, G, X genome groups form a well-

supported clade matching the pattern seen in the chloroplast regions. The B and Y genome groups are once again split from the A/G/X clade but the support is low for the placement of this division in the GBSSI gene region (0.53 posterior probability). *Hibiscus costatus* again forms a polytomy near the base of the phylogeny further indicating an uncertain relationship for this American diploid species.

More extensive sampling of this gene region in *Furcaria* is required (only 20 clones successfully generated) to eliminate GBSSI as a useful region for phylogenetic inference of donor lineages to the J or V genomes of the Australian species.

ITS analysis

A basic local alignment search in GenBank showed ITS sequences were somewhat similar to existing sequences and the highest match was equal to 94% for *H. sabdariffa*. The ITS regions amplified a total of approximately 750 bp, and contained 54.8% variable sites with 28.4% parsimony-informative sites.

The Bayesian analysis of the ITS region (Fig. 5) also shows support for genome groups within *Furcaria* forming clades but some have with low support. The A genome forms two clades due to strong heterozygosity within the diploids of this group (Small et al. unpub). The top-most A genome clade contains 1 *H. asper* copy (diploid), all 3 *H. cannabinus* copies (diploid), 5 tetraploid copies (*H. acetosella, H. meeusei, H. nigricaulis, H. radiatus, H. sabdariffa*), and 1 GJV copy from *H. meraukensis* but with low support (0.56 posterior probability). The lower-most A genome clade contains 2 *H. asper* copies (diploid), 2 *H. greenwayi* copes (diploid), 3 tetraploid copies (*H. meeusei, H. nigricaulis, H. sabdariffa),* and all 3 copies of the GP tetraploid, *H. furcellatus* with high support (1 posterior probability). This may

indicate the P genome of *H. furcellatus* is a modified A genome but the posterior probability is low for this relationship (0.69). The G genome forms a clade with a diploid *H. sudanensis* (diploid), 2 copies of the tetraploid *H. furcelltatus* (GH), and 44 GJV hexaploid copies (23 *H. splendens,* 13 *H. arnhemensis*, 7 *H. marenitensis*, 2 *H. meraukensis*). All clones from the Australian hexaploids form a clade with the G genome species and do not aid in the elucidation of the J or V genome donor species to these lineages. This is likely due to high concerted evolution in the G genome species for the ITS region (Small et al. unpub.) Additional studies have found ITS sometimes does not provides high taxonomic resolution due to non-variable sites and complex indels (Schilling, Mattson, & Floden 2013). The B genome forms a clade with 2 copies of *H. surattensis* (diploid), 1 copy of *H. acetosella* (tetraploid) and 1 copy of *H.* radiatus (tetraploid). *Hibiscus costatus* (diploid) maintains the uncertain phylogenetic pattern seen in the cpDNA and GBSSI analysis. The X genome forms a clade with two diploid species (*H. hiernianus, H. mastersianus*) and two tetraploid species (*H. meeusei, H. nigricaulis*). The Y genome forms a clade with 2 individuals of *H. mechowii* (diploid) and 2 individuals of *H. sabdariffa* (tetraploid). 2 outgroups were used for this phylogeny- *H. macrophyllus* and *Urena procumbens*. No split of the B/Y clades from the A/G/X clades is present in the ITS phylogeny. In total 44 clones were sequenced for the ITS region. Fig. 16 shows this level of sampling would be adequate for finding all the copies in the Australian hexaploids. However, this assumes the gene copies are sequenced with equal probability. Therefore, more sampling might be useful in for the ITS region as well.

Chapter 4: Discussion

Implications

Hibiscus section *Furcaria* is well supported as a distinct clade with multiple genome groups. The previous morphological and cytological work performed by Menzel and others (Menzel 1963, Menzel & Martin 1980, Menzel et al. 1983) is consistent with the phylogenetic relationships found in the present study. The cpDNA phylogeny clearly shows monophyletic groups for the A, B, X, Y, and G genomes and a potential split in the clades giving rise to the B,Y genomes and the A,G,X genomes. This split may help explain the current distribution of species for section *Furcaria* seen in the world, but no clear patterns are seen in the distribution of the diploids on the African continent to support this division (Wilson 1993).

The GJV genome group is shown to be monophyletic using maternal cpDNA and emerges from the G genome group represented by *H. furcellatus* (GP)*, H. rostellatus* (GH)*,* and *H. sudanensis* (G). This also confirms *H. furcellatus* and *H. rostellatus* as relatives of the G genome group. The G copy from *H. sudanensis* forms an unresolved polytomy within the G/GJV clade. One possible explanation for this is that the GBSSI and ITS regions may not diverge enough in the J or V genomes to separate *H. sudanensis* from the Australian hexaploids. The G genome may be homogenized through recombination and maintained during hybridization events. The ITS gene region is known to sometimes provide low phylogenetic resolution due to the presence of pseudogenes, introgression, and extensive variation after duplication events (Álvarez & Wendel 2003) so this region may not be suitable for analysis within *Furcaria* without additional methods.

The G genome diploid lineage represented by *H. sudanensis* is located in a specific area of Central Africa (Zaire, Central African Republic, and South Sudan) but species with the G

genome show a widespread distribution across the Southern Hemisphere, suggesting a Gondwanan distribution. However, the Eumalvoideae clade that contains *Hibiscus* is much too young to be influenced by continental divides, with an estimated divergence time of 58-60 mya (Carvalho et al. 2011). Therefore long-distance dispersal coupled with vicariance within multiple continents likely explains the current distribution of this section.

Hibiscus tiliaceus provides an example of a species within the genus, but not the section, that could explain how a genome may have spread and developed through long-distance dispersal. *Hibiscus tiliaceus* or the sea hibiscus is a pantropical species that likely underwent multiple long-distance dispersal events but retains species continuity through gene flow (Takayama et al. 2008). Tang et al. (2011) show habitat differentiation of estuarine and inland populations in southern China for this species using retrotransposon markers. Gene flow is prominent in the estuarine populations of *H. tiliaceus* and this might explain how polyploid genome groups could occur in the same area and undergo allopolyploidization. Once a species reaches the coastline it may become established and could diverge through vicariance, later that same species or a different one may arrive and hybridize with this diverged species. Cotton studies provide an example of how species with different genome groups can migrate long distances and hybridize to form combinations in novel environments (Wendel & Cronn 2003, Wendel & Grover 2015). *Hibiscus tiliaceus* is not within section *Furcaria* but is distributed in a similar coastline area of northern Australia. No direct conclusions about the biogeography of the J or V genome donor linegaes are evident from this work, but more integrated methods with *in situ* and *ex situ* components in estuarine plant populations may reveal a connection between ploidy and biogeography in section *Furcaria.*

The J and V genomes do not show a clear pair of diploid progenitors. However, some diploids are likely candidates for further study. *Hibiscus rostellatus,* a GH tetraploid is one such candidate species. The H genome may be homologous to the J or V genomes (Small et al. unpub.), as evidence shows *H. rostellatus* can successfully cross with *H. heterophyllus,* a GJV species (Menzel & Martin 1971). *H. rostellatus* may have developed as an autotetraploid (i.e. GG) and diverged from the G genome early on in the evolution of section *Furcaria* to become GH. This GH tetraploid could then have hybridized with another species in *Furcaria* with a different genome group. The ITS phylogeny shows *H. rostellatus* in a clade with the GJV species but increased sampling of *H. rostellatus* is necessary to determine if this grouping is due to the G or H genome.

Hibiscus mechowii is sister to one of the Australian GJV genome group clades in the GBSSI phylogeny but with weak support (0.7 posterior probability). Morphologically this species is an intriguing potential donor for the J or V genome because it is one of only two species that has both an entire set of bracts and no calyx nectaries, two unique characteristics of the GJV Australian species (Wilson 1993). *Hibiscus australensis*, a species with unknown genome group or chromosome count, occurs on the Austral Islands in French Polynesia and also shares this unique morphology (Wilson 2006). The distribution of the Y genome is currently limited to central Sub-Saharan Africa with only one tetraploid species known, *H. sabdariffa* (AY). The evolution of the Y genome into the J or V genome may be possible due to the limited amount of species represented in the Y genome currently but increased sampling of Y genome species is necessary to provide an adequate result.

Future Work

Additional clonal sampling and phylogenetic analyses into other nuclear regions could promote the understanding of evolution and polyploidy in *Hibiscus* section *Furcaria* and lead to definitive evidence of the biogeography within this group. Emphasis should be placed on increased clonal sampling for a single, easily-sequenced hexaploid species and determining the genome groups of species with unknown types- *H. costatus,* Pacific Island species, and other Australian species in order to eliminate or confirm these species as being donors or relatives of the Australian hexaploids.

Next-gen sequencing is key to the study of polyploidy in the future as genomes can be separated from one another through indices using SNPs (Page et al. 2013). The current study developed a small picture of what is occurring in the polyploid genomes of *Hibiscus* section *Furcaria* but sequencing larger portions of the genome could explain the story of evolution in this group more fully. In addition to the development of molecular methods more hybridization studies could be performed to test the validity of potential diploid donors.

