GENETICS OF FLORAL DEVELOPMENT IN *MIMULUS*

by

JONATHAN D. GRUBER

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 Science

June 2000

APPROVED: 

Dr. John H. Willis

An Abstract of the Thesis of

Jonathan D. Gruber **for the degree of** **Bachelor of Sciences**

in the Department of Biology **to be taken** **June 2000**

Title: GENETICS OF FLORAL DEVELOPMENT IN *MIMULUS*

Approved: [redacted]
Dr. John H. Willis

Genetic differences between populations occur as a result of evolution. These changes, however, are not well characterized in any organism. The method of genetically mapping Quantitative Trait Loci (QTLs) offers insight to the genetic basis for adaptive differences but leaves some questions unanswered. To describe candidate genes and genetic loci that might cause QTLs on a linkage map of the related monkey flowers *Mimulus guttatus* and *Mimulus nasutus*, I utilized two approaches. First, I induced mutations in candidate *M. guttatus* genes involved in floral development. Second, I used the Polymerase Chain Reaction (PCR) to amplify DNA fragments of two *Mimulus* genes (polymorphic between the species) that are members of the TCP gene family, which is known to affect floral development in other plants. If the mutant genes and polymorphic gene fragments map to QTLs, they will be implicated as possible causes for the adaptive differences between *M. guttatus* and *M. nasutus*.

ACKNOWLEDGEMENTS

The author is indebted to many people for vital assistance for this work. Thanks to Professor John Willis, especially for his guidance but also for his willing donation of time, effort, and lab resources to these experiments. The other members of the thesis committee, Professors Eric Selker, Sharon Schuman, and Nathan Tublitz, deserve abundant thanks for their comments and advice. Over the last two years, many member of the Willis laboratory, including Dr. Alan Kelly and Dr. Lila Fishman, graduate students Jan Aagaard, Megan Hall, and Noland Martin, and undergraduates Cheryl Gonzales, Emily Morgan, and Kevin Reichelt, have provided vital assistance to these experiments. For experimental advice and scientific inspiration, the author thanks Jan Aagaard especially. These experiments would not have been possible if not for the abundant help and patience of Marcia Trenary and the greenhouse staff. Finally, the author is very grateful for the support of his friends and family, especially Laura Reed for her cheerful work as the author's scientific sounding board.

For two summers the author was supported by a National Science Foundation Research Training Grant (NSF RTG) for Genetic Mechanisms of Evolution, for which he is very grateful.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Adaptation is at the heart of evolution. What mechanisms enable related populations to evolve adaptive differences? Neo-Darwinist thinking would argue that it is the gradual accumulation of small genetic changes over time that lead to adaptive differences between species. This explanation has appeal because most traits are quantitative and continuous: they do not simply occur in several definite forms. On this view, perhaps the slight and subtle variation that exists among individuals within a population is the raw material for adaptive differences between populations and species. On the other hand, it is conceivable that some adaptive differences are due to just a few heritable changes of large effect that are qualitatively different from the standing variation within populations. Using the model system of yellow monkey flowers, I attempted to characterize candidate loci that researchers can use to find the genes responsible for adaptive differences.

The genetics of adaptive differences between species are quite complicated. There are many pieces to the puzzle for researchers to discover. How many genes are involved? It has been an old question in biology, and is now one of modern importance, for humans are faced with large-effect phenomena like antibiotic resistance in bacteria or pesticide resistance in insects, which may or may not be typical instances of evolution. What classes of genes are involved? For instance, the function of many genes is to affect the

expression of other genes. Some researchers have theorized that these regulatory genes are better candidates to influence evolution than "structural" genes, whose function does not involve the expression of other genes (Doebley 1993). What kind of genetic changes occurred at each gene? Changes at a gene can affect many of its properties, such as the timing of the gene's expression, the quantity of gene product, and the structure of that gene product. Each of these genetic changes can have a different effect on the organism.

The "evolution" of maize from its wild relative, teosinte, is one of the only examples when a quantitative difference has been explored so thoroughly that a major contributing gene has been well characterized. This story demonstrates how much information must be known to truly show how morphological transitions occur at the molecular genetic level.

Researchers interested in this process in maize had to synthesize three important data. One was the location on a maize/teosinte chromosome that statistically contributes to the difference in phenotype (the manifestation of a particular genetic makeup, or genotype) between species. The second, from a very separate line of investigation, was the linkage map location of a particular mutant gene that causes a phenotypic change in maize. The third was the identification and molecular characterization of the genes.

Maize (*Zea mays* subspecies *mays*) and teosinte (*Zea mays* ssp. *parviglumis* or *mexicana*) differ dramatically in several morphological characteristics. Most important is the differences in the primary lateral branches. In maize, the primary lateral branches are short and terminate in the female inflorescence (the ear). In teosinte, the primary branches are much longer, and the terminal region is the male inflorescence, a tassel-like

structure. There are also differences in the structure of the ears, as maize makes many cupules (kernels) in multiple rows, whereas teosinte has a single row of fewer than ten cupules. Hybrids between the varieties show an intermediate form (Doebley and Stec 1991, 1993).

To find a region of the maize genome that contributes to the difference in quantitative phenotype between the two varieties, Doebley and Stec (1991, 1993) performed F2 crosses between domesticated corn and teosinte. A typical F2 cross involves two genetically distinct parents (P1 and P2), mated together. Their progeny are the F1 (filial-1) generation, each member of which has inherited half of its alleles (particular versions of a gene) from one parent and the other half from the second parent. Members of the F1 are then mated with each other (or with themselves, for hermaphroditic organisms like corn). The progeny of those matings are the F2 generation. While, on average, each F2 organism still has half of its genes from each parent, an F2 can have any combination of the two parents' alleles at a particular region (locus) on the chromosome. If both alleles are identical, the individual is called "homozygous" at that locus. If the alleles are different, the individual is called "heterozygous." Because of this variation, F2 individuals are generally more different from one another than are the F1 individuals.

Using this scheme, a map of molecular markers (particular regions on the chromosomes whose presence can be assayed) was made (Doebley and Stec 1991, 1993). In the F2, the degree to which parent alleles at two loci are inherited concordantly is dependent on the distance between the loci on a chromosome. Statistics have been

developed to describe this property in terms of relative genetic "distances" between markers, and computer programs make the process of developing an accurate map of these distances relatively easy.

For each plant, measurements for various quantitative traits were taken. Using these measurements and the molecular marker data, one approach for finding genes responsible for the difference is to locate Quantitative Trait Loci (QTLs). QTLs are regions on chromosomes that contribute to the difference between parents in a trait. Molecular markers with no direct effect on phenotype ("neutral") can be used to map QTLs by testing whether there is a statistical association between marker and phenotypes in the F2. For example, if the F2 individuals who carry a certain marker genotype have different phenotypes, on average, than individuals with an alternative marker genotype, that marker may be linked to a QTL.

Doebley and Stec (1991, 1993) used these methods to localize regions on the maize genome that contributed to the difference between wild teosinte and domesticated corn. Some QTLs had big effects, and the researchers wanted to know what the underlying molecular, genetic, and developmental basis was for each QTL. How can they find the actual gene (or genes) at the QTL? It turns out that they already knew of a good candidate gene that warranted further investigation.

Corn, as a vital food crop, has been studied extensively by geneticists for many decades. Many spontaneous mutants have been described. Visible mutations can behave just like molecular markers, in that they can be mapped relative to other visible or molecular markers based on how often they are inherited concordantly. The candidate

maize mutation is the *teosinte branched 1* (*tb1*) mutant, which produces long primary lateral branches tipped by tassels, a phenotype reminiscent of the teosinte ancestor.

The *tb1* mutant gene was known to be located in the same region of the maize genome as a Quantitative Trait Locus of major effect that Doebley and Stec (1991, 1993) had detected. Ultimately, they were able to show that specific differences in the upstream regulatory regions of the *tb1* gene, which they had originally detected as a QTL of major effect, were largely responsible for the domestication of maize (Wang *et al.* 1999). Their other QTLs have not yet been characterized.

The conclusions of the corn researchers were possible only because of the large number of candidate loci that were known in the system. Unfortunately, while work in maize has yielded some amazing insights about the domestication of a staple crop, it is not a natural system. Adaptation under the pressures of natural selection might be a different matter entirely. Moreover, study of a natural adaptive difference between species can be a starting point for investigating higher-level questions of ecological interactions in evolution, a direction that is very limited for maize research.

This study was designed to ask similar questions about evolutionary genetics, but in the natural system of two closely related species of yellow monkey flowers. *Mimulus guttatus* has large, showy flowers, which facilitate bee-mediated pollination. In contrast, *Mimulus*

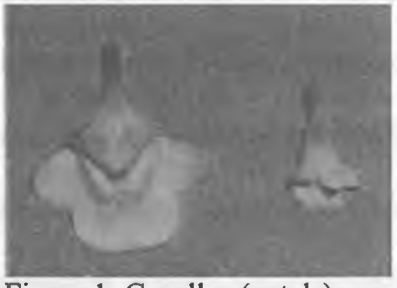


Figure 1. Corollas (petals) from *Mimulus guttatus* (left) and *Mimulus nasutus* (right). *M. nasutus* probably evolved from a *M. guttatus*-like ancestor.

nasutus is an obligate self-fertilizing species. Individuals of this species make very small flowers that typically do not open. "Selfing" is facilitated by the placement of male and female sexual organs very near each other, unlike the positioning of *M. guttatus*' sex organs. *M. nasutus* is generally interfertile with *M. guttatus* in laboratory settings. In the wild, however, even sympatric (co-existing) populations of the two species do not commonly mate.

The ancestral condition for this pair of species was probably one in which individuals commonly mated with other individuals (Vickery 1978). Modern evidence for this is that the geographical range of *M. guttatus* is much larger than any other closely related monkey flower, and encompasses the ranges of many of the derived, sister selfing species, such as *M. nasutus* and *M. platycalyx*. As a result of this primitive-derived relationship, the genetic differences between *M. guttatus* and *M. nasutus* may approximate the genetic changes that must have been necessary for the selfer to evolve out of an outcrossing ancestor.

Our lab is currently conducting a Quantitative Trait Loci study in hybrids between these species using microsatellite and Amplified Fragment Length Polymorphism (AFLP) molecular markers. Traits that are thought to contribute to the adaptive difference in flower morphology between the species were measured quantitatively in an F2 population. Many QTLs have been mapped to chromosomal regions. Unfortunately, nothing is known about what genes actually lie in these locations. The successful aspect of the investigations in maize was working backwards, from candidate loci to the QTL. Because candidate loci are not yet known in the *Mimulus* genome, I decided to begin to

try to identify some. I sought two types of candidates: first, phenotypic mutants of *M. guttatus*, which differed from wild-type in many different aspects of floral morphology, and second, molecular polymorphisms (differences) between *M. guttatus* and *M. nasutus*, as shown through analysis of their DNA sequences for specific genes.

Because one of the major differences between the two species is floral morphology, we were interested in finding the genes responsible for normal flower development. One can easily envision a scenario in which many of the evolutionary changes necessary for *M. nasutus* to evolve involved such genes.

Phenotypic mutants have not been characterized in monkey flowers to any great extent. Thus, one important part of this study involved artificial mutagenesis of *Mimulus guttatus* seeds. The mutants sought were ones that were abnormal for floral development traits.

Genes affecting floral development have been studied in several other model systems, including *Arabidopsis thaliana* and *Antirrhinum majus* (snapdragon). The floral traits they control are many, from floral organ identity (e.g., *plena, deficiens* and *agamous*; Coen and Meyerowitz 1991) to flower symmetry (e.g., *cycloidea, dichotoma*; Luo *et al* 1999).

In an attempt to find such genes in the *Mimulus* system, a large number of *M. guttatus* seeds were treated with a chemical mutagen, and the families of the second mutant generation were examined for abnormalities in floral morphology. Screens for mutants are a common way that genes' functions are elucidated. Given that natural mutation rates are generally low for important genes such as those controlling floral

development, a chemical mutagenic process increased the odds of finding mutants in the laboratory setting.

Another way in which a QTL can be characterized is if an actual gene that is known to affect floral development, rather than a phenotypic mutant, is known to exist in the same region as the QTL. Thus, in a separate line of investigation from the mutant studies, I sought genes that are known to affect floral development in snapdragon in unmutagenized individuals of both *Mimulus guttatus* and *M. nasutus*. When characters show some sort of similarity between related species, they are said to be homologous. The same can be said of genes in different species. It is presumed that, when genes show significant portions that are identical, they are most likely derived from an ancestral gene that also shared those portions, and often the two genes will be considered to be in the same gene "family." Because genes can spontaneously duplicate, "homologous" genes can even be found within one organism (paralogs).

When the sequence of base pairs that define a gene in one organism is known, that information can be used to find homologous genes in other organisms. In this *Mimulus* study, a technique known as Polymerase Chain Reaction (PCR) was used to search for homologs of the *Antirrhinum majus* (snapdragon) genes *cycloidea* (*cyc*) and *dichotoma* (*dich*), both of which are known to contribute to the non-radially symmetric phenotype of snapdragon flowers (Luo *et al.* 1999). The predicted amino acid structure of their gene product is one that is known to bind DNA; thus, they may very well be transcriptional regulators (whose function is to affect the expression of other genes). Genes that share

this specific DNA-binding region are known as TCP genes, which also include some genes in more distantly related plants. *tb1*, in fact, is a TCP gene.

The central question is, "What kinds of genes are responsible for the divergence of *Mimulus nasutus* from a *M. guttatus*-like ancestor?" The significance of the mutants detected in this study will be truly understood when it is discovered whether the loci they are mutant for are located near the same molecular markers as a QTL. The same can be said of the putative TCP genes that were detected with PCR. In either case, it is possible that when a QTL and mutant/TCP locus map to the same region, they may be the same genes, which will warrant further molecular and phenotypic study. That will help to complete the story of how changes at that locus affected the evolution of the selfer out of the outcrossing ancestor.

