MOLECULAR PHYLOGENY OF THE SECTION SIMIOLUS OF THE GENUS *MIMULUS* (SCROPHULARIACEAE)

by

MADISON MACHT

A THESIS

Presented to the Department of Biology
and the Honors College of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Bachelor of Arts

June 1998

APPROVED: 

Dr. John Willis

An Abstract of the Thesis of

Madison Macht for the degree of Bachelor of Arts

in the Department of Biology to be taken June 1998

Title: MOLECULAR PHYLOGENY OF THE SECTION SIMIOLUS OF THE GENUS *MIMULUS* (SCROPHULARIACEAE)

Approved: [redacted]

Dr. John Willis

The evolution of self-fertilizing species from outcrossing ancestors appears to be one of the most common evolutionary transitions in flowering plants. This study seeks to answer fundamental questions on the evolution of self-fertilization within the section Simiolus of the genus *Mimulus* (Scrophulariaceae) by construction of a molecular phylogeny. How many times did self fertilization evolve? Are selfers independently derived or closely related? Are selfers of recent origin? Are selfers evolutionary dead ends, or the basis for speciation? Parsimony and neighbor-joining phylogenies constructed using the 538 bp of the internal transcribed spacer (ITS) region of 18S-26S nuclear ribosomal DNA (nrDNA) supports the existing section-level clades within the genus, but are unable to resolve the relationships between outcrossers and selfers within the Simiolus section. Phylogenies constructed using a 246 bp anonymous region of the genome containing a microsatellite differ from ITS phylogenies and are also unable to resolve outcrosser-selfer relationships. These results support the close relationship of outcrossers and selfers within this section and suggest new avenues for phylogenetic research.

ACKNOWLEDGMENTS

I would like to thank John Willis for his guidance and help throughout this project. I pay special thanks to his overflowing excitement, his patience with me as I learned, and his congratulations when I finished. I also thank Alan Kelly and Mike Pfrender who took time to teach me the principles of molecular biology, and how to work in a lab. I thank Tom Titus for teaching me the details of phylogenetic analysis. I also thank Mike Lynch and the Research Training Grant Program, as well as the Howard Hughes Medical Institute for providing the financial support for this project. I thank my mother, Mimi Macht, for sharing her love of hiking in the wildflowers, and my father, Will Macht, for his advice and his love. Finally, I thank my brother, Marlow, who showed me science and taught me so much more.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Today, we face more and more challenges whose solutions depend on an understanding of the biological sciences. Expanding population and increased consumption of energy, homes, highways and food cause rapid destruction of ecosystems. The solutions to malnutrition, diminishing resources, effects of pesticides, air and water pollution, declining biodiversity, AIDS, cancer, and heart disease have roots in biology. These problems will be solved with the help of biologically informed citizens.

Evolution is the field of biology that links all others. It describes the continual appearance of antibiotics and antibiotic-resistant bacteria. It explains how new viruses continually appear that cause diseases that ravage humanity. It predicts the long-term effects of pollution, determines which foods we eat, and determines what resources will be available in the future. Evolution explains how endangered species became endangered, how they interact with the rest of the natural world, and how they will fare in the future. For many, evolution explains the cause of the existence of humans, and all other creatures, on this planet.

In the 1850's, Charles Darwin challenged contemporary beliefs about our natural world by observing different species. The concept of a species is more abstract than you might think. Generally, a species is a group of organisms with similar structure and function that breed, in nature, only with each other. This concrete separation raises serious questions that will become central to this project.

So, as most scientists start, Darwin started with what he saw. First, he noticed that each species has the capacity to produce more offspring than will survive to maturity. Second, he noticed that individuals of the same species in a population have different traits. Third, he noticed that there is only so much food, water, light, growing space, and other resources available to a population, and therefore they must compete with each other for limited resources. Finally, he observed that different individuals had different reproductive success. Those that have the most favorable combination of characteristics are most likely to survive and reproduce.

Here Darwin suggested a model called natural selection. If each species produces more offspring that can survive, each individual in a species is slightly different, and the offspring compete for limited resources, than certain individuals will win in the struggle for existence, and certain individuals will lose. Those that win earn the right to pass on their characteristics. So, one can say that nature selects certain individuals to reproduce, because they are better fit to survive. In biological terms, these individuals are called the most fit.

Why do living creatures differ?

Though the reasons that one individual is slightly more fit than another would remain unknown for a century, Darwin proposed a continuous occurrence of variations. Most variations are disadvantageous to an organism and the organism dies before it can reproduce. Certain variations, however, make organisms more fit, and therefore are passed on to the offspring. These variations are further modified by more variations, and so the process proceeds until one organism looks quite different from another, and are called new species. The process by which new species evolve from ancestors is called speciation. This entire process that Darwin called "descent with modification" is known as evolution. We only have to look out the window at a forest of trees, pet our dogs, eat a stalk of broccoli, look in the mirror, and ponder the relationships between it all to realize the wonders of evolution.

The age of genes

Today we know the mechanism that causes the variation Darwin observed. The structure of an organism is determined by its genes, which are sections of deoxribonucleic acid more commonly known as DNA. DNA is like a long necklace made up of only four different colored beads. There are so many beads on this necklace that if they were removed they would fill the Rose Bowl stadium in Pasadena, California several times over. The particular order of the beads encodes a message for how to build living things, just like the order of the 26 letters on this page determines my message. Rearanging these beads (or letters) changes the message completely. These genes are handed down from parents to offspring, but not always in the same form. The genes can be altered by

random mutations, causing a change in the blueprint of the offspring. The unification of Darwin's principles and the principles of genetics resulted in neo-Darwinism, which forms the theory that mutation provides the genetic variability that selection acts upon during evolution.

How do we study evolution?

To study evolution we look at the relationships between organisms. To do this, we classify shared derived characteristics between these organisms. Because the basic structures in a common ancestor were modified in different ways as subsequent organisms evolved, we can trace these modifications back to their source. Organisms that share more characteristics are said to be more closely related in an evolutionary sense. Then, we are able to classify organisms into a hierarchical system. Organisms that share major characteristics — for example all plants — are classified together in a large group. However, finer divisions will separate plants into several smaller groups — perhaps flowering plants and non-flowering plants. Each grouping is called a taxon (pl.*taxa*), and the science of naming and classifying organisms is known as taxonomy. Examples of taxa (from fine grouping to broad grouping) are variety, species, genus, family, order, class, phylum, and kingdom. Taxonomy is essential to understanding, describing, and discussing the diversity of life. All the scientific names for organisms are based on their taxonomy. These names are the same all over the world, enabling researchers in Puerto Rico to communicate with researchers in Russia.

Modern classification is often referred to as systematics, because it is systematically based on evolutionary relationships. The branching evolutionary tree that we create is called a phylogeny (literally "production of phyla"). This diagram shows, for each taxa, with which other taxa it shares its most recent common ancestor.

How do you build a phylogeny?

There are several different methods of obtaining phylogenetically informative data. The traditional method is based on the morphology of taxa. Organisms that look more similar are grouped together. However, this is not always the best method. For example, crocodiles may look more like lizards, but they are more closely related to birds.

Crocodiles and lizards have analogous similarities, which are similar characteristics derived from different ancestors.

Recently molecular biology has provided important tools for taxonomists. Because evolution is determined by the mutations of genes, and mutations occur randomly throughout history, certain genes can be used as molecular clocks. In comparing these genes in different organisms we find that more closely related organisms have more similar sequences. Thus, the differences in DNA sequence provide the phylogenetic information. Just as examining ancient layers of rock provides information about the history of the earth, comparing DNA sequences of different organisms provides information about the history of a species. Work in this field makes up the discipline of molecular evolution.

How fast is the clock ticking?

Picking the correct molecular clock is not easy. First, one must find a clock that ticks at the correct rate. The DNA sequences of the organisms being compared must be significantly variable to show evolutionary divergences, but significantly similar to know the same sequences are being compared. The clock must fit in the proper window of time. The rate of molecular evolution is determined by selective pressure. Mutations in some DNA sequences are disadvantageous to the fitness of the organism. Therefore, these mutated sequences do not confer any reproductive advantage to the organism, and are not favored by natural selection. Thus, the functional sequences are under selective pressure to stay that way. These sequences are said to be conserved. However, mutations in most DNA do not effect the fitness of the organism. This is often called junk DNA, and is not under high selective pressure. The trick in constructing a phylogeny based on molecular evolution is to pick a molecular clock that ticks at the correct rate.

The other difficulty in using molecular clocks is aligning the different sequences of different species. Because the outcome of the phylogeny depends on the differences in sequences, an improper alignment would result in false differences. Alignment is made easier by comparing divergent sequences that are flanked by conserved sequences, because it gives us a reference point at which to start.

Why is all this important?

This type of work enables us to observe, explain, and classify the diversity we see in the biological world. This has important industrial and medical applications. It also provides groundwork for future comparative studies in the evolution of species – an issue that Darwin raised over a century ago that has not yet been answered. Finally, this type of research allows us to ask more basic questions about the details of evolution. Basic research is based on the love of knowledge – it does not have direct applications today, but it could help us in the future. Basic research satisfies our wonder and passion to discover the details of the natural world.

Plant breeding systems: Why do some plants mate with themselves?

One of these details of the natural world is an organism's method of reproduction, or breeding system. One of these methods of reproduction is call self-fertilization, or the fusion of male and female gametes from a single genetic individual. The evolution of self-fertilizing species from outcrossing (mating between different genetic individuals) ancestors appears to be one of the most common evolutionary transitions in flowering plants, and its consequences have concerned plant biologists for more than a century (Darwin 1876). Predominant selfing and predominant outcrossing are thought to be alternative stable states of a plant population's mating system suggesting selection against a population that does both (Maynard Smith 1978; Lloyd 1979; Charlesworth 1980; Feldman and Christiansen 1984, Lande and Schemske 1985a). Selfing is promoted by a more direct transmission of genes (Fisher 1941), reproductive assurance, preservation of adaptations to local environments, and reduced expenditure on copulation. The only factor that appears to be strong enough to prevent the evolution of selfers from outcrossing populations is inbreeding depression — the lowered fitness of inbred individuals compared to outcrossers (Charlesworth and Charlesworth 1979).

The level of inbreeding depression in a species, and thus its predominant mating system, depends on the level of recessive gene mutations that decrease an individual's fitness. Recessive gene mutations can be carried through an outcrossing population without being expressed if a functional copy of the gene is present. An organism is said to be heterozygous for a particular gene if it has a mutant copy and a functional copy.