Lastly, more clones are likely needed to ensure full coverage of the gene copies within the hexaploid of *Hibiscus* section *Furcaria.* In order to determine the adequate level of sampling necessary to obtain all gene copies from the Australian hexaploids a probability function was created in R by Dr. Brian O'Meara to assess an adequate level of sampling. Assuming all genome groups are sequenced at the same rate the probability of sampling all genome groups will increase as more clones are sampled. Fig. 12 shows two lines representing the number of gene copies. The red line represents 6 copies while the black line represents 3 gene copies. The number 6 was chosen to represent each copy of a gene that is actually present in the Australian hexaploids (GGJJVV) and 3 was chosen to represent each copy from one pair of the sister

chromosomes in each genome group (G, J, or V) in this polyploid group. The blue line shows a 95% probability of sampling all gene regions. If pairs of sister chromosome copies are the same for the gene regions used (black line) the level of adequate sampling is quite low (below 10). However, if each 6 gene copies are needed then the 95% probability requires close to 25 clones (a level only obtained by ITS copies.) This also assumes each species contains G, J, and V genomes that are not altered in the nuclear gene regions sampled for this study (ITS, GBSSI).

Lastly, the adaptive radiation of the Australian species likely occurred due to open niche space after long-distance dispersal of seeds from one continent to the other combined with vicariance, but the specific geographic or biological factors leading to this speciation are relatively unknown. Perhaps the release from sugar production in calyx nectaries for the allowed for an adaptive advantage in the Australian progenitor that led to speciation but this has yet to be tested. Figures 1 and 2 show these species occur in distinct geographic areas and some have much wider distributions than others. The reason for this ecological separation is currently unknown and could warrant further study as well. Further knowledge of the ecological factors that construct complex polyploid relationships in this group would be greatly beneficial.

References

Adams, K. L., & Wendel, J. F. (2005). Polyploidy and genome evolution in plants. *Current opinion in plant biology*, *8*(2), 135-141.
Álvarez, I., & Wendel, J. F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Molecular phylogenetics and evolution*, *29*(3), 417-434.
Andreuzza S, Siddiqi I (2008) Spindle Positioning, Meiotic Nonreduction, and Polyploidy in Plants. *PLoS Genet*, 4(11): e1000272. doi:10.1371/journal.pgen.1000272
Blanc, G., & Wolfe, K. H. (2004). Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *The Plant Cell*, *16*(7), 1667-1678.
Carvalho, M. R., Herrera, F. a., Jaramillo, C. A., Wing, S. L., & Callejas, R. (2011). Paleocene Malvaceae from northern South America and their biogeographical implications. *American Journal of Botany*, *98*(8), 1337–1355. doi:10.3732/ajb.1000539
Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis.Molecular Biology and Evolution, 17, 540-552.
Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nature Reviews Genetics*, 6 (11): 836-846.
Craven, L. A., Wilson, F. D., & Fryxell, P. A. (2003). A taxonomic review of Hibiscus sect. Furcaria (Malvaceae) in Western Australia and the Northern Territory. *Australian Systematic Botany*, *16*, 185–218. doi:10.1071/SB01046
d'Erfurth I, Jolivet S, Froger N, Catrice O, Novatchkova M, et al. (2008) Mutations in ATPS1 (Arabidopsis thaliana Parallel Spindle 1) lead to the production of diploid pollen grains. *PLoS Genet,* 4:e1000274. doi:10.1371/journal.pgen.1000274.
Greiner, S., Sobanski, J., & Bock, R. (2015). Why are most organelle genomes transmitted maternally? *Bioessays*, *37*(1), 80–94. doi:10.1002/bies.201400110
Hahn MA, van Kleunen M, Müller-Schärer H. (2012). Increased phenotypic plasticity to climate may have boosted the invasion success of polyploid *Centaurea stoebe*. *Plos One* 7:e50284.
Hochreutiner, B. P.G., (1900). Revision du genre Hibiscus. Ann.Cons.*Jard.BotGenev,* 4: 23-191.
Hooke, R. L., Martín-Duque, J. F., & Pedraza, J. (2012). Land transformation by humans: a review. *GSA today*, 22(12), 4-10.
Hull-Sanders H.M., Johnson R.H., Owen H.A., Meyer G.A. (2009). Effects of polyploidy on secondary chemistry, physiology, and performance of native and invasive genotypes of *Solidago gigantea* (Asteraceae). *American Journal of Botany* 96:762-770.
Husband B.C., Baldwin S.J., Suda J. (2013). The incidence of polyploidy in natural plant populations: major patterns and evolutionary processes. In Greilhuber J, Dolezel J, Wendel JF, eds. *Plant Genome Diversity Volume 2*. Vienna: Springer, 255-276.
Husband B.C., (2004). The role of triploid hybrids in the evolutionary dynamics of mixed-ploidy populations. *Biological Journal of the Linnean Society* 82:537-546.
ISAAA. (2013). ISAAA Brief 46-2013: Executive Summary. Global Status of Commercialized Biotech/GM Crops: 2013.
Jiao, Y., (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97-102.
Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of molecular evolution*, 16(2), 111-120.
Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological

character data. *Systematic biology*, *50*(6), 913-925.
Levin, D. A. (1975). Minority cytotype exclusion in local plant populations.*Taxon*, 35-43.
Lutz A.M. (1907). A preliminary note on the chromosomes of Oenothera lamarckiana and one of its mutants, O. gigas. Science 26:151–152
Mason-Gamer, R. J. (2013). Phylogeny of a genomically diverse group of Elymus (Poaceae) allopolyploids reveals multiple levels of reticulation. *PLoS One*, *8*(11), e78449.
Marshall, W. F. (2009). Centriole evolution. *Current opinion in cell biology 21*(1), 14-19.
Menzel, M. Y., Goetz, S. G., & Adamson, W. C. (1983). Some pieces of the African genome puzzle in Hibiscus sect. Furcaria (Malvaceae). *American Journal of Botany*, 285-297.
Menzel, M. Y., & Martin, D. W. (1980). Evidence for the presence of an intercontinental genome in the Australian hexaploid alliance of Hibiscus Sect. Furcaria. *Australian Journal of Botany*, *28*(3), 369-383.
Menzel, M. Y., & Martin, D. W. (1971). Chromosome homology in some intercontinental hybrids in Hibiscus sect. Furcaria. *American Journal of Botany*, 191-202.
Menzel, M. Y., & Wilson, F. D. (1963). Cytotaxonomy of twelve species of Hibiscus section Furcaria. *American Journal of Botany*, 262-271.
Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA pp 1 - 8.
Page, J. T., Gingle, A. R., & Udall, J. A. (2013). PolyCat: A Resource for Genome Categorization of Sequencing Reads From Allopolyploid Organisms. *G3: Genes|Genomes|Genetics*, *3*(3), 517–525. http://doi.org/10.1534/g3.112.005298
Pandit M.K., White S.M., Pocock M.J. (2014). The contrasting effects of genome size, chromosome number and ploidy level on plant invasiveness: a global analysis. *New Phytologist* 203:697-703.
Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R., Moret, B. M., & Stamatakis, A. (2009, May). How many bootstrap replicates are necessary?. In *Research in Computational Molecular Biology* (pp. 184-200). Springer Berlin Heidelberg.
Pfeil, B. E., & Crisp, M. D. (2005). What to do with Hibiscus? A proposed nomenclatural resolution for a large and well known genus of Malvaceae and comments on paraphyly. *Australian Systematic Botany*, *18*(1), 49-60.
Pfeil, B. E., Brubaker, C. L., Craven, L. A., & Crisp, M. D. (2004). Paralogy and orthology in the Malvaceae rpb2 gene family: investigation of gene duplication in Hibiscus. *Molecular Biology and Evolution*, *21*(7), 1428-1437.
Ramsey, J., & Ramsey, T. S. (2014). Ecological studies of polyploidy in the 100 years following its discovery. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*(1648), 20130352.
Ramsey, J. & Schemske, D. W. (2002). Neopolyploidy in flowering plants. *Annu. Rev. Ecol. Syst.* 33, 589–639.
Ramsey J., Schemske D.W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review Of Ecology And Systematics* 29:467-501.
Robertson K., Goldberg E.E., Igic B. (2011). Comparative evidence for the correlated evolution of polyploidy and self-compatibility in Solanaceae. *Evolution* 65:139-155.
Ronquist, F. and J. P. Huelsenbeck. (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Sattler, M. C., Carvalho, C. R., & Clarindo, W. R. (2016). The polyploidy and its key role in plant breeding. *Planta*, *243*(2), 281-296.
Schilling, E.E., N. Mattson, and A. Floden. (2014). Barcoding the Asteraceae of Tennessee, tribe Coreopsideae. *Phytoneuron* 2014-101: 1–6.
Segraves K.A., Thompson J.N. (1999). Plant polyploidy and pollination: floral traits and insect visits to diploid and tetraploid *Heuchera grossularifolia*. *Evolution* 53:1114-1127.
Shaw, J., Shafer, H. L., Leonard, O.R, Kovach, M. J., Schorr, M., & Morris, A.B. (2014). Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: The tortoise and the hare IV. *American Journal of Botany* 101(11): 1987-2004.
Shaw, J., Lickey, E.B., & Small, R.S. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany* 94(3): 275-288.
Small, R. L., Ryburn, J. A., Cronn, R. C., Seelanan, T., & Wendel, J. F. (1998). The tortoise and the hare: choosing between noncoding plastome and nuclear Adh sequences for phylogeny reconstruction in a recently diverged plant group. *American Journal of Botany 85*(9), 1301-1315.
Small, R. (2004). Phylogeny of *Hibiscus* sect. *Muenchhusia* (Malvaceae) Based on Chloroplast rpL16 and ndhF, and Nuclear ITS and GBSSI Sequences. *Syst. Botany* 29, 2.
Soltis D.E., Buggs R.J., Doyle J.J., Soltis P.S. (2010). What we still don't know about polyploidy. *Taxon* 59:1387-1403.
Stebbins, G. L. (1970). Variation and evolution in plants: progress during the past twenty years. In *Essays in evolution and genetics in honor of Theodosius Dobzhansky* (pp. 173-208). Springer US.
Swofford, D. L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
Takayama, K., Tateishi, Y., Murata, J. I. N., & Kajita, T. (2008). Gene flow and population subdivision in a pantropical plant with sea‐drifted seeds Hibiscus tiliaceus and its allied species: evidence from microsatellite analyses. *Molecular Ecology*, *17*(11), 2730-2742.
Tamura, K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+ C-content biases. *Molecular biology and evolution*, *9*(4), 678-687.
Tamura, K. Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0 *Molecular Biology and Evolution,* 30:25-29.
Tang, T., He, L., Peng, F., & Shi, S. (2011). Habitat differentiation between estuarine and inland Hibiscus tiliaceus L. (Malvaceae) as revealed by retrotransposon-based SSAP marker. *Australian Journal of Botany*, *59*(6), 515-522.
Tate, J. A., Soltis, D. E., & Soltis, P. S. (2005). Polyploidy in plants. *The evolution of the genome*, 371-426.
Wendel JF, Grover CE. (2015). Taxonomy and Evolution of the Cotton Genus, Gossypium. In *Cotton, Agronomy Monograph, American Society of Agronomy, Inc., Crop Science Society of America, Inc*., and *Soil Science Society of America, Inc.*
Wendel, J. F., & Cronn, R. C. (2003). Polyploidy and the evolutionary history of cotton. *Advances in agronomy*, *78*, 139-186.
Wilson, F. D. (2006). A distributional and cytological survey of the presently recognized taxa

