ABSTRACT

BARRIER, MARIANNE. Selection and Rapidly-Evolving Genes. (Under the direction of Michael D. Purugganan and William R. Atchley)

Rapidly-evolving genes which exhibit an increased rate of amino acid substitution, resulting in greater amino acid sequence difference, often provide insight into the mechanisms of adaptation and speciation. In this dissertation research, the evolution of rapidly-evolving genes under selection was examined. Homologues to the *Arabidopsis APETALA3* (*ASAP3/TM6*) and *APETALA1* (*ASAP1*) floral regulatory genes and the *CHLOROPHYLL A/B BINDING PROTEIN9* (*ASCAB9*) photosynthetic structural gene were isolated from species in the Hawaiian silversword alliance, a premier example of plant adaptive radiation. The two floral homeotic genes are found in duplicate copies within members of the Hawaiian silversword alliance and appear to have arisen as a result of interspecific hybridization between two North American tarweed species. Molecular phylogenetic analyses suggest that the interspecific hybridization event involved members of lineages that include *Carlquistia muirii* (and perhaps *Harmonia nutans*) and *Anisocarpus scabridus*. Next, rates of regulatory and structural gene evolution in the Hawaiian species were compared to those in related species of North American tarweeds. Molecular evolutionary analyses indicate significant increases in nonsynonymous relative to synonymous nucleotide substitution rates in the *ASAP3/TM6* and *ASAP1* regulatory genes in the rapidly evolving Hawaiian species. By contrast, no general increase is evident in neutral mutation rates for these loci in the Hawaiian species. Finally, a group of potential

rapidly-evolving genes were identified in *Arabidopsis* using a powerful evolutionary expressed sequence tag (EST) approach. One indicator of adaptive selection at the molecular level is an excess of amino acid replacement fixed differences per replacement site relative to the number of silent fixed differences per silent site (ω = Ka/Ks). The evolutionary EST approach was used to estimate the distribution of ω among 304 orthologous loci between *Arabidopsis thaliana* and *Arabidopsis lyrata* to identify genes potentially involved in the adaptive divergence between these two Brassicaceae species. Twenty-one of 304 genes (7%) were found to have an estimated $\omega > 1$ and are candidates for genes associated with adaptive divergence. A hierarchical Bayesian analysis of protein coding region evolution within and between species also indicates that the selection intensities of these genes are elevated compared to typical *A. thaliana* nuclear loci.

SELECTION AND RAPIDLY-EVOLVING GENES

by

MARIANNE BARRIER

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

GENETICS

Raleigh

2002

APPROVED BY:

Dr. Stephanie E. Curtis

Dr. Jeffrey L. Thorne

Dr. Michael D. Purugganan
Co-chair of Advisory Committee

Dr. William R. Atchley
Co-chair of Advisory Committee

DEDICATION

This work would not have been possible without the continued love and support of my parents, Jack and Jean Barrier, who instilled in me the desire to learn and to help others. I thank my sister, Melinda Kroeze and her family: Darryl, Matthew, and Adriana, for providing support and a place to go whenever I needed a break. I thank my best friend, Dina Sieroslawski, for being there almost all of my life and for believing in me. I thank the scientists I worked with at EPA who inspired me to follow my scientific interests. I thank the rest of my family, my friends, and the members of Millbrook Baptist Church for their continued support. I dedicate this in honor of my grandfather, Kelly Howard, and in memory of my grandmothers, Tannie Howard and Nancy Barrier, and my grandfather Omer Barrier.

BIOGRAPHY

Marianne Barrier was born on May 20, 1970 in Endicott, New York. She received her early education in Raleigh, North Carolina, graduating from William G. Enloe High School in 1988. She received her Bachelor of Science in Mathematics with a Computer Science concentration and Art minor from the University of North Carolina at Greensboro in 1992. She worked as a Business Systems Analyst providing computer graphics and systems support for the Environmental Criteria and Assessment Office of the Environmental Protection Agency until beginning graduate school in 1996. During that time she obtained further education with a certificate in Architectural CAD and Drafting in 1994 from Wake Technical Community College in Raleigh and a certificate in 3D Art and Animation in 1995 from the School of Communication Arts, also in Raleigh.

She was admitted to the Genetics Department at North Carolina State University in 1996 as a candidate for the degree of Doctor of Philosophy. She has undertaken her doctoral research under the guidance of Dr. Michael D. Purugganan. On July 1, 2002 she will begin her postdoctoral position with Dr. Philip E. Mirkes in the Department of Pediatrics at the University of Washington Medical School in Seattle, Washington.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Michael Purugganan, who has been a great mentor during my years of doctoral research. I appreciate his guidance and enthusiasm, especially at the beginning when I was a novice in the field of scientific research. I am grateful to all of the members of the Purugganan Lab, past and present, for their friendship and counsel over the years. I would especially like to thank Jane Suddith for all of her help and for taking care of my rotation project, so that it was ready for me when I officially joined the lab.

I thank my co-conspirator, Bill Atchley, for his pearls of wisdom and for allowing me to assist in many educational side-projects. I thank my teaching mentor, Betty Gardner, for her guidance in teaching the genetics laboratory course and for giving me a greater appreciation for the amazing efforts of all teachers. I thank the GRL staff for their assistance when my work kept me there for a couple of months. I thank Rob Robichaux, Bruce Baldwin, and Elizabeth Friar for their assistance with my research projects and guidance during our fieldwork trips in Hawai'i and California.

Finally, I would like to thank everyone in the Genetics Department at NCSU: Stephanie Curtis and the genetics office staff, who keep the department going, Jeff Hawley who keeps the labs running (or moving), and everyone else in the department who has provided supplies or guidance. I greatly appreciate the friendship and support of the "payday" group during our many years together.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Appendix B

LIST OF SYMBOLS AND ABBREVIATIONS

aa	Amino acid
AFLP	Amplified Fragment Length Polymorphism
AGI	*Arabidopsis* Genome Initiative
AP1	*Arabidopsis APETALA1*
AP3	*Arabidopsis APETALA3*
ASAP1	*A. sandwicense APETALA1*
ASAP3/TM6	*A. sandwicense APETALA3/TM6*
ASCAB9	*A. sandwicense CHLOROPHYLL A/B BINDING PROTEIN9*
BAC	Bacterial Artificial Chromosome
bp	Base pair
CAB9	*Arabidopsis CHLOROPHYLL A/B BINDING PROTEIN9*
CAL	*Arabidopsis CAULIFLOWER*
CDR	Complementarity-Determining Regions
CI	Confidence Interval
cpDNA	Chloroplast DNA
DEF	*Antirrhinum DEFICIENS*
EST	Expressed Sequence Tag
ETS	External Transcribed Spacer
FR	Framework Region
HLA	Human Leukocyte Antigens
HPCI	Highest Posterior probability Credibility Interval
indel(s)	Insertion/Deletion mutation(s)
ITS	Internal Transcribed Spacer
K	Number of nonsynonymous (replacement) substitutions
Ka	Proportion of amino acid replacement (nonsynonymous) substitutions
Ks	Proportion of silent (synonymous) substitutions
MCMC	Markov Chain Monte Carlo

MHC	Major Histocompatibility Complex
mya	Million Years Ago
N	Number of nonsynonymous (replacement) substitutions
N_e	Effective population size
PCR	Polymerase Chain Reaction
PI	*Arabidopsis PISTILLATA*
QTL	Quantitative Trait Loci
RACE	Random Amplification of cDNA Ends
rDNA	Ribosomal DNA
RT	Reverse Transcription
S	Number of synonymous (silent) substitutions
SQUA	*Antirrhinum SQUAMOSA*
ssp	Subspecies
SSR	Simple Sequence Repeats
t	Scaled species divergence parameter
V_H	Immunoglobulin heavy-chain variable-region
YAC	Yeast Artificial Chromosome
ε	Sequencing error
γ	Selection intensity on replacement changes
μ	Mean
π	Nucleotide diversity
σ^2	Variance
θ	Nucleotide diversity
θ^R	Mutation rates at replacement sites
θ^S	Mutation rates at silent sites
ω	Ratio of replacement (Ka) to silent (Ks) substitutions (Ka/Ks)

CHAPTER 1.

Introduction and Literature Review

Introduction

Rapidly-evolving genes have been of particular interest in evolutionary research because they often provide insight into the mechanisms of adaptive differences within or between species (e.g., Schmid and Tautz 1997; Purugganan 1998, 2000). One definition of rapidly-evolving genes, which is the definition used throughout this dissertation, refers to those genes which exhibit an increased rate of amino acid substitution resulting in greater amino acid sequence difference. In molecular systematics, this difference can sometimes make these genes useful as markers for determining phylogenetic relationships (e.g., Kim and Jansen 1995; Hedin and Maddison 2001). In this dissertation research, the evolution of those rapidly-evolving genes demonstrating the effects of diversifying selection is examined with the goals of determining which types of genes have undergone diversification and how best to identify them.

This dissertation first examines the molecular evolution of genes in a group of species, the Hawaiian silversword alliance, which has undergone a spectacular adaptive radiation leading to a great variety of morphological and ecological adaptations (Barrier et al. 1999, 2000, 2001). This research clarifies the origin of the ancestral tetraploid Hawaiian silversword alliance species as originating from an interspecific hybridization of two lineages of diploid North American tarweeds (Barrier et al. 1999). The added genomic diversity from the polyploidization event may have provided greater genetic flexibility which, along with selection, resulted in

the spectacular adaptation and speciation that followed the radiation across the Hawaiian archipelago (Stebbins 1971).

Next, the evolution of regulatory genes in the Hawaiian silversword alliance is examined (Barrier et al. 2001), since divergence of regulatory genes is believed to underlie the evolution of morphology (Purugganan 1998, 2000). The regulatory genes analyzed are the homologs of the *Arabidopsis APETALA1 (AP1)* and *APETALA3 (AP3)* genes, members of the MADS-box family of plant transcriptional activators (Riechmann and Meyerowitz 1997). In previous research, *AP1* has been shown to be involved in regulating the initiation of flowering as well as in controlling sepal and petal organ identity in *Arabidopsis* flowers (Weigel 1998). *AP3* has been shown to be involved in petal and stamen development in *Arabidopsis* flowers (Weigel 1998). The homolog to the *Arabidopsis AP3* gene in the silverswords is called *ASAP3/TM6* because phylogenetic analysis indicates it is also an ortholog of the *Lycopersicon TM6* locus (Purugganan 1997; Kramer, Dorit and Irish 1998). In this dissertation research, the homologs to these regulatory genes (called *ASAP1, ASAP3/TM6* in the alliance and tarweed species) exhibit an increased rate of protein evolution in the Hawaiian silversword alliance species when compared to their homologs in the North American ancestral species, which have not undergone a similar type of adaptive radiation. The photosynthetic structural gene *CHLOROPHYLL A/B BINDING PROTEIN9* (*CAB9*) (Pichersky et al. 1991) was also studied as a comparison to the regulatory genes because previous studies failed to show a correlation between rates of molecular evolution at structural genes and rates of morphological evolution (King and Wallace

1975; Sturmbauer and Meyer 1992). Only a slight increase in amino acid evolution is observed in the homolog of this structural gene (called *ASCAB9*) in the Hawaiian silversword alliance species when compared to the homolog in the North American tarweeds.

Finally, an evolutionary expressed sequence tag (EST) approach is used to identify rapidly-evolving genes in the model organism *Arabidopsis thaliana* (Barrier et al. manuscript, Chapter 4). *Arabidopsis lyrata* is used for comparison to identify genes which might be involved in the adaptive divergence between these closely-related species. This approach identifies 21 putatively fast-evolving genes in *Arabidopsis* out of the 304 unique ESTs examined. Population studies for 7 of the potentially fast-evolving genes reveal that these genes display possible evidence of diversifying selection.

Selection and Rapidly-Evolving Genes

Mutation and selection

Selection comes in many forms, but, in general, selection acts to retain or to remove random nucleotide mutations. Rates of mutation vary between genes in an organism as well as between the same genes in different populations. In *E. coli* genes, for example, mutation rates range from 10^{-5} to 10^{-9} mutations per cell per generation (Hartl and Jones 2001). Most mutations in noncoding regions are selectively neutral, so for individual genes the rate of mutation at noncoding sites can be used to estimate the neutral mutation rate for the whole gene region (Kreitman 2000). This estimate cannot be applied to the genome as a whole because across the genome there is significant variation in the levels of polymorphism in noncoding regions, which are strongly correlated with rates of recombination (Begun and Aquadro 1992; Aquadro 1997).

In protein-coding regions, some mutations are neutral or even advantageous, although most are deleterious. Neutral or nearly-neutral mutations are usually removed or maintained by random genetic drift rather than by selection (Ohta 1995, 2000). Typically, deleterious mutations are eliminated quickly by purifying selection while advantageous mutations are kept in the population by some form of positive selection (Ohta 1991a, 1991b).

Linkage extends the effects of selection to other loci linked to the selected site, thus reducing variation at these neutral sites. Hitchhiking or selective sweeps occur

when neutral mutations at linked sites are brought along with the fixation of advantageous mutations (Majewski and Cohan 1999; Barton 2000). In contrast, background selection results in the loss of neutral variants that are linked to deleterious mutations eliminated by purifying selection (Schug et al. 1998; Andolfatto and Przeworski 2001).

Forms of selection

Balancing, diversifying, and directional selection are three specific forms of selection. Balancing selection can be observed at individual residues where two or more variants are maintained in the population (e.g., Erlich and Gyllensten 1991; Verra and Hughes 1999). Erlich and Gyllensten (1991) observed balancing selection of multiple variants at select residues of the antigen binding site (ABS) β-chain, as well as at sites implicated in susceptibility to disease. At each polymorphic residue, they found that the same limited set of amino acids was maintained in all of the primate species studied.

Balancing selection can be the result of overdominance, where the heterozygous condition at a locus is more fit than either homozygote (e.g., Hughes et al. 1990; Garrigan and Hendrick 2001; Aidoo et al. 2002). Sickle-cell anemia is a common example of overdominance, also known as heterozygote advantage. The presence of a single copy of the sickle-cell hemoglobin gene (*HbS*) is believed to provide protection from malaria, while two copies (*HbSS*) produce the sickle-cell disease and no copies (*HbAA*) provide no protection from malaria (Aidoo et al. 2002). Aidoo et al. (2002) examined the protective advantage of the heterozygote (*HbAS*) in malaria endemic regions and confirmed that in children age two to sixteen months old, who are most at risk for malaria, *HbAS* was associated with lower mortality rates.

Diversifying selection, similar to balancing selection, also results in amino acid diversity at a site or region (e.g., Burstein et al. 1982; Tanaka and Nei 1989), but the set of alleles observed can continue to change. By examining the complementarity-

determining regions (CDRs) and framework regions (FRs) of mouse and human immunoglobulin heavy-chain variable-region (V_H) genes, Tanaka and Nei (1989) observed higher nonsynonymous rates than synonymous rates of substitution in the CDRs but the reverse in the FRs. This high degree of V_H gene CDR diversity supports the previous hypothesis of Gojobori and Nei (1984) that these regions are subject to diversifying selection.

Directional selection is evident when specific mutations are driven to fixation in a population while others are eliminated (e.g., Blanco-Gelaz et al. 2001; Hoekstra et al. 2001). Blanco-Gelaz et al. (2001) examined twenty-three closely related alleles (B*2701-23) of the HLA-B27 gene in order to identify the patterns of synonymous and nonsynonymous polymorphisms at this gene in human populations from around the world. This gene has been extensively studied because it has a strong association with human conditions such as ankylosing spondylitis. They discovered a striking correlation between the structural features of B27 subtypes and their ethnic distribution which suggests a model of strong directional evolution of these subtypes from the B*2705 allele.

Rates of substitution as indicators of selection

Molecular evolutionists have sought to develop methods to identify genes, and even nucleotide positions, which may have experienced recent bursts of balancing, directional, or diversifying selection. One way to identify the type of selection acting on a protein sequence is to examine the amount of amino acid replacement and

synonymous substitutions in the protein coding regions. In protein coding sequences, synonymous substitutions are those which do not alter the amino acid coding and are, therefore, more likely to be neutral. Nonsynonymous or replacement substitutions, which do alter amino acid coding, tend to be deleterious. Values often used to analyze mutations in coding regions are: Ks, the estimated number of synonymous changes per synonymous site, and Ka, the estimated number of amino acid replacements per replacement site (Nei and Gojobori 1986).

Many of the commonly used models for estimating Ka and Ks are based on using nucleotides as the evolutionary unit. These include Jukes and Cantor (1969), Kimura (1980), and Nei and Gojobori (1986). These models examine each nucleotide as an individual unit, independent of the surrounding nucleotides and assume uniform base composition. The Jukes and Cantor model (1969) assumes that all kinds of substitutions occur at the same rate, while the Kimura two-parameter model (Kimura 1980) allows for different rates of transitions and transversions. Tests of these models have determined that they tend to underestimate substitution rates, particularly of Ka, so they provide a conservative estimate of the substitution rates. These methods underestimate these rates due to a parsimony approach used to infer the minimal number of nucleotide substitutions between two homologous codons and because they do not fully account for multiple substitutions.

Another way of estimating Ka and Ks is by using the codon as the unit of evolution rather than individual nucleotides (Muse and Gaut 1994). Codon-based models acknowledge that the three sites of a codon are linked because they form a

single coding unit. Two common codon-based models developed by Muse and Gaut (1994) and Goldman and Yang (1994) are based on the maximum likelihood framework of Felsenstein (1981). These likelihood-based models often provide a better estimate of Ka and Ks because they include factors such as transition/transversion rate bias and codon usage bias (Goldman and Yang 1994). Multiple substitutions are handled by assuming that they happen as a series of single substitution events rather than as simultaneous substitutions (Muse and Gaut 1994). They can also use physicochemical distances between amino acids (e.g., Grantham 1974) to account for the selective restraints at the protein level (Goldman and Yang 1994). The method of Muse and Gaut (1994) estimates the equilibrium frequency of each codon based on each of the three positions in the codon. The method of Goldman and Yang (1994) does not examine the three codon positions separately, but rather estimates the equilibrium frequencies for each of the 61 sense codons as a whole (stop codons are excluded). Since these codon-based models tend to have more parameters to estimate, this approach can be computationally intensive.

The ratio of replacement (Ka) to synonymous (Ks) substitutions, ω, (ω = Ka/Ks) is often used as an indicator of selection (Messier and Stewart 1997), with ω values falling into three categories. First, $\omega > 1$ (Ka > Ks) indicates diversifying selection. Examples of diversifying selection are further discussed below. Second, $\omega = 1$ (Ka = Ks) indicates neutrality or no selection, as is the case for pseudogenes (e.g., Cadavid et al. 1996). Third, $\omega < 1$ (Ka < Ks) indicates negative selection, the category into which most nuclear genes fall. This is also known as purifying selection because

selection acts to eliminate deleterious mutations (e.g., Boyson et al. 1999; Bustamante, Townsend and Hartl 2000). In observations of nuclear genes, the average ω value typically falls in the range of purifying selection: 0.14 for monocot and dicot plant genes (Martin et al. 1989; Huang et al. 1992) and 0.16 for mammalian genes (Li et al. 1985).

One way to determine if ω is significantly different from 1 is to perform a Z-test on the difference between Ka and Ks to see if it is significantly greater than 0. A concern about this as a test of selection is that the criterion of $\omega > 1$ for diversifying selection is very stringent. This criterion may miss cases where selection is limited to a region of the gene, so that the evidence of selection is neutralized by neutral or negatively selected regions when the gene is examined as a whole (Sharp 1997; Crandall et al. 1999).

Population-based tests of selection

Selection can also be inferred by examining the levels and patterns of nucleotide divergence within and between species. One molecular population genetic test of neutrality is the McDonald-Kreitman test (McDonald and Kreitman 1991) which determines the evolutionary forces that act on protein coding regions. This test compares the numbers of nonsynonymous, N (or K), and synonymous, S, polymorphisms within species as well as divergence between species. These values are tested in a two-by-two contingency table where, under neutral evolution, the N/S ratios should be equivalent between the two groupings. If neutrality is rejected, then

two selective scenarios are possible. The most common is when there are more fixed replacement differences between species than replacement polymorphisms within species as a result of adaptive protein evolution between species. The opposite case, where within-species replacement polymorphism is greater, may be the result of an accumulation of deleterious mutations within species or as the result of balancing selection. The statistical significance of these contingency tables can be calculated with the two-tailed Fisher's exact test and the G-test of independence. Both of these tests determine whether there is significant difference between the N/S ratios. The McDonald-Kreitman test can also be applied to other categories, such as the direction of mutation from a preferred to an unpreferred codon or vice versa, within and between species (Akashi 1995).

This test of selection examines a single gene and may not always detect weak selection. A newer method by Bustamante et al. (2002) uses the same basic contingency table categories as the McDonald-Kreitman test, but analyzes multiple genes from the same species pair. By pooling the information from a set of genes and making use of the fact that they all share the same interspecies divergence time, a hierarchical Bayesian method is used to estimate selection coefficients for replacement changes under a Poisson Random Field model. The direction and intensity of selection is estimated for each gene based on this model (See Appendix A). This is a more powerful test of selection because it utilizes information about selective forces from all of the genes to determine the selection on individual genes.

Rapidly-evolving genes

The term "rapidly-evolving genes" encompasses circumstances in which genes exhibit rapid genetic change. One example is seen in viruses where many genes tend to evolve rapidly because of high mutation rates and the absence of proofreading and mismatch repair mechanisms (Holland et al. 1982). In another example, repeat regions such as microsatellites tend to be highly variable due to their susceptibility to strand slippage during replication and unequal crossing over (Scherf et al. 1988). Weber and Wong (1993) determined the mutation rate of microsatellites to be 10^{-2} to 10^{-3} per locus per gamete per generation. (See Appendix B for a study of two genes that appear to be hotspots for microsatellite formation.) Finally, the types of rapidly-evolving genes examined throughout this dissertation are those genes which exhibit a greater rate of amino acid substitutions than of synonymous substitutions. These types of genes are often associated with adaptive variation within or between species or populations. In this case, $\omega > 1$, the higher rate of substitution is not just due to an increase in the neutral mutation rate (μ). A general increase in the neutral mutation rate would result in higher rates of substitution not only at nonsynonymous sites but also at noncoding and synonymous sites. Two of the main factors that contribute to increased nonsynonymous variation in these rapidly-evolving genes are a lack of strong structural or functional constraint (e.g., Avedisov et al. 2001; Hedin and Maddison 2001) and diversifying selection.

This dissertation research is based on an interest in genes evolving rapidly due to diversifying selection, because they are more likely involved in species adaptations.

Selection for adaptation is usually in response to various stimuli such as changes in the environment (e.g., Donohue et al. 2001), or reproductive pressures (e.g., Begun et al. 2000; Armbrust and Galindo 2001). Many well-known rapidly-evolving genes are those involved in the immune systems of animals, where being able to identify a wide range of pathogens is vital. For example, in mice and humans diversifying selection was observed in the antigen recognition site of the major histocompatibility complex (MHC) class I loci (Hughes and Nei 1988; Bergstrom and Gyllensten 1995) and in the complementarity-determining (hypervariable) regions (CDR) of immunoglobulins (Burstein et al. 1982; Tanaka and Nei 1989). The defense mechanisms in plants also display regions of diversifying selection. In many dicots, positive selection was observed in plant class I *chitinase*, which attacks plant pathogens (Bishop et al. 2000).

Interactions between predator and prey or pathogen and host can drive selection as each competes for the advantage. In the predatory snail, *Conus abbreviatus*, its "weapon", *conotoxin*, is under strong selection for functional diversity with a rate of evolution higher than that of most other known proteins (e.g., Duda and Palumbi 1999). In a study of the malaria parasites (*Plasmodium* ssp.), the immunogenic regions of the malaria proteins, targeted by host antibodies, display positive selection for the ability to evade the host's defense mechanism (Hughes 1991, 1992). On the host side, in cattle the *CD45* locus, encoding a plasma membrane-associated tyrosine phosphatase, is thought to be subject to strong pathogen-driven selection for diversity (Ballingall et al. 2001).

Rapid protein evolution is also observed following gene duplication (e.g., type II 18S rDNA genes, Carranza et al. 1999; *Acp70A* gene in *Drosophila subobscura*, Cirera and Aguade 1998; mammalian *defensin* gene, Hughes and Yeager 1997) due to either relaxed constraints (Li 1985), advantageous mutations (Ohta 1991a), or both. Many gene families are the result of duplications followed by rapid evolution (e.g., chalcone synthase (CHS) multigene family in *Ipomea*, Durbin et al. 1995; MHC complex, Hughes and Yeager 1997; pregnancy-associated glycoproteins (PAG) in *Artiodactyla*, Garbayo et al. 2000; *PHY* gene family in angiosperms, Alba et al. 2000).

Evolution of the Adaptively Radiating Hawaiian Silversword Alliance

The Hawaiian silversword alliance

The research discussed in this dissertation utilizes two plant systems. One system includes species in the *Compositae* (Asteraceae: Heliantheae—Madiinae): the Hawaiian silversword alliance and their closest living relatives, the North American tarweeds. The Hawaiian silversword alliance was chosen for this study because it has long been considered a prime example of an insular plant adaptive radiation (Carlquist 1980; Carr 1985; Robichaux et al. 1990; Raven et al. 1992; Baldwin and Robichaux 1995; Baldwin 1997). The alliance is a group of species which arose from a single ancestor (Carlquist 1957, 1959a; Witter 1986; Witter and Carr 1988; Crins et al. 1988; Crins and Bohm 1990; Robichaux 1985; Robichaux and Canfield 1985; Robichaux et al. 1990; Baldwin et al. 1990) and are found exclusively on the Hawaiian archipelago (Carr 1985). The founding individual was the product of an interspecific hybridization of two different species of North American tarweeds (Barrier et al. 1999). The alliance ancestor arrived on the Hawaiian Islands no earlier than 5-6 million years ago (mya), around the time of the formation of the oldest of the major islands, Kaua'i (Baldwin and Sanderson 1998). Since this arrival, the species have undergone a spectacular adaptive radiation with explosive speciation, spreading across the islands into a wide array of habitats and diversifying into the thirty morphologically distinct species seen today (Carr 1985, 1999).