Heterozygotes are more common in outcrossing populations. In selfing populations, two mutant copies of a gene tend to combine to make a recessive homozygote. These individuals tend to have reduced fitness, or die, and are thus purged from a selfing population. Populations in which only a select few of the fittest members survive or populations that lack sufficient pollinators tend to have little inbreeding depression and are selected for selfing. This is due to increased efficiency of selection against recessive sublethal and lethal mutations in inbred populations (Lande and Schemske 1985a; Charlesworth and Charlesworth 1987; Charlesworth, Morgan and Charlesworth 1990; Charlesworth *et al* 1991). Large outcrossing populations tend to have high inbreeding depression and are selected for outcrossing.

Genetic models predict that selfers evolved relatively recently from outcrossers, and that they rarely radiate into large numbers of selfing species (Stebins 1957; Lande and Schemske 1985a). Predominantly outcrossing populations are sensitive to pollinator failure and other population crashes that could directly increase the level of inbreeding, and select for a breakdown of self-incompatibility mechanisms. If this happened relatively often, deleterious recessive mutations would be purged from a population, lowering inbreeding depression and selecting for selfing. In contrast, it is thought that a population with a history of selfing and relatively low inbreeding depression can not readily revert to outcrossing. However, since self-fertilization is an effective barrier to gene flow, it is thought that highly selfing populations would promote speciation in some populations. This study seeks to answer fundamental questions on the evolution of self-fertilization. How many times did self-fertilization evolve? Are selfers independently derived or closely related? Are selfers of recent origin? Are selfers evolutionary dead ends, or the basis for speciation?

We test these evolutionary questions by constructing a phylogeny — the branching diagram that shows, for each species, with which other species it shares its most recent common ancestor. Unfortunately, a phylogeny based on visible characteristics would group all selfing species together, a false interpretation if selfers evolve often from outcrossers. A more appropriate phylogeny is based on a comparison of DNA sequences of different species. Evolutionary relationships are determined because more closely

related species have more similar DNA sequences. This technique offers a virtually unlimited source of discrete characters, which helps to resolve ambiguous evolutionary relationships. Unfortunately, there are few studies that use a molecular phylogeny to test these ideas.

Mimulus*: an ideal model system*

I chose wildflowers in the genus *Mimulus* (SCROPHULARIACEAE, Figwort family) as a model system to study these ideas. My research will concentrate on the section Simiolus, which includes about 20 species (a section is a finer division than genus, but broader than species). This plastic section contains a wide variety of outcrossing and selfing flowers arranged into unresolved species relationships. Oregon and California contain most of the section's species diversity, plants are easily collected, transplanted and grown in a greenhouse, and seeds are easily collected and stored. In addition, much is known about inbreeding depression and mating system evolution in this genus (Willis 1993, Fenster and Ritland 1994) and several projects are underway to study the evolution in different portions of the genus (Ritland et. al. 1993; A. Liston, pers. comm., P. Beardsley, pers.comm.). These factors make *Mimulus* an ideal model system.

Morphological evidence suggests the evolution of self-fertilization multiple times. The section Simiolus is characterized by the presence of two groups of closely related groups called species complexes. Each complex consists of a system of geographically well-isolated populations existing in moist habitats and lacking significant crossing barriers (Vickery, 1978). It is thought that members of these species complexes invade new locations as colonizing (Baker and Stebins, 1965) or fugitive (Hutchinson, 1957) species by seeds washed downstream, blown by the wind, or carried by small animals (Lindsay, 1964). Selfing is thought to have evolved multiple times among these complexes because self-compatibility greatly facilitates these methods of life (Antonovics, 1968).

These two species complexes in Simiolus are the *M. guttatus* and the *M. glabratus* complexes. The *M. guttatus* complex includes *M. guttatus*, *M. nasutus*, *M. laciniatus*, *M. platycalyx*, and *M. glauscens* (Vickery 1978). Distribution of this complex ranges from Southern Chihuahua to the Aleutian Islands and from the Pacific Coast to the Rocky

Mountains, but the center of diversity is California. Members of this complex are typically facultative self-pollinators except for *M. nasutus*, which may be cleistogamous or pseudocleistogamous. The base chromosome number of members of this complex is n=14 (Table 1) but Vickery suggests evolution by aneuploidy, M. nasutus (n=13), M. platycalyx (n=15) (1978). Kiang (1973) proposed *M. nasutus* is a derived daughter species of *M. guttatus*, and it is thought that the other species could have arisen in a similar way without such a pronounced shift to selfing (Vickery, 1978).

The *M. glabratus* complex includes *M. andicolus*, *M. glabratus* var. *externus*, *M. glabratus* var. *fremontii*, *M. glabratus* var. *glabratus*, *M. glabratus* var. *utahensis*, *M. glabratus* var. *parviflorus*, and *M. pilosiusculus*. This complex ranges from about 45° north latitude in the Quebec Province south through the Great Basin, Great Plains, and into the highlands of Chiapas and Guatemala. It does not occur elsewhere in Central America, but begins in northern South America and spreads throughout the continent. Members of this complex are facultative and frequent self-pollinators (Vickery, 1978). They tend to self-fertilize more frequently than members of the *M. guttatus* complex. Cytological evidence shows that these members of this complex form a north-to-south heteroploid series, n=15 → n=30 → n=31 → n=46 (Vickery, 1978). This series is important for two reasons: it opposes Häegerup's hypothesis (Löve and Löve, 1957) of increasing polyploidy with increasing climatic rigor from the equator to the north pole, and it suggests that polyploidy is an evolutionary pattern within groups.

Recent evidence has shown that the internal transcribed spacer (ITS) region of 18S-26S nuclear ribosomal DNA (nrDNA) is a useful source of characters for phylogenetic studies in flowering plant families (Baldwin et al 1995). The ITS is a section of DNA that is easily amplified, sequenced, and compared, and has been shown to be significantly variable for species-level comparisons.

Alternative methods for obtaining molecular characters for phylogenetic analysis include using anonymous single copy sequences (Karl and Avise 1993; Karl, Bowen and Avise 1992) and microsatellite DNA. Anonymous loci provide a large number of independent loci but nothing is known about the sequences produced which makes homology identification and evolutionary modeling difficult. Furthermore, PCR primers

may not bind equally in different species. The number of tandem polynucleotide repeats in microsatellite DNA tends to vary from individual to individual due to unequal crossing-over and slip mismatch replication. This enhances the chance of non-parsimonious or parallel evolution of a particular length variant and can lead to misleading phylogenetic results.

MATERIALS AND METHODS

Taxa. Ten species of Mimulus in the section SIMIOLUS and one from the section PARADANTHUS were selected for the ingroup (the taxa of primary interest) of this analysis (Table 1). Ingroup species were selected to include selfers, partial selfers, and outcrossers, and presumed close relatives and distant relatives within and outside of species complexes. M. floribundus was used as the outgroup (a taxa assumed to phylogenetically outside the ingroup and used to root a phylogeny) species because of its presumed ancestral role to Simiolus (Grant 1924) and its morphological variation. ITS sequence data from a sample of M. floribundus from Moll Reservoir, Lassen County, CA was obtained from Justen Whitthall at Oregon State University in Corvallis, OR.

Two species samples, *M. dentatus* and *M. Tilingi*, were obtained from sites in the wild by removing the entire plant with attached soil and placing it in a moist plastic bag. *M. dentatus* was obtained near dark old-growth areas on the edges of small tributaries of the Smith River in Douglas County, Oregon, in the Suislaw National Forest. The sample of M. Tilingi was obtained from the Upper Strawberry Lake in Grant County, Oregon. Both samples were collected and identified by Geoff Smick. The sample of *M. guttatus* was a member of a large population obtained from Iron Mountain in the Western Oregon Cascades by Dr. John Willis during previous experiments (Willis 1993). Dr. Willis also obtained the sample of M. nasutus from a large population near Maupin in Deschutes County, Oregon. The sample of *M. nasutus* had been inbred for 5 generations. Plants were set on metal benches in the greenhouse facilities at the University of Oregon.

All plants were grown in 4in plastic pots with Grower's Gold potting soil and hand watered from above once a day during most days and twice a day during hot, dry, summer days. The biological pesticide Gnatrol was included in water at directed concentrations every other watering session to reduce gnats. Artificial sodium lights were used to keep light on plants for 12 hours.

The remaining members of the ingroup were grown from seed obtained from Dr. Robert Vickery and Dr. David Carr. Seeds were germinated under a mist bench with natural light in 4cm square pots filled with Grower's Gold potting soil. This resulted in high rates of germination (>85%) for all species except *M. laciniatus* which had to be

removed from the ingroup due to lack of a sample. After seedlings had developed cotyledons about 2cm long they were transplanted into 4in pots and treated just as the plants obtained from the field.

TABLE 1: Summary of Ingroup Species: Species were chosen to include members of both species complexes, other members of Simiolus, and members of Paradanthus. The Vickery number identifies the plant was obtained from Dr. Vickery and is in reference to Vickery (1978). *M. micranthus* seed was furnished courtesy of Dr. David Carr, and *M. laciniatus* was not used for future analysis due to lack of germination.

2n	Vick. #	Sect.	Spec.Cmplx	Species	Location
16, 28, 30, 56	-	Sim.	*M. guttatus*	*M. guttatus*	Iron Mountain, OR
-	-	Par.	-	*M. dentatus*	Siuslaw Nat. Forest, Douglas Co., OR
32	13007	Sim.	-	*M. dentilobus*	Camp Creek, AZ
92	5041	Sim.	*M. glabratus*	*M. glabratus* var. *parviflorus*	Illapel, Chile
62	6194	Sim.	*M. glabratus*	*M. glabratus* var. *glabratus*	Veracruz, Mexico
28, 30, 56	-	Sim.	*M. guttatus*	*M. Tilingi*	Upper Strawberry Lake, OR
-	-	Sim.	-	*M. micranthus*	Dos Rios, California, Mendoncino Co.
26, 28	-	Sim.	*M. guttatus*	*M. nasutus*	Shear's Falls, Oregon
30	5752	Sim.	*M. guttatus*	*M. platycalyx*	Crystal Lakes Reservoir, San Mateo Co., CA
60, 62, 64	-	Sim.	*M. guttatus*	*M. luteus*	Illapel, Chile
28	-	Sim.	*M. guttatus*	*M. laciniatus*	*Did not germinate*

Isolation of genomic DNA from plant samples. In order to analyze a plant's DNA, the DNA must first be isolated. Genomic DNA (the complete set of DNA in an organism) was obtained from corolla tissue following the protocols outlined by Lin, J.-Z. and Ritland, K. (1995). Tissue from different individuals of the same species was not mixed. Up to 150 mg of corolla tissue was placed into a 1.5 ml microcentrifuge tube and frozen with dry ice. A small amount of liquid nitrogen was added to the tube, allowed to evaporate, and then the tissue was quickly ground to a fine powder for about 3 minutes using a small plastic pestle mounted to a Fischer Scientific StedFast Stirrer (Model SL

1200). 25 µL of CTAB buffer (100mM Tris, pH 8.0, 1.4M NaCl, 20mM EDTA, 2%CTAB) was added until a paste forms and then 400 µL of CTAB was added with 1 µL of β-mercaptoethanol to break disulfide bridges. The tube was incubated in a 60° C water bath for 20 min and gently inverted during the incubation. Next, the tube was cooled to room temperature and inverted after adding 500 µL of chloroform. Tubes were then centrifuged for 5 min at 13200 rpm. The aqueous phase was transferred to a new 1.5 mL microcentrifuge tube and 300 µL cold isopropanol was added before inverting several times and centrifuged again for 5 min at 13200 rpm. After centrifuging the supernatant was poured off, the remaining pellet was rinsed briefly in 70 % ethanol, and then allowed to dry. The pellet was resuspended in 50 µL TE, and DNA concentration was determined using a Hoechst 33258 dye DNA fluorometer (Hoefer Scientific). Genomic DNA samples were then labeled and stored in a -20°C freezer in 1.5 mL plastic microcentrifuge tubes.