of Hibiscus section Furcaria (Malvaceae). *Bonplandia*, 53-62.
Wilson, F.D. (1994). The genome biogeography of Hibiscus L. section Furcaria DC. *Genetic Resources and Crop Evolution* 41: 13-25.
Wood, T. E., Takebayashi, N., Barker, M.S., Mayrose, I., Greenspoon, P.B., & Rieseberg, L.H. (2009). The Frequency of Polyploid Speciation in Vascular Plants. *Proceedings of the National Academy of Sciences of the United States of America,* 106.33, 13875-79.
Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular evolution*, *39*(3), 306-314.

Appendices

Appendix A: Tables

Table 1 Plant materials used in this study from *Hibiscus* section *Furcaria*. Voucher specimens deposited at the University of TN Herbarium for Australian species sampled. Adapted from Wilson (1993) Table 1.

Species	Accession #	Genome Group	Area of origin
H. acetosella	A59-152	AB	Africa
H. apser	A82-1255, A93-1417	A	Africa
H. arnhemensis	A68-754	GJV	Australia
H. cannabinus	74200 I4, 51-7	A	Africa
H. costatus	A60-243	Unknown, but diploid	America
H. furcellatus	A91-1412	GP	Americas
H. greenwayi	A75-1085, A75-1177	A	Africa
H. hiernianus	A69-803	X	Africa
H. marenitensis	A86-1389	GJV	Australia
H. mastersianus	A75-1136	X	Africa
H. mechowii	A70-814	Y	Africa
H. meeusei	A69-805	AX	Africa
H. meraukensis	A82-1288	GJV	Australia
H. nigricaulis	PI 196180	AX	Africa
H. radiatus	A59-53	AB	India
H. rostellatus	A60-211	GH	Africa
H. sabdariffa	A59-56	AY	Africa
H. splendens	A94-1423, 4-3 BP	GJV	Australia
H. sudanensis	PI 267679	G	Africa
H. surattensis	A75-1139, A59-151	B	Africa, Asia

Table 2 Australian plants grown from seed in *Hibiscus* section *Furcaria*. The 4 specimens used for this study deposited at the University of TN Herbarium have assigned collection numbers.

Species	Germinated	Accesion Number	Collection Number
H. arnhemensis	Yes	A68-754	RSmall (s.n.)
H. fallax	No	LL 1579	
H. forsteri	Yes	A94-1428	
H. fryxelli var. fryxellis	Yes	A86-1398	
H. fryxelli var. mollis	No	A84-1337	
H. marenitensis	Yes	A86-1389	WHoskins001
H. menzeliae	Yes	A67-717	
H. meraukensis	Yes	A82-1288	WHoskins002
H. reflexus	No	A82-1264	
H. splendens	Yes	A94-1423	WHoskins003
H. superbus	Yes	A86-1400	
H. zonatus	No	A84-1369	

Table 3 Percent of variable and parsimony-informative sites from all sequence analyses.

	%Variable Sites	% Parsimony-Informative Sites
cpDNA		
ndhF-rpl32R	9.5	2.8
rpl32F-trnL	6.5	3.0
ndhC-trnV	8.3	3.6
trnK-rps16	8.6	2.1
nDNA		
GBSSI		
exons	13.9	5.1
introns	26.7	9.7
ITS	54.8	28.4

Appendix B: Figures

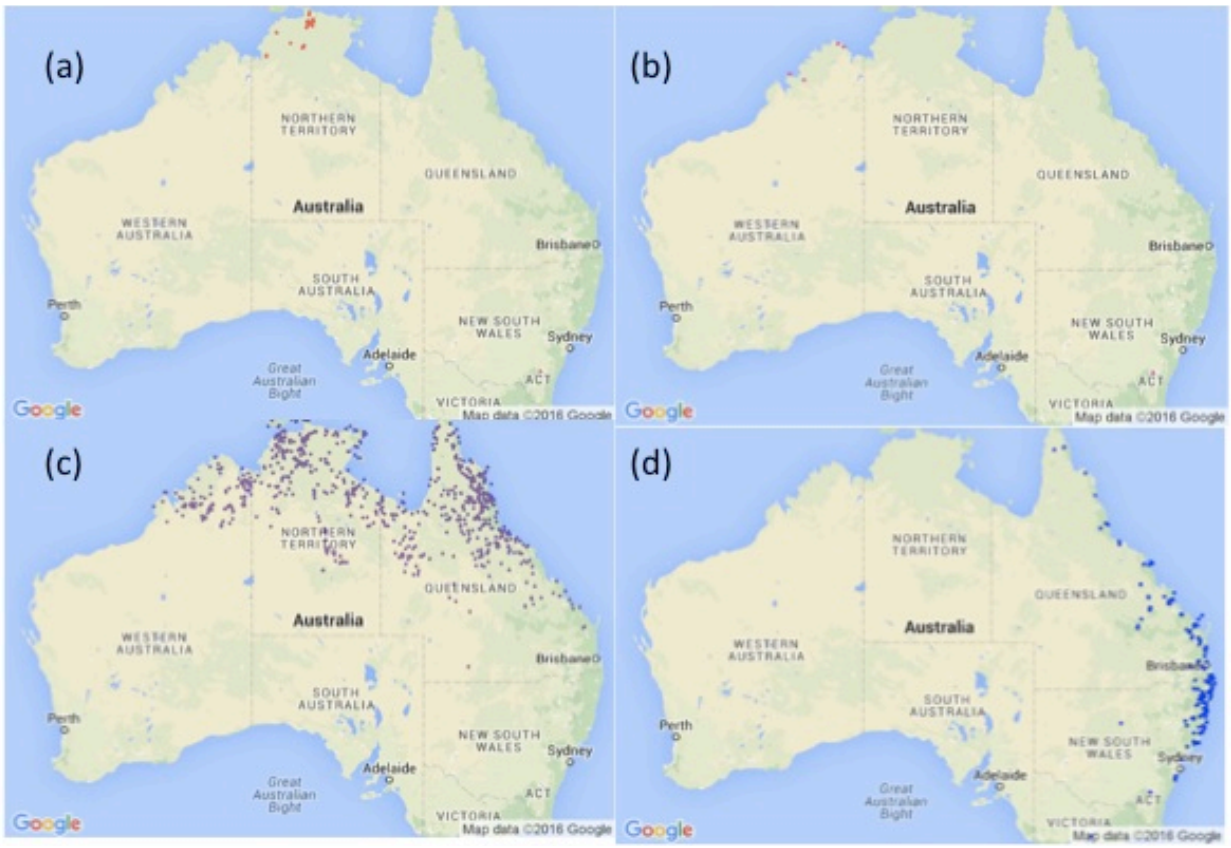


Fig. 1 Species occurrence maps for the four studied species in *Hibiscus* section *Furcaria.* Maps constructed on Living Atlas of Australia (http://www.ala.org.au/) for the four species sampled Australian species in this study. Occurrences recorded by various organizations and government agencies of Australia. (a) red dots represent *H. arnhemensis* (b) pink dots represent *H. marenitensis* (c) purple dots represent *H. meraukensis* (d) blue dots represent *H. splendens*