Despite the striking morphological differences, these species are genetically very similar and have not yet become reproductively isolated by sterility barriers. Vigorous hybrids have been formed between many of the alliance members within and between genera as well as between some members of the Hawaiian silversword alliance and North American tarweeds (Carr and Kyhos 1981, 1986; Carr et al. 1989, 1996; Baldwin et al. 1991, 1998; Baldwin 1997). The significant morphological diversity resulting from this relatively recent adaptive radiation makes this an interesting group for studying the evolutionary genetics of adaptive divergence.

Morphology and ecology of the Hawaiian silversword alliance

In 1985, after almost a decade of research in the area, Gerald D. Carr published a monograph on the Hawaiian Madiinae summarizing all that was known at that time (Carr 1985). What follows is a summary of the morphology and ecology of this group, taken mostly from this monograph with some recent updates.

The Hawaiian silversword alliance is comprised of 30 perennial species divided into three endemic genera: *Argyroxiphium*, *Dubautia*, and *Wilkesia*. This group is found exclusively on the six major islands of the Hawaiian archipelago: Kaua'i, O'ahu, Moloka'i, Lana'i, Maui, and Hawai'i. Islands of the Hawaiian archipelago were formed through the buildup of lava from a "hot spot" in the ocean floor followed by their slow drift to the Northwest. The major islands today range in age from 0.4 million years for the youngest, Hawai'i, at the Southeast end of the chain, to 5.1 million years for the oldest, Kaua'i (Clague and Dalrymple 1987). This

continuing process of rebirth, along with erosion and subsidence, has shaped the vast array of habitats now occupied by the alliance species: dry alpine deserts of cinder and lava, dry scrub and woodland, wet scrub and forest, and bogs. All but five species of the alliance are single-island endemics with half of these single-island species isolated to a single volcano, to an area of the island, or even to a specific lava flow. These various habitats extend from 75 m to 3750 m in elevation (cf. Smathers and Mueller-Dombois 1974) and on average receive from less than 40 cm to more than 1230 cm of annual rainfall.

Along with this environmental diversity, there is a remarkable amount of morphological variation between species. Members of the genera *Argyroxiphium* and *Wilkesia* are either monocarpic or polycarpic rosette plants with the placement of their rosette either at the base of the plant or elevated up to five meters on the woody stem. The members of *Dubautia* are more varied and include trees, shrubs, mat-forming shrubs, cushion plants, and a liana. Stems of the alliance range from subherbaceous stalks only a few millimeters in diameter to massively woody trunks up to at least 5 dm in diameter. Their leaves differ in size (from 5 mm to 500 mm in length), in shape (including linguate, elliptic, lanceolate, linear, ovate, and oblong) and in arrangement on the stems (opposite, whorled, or helical). Their ray or disk flowers are seen in an assortment of colors including white, yellow, and purple. Very sticky resinous secretions from glandular trichomes are what give their North American ancestors their "tarweed" name, but even the amount of resinous secretion varies among the

Hawaiian alliance species. With all of this morphological variation, it is easy to see how difficult it initially was to recognize these species as members of a single group.

A monophyletic group

Early on it was unclear whether the Hawaiian silversword alliance species arose from a single founding event or from several independent arrivals. In the late 1950s, Sherwin Carlquist's examination of the vegetative anatomy of *Argyroxiphium*, *Dubautia*, and *Wilkesia* provided structural evidence that the Hawaiian Madiinae originated from a single colonizing species (Carlquist 1957, 1959a). Later, G. D. Carr and D. W. Kyhos added to this evidence their cytogenetic studies of hybrids. They noted that there is still enough genetic similarity between alliance species that spontaneous interspecific and intergeneric hybrids are found at high frequency in nature (Carr and Kyhos 1981) and are easily produced in the lab (Carr and Kyhos 1986). By studying the chromosomal pairings of these hybrids, they were able to identify eight different reciprocal translocations and an aneuploid reduction, each characterizing a group of alliance species (Carr and Kyhos 1986). Despite these chromosomal alterations, hybrids between groups were not sterile, but produced some proportion of viable pollen. Further evidence to support the relationship of these species from a single founding event include allozymic data (Witter 1986; Witter and Carr 1988), flavonoid variation (Crins et al. 1988; Crins and Bohm 1990), ecophysical properties (Robichaux 1985; Robichaux and Canfield 1985; Robichaux et al. 1990), and chloroplast DNA data (Baldwin et al. 1990). Later, molecular approaches were

used to analyze putative hybrid taxa and their potential parental species (Carr et al. 1989; Baldwin 1997; Baldwin et al. 1998), providing further evidence for a single colonizing ancestral species by showing that the Hawaiian silversword alliance is a genetically cohesive group.

Tarweed ancestry

As the monophyly of the silverswords was being resolved, their ancestry was also in question. As early as 1852, an affinity of *Argyroxiphium* species to the American tarweeds was suggested by A. Gray (1852), but was disputed by D.D. Keck in 1936, based on presumed morphological dissimilarities and the magnitude of the oceanic barrier to migration from mainland North America to Hawai'i (Keck 1936). This uncertainty is understandable since the tarweeds do not share the life-form diversity of the silversword alliance species, but are mostly ephemeral herbs (Baldwin and Sanderson 1998). Even in 1948, relationships to the Hawaiian species were still considered obscure in "Derivation of the flora of the Hawaiian islands" (Fosberg 1948). A decade later, S. Carlquist hypothesized that the Hawaiian silversword alliance arose from the American tarweeds based on his anatomical studies of their wood, trichome, and vegetative structures (Carlquist 1958a, b, 1959b). Carlquist addressed the problem of the extremely long-distance dispersal needed by concluding that the sticky bracts and fruit appendages of the perennial tarweeds were perfect adaptations for animal-mediated dispersal (Carlquist 1966, 1967). The prospect of this method of dispersal is supported by the observation that North American migratory

birds and accidental arrivals are regularly sighted on the Hawaiian Islands (cf. Shallenberger 1981).

As the alliance-tarweed relationship became accepted, the classification schemes of the sunflower family in 1977 easily accommodated the Hawaiian silversword alliance in the subtribe Madiinae of the tribe Heliantheae (Stuessy 1977). This association was further reinforced when Baldwin et al. (1991) used restriction-site comparisons of chloroplast DNA in the Hawaiian alliance and perennial continental tarweeds to reveal a close affinity of the monophyletic Hawaiian silversword alliance not just to the North American tarweeds, but to two specific diploid members of the group, *Kyhosia bolanderi* (formerly *Madia bolanderi*) and *Carlquistia muirii* (formerly *Raillardiopsis muirii*). They further supported their phylogenetic analyses by producing vigorous hybrids of *Dubautia laevigata* X *C. muirii* and (*D. knudsenii* X *D. laxa*) X *K. bolanderi*. This study refined the ancestral relationship of the alliance to specific lines of continental tarweeds by showing that the monophyletic Hawaiian alliance group was nested within a western American lineage of the genera *Kyhosia* and *Carlquistia*. This nested phylogeny was supported further by sequence analyses of ribosomal DNA ITS and ETS regions (Baldwin and Sanderson 1998, Baldwin and Markos 1998). While the ancestry of the alliance was now solidly linked to a branch of the North American tarweeds, the specific nature of the founding Hawaiian silversword alliance individual was still unclear.

Allopolyploidization

Initially researchers believed the polyploid ancestor to the Hawaiian silversword alliance would still be extant on the mainland, so they focused on the existing polyploid continental tarweeds (n = 14) (Carr et al. 1996). Carr et al. (1996) produced vigorous hybrids between polyploid Hawaiian alliance and diploid continental tarweed species and successfully crossed the hybrid of two tarweeds (*Kyhosia bolanderi* X *Carlquistia muirii*) with an alliance species (*D. scabra*). After looking at meioses of hybrid pollen as well as the morphologies of the hybrids and their parents, they proposed that these data suggest an allopolyploid origin for the Hawaiian silversword alliance genome.

Genetic evidence for the allopolyploid origin is uncovered in the research described here in Chapter 2 (Barrier et al. 1999). By examining the sequences of two nuclear regulatory genes in members of the Hawaiian silversword alliance and the North American tarweeds, it is determined that the founding member of the alliance was, in fact, the result of an allopolyploidization event. In the sequencing of these genes, it is discovered that the Hawaiian alliance species have two distinct copies of each gene (called *A* and *B*), while the diploid North American tarweed species has only a single copy of each. Phylogenetic analysis of these genes reveals that the alliance *A* and *B* copies each form a monophyletic group and that each group clusters with specific individuals of the North American tarweeds: *Carlquistia muirii*, *Anisocarpus scabridus* (formerly *Raillardiopsis scabrida*), and possibly *Harmonia nutans* (formerly *Madia nutans*). These individuals are from the same subgroup of

tarweeds Baldwin et al. (1991) initially established as the closest-living ancestors to the alliance. From this it is determined that the founding member of the Hawaiian silversword alliance was the result of an interspecific hybridization between members of the lineages including *C. muirii* and *A. scabridus*. Further support for this comes from previous reports of a vigorous hybrid combination between *C. muirii* and *A. scabridus* (Baldwin 1989; Kyhos, Carr and Baldwin 1990). By determining that the founding member of the Hawaiian silversword alliance was most likely an allopolyploid from the North American tarweeds, this information can now be used to examine possible evolutionary mechanisms for the spectacular adaptive radiation of the alliance species.

Adaptive radiation

Adaptive radiations are particularly interesting because they involve the ecological differentiation of individuals within a lineage and are often associated with increases in morphological diversity and rates of speciation (Givnish and Sytsma 1997). They are especially noticeable on islands where the relative isolation leads to "numerous vacant niches or ecological opportunities for island immigrants" (Carlquist 1974). Volcanic islands like the Hawaiian archipelago can provide a wide range of climatically and physiographically differentiated habitats due to the recurrent volcanic eruptions and constant erosion (e.g., rainfall) (Knapp 1975). The isolation of the islands also means fewer species overall to compete for resources. The likelihood that

a lineage will undergo an adaptive radiation is greatly increased by this absence of competition (Mayr 1942; Simpson 1953; Skelton 1993).

It is believed that the ancestral Hawaiian silversword alliance species arrived on the oldest of the major islands, Kaua'i, and radiated out from there. Baldwin and Sanderson were able to determine a maximum age for the divergence of the Hawaiian silversword alliance by using paleoclimatic and fossil data as an "external" calibration and rate-constant ribosomal DNA sequence evolution of the Internal Transcribed Spacer (ITS) region (Baldwin and Sanderson 1998). Their maximum-age estimate for the most recent common ancestor of the alliance was 5.2 ± 0.8 mya, which falls within the history of the modern high islands (5.1 mya for Kaua'i). This would indicate that the ancestral alliance species most likely arrived on the island of Kaua'i when it was still young and open to colonization. The alliance members then spread out into the various habitats available on the island and made their way southeast to each new island as it arose from the ocean. It is possible to trace the evolution of a character across the islands by examining the patterns of variability of that character. Carr and Kyhos (1981, 1986) hypothesized that a chromosome number of n = 14 was the alliance's ancestral type. They demonstrated that plants with n = 14 occurred on the older islands of the Hawaiian archipelago, while the n = 13 plants were restricted to the younger islands. Molecular phylogenetic data from cpDNA nuclear ribosomal DNA confirmed the dysploid descent from n = 14 to n = 13 and verified monophyly of the n = 13 taxa (Baldwin et al. 1990; Baldwin and Robichaux 1995; see Baldwin 1997). This was later supported by cytogenetic data from Witter and Carr (1998)

which found the n = 14 *Dubautia* species to be more highly divergent than the n = 13 species and therefore older. The aneuploid reduction (from n = 14 to 13) in a monophyletic group of *Dubautia* species was shown to be correlated with several major ecological, morphological, and physiological differences (Robichaux et al. 1990).

The diversification of the Hawaiian silversword alliance into the many species currently observed was most likely due to these isolated populations adapting to their particular habitats, as is seen in studies of other island endemics (Crawford and Stuessy 1997). The genetic similarity between alliance species, despite their ecological and morphological divergence, has also been observed in other studies of plant and animal adaptive radiations (Rieseberg and Swensen 1995). The allopolyploid nature of the Hawaiian silversword alliance members may have added to their ability to adapt to the various habitats.

Polyploidy

Polyploidy is known as a common mechanism of speciation in flowering plants (Grant 1981; Levin 1983; Carlquist 1995; Stebbins 1971). Polyploids often exhibit an increased rate of evolutionary change because of an initially high amount of genetic redundancy due to duplicated gene loci (Stebbins 1971). Allopolyploids have the added benefit of two complete, but dissimilar genomes, providing a wealth of genetic diversity in a single individual (Levin 1983). Wendel (2000) reviewed the processes and mechanisms involved in the evolution of polyploid plant genes and

genomes and discussed how experiments of synthetic allopolyploids have provided much of the knowledge about the 'dynamic nature of polyploidy genomes' (Soltis and Soltis 1995).

In one study of natural allopolyploids, Small, Ryburn and Wendel (1999) studied the nucleotide diversity at the homoeologous *AdhA* loci in allotetraploid cotton species, *Gossypium hirsutum,* using the closely-related *G. barbadense* for comparison. These are two of the five allotetraploid cotton species believed to have come from the same allopolyploidization event 1-2 mya. The pair of nuclear *AdhA* loci represent the two diploid lineages (A and D) which formed the ancestral allotetraploid plant. Instead of increased diversity from the polyploidization, Small, Ryburn and Wendel (1999) observed estimates of nucleotide diversity for *AdhA* that were lower than that of any other plant nuclear gene described. This is believed to be due to a history of genetic bottlenecks, rapid population expansion and a slow mutation rate. The nucleotide diversity for *AdhA* from the D-subgenome is approximately twice as high as that from the A-subgenome, indicating that the two subgenomes of these allotetraploids may be subject to different evolutionary pressures.

In another study, Song et al. (1995) produced reciprocal synthetic allopolyploids in *Brassica* in order to study the effects of polyploidy over five generations of selfing. Using Southern hybridizations, in each generation they detected a high frequency of genomic changes including recovery and loss of parental fragments and frequent appearances of novel fragments. Similar experiments of synthetic allopolyploids in wheat (Feldman et al. 1997; Liu, Vega and Feldman 1998;

Liu et al. 1998) also produced similar striking examples of polyploidy-induced genomic changes, particularly with rapid, non-random sequence elimination. In the case of the Hawaiian silversword alliance, the profound adaptive radiation may have been influenced by a similar diversification of the allopolyploid genome.

Identifying Rapidly-Evolving Genes in *Arabidopsis*

The latter part of this dissertation research focuses on the mustard plant, *Arabidopsis thaliana (*Brassicaceae*)* and its relative, *Arabidopsis lyrata*. Here an evolutionary Expressed Sequence Tag (EST) approach (Swanson et al. 2001) is used to identify potential fast-evolving genes in *Arabidopsis*. A subset of these rapidly-evolving ESTs is then chosen and a population analysis performed to examine the variation within *A. thaliana* as well as between species. This EST approach allows for the identification of a set of genes which may be involved in the divergence of these species.

Evolutionary EST approach

An evolutionary EST approach is a method of comparing sequences of expressed genes from two closely-related species in order to identify genes involved in functional divergence. Swanson et al. (2001) used this approach to identify rapidly-evolving male reproductive proteins in *Drosophila*. In their study, EST sequences from the male accessory gland of *Drosophila silmulans* were compared to their orthologs in *D. melanogaster*, using the complete genomic sequence now available (Adams et al. 2000). Of 176 unique genes isolated, nineteen were identified as showing evidence of rapid divergence between these closely related species by an excess of replacement over synonymous changes.

For this dissertation research, a cDNA library is created from *Arabidopsis lyrata* flowering material and a collection of ESTs sequenced. Homologous sequences from *Arabidopsis thaliana* are found by searching online databases with NCBI's BLAST software. By observing the pairwise differences for each of the 304 unique genes isolated, 21 potential rapidly-evolving genes with an ω value (Ka/Ks) greater than one are identified. Seven of these genes are then sequenced in a set of *A. thaliana* and *A. lyrata* individuals in order to verify the between-species differences and further examine the pattern of differences both within *A. thaliana* and between species. The study identifies a group of genes which may be undergoing positive selection in *Arabidopsis*.

Arabidopsis – a model for plant research

Arabidopsis thaliana (L.) Heynh. is the chosen organism because it is a well-developed model organism for plant research with a long history. Johannes Thal, a 16th century German physician, first described *Arabidopsis* in a Flora Book of the Harz Mountains in 1577 and was later honored with the "*thaliana*" name (Redei 1992). It wasn't until 1907 that the first *Arabidopsis* research was conducted by Friedrich Laibach, who published a significant paper on the individuality and continuity of chromosomes after correctly counting the 10 somatic and 5 meiotic chromosomes of *Arabidopsis* (Redei 1992). Later, in a 1943 publication, Laibach listed many of the same favorable features of *A. thaliana* which have allowed it to develop into such a strong model system for plants today (Redei 1992).

Arabidopsis thaliana is a weed in the mustard family comprised of a rosette of small leaves with a main stem topped by an inflorescence. The plant is very small with a height up to just 30 or 40 cm and mature flowers about 3 mm long and 1 mm in diameter (Meyerowitz 1987) which allows for cultivation in limited space. *Arabidopsis* has the benefit of rapid development in the lab, where seeds germinate in a few days and sometimes grow to flowering in just four to five weeks (Meyerowitz 1987). The mature *Arabidopsis* plant can then produce more than 10,000 seeds (Meyerowitz 1987). The standard *Arabidopsis* flower has four whorls from outside to inside containing: four sepals, four white petals, six stamens (two short and four long), and an ovary of two carpals in the center. The simple structures of this plant have allowed for the development of basic models for characterizing gene interactions like the ABC model of flower development (Bowman et al. 1991). A large number of genes involved in *Arabidopsis* development, from germination to floral structures, have been characterized through the analysis and mapping of natural and induced mutants (Meyerowitz 1987).

Population genetics of *Arabidopsis*

Arabidopsis thaliana is believed to be native to Western Eurasia (Hoffman 2002). Currently, *A. thaliana* ecotypes are widely distributed in open or disturbed habitats in temperate regions of Eurasia, North Africa, and North America (O'Kane and Al-Shebaz 1997; Mitchell-Olds 2001). Many molecular studies have examined the population genetics of *Arabidopsis* using individual genes (Purugganan and

Suddith 1998, 1999; Aguade 2001) groups of genes (Olsen et al. 2002; Bustamante et al. 2002), microsatellite markers (Clauss, Cobban and Mitchell-Olds 2002), and AFLP polymorphisms (Miyashita, Kawabe and Innan 1999).

Population genetic studies of *Arabidopsis thaliana* at individual loci provide evidence of the evolutionary forces acting at these specific genes. Purugganan and Suddith (1998) observed an excess of intraspecific replacement polymorphisms in alleles of the *Arabidopsis CAULIFLOWER (CAL)* floral regulatory gene, with 16 of the 21 coding-region polymorphisms being replacement substitutions. Comparisons between *A. thaliana* and *A. lyrata CAL* alleles revealed 13 fixed replacement differences out of the 28 total fixed differences between these closely-related species. The McDonald-Kreitman test demonstrated that this data is significantly different from expectations of neutrality, suggesting positive selection may be acting at this locus in *Arabidopsis thaliana.* A similar pattern of increased intraspecific replacement polymorphism was observed in two other floral regulatory genes, *APETALA3 (AP3)* and *PISTILLATA (PI)* by Purugganan and Suddith in 1999. The results from both of these studies are suggestive of a recent species expansion. In contrast, a population genetic study of two genes, *FAH1* and *F3H*, coding for enzymes in the phenylpropanoid pathway, revealed no evidence for significant replacement polymorphism (Aguade 2001).

Studies of individual genes can only provide snapshots of the evolutionary processes occurring throughout the genome. In order to observe the overall evolutionary trends of a genome, groups of loci must be examined across the genome.

A study by Olsen et al. (2002) examined six regulatory genes involved at different stages of the floral development pathway. They observed different patterns of molecular evolution in these regulatory genes along the pathway, with the two earlier acting genes showing evidence of adaptive evolution. Bustamante et al. (2002) examined a group of 12 nuclear genes of diverse function in *Arabidopsis* and a group of 34 genes in *Drosophila* in order to compare the pattern and intensity of selection in the two genomes. This study revealed contrasting trends of gene substitution, possibly attributable to the different mating systems. They found evidence for predominantly beneficial gene substitutions in the outcrossing *Drosophila* species, but mostly detrimental substitutions in the self-fertilizing *Arabidopsis.* A similar effect of mating system was also observed by Miyashita, Kawabe and Innan (1999) in an Amplified Fragment Length Polymorphism (AFLP) analysis of 374 polymorphic bands in *Arabidopsis*. In observing the levels and patterns of DNA variation in *Arabidopsis*, they observed an excess of singleton variation, often associated with inbreeding. Their data also suggested a recent spread of this species with limited migration between populations.

Relatives of *Arabidopsis thaliana*

A. thaliana was compared to the closely-related species *A. lyrata* in the dissertation research in order to eliminate the problem of multiple substitutions which obscure selection as species become more distant (Swanson et al. 2001). From phylogenetic analysis of rDNA ITS regions, Koch et al. (1999) produced maximum

parsimony and Neighbor-joining trees that agree that *A. halleri* and the three subspecies of *A. lyrata* are most closely related to *A. thaliana*. Koch et al. (2000) determined that *A. thaliana* diverged from *A. halleri* and *A. lyrata* approximately 5 million years ago (mya), relatively recently on the evolutionary timescale.

Despite their close relationship, *A. thaliana* and *A. lyrata* do exhibit significant differences in chromosome number (*A. thaliana* 2n = 10, *A. lyrata* 2n = 16), mating type, ecology, and life history (Clauss, Cobban and Mitchell-Olds 2002). *A. thaliana* is 99.7% selfing (Abbott and Gomes 1989) while *A. lyrata* is 99% outcrossing and mostly self-incompatible (Savolainen et al. 2000). In a population genetic study using microsatellite markers (Clauss, Cobban and Mitchell-Olds 2002), the effects of the different mating systems could be seen in the genetic variability of these species. The selfing *A. thaliana* was shown to have only 8% polymorphic loci while loci of the self-incompatible *A. lyrata* ssp. *petraea* were 86% polymorphic. This is consistent with predictions that outcrossing species maintain higher levels of genetic variation (Charlesworth and Charlesworth 1995).

Tools for genetic research in *Arabidopsis*

Currently, the *Arabidopsis thaliana* genetic toolkit includes DNA libraries, mapping populations, YACs, BACs, and mutants all available for molecular analysis from stock centers (Koch et al. 1999). Extensive physical and genetic maps of all five chromosomes are available online at the *Arabidopsis* Genome Initiative (AGI) website www.Arabidopsis.org. In 2000, 115.4 Mb of the 125 Mb *A. thaliana* genome

(Columbia ecotype) was sequenced (The *Arabidopsis* Genome Initiative 2000) and is also available on the AGI website.

With the genome sequenced and available online, the job of identifying and properly annotating genes in the *Arabidopsis* genome has been taken on by many groups such as The Institute for Genomic Research (TIGR – www.tigr.org/tdb), the Munich Information center for Protein Sequences (MIPS – mips.gsf.de) and the Kazusa *Arabidopsis* data opening site (KAOS – www.kazula.or.jp/kaos). These groups and others provide various database tools and options for analyzing and visualizing regions of the *Arabidopsis* genome. As of May 2002, the MIPS *A. thaliana* database (MAtDB) website (http://mips.gsf.de/proj/thal/db/index.html) lists 25,628 gene/protein entries, but only 2565 of the entries have been classified so far. As more genomes are sequenced, comparative genomics will allow more genes to be identified through DNA and protein sequence homology.

Literature Cited

Abbott, R. J. and Gomes, M. F. (1989) Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. *Heredity* **62**:411-418.

Adams, M. D., Celniker, S. E., Holt, R. A., et al. (2000) The genome sequence of *Drosophila melanogaster. Science* **287**: 2185-2195.

Aguade, M. (2001) Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the *FAH1* and *F3H* genes, in *Arabidopsis thaliana*. *Mol. Biol. Evol.* **18**: 1-9.

Aidoo, M., Terlouw, D. J., Kolczak, M., McElroy, P. D., ter Kuile, F. O., Kariuki, S., Nahlen, B. L., Lal, A. A. and Udhayakumar, V. (2002) Protective effects of the sickle cell gene against malaria morbidity and mortality. *The Lancet* **359**: 1311-1312.

Akashi, H. (1995) Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in *Drosophila* DNA. *Genetics* **139**: 1067-1076.

Alba, R., Kelmenson, P. M., Cordonnier-Pratt, M. M. and Pratt, L. H. (2000) The phytochrome gene family in tomato and the rapid differential evolution of this family in angiosperms. *Mol. Biol. Evol.* **17**: 362-373.

Andolfatto, P. and Przeworski, M. (2001) Regions of lower crossing over harbor more rare variants in African populations of *Drosophila melanogaster*. *Genetics* **158**: 657-665.