MATERIALS AND METHODS

Mutant screen

Seed stocks

A population of annual *Mimulus guttatus* plants on Iron Mountain (44°25'00" north x 122°07'00" west) was sampled. A line of inbred, homozygous, genetically identical plants was formed from this population by repeated self-fertilization. IM62 was the name of the line used for mutagenesis (as well as the QTL study).

Mutagenesis by EMS

I subjected 3600 genetically identical IM62 seeds to treatment with Ethyl Methane Sulfonate (EMS) as described for *Arabidopsis thaliana* by Leyser and Furner (http://www.arabidopsis.org/comguide/chap_1_plants/6_EMS_mutagenesis.html). The protocol includes directions for varying the strength of treatment, either by time of exposure or strength of mutagen. The following specific values were used: the concentration of EMS was made to be 85 mM, and the seeds were treated for four hours.

They dried for three days, were counted (for logistical potting purposes) over the next three days, and were planted a week following treatment.

Planting

The mutagenized seeds, along with 340 negative control seeds (untreated IM62 seeds), were planted at a density of 10 per 2-inch pot and kept in a greenhouse. Germination rates were assessed by counting the number of seedlings per pot, then the individuals were transplanted so that just one occupied each pot. This generation was designated M1 (Mutant 1).

Self-fertilization of M1

EMS works by altering the pairing properties of guanine (G, one of the four nucleotides in the DNA "alphabet"), but altering the properties of every guanine would kill any organism, so whether any particular guanine base is hit is an issue of chance. Therefore, it is extremely unlikely that a mutation could be caused in both copies of a given wild-type (non-mutant) gene. Thus we expected many of these mutations to be recessive, which means that they can be masked by the presence of a wild-type allele. The only way to ensure that mutants were generated whose phenotype was due solely to mutant alleles was to cross the M1 plants by themselves. For single loci, it is expected that such crosses will result in 25% mutant plants. Though in the wild *Mimulus guttatus*

plants normally outcross, artificial self-fertilization is simple and effective. Therefore, 1-4 flowers per M1 plant, as time allowed, were self-pollinated using forceps to place the plant's own pollen on its stigma. Pollinated flowers were taped. When the fruit had ripened and dried, and loose seeds were visible in the calyx, the calyx was clipped and placed in a seed envelope, marked with the individual's identification number, the date of collection, and the generation that the seeds would compose (M2). This process was used to obtain seed from 600 plants.

Planting of M2 generation

M2 seeds were planted in 4-inch square flat inserts. Up to 50 seeds were sewn per insert (to ensure that, under the risk of partial lethality, at least 10-15 plants would grow), and the populations were culled to approximately 15 plants upon germination. With 15 individuals per family, there is only a 1.3% chance that none of the plants would carry a recessive homozygote genotype (which was being sought). Higher total numbers would have overtaxed the resources available in each insert. About 9000 plants were expected, although some families showed very poor germination rates.

Screening the M2

The flowers and plants of all 600 families (a family is all the M2 progeny of one selfed M1 individual) were examined for abnormalities. Traits examined included petal

shape and number, plant and flower color, general flower size, placement of sexual organs and tissues, and general patterns of abnormality (e.g., fused stems). Observations were recorded and interesting individuals were photographed. Open-pollinated plants (presumably self-pollinated, but may be outcrossed through accidental contact with other plants) had their seed pods collected.

Segregation analysis

Seventeen noteworthy (and three "families" of IM62 negative controls) M2 families had their remaining seed stock planted at 1 seed per 2-inch pot. Up to 32 seeds were planted as supplies allowed. All 17 families had demonstrated robust mutant phenotypes in the initial screen; that is, there were several plants in the family with the same abnormality, yet an equal or greater number in the family appeared wild-type (79 families total fit this category). Only 17 families were retested for two reasons: first, at 32 pots per family, resources and time limit how many plants can be dealt with concurrently; second, the point of the experiment was to identify the mutations as being heritable in a Mendelian fashion; if it was clear that a sizable fraction do segregate as such, then the replanting of more mutant families in the future is a more reasonable proposition.

After three or more flowers had bloomed per plant, they were scored for mutant traits. Chi-squared analysis, a statistical test which assesses the probability of getting the observed data under the expectations of a given model, was performed to test for 3:1 Mendelian segregation, which should be the result if the mutant trait is caused by a

mutation at a single locus. Such mutations were the principal targets of this screen. If possible, mutants were self-pollinated to generate seed for future use. If mutants were sterile, three or more siblings within the family were self-pollinated.

PCR amplification of putative TCP genes

Extraction of DNA

DNA was extracted as in Doyle and Doyle (1987) from leaf and flower material from *Antirrhinum majus* (snapdragon, courtesy Daisies Flowers and Gifts), *Mimulus guttatus* (IM62 inbred line, see above), and *M. nasutus* (Sherar's Fall's inbred line, which underwent a similar number of inbreeding events in the greenhouse). The extraction buffer was modified CTAB; the extraction was either chloroform only or a combination of 50:50 chloroform:phenol followed by another pure chloroform extraction.

Designing of primers

PCR (polymerase chain reaction) is generally used to generate many (millions) of copies of a specific fragment of DNA from template DNA, which is generally a larger sequence. Which part of the template DNA is amplified depends on the placement of a forward and reverse primer. The majority of the DNA amplified in a reaction will be the forward primer (or a similar region of the template DNA) on one end, then the template

DNA that lies between the flanking primer regions, and then the reverse primer (or a similar region of the template DNA). Therefore, the design of primers is an important task. Möller et al. (1999) used primers designed from the *Antirrhinum* (snapdragon) cycloidea gene to amplify a putative *cycloidea* homolog in members of the Gesneriaceae (a flower family including African violets). Because *Antirrhinum* and *Mimulus* are members of the Scrophulariaceae family, yet the primers were successful in another family, we concluded that they would probably work equally well in *Mimulus*. These primers were ordered from Operon. The forward primer, 5'-ATGCTAGGTTTCGACAAGCC, was designated GcycFS; the reverse primer 5'-ATGAATTTGTGCTGATCCAAAATG, was designated GcycR, as in Möller *et al.* (1999).

Amplification by PCR

These primers were used to amplify regions from *Antirrhinum majus* (positive control) *Mimulus guttatus*, and *M. nasutus* genomic DNA, according to the following "mastermix" recipe (per 25 µL reaction): 2.5 µL 10x Taq salts, 2.0 µL 2.5 mM dNTP, 0.25 µL 33 µM GcycFS, 0.25 µL 33 µM GcycR, 0.30 µL Taq polymerase in glycerol, 18.7 µL sterile H_2O. Each sample also received 20 ng genomic DNA in 1 µL water, except for the negative control.

PCR works by a cycle of melting (using a high temperature) the complementary DNA strands apart, followed by cooling to a temperature at which the primers can anneal

(pair up) to the regions of template DNA with which they complement. A special high-temperature-tolerant polymerase (an enzyme that copies DNA from a complementary strand) then extends off the primer according to the strand that the primer has annealed to. After this has a chance to occur, the temperature is raised again so that the strands melt apart, and the process is repeated. The thermocycler protocol used for these (and all subsequent) reactions was 1) three minutes at 94°C, 2) 30 seconds at 94°C, 3) 30 seconds at 55°C (annealing temperature), 4) two minutes at 72°C, 5) GOTO step two 44 times, 6) ten minutes at 72°C, 7) 4°C until removal.

During the testing phases, four different tubes were sent through the protocol. One contained mastermix only (negative control), one contained mastermix and *Antirrhinum majus* DNA (positive control), one contained mastermix and *Mimulus guttatus* DNA, and the last contained mastermix and *M. guttatus* DNA.

To each sample, 2 µL loading dye buffer was added. Up to 20 µL of the resultant mixture was loaded into a 3% agarose gel. When the gel is placed in an electric field, DNA moves toward the positive pole (as it is negatively charged), but because smaller fragments move through the gel's matrix faster than larger fragments, the DNA separates by size. Ethidium bromide, which binds to DNA and glows under UV light, is included in the gel. When the gel is placed on a UV table, the regions where DNA of a certain size has traveled glow as a "band."

Isolation of bands

After the initial testing of the primers, the positive control was omitted to avoid possible contamination of the *Mimulus* samples. The same mix was used again on the two types of *Mimulus* DNA, as well as a negative control tube. Again the samples were run in a gel as above. Each of the two bands resulting from the first round of PCR amplification was sampled by picking a bit of the glowing gel. Each pick was resuspended in up to 100 µL warm sterile H_20. This was used as the template DNA for another amplification, with the same mastermix recipe and thermocycler protocol, so that we might obtain samples that included only one DNA fragment or the other. Generally, several replicates of each sample (of which there were four: two bands from *M. guttatus*, two bands from *M. nasutus*) were amplified to ensure sufficient DNA yields for sequencing. A 5 µL sample from each tube was run on a 2% agarose gel to verify that one band was present and alone (or, much stronger than the other band).

Sequencing of PCR fragments

Replicates whose subsamples passed this test were pooled together and purified using QIAgen's QIAquick column purification using a microcentrifuge, according to the manufacturer's instructions for higher concentration yield. The DNA concentration was measured using a Hoefer Scientific Instruments DNA Fluorometer Model TKO 100. The samples (60 ng of each) were sequenced on an ABI Prism 377 DNA Sequencer, using the

Big Dye Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA Polymerase, FS. The primer used for sequencing was GcycR.

Editing of sequence data

Because the reverse primer was used as the sequencing primer, the resulting data had to be converted to its reverse complement to be compared with the published sequences of TCP genes. (Genes are generally read from 5' to 3' in the direction that a cell's machinery uses to translate the DNA into a polypeptide of amino acids.) This was performed using the Lasergene Navigator software package's EditSeq software. The converted sequence was hand-edited to eliminate the primer regions. Sequences of other TCP genes (*dichotoma*, *cycloidea1A*, *cyc1B*, *cyc2*, *cyc3*, and *cyc4*) were downloaded from the National Center for Biotechnology Information Genbank database; also, a Gesneriaceae (the plant family that includes African violets) sequence was kindly provided by Q. C. B. Cronk. All were edited to contain only the fragment expected if the Gcyc primers had been used. EditSeq was also used to translate the *Mimulus* sequences into predicted amino acid structures, using the standard genetic code.

Most genes code for proteins, which are long strings of amino acids chemically bonded. The genetic code, which is used to convert the sequence of nucleotide bases into a sequence of amino acids, is conserved in almost all organisms. Each amino acid has at least one corresponding nucleotide "codon," which is three bases long. Each three-base codon is read as a single unit by the translational machinery, which makes the protein.

Out of 64 possible codons, three tell the machinery to stop making protein. Because of those, in a functional gene, usually one of the six possible reading frames (three going one direction, three going the other) results in a long stretch of amino acids without any stop codons, while the other five are punctuated by stop codons. This best frame is known as an Open Reading Frame (ORF). All four DNA fragments sequenced in this study demonstrated an ORF, starting at the second base and extending to nearly the end of the fragment. (It was using the ORF that we found that the fragments coded for the TCP and R domains that we were expecting.)

Alignment of sequences

Clustal X (version 1.8) was used to align the *Mimulus* sequences with other TCP genes. Depending on the alignment, certain sequences were edited by hand. A bootstrapped neighbor-joining tree (a diagram of how closely sequences are related based on their relative similarities) was also generated by the program and visualized using Njplot (bundled with Clustal X). The following comparisons were performed: full nucleotide sequences, TCP domain only nucleotide sequences, and TCP plus R domains only nucleotide sequences.

The trees show relatedness of sequences by assuming that sequences that are more similar now shared a common ancestor more recently than sequences that are more different. Thus, the nodes (branching points) represent the last point at which the

sequences at the tips of the branches had a common ancestral gene. All of the sequences sharing a node are said to be within the same clade (clades are nested and hierarchical).

The numbers at the nodes are "bootstrap" values. Bootstrapping is a technique in which a randomly generated subsample of the nucleotides within each gene, compared with the same subsample from the other sequences, is used to generate a phylogenetic tree. This is repeated 1000 times. The bootstrap values represent how many times, out of 1000, the sequences found below the node were together in their own clade, separate from other sequences. Thus, bootstrap values represent the level of confidence that we have in each grouping.

RESULTS AND DISCUSSION

Mutant screen

Toxicity of EMS treatment

The set of 1690 M1 seeds that was precisely planted at 10 seeds per pot showed a germination rate of 82.84% (1690/2040). The other 1460 M1 seeds were also planted, but more haphazardly. In a negative control flat, 272 of 340 unmutagenized seeds (80.00%) germinated. However, many M1 plants proved to be unusually difficult to self-fertilize, resulting in a relatively low number of M2 families (600/3500, 17.14%). Generally, they failed to make sufficient pollen, or pollinations resulted in small, inviable, aborted seed.

Mutants detected in M2

Out of 600 M2 families surveyed, 401 showed some sort of abnormality (66.83%). Of these, 79 families (13.17%) contained several plants in which consistent differences from wild-type controls were found. Categorizing the mutants was an arbitrary process, and some families fit into several categories. Figure 2 describes the distribution of mutants.