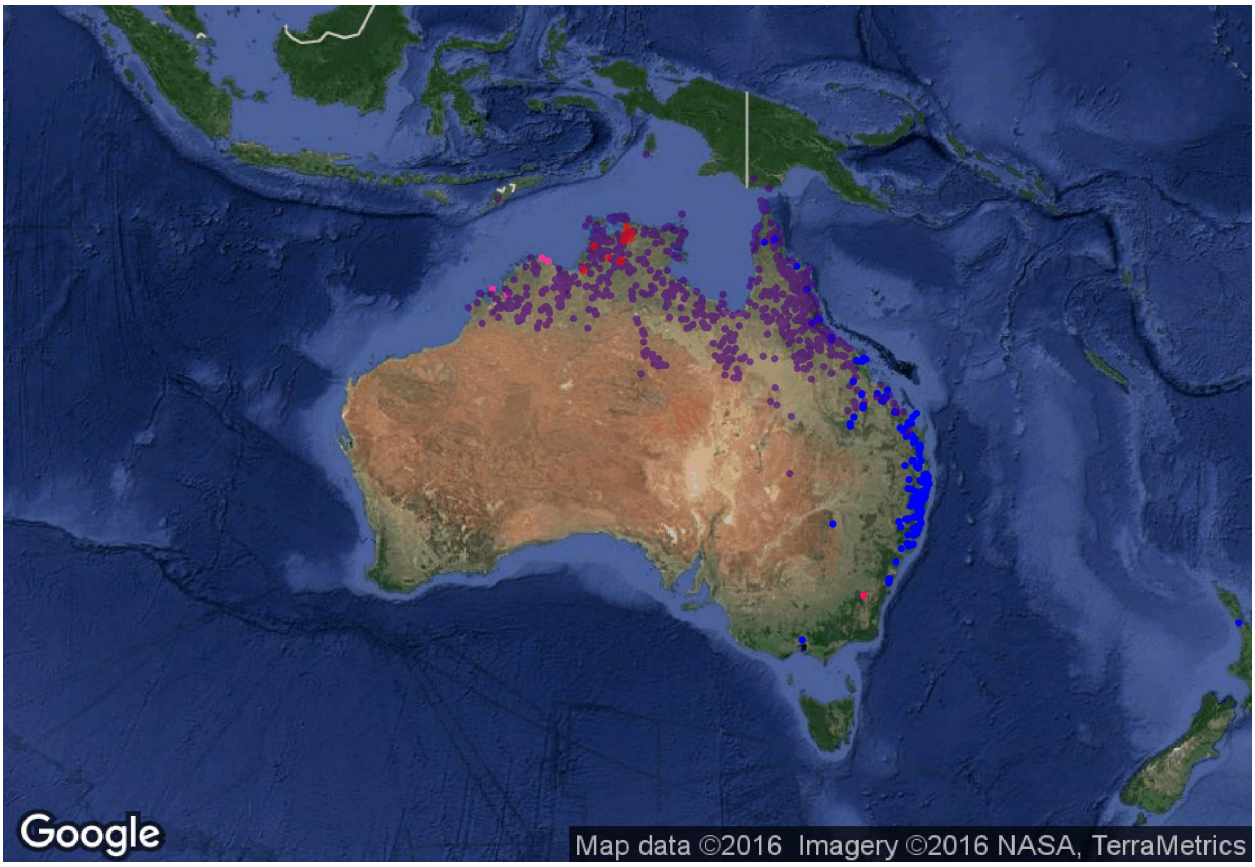


Fig. 2 Total species occurrence map for the four studied species in *Hibiscus* section *Furcaria.* Occurrences recorded by various organizations and government agencies of Australia. (a) red dots represent *H. arnhemensis* (b) pink dots represent *H. marenitensis* (c) purple dots represent *H. meraukensis* (d) blue dots represent *H. splendens*

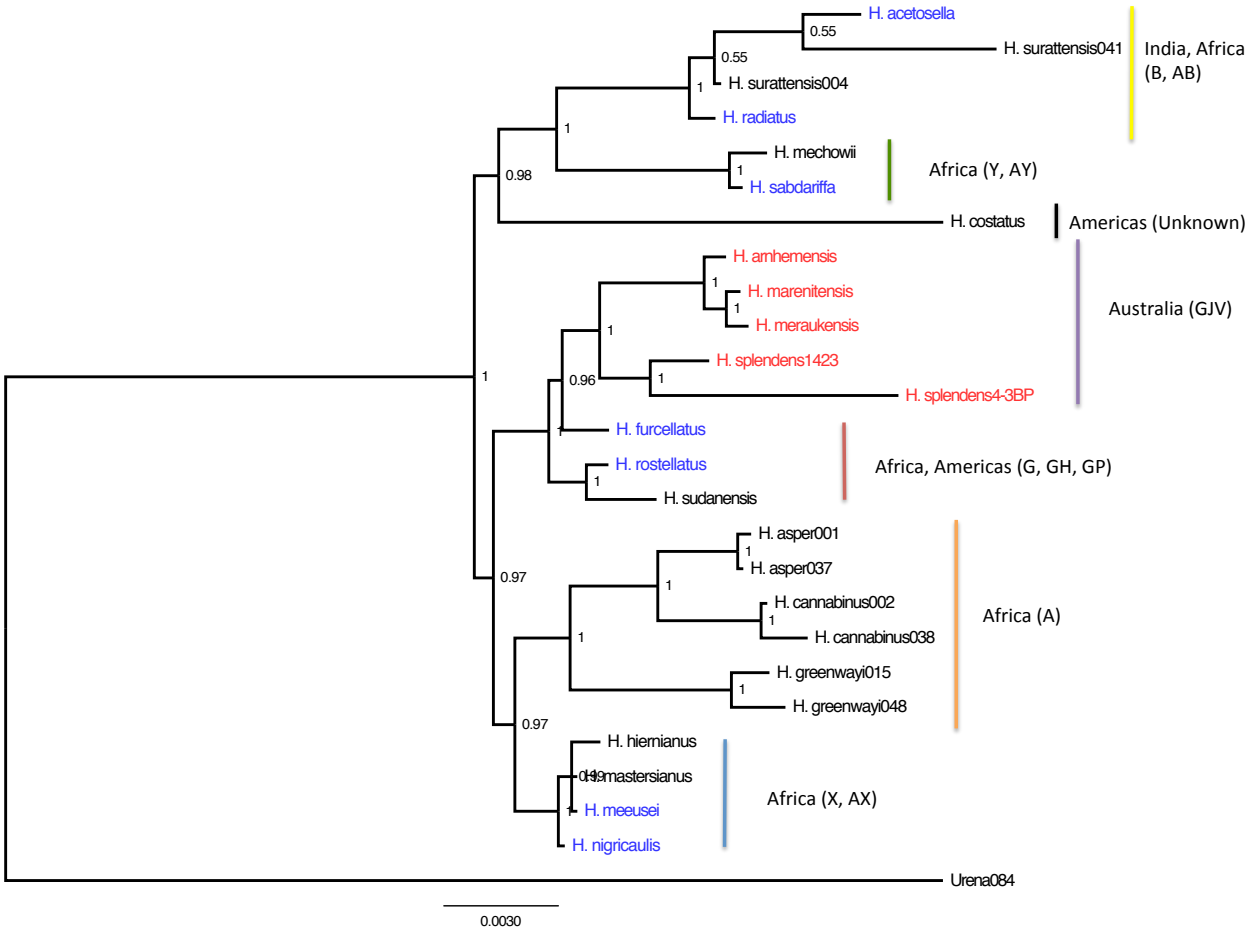


Fig. 3 Bayesian analysis of 4 cpDNA regions with genome groups, geography, and ploidy level Colored bars represent genome groups and their location. Colored taxa correspond to ploidy level (black=diploid, blue=tetraploid, red=hexaploid). This Bayesian analysis shows a phylogeny using 4 regions (*ndhF-rpl32R, rpl32F-trnL, ndhC-trnV, rps16-trnK*) of *Hibiscus* L. section *Furcaria* with posterior probabilities at nodes. This phylogeny shows support for an association between genome groups and biogeographic speciation of *Hibiscus* section *Furcaria*. The Australian GJV group nested within the maternal G group. Bar below represents average substitutions per site.