Aquadro, C. F. (1997) Insights into the evolutionary process from patterns of DNA sequence variability. *Curr. Opin. Genet. Dev.* **7**: 835-840.

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana. Nature* **408**: 796-815.

Armbrust, E. V. and Galindo, H. M. (2001) Rapid evolution of a sexual reproduction gene in centric diatoms of the genus *Thalassiosira*. *Appl. Environ. Microbiol.* **67**: 3501-3513.

Avedisov, S. N., Rogozin, I. B., Koonin, E. V. and Thomas, B. J. (2001) Rapid evolution of a *cyclin A* inhibitor gene, roughex, in *Drosophila. Mol. Biol. Evol.* **18**: 2110-2118.

Baldwin, B. G. (1989) Chloroplast DNA phylogenetics and biosystematic studies in Madiinae (Asteraceae). Ph.D. dissertation, University of California, Davis.

Baldwin, B. G. (1997) Adaptive radiation of the Hawaiian silversword alliance: congruence and conflict of phylogenetic evidence from molecular and non-molecular investigations. In: Molecular evolution and adaptive radiation. Givnish, T. J. and Sytsma, K. J. (eds.) Cambridge University Press, Cambridge, England. Pp. 103-128.

Baldwin, B. G. and Markos, S. (1998) Phylogenetic utility of the external transcribed spacer (ETS) of 18S-26S rDNA: Congruence of ETS and ITS trees of Calycadenia (Compositae). *Mol. Phylogenet. Evol.* **10**: 449-463.

Baldwin, B. G. and Robichaux, R. H. (1995) Historical biogeography and ecology of the Hawaiian silversword alliance: New molecular phylogenetic perspectives. In: Hawaiian biogeography: evolution on a hot spot archipelago. Wagner, W. L., Funk, V. A. (eds.), Smithsonian Institution Press, Washington, D.C. Pp. 259-287.

Baldwin, B. G. and Sanderson, M. J. (1998) Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proc. Natl. Acad. Sci. USA* **95**: 9402-9406.

Baldwin, B. G., Kyhos, D. W. and Dvorak, J. (1990) Chloroplast DNA evolution and adaptive radiation in the Hawaiian silversword alliance (Asteraceae-Madiinae). *Ann. Mo. Bot. Gard.* **77**: 96-109.

Baldwin, B. G., Kyhos, D. W., Dvorak, J. and Carr, G. D. (1991) Chloroplast DNA evidence for a North American origin of the Hawaiian silversword alliance (Asteraceae). *Proc. Natl. Acad. Sci. USA* **88**: 1840-1843.

Baldwin, B. G., Kyhos, D. W. and Dvorak, J. (1998) Chloroplast DNA evolution of the Hawaiian Madiinae (Compositae). *Amer. J. Bot.* **75**: 152.

Ballingall, K. T., Waibochi, L., Holmes, E. C., Woelk, C. H., MacHugh, N. D., Lutje, V. and McKeever, D. J. (2001) The CD45 locus in cattle: allelic polymorphism and evidence for exceptional positive natural selection. *Immunogenetics* **52**: 276-283.

Barrier, M., Baldwin, B. G., Robichaux, R. H. and Purugganan, M. D. (1999) Interspecific hybrid ancestry of a plant adaptive radiation: Allopolyploidy of the Hawaiian silversword alliance (Asteraceae) inferred from floral homeotic gene duplications. *Mol. Biol. Evol.* **16**: 1105-1113.

Barrier, M., Friar, E., Robichaux, R. H. and Purugganan, M. D. (2000) Interspecific evolution in plant microsatellite structure. *Gene* **241**: 101-105.

Barrier, M., Robichaux, R. H. and Purugganan, M. D. (2001) Accelerated regulatory gene evolution in an adaptive radiation. *Proc. Natl. Acad. Sci. USA* **98**: 10208-10213.

Barton, N. H. (2000) Genetic hitchhiking. *Phil. Trans. R. Soc. Lond. B* **355**: 1553-1562.

Begun, D. J. and Aquadro, C. F. (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**: 519-520.

Begun, D. J., Whitley, P., Todd, B. L., Waldrip-Dail, H. M. and Clark, A. G. (2000) Molecular population genetics of male accessory gland proteins in *Drosophila*. *Genetics* **156**: 1879-1888.

Bergstrom, T. and Gyllensten, U. (1995) Evolution of MHC class II polymorphism: the rise and fall of class II gene function in primates. *Immunol. Rev.* **143**: 13-31.

Bishop, J. G., Dean, A. M. and Mitchell-Olds, T. (2000) Rapid evolution in plant chitinases: molecular targets of selection in plant-pathogen coevolution. *Proc. Natl. Acad. Sci. USA* **97**: 5322-5327.

Blanco-Gelaz, M. A., Lopez-Vazquez, A., Garcia-Fernandez, S., Martinez-Borra, J., Gonzalez, S. and Lopez-Larrea, C. (2001) Genetic variability, molecular evolution, and geographic diversity of *HLA-B27*. *Hum. Immunol.* **62**: 1042-1050.

Bowman, J. L., Smyth, D. R. and Meyerowitz, E. M. (1991) Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* **112**: 1-20.

Boyson, J. E., Iwanaga, K. K. et al. (1999) Evolution of a new nonclassical MHC class I locus in two Old World primate species. *Immunogenetics* **49**: 86-98.

Burstein, Y., Breiner, A. V. et al. (1982) Recent duplication and germ-line diversification of rat immunoglobulin kappa chain gene joining segments. *Proc. Natl. Acad. Sci. USA* **79**: 5993-5997.

Bustamante, C. D., Townsend, J. P. and Hartl, D. L. (2000) Solvent accessibility and purifying selection within proteins of *Escherichia coli* and *Salmonella enterica*. *Mol. Biol. Evol.* **17**: 301-308.

Bustamante, C. D., Nielsen, R., Sawyer, S., Olsen, K. M., Purugganan, M. D. and Hartl, D. L. (2002) The cost of inbreeding in *Arabidopsis*. *Nature.***416**: 531-534.

Cadavid, L. F., Hughes, A. L. and Watkins, D. I. (1996) MHC class I-processed pseudogenes in New World primates provide evidence for rapid turnover of MHC class I genes. *J. Immunol.***157**: 2403-2409.

Carlquist, S. (1957) Leaf anatomy and ontogeny in *Argyroxiphium* and *Wilkesia* (Compositae) *Am. J. Bot.* **44**: 696-705.

Carlquist, S. (1958a) Wood anatomy of Heliantheae (Compositae). *Tropical Woods* **108**: 1-30.

Carlquist, S. (1958b) Structure and ontogeny of glandular trichomes of Madiinae (Compositae). *Am. J. Bot.* **45**: 675-682.

Carlquist, S. (1959a) Vegetative anatomy of *Dubautia, Argyroxiphium,* and *Wilkesia* (Compositae). *Pacific Sci.* **13**: 195-210.

Carlquist, S. (1959b) Studies on Madiinae: Anatomy, cytology, and evolutionary relationships. *Aliso* **4**: 171-236.

Carlquist, S. (1966) Biota of long-distance dispersal. II. Loss of dispersibility in pacific Compositae. *Evolution* **20**: 30-48.

Carlquist, S. (1967) Biota of long-distance dispersal. V. Plant dispersal to pacific islands. *Bull. Torrey Bot. Club* **94**: 129-162.

Carlquist, S. (1974) Island Biology. Columbia University Press, New York.

Carlquist, S. (1980) Hawaii: A Natural History. Pacific Tropical Botanical Garden, Lawai, Kauai.

Carlquist, S. (1995) Introduction. In: Hawaiian biogeography – evolution on a hot spot archipelago. Wagner, W. L., Funk, V. A. (eds.) Smithsonian Institution, Washington. P1.

Carr, G. D. (1999) A reassessment of *Dubautia* (Asteraceae: Heliantheae—Madiinae) on Kaua'i. *Pacific Science* **53**: 144-158.

Carr, G. D. (1985) Monograph of the Hawaiian Madiinae (Asteraceae): *Argyroxiphium, Dubautia,* and *Wilkesia*. *Allertonia* **4**: 1-123.

Carr, G. D. and Kyhos, D. W. (1981) Adaptive radiation in the Hawaiian silversword alliance (Compositae-Madiinae). I. Cytogenetics of spontaneous hybrids. *Evolution* **35**: 543-556.

Carr, G. D. and Kyhos, D. W. (1986) Adaptive radiation in the Hawaiian silversword alliance (Compositae-Madiinae). II. Cytogenetics of artificial and natural hybrids. *Evolution* **40**: 959-976.

Carr, G. D., Robichaux, R. H., Witter, M. S. and Kyhos, D. W. (1989) Adaptive radiation of the Hawaiian silversword alliance (Compositae-Madiinae): a comparison with Hawaiian picture-winged *Drosophila*. In: Genetics, Speciation, and the Founder Principle. Giddings, L. V., Kaneshiro, K. Y. and Anderson, W. W. (eds.) Oxford University Press, Oxford. Pp. 79-97.

Carr, G. D., Baldwin, B. G. and Kyhos, D.W. (1996) Cytogenetic implications of artificial hybrids between the Hawaiian silversword alliance and North American tarweeds (Asteraceae, Heliantheae-Madiinae). *Am. J. Bot.* **83**: 653-660.

Carranza, S., Baguna, J. and Riutort, M. (1999) Origin and evolution of paralogous rRNA gene clusters within the flatworm family *Dugesiidae* (Platyhelminthes, Tricladida). *J. Mol. Evol.* **49**: 250-259.

Charlesworth, D. and Charlesworth, B. (1995) Quantitative genetics in plants: the effect of the breeding system on genetic variability. *Evolution* **49**: 911-920.

Cirera, S. and Aguade, M. (1998) Molecular evolution of a duplication: the sex-peptide (Acp70A) gene region of *Drosophila subobscura* and *Drosophila madeirensis. Mol. Biol. Evol.* **15**: 988-996.

Clague, D. A. and Dalrymple, G. B. (1987) The Hawaiian-Emperor volcanic chain, part I: geologic evolution. In: Volcanism in Hawaii, U. S. Geological Survey Professional Paper 1350. Decker, R. W., Wright, T. L. and Stauffer, P. H. (eds.) Washington, DC: U.S. Government Printing Office. Pp. 5-54.

Clauss, M. J., Cobban, H. and Mitchell-Olds, T. (2002) Cross-species microsatellite markers for elucidating population genetic structure in *Arabidopsis* and *Arabis* (Brassicaceae). *Mol. Ecol.* **11**: 591-601.

Crandall, K. A., Kelsey, C. T., Imamichi, H., Lane, H. C. and Salzman, N. P. (1999) Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. *Mol. Biol. Evol.* **16**: 372-382.

Crawford, D. J. and Stuessy, T. F. (1997) Plant speciation in oceanic islands. In: Evolution and diversification of land plants. K. Iwatsuki, P. H. Raven (eds.) Tokyo: Springer. Pp 249-267.

Crins, W. J., Bohm, B. A. and Carr, G. D. (1988) Flavonoids as indicators of hybridization in a mixed population of lava-colonizing Hawaiian tarweeds (Asteraceae, Heliantheae, Madiinae). *Syst. Bot.* **13**: 567-571.

Crins, W. J. and Bohm, B. A. (1990) Flavonoid diversity in relation to systematics and evolution of the tarweeds. *Ann. Mo. Bot. Gard.* **77**: 73-83.

Donohue, K., Pyle, E. H., Messiqua, D., Heschel, M. S. and Schmitt, J. (2001) Adaptive divergence in plasticity in natural populations of *Impatiens capensis* and its consequences for performance in novel habitats. *Evolution* **55**: 692-702.

Duda, T. F., Jr. and Palumbi, S. R. (1999) Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc. Natl. Acad. Sci. USA* **96**: 6820-6823.

Durbin, M. L., Learn, G. H., Huttley, G. A. and Clegg, M. T. (1995) Evolution of the chalcone synthase gene family in the genus *Ipomoea*. *Proc. Natl. Acad. Sci. USA* **92**: 3338-3342.

Erlich, H. A. and Gyllensten, U. B. (1991) The evolution of allelic diversity at the primate major histocompatibility complex class II loci. *Hum. Immunol.* **30**: 110-118.

Feldman, M., Liu, B., Segal, G., Abbo, S., Levy, A. A. and Vega, J. M. (1997) Rapid elimination of low-copy DNA sequences in polyploidy wheat: a possible mechanism for differentiation of homoeologous chromosomes. *Genetics* **147**: 1381-1387.

Felsenstein, J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**: 368-376.

Fosberg, F. R. (1948) Derivation of the flora of the Hawaiian islands. In: Insects of Hawaii. Zimmerman, E. C. University of Hawaii Press, Honolulu. **1**: 107-119.

Garbayo, J. M., Green, J. A. et al. (2000) Caprine pregnancy-associated glycoproteins (PAG): their cloning, expression, and evolutionary relationship to other PAG. *Mol. Reprod. Dev.* **57**: 311-322.

Garrigan, D. and Hendrick, P. W. (2001) Class I MHC polymorphism and evolution in endangered California Chinook and other Pacific salmon. *Immunogenetics* **53**: 483-489.

Givnish, T. J. (1997) Adaptive radiation and molecular systematics: issues and approaches. In: Molecular evolution and adaptive radiation. Givnish, T. J. and Sytsma, K. J. (eds.) Cambridge University Press, Cambridge, England. Pp. 1-54.

Gojobori, T. and Nei, M. (1984) Concerted evolution of the immunoglobulin VH gene family. *Mol. Biol. Evol.* **1**: 195-212.

Goldman, N. and Yang, Z. (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**: 725-736.

Grant, V. (1981) Plant speciation, 2nd ed. Columbia University Press, New York.

Grantham, R. (1974) Amino acid difference formula to help explain protein evolution. *Science* **185**:862-864.

Gray, A. (1852) Quarterly meeting, August 8, 1849. *Proc. Am. Acad. Arts* **2**: 159-160.

Hartl, D. L. and Jones, E. W. (2001) Genetics: analysis of genes and genomes. 5th ed. Jones and Bartlett Publishers, Inc.

Hedin, M. C. and Maddison, W. P. (2001) Phylogenetic utility and evidence for multiple copies of elongation factor-1alpha in the spider genus Habronattus (Araneae: Salticidae). *Mol. Biol. Evol.* **18**: 1512-1521.

Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hoang, A., Hill, C. E., Beerli, P. and Kingsolver, J. G. (2001) Strength and tempo of directional selection in the wild. *Proc. Natl. Acad. Sci. USA* **98**: 9157-9160.

Hoffman, M. H. (2002) Biogeography of *Arabidopsis thaliana* (L.) Heynh (Brassicaceae). *J. Biogeography* **29**: 125-134.

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S. and Vandepol, S. (1982) Rapid evolution of RNA genomes. *Science* **215**: 1577-1585.

Huang, N., Stebbins, G. L. and Rodriguez, R. L. (1992) Classification and Evolution of *α*-Amylase Genes in Plants *Proc. Natl. Acad. Sci. USA* **89**:7526-7530.

Hughes, A. L. (1991) Circumsporozoite protein genes of malaria parasites (*Plasmodium* ssp.): evidence for positive selection on immunogenic regions. *Genetics* **127**: 345-353.

Hughes, A. L. (1992) Positive selection and interallelic recombination at the *merozoite surface antigen-1* (*MSA-1*) locus of *Plasmodium falciparum*. *Mol. Biol. Evol.* **9**: 381-393.

Hughes, A. L. and Nei, M. (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**: 167-170.

Hughes, A. L. and Yeager, M. (1997) Coordinated amino acid changes in the evolution of mammalian *defensins*. *J. Mol. Evol.* **44**: 675-682.

Hughes, A. L., Ota, R. and Nei, M. (1990) Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. *Mol. Biol. Evol.* **7**: 515-524.

Jukes, T. H. and Cantor, C. R. (1969) Evolution of protein molecules. Pp. 21-132. In: Mammalian protein metabolism. Vol. 3. Munro, H. N. (ed.) Academic Press, New York.

Keck, D. D. (1936) The Hawaiian silverswords: Systematics, affinities, and phytogeographic problems of the genus *Argyroxiphium*. *Occas. Pap. Bishop Mus.* **11**: 1-38.

Kim, K. J. and Jansen, R. K. (1995) *ndhF* sequence evolution and the major clades in the sunflower family. *Proc. Natl. Acad. Sci. USA* **92**: 10379-10383.

Kimura, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**: 111-120.

King, M. C. and Wilson, A. C. (1975) Evolution at 2 levels in humans and chimpanzees. *Science* **188**: 107-116.

Knapp, R. (1975) Vegetation of the Hawaiian Islands. *Hawaiian Bot. Soc. Newsletter* **14**: 95-118.

Koch, M., Bishop, J. and Mitchell-Olds, T. (1999) Molecular systematics and evolution of *Arabidopsis* and *Arabis*. *Plant Biol.* **1**: 529-537.

Koch, M. A., Haubold, B. and Mitchell-Olds, T. (2000) Comparative evolutionary analysis of *chalcone synthase* and *alcohol dehydrogenase* loci in *Arabidopsis, Arabis,* and related genera (Brassicaceae). *Mol. Biol. Evol.* **17**: 1483-1398.

Kreitman, M. (2000) Methods to detect selection in populations with applications to the human. *Annu. Rev. Genom. Hum. G.* **1**: 539-559.

Kyhos, D. W., Carr, G. D. and Baldwin, B. G. (1990) Biodiversity and cytogenetics of the tarweeds (Asteraceae). *Ann. Mo. Bot. Gard.* **77**: 84-95.

Levin, D. A. (1983) Polyploidy and novelty in flowering plants. *Am. Nat.* **122**: 1-25.

Li, W. –H. (1985) In: Population Genetics and Molecular Evolution. Ohta, T. and Aoki, K. (eds.) Springer, Berlin.

Li, W. –H., Wu, C. –I. and Luo, C. –C. (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* **2**: 150-174.

Liu, B., Vega, J. M. and Feldman, M. (1998) Rapid genomic changes in newly synthesized amphiploids of *Trilicum* and *Aegilops*. II. Changes in low-copy coding DNA sequences. *Genome* **41**: 535-542.

Liu, B., Vega, J. M., Segal, G., Abbo, S., Rodova, M. and Feldman, M. (1998) Rapid genomic changes in newly synthesized amphiploids of *Trilicum* and *Aegilops*. I. Changes in low-copy non-coding DNA sequences. *Genome* **41**: 272-277.

Macdonald, G. A., Abbott, A. T. and Peterson, F. L. (1983) Volcanoes in the Sea. Ed. 2 University of Hawaii Press, Honolulu.

Majewski, J. and Cohan, F. M. (1999) Adapt globally, act locally: the effect of selective sweeps on bacterial sequence diversity. *Genetics* **152**: 1459-1474.

Martin, W., Gierl, A. and Saedler, H. (1989) Molecular evidence for pre-Cretaceous angiosperm origins. *Nature* **339**: 46-48.

Mayr, E. (1942) Systematics and the origin of species. Columbia University Press, New York.

McDonald, J. H. and Kreitman, M. (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila.* *Nature* **351**: 652-654.

Messier, W. and Stewart, C. B. (1997) Episodic adaptive evolution of primate lysozymes. *Nature* **385**: 151-154.

Meyerowitz, E. M. (1987) *Arabidopsis thaliana. Ann. Rev. Genet.* **21**: 93-111.

Mitchell-Olds, T. (2001) *Arabidopsis thaliana* and its wild relatives: a model system for ecology and evolution. *Trends Ecol. Evol.* **16**: 693-700.

Miyashita, N. T., Kawabe, A. and Innan, H. (1999) DNA variation in the wild plant *Arabidopsis thaliana* revealed by amplified fragment length polymorphism analysis. *Genetics* **152**: 1723-1731.

Muse, S. V. and Gaut, B. S. (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol. Biol. Evol.* **11**: 715-724.

Nei, M. and Gojobori, T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**: 418-426.

O'Kane, S. L. and Al-Shebaz, I. A. (1997) A synopsis of *Arabidopsis* (Brassicaceae). *Novon* **7**: 323-327.

Ohta, T. (1991a) Multigene families and the evolution of complexity. *J. Mol. Evol.***33**: 34-41.

Ohta, T. (1991b) Role of diversifying selection and gene conversion in evolution of major histocompatibility complex loci. *Proc. Natl. Acad. Sci. USA* **88**: 6716-6720.

Ohta, T. (1995) Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *J. Mol. Evol.* **40**: 56-63.

Ohta, T. (2000) Mechanisms of molecular evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **355**: 1623-1626.

Olsen, K. M., Womack, A., Garrett, A. R., Suddith, J. I. and Purugganan, M. D. (2002) Contrasting evolutionary forces in the *Arabidopsis thaliana* floral developmental pathway. *Genetics* **160**: 1641-1650.

Pichersky, E., Subramaniam, R., White, M. J., Reid, J., Aebersold, R. and Green, B. R. (1991) *Mol. Gen. Genet.* **227**: 277-284.

Purugganan, M. D. (1998) The molecular evolution of development. *BioEssays* **20**: 700-711.

Purugganan, M. D. (2000) The molecular population genetics of regulatory genes. *Mol. Evol.* **9**: 1451-1461.

Purugganan, M. D. and Suddith, J. I. (1998) Molecular population genetics of the *Arabidopsis CAULIFLOWER* regulatory gene: Non-neutral evolution and wild variation in floral homeotic function. *Proc. Natl. Acad. Sci. USA* **95**: 8130-8134.

Purugganan, M. D. and Suddith, J. I. (1999) Molecular population genetics of floral homeotic loci: Departures from the equilibrium-neutral model at the *AP3* and *PI* genes of *Arabidopsis thaliana. Genetics* **151**: 839-848.

Raven, P. H., Every, R. F. and Eichhorn, S. E. (1992) Biology of plants. 5th ed. Worth, New York.

Redei, G. P. (1992) A heuristic glance at the past of *Arabidopsis* genetics. In: Methods in *Arabidopsis* research. Koncz, C., Chua, N. –H. and Schell, J. (eds.) World Scientific Publishing Co. Pte. Ltd. Pp. 1-15.

Riechmann, J. L. and Meyerowitz, E. M. (1997) MADS domain proteins in plant development. *Biol. Chem.* **378**: 1079-1101.

Rieseberg, L. H. and Swensen, S. M. (1995) Conservation genetics of endangered island plants in conservation genetics: case histories from nature. Avise, J. C. and Hamrick, J. L. (eds.) Chapman and Hall, New York. Pp. 304-334.

Robichaux, R. H. (1985) Tissue elastic properties of a mesic forest Hawaiian *Dubautia* species with 13 pairs of chromosomes. *Pac. Sci.* **39**: 191-194.

Robichaux, R. H. and Canfield, J. E. (1985) Tissue elastic properties of 8 Hawaiian *Dubautia* species that differ in habitat and diploid chromosome number. *Oecologia* **66**: 77-80.

Robichaux, R. H., Carr, G. D., Liebman, M. and Pearcy, R. W. (1990) Adaptive radiation of the Hawaiian silversword alliance (Compositae-Madiinae)—ecological, morphological and physiological diversity. *Ann. Mo. Bot. Gard.* **77**: 64-72.

Savolainen, O., Langley, C. H., Lazzaro, B. P. and Fr, H. (2000) Contrasting patterns of nucleotide polymorphism at the *alcohol dehydrogenase locus* in the outcrossing *Arabidopsis lyrata* and the selfing *Arabidopsis thaliana. Mol. Biol. Evol.* **17**: 645-655.

Scherf, A., Hilbich, C., Sieg, K., Mattei, D., Mercereaupuijalon, O. and Mullerhill, B. (1988) The 11-1 gene of *Plasmodium falciparum* codes for distinct fast evolving repeats. *EMBO J.* **7**: 1129-1137.

Schmid, K. J. and Tautz, D. (1997) A screen for fast evolving genes from *Drosophila*. *Proc. Natl. Acad. Sci. USA* **94**: 9746-9750.

Schug, M. D., Hutter, C. M., Noor, M. A. F. and Aquadro, C. F. (1998) Mutation and evolution of microsatellites in *Drosophila melanogaster*. *Genetica* **103**: 359-367.

Shallenberger, R. J. (ed.) (1981) *Hawaii's Birds.* Hawaiian Audubon Soc., Honolulu.

Sharp, P. M. (1997) In search of molecular Darwinism. *Nature* **385**: 111-112.

Simpson, G. G. (1953) The major features of evolution. New York, NY: Columbia University Press.

Skelton, P. W. (1993) Adaptive radiation: definition and diagnostic tests. Pp. 45-58. In: Evolutionary patterns and processes. Lees, D. R. and Edwards, D. (eds.) London, England: Academic Press.

Small, R. L., Ryburn, J. A. and Wendel, J. F. (1999) Low levels of nucleotide diversity at homoeologous *Adh* loci in allotetraploid cotton (*Gossypium* L.). *Mol. Biol. Evol.* **16**: 491-501.

Smathers, G. A. and Mueller-Dombois, D. (1974) Invasion and recovery of vegetation after a volcanic eruption in Hawaii. *National Park Service Scientific Monograph Series* No. 5.

Soltis, D. E. and Soltis, P. S. (1995) The dynamic nature of polyploidy genomes. *Proc. Natl. Acad. Sci. USA* **92**: 8089-8091.

Song, K. M., Lu, P., Tang, K. L. and Osborn, T. C. (1995) Rapid genomic change in synthetic polyploids of *Brassica* and its implications for polyploidy evolution. *Proc. Natl. Acad. Sci. USA* **92**: 7719-7723.