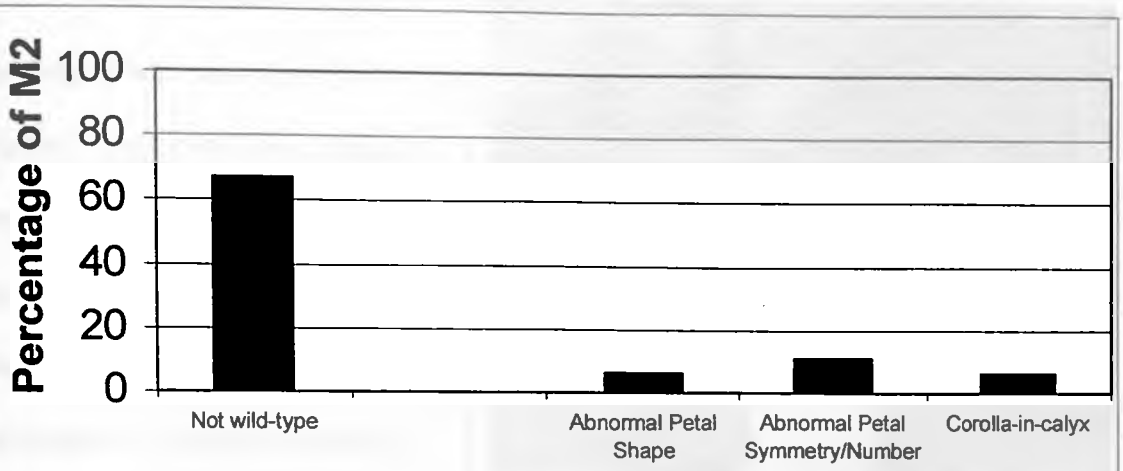
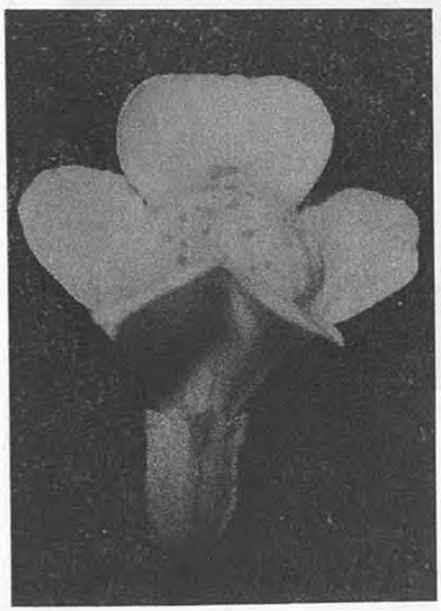


Figure 2. Distribution of mutants in the M2. The three categories at right were the most common classes of mutants. Other common phenotypes included leaf or stem abnormalities and reduced overall plant size.



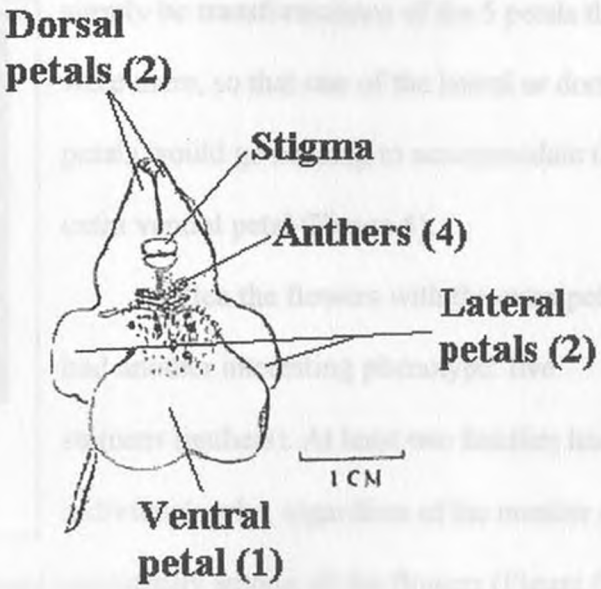


Figure 3. Morphology of a wild-type *M. guttatus* flower. The green structure at the bottom of the photograph (dorsal view) is the calyx, which will form the seed pod after pollination. Diagram from Vickery (1978), labels added.

Wild-type features of a monkey flower can be seen in Figure 3. By far the most common mutant phenotype I observed in the screen of M2 mutants was a doubling of the center ventral petal (see Figure 4). Plants exhibiting these phenotypes always had some wild-type flowers in addition to one or more abnormal ones. Frequently the only defect of the flowers would be the duplication of the petal (going from 5 to 6 petals); however, occasionally there would simply be transformations of the 5 petals that were there, so that one of the lateral or dorsal petals would go missing to accommodate the extra ventral petal (Figure 5).

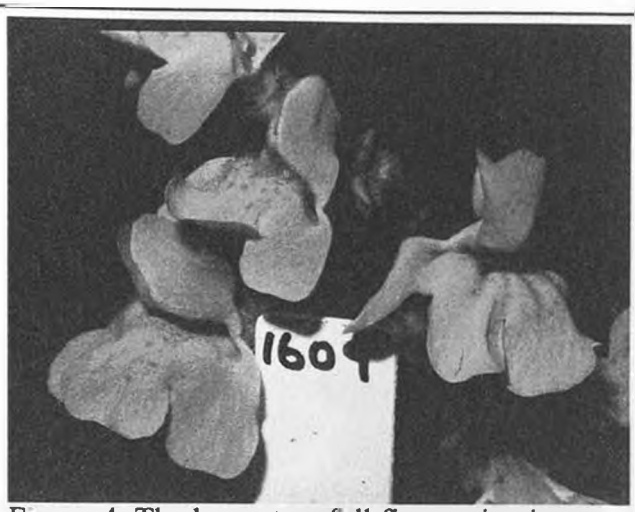


Figure 4. The lower two full flowers in view have the doubled ventral lobe.

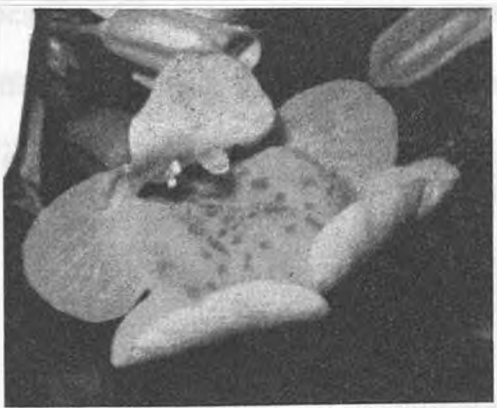


Figure 5. This flower only has one dorsal petal, but also possesses an extra ventral petal.

Often the flowers with the extra petal had another interesting phenotype: five stamens (anthers). At least two families had individuals who, regardless of the number of petals for a given flower, had five anthers consistently among all the flowers (Figure 6; however, these were not recovered in the segregation study, see below). The locus controlling such a trait would be of special interest. First of all, the properties of the

anther are very important to the difference in rates of self-fertilization between *Mimulus guttatus* and *M. nasutus*, making this a plausible candidate locus for the QTL study. Second, developmental studies in *Antirrhinum majus* (snapdragon) have shown that, initially, five stamens begin developing, but the dorsal one has its development arrested early (Luo *et al.* 1996), which corresponds to the location of the extra anther in the mutant *M. guttatus*. Thus it is likely that the mutation affects the monkey flower's ability to arrest the development of this structure, rather than causing its spontaneous generation in the wrong place.

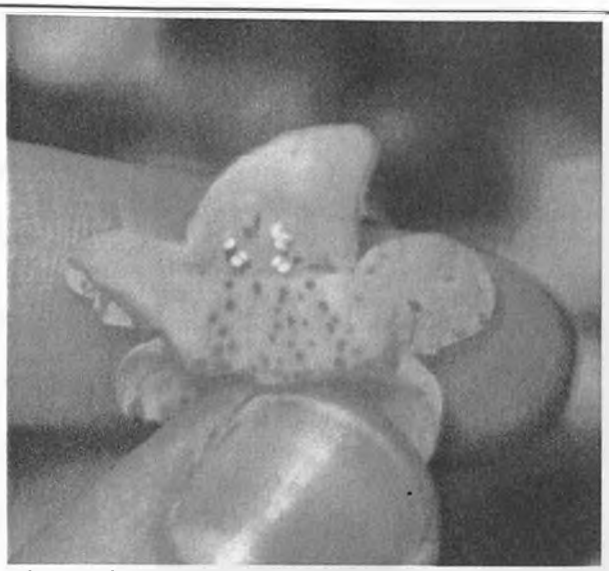


Figure 6. A mutant with five anthers, two on the left and three (one of which is actually dorsal) on the right.

Figure 7. This family included members that made some wild-type flowers and some that were missing both lateral petals.

Loss of a petal was also a common theme in the M2. Generally, when a dorsal petal was missing (as in Figure 5), the other petals were located at greater angles to each other to accommodate. However, mutants of the lateral petals were also found

(Figures 7-9), and their absences simply left vacancies in the circumference of the flower. While sometimes this trait did not show up in all flowers of a plant (Figure 7), the mutants in some families were missing both petals consistently (Figure 8). Others were consistent in missing only the right or only the left lateral petal (Figure 9). Lastly, the mutants of one family were consistent only in missing at least one lateral petal (sometimes both), but which petal was missing was not uniform.

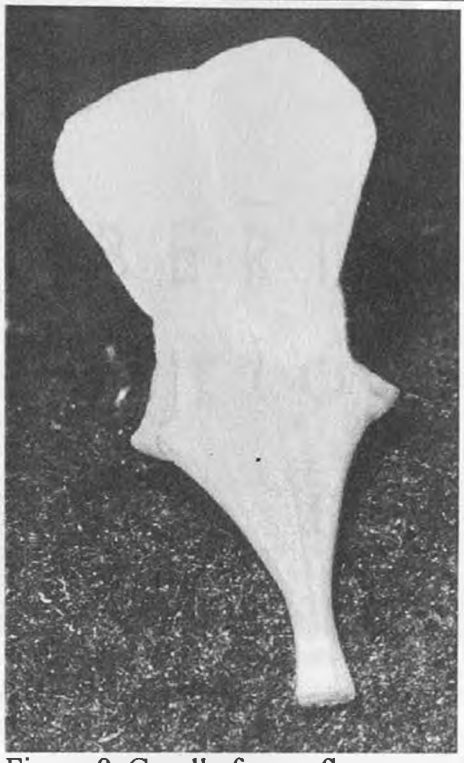


Figure 8. Corolla from a flower missing its lateral lobes (view of ventral side).

Unfortunately, these types of mutants were not studied for segregation of mutant alleles. If there is a genetic basis for these mutations, however, it indicates that different signals specify the action at the right and left sides of a developing flower. Whether these are different genes or simply the effects of different mutation alleles of the same gene can be determined by assessing the phenotypes of asymmetrical

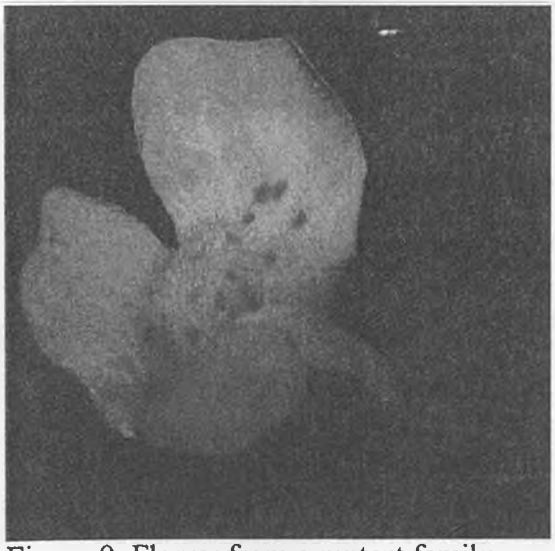


Figure 9. Flower from a mutant family missing the right lateral petal consistently.

families crossed together.

Another common occurrence in the mutants was that some flowers simply made petals with bizarre shapes. Sometimes these were abnormal but at least uniform within and between flowers of a plant (Figure 10). On other occasions, the strange shapes differed within and between flowers (Figure 11), indicating a more general loss of petal-shape control.

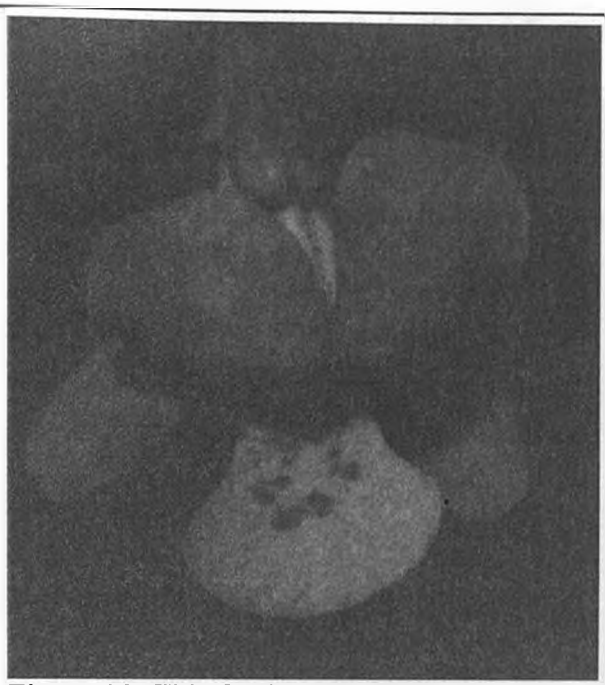


Figure 10. This family's mutants had an abnormal, but consistent, shape to the ventral and lateral petals.

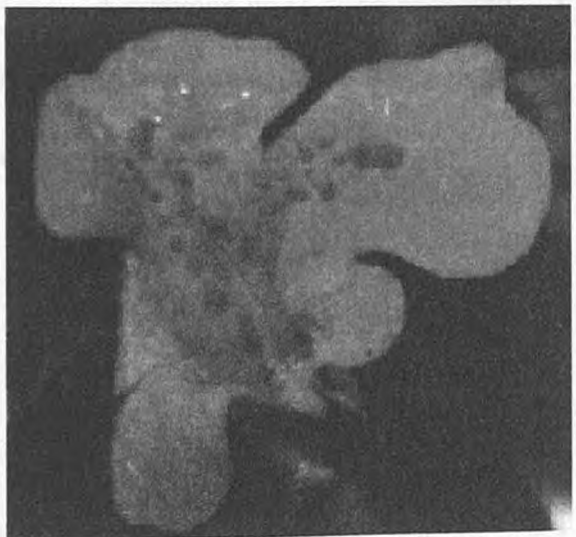


Figure 11. The corolla of this flower was completely misshapen.

One family of mutants had a phenotype unlike any of the others. In the mutants of the family, the anthers (male sex organs) in every flower were replaced by either strips or very narrow cones of corolla (petal) tissue (Figure 12). These replacements were uniform and precise: exactly four, exactly where the anthers should be, and in every

flower of affected plants. See the Segregation Studies section (below) for a discussion about this mutant's similarity to floral mutants in other organisms.