Amplification of ITS fragments. ITS regions were amplified by the polymerase chain reaction (PCR) using 25S-25R (5'TATGCTTAAACTCAGCGGG3') (Nickrent et al. 1994) and ITS-5* (5'GGAAGGAGAAGTCGTAACAA3') (Liston et al. 1996). Primers were obtained from the Dr. Aaron Liston laboratory at the Department of Botany at Oregon State University in Corvallis, Oregon. Melting points of these primers were determined with the OLIGO 4.0 program for the Macintosh (Tm = 62.3°C, 58.5°C; delta-G(25°C) = -38.0 kcal/mol, -35.4 kcal/mol respectively). Negative controls were used to avoid contamination. Reactions were 100uL, and included 70uL ddH2O, 10uL 10x Buffer (Promega), 8uL MgCl2 (25mM stock), 4uL dNTPs (2.5mM stock), 2uL of each primer, 2uL BSA (10mg/mL), 2uL DMSO, 30 ng template, and 1.5uL Taq Polymerase (Promega). PCR conditions were:

Step 1	94° C	1 min
Step 2	94° C	45 sec
Step 3	55° C	30 sec
Step 4	72° C	90 sec
Step 5	Go to Step 2	39 times
Step 6	72° C	5 min
Step 7	4° C	30 sec
Step 8	END	

5uL of each sample was assayed for amplification and length by electrophoresis on a 1% agarose gel stained with ethidium bromide using 1x TBE as the gel buffer. Length was determined by comparison to a known 100 base pair ladder. The remaining 95uL of successful samples was purified by QIAquick Gel Extraction (Qiagen) following the protocols outlined in the instruction manual for gel extraction using a microcentrifuge. Purified ITS fragments were eluted in 30 µL buffer EB to increase concentration. Concentration was checked by the same DNA fluorimeter, and samples were labeled and stored in plastic 1.5 mL microcentrifuge tubes in a -20°C freezer. After use, samples were transferred to a -80°C freezer for long term storage.

Purification of anonymous sequences from plasmid library

Bacterial infection. Petri plates were prepared by first making 200 µL LB bacterial medium in a 500 µL Erlenmeyer flask and sterilized in a slow exhaust Autoclave for 15 min. LB medium contains: 2g tryptone, 1g yeast extract, 2g NaCl, 3g Agar in distilled deionized water (ddH20). The medium was cooled to 50°C in a water bath, and 160 µL IPTG (the lacZ inducer) and 160 µL X-Gal (blue indicator precursor) were added. Petri plates were filled with 25 mL LB medium using sterile technique and allowed to sit overnight.

The XL1-Blue MRF' strain of *E. coli* was applied to the plates with a sterilized metal loop and colonies were allowed to grow for 24 hours in a 37°C room. Cultures were prepared from these colonies in 25 mL bacteria tubes in 3 mL of pre-sterilized LB broth for 6-7 hours. When ready to plate, 100 µL top agar was prepared by adding the following to ddH20: 1 g tryptone, 0.5 g yeast extract, 1 g NaCl, 1.5 g agar, and 0.65 g agarose. The top agar mixture was allowed to sit in a 50°-55°C water bath until ready to pour.

We used a pre-prepared plasmid library of *M. nasutus* genomic fragments (200-600 bp) as the source of anonymous sequences. Four 200 µL aliquots of the bacterial culture were incubated at 37°C for 15 min in 15 mL glass bacteria tubes with varying amounts (2 µL, 1 µL, 1/10 µL, 1/100 µL) of this S5.2 lambda ZAPII library (>8000 pfu/µL). To get the two most dilute concentrations the library was diluted with SM buffer. 100 µL of this

buffer contains 100 µL NaCl, 0.2g MgSO4 x 7H20, 5 mL of 1M Tris pH 7.5, and 0.01g gelatin.

Plating the phage-infected cells was done by pipetting 3 µL warm top agarose into the 15 mL tubes containing the cells and then quickly pouring the contents on a pre-labeled plate. Plates were incubated overnight at 37°C.

Phagemid production. Phagemids were produced by first excising twelve plaques from the agar plates and transferring them into a sterile microfuge tube containing 200 µL SM buffer and 10 µL chloroform. These tubes were incubated for 2-3 hours at room temperature and labeled Anon1-12. Next, 40 µL of this mixture was incubated for 15 min at 37°C in a 5 mL glass bacteria tube with 100 µL XL1-Blue cells and 1 µL of R408 helper phage (>1x10^6 pfu/ml). After this initial incubation 2 mL of LB broth was added and the mixture was allowed to incubate for 3 hours at 37°C with shaking. Tubes were then heated at 70°C for 20 minutes to kill the bacteria, transferred to a 1 mL microfuge tube and centrifuged at 13000 rpm for 5 min. The supernatant containing the phagemids were decanted to a new, sterile 1 mL microfuge tube, labeled Anon1-12, and placed in a -20°C freezer for short-term storage.

Plasmid rescue. In a 5 mL glass tube 5 µL phagemid stock was combined with 100 µL XL1-Blue MRF' cells and allowed to incubate for 15 min at 37°C. This was diluted to 1 mL with LB broth and 10 µL was mixed with 50 µL LB broth and spread onto LB plates containing ampicillin (100 µg/mL) and allowed to incubate overnight. DNA from the 12 different plaques was purified following the Promega Wizard Plus protocols, eluting in 50 µL ddH20, and labeled Anon1-12. DNA concentrations were checked with a fluorometer.

Amplification of anonymous fragments. Six of the 12 samples were chosen to be sequenced using plasmid primers, and then forward and reverse PCR primers were designed for 5 of these anonymous sequences using the PRIMER 0.5 program for the Macintosh. The two primers used for the data in this study were Anon11F (5'-CAACCACCGTGCCAAATTC-3') and Anon11R (5'-CCGCCGGAGCCCGGTCC-3'). Primer design was done to optimize fragment length, primer stability, and primer function. All Anon primers were ordered from Operon Technologies, in., Alameda,

California. These primers were used in 75mL PCR reactions containing 54 µL H20, 7.5 µL 10x buffer (Promega), 1.0 µL dNTP's (2.5mM stock), 3 µL of each primer (5µM stock), 6 µL MgCl2 (25mM stock), 0.5 µL Taq polymerase, and 20 ng template. Initially, five ingroup species were used to test the efficacy of these primers in *M. nasutus*, and other species. PCR conditions were:

Step 1	94° C	90 sec
Step 2	94° C	40 sec
Step 3	51° C	60 sec
Step 4	74° C	40 sec
Step 5	Go to Step 2	29 times
Step 6	74° C	2 min
Step 7	4° C	30 sec
Step 8	END	

Two fragments, Anon 10 and Anon 11 were assayed on a 1% agarose gel by electrophoresis using 1x TBE as the gel buffer and purified using the QIAquick Gel Extraction (Qiagen) following the protocols outlined in the instruction manual for gel extraction using a microcentrifuge. These fragments were sequenced and examined for homology between the five species. The Anon10 fragment showed little or no variation between species (data not shown) and was not pursued due to lack of phylogenetic information. The Anon11 fragment showed considerable species variation and was PCR amplified, purified, and sequenced for the remaining members of the ingroup, and for the outgroup, *M. floribundus.*

DNA Sequencing. Automated fluorescent sequencing was done by the DNA sequencing facility at the University of Oregon Institute of Molecular Biology with an ABI Prism model 377, version 2.1.1 (Perkin-Elmer). Sequencing was done for 25 cycles under the following conditions:

Denaturing	96° C	10 sec
Annealing	50° C	5 sec
Polymerization	60° C	4 min

Due to high G-C content and the formation of a hairpin-loop about 80 bp from the 5' end of ITS 1, glycerol (5%) and/or DMSO (5-15%) were used in the reaction mixtures to reduce this secondary structure formation. These procedures did not result in a sequence read past this hairpin-loop. When appropriate, unknown base calls (N) in a sequence read

in one direction were manually replaced with the known base call at that position from a sequence read from the other direction.

Sequence Alignment. Sequences were tested for homology with ITS regions from other species by a BLASTN 1.4.11 search (Altschul et al, 1990). Sequences were then aligned using MALIGN v.1.99 (Wheeler and Gladstein), a parsimony-based multiple sequence alignment program. Specifying *M. floribundus* as the outgroup for the alignments was the only phylogenetically-based weighting of the alignment. Alignments were determined for gap cost / transversion cost ratios of one through seven to account for various insertion and deletion (indel) possibilities. Base ambiguities were treated as missing data to prevent phylogenetic weighting of these characters.

All seven different gap-cost alignments were used in tree construction, but each alignment yielded similar results in topology, branch lengths, and bootstrap values (data now shown). For both data sets we used alignments with gap cost / transversion ratios of five based on an arbitrary decision and alignment checks by eye.

Tree construction. Phylogenetic trees were constructed using two optimality criteria: Fitch parsimony (assuming unordered character states) using PAUP 3.1.1 (Swofford, 1993) and neighbor joining using MEGA 1.0 (Sudhir, Tamura, and Nei, 1993). Different optimality criteria were used for purposes of comparison. *M. floribundus* was specified as the outgroup in all analyses.