Stebbins, G. L. (1971) Chromosome evolution in higher plants. Arnold, London.

Stuessy, T. F. (1977) Heliantheae – systematic review. In: The biology and chemistry of the Compositae. Heywood, V. H., Harbone, J. B. and Turner, B. L. (eds.) Academic Press, London. **2**:621-671.

Sturmbauer, C. and Meyer, A. (1992) Genetic divergence, speciation and morphological stasis in a lineage of African cichlid fishes. *Nature* **358**: 578-581.

Swanson, W. J., Clark, A. G., Waldrip-Dail, H. M., Wolfner, M. F. and Aquadro, C. F. (2001) Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **98**: 7375-7379.

Tanaka, T. and Nei, M. (1989) Positive Darwinian selection observed at the variable-region genes of immunoglobulins. *Mol. Biol. Evol.* **6**: 447-459.

Verra, F. and Hughes, A. L. (1999) Natural selection on apical membrane antigen-1 of *Plasmodium falciparum. Parassitologia* **41**: 93-95.

Weber, J. L. and Wong, C. (1993) Mutation of human short tandem repeats. *Hum. Mol. Genet.* **2**: 1123-1128.

Weigel, D. (1998) From floral induction to floral shape. *Curr. Opin. Plant Biol.* **1**: 55-59.

Wendel, J. F. (2000) Genome evolution in polyploids. *Plant Mol. Biol.* **42**: 225-249.

Witter, M.S. (1986) Thesis. University of Hawaii, Honolulu.

Witter, M. S. and Carr, G. D. (1988) Adaptive radiation and genetic differentiation in the Hawaiian silversword alliance (Compositae: Madiinae). *Evolution* **42**: 1278-1287.

CHAPTER 2.

Interspecific Hybrid Ancestry of a Plant Adaptive Radiation: Allopolyploidy of the Hawaiian Silversword Alliance (Asteraceae) Inferred From Floral Homeotic Gene Duplications

Marianne Barrier, Bruce G. Baldwin, Robert H. Robichaux

and Michael D. Purugganan

Molecular Biology and Evolution 16 (8): 1105-1113 (1999).

Note: Species name changes since publication are listed here old name -> new name. *Madia/Raillardiopsis* group -> "*Madia*" group; *Madia bolanderi -> Kyhosia bolanderi*; *Madia nutans -> Harmonia nutans*; *Raillardiopsis muirii -> Carlquistia muirii*; *Raillardiopsis scabrida -> Anisocarpus scabridus*

Abstract

The polyploid Hawaiian silversword alliance (Asteraceae), a spectacular example of adaptive radiation in plants, was shown previously to have descended from North American tarweeds of the *Madia/Raillardiopsis* group, a primarily diploid assemblage. The origin of the polyploid condition in the silversword alliance was not resolved in earlier biosystematic, cytogenetic, and molecular studies, apart from the determination that polyploidy in modern species of *Madia/Raillardiopsis* arose independent of that of the Hawaiian group. We determined that two floral homeotic genes, *ASAP3/TM6* and *ASAP1*, are found in duplicate copies within members of the Hawaiian silversword alliance and appear to have arisen as a result of interspecific hybridization between two North American tarweed species. Our molecular phylogenetic analyses of the *ASAP3/TM6* loci suggest that the interspecific hybridization event in the ancestry of the Hawaiian silversword alliance involved members of lineages that include *Raillardiopsis muirii* (and perhaps *Madia nutans*) and *Raillardiopsis scabrida*. The *ASAP1* analysis also indicates that the two species of *Raillardiopsis* are among the closest North American relatives of the Hawaiian silversword alliance. Previous biosystematic evidence demonstrates the potential for allopolyploid formation between members of the two North American tarweed lineages; a vigorous hybrid between *R. muirii* and *R. scabrida* has been produced that formed viable, mostly tetraporate (diploid) pollen, in keeping with observed meiotic

failure. Various genetic consequences of allopolyploidy may help to explain the phenomenal evolutionary diversification of the silversword alliance.

Introduction

Adaptive radiations are among the most spectacular processes in organismal evolution. Radiations are characterized by rapid bursts of evolutionary innovation and are associated with increased speciation rates, elevated levels of morphological diversity, and marked differentiation in ecological characteristics (Givnish 1997). Some of the most dramatic examples of adaptive radiations are found in the Hawaiian archipelago, where isolation, continued island formation, and a high diversity of environmental settings promote species radiations in both plant and animal groups.

The Hawaiian silversword alliance (Asteraceae: Heliantheae–Madiinae) is a premier example of adaptive radiation in plants (Carr 1985; Robichaux et al. 1990; Baldwin and Robichaux 1995; Baldwin 1997). The alliance comprises 30 perennial species in three endemic genera: *Argyroxiphium*, *Dubautia*, and *Wilkesia* (Carr 1985 1998a, 1998b). The species are distributed on six of the eight main islands of the Hawaiian archipelago (Kaua'i, O'ahu, Moloka'i, Lana'i, Maui, and Hawaii), with all but five species being single-island endemics (Carr 1985). Species in the alliance grow in a wide range of habitats, including exposed lava, dry scrub, mesic forests, wet forests, and bogs (Baldwin and Robichaux 1995; Baldwin 1997). The species also display an impressive array of morphological growth forms; including rosette plants, cushion plants, subshrubs, shrubs, trees, and lianas (Baldwin and Robichaux 1995; Baldwin 1997). Furthermore, the silversword alliance is characterized by a high level

of chromosomal repatterning, with eight genomic rearrangements distinguished by reciprocal translocations and an aneuploid reduction (Carr and Kyhos 1986).

Understanding the origin of species-rich insular groups is crucial to our attempts to reconstruct evolutionary patterns of ecological and morphological change that characterize adaptive radiations (Givnish 1997). Both cpDNA and rDNA ITS studies have confirmed Carlquist's (1959) hypothesis that the closest relatives of the Hawaiian silversword alliance can be found among the North American tarweeds (Asteraceae: Heliantheae–Madiinae) in the paraphyletic *Madia/Raillardiopsis* group (Baldwin et al. 1991; Baldwin 1992, 1996, 1997; Baldwin and Robichaux 1995). A major unanswered question concerns evolution of the polyploid condition found throughout the Hawaiian silversword alliance (Baldwin et al. 1991; Baldwin and Kyhos 1996; Baldwin 1997). Both cytogenetic (Carr and Kyhos 1986; Kyhos, Carr and Baldwin 1990; Carr, Baldwin and Kyhos 1996) and allozymic (Witter and Carr 1988) data indicate that the Hawaiian species are tetraploids (n = 13, 14), in contrast to the basally diploid condition (n = 6–9) in each of the most closely related North American lineages of *Madia/Raillardiopsis* (Baldwin 1996). Resolution of the origin of polyploidy in the silversword alliance is important for understanding the genomic constitution of the founder species and its descendants, information that could help to identify factors that facilitated this spectacular adaptive radiation.

We report here the isolation of orthologs of two *Arabidopsis* floral homeotic genes, *APETALA1* and *APETALA3/TM6*, from members of the Hawaiian silversword alliance and the North American tarweeds. PCR-based sampling indicates that these

floral regulatory genes are each present in two copies in the Hawaiian species, but in only one copy in the North American tarweeds examined. Our results suggest that the Hawaiian silversword alliance is an allopolyploid group that descended from a hybrid between members of extant tarweed lineages.

Materials and Methods

Sampling and nucleic acid isolation

Genomic DNA samples were isolated from young leaf tissue collected from 10 species of the Hawaiian silversword alliance and 7 species of North American tarweeds. The Hawaiian species were selected to represent each of the four major lineages in the silversword alliance as previously identified from rDNA ITS trees (Baldwin and Robichaux 1995; Baldwin 1996). Four North American species (*Madia bolanderi*, *Madia nutans*, *Raillardiopsis scabrida*, and *Raillardiopsis muirii*) were chosen to represent each of the four major lineages in the *Madia/Raillardiopsis* group as also identified from rDNA ITS trees (Baldwin 1996). Three other North American tarweed species (*Adenothamnus validus*, *Raillardella pringlei*, and *Osmadenia tenella*) are known to fall outside the clade comprising *Madia/Raillardiopsis* and the silversword alliance (Baldwin 1996) and were included to serve as the outgroup in the analyses. Genomic DNA was isolated from leaf tissue of *Argyroxiphium sandwicense* ssp. *macrocephalum* using a modified CTAB protocol that reduced the amount of pectin and secondary-product contamination (Friar, Robichaux and Mount 1996). For all of the North American species and the remaining Hawaiian species, genomic DNA was isolated using the methods of Palmer (1986) or Doyle and Doyle (1987) and further purified on $CsCl_2$ gradients. Total floral RNA was isolated from immature and mature capitula of *A. sandwicense* ssp. *macrocephalum* using a phenol/SDS extraction procedure (Ausubel et al. 1992).

Isolation of the *Argyroxiphium sandwicense AP3/TM6* and *AP1* genes

MADS-box loci can be readily identified using sets of degenerate PCR primers that specifically amplify MADS-box sequences (Rounsley, Ditta and Yanofsky 1995). The isolation of MADS-box sequences from *A. sandwicense* ssp. *macrocephalum* by this method allowed us to obtain sequence information that permitted the design of 5' nonoverlapping, nested PCR primers specific to the *ASAP3/TM6* and *ASAP1* genes. These gene-specific primers were used to isolate cDNAs utilizing the PCR-based rapid amplification of cDNA ends (RACE) protocol (Frohman, Dush and Martin 1988). First-strand cDNA was synthesized from *A. sandwicense* ssp. *macrocephalum* floral RNA with AMV reverse transcriptase following standard protocols (Boehringer Mannheim), and amplified cDNA was cloned using the TA procedure (Invitrogen) and sequenced.

PCR primers to amplify *ASAP3/TM6* and *ASAP1* genomic regions were designed based on the cDNA sequences. The primers ASAP3-2 and ASAP3-3R allowed PCR amplification of a region spanning exons 1–4 of the *ASAP3/TM6* gene. Primers were designed to isolate both size-differentiated copies of this gene (1.1 kb vs. 1.4 kb) in the Hawaiian silversword alliance species. For *ASAP1*, primers ASAP1/3X and ASAP1/8XR were designed to isolate both *A* and *B* copies of this gene; the length difference between the duplicate copies was too small, however, for reliable size differentiation. To isolate the two copies of *ASAP1*, gene-specific primer pairs (ASAP1-F1A/ASAP1-RB and ASAP1-AF/ASAP1-R) were constructed to amplify sequences from exons 3–8 of different duplicate copies of the gene within the

Hawaiian species. The primers were used in PCR amplifications using the error-correcting rTth polymerase formulation (Perkin-Elmer) in standard buffer with 40 cycles of 95°C for 1 min, 55°C for 1 min, and 72°C for 4 min. The nucleotide error rate for this formulation is less than 1 bp in 7 kb of sequence (unpublished observations). In experiments designed to determine whether other copies of these loci are present, an annealing temperature 5–10°C below primer melting temperature was used. PCR-amplified DNA was cloned using the TA cloning kit (Invitrogen) and sequenced using automated sequencers (Iowa State University Sequencing Facility, NCSU DNA Sequencing Facility). Sequencing was done with nested primers, with multiple sequencing reactions conducted for divergent sequences. All sequence changes were rechecked visually against sequencing chromatograms and are deposited in GenBank (accession numbers AF147210–AF147258).

Phylogenetic analyses

Nucleotide sequences were aligned visually. Phylogenetic analyses were conducted using maximum-parsimony techniques implemented in PAUP*4.0d54 (Swofford 1998). Both substitution and insertion/deletion (indel) differences were used and weighted equally in the analyses, with the indels separately coded (as additional characters) to reflect nonindependence of continuous gaps. Parsimony analyses were conducted using the heuristic search procedure, with random taxon addition (10 replicates), tree bisection-reconnection branch swapping, and MULPARS in effect. Clade support was estimated by parsimony analysis of 500 bootstrap replicates of the data set using the search procedures outlined above.

Interspecific hybridizations

In addition to a hybrid combination previously reported between *R. muirii* and *R. scabrida* (Baldwin 1989; Kyhos, Carr and Baldwin 1990), hybrids were produced in the combinations *Madia madioides* X *R. scabrida* and *Dubautia knudsenii* X (*M. madioides* X *R. scabrida*) (unpublished data). Cross-pollinations were performed by rubbing the stylar surfaces of pollen-shedding capitula together. Full cypselae from crosses were hydrated overnight on wet filter paper, and F1 embryos were surgically excised and germinated on wet filter paper. The F1 plants were grown to flowering maturity under cool, low-humidity greenhouse conditions at the University of California–Berkeley. Pollen fertility and pollen morphology of hybrid plants were examined by staining with cotton blue in lactophenol. Chromosomal association at

meiotic metaphase I was examined, using phase microscopy, in squashed microsporocytes stained with acetocarmine mixed with Hoyer's solution.

Results and Discussion

Floral homeotic genes as sources of characters for plant phylogeny reconstruction

In recent years, the molecular genetics of floral and inflorescence development have been the subject of intense study (Riechmann and Meyerowitz 1997). Much of this work has been carried out on *Arabidopsis thaliana* (Brassicaceae), in which several genes that control flower development have been identified (Riechmann and Meyerowitz 1997). The loci are referred to as floral homeotic genes, with mutational lesions in the genes resulting in the formation of aberrant organ types in flowers.

Molecular studies have revealed that most of the floral homeotic genes isolated to date belong to the MADS-box regulatory gene family of transcriptional activators (Ma, Yanofsky and Meyerowitz 1991; Purugganan et al. 1995; Riechmann and Meyerowitz 1997). In *A. thaliana*, at least 25 members of this gene family have been isolated, and most of these appear to regulate differing aspects of flower development (Purugganan et al. 1995; Riechmann and Meyerowitz 1997). Many of the duplications that resulted in the growth of this gene family appear to have occurred fairly early in the evolutionary history of the vascular plants (Purugganan 1997), and the different paralogous genes are readily distinguishable. Some members of the gene family, however, appear to have arisen from more recent duplications, possibly as a result of polyploidization events that occurred during evolution (Mena et al. 1996; Lowman and Purugganan 1999). The MADS-box floral homeotic genes are among the fastest-

evolving plant nuclear loci observed thus far, with variable regions that are evolving at 3.7 x 10^{-9} nonsynonymous substitutions per site per year (Purugganan et al. 1995). This rate is comparable to the mean rate for synonymous substitutions for other plant nuclear genes (4.7 x 10^{-9} substitutions per site per year) (Purugganan et al. 1995). The rapid evolution of the MADS-box floral homeotic genes, including their introns, makes them attractive as potential sources of new molecular characters for plant phylogenetic investigations at lower taxonomic levels.

AP3/TM6 and *AP1* orthologs in the Hawaiian silversword alliance

Homologs of the *Arabidopsis* floral homeotic genes *AP3* and *AP1* were successfully isolated from developing flowers of *A. sandwicense* ssp. *macrocephalum* utilizing RT-PCR techniques. The isolated *AP3* homolog cDNA is 837 bp long and contains a long open reading frame encoding a putative protein of 226 amino acids (aa) that encodes a MADS domain (see Figure 1). The protein shows strong similarity to the *Arabidopsis AP3* and *Antirrhinum DEFICIENS* proteins, with overall similarities of 84% and 88%, respectively, at the peptide level. Phylogenetic analysis indicates that the isolated gene is a member of the *AP3* group of genes (Purugganan et al. 1995; Purugganan 1997) and appears to be an ortholog of the tomato *TM6* locus (Kramer, Dorit and Irish 1998). We refer to the gene as *ASAP3/TM6*.

The isolated *A. sandwicense* ssp. *macrocephalum AP1* (or *ASAP1*) cDNA is 839 bp long, with an open reading frame encoding a putative protein of 231 aa and possessing a 57-aa MADS domain (see Figure 1). The encoded protein shows strong

similarity (80%–85%) to the *Arabidopsis AP1* and *Antirrhinum SQUAMOSA* proteins, and phylogenetic analysis confirms that *ASAP1* is the ortholog of the *Arabidopsis AP1* locus (results not shown).

A

	1	10		20		30
ASAP3/TM6-A	**MGRGR**	**VETRK**	**IENNT**	**NRQVT**	**YSKRR**	**NGIFK**
AP3	**.A..K**	**IQIKR**	**...Q.**	**.....**	**.....**	**..L..**
DEF	**.A..K**	**IQIKR**	**...Q.**	**.....**	**.....**	**..L..**
TM6	**???.K**	**I.IK.**	**...S.**	**.....**	**.....**	**.....**

	31	40		50		57
ASAP3/TM6-A	**KAHEL**	**TVLCD**	**AKVPL**	**IMFSN**	**TGKFH**	**EY**
AP3	**.....**	**.....**	**.R.SI**	**....S**	**SN.L.**	**..**
DEF	**.....**	**S....**	**...SI**	**..I.S**	**.Q.L.**	**..**
TM6	**.RK..**	**.....**	**..IS.**	**..L.S**	**.R.Y.**	**..**

B

	1	10		20		30
ASAP1	**MGRGK**	**VQLRR**	**IENKI**	**NRQVT**	**FSKRR**	**GGLLK**
AP1	**....R**	**...K.**	**.....**	**.....**	**.....**	**A....**
SQUA	**.....**	**...K.**	**.....**	**.....**	**.....**	**.....**

	31	40		50		57
ASAP1	**KAHEI**	**SVLCD**	**AEVAL**	**IVSSS**	**KGKLF**	**EF**
AP1	**.....**	**.....**	**.....**	**V.F.H**	**.....**	**.Y**
SQUA	**....L**	**.....**	**.....**	**..F.N**	**.....**	**.Y**

Figure 1. MADS-box sequences of isolated floral homeotic gene orthologs. The amino acid sequences inferred from the *Argyroxiphium sandwicense* ssp. *macrocephalum (A) ASAP3/TM6-A* and *(B) ASAP1-A* cDNA sequences are shown and compared with homologs in *Arabidopsis thaliana* (*AP3*, *AP1*), *Antirrhinum majus* (*DEF*, *SQUA*), and *Lycopersicon esculentum* (*TM6*).

Duplicate copies of the floral homeotic genes in the Hawaiian silversword alliance

Using primers designed from cDNA sequences, *ASAP3/TM6* genomic sequences spanning exons 1–4 were amplified by PCR. Two bands of 1.1 and 1.4 kb were obtained from species in the Hawaiian silversword alliance, suggesting that *ASAP3/TM6* is present in two distinct copies in each species. In contrast, only one PCR band was amplified from each North American tarweed species examined. The 1.1-kb fragment from the Hawaiian species is designated *ASAP3/TM6-A*, and the larger 1.4-kb product is *ASAP3/TM6-B*. We successfully isolated both gene copies of *ASAP3/TM6* from nine Hawaiian silversword alliance species, but only *ASAP3/TM6-A* from *Dubautia plantaginea*. The two gene copies from *A. sandwicense* ssp. *macrocephalum* are minimally divergent from one another (93% similarity). The introns of *ASAP3/TM6-A* and *ASAP3/TM6-B* have diverged by 20 indels ranging in size from 1 to 369 bp (see Figure 2). In contrast, orthologous sequences within the *ASAP3/TM6-A* and *ASAP3/TM6-B* gene groups in the Hawaiian taxa are 99% and 97%–98% similar, respectively.

The *ASAP1* gene was also amplified using primers designed from cDNA sequences and (like *ASAP3/TM6*) was present in duplicate copies in the genomes of the Hawaiian species. PCR amplification of exons 3–8 of *ASAP1* yielded two copies that were both approximately 1.8 kb in length. In contrast, PCR amplification produced only one gene copy in each of the North American species examined. The duplicate copies of the gene in the Hawaiian species, which differ only slightly in size, are designated *ASAP1-A* and *ASAP1-B*. We successfully isolated both copies from

nearly all Hawaiian silversword alliance species, with the exception of the *B* copy from *Dubautia laevigata* and *Dubautia latifolia*. We also isolated the *ASAP1* loci from a representative of each of the four major lineages within the North American *Madia/Raillardiopsis* group, as well as from *O. tenella*. As with *ASAP3/TM6-A* and *ASAP3/TM6-B*, the two gene copies of *ASAP1* from *A. sandwicense* ssp. *macrocephalum* show 93% similarity. The two gene copies are differentiated by 30 indels that range in size from 1 to 61 bp (see Figure 2). Orthologous sequences within the *ASAP1-A* and *ASAP1-B* gene groups in the Hawaiian taxa were 98%–99% and 95%–96% similar, respectively.

The presence of two distinct *ASAP3/TM6* and *ASAP1* genes in the Hawaiian species is not unexpected—as mentioned above, both cytogenetic (Carr and Kyhos 1986; Kyhos, Carr and Baldwin 1990; Carr, Baldwin and Kyhos 1996) and allozyme (Witter and Carr 1988) studies indicate that the Hawaiian species are tetraploids (n = 13, 14), in contrast to the basally diploid condition of each of the most closely related North American tarweed lineages (n = 6–9) (Baldwin 1996). We utilized a permissive PCR-based approach to examine whether the members of the North American *Madia/Raillardiopsis* group possessed only a single gene copy of these two floral homeotic loci. A PCR-based approach was necessitated by the difficulties in conducting DNA blot analyses with these species, which appear to have both large genomes and high methylation levels (unpublished data).

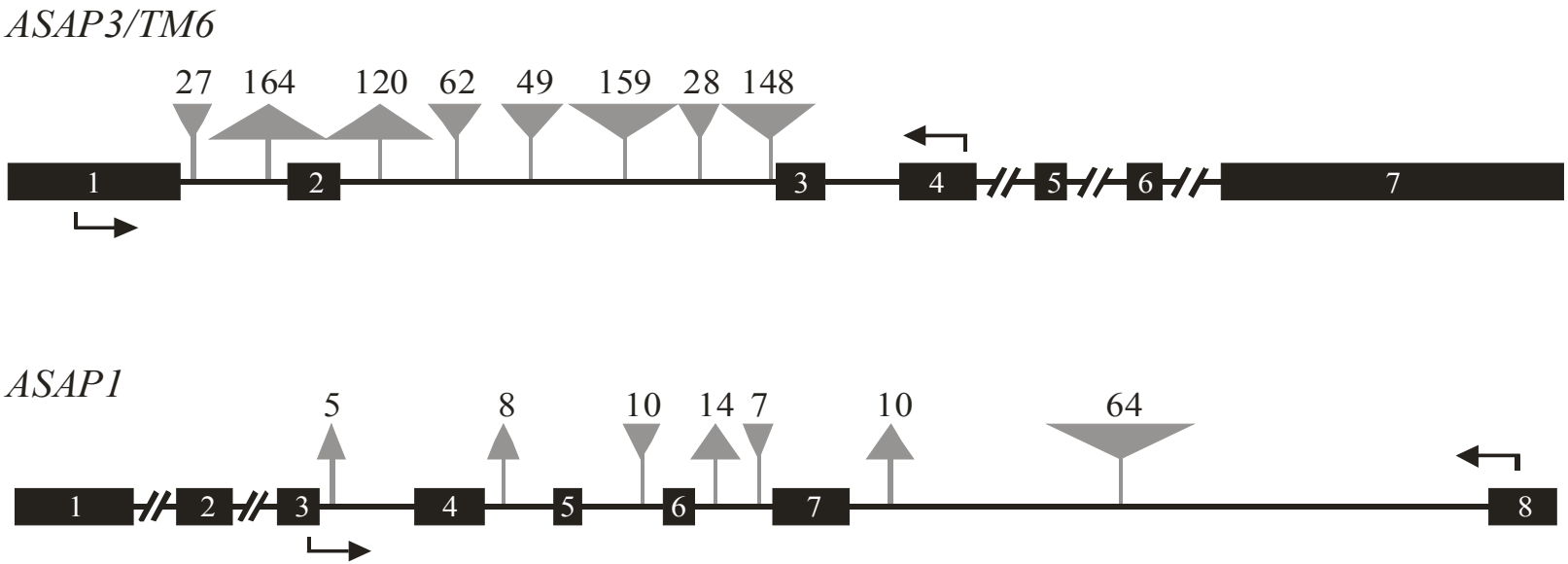

Figure 2. Gene maps of *(A) ASAP3/TM6* and *(B) ASAP1* loci. Exons are shown as numbered boxes. The gene maps depicted are for the *A* copies of the loci. Upright and inverted triangles represent major deletions and insertions, respectively, that characterize the *B* duplicate copies. Numbers above the triangles indicate the sizes of the indels. Arrows show positions of PCR primers used to isolate genomic sequences. A 100-bp scale bar is provided. Relative sizes of exons and introns in the amplified regions were derived from comparison of genomic and cDNA sequences. Exon sizes outside the amplified regions are estimates based on comparison with data from *Arabidopsis thaliana* orthologs.

For permissive PCR, primers were designed based on the conserved regions of both *A* and *B* copies of each gene and used to amplify sequences in the *Madia/Raillardiopsis* species using permissive annealing temperatures (below primer melting temperature [5–10°C]). The design of these primers around conserved regions ensures the amplification of both gene copies if they are present in the North American taxa. This permissive PCR approach resulted in the detection of nine new sequences of >500 bp. Analyses of these novel amplicons against GenBank identified two of these sequences: (1) a *Tyl/copia*-like retrotransposon and (2) a plant *extensin*-like gene. The other seven sequences could not be identified from database searches. No duplicates of the *ASAP1* or *ASAP3/TM6* genes were detected in the *Madia* and *Raillardiopsis* genomes using this permissive PCR approach. Together with the previous cytogenetic and allozyme studies, these results suggest that these floral homeotic loci are present in only single copies in the immediate North American relatives of the Hawaiian silversword alliance.