Many times the bizarre shapes of flowers were clearly the result of more than one flower trying to develop at a time. Sometimes they grew into two nearly-complete flowers within one calyx (Figure 13), but usually it was just one large circle of petals, often with extra stigmata, ovaries, and many extra stamens (Figure 14).

A common phenotype observed in mutants was replacement of a region

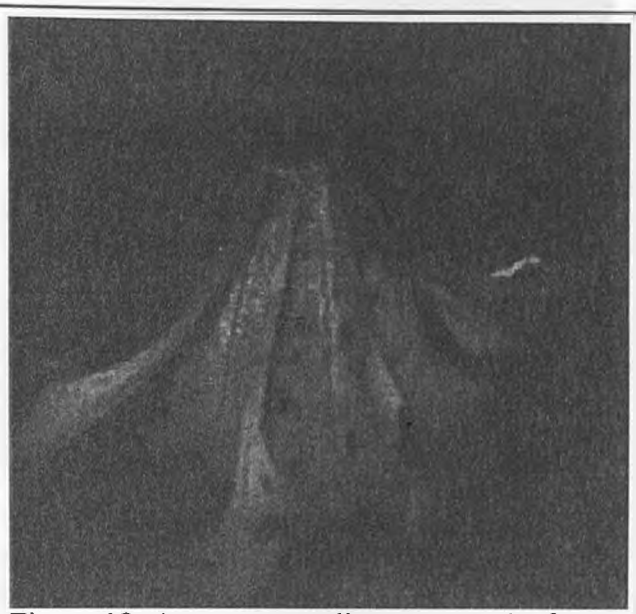


Figure 12. A mutant corolla cut open; the four structures of corolla tissue replace the anthers of a wild-type flower.

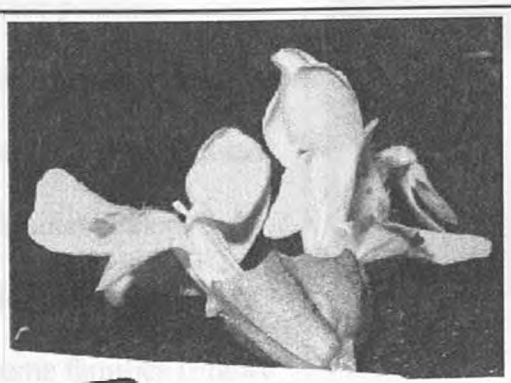


Figure 13. Two complete corollas growing out of one calyx.

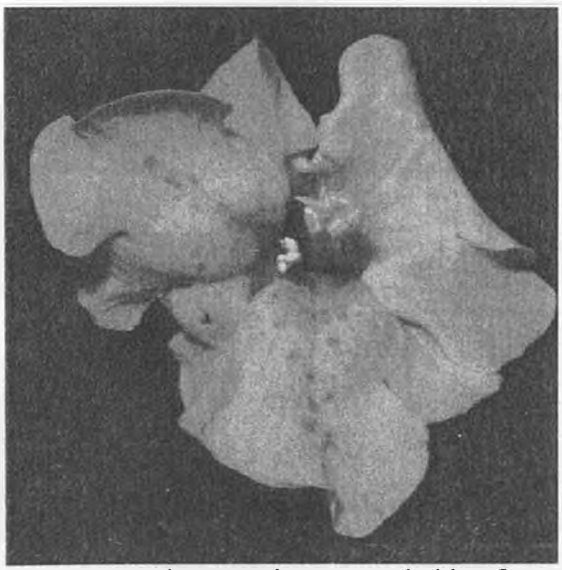


Figure 14. The complete ventral side of at least two flowers grew as a single corolla structure. Note also that the stigma has several extra lobes (wild-type is two).

of the calyx with corolla-like (petal) tissue. This is observed at low frequency in our unmutagenized greenhouse populations (see the control population data in Table 1). However, some M2 families exhibited higher-than normal frequency of this phenotype (Figure 15). In some cases this was so pronounced that corolla-like properties, like hairs or spots, appeared on that region of the calyx.

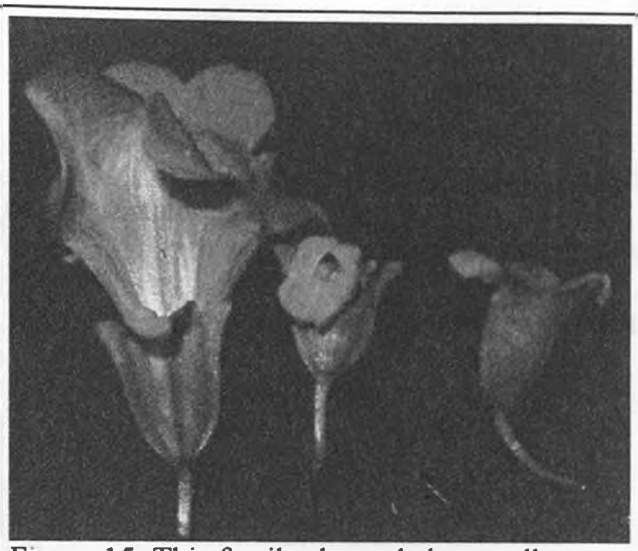


Figure 15. This family showed abnormally strong "corolla-in-calyx" phenotype, in which a region of the calyx is replaced with corolla tissue.

Obviously, the mutagenesis of these plants was not able to just target mutants of floral morphology; many other mutations were possible. Color changes were a theme of some of the mutant families. Paler flowers were seen in some families (Figure

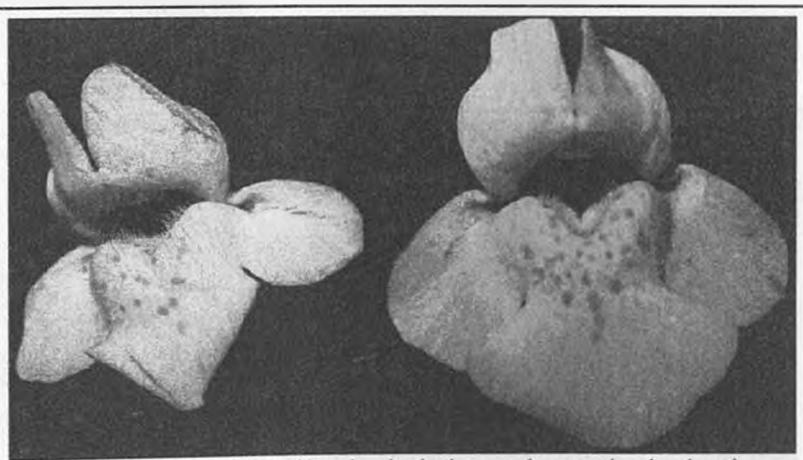


Figure 16. Some families included members who had paler yellow flowers (left) than wild-type siblings.

16). Also, some families showed streaking of near-whiteness in leaves, stems, and calicies (Figure 17).

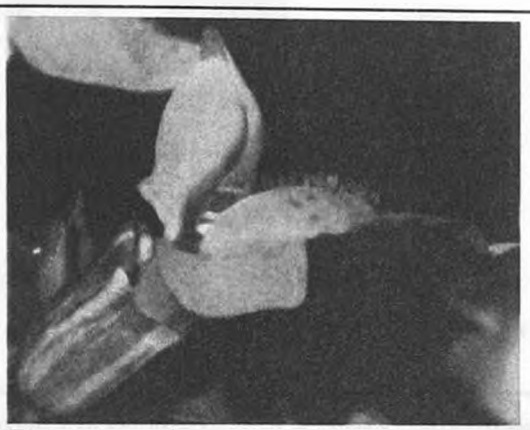


Figure 17. The calicies of this plant had strong streaks of paler green color.

General size of the plants was also affected by the mutagenesis. One family of plants grew to such sizes (Figure 18) that we were concerned about contamination, although the replanting success of the phenotype makes this unlikely (see Segregation Studies). The predominant size differences, though, were in families that had a few plants smaller than the rest. In some cases these were normal-shaped and just smaller and thus it was difficult to say confidently that they are actually

Figure 18. Most plants in this family were much taller than negative controls, and the flowers were also markedly larger.

Figure 19. The mutant (left) had much smaller flowers than its wild-type sibling (right).

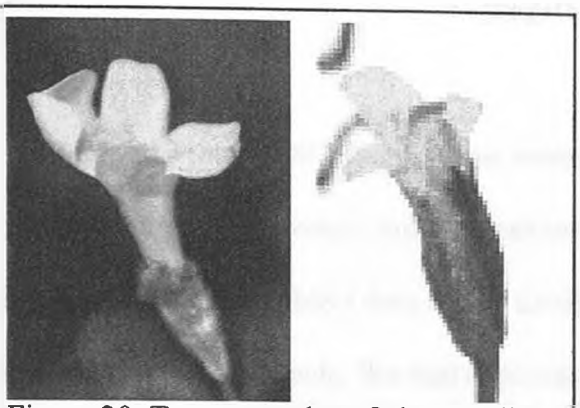


Figure 20. Two examples of abnormally small flowers.

mutants (Figure 19). In several other cases, however, the smaller plants had other blatant abnormalities (Figure 20).

Overall, the plants in the M2 screen were generally sickly. Beyond all these mutations already described, many other abnormalities occurred. Sometimes only one occurrence of a really malformed flower or stem was present in a family; other times, there were some plants with many more problems than others. Many times, excess flower tissue caused the corolla or calyx to rip, a flower was replaced by a chaotic mass of tissue, stems were fused, or any one of many other abnormalities occurred (examples in Figure 21).

Figure 21. General-integrity mutants. A second flower grew from an abnormal calyx (left). Occasionally a mutant would make a jumbled ball of all kinds of tissues (right).

Segregation studies

The 17 mutant M2 families that were replanted for segregation studies all fit the "robust" category. However, not all expected mutant phenotypes were found in the segregation studies. Table 1 lists the 13 families (four had not bloomed by the time of this writing) and three controls. We had expected that most mutations would show up in approximately ¼ of the seeds planted, as Mendelian ratios predict for phenotypes due to a recessive allele. As Table 1 shows, this hypothesis could not be rejected for 11 of the 13 families. However the sample sizes (approximately 25 to 30 plants per family) prevents us from being certain that the recessive model is the only one that cannot be rejected.

One of the families, 58, that failed the chi-squared test for a recessive allele hypothesis did not fail when the same test was applied using a dominant hypothesis (that mutants should represent ¾ of the progeny). This should have been noticed in the M1 parent, which was probably heterozygous for the mutation; however, the mutation does not manifest itself in any abnormalities except for thick stems and unusually large flowers, so it is possible that I did not notice it.

It is important to keep the segregation data in perspective. Clearly, the mutagenesis adversely affected the overall integrity of the plants. Furthermore, this was increased in the M2, indicating that it was due to segregation of recessive deleterious alleles. My inability to recover certain mutants at all in the second planting of the M2 indicates that specific phenotypes I identified in the first planting were not necessarily due to a gene that solely affects that character. Perhaps what I initially observed were

Family	Mutant Phenotype	Germ. rate	Plants Scored	Observed "Mutants"	X^2	Reject model?
IM control	corolla tissue in calyx	100.00%	32	2	6.000	yes
IM control	corolla tissue in calyx	84.38%	27	2	5.069	yes
IM control	calyx spotting	87.50%	28	4	1.714	no
58	huge plants	28.13%	9	6	8.333	yes♦
502	top half of corollas in shreds	78.13%	25	7	0.120	no
578	side petals perpendicular to center*	50.00%	16	2	1.852	no
1280	one jumbled flower per plant*	78.13%	25	3	2.253	no
1307	strikingly small	78.13%	25	3	2.253	no
	double flowers		25	1	5.880	yes
1406	corolla tissue in calyx*	68.75%	22	4	0.545	no
1422	top half of corollas in shreds*	93.75%	30	12	3.600	no¤
1554	exploded calyx, extra anthers	25.00%	8	2	0.000	no
	veins greener than surrounding leaf		8	1	0.667	no
1576	one flower each w/ dbl bottom petal	90.63%	29	2	5.069	yes
1607	three buds/leaves per node	76.92%	20	3	1.067	no
1971	each has one flower w/ a split corolla	90.63%	29	3	3.322	no¤
3035	one flower each missing top petals*	71.88%	23	2	3.261	no¤
3308	*plena*-like replacement of anthers	68.00%	17	2	1.588	no

Table 1. Segregation studies of mutant families. The model tested was that of a segregating recessive allele at a single locus. Rejection of the model occurs when the chances of getting these data (p) are 5% or less.

* "Mutant" phenotype not observed in the original screen.

♦ A model of a segregating dominant allele cannot be rejected.

¤ $0.1 < p < 0.05$, thus the difference between the model's expectation and the observed data is marginally significant and acceptance of the model is dubious.

effects of genes that affect overall plant health. An experiment designed to compare the frequency of plants that had a specific abnormality with the frequency of plants showing any abnormality would explore this possibility.

It is noteworthy that just three families out of 13 exceeded the number of mutants expected under a model of a segregating recessive allele. We suspect that a primary reason for this is that many mutations are pleiotropic (they have effects on several different traits of the plants). One of these effects could be partial lethality, when mutant seeds simply have a lower germination rate than wild-type sibling seeds. Thus, if the seeds that harbor the morphological mutation are also the seeds subject to lower germination rates, the result would be a smaller number of mutants than expected. This is

certainly compatible with the data, in which the negative control plants had higher overall germination rate than 10 of the mutant families.

The low mutant count in many families could also be a function of "incomplete penetrance." This term refers to a situation in which organisms with a particular genotype simply have an increased (but not 100%) chance of experiencing the mutant phenotype. Some plants in the M2 segregation studies have not made as many flowers as their siblings had the opportunity to in the initial M2 screening. On a per-plant basis, we would expect that only after a sufficient number of flowers have bloomed on a mutant plant will it have more abnormal flowers than a control plant (or non-mutant sibling), a condition that we have not applied to the second planting.