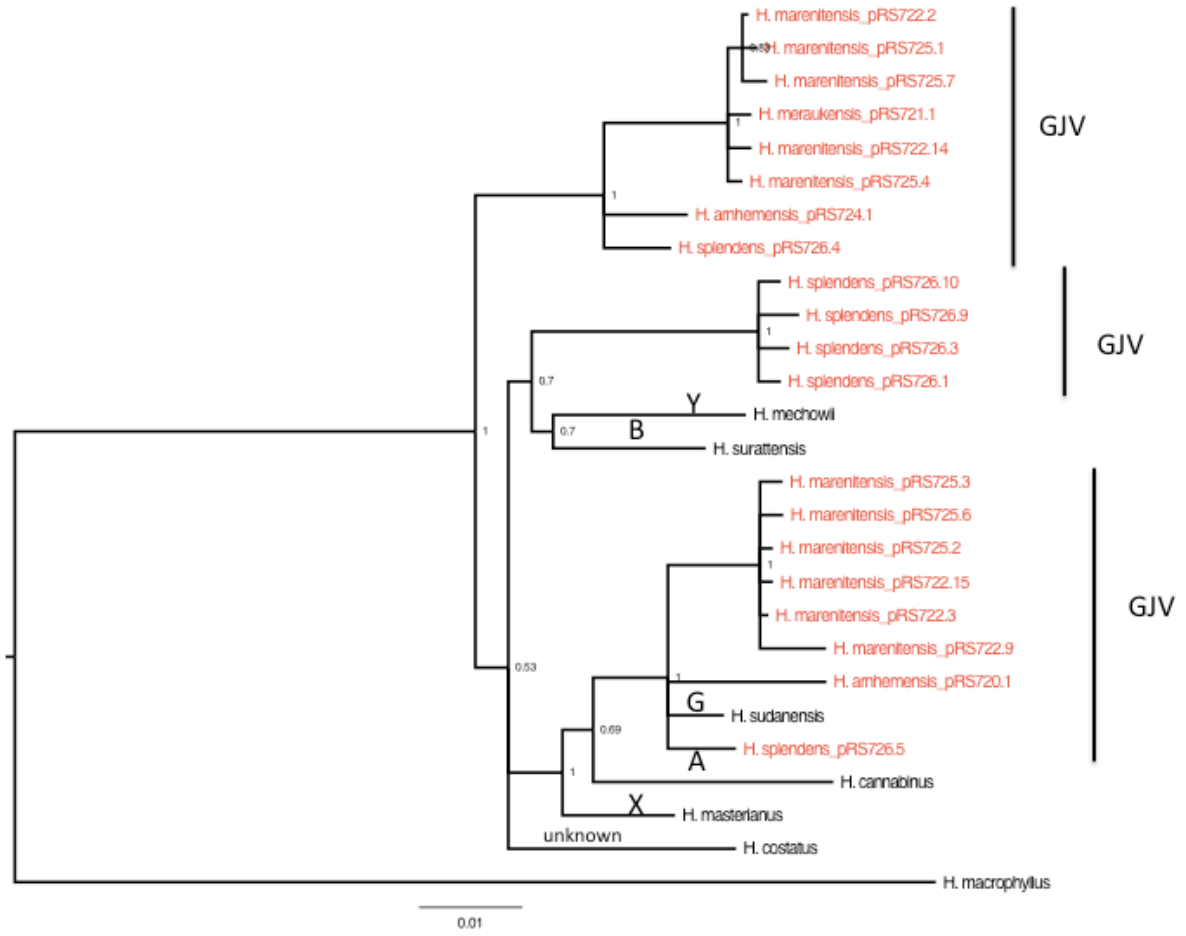


Fig. 4 Bayesian analysis of GBSSI region
This Bayesian analysis of *Hibiscus* L. section *Furcaria* uses GBSSI (granule-bound starch synthase I) and displays posterior probabilities at nodes. This phylogeny shows high support for *H. sudanensis* representing a donor lineage to the Australian species and low support for *H. mechowii* (Y group) or *H. surratensis* representing a donor lineage to the Australian species. Two other groups of Australian species are contained within the phylogeny but it is unclear what donor diploid species are contributing to their hexaploidy. Black taxa = diploid species, red taxa= hexaploid species. Bar below represents average substitutions per site.

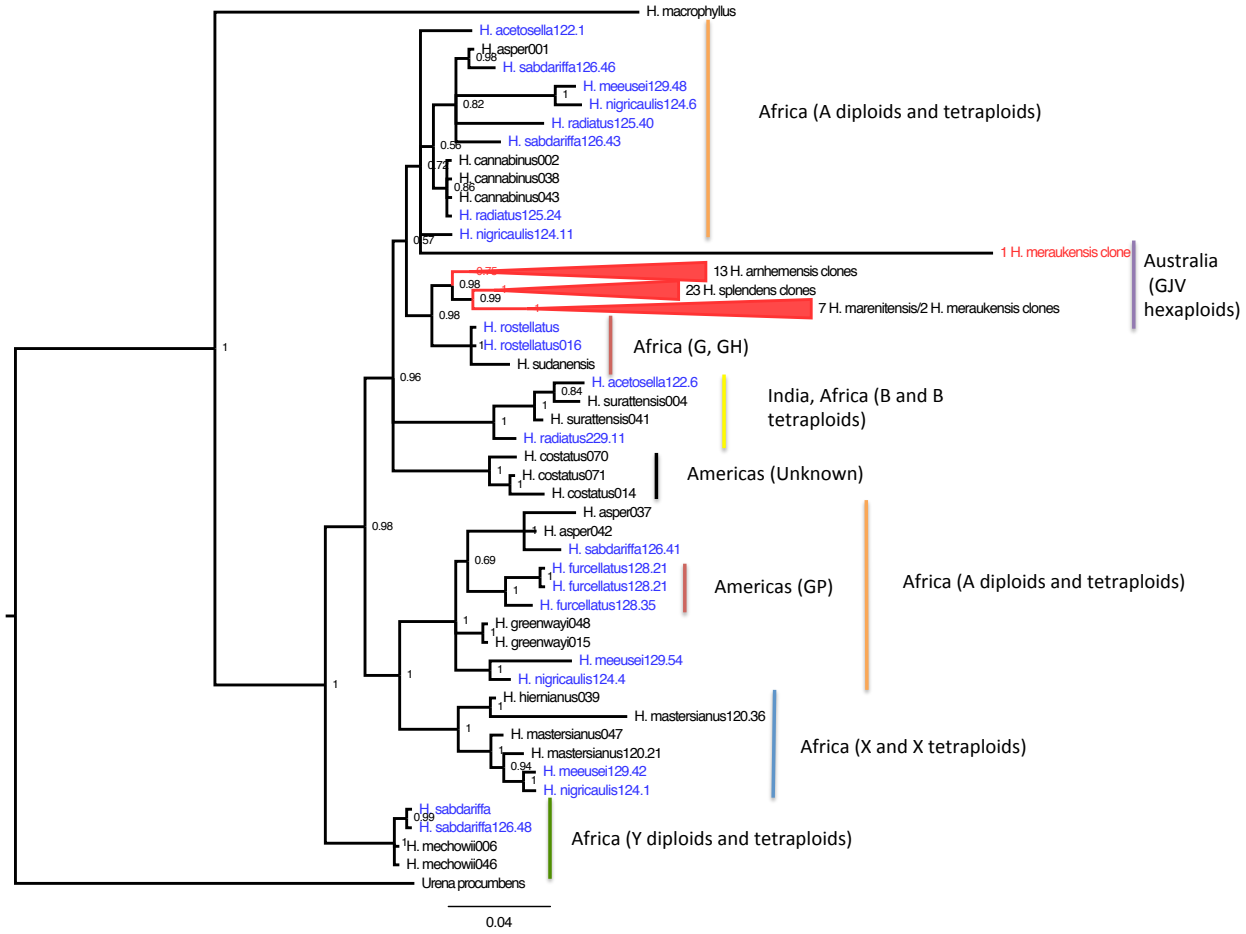


Fig. 5 Bayesian analysis of ITS region
This Bayesian analysis of *Hibiscus* L. section *Furcaria* uses ITS (internal transcribed spacer) with posterior probabilities displayed at nodes. Colored bars represent genome groups and their location. Colored taxa correspond to ploidy level (black=diploid, blue=tetraploid, red=hexaploid). The phylogeny shows support clades for most genome groups and high support for *H. sudanensis* representing a donor lineage to the Australian species but no other lineages. Bar below represents average substitutions per site.