ASAP3/TM6 gene tree

Figure 3A depicts a gene phylogeny for the *ASAP3/TM6* loci obtained from maximum-parsimony analysis. The tree demonstrates that the *ASAP3/TM6-A* and *ASAP3/TM6-B* genes from the Hawaiian silversword alliance each form a monophyletic group. The *ASAP3/TM6-A* sequences group together with 83% bootstrap support, and the *ASAP3/TM6-B* copies constitute a monophyletic group with 100% support. The phylogeny of the loci also suggests that the presence of the *A* and

B copies in the Hawaiian species is the result of evolutionary reticulation, rather than duplication following divergence from the North American tarweeds; the *A* and *B* copies are not resolved as sister to one another, but rather as sister to genes from different species or species groups of North American tarweeds. The *ASAP3/TM6-A* genes appear to be closely related to the gene in *R. scabrida*, with the association showing 100% bootstrap support. In contrast, the *ASAP3/TM6-B* sequences appear most closely related to the sequences of *R. muirii* and *M. nutans*, with 89% bootstrap support. In the *ASAP3/TM6* phylogeny, the *R. muirii* and *M. nutans* genes also group together with 89% bootstrap support.

A large number of indels differentiate the *ASAP3/TM6* loci in the different species examined. A minimum of 176 indel events in the *ASAP3/TM6* introns can account for the length differences observed and scored for analysis. Many of the indels are found in only one species, but several appear to be shared by multiple taxa; e.g., the indels diagnosing the *ASAP3/TM6-A* and *ASAP3/TM6-B* gene lineages in the silversword alliance. The indels range in size from 1 bp to a large deletion of 1.2 kb restricted to *M. nutans*. Five microsatellite loci also contribute to some of the length variation.

Several indels provide support for distinct North American tarweed ancestries of the *ASAP3/TM6-A* and *ASAP3/TM6-B* genes (see Figure 3A). Two indels, a 4-bp insertion and a 3-bp deletion in intron sequences, unambiguously group the *R. scabrida* locus with the Hawaiian *ASAP3/TM6-A* sequences. Moreover, at least four

indels, ranging in size from 10 to 159 bp, are shared between the *R. muirii* gene and the Hawaiian *ASAP3/TM6-B* sequences.

Length variation in *ASAP3/TM6* exhibits homoplasy that partly complicates identification of the North American tarweed gene lineages most closely related to those of the Hawaiian silversword alliance. None of the four length variants shared between the Hawaiian *ASAP3/TM6-B* sequences and the *R. muirii* gene is found in the *M. nutans* sequence, although a separate 7- bp deletion is shared by all three lineages. Three indels can also be found that are shared by the *ASAP3/TM6-B* genes and the sequences from *R. muirii* and *M. bolanderi*.

ASAP1 gene tree

The *ASAP1* gene provides an additional set of nuclear sequences for analysis of relationships between the Hawaiian silversword alliance and the North American tarweeds. Like the *ASAP3/TM6* gene, *ASAP1* was found in duplicate copies in the Hawaiian species, yet appears to be present in only one copy in the genomes of North American members of the *Madia/Raillardiopsis* group. Each of the copies found in the Hawaiian species, *ASAP1-A* and *ASAP1-B*, forms a monophyletic group with 100% bootstrap support (see Figure 3B). Moreover, a larger clade containing the genes of *R. scabrida* and *R. muirii* with both Hawaiian *ASAP1-A* and *ASAP1-B* loci is supported by bootstrap analysis at the 87% level. Thus, the *ASAP1* tree corroborates the finding from the *ASAP3/TM6* phylogeny that the genes of *R. scabrida* and *R. muirii* are more

closely related to those of the Hawaiian species than are the genes of other sampled species of the *Madia/Raillardiopsis* group (see Figure 3B).

Unlike the *ASAP3/TM6* phylogeny, however, the *ASAP1* phylogeny does not provide strong resolution of the relationships among the Hawaiian *ASAP1-A* and *ASAP1-B* genes and the *ASAP1* genes of *R. muirii* and *R. scabrida*. Several indels provide evidence for a lineage comprising the *R. muirii* gene and *ASAP1-B* sequences; three indels ranging in size from 3 to 8 bp are shared by only the *R. muirii ASAP1* gene and the *ASAP1-B* sequences. However, a 9-bp insertion is shared by the *R. muirii* gene and both Hawaiian gene lineages, and two small deletions place both the *R. muirii* and *R. scabrida ASAP1* sequences closer to the *ASAP1-A* lineage than to the *ASAP1-B* lineage. The incomplete resolution within the lineage of Hawaiian and *Raillardiopsis* genes may be due to the effects of homoplasy accompanying the high rate of molecular evolution of the *ASAP1* gene, which appears to evolve faster than the *ASAP3/TM6* locus (unpublished data).

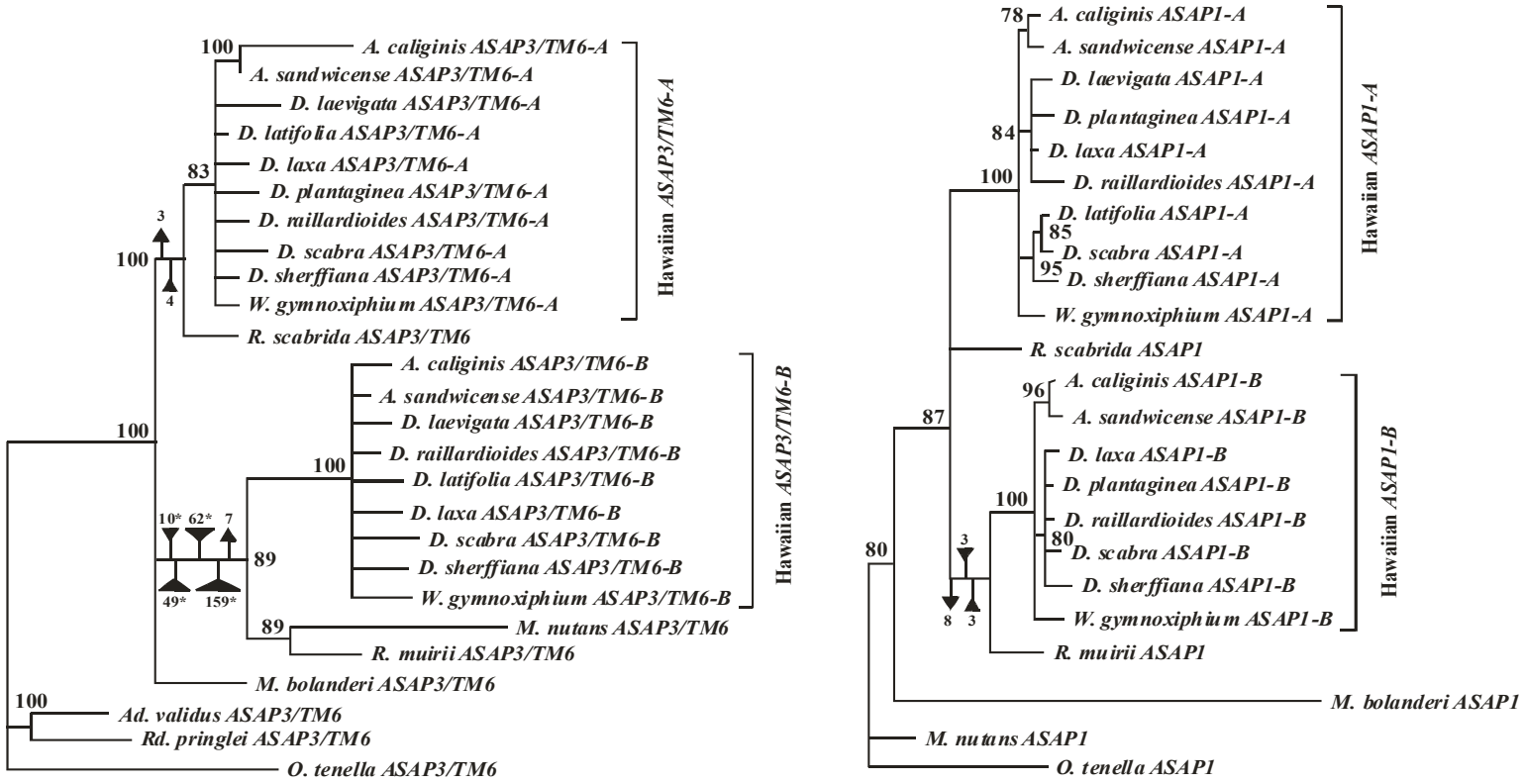

Figure 3. Phylogenetic reconstruction of *(A) ASAP3/TM6* (tree length = 343 steps, consistency index [CI] = 0.860) and *(B) ASAP1* loci (tree length = 211 steps, CI = 0.863) in the Hawaiian silversword alliance and North American tarweeds using maximum-parsimony analyses. Numbers next to the nodes give bootstrap support for branches. Indels (>1 bp) that support the grouping of *Raillardiopsis scabrida* and *Raillardiopsis muirii* with *A* and *B* gene copies are mapped onto the phylogeny. The starred indels in the *ASAP3/TM6* tree are those length variants found in both *R. muirii* and *ASAP3/TM6-B*, but not *Madia nutans*. Abbreviations: A = *Argyroxiphium*, D = *Dubautia*, W = *Wilkesia*, R = *Raillardiopsis*, M = *Madia*, Ad = *Adenothamnus*, Rd = *Raillardella*, and O = *Osmadenia*.

A hybrid (Allopolyploid) ancestry of the Hawaiian silversword alliance

The origin of the Hawaiian silversword alliance, considered ‘‘the best example of adaptive radiation in plants’’ (Raven, Evert and Eichorn 1992), was ill-defined until Carlquist (1959) suggested that the group was related to North American tarweeds. Molecular systematic investigations using both cpDNA restriction site data (Baldwin et al. 1991) and rDNA ITS sequences (Baldwin and Robichaux 1995; Baldwin 1997) have confirmed and extended Carlquist’s (1959) hypothesis by showing that the Hawaiian species are phylogenetically nested within the paraphyletic *Madia/Raillardiopsis* group and originated after considerable diversification of the North American tarweeds (Baldwin 1996; Baldwin and Sanderson 1998). Previous investigations have not, however, determined whether the species of the silversword alliance are auto- or allopolyploids (see Baldwin 1997).

Our molecular phylogenetic analyses using two nuclear-encoded floral homeotic genes provide evidence that species in the Hawaiian silversword alliance are allopolyploids that descended from a hybrid between species of two extant lineages in the *Madia/Raillardiopsis* group. Based on the *ASAP3/TM6* analysis, the interspecific hybridization event in the ancestry of the Hawaiian silversword alliance involved members of lineages that include *R. muirii* (and perhaps *M. nutans*) and *R. scabrida*. The *ASAP1* analysis also suggests that the two species of *Raillardiopsis* are among the closest North American relatives of the silversword alliance. Moreover, the earlier finding that *R. muirii* cpDNA is more closely related to the Hawaiian cpDNAs than is *R. scabrida* cpDNA (Baldwin et al. 1991) suggests that a member of the lineage

represented by *R. muirii* was the maternal parent of the hybrid ancestor of the silversword alliance (see Figure 4). The phylogenetic position of the lineage represented by *M. nutans* remains uncertain because of apparent topological conflict between the *ASAP3/TM6* and *ASAP1* trees. The *ASAP3/TM6* tree placement of *M. nutans* and *R. muirii* as sister taxa receives support, however, from the rDNA ITS phylogeny, which is topologically congruent with the *ASAP3/TM6* results (Figure 4).

Phylogenetic analyses of nuclear genes that have not undergone interlocus concerted evolution have proven crucial to our identification of a hybrid ancestry of the silversword alliance. Both cpDNA and rDNA ITS analyses, which have provided the most widely used molecular data for investigations of plant phylogeny, can sometimes prove insufficient for identifying reticulate evolution (see Figure 4). Uniparental cpDNA trees are useful for aiding resolution of hybridization events only in comparison with phylogenetic data from the nuclear genome (e.g., Baldwin 1997). Analyses of rDNA ITS sequences in the Hawaiian silversword alliance did not reveal evidence of divergent or recombinant ITS copies within species (Baldwin and Robichaux 1995; Baldwin 1997; Baldwin and Sanderson 1998), a result that may be explained by rapid, unidirectional concerted evolution following the ancient hybridization event resolved here (Hillis et al. 1991; Baldwin and Sanderson 1998). Strongly conflicting signal between cpDNA and rDNA ITS data, another possible indication of hybrid ancestry, was not found from comparisons of the two character sets using the partition homogeneity test (Farris et al. 1995), as implemented in PAUP* (Baldwin 1992; unpublished data).

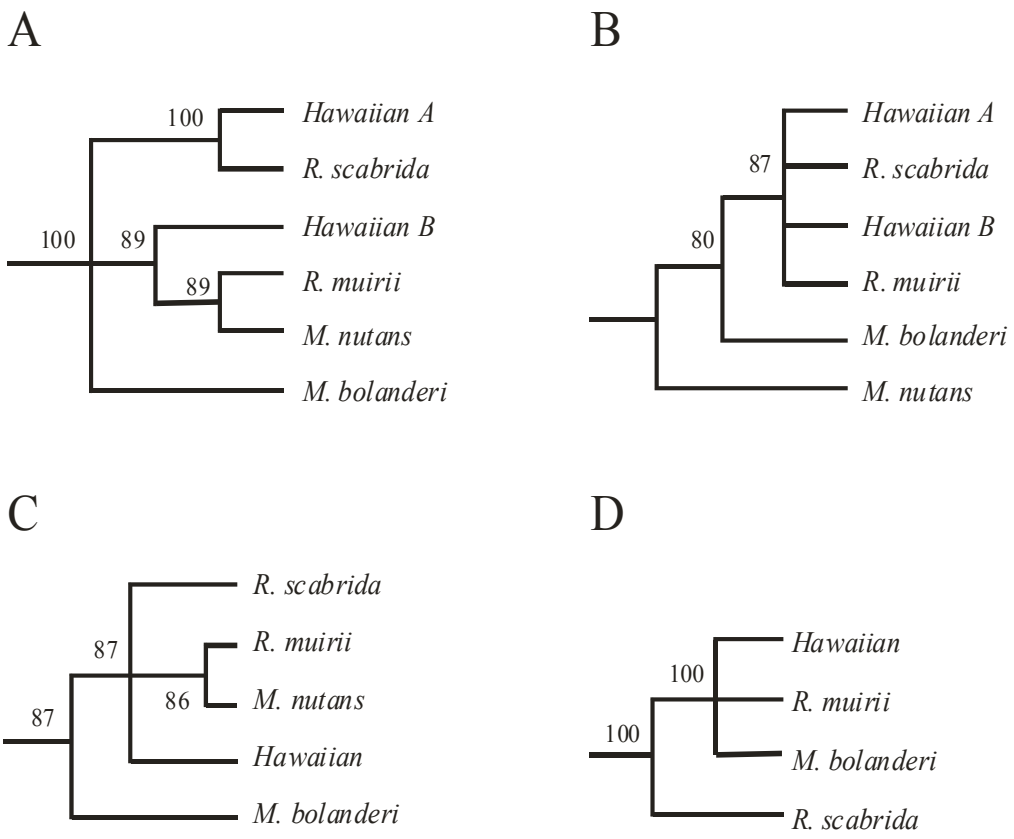

Figure 4. Comparison of nuclear and cpDNA trees of the Hawaiian silversword alliance and North American tarweeds. The trees summarize relationships resolved from analyses of *(A) ASAP3/TM6*, *(B) ASAP1*, *(C)* rDNA ITS, and *(D)* cpDNA restriction site data. Tree C is the result of a reanalysis of the rDNA ITS data (Baldwin 1992), with sampling restricted to the same set of ingroup taxa examined in the present study (unpublished data). Tree C is topologically congruent with published rDNA ITS trees of similar sampling (Baldwin 1992), except that *Madia stebbinsii* was included in the earlier study instead of the closely related *Madia nutans*. Tree D was published previously (Baldwin et al. 1991). Numbers next to the nodes give bootstrap support levels. All nodes with less than 50% support are shown collapsed.

Plausibility of the inferred evolutionary reticulation based on evidence from artificial hybrids

Although up to 5–6 Myr may have elapsed since the onset of diversification of the silversword alliance (Baldwin and Sanderson 1998), enough genetic similarity remains between *R. muirii* and *R. scabrida*, representatives of the two apparent ancestral lineages of the Hawaiian group, to permit formation of hybrids (see Figure 5). A cross between *R. muirii* and *R. scabrida* yielded a vigorous hybrid individual of low fertility (13.5% pollen stainability) (Baldwin 1989). The high proportion of unstained, inviable pollen probably arose from the extensive meiotic failure observed at the chromosomal level (Baldwin 1989; Kyhos, Carr and Baldwin 1990). One third of the stainable, presumably viable pollen grains in this hybrid were unusually large and tetraporate (rather than triporate), a condition correlated with unreduced ploidy in Asteraceae (Baldwin 1989; Kyhos, Carr and Baldwin 1990). Although production of synthetic allopolyploids was not attempted, the *R. muirii* X *R. scabrida* individual (see Figure 5) displayed sufficient vigor and apparent diploid-pollen fertility to have promising prospects for generating an allopolyploid lineage. The hypothesis of an allopolyploid origin of the silversword alliance based on hybridization between lineages including *R. muirii* and *R. scabrida* thus appears to be biologically plausible.

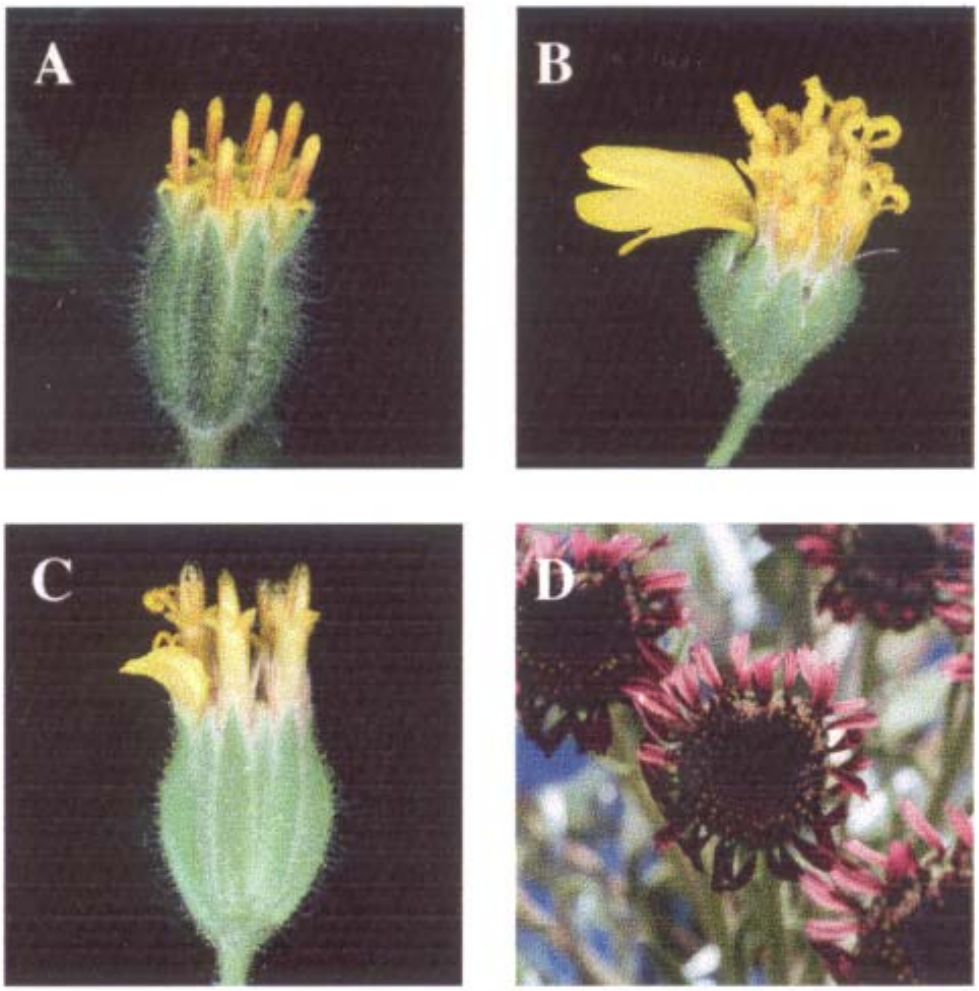

Figure 5. Flowering capitula of *(A) Raillardiopsis muirii*, *(B) Raillardiopsis scabrida*, and *(C)* their interspecific hybrid (*R. muirii* X *R. scabrida*). Note the presence of the ray flower with expanded limb in *R. scabrida*, the absence of ray flowers in *R. muirii*, and the presence of one ray flower with a short limb in the hybrid. The capitulum of the Haleakala silversword, *Argyroxiphium sandwicense* ssp. *macrocephalum (D)*, is included for comparison.

Hybrids between members of the silversword alliance and each of four perennial species of the *Madia/Raillardiopsis* group have also been produced (Baldwin 1989; Kyhos, Carr and Baldwin 1990; Baldwin et al. 1991; Carr, Baldwin and Kyhos 1996). The intergeneric hybrid progeny between *R. muirii* and *D. laevigata* were vigorous with high pollen fertility (pollen stainability 49%), although most of the stainable grains were large, tetraporate, and presumably unreduced (Carr, Baldwin and Kyhos 1996). A cross between *R. scabrida* and a *D. knudsenii* X *Dubautia laxa* hybrid produced a single plant, which died before flowering (Carr, Baldwin and Kyhos 1996). Hybrids have also been produced between *D. knudsenii* and a *M. madioides* X *R. scabrida* hybrid. Based on results of rDNA ITS analysis (Baldwin 1996), *M. madioides* and *R. scabrida* are sister species belonging to the same lineage in the *Madia/Raillardiopsis* group. Hybrids between *D. knudsenii* and the *M. madioides* X *R. scabrida* hybrid were vigorous, with low pollen fertility (pollen stainability ca. 1%) and mostly univalents at meiotic metaphase I. Despite low fertility of the plants, the ability to produce these vigorous intergeneric hybrids indicates that the Hawaiian species and their closest relatives among North American tarweeds continue to retain considerable genetic similarity after millions of years of evolutionary divergence.

Genomic constitution of the Hawaiian founder and extinction implications

Our evidence for allopolyploidy of the silversword alliance leads us to conclude that the tarweed founder, not just the most recent common ancestor, of the Hawaiian group was an allopolyploid. Otherwise, we must hypothesize (1) ancient

long-distance dispersal of two (diploid) North American tarweed species to the same location in the Hawaiian Islands, (2) hybridization and formation of an allopolyploid lineage in situ, and (3) subsequent extinction of the two diploid species in Hawai'i (Baldwin 1997). The alternative scenario of dispersal of an allopolyploid species from western North America is simpler, requiring only one dispersal to the Hawaiian Islands and extinction of one lineage—the North American allopolyploid group (Baldwin 1997). Although multiple polyploid lineages are known within the n = 8 *Madia/Raillardiopsis* lineage, strongly-supported rDNA ITS relationships show that the extant North American polyploids most closely related to the Hawaiian species originated well after divergence of the silversword alliance from the North American tarweeds (Baldwin 1996; unpublished data). We conclude that despite the phenomenal evolutionary success of the silversword alliance ancestor in the Hawaiian setting, allopolyploids of the same genomic constitution in North America proved to be an evolutionary dead end. A similar allopolyploid extinction in western North America following successful long-distance dispersal (to South America) appears to have occurred in the Asteraceae genus Blennosperma (Ornduff 1961).

Genetic consequences of allopolyploidy in the Hawaiian silversword alliance

The inferred hybrid constitution of the Hawaiian founder may have promoted adaptive radiation of the silversword alliance. The presence of two divergent genomes in the colonizing ancestor of the Hawaiian group may have endowed the ancestor with more genetic variation and a greater ability to respond to selection (Jiang et al. 1998)

than would be found in a diploid progenitor. Any intergenomic (allosyndetic) chromosome pairing that may have occurred in the early Hawaiian allopolyploids would have allowed for wide genetic segregation and a broad array of progeny phenotypes (see Grant 1975), thereby conceivably enhancing the prospects for success in diverse island habitats. The impact of such extensive recombination on diversification of the silversword alliance may have been great, particularly in light of the ecological and morphological diversity of the North American lineages implicated in the formation of the hybrid ancestor of the Hawaiian group.

In general, the increased number of different loci correlated with allopolyploid formation provides novel avenues for molecular evolutionary and phenotypic divergence (Soltis and Soltis 1995). Allopolyploidy may impart upon plants a greater degree of genetic variation (Levin 1983; Soltis and Soltis 1995), increased redundancy of developmental pathways (Gibson and Spring 1998), and potential for structural and functional divergence of duplicated loci (Ohno 1970). We conclude that a variety of genetic factors associated with allopolyploidy may have been critically important in the spectacular adaptive radiation of the Hawaiian silversword alliance.

Acknowledgments

We thank S. Bainbridge, M. Bruegmann, J. Canfield, G. Carr, J. Dvorak, E. Friar, R. Hobdy, L. Johnson, D. Kyhos, A. Medeiros, S. Perlman, M. Porter, the late L. Stemmermann, R. Warshauer, B. Wiegmann, and K. Wood for helpful discussions and field support. We also thank D. Swofford for generously providing us with test versions of PAUP* 4.0 and for granting permission to publish our results from analyses using PAUP*. This work was supported in part by NSF grant DEB-9708540 to M. D. P. and R. H. R., an NSF National Young Investigator Award (DEB-9458237) to B. G. B., and an Alfred P. Sloan Foundation Young Investigator Award to M. D. P.