One last possible reason that consistently fewer mutants were observed than expected under a single-locus, segregating recessive allele model is that the model is wrong. The data do not support this, inasmuch as the chi-squared tests were not able to rule out the simplest model. This test, however, is incapable of "confirming" that any model is correct; it can only reject models that are probably incorrect. After single-locus models are exhausted (all families only fit one, if any, single-locus model, usually of a recessive allele), one might turn to a model invoking two loci. If, for example, the mutant phenotype is observed only in individuals who are homozygous recessive at two mutant loci, only 1/16 of the progeny in the M2 should demonstrate the mutant phenotype. This was not the major focus of my analysis for several reasons. First, the analysis and future utility of single-locus mutants is abundantly easier. Second, the simplest explanation, of a single-locus recessive allele, cannot be rejected. Lastly, the scenarios leading to a model

more complex than a single-allele are exponentially more improbable than the simple scenario.

Many mutants were identified in the M2 that are candidate loci for the QTL study. In the 13 families that were scored in the segregation study, four show particular promise. The first two, 502 and 1422, have almost identical phenotypes of corollas that have their distal (outermost) portions shredded into strips (similar in morphology to Figure 22). Therefore, any future study should include mating between mutants from each family. If they are mutant at different loci affecting the same trait, they should complement and show wild-type phenotype, but if they are mutants in the same gene, most likely they will not complement, and only mutant progeny will be observed.

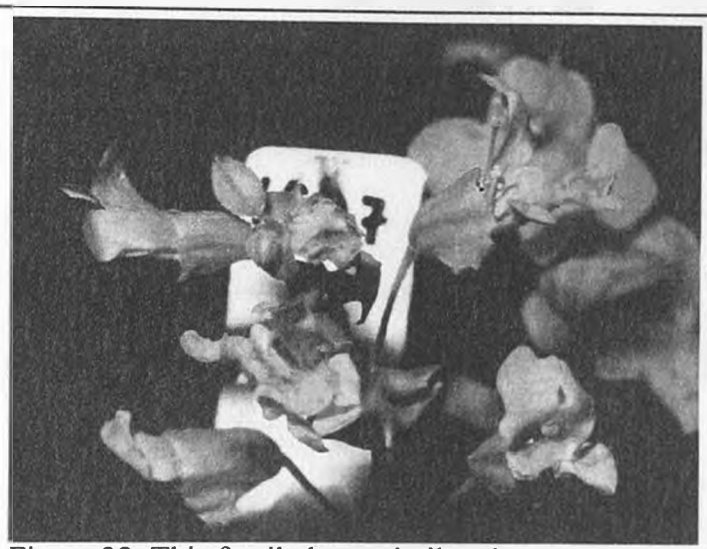


Figure 22. This family has a similar phenotype to families 502 and 1422, in which corollas that emerge are torn and disorganized.

Despite the fact that the mutant does not resemble *M. nasutus* any more than it does *M. guttatus*, all loci should be considered candidates for future analysis. This is because of the many ways in which a gene can be altered slightly by mutation, so what I observe might occur at virtually the same locus as a change that happened in the evolution of *M. nasutus*. For example, perhaps the phenotype in these two families is

caused by turning off a gene entirely, one that normally tells the flower what cellular structure the petals should be composed of. A more subtle change could tell *M. nasutus* to make smaller flowers. At this point, there is no reason to exclude possibilities like this.

The mutation(s) can be mapped relative to the QTLs and molecular markers by crossing mutants with a wild-type *Mimulus nasutus*, self-fertilizing the F1, and genotyping the mutants in the F2. If they map to an area of interest (at a QTL), a search for homologous mutants in *Antirrhinum* or other model plant systems will be relevant and warranted. Strategies of molecular investigation would then be employed.

The most striking difference between *Mimulus guttatus* and *M. nasutus* is, of course, size. Thus the family (58) that contained individuals that are noticeably bigger presents an excellent candidate locus. In particular, if it segregates faithfully (if its progeny can be classified unambiguously as "small" or "large," rather than size being continuously distributed among the progeny) in follow-up studies, it will be very valuable. In many organisms, loci that actually affect size are extremely hard to characterize because so many genes affect the trait. Moreover, just because the induced mutation has increased the size of some progeny certainly does not mean that other alleles at the locus could not cause smaller offspring.

The fourth promising mutant, 3308, is exciting in spite of its lower-than-expected recovery in the M2 segregation study and the limitations of being sterile when mutant. That is because it shows a remarkably similar phenotype to the *plena* mutant in *Antirrhinum majus* (Figure 23), which in turn has close similarity to the *agamous* mutant from *Arabidopsis thaliana* (Riechmann and Meyerowitz 1997). When a gene of this type

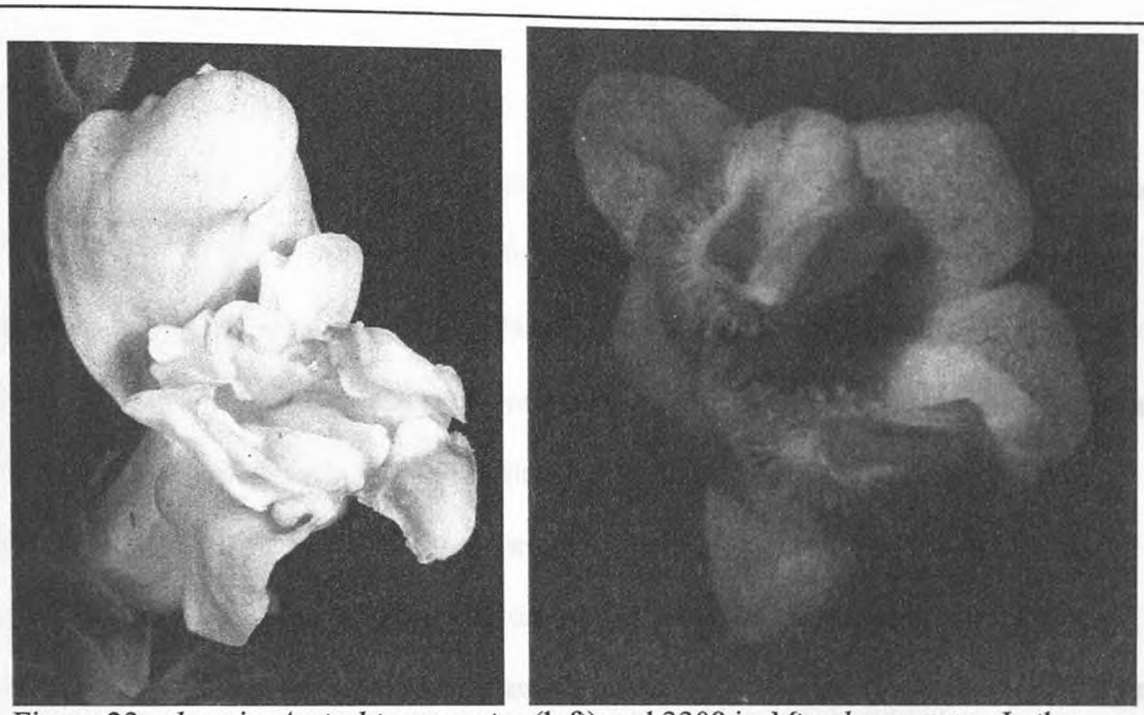


Figure 23. *plena* in *Antirrhinum majus* (left) and 3308 in *Mimulus guttatus*. In the location where the flower normally makes the male sex organs, the anthers and filaments, it instead makes structures of petal tissue. (*plena* photograph from Lönnig 1997)

is mutant, the stamens (male sex organs) are converted to strips or cones of petal tissue, and the pistil (female sex organ) is converted to variable structures, often the sepals. This mutant may allow us to ascertain whether the gene responsible was important for the evolution of *Mimulus* nasutus. One current theory argues that changes in transcription factors in plants are excellent candidates to effect evolutionary change (Doebley and Lukens 1998), and the *plena* mutation is believed to be a transcription factor (Bradley *et al.* 1993). If we can find that it did influence evolution in our system it will support the theory.

plena, and a homolog (a gene in one species that is related functionally and evolutionarily to a gene in another species) from *Arabidopsis thaliana* called *agamous*, are known to be a "C-class" gene under the ABC model of floral development (Coen and Meyerowitz 1991). This model has been developed to explain how the placement of the proper floral organs is accomplished. Briefly, flowers can be viewed as a set of four concentric circles (like a bull's eye), known as whorls. The difference in gene expression between these whorls specifies their identities. In monkey flowers, the outermost whorl normally forms the calyx. Next in is the petals, then the anthers, then the stigma. The ABC model predicts that each gene is expressed in two adjacent whorls: A in the calyx and the petals, B in the petals and the anthers, and C in the anthers and the stigma (Figure 24). Thus four fates can be specified: if only A is expressed, the whorl becomes the calyx; if A and B are expressed, the whorl becomes the petals, and so on. When a C-class gene, such as *plena*, is mutant, the third whorl assumes the identity of the second, because each is affected primarily by the B-class genes. This is clearly what may be occurring in mutant 3308.

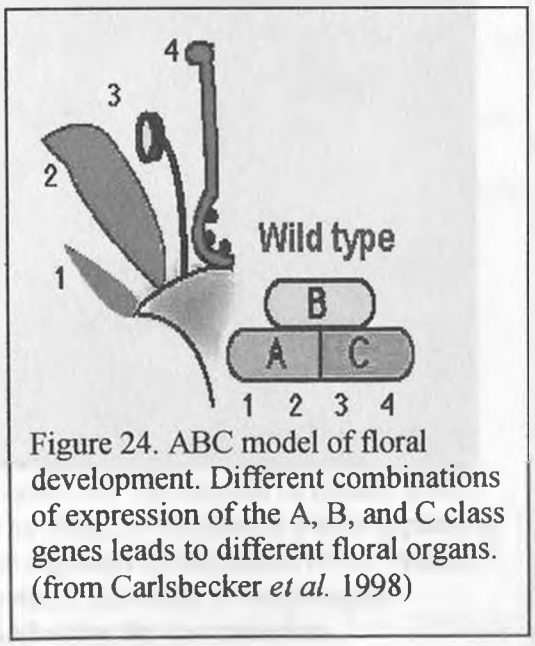


Figure 24. ABC model of floral development. Different combinations of expression of the A, B, and C class genes leads to different floral organs. (from Carlsbecker *et al.* 1998)

There are several routes to analyzing this mutation. Because 2/3 of the siblings of the sterile mutants are predicted to be heterozygous, if several are self-fertilized, at least one should result in a 3:1

segregation of the mutation in its offspring. The crossing scheme would be a bit more complicated than for a simple, fertile mutant, but the mutation could still be mapped by similar techniques.

Because *plena* and *agamous* sequences are both available (Bradley *et al.* 1993; Yanofsky *et al.* 1990), the technique of PCR amplification of homologous fragments (as was performed for *cycloidea/dichotoma* homologs) in *Mimulus* is an option for future study. The mutant can be sequenced along with the wild-type DNA of both species of monkey flower. If a polymorphism is detected between the sequence of *Mimulus guttatus* and *Mimulus nasutus*, the locus can be mapped (see Future work: Mapping *Mimulus* loci A and B). But, if the sequences are the same between *M. guttatus* and *M. nasutus*, the locus can still be mapped because we have still have a [phenotypic] polymorphism at our disposal.

There are many more mutants that warrant segregation study. Unfortunately, these are time-consuming studies because of the large family sizes and precision required. With mutant seed already produced, however, future researchers start several months of full-time work ahead of where I did. I am confident that many more loci can be map-worthy just a few months after more segregation studies are initiated.

PCR screen for TCP genes

Using genomic DNA from both *Mimulus guttatus* and *M. nasutus* as template, two PCR products differing in length by several hundred nucleotides (in base pairs, bp)

were detected (Figure 25) and sequenced. In *Mimulus guttatus*, the fragments were 758 bp ("MG-A") and 573 bp ("MG-B"). In *Mimulus nasutus*, similar fragments were found, of lengths 755 bp ("MN-A") and 581 bp ("MN-B"). Alignments are shown in Figures 26 and 27. Predicted protein structures and alignments are shown in Figures 28 and 29. One caveat: toward the 3' end of the sequences, the sequencing signal became more ambiguous; any data from that end must be considered warily.

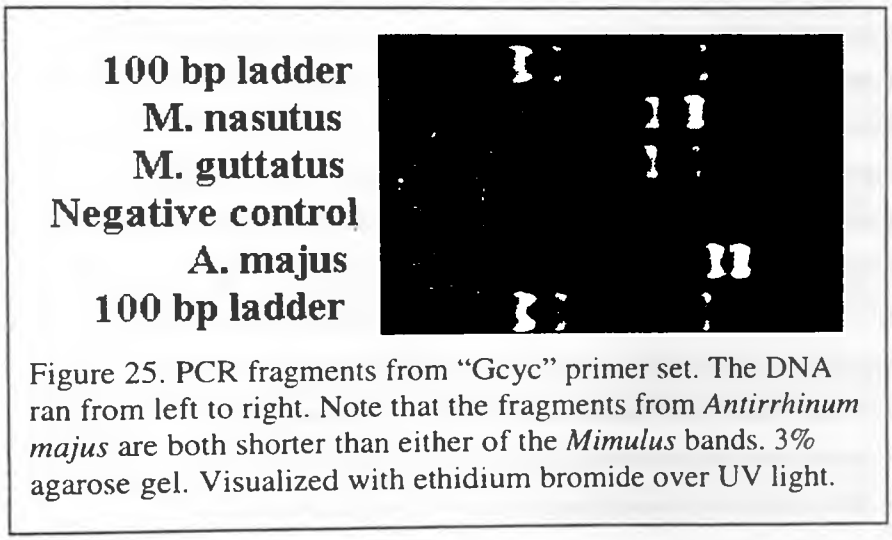


Figure 25. PCR fragments from "Gcyc" primer set. The DNA ran from left to right. Note that the fragments from *Antirrhinum majus* are both shorter than either of the *Mimulus* bands. 3% agarose gel. Visualized with ethidium bromide over UV light.