Parsimony analyses: Trees were built using a simple heuristic search with one tree held at each step during stepwise addition. Tree-bisection-reconnection (TBR) branch-swapping was performed with the MULPARS option in effect. All characters were given equal weights. The steepest descent option was not in effect, branches having maximum lengths of zero were collapsed to yield polytomies, and no topological constraints were enforced during initial tree construction. Sets of equally parsimonious trees were summarized using strict consensus. Bootstrap analysis was performed with 100 replicates of heuristic searches using 50 random addition sequences to estimate the relative robustness (confidence limits) of individual clades. Comparisons of the two data sets (ITS and anonymous sequences) were performed by using the "load constraints"

option. The consensus tree for one data set was used as a topological constraint during tree reconstruction using the other data set.

Treatment of indels: Indels can be treated differently in a parsimony analysis depending on how they are coded. They can be treated as missing data or as a fifth character state. The latter method has the advantage of retaining the putative phylogenetic information in these insertion and deletion events, while the former method places more weight on the substitutions and increases ambiguities. In both the ITS and the anonymous sequence data sets we feel that these gaps provide important information for this analysis, and all gaps were treated as a fifth character state. Ambiguous nucleotide positions (N or ?) were treated as missing data.

Neighbor joining analyses: Sequences were treated as non protein-coding data because the phylogenetically informative sites lie in the ITS regions, not the 5.8S rRNA. Both transitions and transversions were used due to a limited number of useful sites. Trees and distance matrices were constructed with the Jukes-Cantor method of distance estimation and pairwise deletion of gaps. This means that gaps are not counted in the analysis when they occur between any pairwise comparison of taxa. Bootstrap analysis was performed with 500 replicates.

Results

DNA Sequencing and Alignment. ITS fragments were approximately 650 base pairs, consistent with lengths of other angiosperms (Baldwin et al., 1995). However, only 550 base pairs were read when sequenced from the 3' end, and only 120 base pairs were read when sequenced from the 5' end. These results, and the putative model of ITS secondary structure suggest the formation of a 15-20 base pair Watson-Crick complementary base pair hairpin about 80 base pairs downstream from the 3' end of ITS 1. Sequencing through these regions is a common problem (Baldwin et al., 1995), and neither the addition of DMSO or glycerol enabled us to obtain a continuous sequence read. To maintain the confidence of the sequence reads ITS 1 sequences to the 5' end of the hairpin were not used, although crude alignments suggest the ITS 1 to be about 240 base pairs.

Sequences of the ITS 1, ITS2 and the nrDNA coding regions found closest homology in the BLAST search with the same nrDNA of *Clerodendrum aculeatum*. Sequences were aligned with these regions indicated (Figure 1). The portion of the ITS 1 that was used was 153 characters long, had 45 variable sites (29.4%), 5 of which were parsimony informative (3.3%). One or more gaps were needed at 11 sites (7.2%). The 5.8S nrDNA was 163 characters long, had 24 variable sites (14.7%), 1 of which was parsimony informative (0.6%). No gaps were needed in this section to align the sequences. The complete ITS 2 was 218 characters long, had 21 variable sites (9.7%), 11 of which were parsimony informative (5.0%). One or more gaps were needed at 7 sites (3.2%) to align the ITS2. Four sites of the 28S nrDNA were included for alignment purposes, and all four were conserved among taxa. Overall, the portion of the sequences used in the alignments ranged from 526 – 538 base pairs, there were 538 character sites including gaps, 90 variable sites (16.7%), 25 parsimony informative sites (4.6%) and a G-C content of 62%. (Table 2). Sequence divergence for this data set ranged from 0.78 to 12.0%.

Anonymous sequences ranged from 230 — 246 base pairs (Table 2), and contained a 50 base pair sequence of only A and T bases thought to be a microsatellite. Out of 246 character sites, 86 were variable across the taxa (35.0%), 30 were parsimony informative

FIGURE 1: Aligned nucleotide sequences of partial ITS 1, complete 5.8S nrDNA, and complete ITS 2 for ingroup and outgroup taxa (Table 1). Sequences begin with the 5' end and vertical columns represent the 538 nucleotide positions. Ambiguous sites are represented by "N" (see Table 2 for ambiguous site frequency) and are treated as missing data, while gaps inserted by the alignment program are indicated by "-" and treated as phylogenetically useful information.

```
                                          ITS 1
M. micranthus  ---CCCACCCCNGCCGATGCGCACGAAATTGCGCCTCGTGCGGACTAACGTACCCCGGCGCGGAATGCGCCAAGGAAAACTCAACGAAGCGCCCGCCCCC
M. dentatus    CTGCCCCTCCCGGCCGATGCGCACGAAACTGCGCCTCGTGCGGACTAACGTACCCCGGCGCGGAATGCGCCAAGGAAAACTCAACGAAGCGTCCG-CCCC
M. glabratus g CGGACCCAACCGGCGGATGCGCACGAAACTGCGCCTCGTGCGGACTAACGTA-CCCGGCGCGGAATGCGCCAAGGAAAACTCAACGAAGCGCCCGCCCCC
M. dentilobus  CCGACCCTCCCCGCCGACGCGCACGAAACTGCGCCTCGTGCGGACTAACGTACCCCGGCGCGGAATGCGCCAAGGAAAACTCAACGAAGCGTCCGCCCCC
M. platycalyx  CCGACCCAACCCGCCGATGCGCAGGAAATTG-GCCTTGTTCGGACTAACGTACCCCGG-GCGGAATGGGCCAAGGAAAACTCAACGAAGGGTCCGCCCCC
M. glabratus p CCGACCCAAACCGCCGATGCGCACGAAACTGCGCCTCGTGCGGACTAACGTACCCCGGCGCGGAATGCGCCAAGGAAAACTCAACGAAGCGTCCGCCCCC
M. Tilingi     CCGACCCAACCGGTCAATGCGCACAAAATTGCGCCTCGTGCGGACTAAGGTACCCCGGCCCGAAATGCGCCAAGGAAAACTCAACAAAGGGTCCGCCCCC
M. guttatus    CCGACCCAACCCGCCGATGCGCACGAAACTGCGCCTCNTGCGGACTAACGTACNCCGGCGCGGAATGCGCCAAGGAAAACTCNNCGAAGNGCCCGCCCCC
M. luteus      -CCCACCAACCCGCCGATGCGCACGAAACTGCGCCTCGTGCGGACTAACGTACCCCGGCGCGGAATGCGCCAAGGAAAACTCAACGAAGCGTCCGCCCCC
M. nasutus     CNGACCCAACCCGCCGATGCGCACGAAACTGCGCCTCGTGCGGACTAACGTACCCCGGCGCGGAATGCGCCAAGGAAAACTCAACGAAGCGCCCGCCCCC
M. floribundus ------CTCCCCGCCGATGCGCACGAAACTGCGCACCGTGCGGACTAACGTACCCCGGCGCGGAACGCGCCAAGGAAAACTCAACGAAGCGTCCG-CCCC

                                                                                 5.8 S nrDNA
M. micranthus  CCGTTGCCTCGTACGCGACGTGCGCGGGGGTGCGAGACGTGTCTTGAATGTCAAAACGACTCTCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAA
M. dentatus    CCGTTGCCTCGTTCGCGACGTGCGCGGGGGTGC-AGACGTGTCTTGAATGTCAAAACGACTCTCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAA
M. glabratus g CCGTTGCCTCGTACGCGACGTGCGCGGGGGTGCGAGACGTGTCTTGAATGTCAAAACGACTCTCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAA
M. dentilobus  CCGTTGCCTCGTTCGCGACGTGCGCGGGGGTGCGAGACGTGTCTTGAATGTCAAAACGACTCTCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAA
M. platycalyx  CCGTTGCCTTGTTTCCCANNTGNCCGGGGGAACCAAAAGTGTTTTGAANGTCAAAACGACTTTTGGCAACGGANATTTTGGNTTTTGNATTGATTAAGAA
M. glabratus p CCGTTGCCTCGTTCGCGACGTGCGCGGGGGTGCGAGACGTGTCTTGAATGTCAAAACGACTCTCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAA
M. Tilingi     CCGTTGCCTGGTTCCCAATGTGCCCGGGGAACCAAAACGTGTCTTAAATGTCAAAACGACTTTGGGCAACGAATATCTGGGCTCTGGCATCNATNAAGAA
M. guttatus    CCGTTGCCTCGTACACGACGTGCGCGGGGGTGCGAGACGTGTCTTGAATGTCAAAACGACTCTCGGCANCGGATATNTCGGCTCTCGCATNGATNAAGAA
M. luteus      CCGTTGCCTCGTTCGCGACGTGCGCGGGGGTGCGAGACGTGTCTTGAATGTCAAAACGACTCTCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAA
M. nasutus     CCGTTGCCTCGTACGCGACGTGCGCGGGGGTGCGAGACGTGTCTTGAATGTCAAAACGACTCTCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAA
M. floribundus CCGTTGCCTCGTTCGCGACGTGCGCGGGGGTGCCGGACGTGTCTTGAATTTCAAAACGACTCTCGGCAACGGATATCTCGGCTCTCGCATCGATGAAGAA

                                          5.8 S nrDNA
M. micranthus  CGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACTAGGCCGAGGGCACGTC
M. dentatus    CGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACTAGGCCGAGGGCACGTC
M. glabratus g CGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACTAGGCCGAGGGAACGTC
M. dentilobus  CGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCCAGAAGCCACTAGGCCGAGGGCACGTC
M. platycalyx  CGNAGCGAAANNCNAAACTTGGNGTNAANTGCANAATNCCGTNAACCNTTGAGTNTTTNAACGCAAGTTGGGCCCGAAGCCAATAGGCCGAGGGCAAGTT
M. glabratus p CGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACTAGGCCGAGGGCACGTC
M. Tilingi     NGTAGCAAAATGCAATACTTGGTGTNAATTGCANAATCCCGTGAACCATCAAGTCTTTGAACNCAAGTTGCGCCCNAAGCCANTAGCCCAAGGGCAGGTC
M. guttatus    CGTAGNGAAATNCGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATNGAGTCTTTGAACGCAAGTTGCGCCCNAAGCCACTAGACCNAGNGCACGTN
M. luteus      CGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACTAGGCCGAGGGCACGTC
M. nasutus     CGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACTAGGCCGAGGGCACGTC
M. floribundus CGTAGCGAAATGCGATACTTGGTGTGAATTGCAGAATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCACTAGGCCGAGGGCACGTC
```