Literature Cited

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1992) Short protocols in molecular biology. John Wiley, New York.

Baldwin, B. G. (1989) Chloroplast DNA phylogenetics and biosystematic studies in Madiinae (Asteraceae). Ph.D. dissertation, University of California, Davis.

Baldwin, B. G. (1992) Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. *Mol. Phylogenet. Evol.* **1**: 3–16.

Baldwin, B. G. (1996) Phylogenetics of the California tarweeds and the Hawaiian silversword alliance (Madiinae: Heliantheae sensu lato). In: Compositae: systematics. Hind, D. J. N. and Beentje, H. J. (eds.) Royal Botanic Gardens, Kew, England. Pp. 377-391.

Baldwin, B. G. (1997) Adaptive radiation of the Hawaiian silversword alliance: congruence and conflict of phylogenetic evidence from molecular and non-molecular investigations. In: Molecular evolution and adaptive radiation. Givnish, T. J. and Sytsma, K. J. (eds.) Cambridge University Press, Cambridge, England. Pp 103-128.

Baldwin, B. G., Kyhos, D. W., Dvorak, J. and Carr, G. D. (1991) Chloroplast DNA evidence for a North American origin of the Hawaiian silversword alliance (Asteraceae). *Proc. Natl. Acad. Sci. USA* **88**: 1840–1843.

Baldwin, B. G. and Robichaux, R. H. (1995) Historical biogeography and ecology of the Hawaiian silversword alliance: New molecular phylogenetic perspectives. In: Hawaiian biogeography: Evolution on a hot spot archipelago. Wagner, W. L. and Funk, V. A. (eds.) Smithsonian Institution Press, Washington, D.C. Pp. 259-287.

Baldwin, B. G. and Sanderson, M. J. (1998) Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proc. Natl. Acad. Sci. USA* **95**: 9402–9406.

Carlquist, S. (1959) Studies on Madiinae: Anatomy, cytology and evolutionary relationships. *Aliso* **4**: 171–236.

Carr, G. D. (1985) Monograph of the Hawaiian Madiinae (Asteraceae): *Argyroxiphium*, *Dubautia* and *Wilkesia*. *Allertonia* **4**: 1–123.

Carr, G. D. (1998a) A new species of *Dubautia* (Asteraceae-Madiinae) from Kaua'i, Hawaiian Islands. *Novon* **8**: 4–7.

Carr, G. D. (1998b) Another new species of *Dubautia* (Asteraceae-Madiinae) from Kaua'i, Hawaiian Islands. *Novon* **8**: 8–11.

Carr, G. D., Baldwin, B. G. and Kyhos, D. W. (1996) Cytogenetic implications of artificial hybrids between the Hawaiian silversword alliance and North American tarweeds (Asteraceae, Heliantheae-Madiinae). *Am. J. Bot.* **83**: 653–660.

Carr, G. D. and Kyhos, D. W. (1986) Adaptive radiation in the Hawaiian silversword alliance (Compositae-Madiinae): II. Cytogenetics of artificial and natural hybrids. *Evolution* **40**: 969–976.

Doyle, J. J. and Doyle, D. L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**: 11–15.

Farris, J. S., Kallersjo, M., Kluge, A. G. and Bult, C. (1995) Constructing a significance test for incongruence. *Syst. Biol.* **44**: 570–572.

Friar, E. A., Robichaux, R. H. and Mount, D. W. (1996) Molecular genetic variation following a population crash in the endangered Mauna Kea silversword, *Argyroxiphium sandwicense* ssp. *sandwicense* (Asteraceae). *Mol. Ecol.* **5**: 687–691.

Frohman, M. A., Dush, M. and Martin, G. R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85**: 8998–9002.

Gibson, T. J. and Spring, J. (1998) Genetic redundancy in vertebrates: polyploidy and persistence of genes encoding multidomain proteins. *Trends Genet.* **14**: 46–49.

Givnish, T. J. (1997) Adaptive radiation and molecular systematics: issues and approaches. In: Molecular evolution and adaptive radiation. Givnish, T. J. and Sytsma, K. J. (eds.) Cambridge University Press, Cambridge, England. Pp. 1-54.

Grant, V. (1975) Genetics of flowering plants. Columbia University Press, New York.

Hillis, D. M., Moritz, C., Porter, C. A. and Baker, R. J. (1991) Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science* **251**: 308–310.

Jiang, C. -X., Wright, R. J., El-zik, K. M. and Patterson, A. H. (1998) Polyploid formation created unique avenues for response to selection in Gossypium (cotton). *Proc. Natl. Acad. Sci. USA* **95**: 4419–4424.

Kramer, E. M., Dorit, R. L. and Irish, V. F. (1998) Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the *APETALA3* and *PISTILLATA* MADS-box gene lineages. *Genetics* **149**: 765–783.

Kyhos, D. W., Carr, G. D. and Baldwin, B. G. (1990) Biodiversity and cytogenetics of the tarweeds (Asteraceae). *Ann. Mo. Bot. Gard.* **77**: 84–95.

Levin, D. A. (1983) Polyploidy and novelty in flowering plants. *Am. Nat.* **122**: 1–25.

Lowman, A. C. and Purugganan, M. D. (1999) Duplication of the *Brassica oleracea APETALA1* gene and the evolution of domesticated cauliflower. *J. Hered.* (in press).

Ma, H., Yanofsky, M. F. and Meyerowitz, E. M. (1991) *AGL1-AGL6*, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. *Genes Dev*. **5**: 484–495.

Mena, M., Ambrose, B. A., Meeley, R. B., Briggs, S. P., Yanofsky, M. F. and Schmidt, R. J. (1996) Diversification of the C-function activity in maize flower development. *Science* **274**: 1537–1540.

Ohno, S. (1970) Evolution by gene duplication. Springer-Verlag, New York.

Ornduff, R. (1961) Experimental studies in two genera of Helenieae (Compositae): Blennosperma and Lasthenia. *Q. Rev. Biol.* **38**: 141–150.

Palmer, J. D. (1986) Isolation and structural analysis of chloroplast DNA. *Methods Enzymol.* **118**: 167–186.

Purugganan, M. D. (1997) The MADS-box floral homeotic gene lineages predate the origin of seed plants: phylogenetic and molecular clock estimates. *J. Mol. Evol.* **45**: 392–396.

Purugganan, M. D., Rounsley, S. D., Schmidt, R. J. and Yanofsky, M. F. (1995) Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family. *Genetics* **140**: 354–356.

Raven, P. H., Evert, R. F. and Eichorn, S. E. (1992) Biology of plants. 5th ed. Worth, New York.

Riechmann, J. L. and E. M. Meyerowitz (1997) MADS domain proteins in plant development. *Biol. Chem.* **378**: 1079–1101.

Robichaux, R. H., G. D. Carr, M. Liebman and R. W. Pearcy (1990) Adaptive radiation of the Hawaiian silversword alliance (Compositae-Madiinae)—ecological, morphological and physiological diversity. *Ann. Mo. Bot. Gard.* **77**: 64–72.

Rounsley, S. D., Ditta, G. S. and Yanofsky, M. F. (1995) Diverse roles for MADS-box genes in *Arabidopsis* development. *Plant Cell* **7**: 1259–1269.

Soltis, D. E. and Soltis, P. S. (1995) The dynamic nature of polyploid genomes. *Proc. Natl. Acad. Sci. USA* **92**: 8089–8091.

Swofford, D. L. (1998) PAUP* 4.0 beta. Sinauer Associates, Sunderland, Mass.

Witter, M. S. and Carr, G. D. (1988) Adaptive radiation and genetic differentiation in the Hawaiian silversword alliance (Compositae: Madiinae). *Evolution* **42**: 1278–1287.

CHAPTER 3.

Accelerated Regulatory Gene Evolution in an Adaptive Radiation

Marianne Barrier, Robert H. Robichaux and Michael D. Purugganan

Proceedings of the National Academy of Sciences, USA 98 (18): 10208-10213 (2001)

Abstract

The disparity between rates of morphological and molecular evolution remains a key paradox in evolutionary genetics. A proposed resolution to this paradox has been the conjecture that morphological evolution proceeds via diversification in regulatory loci, and that phenotypic evolution may correlate better with regulatory gene divergence. This conjecture can be tested by examining rates of regulatory gene evolution in species that display rapid morphological diversification within adaptive radiations. We have isolated homologues to the *Arabidopsis APETALA3* (*ASAP3/TM6*) and *APETALA1* (*ASAP1*) floral regulatory genes and the *CHLOROPHYLL A/B BINDING PROTEIN9* (*ASCAB9*) photosynthetic structural gene from species in the Hawaiian silversword alliance, a premier example of plant adaptive radiation. We have compared rates of regulatory and structural gene evolution in the Hawaiian species to those in related species of North American tarweeds. Molecular evolutionary analyses indicate significant increases in nonsynonymous relative to synonymous nucleotide substitution rates in the *ASAP3/TM6* and *ASAP1* regulatory genes in the rapidly evolving Hawaiian species. By contrast, no general increase is evident in neutral mutation rates for these loci in the Hawaiian species. An increase in nonsynonymous relative to synonymous nucleotide substitution rate is also evident in the *ASCAB9* structural gene in the Hawaiian species, but not to the extent displayed in the regulatory loci. The significantly accelerated rates of regulatory gene evolution in the Hawaiian species

may reflect the influence of allopolyploidy or of selection and adaptive divergence. The analyses suggest that accelerated rates of regulatory gene evolution may accompany rapid morphological diversification in adaptive radiations.

Introduction

Rates of morphological evolution are generally not correlated with rates of molecular evolution. This paradoxical observation was highlighted early by Wilson and coworkers (King and Wilson 1975; Wilson 1977; Cherry, Case and Wilson 1978), and subsequent molecular studies in species groups that have undergone recent adaptive radiations, such as African rift lake cichlids (Meyer et al. 1990), columbines (Hodges and Arnold 1994), and the Hawaiian silversword alliance (Baldwin and Robichaux 1995), have documented marked incongruities in rates of morphological and molecular evolution. A proposed resolution to this paradox has been the conjecture that evolutionary changes in regulatory genes, rather than large-scale diversification in structural genes, may be responsible for interspecific variation in organismal morphologies (King and Wilson 1975; Wilson 1977; Cherry, Case and Wilson 1978; Lowe and Wray 1997; Doebley and Lukens 1998; Purugganan 1998). This conjecture is reinforced by molecular developmental studies that indicate that dramatic shifts in organismal structure may arise from mutations at key regulatory loci (Lowe and Wray 1997; Doebley and Lukens 1998; Purugganan 1998). One test of this conjecture is to examine whether there is a significant acceleration in rates of regulatory gene evolution in species that display rapid morphological diversification within adaptive radiations.

The Hawaiian silversword alliance (Asteraceae: Heliantheae—Madiinae) is a premier example of plant adaptive radiation (Carr 1985; Baldwin and Robichaux

1995; Baldwin 1997). The silversword alliance comprises 30 perennial species in three endemic genera: *Argyroxiphium*, *Dubautia*, and *Wilkesia*. The species are distributed on six of the eight main islands of the Hawaiian archipelago (Kaua'i, O'ahu, Moloka'i, Lana'i, Maui, and Hawai'i), with all but five species being single-island endemics. Species in the silversword alliance grow in a wide range of habitats and exhibit a wide array of growth forms, including rosette plants, cushion plants, shrubs, trees, and lianas. They also display great variation in reproductive traits, especially in inflorescence and floral morphology and capitulescence (or flowering stalk) architecture. This is clearly evident in quantitative phenotypes, such as the numbers of flowers in the capitula, the numbers of capitula per flowering stalk, the architectural organization of flowering stalks, and the sizes and morphologies of floral organs (Carr 1985). Even sibling species that appear to have diverged less than 500,000 years ago on the Island of Hawai'i display significant differences in quantitative inflorescence and floral characters (A. L. Lawton-Rauh, R. H. R. and M. D. P., unpublished observations).

Molecular phylogenetic analyses using chloroplast DNA and rDNA internal transcribed spacer (ITS) sequences (Baldwin et al. 1991; Baldwin and Robichaux 1995; Baldwin 1997) have confirmed an early hypothesis that the closest relatives of the Hawaiian silversword alliance can be found among North American tarweeds (Asteraceae: Heliantheae—Madiinae) in the ''*Madia*'' lineage (Baldwin 1996). Both cytogenetic (Carr and Kyhos 1986) and allozyme (Witter and Carr 1988) data indicate that the Hawaiian species are tetraploids (n = 13–14), in contrast to the basally diploid

condition (n = 6–9) in the most closely related North American species within the ‘‘*Madia*’’ lineage (Baldwin 1996; Barrier et al. 1999). Results of our recent phylogenetic analyses using two floral homeotic genes have led us to conclude that the Hawaiian silversword alliance descended from an interspecific hybrid between members of the *Anisocarpus scabridus* and *Carlquistia muirii* lineages of North American tarweeds (Barrier et al. 1999). Likelihood estimates based on sequence data from the rDNA ITS locus suggest that the most recent common ancestor of the Hawaiian species existed 5.2 ± 0.8 million years ago (mya), contemporaneous with the origin of the Island of Kaua'i (Baldwin and Sanderson 1998). By contrast, the earliest date for the diversification of the North American tarweeds appears to be in the mid-Miocene, 15 mya (Baldwin and Sanderson 1998).

We have isolated homologues to the *Arabidopsis APETALA3* (*ASAP3/TM6*) and *APETALA1* (*ASAP1*) floral regulatory genes and the *CHLOROPHYLL A/B BINDING PROTEIN9* (*ASCAB9*) photosynthetic structural gene from species in the Hawaiian silversword alliance and from related species of North American tarweeds (Barrier et al. 1999). Isolation of the genes from both the rapidly evolving Hawaiian species and their North American relatives provides an opportunity to compare rates of gene evolution between lineages that differ greatly in rates of morphological diversification. Molecular evolutionary analyses indicate significant increases in nonsynonymous relative to synonymous nucleotide substitution rates in the *ASAP3/TM6* and *ASAP1* regulatory genes in the rapidly evolving Hawaiian species. By contrast, no general increase is evident in neutral mutation rates for these loci in

the Hawaiian species. An increase in nonsynonymous relative to synonymous nucleotide substitution rate is also evident in the *ASCAB9* structural gene in the Hawaiian species, but not to the extent displayed in the regulatory loci. These results suggest that the adaptive radiation of the Hawaiian silversword alliance has been accompanied by accelerated rates of regulatory gene evolution, possibly resulting from allopolyploidy or from selection and adaptive divergence.

Materials and Methods

Isolation of genes

The *ASAP3/TM6* and *ASAP1* regulatory genes were isolated as described (Barrier et al. 1999). The *ASCAB9* structural gene was first identified as an expressed sequence tag (EST) in *Argyroxiphium sandwicense* ssp. *macrocephalum*. Ten Hawaiian species were selected to represent the four major lineages in the silversword alliance as previously identified from rDNA internal transcribed spacer (ITS) trees (Baldwin and Robichaux 1995; Baldwin 1997; Barrier et al. 1999). Five North American tarweed species (*A. scabridus*, *C. muirii*, *Harmonia nutans*, *Kyhosia bolanderi*, and *Madia sativa*) were chosen from within the ‘‘*Madia*’’ lineage as identified from rDNA ITS trees (Baldwin 1996). Six other North American tarweed species (*Adenothamnus validus*, *Calycadenia multiglandulosa*, *Centromadia pungens*, *Deinandra lobbii*, *Osmadenia tenella*, and *Raillardella pringlei*) known to fall outside the ‘‘*Madia*’’ lineage (Baldwin 1996) were also included.

PCR primers to amplify *ASAP3/TM6*, *ASAP1*, and *ASCAB9* genomic regions were designed based on cDNA sequences. The primers were used in PCR amplification reactions using the error-correcting rTth polymerase formulation (Perkin–Elmer) or Pwo polymerase (Boehringer Mannheim) in standard buffer with cycling conditions recommended by the manufacturer. The nucleotide error rate for these error-correcting polymerases is less than 1 bp in 7 kb of sequence (J. I. Suddith and M. D. P., unpublished observations). PCR-amplified DNA was cloned by using

the TA cloning kit (Invitrogen) and sequenced by using automated sequencers (Iowa State University Sequencing Facility and North Carolina State University DNA Sequencing Facility). Sequencing was done with nested primers, with multiple sequencing reactions conducted for divergent sequences. All sequence changes were rechecked visually against sequencing chromatograms and deposited in the GenBank database (accession nos. AF147210–AF147258 and AF398723–AF398755).

Reverse transcription (RT)-PCR assay of regulatory gene expression

Total RNA was isolated from developing inflorescences (or capitula) of *Dubautia arborea* by using an RNA plant extraction kit (Qiagen, Chatsworth, CA). Poly(A+)RNA was purified from total RNA with the Oligotex suspension system (Qiagen), and oligo(dT)-primed first-strand cDNA was synthesized by using AMV reverse transcriptase (Promega). Products of the cDNA syntheses were used as a template in RT-PCR reactions using gene-specific primers. For *ASAP3/TM6*, a universal ASAP3XF.1 forward primer (5'-TACAAACAGGCAGGTGACATAC-3') in exon 1 and the *A* copy-specific ASAP3XRB.2 (5'-CTTCTAAGTTTATTGTTGCTC-3') or *B* copy-specific ASAP3XRA.2 (5'-CTTCTAAGTTTGTTGTTGCTA-3') reverse primers in exon 4 were used. For *ASAP1*, a universal ASAP1XF.1 forward primer (5'-GAAACCACAGGCACTATATGGG-3') spanning exons 3 and 4, and the *A* copy-specific ASAP1XAR.1 (5'-CGTTGCTTCTTCCGTCACCTCC-3') or *B* copy-specific ASAP1XBR.1 (5'-CGTTGCTTCTTCCGTCACCTCG-3') reverse primers in exon 8 were used (see Figure 1). PCR amplification reaction conditions were 40 cycles of 95°C for 1 min, 52°C for 1 min, and 72°C for 1 min, using Taq polymerase (Promega). Products were diluted, fractionated on 1% agarose gels, and blotted on nylon filter membranes (Amersham Hybond, Piscataway, NJ). Gel blots were separately probed with ^{32}P-labeled *ASAP3/TM6* and *ASAP1* gene probe mixtures.

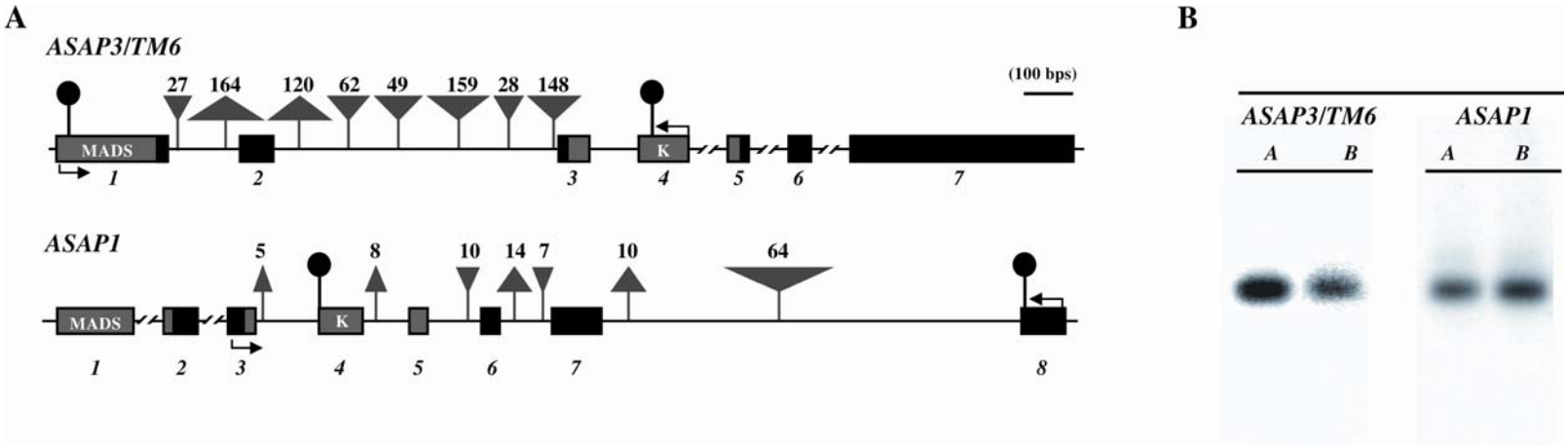

Figure 1. *(A)* Gene maps of *ASAP3/TM6* and *ASAP1* loci. Exons are shown as numbered boxes. The gene maps depicted are for the *A* copies of the loci. Upright and inverted triangles represent major deletions and insertions, respectively, that characterize the *B* duplicate copies. Numbers above the triangles provide the sizes of the indels. Arrows show positions of PCR primers used to isolate genomic sequences. Circles indicate approximate location of copy-specific RT-PCR primers used in gene expression assays. A 100-bp scale bar is provided. Relative sizes of exons and introns in the amplified regions were derived from comparison of genomic and cDNA sequences. Exon sizes outside the amplified regions are estimates based on comparison with data from *A. thaliana* orthologues. *(B)* Expression of the *ASAP3/TM6* and *ASAP1* floral regulatory genes in developing inflorescences of *D. arborea*. Expression was assayed with gene-specific primers (for the duplicate *A* and *B* copies) that amplified ~300 nucleotides of cDNA in RT-PCR reactions. Control reactions using cloned *A* and *B* gene copies indicate the primers are copy-specific. Identity of amplified products was confirmed by sequencing.

Rate analyses

Pairwise nonsynonymous (Ka) and synonymous (Ks) nucleotide substitutions in the coding regions of the *ASAP3/TM6*, *ASAP1*, and *ASCAB9* loci were calculated according to the method of Nei and Gojobori (1986), as implemented in MEGA (Kumar, Tamura and Nei 1993). This method gives unbiased estimates when evaluating sequences of low divergences (Nielsen and Yang 1998). Pairwise Ka/Ks values were calculated for orthologues of *ASAP3/TM6*, *ASAP1*, and *ASCAB9*. The individual pairwise Ka and Ks values reported here can be found in Tables 2–11, which are published as supplemental data on the PNAS web site, www.pnas.org (see Appendix C). Significance values were estimated by using 1,000 two-sample bootstrap replicates generated from the joint distribution of Ka/Ks ratios for the Hawaiian and North American species pairs. The Tajima relative rate test was used to examine rate variation between genes in the Hawaiian species and their closest North American relatives, with *C. pungens* as an outgroup (Tajima 1993). This test has been shown to perform as well as maximum likelihood methods for noncoding region data. Gene phylogenies were reconstructed with PAUP* 4.0b5 (Swofford 2000) as described (Barrier et al. 1999). Ancestral nucleotide character states were reconstructed by using BASEML in the PAML program package (Yang 2000), with the HKY85 nucleotide substitution model (Hasegawa, Kishino and Yano 1985), and assuming no molecular clock.

Results and Discussion

Regulatory and structural genes in the Hawaiian and North American species

Homologues to the *Arabidopsis APETALA3* and *APETALA1* floral regulatory genes were isolated from developing flowers of *A. sandwicense* ssp. *macrocephalum*. The two genes are designated as *ASAP3/TM6* and *ASAP1*, respectively (Barrier et al. 1999). The two loci are members of the MADS-box family of plant transcriptional activators, many of which are known to regulate floral and inflorescence development in angiosperms (Riechmann and Meyerowitz 1997). Phylogenetic analysis indicates that the *ASAP3/TM6* gene is a member of the *AP3* floral homeotic gene group, and appears to be an orthologue of the *Lycopersicon TM6* locus (Purugganan 1997; Kramer, Dorit and Irish 1998). The *ASAP1* gene is an orthologue of the *Arabidopsis AP1* locus. Genetic studies in *Arabidopsis thaliana*, *Antirrhinum majus*, and *Zea Mays* indicate that the developmental functions of the floral regulatory genes exhibit broad conservation across the angiosperms (Weigel 1998). *AP3* orthologues control petal and stamen development, and *AP1* orthologues regulate the establishment of floral primordia and control sepal and petal organ identity (Weigel 1998). Molecular genetic studies in *Arabidopsis* further indicate that changes in the activity of *AP1* are correlated with variation in inflorescence branch number and flowering time (Mandel and Yanofsky 1995), and QTL mapping studies show that changes in petal and stamen sizes in *Arabidopsis* map to a region that includes *AP3* (Juenger, Purugganan and Mackay 2000). Because species in the silversword alliance vary in the number of

flowers per capitula, flowering time, and the sizes of petals and stamens (Carr 1985), both *ASAP3/TM6* and *ASAP1* are candidates for genes that underlie floral and inflorescence morphological and phenological variation among the Hawaiian species.

Gene fragments encompassing exons 1 to 4 of *ASAP3/TM6* and exons 3 to 8 of *ASAP1* were isolated and sequenced from the Hawaiian and North American species (Barrier et al. 1999). The isolated fragments of the *ASAP3/TM6* gene contain the coding region for the first 124 amino acids (aa) of the 227-aa encoded protein, including the MADS-box DNA-binding and putative K-box dimerization domains (Riechmann and Meyerowitz 1997). The isolated fragments of the *ASAP1* gene contain a region encoding 128 aa of the 242-aa protein, including the K-box and C-terminal domains. The C-terminal domain of MADS-box regulatory proteins is believed to contain the transcriptional activation domain (Riechmann and Meyerowitz 1997). The isolation of different regions of the *ASAP3/TM6* and *ASAP1* genes allows sequence evolution to be examined across different domains of these transcriptional activators.