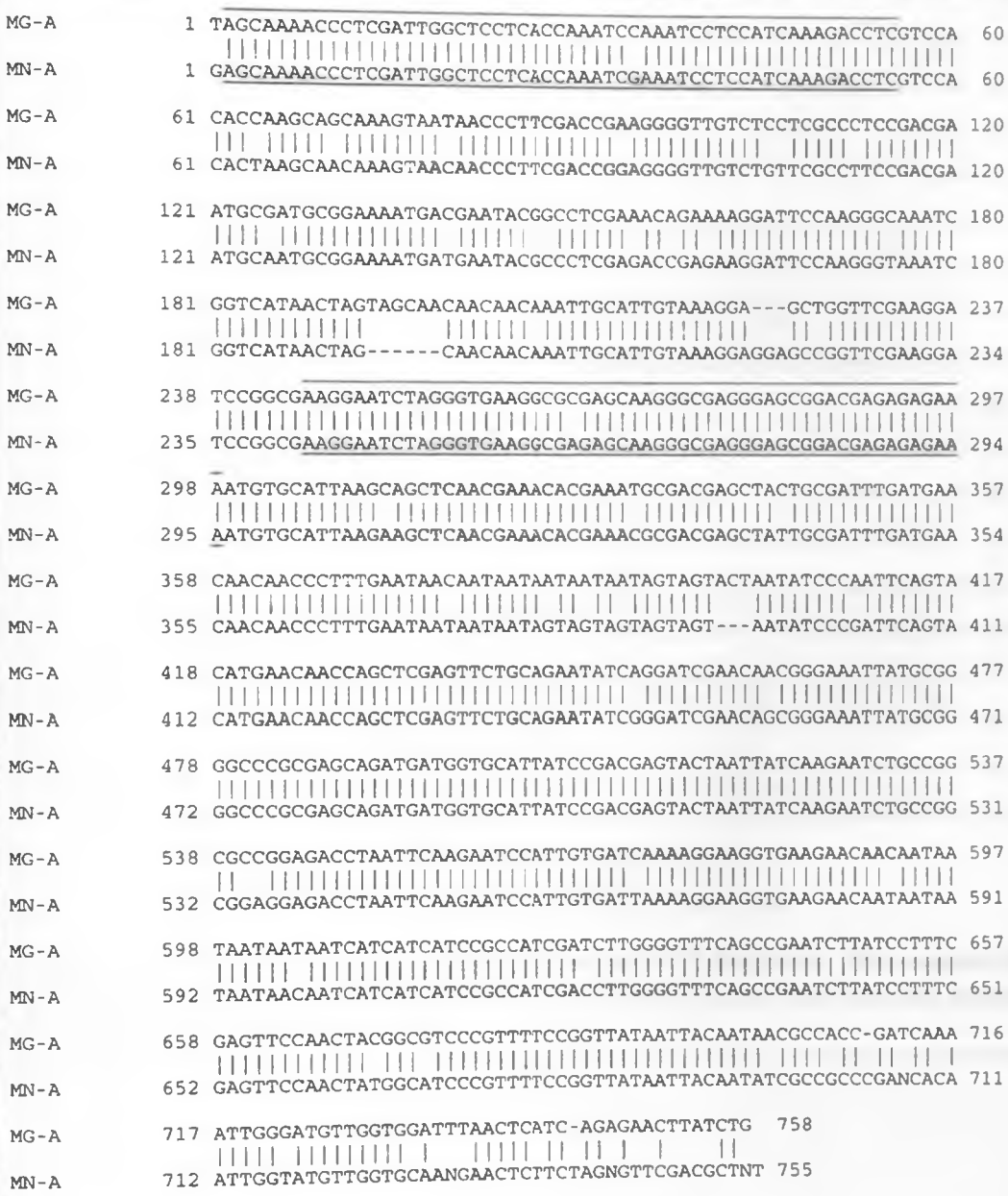


Figure 26. Nucleotide sequences of A, the longer PCR fragment. The sequences from the two species are 92% identical. The TCP (first) and R (second) domains are marked with lines.

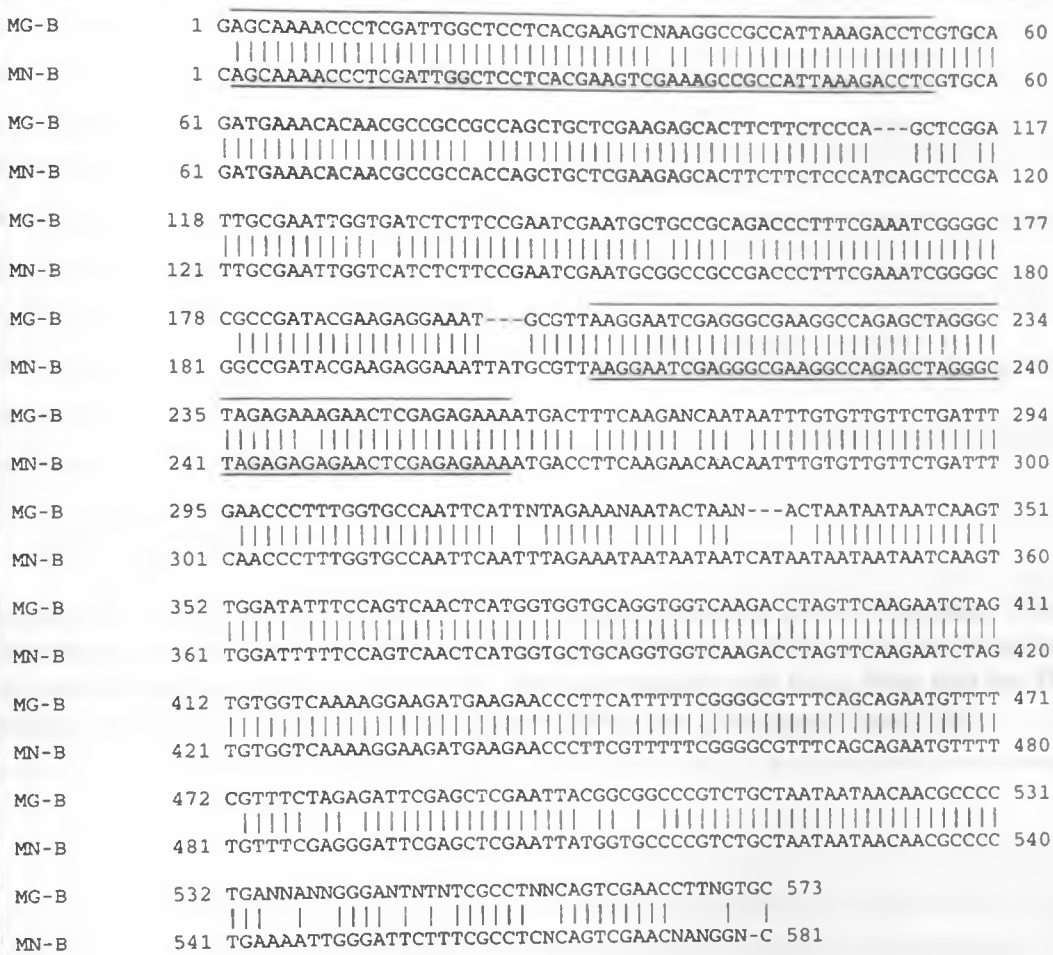


Figure 27. Nucleotide sequences of B, the shorter PCR fragment. The sequences from the two species are 92% identical. TCP (first) and R (second) domains are marked with lines.

```
MG-A.protein     1 SKTLDWLLTKSKSSIKDLVHTKQQSNNPSTEGVVSSPSDECDAENDEYGLETEKDSKGKS  60
                   ||||||||||||||||||||||||||||| |||.||||||:||||||.||||||||||
MN-A.protein     1 SKTLDWLLTKSKSSIKDLVHTKQQSNNPSTGGVVCSPSDECNAENDEYALETEKDSKGKS  60

MG-A.protein    61 VITSSNNNKLHCKG-AGSKDPAKESRVKARARARERTREKMCIKQLNETRNATSYCDLMN 119
                   ||||  |||||||| |||||||||||||||||||||||||||:|||||||||||||||
MN-A.protein    61 VITS--NNKLHCKGGAGSKDPAKESRVKARARARERTREKMCIKKLNETRNATSYCDLMN 118

MG-A.protein   120 NNPLNNNNNNNNSSTNIPIQYMNNQLEFCRISGSNNGKLCGAREQMMVHYPTSTNYQESAG 179
                   ||||||||.. ||:|||||||||||||||||||.|||||||||||||||||||||||||
MN-A.protein   119 NNPLNNNNSS-SSSNIPIQYMNNQLEFCRISGSNSGKLCGAREQMMVHYPTSTNYQESAG 177

MG-A.protein   180 AGDLIQESIVIKRKVKNNNNNNNNHHHPPSILGFQPNLILSSSNYGVPFSGYNYNNATD-- 237
                   .|||||||||||||||||||||||||| ||||||||||||||:|||||||| |:
MN-A.protein   178 GGDLIQESIVIKRKVKNNNNNNNNHHHPPSTLGFQPNLILSSSNYGIPFSGYNYNIAARXQ 237

MG-A.protein   238 -QNWDVGGFNSSENLS 252
                      |    |
MN-A.protein   238 LVCWCXELF^       246

. = similar hydrophobic properties
: = similar molecular mass
^ = stop codon
```

Figure 28. Amino acid sequences of A, the longer predicted protein fragment. The sequences are 90% identical. The TCP (first) and R (second) domains are conserved between *Mimulus* species at this level; they are marked with lines. Note that the TCP domain of A differs from that of B (Figure 29) at two positions (13 and 14).

```
MG-B.protein     1 SKTLDWLLTKSKAAIKDLVQMKHNAAASCSKSTSSP-SSDCELVISSESNAAADPFEIGA  59
                   ||||||||||||||||||||||||||||#|||||||| ||||||||||||||||||||||
MN-B.protein     1 SKTLDWLLTKSKAAIKDLVQMKHNAATSCSKSTSSPISSDCELVISSESNAAADPFEIGA  60

MG-B.protein    60 ADTKRK-CVKESRAKARARARERTREKMTFKXNNLCCSDLNPLVPIHFRNNTNTNNN-QV 117
                   |||||| |||||||||||||||||||||| |||||||#||||||#||||:| ||| ||
MN-B.protein    61 ADTKRKLCVKESRAKARARARERTREKMTFKNNNLCCSDFNPLVPIQFRNNNNHNNNNQV 120

MG-B.protein   118 GYFQSTHGGAGGQDLVQESSVVKRKMKNPSFFGAFQQNVFVSRDSSSNYGGPSANNNNAP 177
                   |#|||||.|||||||||||||||||||||||||||||||||||||:|||||||||
MN-B.protein   121 GFFQSTHGAAGGQDLVQESSVVKRKMKNPSFFGAFQQNVFVSRDSSSNYGAPSANNNNAP 180

MG-B.protein   178 X^             178

MN-B.protein   181 ENWDSFASQSNXG 193

: = similar hydrophobic properties
# = similar molecular mass
^ = stop codon
```

Figure 29. Amino acid sequences of B, the shorter predicted protein fragment. The sequences are 92% identical. The TCP (first) and R (second) domains are conserved between *Mimulus* species at this level; they are marked with lines. Note that the TCP domain of B differs from that of A (Figure 28) at two positions (13 and 14).

Evidence that A and B are functional genes

Each fragment from both species contained regions of high similarity to the two domains that are typical of TCP genes (the TCP domain and the R domain; Cubas *et al.* 1999). The longer fragments had nearly 200 bp between the TCP and R domains (MG-A: 189 bp; MN-A: 186 bp). The shorter fragments also had shorter intervening segments between domains (MG-B: 147 bp; MN-B: 153 bp). The remainder of the difference between the two size fragments is found after the R domain.

These two domains are known to exist in functional genes in *Antirrhinum majus*, *Zea mays*, *Linaria vulgaris*, and *Arabidopsis thaliana* (Luo *et al.* 1996, Doebley *et al.* 1997, Cubas *et al.* 1997, and Cubas *et al.* 1999). The TCP domain is predicted to form a non-canonical basic Helix-Loop-Helix secondary protein structure (proteins have their functionality in the shape they form when they fold up), which is most likely involved in binding to DNA, probably as a transcription factor (Cubas *et al.* 1999). The function of some of these genes is known; for instance, *cycloidea* and *dichotoma* in *A. majus* are required for proper establishment of floral asymmetry, and without them the

Figure 30. The corolla of mutant 2925. The shape of the lateral petals appears to be replaced by the shape of the ventral petal, a trait seen in snapdragons harboring the *cyc* mutation.

flowers are "peloric," showing radial symmetry (Luo *et al.* 1999). A clue to their action is that they negatively regulate the effects of another gene affecting floral asymmetry, *divaricata* (Almeida *et al.* 1997). The predicted DNA binding properties of TCP genes certainly offers a possible mechanism for this effect.

When *cycloidea* is mutant on its own, the lateral petals assume a ventral identity, giving the flower a semipeloric shape (Luo *et al.* 1999). At least one of the mutants from the screen of M2 *Mimulus guttatus* exhibited a similar phenotype of transformed lateral petals (2925; see Figure 30). Because distinct 2-lipped flowers characterize the entire Schrophulariaceae family (to which *Mimulus* and *Antirrhinum* belong) (Hyam and Pankhurst 1995), we suspect that a mechanism similar to what is known in snapdragon would cause the phenotype in *Mimulus*.

There are several arguments in support of the idea that the fragments I amplified are part of extant, functional genes. Comparison of homologous sequences between *Mimulus* species (e.g., MG-A versus MN-A) yields several clues about the current state of these sequences in *Mimulus*. Upon simple examination, we have no evidence of any impediments to the expression of these fragments. That is, the only detectable stop codons are far down at the 3' end of MN-A and MG-B, and may not actually exist at all (as artifacts of poor sequence resolution). Though the lack of a stop codon makes functionality at least feasible, on its own it is not overwhelming evidence that we are dealing with two functional genes.