FIGURE 1: (CONT.)

```
                                                ][ITS 2
M. micranthus  TGCCTGGGCGTCACGCATTGCGTCG-CCCCCCCCTCTCGCTCCGTCGGGGCGGGATCGGGGGGCGGATATTGGTCTCCCGTGCGCTCGCGCGCGCGGCTG
M. dentatus    TGCCTGGGCGTCACGCATTGCGTCG---CCCCCTTCTCGCTCCATCGGGGCGGGATC-GGGCGCGGATANTGGTCTCCCGTGCGCTCGCGCGCGCGGCTG
M. glabratus g TGCCTGGGCGTCACGCATTGCGTCG--CCCCCCCTCTCGCTCCGTCGGGGCGGGATCGGGGGGCGGATATTGGTCTCCCGTGCGCTCGCGCGCGCGGCTG
M. dentilobus  TGCCTGGGCGTCACGCATTGCGTCG--CCCCCCCTCTCGCTCCGTCGGGGCGGGATC-GGGGGCGGATATTGGTCTCCCGTGCGCTCGCGCGCGCGGCTG
M. platycalyx  TGCCTGGGGGTNAAGCAATGGGTTGCCCCCCCCCCTCTCGCTCCGTCGGGGCGGGATCGGGGGGCGGATATTGGTCTCCCGTGCGCTCGCGCGCGCGGCTG
M. glabratus p TGCCTGGGCGTCACGCATTGCGTCG-CCCCCCCCTCTCGCTCCGTCGGGGCGGGATCGGGGGGCGGATATTGGTCTCCCGTGCGCTCGCGCGCGCGGCTG
M. Tilingi     TGCCTGGGCGTCACGCATTGGCTGGCCCCCCCCCTCTCGCTCCGTCGGGGCGGGATCGGGGGGCGGATATTGGTCTCCCGTGCGCTCGCGCGCGCGGCTG
M. guttatus    TGCNTGGGCGTCACGCANTGNGTCG-CCCCCCCCTCTCGCTCNGTCNGGGCGGGATNNGGGGGCGGATATTGGTCTCCCGTGCGCTCGCGCGCGCGGCTG
M. luteus      TGCCTGGGCGTCACGCATTGCGTCG-CCCCCCCCTCTCGCTCCGTCGGGGCGGGATCGGGGGGCGGATATTGGTCTCCCGTGCGCTCGCGCGCGCGGCTG
M. nasutus     TGCCTGGGCGTCACGCATTGCGTCG-CCCCCCCCTCTCGCTCCGTCGGGGCGGGATCGGGGGGCGGATATTGGTCTCCCGTGCGCTCGCGCGCGCGGCTG
M. floribundus TGCCTGGGCGTCACGCATTGCGTCG---CCCCCTTCTCGCTCCACCGGGGCGGGATC-GGGCGCGGATATTGGTCTCCCGTGCGCTCGAGCGCGCGGCTG

                                  ITS 2
M. micranthus  GCCCAAATGAGATCCCTCGGCGATGCACGTCACGAGCAGTGGTGGTTGAATTATCGACTCGCTTGCTGCGGTGCTTGACGGCATCGTCCGCTCGGGCATC
M. dentatus    GCCCAAATGAGATCCCTCGGCGATGCATGTCACGAGCAGTGGTGGTTGAATTCTCGACTCGCTTGCTGTAATGCTTGATGGCATCGTCCGCTCGGGCATC
M. glabratus g GCCCAAATGAGATCCCTCGGCGATGCACGTCACGAGCAGTGGTGGTTGAATTCTCGACTCGCTTGCTGCGGTGCTTGGCGGCATCGTCCGCTCGGGCATC
M. dentilobus  GCCCAAATGAGATCCCTCGGCGATGCACGTCACGAGCAGTGGTGGTTGAATTCTCGACTCGCTTGCTGCGGTGCTTGACGACATCGTCCGCTCGGGCATC
M. platycalyx  GCCCAAATGAGATCCCTCGGCGATGCACGTCACGAGCAGTGGTGGTTGAATTATCGACTCGCTTGCTGCGGTGCTTGACGGCATCGTCCGCTCGGGCATC
M. glabratus p GCCCAAATGAGATCCCTCGGCGATGCACGTCACGAGCAGTGGTGGTTGAATTCTCGACTCGCTTGCTGCGGTGCTTGACGGCATCGTCCGCTCGGGCATC
M. Tilingi     GCCCAAATGAGATCCCTCGGCGATGCACGTCACGAGCAGTGGTGGTTGAATTATCGACTCGCTTGCTGCGGTGCTTGACGGCATCGTCCGCTCGGGCATC
M. guttatus    GCCCAAATGAGATCCCTCGGCGATGCACGTCACGAGCAGTGGTGGTTGAATTCTCGACTCGCTTGCTGCGGTGCTTGACGGCATCGTCCGCTCGGGCATC
M. luteus      GCCCAAATGAGATCCCTCGGCGATGCACGTCACGAGCAGTGGTGGTTGAATTCTCGACTCGCTTGCTGCGGTGCTTGACGGCATCGTCCGCTCGGGCATC
M. nasutus     GCCCAAATGAGATCCCTCGGCGATGCACGTCACGAGCAGTGGTGGTTGAATTATCGACTCGCTTGCTGCGGTGCTTGACGGCATCGTCCGCTCGGGCATC
M. floribundus GCCCAAATGAGATCCCTCGGCGATGCATGTCACGAGCAGTGGTGGTTGAATTCTCGACTCGCTTGCTGTAATGCTTGATGGCATCGTCCGCTCGGGCATC

                                  ][28 S nrDNA
M. micranthus  AACAACGACCCAACGGCGCCACATCGGCGCCTTCGATT
M. dentatus    ACCAACGACCCAACGGCGCCAC-TTGGCGCCTTCGATT
M. glabratus g ACCAACGACCCAACGGCGCCACATCGG-G-CTTCGATT
M. dentilobus  ACCAACGACCCAACGGCGCCACATCGGCGCCTTCGATT
M. platycalyx  AACAACGACCCAACGGCGCCACATCGGCGCCTTCGATT
M. glabratus p ACCAACGACCCAACGGCGCCACATCGGCGCCTTCGATT
M. Tilingi     ACCAACGACCCAACGGCGCCACATCGGCGCCTTCGATT
M. guttatus    ACCAACGACCCAACGGCGCCACATCGGCGCCTTCGATT
M. luteus      ACCAACGACCCAACGGCGCCACATCGGCGCCTTCGATT
M. nasutus     AACAACGACCCAACGGCGCCACATCGGCGCCTTCGATT
M. floribundus ACCAACGACCCAACGGCGCCCC-CCGGCGCCTTCGATT
```

FIGURE 2: Aligned nucleotide sequences of anonymous microsatellite flanking regions of ingroup and outgroup taxa (Table 1). Vertical columns represent the 246 nucleotide positions, beginning with the 5' end. Ambiguous sites are represented by "N" (see Table 2 for ambiguous site frequency) and are treated as missing data, while gaps inserted by the alignment program are indicated by "-" and treated as phylogenetically useful information.

```
M. nasutus      AATTTCCTACGCGCTCGTATTTTTTATTATTTTTTTCGGGTCGGAAAATTTTCGGACGTTCGGCACCGGCCCAGCC--GGTGTGCGGGGCTGCCTTGCCA
M. Tilingi      AATTTCCTACGCGCTCGTATTTTTTATTATTTTTTTCGGGTCGGAAAATTTTCGGACTTTCGGCACCGGCCCAGCCAAGGTGTGCGGGGCTGCCTTGCCA
M. guttatus     AATTTCCTACGCNCTCGTATTTTTTATTATTTTTTTCGGGTCGGAAAATTTTCGGACTTTCGGCACCGGCCCAGCCAATGTGTGCGGGGCTGCCTTGCCA
M. platycalyx   AATTTCCTACGCGCTCGTATTTTTTATTATTTTTTTCGGGTCGGAAAATTTTCGGACTTTCGGCACCGGGTGGGCC--GGTGTGCGGGGCTGCCTTGCCA
M. micranthus   AATTTCCTACGCGCTCGTATTTTTTATTATTTTTTTCGGGTCGGAAAATTTTCGGACGTTCGGCACCGGCCCAGCC--GGTGTGCGGGGCTGCCTTGCCA
M. dentilobus   AATTTCCTACGNGCTCGTATAATTTATTATTTTTTTCGGGTCGNAAAATTATNGGACGTTCGGCACCTGTNTATCC--GGTGTGCGGAGTTGCCTTGCCC
M. glabratus p  AATTTCCTACGTCCTCGTATTTTTTATTATTTTTTTCGGGTCGGAAAATTTTCGGACTTTCGGCACCGGCCCGGGC-TGGTGTGCGGGGCTGCCTTGCCC
M. luteus       AATTTCCTACGTCCTCGTATTTTTTATTATTTTTTTCGGGTCGGAAAATTTTCGGACTTTCGGCACC-GTCGGGGC-TGGTGTGCGGGGCTNCCTTGCCC
M. glabratus g  AATTTCCTACGTCCTCGTATTTTTTATTATTTTTTTCGGGTCGGAAAATTTTCGGCCGTTCGGCAACTGCTGGGCC--GGTGTGCGGGGCTGCTTTGTCA
M. dentatus     AATTTCCTACGCGTTCGTATTTTTTATTATTTTTTTCGGGTCGGAAAATTTTCGGACGTTCGGCACCGGGCCAGCCAANNTGTGGGCGGCTGCNTNNCCA
M. floribundus  AATTTCCTACGCGCTCGTATTTTTTATTATTTTTTTCGGGTCGGAAAATTTTCGGACGTTCGGCACCGGCCCAGCC--GGTGTGCGGGGCTGNCTTGCCA

M. nasutus      TAACTAGTCG----C--C-TTTTTCGCACCGCACCCGA-------CGCCCGGCGGCCGGCNAAGGCAAAGGNAAAGGNTAAGGNTAAGGCTATTATTATA
M. Tilingi      TAACTAGTCG----C--C-TTTTTCGCACCGCACCCGACGCCCGCCGCCCGGCGGCCGGCGAAGGCAAAGGCAAAGGCTAAGGCTAAGGCTATTATTATA
M. guttatus     ACACTAGTCGCCCCCGACGTTTTTCGCACCGCACCCGACGCCCGCCGCCCGGCGGCCGGCGAAGGCAAAGGCAAAGGCTAAGGCTAAGGCTATTATTATA
M. platycalyx   AAACTAGTCG----C--C-TTTTTCGCACCGCACCCGA-------CGCCCGCCGCCCGGCGAACGGAGAAGGCAAAGGTAAAGGCNAAGGCNNNAGCNNT
M. micranthus   TAACTAGTCG----C--C-TTTTTCGCACCGCACCCGA-------CGCCCGGCGGCCGGCGAAGGCAAAGGCAAAGGCTAAGGCTAAGGCTATTATTATA
M. dentilobus   TATCTAGTCG----C--G-CNTNTCGCGCCGCGCGCCN-------NGCGCCCNGCCCGGGCAAAGCCAANGCNTANGNTTANGNTAANNATANTNATATA
M. glabratus p  AAACTTGTCGCCCCCCGACGTTTTTCGCACCGCACCCGACGCCCGCCGCCCGGCGNACGNAAAAGGNAAAGGNAAAGGTTAAGGTTNANGNTNTTATNNTN
M. luteus       AAACTTGTCGCCCCCTACGTTTTTCGCACCGCACCCGACGCCCGCCGCCCGGCGGCCGGCGAAGGCAAAGGCAAAGGCTAAGGCTNANGNTATTATTATA
M. glabratus g  TAACTAGTCGCCCCCGACGTTTTTCGCAACGCACCCGA-------CGCCCGCCGCCCGGCGAACGCAGAAGGCAAGGGTAAGGGTTAGGGNNTNAGCNTN
M. dentatus     TAACTACTTNCCCATAACGTTTTTCGTTTTTCNCNCCNCNCCCGACGCCCGGCGGCCGGCGAAGGCAAAGGCAAAGGCTAAGGCTAAGGCTATTATTATA
M. floribundus  TAACTAGTCG----C--C-TTTTTCGCACCGCACCCGA-------CGCCCGGCGGCCGGCGAAGGCAAAGGCAAAGGCTAAGGCTAAGGCTATTATTATA

M. nasutus      TATATATAAATATATATAATAAAAATTTATTTATTTNGNACCGGGC
M. Tilingi      TATATATAAATATATATAATAAAAATTTATTTATTTCGGACCGGGC
M. guttatus     TATATATAAATATATATAATAAAAAAATATTTATTTCGGACCGGGC
M. platycalyx   TATTTTTTATTTTTTATTATATATATTTATNTTTTNNNNANCGNGC
M. micranthus   TATATATAAATATATATAATAAAAATTTATTTATTTCGGACCGGGC
M. dentilobus   TATATATAAATATATATAATAAAAATTTATTTATTTCGNNCCGGGC
M. glabratus p  TATATNTNANTATATATAATAAAAANTTATTTATTTNGNNCCGGGC
M. luteus       TATATATAAATATATATAATAAAAATTTATTTATTTNGGACCGGGC
M. glabratus g  TATNTNTTANTNTTTATTATANATATTTATTTNTTNCGGACCGNGC
M. dentatus     TATATATAAATATATATAATAAAAATTTATTTATTTGGNACCGGGC
M. floribundus  TATATATAAATATATATAATAAAAATTTATTTATTTCGGACCGGGC
```