The *ASAP3/TM6* and *ASAP1* genes are present in duplicate copies (designated as the *A* and *B* copies) in the tetraploid Hawaiian species (Barrier et al. 1999). Duplication of genes may lead to inactivation of one gene copy via pseudogene formation (Walsh 1995). RT-PCR assays demonstrate, however, that the *A* and *B* copies of both genes are transcriptionally active in developing inflorescences of *Dubautia* (see Figure 1), indicating that pseudogene formation has not generally occurred at these regulatory loci. Only single copies of the genes have been detected in

North American species within the ‘‘*Madia*’’ lineage. Phylogenetic analysis indicates that the *A* and *B* copies of the genes in the Hawaiian species are most closely related to genes in *A. scabridus* and *C. muirii* from North America (Barrier et al. 1999). For the *ASAP3/TM6* gene, the ranges of sequence divergence are 0.2–1.0% in the Hawaiian species (including both the *A* and *B* copies) and 1.7–5.7% in the North American species. For the *ASAP1* gene, the ranges of sequence divergence are 0.2–1.4% in the Hawaiian species and 2.1–5.2% in the North American species. Plots of estimated vs. observed nucleotide substitutions indicate that neither nonsynonymous nor synonymous sites at these loci are saturated (M. B. and M. D. P., unpublished observations).

The orthologue to the *CHLOROPHYLL A/B BINDING PROTEIN9* photosynthetic structural gene was also isolated from *A. sandwicense* ssp. *macrocephalum*, and is designated as *ASCAB9*. The nuclear *CAB9* gene encodes a chloroplast localized protein ~ 260 aa in length, which is a portion of the CP26/29 CAB precursor protein in photosystem II (Pichersky et al. 1991). A 1.2-kb genomic region encompassing intron 3 to the 3’ untranslated region (UTR) (and including exons 4 to 6) was isolated from ten species in the Hawaiian silversword alliance and six species of North American tarweeds. The Hawaiian species appear to possess three copies of the *ASCAB9* gene. Sequence analyses suggest that intergenic recombination between two of the copies (designated as the *A* and *B* copies) may have given rise to the third copy. All three copies were included in our analyses of rates of sequence evolution. Only one copy of the *ASCAB9* gene has been detected in each of the North

American species. For the *ASCAB9* gene, the ranges of sequence divergence are 0.0–0.7% in the Hawaiian species and 0.1–7.4% in the North American species. Because of the small number of nucleotide changes at this locus, especially in the Hawaiian species, the phylogeny of the *ASCAB9* gene is not as highly resolved as the phylogenies of the *ASAP3/TM6* and *ASAP1* genes.

Accelerated gene evolution in the Hawaiian species

Molecular evolutionary analyses reveal that the *ASAP3/TM6* and *ASAP1* regulatory genes in species of the Hawaiian silversword alliance are evolving faster than their orthologues in the North American tarweed species. This accelerated evolution is evident when comparing the relative levels of nonsynonymous and synonymous nucleotide substitutions in the coding regions of these loci. Protein-encoding genes evolving at the neutral rate have Ka/Ks ratios equal to 1 (Li, Wu and Luo 1985). Most genes, however, are subject to strong purifying selection, resulting in lower nonsynonymous relative to synonymous substitution rates (i.e., Ka/Ks < 1) (Li, Wu and Luo 1985). The mean Ka/Ks values among orthologues for the *ASAP3/TM6* and *ASAP1* loci in the North American species are 0.12 ± 0.11 (mean ± SD) and 0.29 ± 0.12, respectively (see Figure 2). These values are comparable to the mean Ka/Ks value of 0.14 observed for several other plant nuclear loci, and 0.11–0.19 observed for MADS-box floral homeotic genes from other species (Purugganan et al. 1995).

In contrast to the loci in the North American species, the floral regulatory genes display elevated Ka/Ks ratios in the Hawaiian species. The *ASAP3/TM6* gene

has a mean Ka/Ks value of 0.79 ± 0.52, and the *ASAP1* gene has a mean Ka/Ks value of 0.98 ± 0.64 (see Figure 2). Because there are two copies of each gene in the Hawaiian species, Ka/Ks ratios were calculated separately for orthologues of the *A* and *B* copies (see Tables 2–7 – Appendix C); Figure 2 reports the joint distribution of Ka/Ks ratios for both copies in the Hawaiian species. For each gene, the mean Ka/Ks values for the duplicate copies, when calculated separately, are similar. The increases in Ka/Ks values for the regulatory loci in the Hawaiian species are significantly different from that expected by chance, as assessed by a bootstrap resampling test ($P < 0.001$). Moreover, all but two pairwise comparisons among the genes in the North American species have Ka/Ks < 0.5. By contrast, nearly 30% of the pairwise comparisons among the genes in the Hawaiian species have Ka/Ks > 1. For example, eight of the nine coding region changes between the isolated *ASAP1-A* gene regions of *Dubautia scabra* and *Dubautia laevigata* are nonsynonymous substitutions. High Ka/Ks values have previously been observed for genes under diversifying or directional selection, such as the self-incompatibility, gamete recognition, and MHC loci (Swanson and Vacquier 1995; Richman, Uyenoyama and Kohn 1996; Hughes and Yeager 1998).

Unlike the two regulatory genes, the *ASCAB9* structural gene does not exhibit a substantial increase in Ka/Ks ratios in the Hawaiian species. The mean Ka/Ks values for the *ASCAB9* gene in the North American and Hawaiian species are 0.14 ± 0.17 and 0.21 ± 0.30, respectively (see Figure 2 and Tables 8–11 – Appendix C). The mean Ka/Ks values do not include pairwise comparisons that have no synonymous

substitutions. Reflecting the lower levels of sequence divergence at the *ASCAB9* locus, several pairwise comparisons, especially among the Hawaiian species, lack synonymous substitutions, although some have one nonsynonymous substitution.

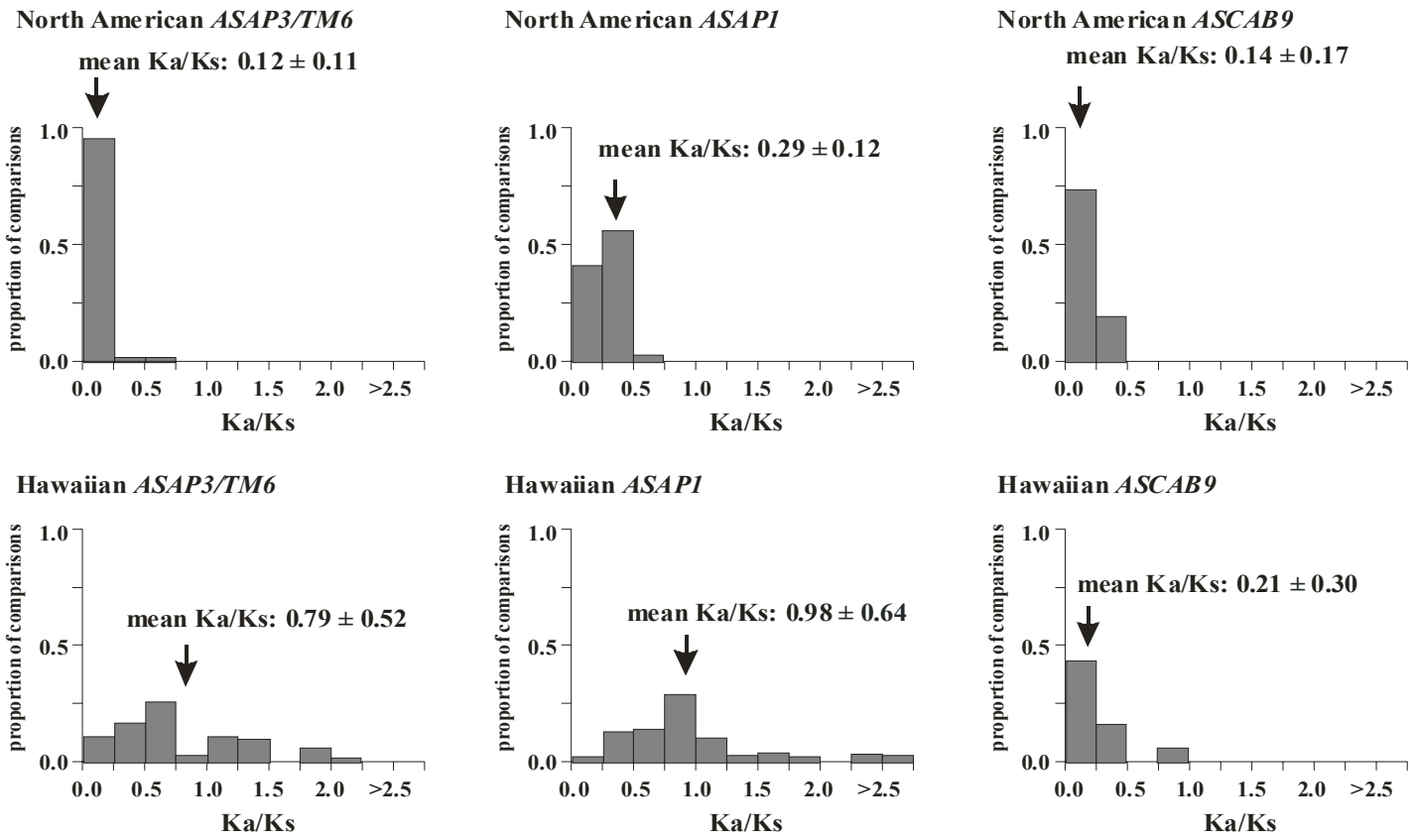

Figure 2. Distribution of Ka/Ks values for the *ASAP3/TM6*, *ASAP1*, and *ASCAB9* genes in the North American tarweeds and Hawaiian silversword alliance. Mean values and standard deviations are indicated. Pairwise comparisons that had no synonymous substitutions are not shown in the histograms and were not included in the analyses. Individual pairwise Ka and Ks estimates are shown in Tables 2–11, which are published as supplemental data on the PNAS web site (see Appendix C).

Accelerated rate confined to the Hawaiian species

Accelerated evolution of the proteins encoded by the floral regulatory genes appears to be confined to the Hawaiian silversword alliance. The number of coding region substitutions was inferred along each branch of the *ASAP3/TM6* and *ASAP1* gene phylogenies by using maximum likelihood ancestral state reconstructions under the HKY85 model of nucleotide substitutions (see Figure 3). The total number of nonsynonymous and synonymous substitutions inferred within the gene phylogenies was then partitioned between the Hawaiian and North American species.

For both the *ASAP3/TM6* and *ASAP1* regulatory genes, the number of nonsynonymous substitutions exceeds the number of synonymous substitutions inferred along branches of the gene phylogenies circumscribed by the Hawaiian silversword alliance (see Table 1). By contrast, most coding region nucleotide substitutions inferred along phylogenetic branches within the North American tarweeds are synonymous. For the *ASAP3/TM6* locus, for example, 25 of the 40 coding region changes (62%) in the Hawaiian species are nonsynonymous substitutions. By contrast, only 5 of the 36 coding region changes (14%) in the North American species are nonsynonymous substitutions. The excess of nonsynonymous over synonymous substitutions in the Hawaiian species compared with the North American species is significant for both the *ASAP3/TM6* locus (Fisher's Exact Test, P = 0.00001) and the *ASAP1* locus (P = 0.002; see Table 1).

The distribution of nonsynonymous and synonymous nucleotide substitutions inferred within the *ASCAB9* gene phylogeny also indicates a significant difference

between the Hawaiian and North American species (Fisher's Exact Test, P = 0.007), with the former having a higher proportion of nonsynonymous substitutions (see Table 1). Unlike the two regulatory genes, however, the *ASCAB9* structural gene does not have an excess of nonsynonymous over synonymous substitutions along branches of the gene phylogenies circumscribed by the Hawaiian silversword alliance.

Neutral mutation rates do not display significant acceleration in the Hawaiian species

The large increases in nonsynonymous relative to synonymous nucleotide substitution rates in the *ASAP3/TM6* and *ASAP1* loci in the Hawaiian species are not correlated with a general acceleration of the neutral mutation rate. Relative rate tests (Tajima 1993) indicate that the regulatory loci do not display a significant increase in nucleotide substitution rates for the largely synonymous third codon positions in the Hawaiian species compared with the North American *Anisocarpus* and *Carlquistia* species (Tajima's test, $P > 0.1$). For both regulatory genes, there is an increase in the rate of intron evolution among the North American and Hawaiian species (Tajima's test, $P < 0.005$), but this involves a rate increase for the more derived loci of North American and Hawaiian species relative to loci of North American species found at more basal positions within the gene phylogenies (see Figure 3). Thus, the increase is not confined to loci in the Hawaiian species. For the *ASCAB9* gene, there is neither an increase in nucleotide substitution rates for the third codon positions (Tajima's test, P

> 0.1) nor an increase in the rate of intron evolution (Tajima's test, P > 0.1) in the Hawaiian species compared with the North American species.

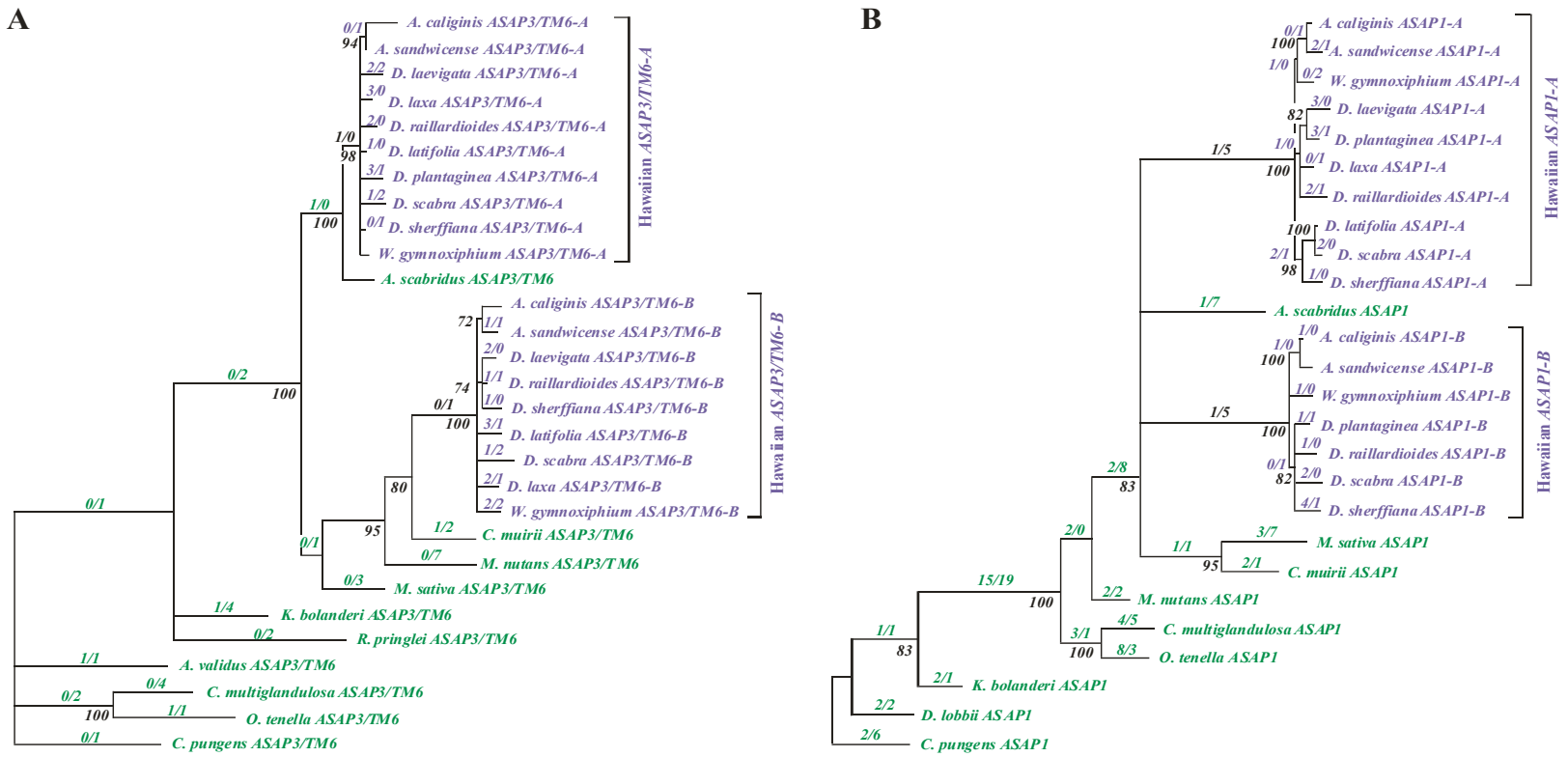

Figure 3. Phylogenies of the *(A) ASAP3/TM6* and *(B) ASAP1* genes in the North American tarweeds (green) and Hawaiian silversword alliance (blue). The phylogenies were reconstructed by using coding region and noncoding region sequences. The numbers of nonsynonymous (N) and synonymous (S) nucleotide substitutions in the coding regions inferred along each branch by maximum likelihood ancestral state reconstructions are given as a ratio (N/S). Levels of bootstrap support are shown next to the nodes. Nodes with less than 70% bootstrap support, and with no coding region nucleotide substitutions inferred along the subtending branch, are collapsed. For the North American species, the generic abbreviations are: A, *Anisocarpus*; Ad, *Adenothamnus*; C, *Carlquistia*; Ca, *Calycadenia*; Ce, *Centromadia*; D, *Deinandra*; H, *Harmonia*; K, *Kyhosia*; M, *Madia*; O, *Osmadenia*; and R, *Raillardella*. For the Hawaiian species, the generic abbreviations are: A, *Argyroxiphium*; D, *Dubautia*; and W, *Wilkesia*.

Table 1. Contingency analyses of coding region substitutions in the Hawaiian and North American genes

	ASAP3/TM6		*ASAP1*		*ASCAB9*	
	HI	NA	HI	NA	HI	NA
Nonsynonymous	25	5	28	50	7	3
Synonymous	15	31	11	64	9	31
	P = 0.00001***		P = 0.002**		P = 0.007**	

The numbers of nonsynonymous and synonymous nucleotide substitutions in the coding regions were inferred along branches of the gene phylogenies using maximum likelihood ancestral state reconstructions (see Figure 3). Inferred substitutions along the phylogenetic branches at the bases of the *A* and *B* copies of the *ASAP3/TM6* and *ASAP1* loci in the Hawaiian species (denoted in black in Figure 3) were not included in the analyses, as the substitutions may have occurred in genes of either North American or Hawaiian species. HI, Hawaiian; NA, North American. Significance was calculated by using Fisher's Exact Test (**, $P < 0.01$; ***, $P < 0.001$).

Accelerated regulatory gene evolution and adaptive radiation of the Hawaiian silversword alliance

Although both regulatory and structural genes display an increase in molecular evolutionary rates in the Hawaiian silversword alliance, there is a marked difference between the two gene classes in nonsynonymous relative to synonymous nucleotide substitution rates. Phylogenetic mapping of coding region changes indicates a significant excess of nonsynonymous over synonymous substitutions for the *ASAP3/TM6* and *ASAP1* regulatory genes, but not for the *ASCAB9* structural gene. It is unclear that the results for the *ASCAB9* locus are representative of other structural genes. However, the results for this locus are in general agreement with previous studies indicating limited genetic differentiation in structural allozyme loci between species in the Hawaiian silversword alliance (Witter and Carr 1988). Thus, accelerated regulatory gene evolution, more than structural gene evolution, appears closely correlated with the rapid morphological diversification of the Hawaiian silversword alliance.

The significantly accelerated rates of regulatory gene evolution in the Hawaiian species may reflect the influence of allopolyploidy or of selection and adaptive divergence. The Hawaiian silversword alliance appears to have descended from an interspecific hybrid between members of two extant North American tarweed lineages (Barrier et al. 1999). Polyploidy, which is believed to be an important avenue for plant species formation (Stebbins 1950; Wendel 2000; Soltis and Soltis 2000), may be associated with increases in genetic heterozygosity (Soltis and Soltis 1989),

chromosomal diversification (Song et al. 1995), and phenotypic variation (Roose and Gottlieb 1976; Jiang et al. 1998). Gene duplication events, such as those observed under polyploidy, may also result in increased rates of nonsynonymous nucleotide substitutions (Li 1985; Ohta 1994). Yet, allopolyploidy might be expected to have genome-wide effects, evident in both regulatory and structural genes. Given the significant excess of nonsynonymous over synonymous substitutions for the *ASAP3/TM6* and *ASAP1* loci, but not for the *ASCAB9* locus, among the Hawaiian species, it is unlikely that allopolyploidy alone has been the key factor influencing regulatory gene evolution during the adaptive radiation.

Many of the pairwise interspecific Ka/Ks values for the *ASAP3/TM6* and *ASAP1* loci in the Hawaiian species are greater than 1 (see Figure 2), which strongly suggests that selection and adaptive divergence may have operated to shape the structure of these transcriptional activators (Swanson and Vacquier 1995; Richman, Uyenoyama and Kohn 1996; Hughes and Yeager 1998). The Hawaiian species differ greatly in a suite of reproductive traits, including floral organ size and morphology, inflorescence size, and capitulescence architecture (Carr 1985). Diversifying or directional selection, acting on variation in these and other reproductive traits, may have led to adaptive divergence of the floral regulatory genes. Further studies may help to clarify whether specific molecular changes at these candidate regulatory genes underlie variation in specific reproductive traits.

In summary, our results suggest that rapid morphological diversification during the adaptive radiation of the Hawaiian silversword alliance has been accompanied by

accelerated evolution of genes that regulate developmental processes. Thus, our results may help to resolve a key paradox in evolutionary genetics. Whereas rates of morphological evolution are generally not correlated with rates of structural gene evolution (King and Wilson 1975; Wilson 1977; Cherry, Case and Wilson 1978), they may be correlated with rates of regulatory gene evolution.

Acknowledgements

We gratefully acknowledge Bruce Baldwin and Amy Lawton-Rauh for providing materials and unpublished data, and Jane Suddith and Abee Boyles for technical support. This work was supported in part by a grant from the National Science Foundation (to M. D. P. and R. H. R.) and an Alfred P. Sloan Foundation Young Investigator Award (to M. D. P.).

Literature Cited

Baldwin, B. G. (1996) Phylogenetics of the California tarweeds and the Hawaiian silversword alliance (Madiinae: Heliantheae sensu lato). In: Compositae: systematics. Hind, D. J. N. and Beentje, H. J. (eds.) Royal Botanic Gardens, Kew, England. Pp. 377–391.

Baldwin, B. G. (1997) Adaptive radiation of the Hawaiian silversword alliance: congruence and conflict of phylogenetic evidence from molecular and non-molecular investigations. In: Molecular evolution and adaptive radiation. Givnish, T. J. and Sytsma, K. J. (eds.) Cambridge University Press, Cambridge, England. Pp. 103–128.

Baldwin, B. G., Kyhos, D. W., Dvorak, J. and Carr, G. D. (1991) Chloroplast DNA evidence for a North American origin of the Hawaiian silversword alliance (Asteraceae). *Proc. Natl. Acad. Sci. USA* **88**: 1840–1843.

Baldwin, B. G. and Robichaux, R. H. (1995) Historical biogeography and ecology of the Hawaiian silversword alliance: New molecular phylogenetic perspectives. In: Hawaiian Biogeography: Evolution on a hot spot archipelago. Wagner, W. L. and Funk, V. A. (eds.), Smithsonian Institution Press, Washington, D.C. Pp. 259–287.

Baldwin, B. G. and Sanderson, M. J. (1998) Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proc. Natl. Acad. Sci. USA* **95**: 9402–9406.

Barrier, M., Baldwin, B. G., Robichaux, R. H. and Purugganan, M. D. (1999) Interspecific hybrid ancestry of a plant adaptive radiation: Allopolyploidy of the Hawaiian silversword alliance (Asteraceae) inferred from floral homeotic gene duplications. *Mol. Biol. Evol.* **16**: 1105–1113.

Carr, G. D. (1985) Monograph of the Hawaiian Madiinae (Asteraceae): *Argyroxiphium, Dubautia,* and *Wilkesia. Allertonia* **4**: 1–123.

Carr, G. D. and Kyhos, D. W. (1986) Adaptive radiation in the Hawaiian silversword alliance (Compositae-Madiinae). II Cytogenetics of artificial and natural hybrids. *Evolution* **40**: 959–976.

Cherry, L. M., Case, S. M. and Wilson, A. C. (1978) Frog perspective on morphological difference between humans and chimpanzees. *Science* **200**: 209–211.

Doebley, J. and Lukens, L. (1998) Transcriptional regulators and the evolution of plant form. *Plant Cell* **10**: 1075–1082.

Hasegawa, M., Kishino, H. and Yano, T. J. (1985) Dating of the human ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**: 160–174.

Hodges, S. A. and Arnold, M. L. (1994) Columbines – A geographically widespread species flock. *Proc. Natl. Acad. Sci. USA* **91**: 5129–5132.

Hughes, A. L. and Yeager, M. (1998) Natural selection at major histocompatibility complex loci of vertebrates. *Annu. Rev. Genetics* **32**: 415–435.

Jiang, C.-X., Wright, R. J., El-Zik, K. M. and Paterson, A. H. (1998) Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). *Proc. Natl. Acad. Sci. USA* **95**: 4419–4424.

Juenger, T., Purugganan, M. D. and Mackay, T. F. C. (2000) Quantitative trait loci for floral morphology in *Arabidopsis thaliana. Genetics* **156**: 1379–1392.