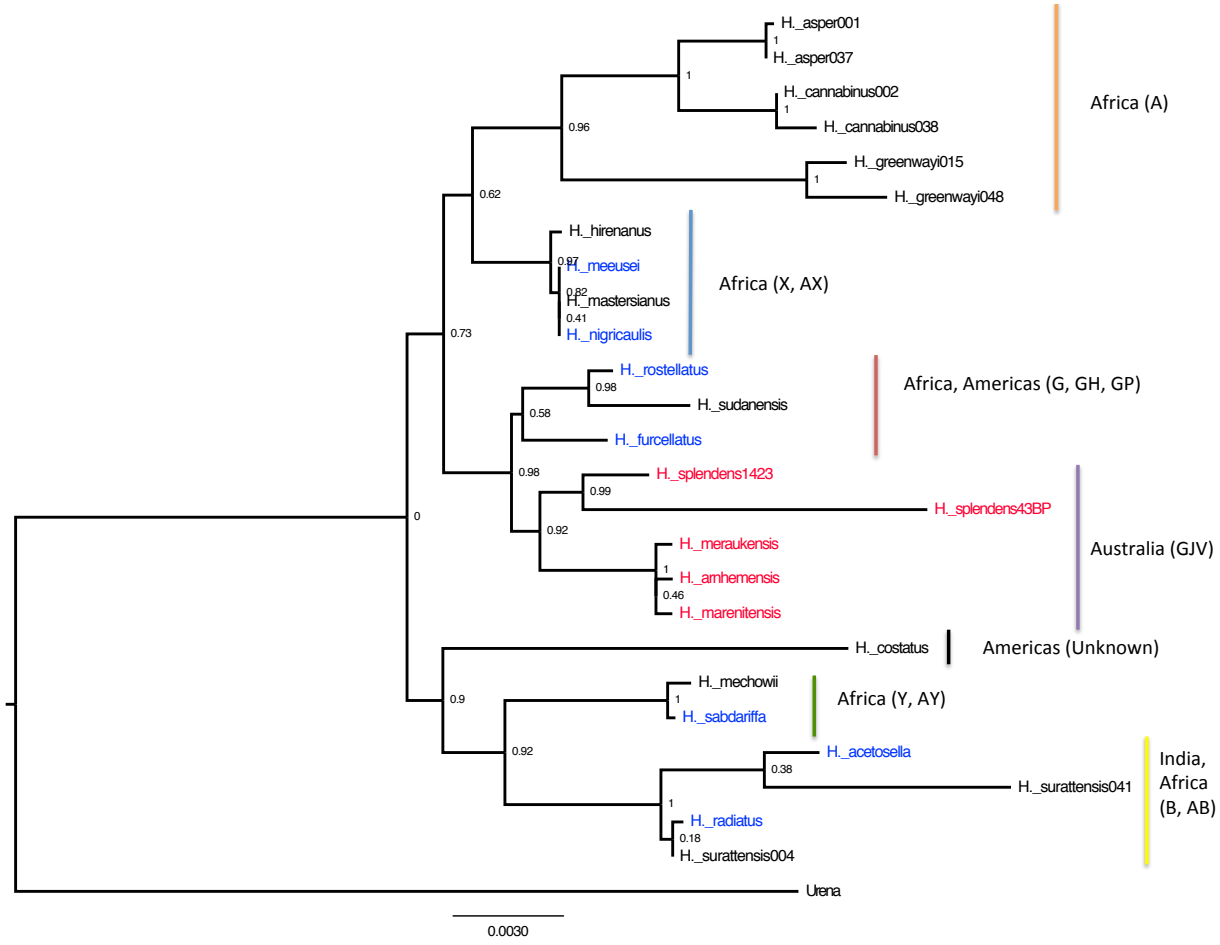


Fig. 6 Maximum likelihood analysis of 4 cpDNA regions
The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura 3-parameter model. The tree with the highest log likelihood (-11083.9067) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 26 nucleotide sequences. There were a total of 5407 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. Bar represents average substitutions per site.

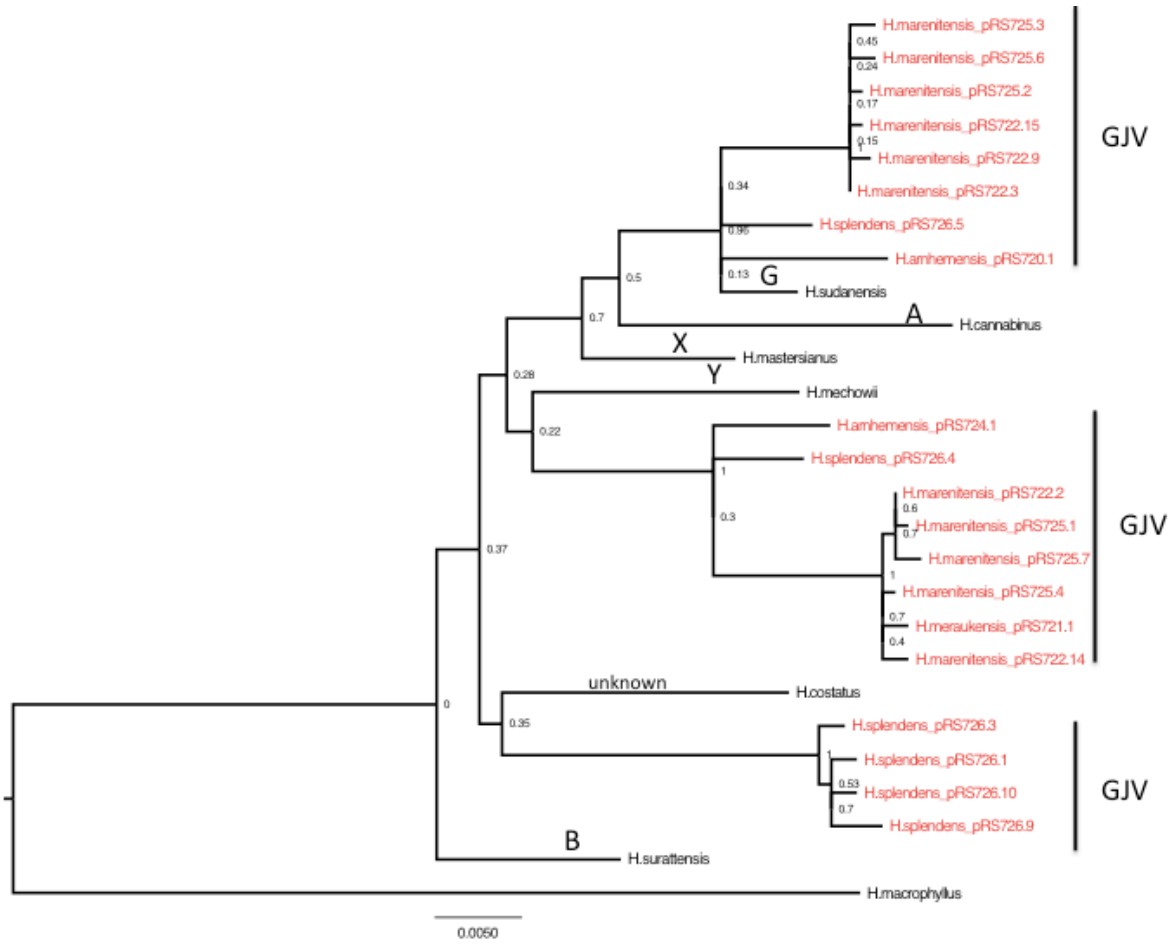


Fig.7 Maximum likelihood analysis of GBSSI region
The evolutionary history was inferred by using the Maximum Likelihood method based on the General Time Reversible mode. The tree with the highest log likelihood (-3788.5858) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 27 nucleotide sequences. There were a total of 1346 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. Bar represents average substitutions per site.

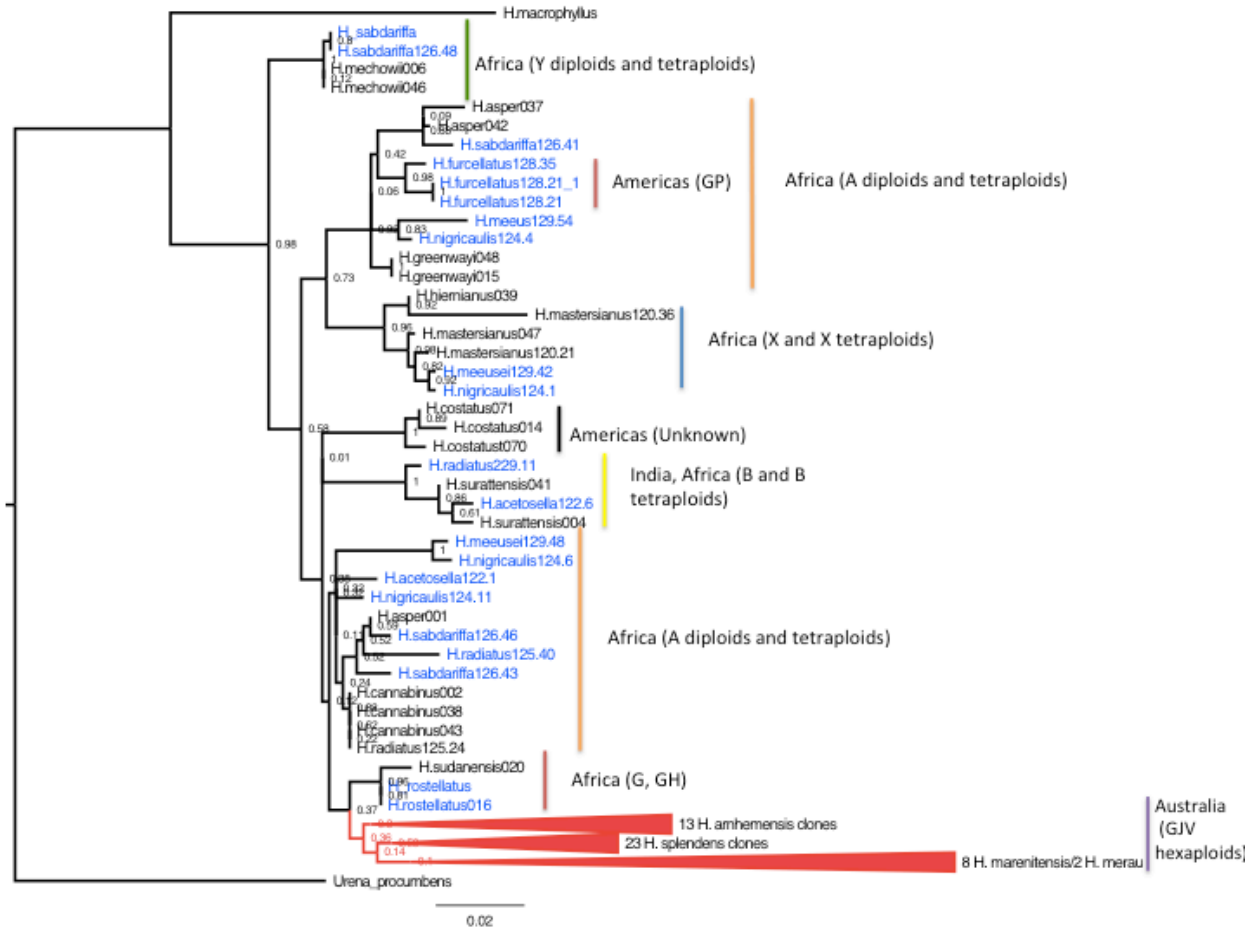


Fig.8 Maximum likelihood analysis of ITS region
The evolutionary history was inferred by using the Maximum Likelihood method based on the General Time Reversible model. The tree with the highest log likelihood (-5035.2136) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 95 nucleotide sequences. There were a total of 639 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. Bar represents average substitutions per site.