(12.2%), and G-C content was 47% (Figure 2). One or more gaps were needed at 6.9% of the sites. The variable replication of microsatellite sequences in different individuals is thought to account for the length variation, and cause the higher gap frequency in these sequence alignments than in the ITS data. Sequence divergence for this data set ranged from 0.0 to 28%.

TABLE 2: Summary of Length Variation of both ITS and Anonymous Fragments: Both ends of all sequences were aligned, so length variation is parsimony informative. However, note that lengths of complete ITS regions (5'— ITS 1— 5.8 S – ITS 2 – 3') would include approximately 100 more base pairs to the 5' end of the hairpin in ITS 1.

Species	ITS frag. (bp)	% ambig.	Anon frag. (bp)	% ambig.
M. guttatus	537	4.3	246	0.4
M. dentatus	531	0.2	245	4.5
M. dentilobus	535	0	231	8.2
M. glabratus var. *parviflorus*	537	0	242	7.4
M. glabratus var. *glabratus*	533	0	237	5.5
M. Tilingi	538	1.7	239	0
M. micranthus	534	0.2	230	0
M. nasutus	537	0.2	230	2.6
M. platycalyx	536	3.9	230	5.7
M. luteus	537	0	244	2.0
M. floribundus	526	0	230	0.4

Tree construction. Parsimony analysis of ITS sequence data generated 12 maximally parsimonious trees. Each of these trees required 143 evolutionary steps. Collapsing these trees via strict consensus generated a final tree (Figure 3) with a consistency index (CI) of 0.846, a homoplasy index (HI) of 0.154, and a retention index (RI) 0.569. Excluding uninformative characters, the CI was 0.656, and the HI was 0.344. The rescaled consistency index (RC) was 0.481. Bootstrap values for the consensus tree ranged from 79 to 100%. Neighbor joining analysis of the ITS data yielded slightly more resolved tree with similar topology (Figure 4). Bootstrap values ranged from 36 to 100%.

FIG. 3. Strict consensus of the 12 most parsimonious trees of 143 steps derived from heuristic analyses (unweighted) of ITS 1, 5.8S, and ITS 2 sequence data (Fig. 1), with indels coded as phylogenetically useful information. Large numbers in italics above branch points are bootstrap percentage values for clades in both the strict consensus and bootstrap majority rule trees. Small numbers above branches are branch lengths.

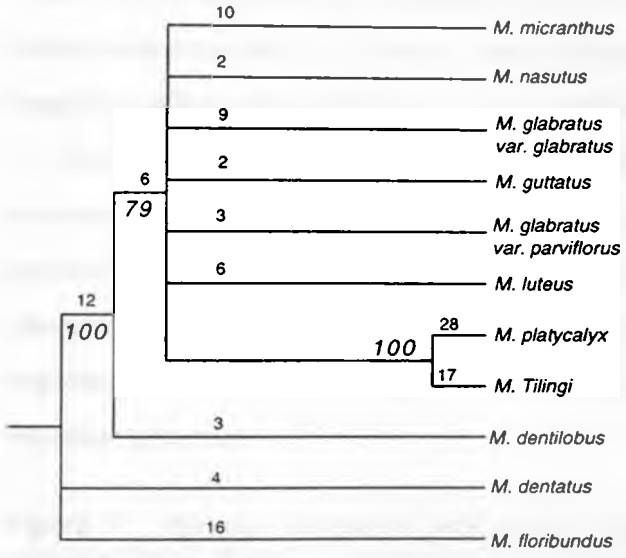


FIG. 4. Tree derived from neighbor joining methods using ITS 1, 5.8S, and ITS 2 sequence data (Fig. 1), using the Jukes-Cantor method of distance estimation and paiwise deletion of gaps. Numbers in italics above branch points are bootstrap percentage values for clades, and small numbers above branches are branch lengths.

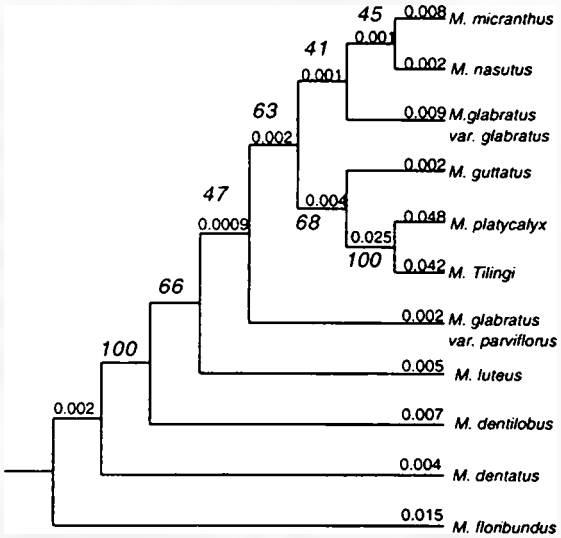


FIG. 5. Strict consensus of the 2 most parsimonious trees of 135 steps derived from heuristic analyses (unweighted) anonymous sequence data (Fig. 2), with indels coded as phylogenetically useful information. Large numbers in italics above branch points are bootstrap percentage values for clades in both the strict consensus and bootstrap majority rule trees. Small numbers above branches are branch lengths.

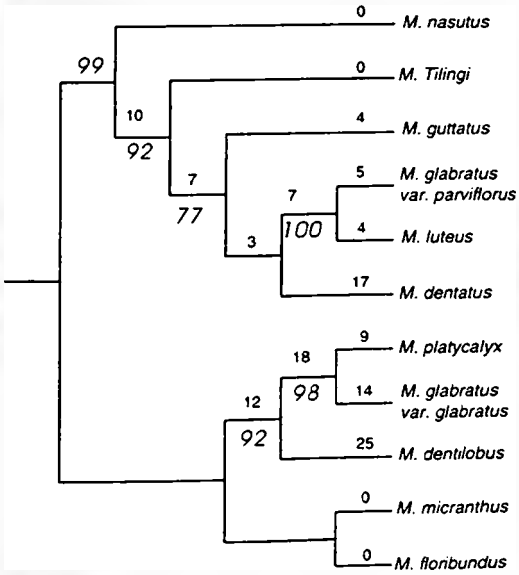
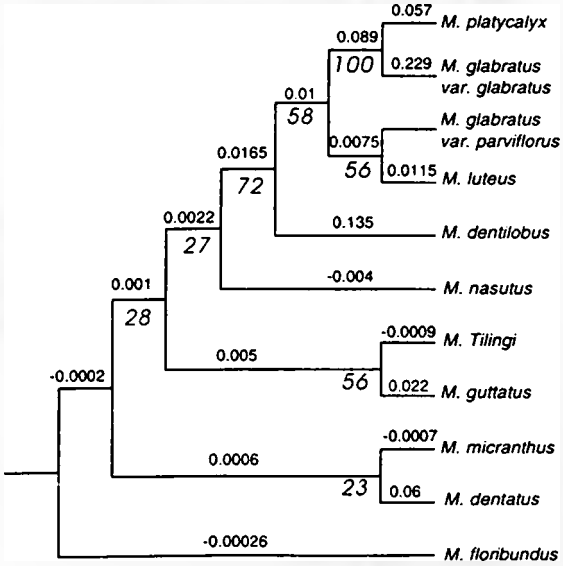


FIG. 6. Tree derived from neighbor joining methods using anonymous sequence data (Fig. 2), using the Jukes-Cantor method of distance estimation and paiwise deletion of gaps. Numbers in italics above branch points are bootstrap percentage values for clades, and small numbers above branches are branch lengths.