The other arguments for functionality are possible only because we have, for A and B, two closely related sequences. The basic premise for all these arguments is that, at

every polymorphic location in the sequence, the data from either *M. guttatus* or *M. nasutus* represents the ancestral state. The other species' sequence, therefore, represents a derived state as a result of mutation. Ever since their lineages split, the sequences have been evolving independently. Thus, the situation is not one in which the entire (for example) MN-B sequence is ancestral. But, at every point where it differs from MG-B, we assume that one evolved and the other probably did not.

Given that premise, the next line of evidence that these are functional genes is the length of the gaps (which could have been either insertions or deletions, but I will refer to all events as "insertions" for simplicity). At all points (until near the 3' end) where the alignment suggests that bases were inserted, they occur in multiples of three.

It is logical that a functional gene would grow or shrink only in multiples of three. Though an amino acid is added for each triplet (and possibly an adjacent amino acid changes), this kind of evolution preserves the reading frame (so that all amino acids following the mutation remain the same). Presumably this cannot occur in highly conserved domains, such as TCP or R, that have specific functions and do not tolerate change well. But, even in an area of the sequence that will tolerate change on its own, if the reading frame is not preserved, all amino acids downstream from the change will also change. That is trouble if a critical domain is downstream. Therefore, in functional genes, we expect not to see gaps in highly conserved domains, and any gaps in the more variable regions must preserve the reading frame. That is exactly what is seen in both comparisons.

We can attribute one possible cause to several of these events; they appear to be insertions or deletions of a three-base unit that is repeated several times in the flanking DNA immediately adjacent to the gap. DNA regions such as this are known as microsatellites, and the insertion or deletion of a repeat unit is a relatively common event, evolutionarily speaking. Although the sequences of these fragments contain relatively few repeat units, they are still more prone to processes, such as DNA slippage (the DNA replication machinery gets confused as to how many of the units it has duplicated so far, and consequently adds or omits some of the units), that are thought to be responsible for the high variation in microsatellites (Schlötterer 1998).

Undoubtedly plants with ancestral states of the *Mimulus guttatus* and *M. nasutus* fragments experienced deleterious mutations in these genes (such as a regular codon mutating to a stop codon, or an insertion in the wrong place or of the wrong number of bases). But, as functional genes, we expect that these plants were probably selected against, and their genomes did not survive. If the A and B sequences were actually pseudogenes (stretches of DNA that are no longer functional, but retain some characteristics of genes because they are descended from functional genes), there should have been no constraint on such mutations surviving to modern populations. The fact that we do not find this in these sequences is good evidence that these are, or were until very recently, functional genes.

Similarly, if there were no functional constraint on these genes, we might expect that they would diverge rather rapidly. The approximately 90% conservation of sequence

between homologous pairs indicates that, if they are pseudogenes, they were released from selection very recently.

The final line of reasoning that suggests that these fragments are pieces of functional genes is the synonymous versus non-synonymous substitution rates between the species. There are 20 amino acids that a gene can code for; however, there are 64 possible codons. Three are reserved as stop codons. The other 61 each code for an amino acid; thus, some amino acids correspond to anywhere from just one codon to a total of six. In the cases in which multiple codons exist, it is possible for a molecular mutation to occur, yet not change the gene product at all. These are known as synonymous substitutions. However, a given codon can change into any one of nine other codons with a single base substitution; at most, five of these will be synonymous substitutions (on average, approximately two will be synonymous). In a stretch of DNA that is not part of a functional gene, there are no constraints on what base changes will be selected for, so non-synonymous substitutions should not be less common than synonymous substitutions, which are selected against in neither functional nor non-functional sequences.

On the alignments of the nucleotide sequences, the number of synonymous substitutions was compared to the number of non-synonymous substitutions. The dubious sequence, after base 709 in MG-A (base 703 in MN-A) and after base 534 in MG-B (base 543 in MN-B), was not analyzed. In both comparisons, the TCP and R domains were only subject to synonymous substitutions (two in the domains of A, three the domains of B). Out of 37 polymorphic bases in the A sequences, 22 (59.49%) were synonymous. The

result was more dramatic in the comparison of the 25 polymorphisms between B sequences, as 17 (68.00%) were synonymous substitutions. In both cases, a greater number of synonymous substitutions were found than would be expected by chance if all substitutions were equally likely. This is the final evidence I present that shows that these fragments are probably part of two functional genes.

Homology of *Mimulus* A and B loci to other TCP genes

Similarity of these sequences to known (and putative) floral symmetry genes was analyzed using the Clustal X alignment software. Table 2 shows the sources of the sequences used. Figures 31, 32, and 33 show the resultant trees (when default parameters are used) for the whole sequence of nucleotides, the TCP domain nucleotides alone, or the TCP domain nucleotides in conjunction with the R-domain nucleotides, respectively. The *Mimulus* A segments were always most similar to each other, as were the B segments. Often, but not always, the A and B clades are more related to each other than to any of the other clades, indicating that the ancestor of *Mimulus* diverged from the ancestor of *Antirrhinum* before the A-B divergence within the *Mimulus* genome and the *cycloidea-dichotoma* divergence within the *Antirrhinum* genome.

Abbreviation	Species	Common name	Sequence Name	Sequence Source
AGcyc4	Antirrhinum graniticum	snapdragon relative	cycloidea 4	genomic DNA
AMcyc1A	Antirrhinum majus	snapdragon	cycloidea 1A	cDNA
AMcyc1B	Antirrhinum majus	snapdragon	cycloidea 1B	genomic DNA
AMcyc4	Antirrhinum majus	snapdragon	cycloidea 4	genomic DNA
AMdich	Antirrhinum majus	snapdragon	dichotoma	cDNA
HRcyc	Haberlea rhodopensis		cycloidea	genomic DNA
LVcyc1B	Linaria triornithophora	toadflax	cycloidea 1B	genomic DNA
MG-A	Mimulus guttatus	yellow monkey flower	A (longer)	genomic DNA
MG-B	Mimulus guttatus	yellow monkey flower	B (shorter)	genomic DNA
MN-A	Mimulus nasutus	yellow monkey flower	A (longer)	genomic DNA
MN-B	Mimulus nasutus	yellow monkey flower	B (shorter)	genomic DNA
MOcyc2	Misopates orontium		cycloidea 2	genomic DNA
MOcyc3	Misopates orontium		cycloidea 3	genomic DNA

Table 2. Sources of data used for phylogenetics. Only sequences derived from cDNA should be regarded confidently as actual genes. cDNAs are derived from mRNA, which is the result of the gene being expressed. Genomic DNA could harbor pseudogenes that have no actual function.

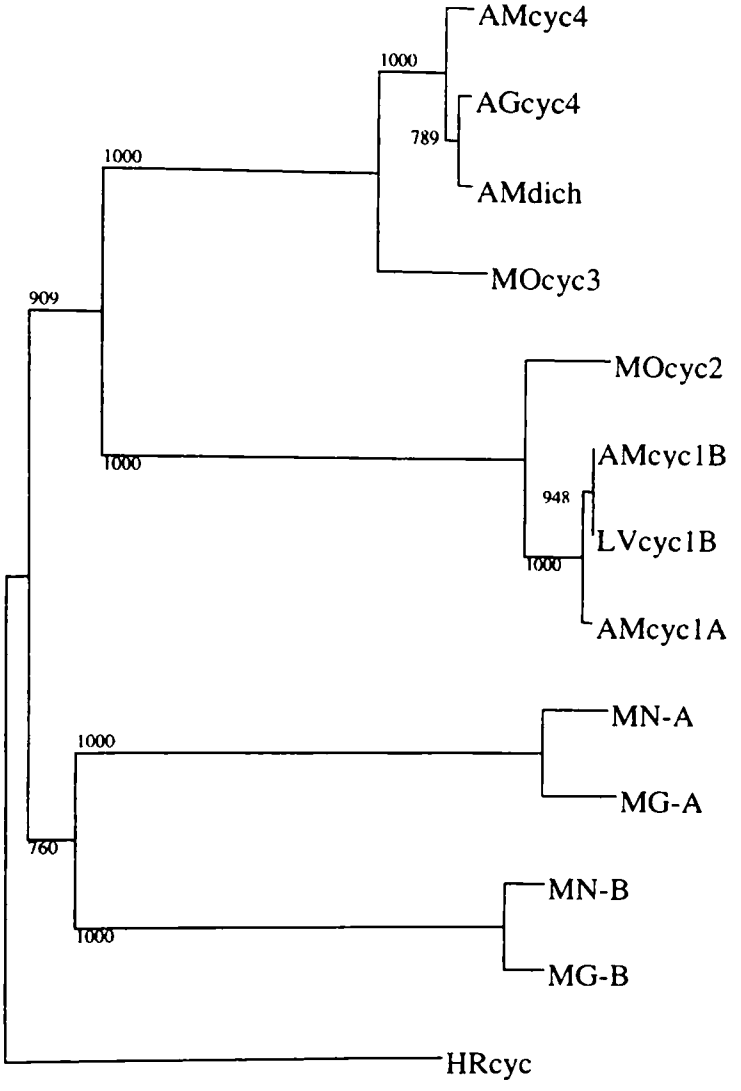


Figure 31. Cladogram of full sequences of putative TCP genes. The "cycloidea" sequence from *Haberlea rhodopensis* (Cronk, personal communication) was used to root the tree because of its more distant relationship to *Antirrhinum* and *Mimulus* than they have with each other. Bootstrap values are out of 1000.

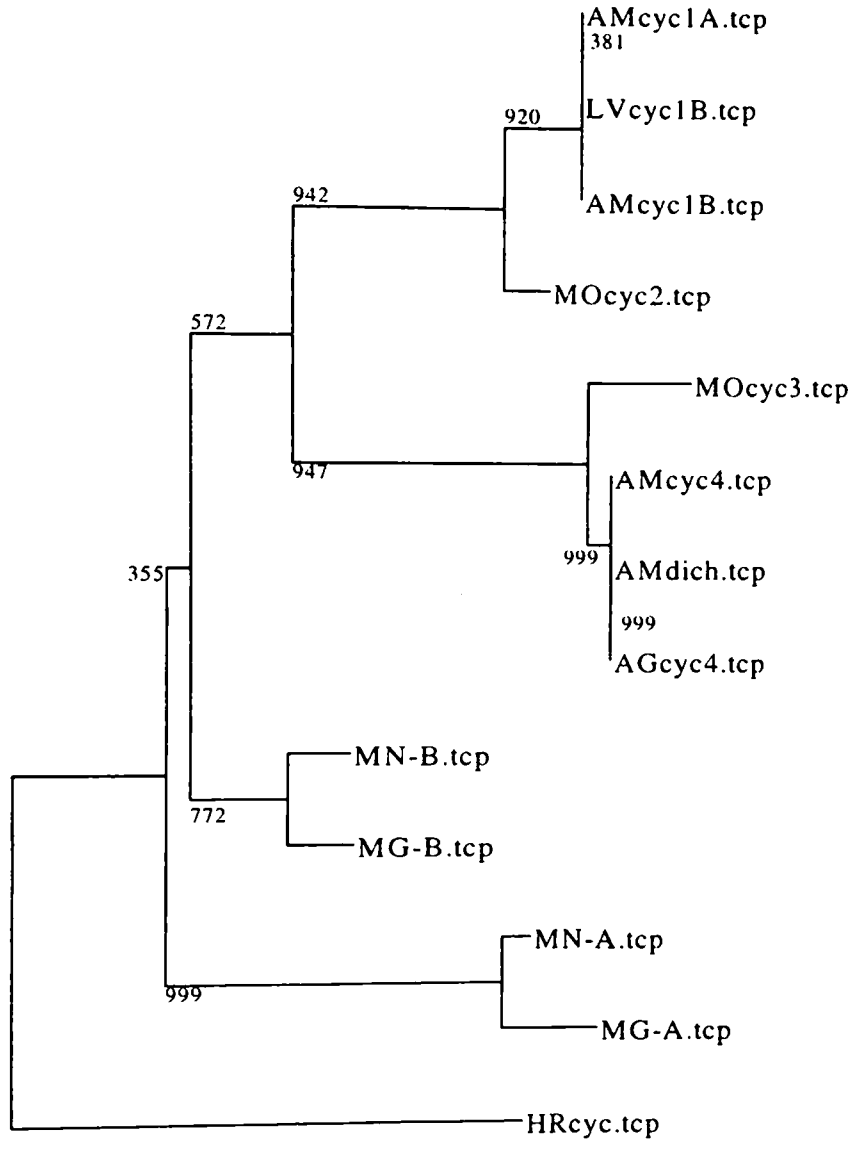


Figure 32. Cladogram of nucleotide sequences from the TCP domain only. Again, sequence from *H. rhodopensis* (Cronk, personal communication) was used to root the tree, and bootstrap values are out of 1000.