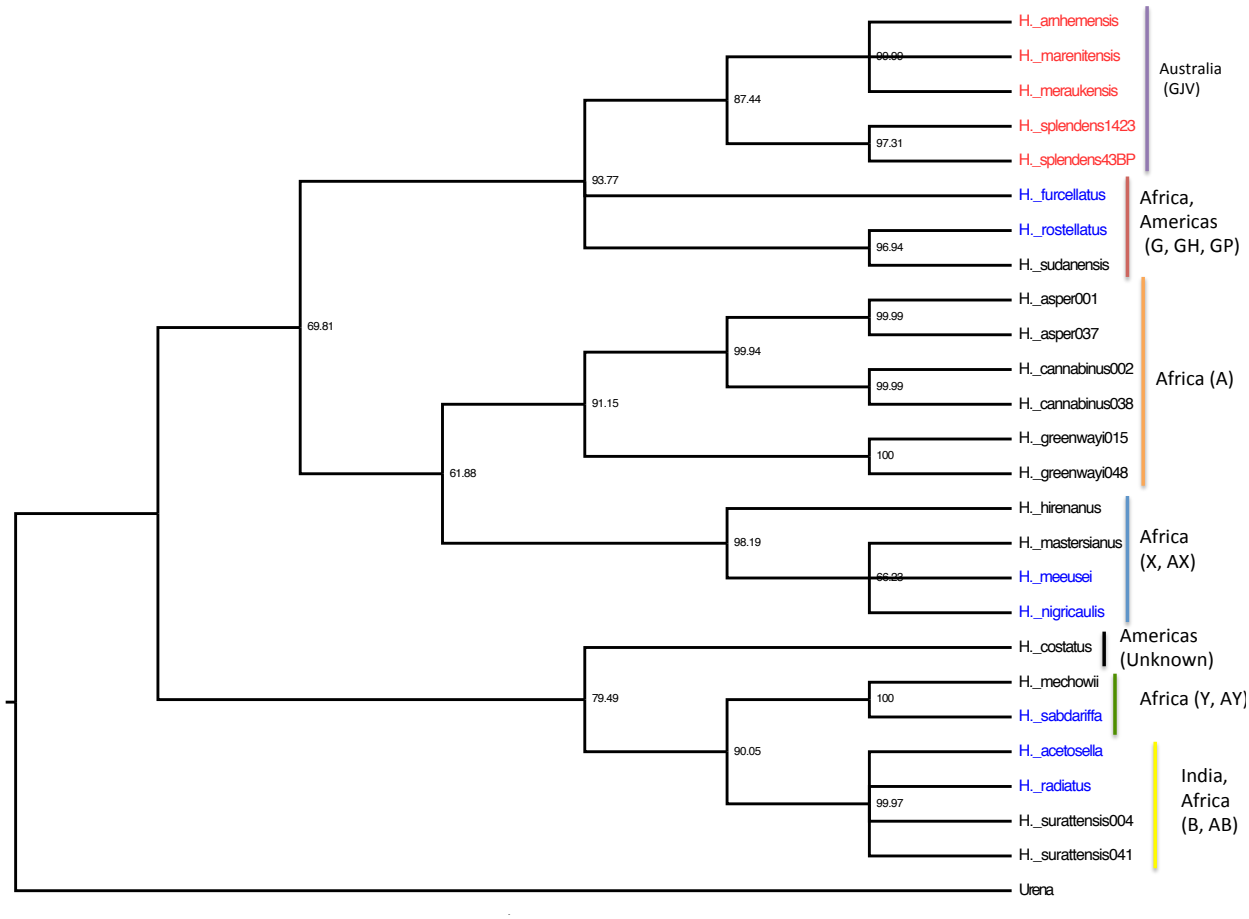


Fig. 9 Parsimony analysis of 4 cpDNA regions using "fast" stepwise addition with 10000 replicates and retaining groups with greater than 50% frequency. Performed in PAUP*4.0 (Swofford 2003).

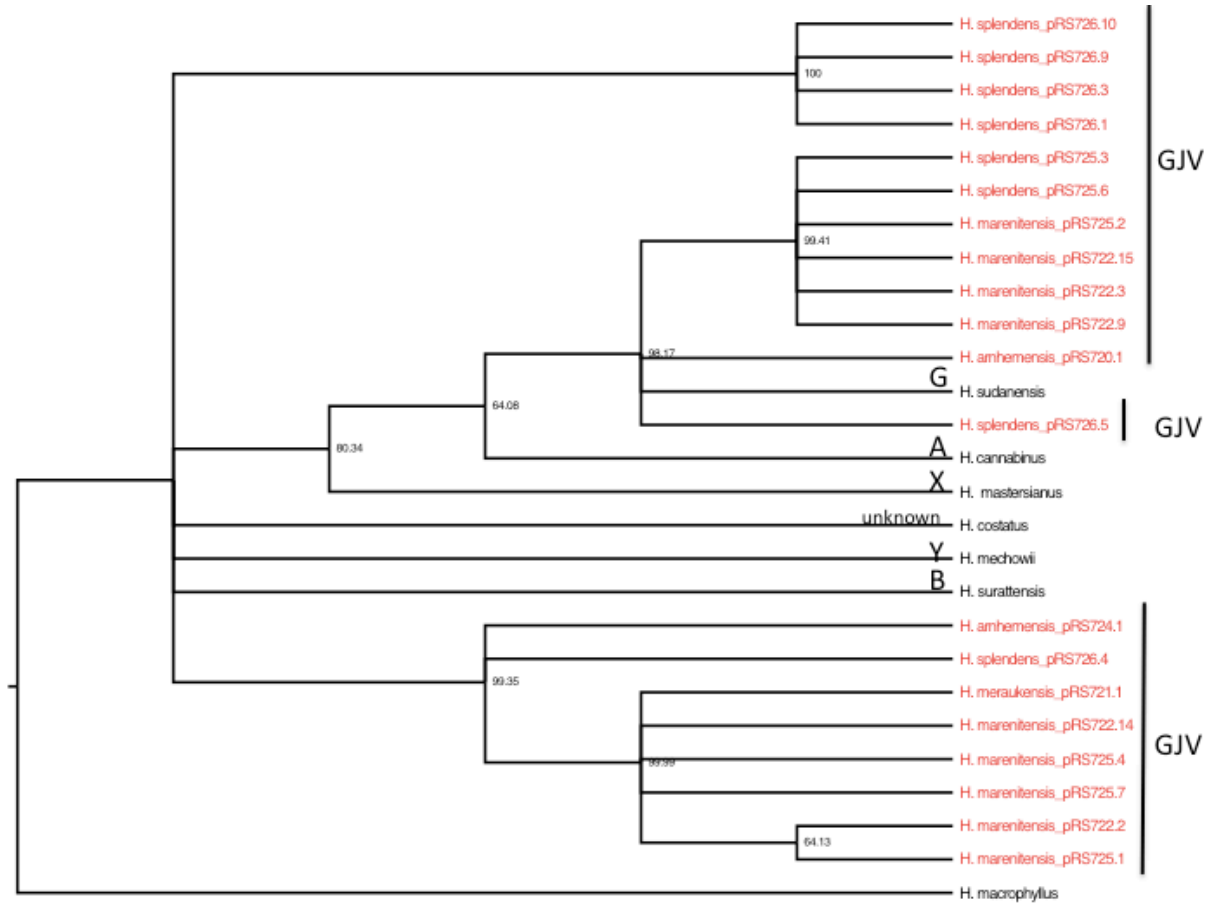


Fig. 10 Parsimony analysis of GBSSI region using "fast" stepwise addition with 10000 replicates and retaining groups with greater than 50% frequency. Performed in PAUP*4.0 (Swofford 2003).