Parsimony analysis of anonymous sequence data generated 2 maximally parsimonious trees each requiring 135 steps. These trees were collapsed to yield a consensus tree with CI = 0.852, HI = 0.148, and RI = 0.798 (Figure 5). After exclusion of uninformative characters, CI = 0.744, HI = 0.256, and the RC = 0.680. Bootstrap values for the consensus tree ranged from 77 to 100%. Neighbor joining analysis of the anonymous sequence data yielded a tree with a different topology, negative branch lengths, and bootstrap values from 23 to 100% (Figure 6).

Because the two data sets generated such different trees they were not combined into one data set. Constraining the anonymous data with the ITS parsimony tree revealed that out of 654,729, 075 distinct rooted binary trees for the 11 taxa, 10,395 (0.001%) were compatible with the constraint tree. Constraining the ITS data with the anonymous sequence parsimony tree revealed that only 3 trees (4.58x10^-7 %) were compatible with the constraint tree.

Figure 7: Pairwise divergence between ITS region sequences from the 11 ingroup and 1 outgroup DNAs (Table 1). Sequences were compared only at unambiguously aligned positions that lacked gaps or missing information. The numbers on the left and on the top refer to the species list in the box to the left of the matrix (note that these are different from the numbers in the anonymous sequence matrix).

		1	2	3	4	5	6	7	8	9	10	11
1. M. micranthus	1		0.0311	0.0191	0.0210	0.0850	0.0190	0.0803	0.0179	0.0171	0.0094	0.0492
2. M. dentatus	2			0.0350	0.0289	0.1080	0.0289	0.0959	0.0363	0.0289	0.0328	0.0213
3. M. glab g	3				0.0229	0.0939	0.0152	0.0825	0.0139	0.0190	0.0114	0.0514
4. M. dentilobus	4					0.0891	0.0113	0.0821	0.0178	0.0151	0.0170	0.0410
5. M. platycalyx	5						0.0800	0.0896	0.0703	0.0845	0.0780	0.1205
6. M. glab p	6							0.0734	0.0098	0.0075	0.0094	0.0410
7. M. Tilingi	7								0.0723	0.0778	0.0757	0.1145
8. M. guttatus	8									0.0138	0.0078	0.0492
9. M. luteus	9										0.0132	0.0390
10. M. nasutus	10											0.0471
11. M. floribun												

Figure 8: Pairwise divergence between anonymous region sequences from the 11 ingroup and 1 outgroup DNAs (Table 1). Sequences were compared only at unambiguously aligned positions that lacked gaps or missing information. The numbers on the left and on the top refer to the species list in the box to the left of the matrix (note that these are different from the numbers in the ITS matrix).

		1	2	3	4	5	6	7	8	9	10	11
1. M. nasutus	1		0.0045	0.0274	0.1677	0.0000	0.1500	0.0494	0.0471	0.1290	0.0630	0.0000
2. M. Tilingi	2			0.0213	0.1758	0.0044	0.1593	0.0615	0.0444	0.1472	0.0737	0.0044
3. M. guttatus	3				0.1945	0.0267	0.1905	0.0647	0.0567	0.1648	0.1005	0.0268
4. M. platycalyx	4					0.1816	0.2854	0.1637	0.1795	0.0808	0.2620	0.1826
5. M. micranthus	5						0.1534	0.0640	0.0460	0.1416	0.0665	0.0000
6. M. dentilobus	6							0.1929	0.1698	0.2452	0.2112	0.1543
7. M. glab p	7								0.0364	0.1403	0.1528	0.0643
8. M. luteus	8									0.1409	0.1237	0.0462
9. M. glab g	9										0.2275	0.1423
10. M. dentatus	)											0.0668
11. M. floribun												

DISCUSSION

Sequence comparisons. The organization of the ITS region in the species examined appears to be similar to other groups of angiosperms. The length of the ITS 2 region in these *Mimulus* species is the same as several other angiosperm families (ex. Solanaceae, Poaceae) as reviewed in Baldwin (1995). While exact length of ITS 1 is not known for these *Mimulus* species, crude estimates of 240 base pairs suggest the length similarity of the *Mimulus* ITS 1 sequence with other angiosperm families. As expected, the length of 5.8S nrDNA is conserved in all examined families (Baldwin 1995). The relatively low level of length variations in ITS sequences is convenient for phylogenetic studies where identification of positional homologies is essential.

The only suggestion that ITS regions of *Mimulus* taxa diverge from other angiosperms comes from the strong hairpin formation 80 base pairs downstream from the 5' end. While secondary structure formation has been a problem in other analyses using the ITS region, estimates suggest the hairpin in *M. guttatus* ITS 1 has a stronger ΔG than other angiosperm families (J. Whitthall, pers. comm.). This structure might be essential for proper post-transcriptional modification of these loci. ITS 1 has a higher percentage of variable sites, parsimony informative sites, and gap sites than the ITS 2, consistent with sequence data and secondary structure variation from other angiosperms (Baldwin 1995; Kim and Jansen, 1994).

The overall variation in the anonymous sequence data was much higher, and G-C content much lower than the ITS sequence data. This result is not surprising, given the structure and function of the microsatellite region. Microsatellite DNA has high length polymorphism, and is not thought to have any functional purpose. This combines to limit both the selective pressure and the phylogenetic utility in interspecific studies of these sequences. Alignment of these sequences did not suffer from their unknown location or function.

Phylogenetic resolutions. The ITS trees (Fig. 3, 4) from both optimality criteria give similar evolutionary predictions for this section of the genus. Four robust clades

exist in both trees. First, trees consistently place *M. floribundus* and *M. dentatus* in a separate clade. This is expected based on morphological data and molecular data (J. Whittall, pers. comm.) and prior placements of these species in the Paradanthus section (Grant, 1924). Although *M. floribundus* was used as the outgroup, this clade was resolved in analyses without an outgroup (data not shown). These taxa exist in a polytomy in the parsimony analysis, and are separated by a small distance in the neighbor joining analysis. If branch lengths equal to the number of synapomorphies separating taxa are roughly proportional to evolutionary distance, *M. floribundus* would be a more distant relative than *M. dentatus* to the Simiolus clade. We suspect that the level of divergence in the ITS data is insufficient to resolve this relationship because the ITS sequences are relatively conserved. For the purposes of this analysis, however, the placement of *M. floribundus* and *M. dentatus* together and in a separate clade from Simiolus is encouraging.

Second, both trees place *M. dentilobus* on a separate branch from the Paradanthus clade and the other members of the Simiolus clade. Grant (1924) groups *M. dentilobus* in Simiolus but not in either the *M. guttatus* of *M. glabratus* species complexes. This, and morphological data agree with the placement on our tree.

Third, the grouping of the eight remaining species into one unresolved clade supports the existing Simiolus group and species complex definitions. The apparent resolution of this clade in the neighbor-joining tree is not supported by the low bootstrap estimates, as 75 percent is a common cut-off point for credibility. The topology shown in this tree is included only as a hypothetical arrangement, and more weight should be placed on the parsimony tree topology. We assume that this clade is mostly unresolved due to lack of variation in the ITS data due to close evolutionary relationships within the clade. The relatively short branches support this belief. The distance matrix for these data (Figure 7) show the short distance between *M. guttatus* and *M. nasutus*, which supports the belief that the selfing *M. nasutus* evolved relatively recently from an outcrossing relative.

The only resolution within this eight-membered clade comes from the *M. platycalyx-M. Tilingi* group, which receives 100 percent bootstrap estimates and has quite long branches (or distances) in both trees. The robust presence of the *M. platycalyx-M. Tilingi*

group warrants a variety of explanations. First, this relationship is supported by cytological evidence, which would place *M. platycalyx*, *M. Tilingi*, and possibly *M. guttatus* nearby on the tree. Perhaps this is an example of selfer (*M. platycalyx*) evolution from an outcrosser (*M. Tilingi*). If so, this would support the belief that selfers are of recent origin. The close relationships within this clade support the belief that the clade is undergoing a relatively recent adaptive radiation. An alternative explanation for the *M. platycalyx-M. Tilingi* group could be long branch attraction, in which long, unbranched lineages tend to be mistakenly "attracted" to one another in the analysis. Sampling more taxa or characters may ameliorate this problem.

While the relatively low level of sequence variation prevents a sharp resolution of the tree, the relatively high consistency and retention indices and low homoplasy index support the resolutions that can be obtained from this tree. Unfortunately, these resolutions do not offer conclusive evidence for the origin of self-fertilization within Simiolus.

Trees obtained from different optimality criteria using anonymous sequence data show different topologies. This fact alone limits the credibility of the trees and confuses possible interpretations. The only shared resolution is the *M. glabratus* var. *glabratus-M. luteus* clade. More interesting is the incompatibility of anonymous sequence trees with the ITS trees. This low compatibility suggests alternative methods of evolution for the two sequences. Given the relatively fast evolutionary rates of the microsatellite sequences and their flanking regions and the high variation between individuals, this is not surprising. We propose the non-parsimonious evolution of these sequences, and tend to avoid basing evolutionary conclusions on these data before obtaining more convincing evidence. Given the widespread utility of the ITS region in species-level comparisons, we put more weight into these analyses.

Difficulty in interpreting the anonymous sequence data stems from the problem of sequence variation in different individuals. In this study, sequence samples were determined from only one individual, but it is quite possible that different individuals in a population might have different alleles, even for more conserved regions like the ITS. It is unclear how this problem should be addressed in phylogenetic analyses. A more comprehensive study might include levels of sequence variation in a particular sequence

for many individuals within various populations. A less expensive technique might include samples of genomic DNA from many individuals. Amplification and sequencing of sequences of interest might be the consensus of all variations within that population.

Conclusion

The ITS region has apparent value as a source of phylogenetic data in the sections Paradanthus and Simiolus in *Mimulus*. These sequences suggest existing Paradanthus and Simiolus clades, and the close relationships between selfers and outcrossers in the *M. guttatus* and *M. glabratus* species complexes. Some anonymous genomic regions containing microsatellite DNA and its flanking regions appear to evolve non-parsimoniously and thus have limited phylogenetic utility. The origin of self-fertilizing members of this section remains unclear.

Future Directions

The remaining 80-100 base pairs on the 5' end of ITS 1 should be resolved to complete the ITS analysis of these taxa. Other possibilities for sequencing through particularly strong hairpins were not undertaken in this study due to time and money constraints. These possibilities include using formamide, dimethyl sulphoxide, detergents (Winship 1989; Zhang et al., 1992; Bechmann et al., 1990; and Kim and Jansen, 1994), and thermostable polymerases and higher sequencing temperatures. Further analysis should include a kingdom-wide survey in the plant kingdom for rDNA hairpins of similar strength to understand their secondary structure and function in the ribosome.