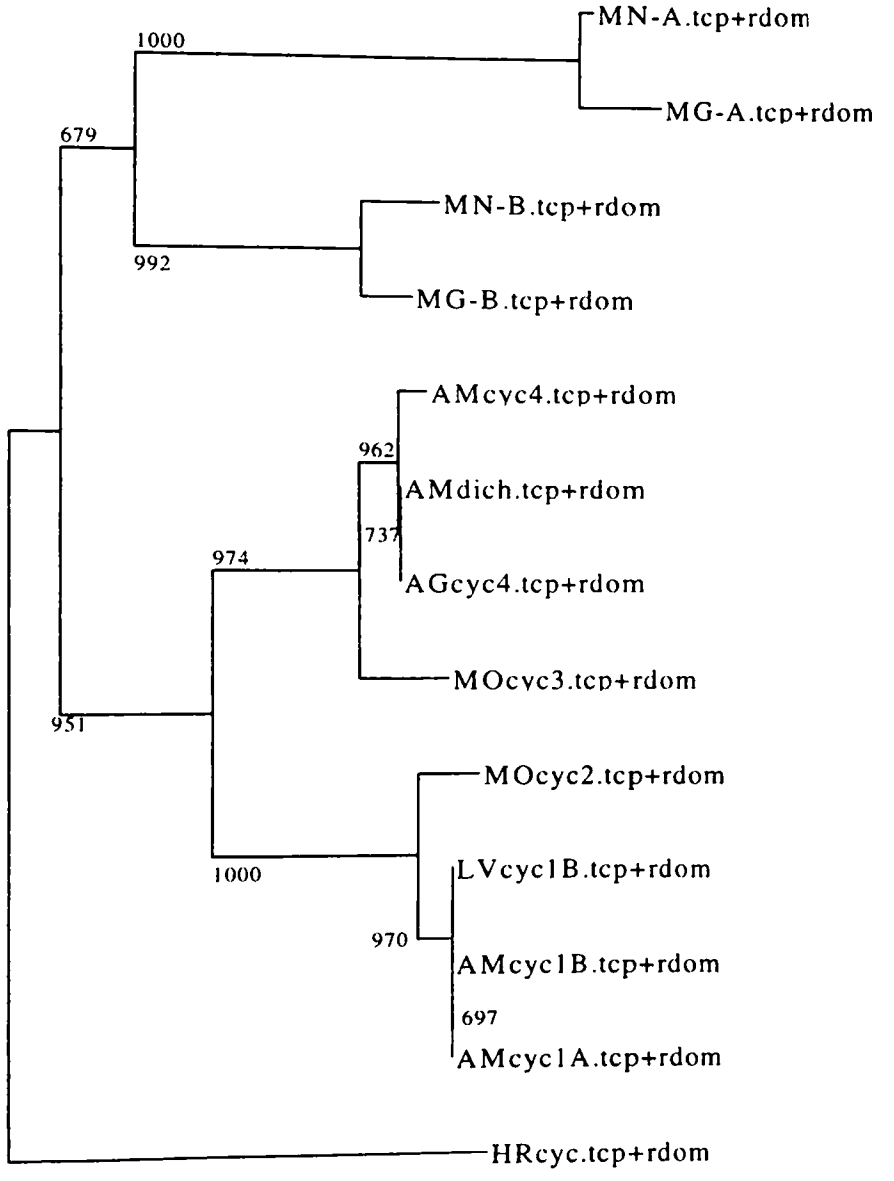


Figure 33. Cladogram of nucleotide sequences from the conserved regions (TCP and R domains) only. Sequence from *H. rhodopensis* (Cronk, personal communication) was used to root the tree, and bootstrap values are out of 1000.

The sequences of these PCR genes were analyzed with some apprehension. Vieira *et al.* (1999) claimed that at least five cycloidea sequences existed in the gene family (*cyc1A*, *cyc1B*, *cyc2*, *cyc3*, and *cyc4*). They suggested that the gene duplications leading to these genes must have occurred at least five million years ago, based on the existence of putative homologous sequences for several of the genes in two clades that diverged at that time (genera *Antirrhinum* and *Digitalis*). We were thus surprised to find little evidence of "appropriately" sized fragments in *Mimulus*.

However, the *cycloidea/dichotoma* split, which defines the two major clades of *Antirrhinum* TCP genes (see Figures 31 and 33), probably occurred after the divergence of *Mimulus* and *Antirrhinum* from a common ancestor (Cronk, personal communication). Therefore we should not expect any TCP genes from *Mimulus* to resemble *dichotoma* any more than *cycloidea*, an expectation that two of the three gene trees support.

My phylogenies (Figures 31, 32, and 33) were generated from alignments performed by Clustal X (version 1.8). The gap creation/extension parameters were left at default levels, which were intermediate. For the full sequence analysis, adjusting the parameters did not increase the overall bootstrap support for any particular tree (data not shown). Figures 31 and 32 would not have been affected by any gap parameters because they involved only conserved domains that contained no insertions or deletions in any sequence. The bootstrap value supporting the *Mimulus* clade was the lowest value on the full-sequence tree, which could be the result of so much divergence in the non-domain regions that they caused the whole sequence to be ambiguous. So the second two trees, based on the TCP or a combined TCP plus R domain sequence, were generated. Domain-

based comparisons sometimes work better because the possible changes that the gene could have gone through are more constrained than other sequence. Not only can bases in domains not evolve to form a stop codon (which is true of the whole sequence), but any non-synonymous substitutions they make must create an amino acid with similar-enough properties to the one it replaces that the function of the gene is not deleteriously affected.

Figure 32, then, shows the tree generated from the first 61 bases of the sequences, which corresponds to the last 18 amino acids of the TCP domain, plus the two amino acids directly after it (Cubas *et al.* 1999). This tree, unfortunately, was less clearly resolved than the full sequence data, and places the longer A fragments as a more distant relative of the *cycloidea/dichotoma* clades than the shorter B fragments are. There is very poor bootstrap support for removing the A fragments from the clade of everything else.

The complete R domain sequences were contained in the region we sequenced. Thus we added them directly to the 3' end of the TCP sequence. Figure 33 shows the resulting phylogenetic tree. This tree is nearly identical to that based on the entire sequence (Figure 31). While the bootstrap support for placing the *Mimulus* sequences and the *cycloidea/dichotoma* sequences as sister clades is still less than ideal, it appears to represent the maximum amount of information that the data will tell us.

It is interesting to note that "cyc3" and "cyc4" appear to be mistakenly named by Vieira *et al.* (1999). Their basic approach, of PCR-amplifying as many sequences as TCP-gene based primers would allow, was very similar to mine. They sampled the genera *Antirrhinum* (snapdragon), *Linaria* (toadflax), *Misopates*, *Cymbalaria*, and *Digitalis* (foxglove). They did not include the *dichotoma* sequence in their phylogenetic

analysis; presumably the full sequence was not available to them, though they do mention its existence as another TCP gene. But, it would be more appropriate to name these sequences based on their relatedness to *dichotoma*, as all three of my phylogenetic trees suggest.

Future work

Mapping *Mimulus* loci A and B

There are two areas of future research that will be aided by my PCR work. The first was the original point of the amplifications, to find candidate loci that may map to QTLs. For that, we would need to genotype hundreds of F2 individuals, which can have genomic DNA from one or both parents at the locus (here, the parents are the inbred *Mimulus guttatus* and *Mimulus nasutus* lines used in the QTL study). Because we already know the sequence, we will be able to design simple and relatively cheap assays for the presence or absence the *M. guttatus* or *M. nasutus* alleles. This can be accomplished in several ways (both of which depend on redesigning primers that will only amplify the A or the B fragment, not both). First, small differences in sequence length (such as between the A or B fragments of each species) can be assayed when the primers used are labeled for use in an acrylamide gel electrophoresis protocol. Second, certain sites in the sequences can be recognized and cut by a class of enzymes known as restriction endonucleases. If a recognition sequence can be found on one species' fragment that does

not exist on the other (due to base substitutions or insertion/deletions), the cut DNA will run out as (at least) two bands of dramatically different sized DNA than the uncut sequence. This will allow us to genotype all the individuals quickly. With that data, the map-generating software that has been used for the QTL analysis (MAPMAKER-QTL) will be used to deduce the molecular location of the loci on the chromosomes relative to other markers. Upon mapping, we will be able to tentatively say whether or not these fragments, which are putative floral development genes, may have been involved in the evolutionary adaptation between *M. guttatus* and *M. nasutus*.

Phylogenetics in the genus *Mimulus*

The other area of future research that this project lends itself to is construction of a phylogeny. There are at least 100 species of monkey flower (Holmgren 1984), not all of whose relatednesses are known. The considerable amount of sequence divergence between these two closely related species in my study indicates that we might find other polymorphisms at the same loci in other monkey flowers. A phylogeny of monkey flowers will have many uses for researchers interested in their ecology and evolutionary history.

The PCR primers that yielded these fragments were designed for the purpose of assessing interspecies relatednesses in the family Gesneriaceae, which includes African violets (Möller *et al.* 1999). That study found that the constraints on the evolution of single-copy nuclear genes (as opposed to non-functional DNA) make such genes useful

for phylogenetics at higher levels of species divergence. It is possible that the presence of both highly constrained regions (the TCP and R domains), as well as apparently less-constrained regions, will make these regions useful for both closely and distantly related species of *Mimulus*.

Conclusion

In this study I was able to characterize loci, both phenotypic and molecular, that will serve as candidate loci for the genes responsible for the adaptive differences between *Mimulus guttatus* and *Mimulus nasutus*. The phenotypic mutants varied in type, from simple size changes to homeotic transformations of floral organs, and most segregated in a Mendelian fashion. Both molecular loci, A and B, appear to be fragments of functional genes of the TCP gene family, and differ by about 9% between *M. guttatus* and *M. nasutus*. Map locations of all these loci, combined with the QTL map experiment, may enable future work in the Willis Lab to fully describe the evolutionary genetic changes that were responsible for the adaptive differences that distinguish *M. guttatus* from *M. nasutus*.

WORKS CITED

Almeida, Jorge, Margarida Rocheta, and Lisete Galego (1997). "Genetic Control of Flower Shape in *Antirrhinum*." *Development* 124:1387-1392.

Bradley, Desmond, Rosemary Carpenter, Hans Sommer, Nigel Hartley, and Enrico Coen (1993). "Complementary Floral Homeotic Phenotypes Result from Opposite Orientations of a Transposon at the *plena* locus of *Antirrhinum*." Cell 72:85-95.

Carlsbecker, Annelie, Jens Sundström, Mats Svensson, Karolina Tandre, Marie Svenson, Liz Izquierdo, Anders Kvarnheden (1998). "MADS-Box Genes." *Molecular and Developmental Biology of Plants*. Uppsala University. 01 May 2000. <http://www.fysbot.uu.se/fysbot/MADS/mads.html>

Coen, Enrico and Elliot M. Meyerowitz (1991). "The War of the Whorls: Genetic Interactions Controlling Flower Development." *Nature* 353:31-37.

Cronk, Q.C. B. Unpublished sequence of Gcyc from *Haberlea rhodopensis*, in electronic correspondence to Jan Aagaard (2000).

Cubas, Pilar, Nick Lauter, John Doebley, and Enrico Coen (1999). "The TCP Domain: a Motif Found in Proteins Regulating Plant Growth and Development." *The Plant Journal* 18:215-222.

Doebley, John and Adrian Stec (1991). "Genetic Analysis of the Morphological Differences Between Maize and Teosinte." *Genetics* 129:285-295.

Doebley, John and Adrian Stec (1993). "Inheritance of the Morphological Differences Between Maize and Teosinte: Comparison of Results for Two F_2 Populations." *Genetics* 134: 559-570.

Doebley, John (1993). "Genetics, Development, and Plant Evolution." *Curr. Opin. Genet. Dev.* 3:865-872.

Doebley, John, Adrian Stec, and Lauren Hubbard (1997). "The Evolution of Apical Dominance in Maize." *Nature* 386:485-488.

Doebley, John and Lewis Lukens (1998). "Transcriptional Regulators and the Evolution of Plant Form." *The Plant Cell* 10:1075-1082.

Doyle, Jeffrey J. and Jane L. Doyle (1987). "A Rapid DNA Isolation Procedure for Small Quantities of Leaf Tissue." *Phytochemical Bulletin* 19:11-15.

Holmgren, N. (1984). "Scrophulariaceae." *Intermountain Flora: Vascular Plants of the Intermountain West*, vol. IV (New York Botanical Garden: New York).

Hyam, Roger and Richard J. Pankhurst (1995). *Plants and Their Names: a Concise Dictionary* (Oxford UP: New York):322 and 456.

Lönnig, Wolf E. (1997). "Snapdragon Mutants – plena (ple)." *Snapdragon Home Page*, ed. Kurt Stüber. Max-Planck-Institut für Züchtungsforschung, Köln. 08 May 2000. <http://www.mpiz-koeln.mpg.de/~stueber/snapdragon/mutants/ple.html>

Luo, Da, Rosemary Carpenter, Coral Vincent, Lucy Copsey, and Enrico Coen (1996). "Origin of Floral Asymmetry in *Antirrhinum*." *Nature* 383:794-799.

Luo, Da, Rosemary Carpenter, Lucy Copsey, Coral Vincent, Jennifer Clark, and Enrico Coen (1999). "Control of Organ Asymmetry in Flowers of *Antirrhinum*." *Cell* 99:367-376.

Leyser, H.M. Ottoline and Ian J. Furner. "EMS Mutagenesis of *Arabidopsis*." *The Arabidopsis Information Resource*. Leland Stanford Junior University. 25 May 1999.
<http://www.arabidopsis.org/comguide/chap_1_plants/6_EMS_mutagenesis.html>

Möller, Michael, Martha Clokie, Pilar Cubas, and Quentin C.B. Cronk (1999). "Integrating Molecular Phylogenies and Developmental Genetics: a Gesneriaceae Case Study." *Molecular Systematics and Plant Evolution*, eds. P.M. Hollingsworth, R.M. Bateman, and R.J. Gornall (Taylor & Francis: London):375-402.

Riechmann, José Luis and Elliot M. Meyerowitz (1997). "MADS Domain Proteins in Plant Development." *Biol. Chem.* 378:1079-1101.

Schlötterer, Christian (1998). "Genome Evolution: Are Microsatellites Really Simple Sequences?" *Current Biology* 8:R132-R134.

Vickery, Robert K., Jr. (1978). "Case studies in the evolution of species complexes in *Mimulus*." *Evol. Biol.* 11:405-507.

Vieira, Cristina, Jorge Vieira, and Deborah Charlesworth (1999). "Evolution of the Cycloidea Gene Family in *Antirrhinum* and *Misopates*." *Mol. Biol. Evol.* 16:1474-1483.

Wang, Rong-Lin, Adrian Stec, Jody Hey, Lewis Lukens, and John Doebley (1999). "The Limits of Selection During Maize Domestication." *Nature* 398:236-239.

Yanofsky, Martin, Hong Ma, John L. Bowman, Gary Drews, Ken Feldmann and Elliot Meyerowitz (1990). "The Protein Encoded by the *Arabidopsis* Homeotic Gene AGAMOUS Resembles Transcription Factors." *Nature* 246:35-39.