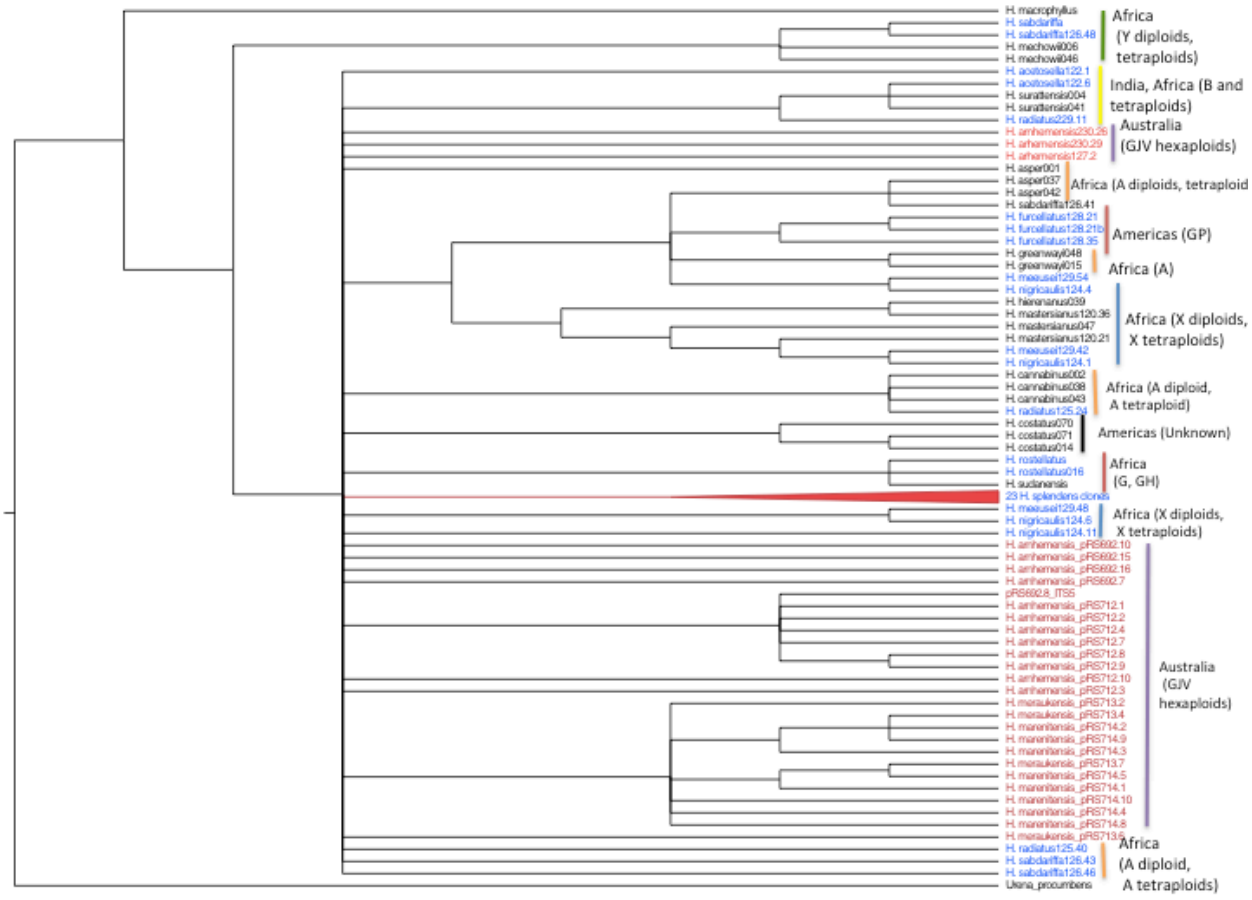


Fig. 11 Parsimony analysis of ITS region using "fast" stepwise addition with 10000 replicates and retaining groups with greater than 50% frequency. Performed in PAUP*4.0 (Swofford 2003).

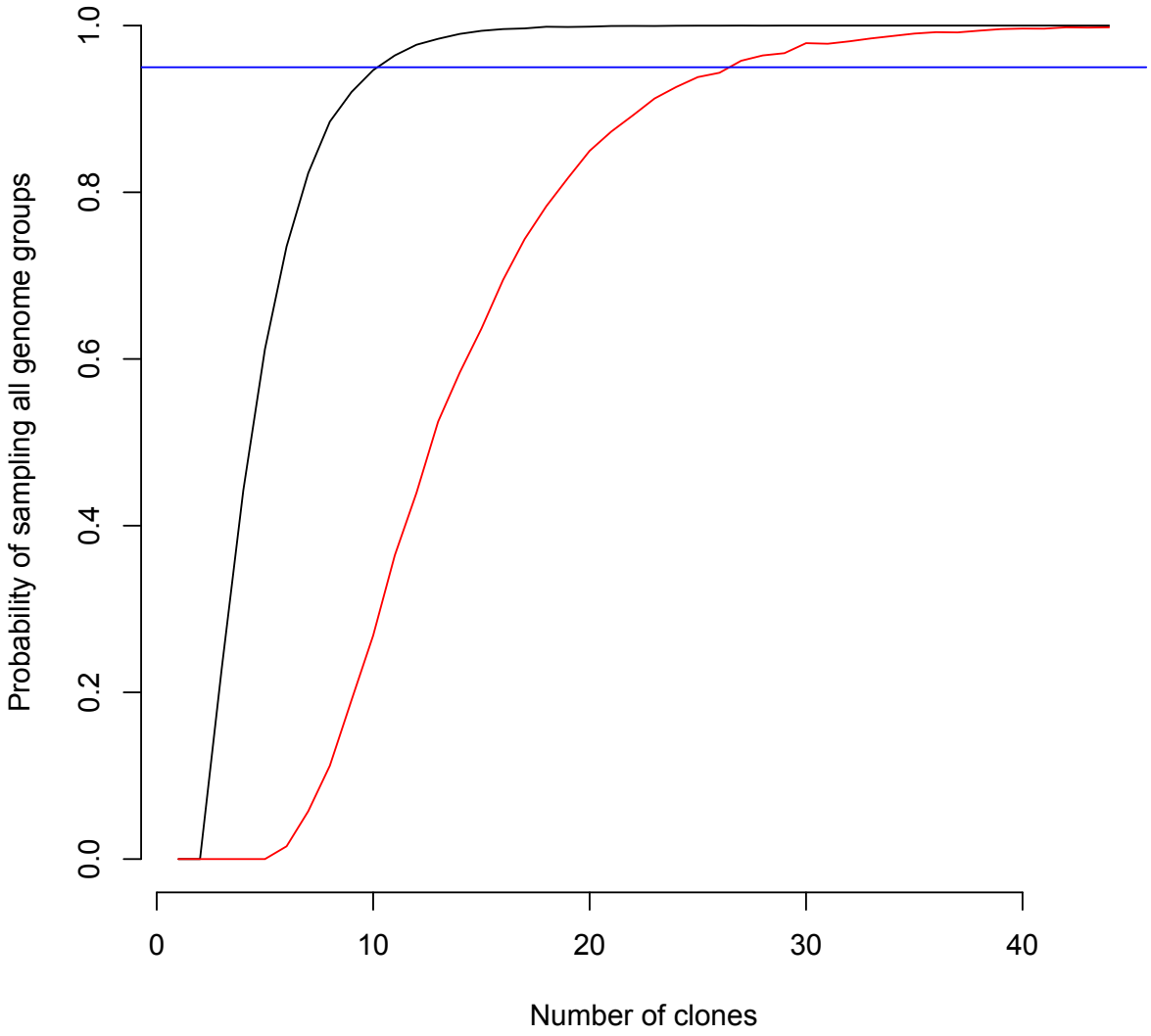


Fig. 12 Probability plot for sampling all genome groups in the Australian hexploids in *Hibiscus* sect. *Furcaria* as a function of the number of clones sampled.
This plot shows two lines representing the number of gene copies. The red line represents 6 copies while the black line represents 3 gene copies. The number 6 was chosen to represent each copy of a gene that is actually present in a hexaploid (GGJJVV) and 3 was chosen to represent each copy from one pair of homologous genome groups (G, J, or V) in this polyploid group. The blue line represents a 95% probability of finding all copies.

Vita

Whitaker Matthew Hoskins was born in Nashville, Tennesse to Barbara and Stephen Hoskins. He attended Granberry Elementary, Glendale Middle, McMurray Junior High, and Hume-Fogg Academic Magnet in Nashville. During his time at Hume-Fogg he was introduced to the world of biology, including but not limited to the concept of evolution, naturalist writings, genetics, and the human body by two wonderful teachers- Barbara Allen and Brenda Royal. After graduation he attended the University of Tennessee- Knoxville where he received a Bachelor's of Science degree in Ecology and Evolutionary Biology. During his time at the University of Tennessee he delved into mycology and participated in research with Dr. Karen Hughes and Dr. Brandon Matheny. Learning so much about the biological world inspired him to pursue his teaching degree so he began the Science Education program at UTK and participated in an internship where he taught Honors Biology and Physical Science at Farragut High School. Upon receiving his Master's in Science Education he worked at Siegel Middle School as a 7^{th} grade science teacher for 1 year. This was a challenging time for him as he learned much on classroom management. He then returned as a Master's degree student and dedicated teaching assistant in the Ecology and Evolutionary Biology Department and continues his love for learning and all living things under the advisement of Dr. Randall Small. He now studies the phylogenetic patterns of polyploid plant groups and hopes to go on to a biological research career at a university or government organization.