Further efforts could concentrate on other anonymous sequences and microsatellite flanking regions. A genus-wide phylogeny could be constructed using chloroplast DNA sequences with slower evolutionary rates than the ITS. Phylogenetic analysis of the data in this study should be completed using maximum likelihood as an optimality criterion. This method can incorporate variable evolutionary rates and different evolutionary models.

REFERENCES

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-10. <http://ncbi.nlm.nih.gov>.

Antonovics, J. 1968. Evolution in closely adjacent plant populations. V. Evolution of self-fertility. Heredity 23:219-238.

Baker, H.G., and Stebbins, G.L. (eds.). 1965. The Genetics of Colonizing Species. Academic Press: New York.

Baldwin, B.G. *et al* . 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Ann. Missouri Bot. Gard. 82:247-277.

Beardsley, Paul. Personal communication. May 25, 1998.

Bechmann, B., W. Luke, and G. Hunsmann. 1990. Improvement of PCR amplified DNA sequencing with the aid of detergents. Nucl. Acids Res. 18:1309.

Charlesworth B., Morgan, M.T. and Charlesworth, D. 1991. Multilocus models of inbreeding depression with synergistic selection and partial self-fertilization. Genet. Res. Camb. 57:177-194.

Charlesworth D., Morgan, M.T. and Charlesworth, B. 1990. Inbreeding depression, genetic load and the evolution of outcrossing rates in a multi-locus system with no linkage. Evolution 44:1469-1489.

Charlesworth, B. 1980. The cost of sex in relation to mating system. J. Theoret. Biol. 84:655-671.

Charlesworth, D. and Charlesworth, B. 1987. Inbreeding depression and its evolutionary consequences. Ann. Rev. Ecol. Syst. 18:237-268.

Darwin, C R. 1876. The Effects of Cross and Self-Fertilization in the Vegetable Kingdom. John Murray: London.

Feldman, M.W. and Christiansen, F.B. 1984. Population genetic theory of the cost of inbreeding. Amer. Nat. 123:642-653.

Fenster, C.B., and Ritland, K. 1994. Evidence for natural selection on mating system in *Mimulus* (Scrophulariaceae). Int. J. Plant Sci. 155:588-596.

Fisher, R.A. 1941. Average excess and average effect of a gene substitution. Ann. Eugen. 11:53-63.

Hutchinson, G.E. 1957. Concluding remarks. Cold Spring Harbor Symp. Quant. Biol. 22:415-427.

Jukes, T.H., and C.R. Cantor. 1969. Evolution of protein molecules. In H.N. Munro (ed.), Mammalian Protein Metabolism. Academic Press: New York.

Karl, S.A, and J.C. Avise. 1993. PCR-based assays of Mendelian polymorphisms from anonymous single-copy nuclear DNA: Techniques and applications for population genetics. Mol. Biol. Evol. 10:342-361.

Karl, S.A., Bowen, B.W., and Avise, J.C. 1992. Global population genetic structure and male-mediated gene flow in the green turtle (*Chelonia mydas*): RFLP analyses of anonymous nuclear loci. Genetics. 131:163-173.

Kiang, Y.T. 1973. Floral structure hybridization and evolutionary relationships of two species of *Mimulus*. Rhodora 75: 225-238.

Kim, K.-J., and R.K. Jansen. 1994. Comparisons of phylogenetic hypotheses among different data sets in dwarf dandelions (*Krigia*): Additional information from internal transcribed spacer sequences of nuclear ribosomal DNA. Pl. Syst. Evol. 190: 157-185.

Kumar, S., K. Tamura, and M. Nei. 1993. MEGA: Molecular Evolutionary Genetics Analysis, version 1.0. The Pennsylvania State University, University Park, PA 16802.

Lande, R. and Schemske, D.W. 1985a. The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. Evolution 39: 24-40.

Lin, J.-Z. and Ritland, K. 1995. Flower petals allow simpler and better isolation of DNA for plant RAPD analyses. Plant Mol. Biol. Reporter. 13:215-218.

Lindsay, D.W. 1964. Natural dispersal of *Mimulus guttatus*. Proc. Utah Acad. Sci. Arts Lett. 41:237-241.

Liston, A., W.A. Robinson, and J.M. Oliphant. 1996. Length variation in the nuclear ribosomal DNA internal transcribed spacer region of non-flowering seed plants. Systemic Botany 21: 109-120.

Lloyd, D.G. 1979. Some reproductive factors affecting the selection of self-fertilization in plants. Amer. Nat. 113:67-79.

Löve, A., and Löve, D.1957. Arctic polyploidy. Proc. Genet. Soc. Can. 2:23-27.

Maynard Smith, J. 1978. The evolution of sex. Cambridge University Press, Cambridge.

Nickrent, D.L., K.P. Schuette, and E.M. Starr. 1994. A molecular phylogeny of Arceuthobium (Viscaceae) based on nuclear ribosomal DNA internal transcribed spacer sequences. American Journal of Botany 81: 1149-1160.

Ritland, C.E., Ritland, K., and Straus, N.A. 1993. Variation in the ribosomal internal transcribed spacers (ITS 1 and ITS 2) among eight taxa of the *Mimulus guttatus* species complex. Mol. Biol. Evol. 10:1273-1288.

Stebbins, L.G. 1957. Self-fertilization and population variability in the higher plants. Amer. Natur. 91:337-354.

Swofford, D.L. 1993. PAUP (Phylogenetic Analysis Using Parsimony) version 3.1, Program and documentation. Champaign, IL.

Vickery, R.K., Jr. 1978. Case studies in the evolution of species complexes in *Mimulus*. Evolutionary Biology 11: 404-506.

Wheeler, W.C., and D.S. Gladstein. MALIGN: a multiple sequence alignment program. J. Heredity. In Press.

Whittall, Justen. Personal communication. June 1, 1998.

Willis, J.H. 1993. Partial self-fertilization and inbreeding depression in two populations of *Mimulus guttatus*. Heredity. 71:145-154.

Winship. P.R. 1989. An improved method for direct sequencing PCR amplified material using dimethyl sulphoxide. Nucl. Acids Res. 17: 1266.

Zhang, W., C. Reading, and A.B. Deisseroth. 1992. Improved PCR sequencing with formamide. Trends Genet. 8: 332.

GLOSSARY

Alignment The juxtaposition of amino acids or nucleotides in homologous molecules to maximize similarity or minimize the number of inferred changes among the sequences; used to infer positional homology.

Allele A particular form of a gene at a particular locus.

Autapomorphy A derived character state unique to a particular taxon.

Bootstrapping A measure of robustness of a phylogeny that seeks the number of times a particular branch appears in many replication of a data set.

bp Base pair; refers to the four bases of DNA and the pairs they form (A-T, C-G)

Character A variable feature in a taxon or DNA sequence that groups one separately from two or more others.

Concerted evolution The generation and maintenance of the same DNA sequence among members of a family of sequence repeats within a species or a population; the mechanisms of concerted evolution remain unclear.

Conserved sequences Sequences under relatively high selective pressure that tend to be more similar among different organisms

Consistency index A measure of the amount of homoplasy exhibited by a character or set of characters on a tree, defined as the sum of the minimum individual character ranges divided by the observed number of changes. If there is no homoplasy, these quantities will be equal, so the consistency index achieves is maximum value of one.

Distance A measure of the difference between two objects, usually dimensionless and measured from 0 to infinity; often presented in a distance matrix.

DMSO Dimethlysulfoxide

Downstream to the 3' end of the target sequence; refers to the 5' → 3' polarity of a DNA sequence.

Electrophoresis The separation of charged molecules in an electric field.

Evolution The field of biology that links all others under a theoretical umbrella; a change in the gene pool of a population over time.

Gaps Editing symbols that are inserted into sequences in the process of alignment to compensate for putative insertion and deletion events (usually denoted by "-")

Genome The complete set of DNA for a particular organism.

Genomic library A mixture of DNA fragments inserted into a virus or plasmid genome that together represent virtually all of an organism's DNA.

Heterozygous An organism is said to be heterozygous for a particular gene if its two copies of the gene are different; for example, a mutant copy and a functional copy

Heuristic method Any analysis procedure that does not guarantee finding the optimal solution to a problem (usually used because it is faster than more exact methods).

Homology Common ancestry.

Homoplasy A collection of phenomena that leads to similarities in characters for reasons other than inheritance from a common ancestor. Includes convergence, parallelism, and reversal.

Inbreeding depression Lowered fitness of inbred individuals compared to outcrossers.

Indel Insertion/ deletion event.

Ingroup The taxa of primary interest.

Jukes-Cantor model The DNA substitution model of T.H. Jukes and R.C. Cantor (1969) that assumes all possible nucleotide substitutions are equally likely.

Microsatellites A subset of variable nucleotide tandem repeats (VNTRs) characterized by very short (2-5 bp) tandem repeats with a high rate of variation in copy number among individuals. These loci tend to be randomly distributed throughout the genome and are subject to replication slippage that leads to length variation.

Neighbor joining A heuristic method for obtaining a point estimate of a minimum evolution tree by grouping closest neighbors.

Neo-Darwinism The unification of Darwin's models of evolution with the understanding of genetics.

Outcrossing The fusion of male and female gametes from different genetic individuals.

Outgroup One or more taxa assumed to be phylogenetically outside the ingroup and used to root a phylogenetic tree.

Parsimony The most simple, or most elegant explanation; an evolutionary model that proposes that evolution acts in the shortest number of steps.

Phylogeny A branching diagram that shows, for each species, with which other species it shares its most recent common anscestor; an evolutionary tree.

Plasmid A self-replicating extrachromosomal circular DNA.

Polymerase chain reaction (PCR) A process for amplifying a target DNA sequence manyfold using specific oligonucleotide primers.

Polytomy A node in a tree that connects more than three branches, often resulting due to a lack of data available to infer the phylogeny.

Recessive homozygote Occurs more often in selfing populations; when two mutant copies of a gene that are normally not expressed to combine to give a mutant phenotype.

Selective pressure The pressure on properly functioning genes to stay that way because organisms that have them are favored by natural selection.

Self-fertilization The fusion of male and female gametes from a single genetic individual.

Speciation The process by which new species are formed.

Species A somewhat ambiguous and often redefined concept that describes one biological entity.

Synapomorphy A shared derived character state that is indicative of a phylogenetic relationship among two or more taxa. Used as a measure for branch lengths in a parsimony analysis.

Upstream to the 5' end of the target sequence.