King, M. C. and Wilson, A. C. (1975) Evolution at 2 levels in humans and chimpanzees. *Science* **188**: 107–116.

Kramer, E. M., Dorit, R. L. and Irish, V. F. (1998) Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the *APETALA3* and *PISTILLATA* MADS-box gene lineages. *Genetics* **149**: 765–783.

Kumar, S., Tamura, K. and Nei, M. (1993) MEGA: Molecular Evolutionary Genetics Analysis. Pennsylvania State Univ., University Park, PA, Version 1.0.

Li, W. H. (1985) In: Population Genetics and Molecular Evolution. Ohta, T. and Aoki, K. (eds.) Springer, Berlin.

Li, W. -H., Wu, C. -I. and Luo, C. -C. (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* **2**: 150–174.

Lowe, C. J. and Wray, G. A. (1997) Radical alterations in the roles of homeobox genes during echinoderm evolution. *Nature* (London) **389**: 718–721.

Mandel, M. A. and Yanofsky, M. F. (1995) A gene triggering flower formation in *Arabidopsis*. *Nature* (London) **377**: 522–524.

Meyer, A., Kocher, T. D., Basasibwaki, P. and Wilson, A. C. (1990) Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature* (London) **347**: 550–553.

Nei, M. and Gojobori, T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**: 418–426.

Nielsen, R. and Yang, Z. (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**: 929–936.

Ohta, T. (1994) Further examples of evolution by gene duplication revealed through DNA sequence comparisons. *Genetics* **138**: 1331–1337.

Pichersky, E., Subramaniam, R., White, M. J., Reid, J., Aebersold, R. and Green, B. R. (1991) *Mol. Gen. Genet.* **227**: 277–284.

Purugganan, M. D. (1997) The MADS-box floral homeotic gene lineages predate the origin of seed plants: phylogenetic and molecular clock estimates. *J. Mol. Evol.* **45**: 392–396.

Purugganan, M. D. (1998) The molecular evolution of development. *BioEssays* **20**: 700–711.

Purugganan, M. D., Rounsley, S. D., Schmidt, R. J. and Yanofsky, M. F. (1995) Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family. *Genetics* **140**: 345–356.

Richman, A. D., Uyenoyama, M. and Kohn, J. (1996) Allelic diversity and gene genealogy at the self-incompatibility locus in the solanaceae. *Science* **273**: 1212–1216.

Riechmann, J. L. and Meyerowitz, E. M. (1997) MADS domain proteins in plant development. *Biol. Chem.* **378**: 1079–1101.

Roose, M. L. and Gottlieb, L. D. (1976) Genetic and biochemical consequences of polyploidy in tragopogon. *Evolution* **30**: 818–830.

Soltis, D. E. and Soltis, P. S. (1989) Genetic consequences of autopolyploidy in *tolmiea* (Saxifragaceae). *Evolution* **43**: 586–594.

Soltis, P. S. and Soltis, D. E. (2000) The role of genetic and genomic attributes in the success of polyploids. *Proc. Natl. Acad. Sci. USA* **97**: 7051–7057.

Song, K. M., Lu, P., Tang, K. L. and Osborn, T. C. (1995) Rapid genomic change in synthetic polyploids of *Brassica* and its implications for polyploidy evolution. *Proc. Natl. Acad. Sci. USA* **92**: 7719–7723.

Stebbins, G. L. (1950) Variation and Evolution in Plants. Columbia Univ. Press, New York.

Swanson, W. J. and Vacquier, V. D. (1995) Extraordinary divergence and positive Darwinian selection in a fusagenic protein coating the acrosomal process of abalone spermatozoa. *Proc. Natl. Acad. Sci. USA* **92**: 4957–4961.

Swofford, D. L. (2000) PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods) Sinauer, Sunderland, MA. Version 4.0b4a.

Tajima, F. (1993) Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* **135**: 599–607.

Walsh, J. B. (1995) How often do duplicated genes evolve new functions. *Genetics* **139**: 421–428.

Weigel, D. (1998) From floral induction to floral shape. *Curr. Opin. Plant Biol.* **1**: 55–59.

Wendel, J. F. (2000) Genome evolution in polyploids. *Plant Mol. Biol.* **42**: 225–249.

Wilson, A. C. (1977) In: Molecular Evolution. Ayala, F. J. (ed.) Sinauer, Sunderland, MA. P. 225.

Witter, M. S. and Carr, G. D. (1988) Adaptive radiation and genetic differentiation in the Hawaiian silversword alliance (Compositae: Madiinae). *Evolution* **42**: 1278–1287.

Yang, Z. (2000) Phylogenetic Analysis by Maximum Likelihood (PAML) University College, London. Version 3.0.

CHAPTER 4.

Selection on Rapidly-Evolving Genes in the *Arabidopsis* Genome

Manuscript authored by

Marianne Barrier, Carlos D. Bustamante and Michael D. Purugganan

Abstract

Genes that have undergone positive or diversifying selection are likely to be associated with adaptive divergence between species. One indicator of adaptive selection at the molecular level is an excess of amino acid replacement fixed differences per replacement site relative to the number of silent fixed differences per silent site (ω = Ka/Ks). We used an evolutionary expressed sequence tag (EST) approach to estimate the distribution of ω among 304 orthologous loci between *Arabidopsis thaliana* and *Arabidopsis lyrata* to identify genes potentially involved in the adaptive divergence between these two Brassicaceae species. We find that 21 of 304 genes (7%) have an estimated $\omega > 1$, and are candidates for genes associated with adaptive divergence. Molecular population genetic analysis of seven of these putative rapidly-evolving loci indicate that, despite their high levels of between-species nonsynonymous divergence, these genes do not have elevated levels of intraspecific replacement polymorphisms. A hierarchical Bayesian analysis of protein coding region evolution within and between species also indicates that the selection intensities of these genes are elevated compared to typical *A. thaliana* nuclear loci.

Introduction

The genetic architecture of species differences has been the subject of intense study in the last few years (Orr and Coyne 1992; Haag and True 2001; Wu 2001). There has been a concerted effort to identify genes responsible for adaptive differences between species to examine the genetic mechanisms that accompany evolutionary diversification (Haag and True 2001) and even speciation (Wu 2001). Adaptive morphological and physiological differences between species should leave a signature of positive selection at the molecular level, and permit an analysis of evolutionary divergence at both the molecular genetic and phenotypic levels (Haag and True 2001). By investigating loci whose sequences have been shaped by positive selection, it may be possible to unravel the evolutionary genetic mechanisms that underlie adaptive divergence between species and the origins and evolution of species differences.

One indicator of adaptive selection at the molecular level is an excess of amino acid replacement fixed differences per replacement site relative to the number of silent fixed differences per silent site (ω = Ka / Ks) [Li, Gojobori and Nei 1981; Li, Wu and Luo 1985; Hughes and Yeager 1998]. Purifying selection on amino acid variation, for example, causes a decrease in the rate of amino acid fixation and thus an inferred $\omega < 1$. If most amino acid variation is neutral, such as in pseudogenes, $\omega \sim 1$ (Li, Gojobori and Nei 1981). Strong diversifying or positive selection operating on amino acid variation is associated with $\omega > 1$ (Hughes and Yeager 1998; Anisimova, Bielawski

and Yang 2001). Empirically, the distribution of ω varies radically among different classes of genes: among plant nuclear genes, $\omega \sim 0.14$ (Purugganan et al. 1995), while for many mammalian pseudogenes ω has been shown to cluster around 1.0 (Bustamante, Nielsen and Hartl 2002). In contrast, values of $\omega > 1$ have been observed in gamete recognition protein-coding genes (Swanson and Vacquier 1995), loci associated with host-parasite interaction (Hughes 1991), and genes involved in adaptation to specific environments (Messier and Stewart 1997). The elevated values of ω in these latter genes, which we refer to as rapidly-evolving genes, are believed to arise from selection for divergence in protein structure and function.

Identifying genes with increased values of ω will be facilitated by evolutionary genomic approaches that permit investigators to sample and compare large numbers of genes between species genomes to search for those loci characterized by rapid evolution (Swanson et al. 2001; Schmid and Aquadro 2001). Although whole genome sequences are not generally available from two closely related species, rapidly-evolving genes can be identified using expressed sequence tags (ESTs). An evolutionary EST approach has been used, for example, in demonstrating that genes encoding putative accessory gland-specific proteins in *Drosophila* species evolve faster than other loci in the genome, possibly as a result of selection pressures associated with mate choice and intersexual genomic conflict (Swanson et al. 2001).

Here we report the use of ESTs to identify putative rapidly-evolving genes in the *Arabidopsis* genome. Expressed sequence tags from developing inflorescences of *Arabidopsis lyrata* were compared to the whole genome sequence of the model plant

A. thaliana to estimate the distribution of nonsynonymous and synonymous substitution differences between these two Brassicaceae species. Twenty-one putative rapidly-evolving genes with ω greater than 1 were identified between these two plant taxa. These genes have accelerated rates of nonsynonymous substitutions that may be associated with adaptive evolution since the divergence of these two species approximately 5 mya (Kusaba et al. 2001). Molecular population genetic analysis of seven of these putative rapidly-evolving loci confirms that the selection intensities on protein sequence change in these genes are significantly higher than those for typical *Arabidopsis* nuclear loci.

Materials and Methods

Isolation and sequencing of expressed sequence tags

Seeds from individuals of a population of *Arabidopsis lyrata* in Karhumaki, Russia were obtained from Outi Savolainen and Helmi Kuittinen. Total RNA was extracted from the *A. lyrata* inflorescences using the RNeasy Plant Mini Kit (Qiagen) and a cDNA library constructed in the plasmid vector pCMV-PCR using the PCR cDNA Library Construction Kit (Stratagene). Plasmid DNA from cDNA clones were isolated using the REAL Prep96 BioRobot Kit (Qiagen) with the BioRobot 9600 (Qiagen), and sequenced using an ABI Prism 3700 96-capillary automated sequencer (Perkin-Elmer). Sequences were edited based on Phred (Ewing et al. 1998) quality scores, with a Phred scoring threshold of 20. Ambiguous base calls were visually confirmed against the chromatograms. These EST sequences are deposited in GenBank (accession numbers BQ834040 – BQ834596, BQ839827 – BQ839830).

Analyses of ESTs

A. thaliana sequences homologous to the high-quality *A. lyrata* EST sequences were identified by BLAST analysis against the *A. thaliana* whole genome coding database found at The *Arabidopsis* Information Resource database (www.*Arabidopsis*.org) using a maximum expected value (E) of e^{-5}. The GenBank non-redundant nucleotide sequence database was also searched to find the closest matching *A. thaliana* genomic BAC clone sequence. The top matches from each

database were visually aligned with their matching *A. lyrata* EST sequence. Calculations were made based on pairwise comparisons between the *A. lyrata* EST and *A. thaliana* coding region genomic DNA sequence. EST-based nonsynonymous and synonymous distances were calculated using the modified Nei and Gojobori method (Nei and Gojobori 1986) as implemented in MEGA 2.1 (Kumar, Tamura and Nei 1993). Since there are several possible sources of sequence error in the experimental acquisition of the EST sequence data, we refer to these estimates of nonsynonymous and synonymous substitutions as Ka* and Ks*, respectively. In general, Ka* = Ka + ε and Ks* = Ks + ε, where ε is an error term that reflects the empirical error in sequence determination. Re-isolation and sequencing of previously identified genes shows no sequencing errors, and suggests that ε is negligible (M. Barrier, unpublished results).

Isolation and sequencing of alleles

Seven genes with Ka*/Ks* values greater than 1 were chosen for molecular population genetic analysis. Primers were designed to amplify 1-2 kb regions of these genes based on the *A. thaliana* sequences in the selected comparisons (see Table 1; Supplementary Information – Appendix D). Leaf tissue from 10-13 *A. thaliana* ecotypes was obtained from single-seed propagated material provided by the *Arabidopsis* Biological Resource Center. DNA was isolated from these *A. thaliana* ecotypes as well as 2-5 *A. lyrata* individuals (see above) using the DNeasy Plant Mini Kit (Qiagen). PCR of *A. thaliana* samples was performed with Taq DNA polymerase

(Eppendorf) using standard protocols. *A. thaliana* samples were sequenced directly via cycle sequencing with primers in both directions. PCR of *A. lyrata* samples was performed with the error-correcting Tgo polymerase (Roche) using the manufacturer's amplification protocol. The error rate of error-correcting polymerases is less than 1 in 7000 basepairs (Olsen et al. 2002). Amplified *A. lyrata* products were cloned into pCR 4Blunt-TOPO vector using the Zero Blunt TOPO PCR Cloning Kit (Invitrogen). Plasmid miniprep DNA was isolated using the QIAprep Miniprep Kit (Qiagen), and sequenced via cycle sequencing. DNA sequencing was conducted with a Prism 3700 96-capilary automated sequencer (Perkin-Elmer). GenBank accession numbers for this population data are AY140430 – AY140531.

Molecular population genetic data analysis

Sequences of *A. thaliana* and *A. lyrata* populations were aligned and visually corrected. All polymorphisms were visually checked against chromatographs or via re-sequencing. Analysis of polymorphism and divergence was carried out using DnaSP 3.5 (Rozas and Rozas 1999). *A. thaliana* species-wide silent site nucleotide diversity, π (Nei 1987), and θ (Watterson 1975) were estimated. The McDonald-Kreitman test (McDonald and Kreitman 1991) was performed to test for neutral evolution in the protein-coding region. A hierarchical Bayesian method is utilized to analyze McDonald-Kreitman-type tables for 12 typical (Bustamante et al. 2002) and 7 putative rapidly-evolving *A. thaliana* genes to estimate selection coefficients for replacement

changes under a Poisson Random Field model (Sawyer and Hartl 1992). Details of the analytical approach utilized here are described in the accompanying Appendix *(A)*.

Results and Discussion

Expressed sequence tags in *Arabidopsis lyrata*

A small collection of ESTs were isolated and sequenced to obtain genes expressed in developing inflorescences in *A. lyrata*. From a cDNA library of about 2800 colonies, 768 clones were sequenced. From this sequence collection, 561 good-quality sequences of least 200 basepairs in length were subjected to further analysis. *A. thaliana* orthologues to these *A. lyrata* ESTs were identified by BLAST searches of the whole genome *A. thaliana* coding sequence database. Duplicate EST matches for *A. thaliana* coding sequences were eliminated. The GenBank non-redundant nucleotide database was also searched for *A. thaliana* BAC clone sequences containing genes homologous to the *A. lyrata* EST sequences. By aligning genomic BAC clone sequence with the *A. lyrata* and *A. thaliana* coding sequences, the boundaries of non-coding regions were located. After eliminating alignments with fewer than 150 bps of coding sequence, 304 unique ESTs remained for further analysis (see Table 2; Supplementary Information –Appendix D).

The *A. lyrata* EST sequences were assigned to different functional categories using the classifications of the orthologous *A. thaliana* sequences from the TAIR database (*Arabidopsis* Genome Initiative 2000). Unclassified genes and those whose classifications were ambiguous were not included in this comparison. Of the more than 25,000 sequences in the *A. thaliana* coding sequence database, only slightly more than 4000 have thus far been unambiguously classified (*Arabidopsis* Genome

Initiative 2000). Eighty-seven of the 304 unique *A. lyrata* ESTs matched an *A. thaliana* coding sequence that has yet to be classified. In determining the count for each functional category, multiple categories listed for a single sequence were each counted as an equal fraction of the sample count. Based on this analysis, the range of functional categories represented by the 304 unique ESTs appears to be representative of those observed for the entire *A. thaliana* gene set (see Figure 1).

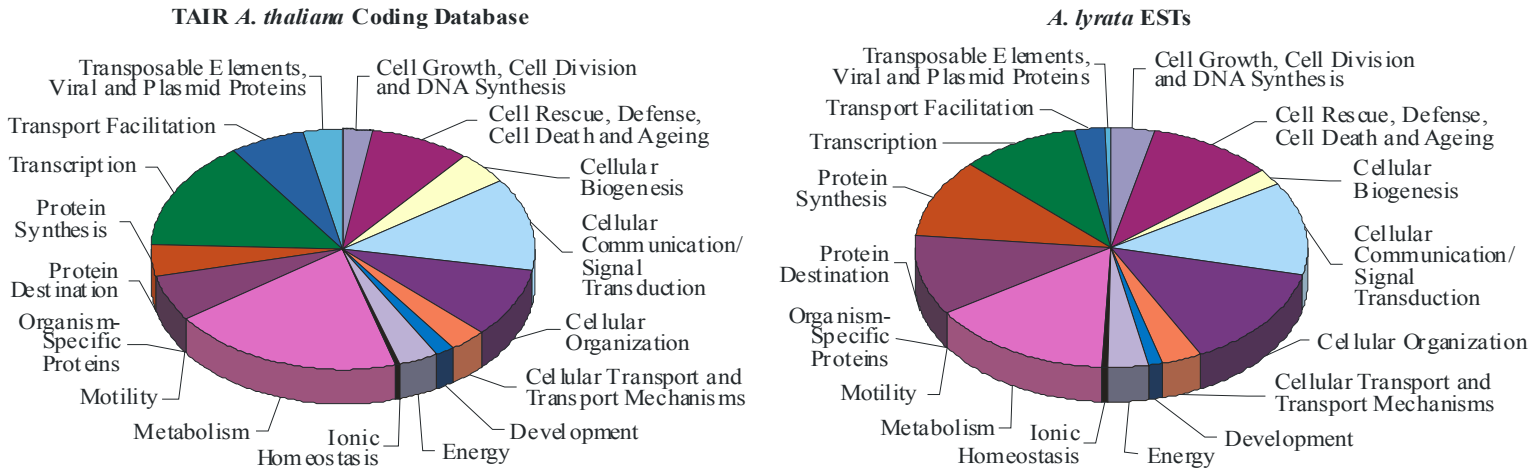

Figure 1. Comparison of functional classifications of genes in the *A. thaliana* genome and the *A. lyrata* evolutionary ESTs.

Distribution of nucleotide substitution rates between *A. thaliana* and *A. lyrata*

Comparisons of the coding sequences of the 304 ESTs from *A. lyrata* with the whole genome sequence from *A. thaliana* allow us to compare the distribution of the rates of nucleotide substitution between these two species. The distributions of both synonymous and nonsynonymous substitutions are shown in Figure 2. The distribution of the number of synonymous nucleotide substitutions ranges from 0.000 to 0.550 synonymous substitutions per synonymous site. The distribution of Ks* has one mode between 0.10 - 0.15, and has a long tail. The mean synonymous substitution distance, which may be considered an estimate of the number of neutral substitutions, is 0.116 ± 0.004 (see Figure 2A), which is comparable to estimates observed in previous comparisons of *A. thaliana*/*A. lyrata* loci. If we assume a divergence time of 5 mya for the two species (Kusaba et al. 2001), then this indicates that the average neutral mutation rate for *Arabidopsis* nuclear genes is approximately 1.2×10^{-8} substitutions/site/year.

The distribution of the number of nonsynonymous substitutions between the two species differs from that observed for synonymous substitutions. The nonsynonymous distance distribution has a mode of less than 0.050 nonsynonymous substitutions/nonsynonymous site, and the frequency decreases with increasing nonsynonymous substitution distances until ~ 0.150 (see Figure 2B). The range of Ka* is 0.000 to 0.159 nonsynonymous substitutions/nonsynonymous site, with a mean of 0.026 ± 0.002. As expected, the mean Ka* is less than Ks* which reflects that action of purifying selection that prevents many nonsynonymous mutations from

reaching fixation between species. The mean rate of nonsynonymous substitutions in nuclear genes between the two species is 0.26×10^{-8} substitutions/site/year, which is ~ 5-fold lower than the average neutral mutation rate.

The distribution in evolutionary rates between species assumes that the comparisons are for orthologues between the two species. Identifying the correct orthologues between species will be confounded by gene duplications immediately at or prior to the most recent common ancestor of *A. thaliana* and *A. lyrata*, followed by deletion of one of the duplicate gene copies in the *A. thaliana* genome. It is unclear how prevalent this situation is, but it is likely to be rare.

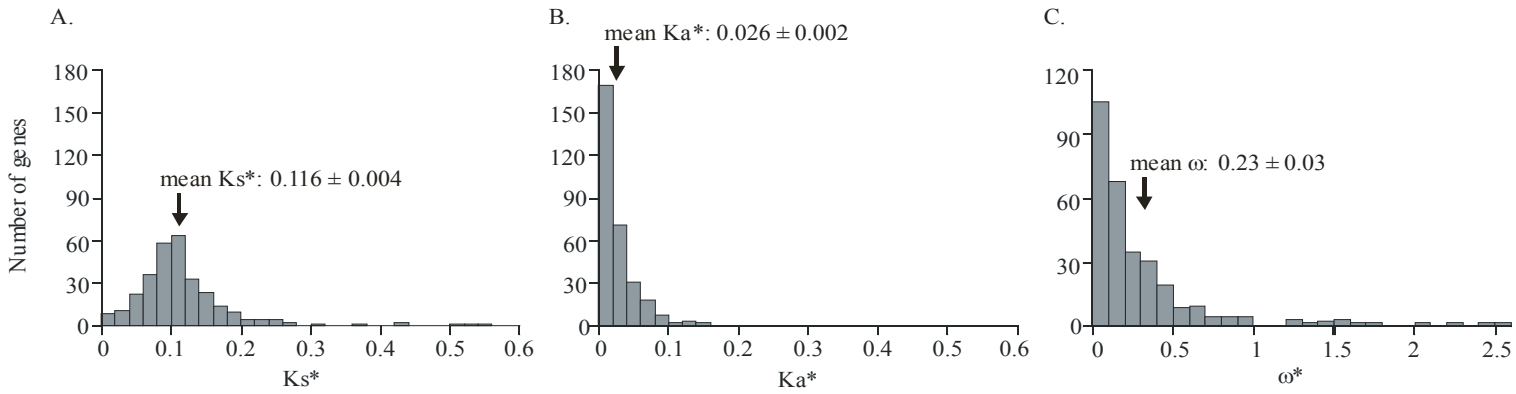

Figure 2. Distributions of *(A)* synonymous (Ks*) and *(B)* synonymous (Ka*) sequence distances, as well as *(C)* ω* estimates, among 304 orthologues between *A. thaliana* and *A. lyrata.*

Variation in selective constraints among *Arabidopsis* genes

The historical action of selection on a gene can be inferred from the relative ratio of nonsynonymous to synonymous substitutions, ω = Ka/Ks (Li, Gojobori and Nei 1981; Li, Wu and Luo 1985; Hughes and Yeager 1998; Anisimova, Bielawski and Yang 2001). The evolutionary EST data can be used to estimate the distribution of the selection parameter ω* between *A. thaliana* and *A. lyrata*. The value of ω* = Ka*/Ks* is estimated for each of the 304 orthologous pairs between the two species, corrected for sequencing errors. The estimates of ω* range from 0.00 to 2.59; the distribution is similar to that of Ka* in that most genes have a low ω* value, and the distribution decreases with increasing ω* (see Figure 2C). The mean value of ω*, obtained by bootstrap resampling of Ka* and Ks* pair values 10,000 times (Bustamante, Nielsen and Hartl 2002), is 0.23 [95% confidence interval (CI): $0.197 \leq \omega^* \leq 0.256$].

Of the 304 orthologous gene pairs in this evolutionary EST study, 21 genes (~7%) have ω* values greater than 1, suggesting that these genes may have accelerated rates of nonsynonymous substitutions associated with the action of diversifying or positive selection. These loci include genes that encode RNA and zinc-finger helicases, *extensin* and *Rac*-like proteins, and zinc finger transcription factors. Nearly half of these putative rapidly-evolving loci (9 out of 21 genes) encode proteins of unknown function.

Molecular population genetics of putative rapidly-evolving *Arabidopsis* genes

It is not clear whether the genes of $\omega > 1$ in the EST analysis actually represent loci that have increased selection intensities. Indeed, the high ω estimates for these genes may simply represent sampling error. In an effort to confirm the roles of positive or diversifying selection in the evolution of putative rapidly-evolving genes, we examined the levels and patterns of nucleotide variation at these loci within and between species (Schmid and Aquadro 2001; McDonald and Kreitman 1991). Molecular population genetic analysis provides methods of selection analysis that determine whether genes are evolving according to the predictions of the neutral theory or have been subject to adaptive selection (Nielsen 2001). Analysis of within-species nucleotide diversity in *A. thaliana* was undertaken for seven genes identified from the evolutionary EST analysis as having ω^* greater than 1 (see Table 3). These genes have a range of ω^* from 1.28 to 2.03, and include the *Arabidopsis NAC2* transcriptional activator (Xie et al. 2000), and genes homologous to the yeast *MSP1* locus (Nakai et al. 1993) and the human *p55.11* protein-coding gene (Boldin, Mett and Wallach 1995). Four others are hypothetical or putative proteins predicted in the *A. thaliana* genome annotation, and represent genes of unknown function. All these genes are found in the *A. thaliana* EST sequence database, indicating that they are expressed in the developing plant. The genome annotation of the correct reading frame for one gene (AT2G04410) is ambiguous; we have relied on the original

genome annotation to identify the reading frame for this locus and this choice does not significantly affect our results.

Alleles of each of these genes were isolated and sequenced from 10 - 13 ecotypes in *A. thaliana* and 2-5 individuals from a Russian population of *A. lyrata*. The latter sample was included to provide an interspecific comparison; the sample sizes from the *A. lyrata* population were too small to permit a meaningful assessment of diversity for these genes in this species. The sequenced portions range in size from approximately 0.4 to 1.6 kb and include exon sequences that encompass the protein-coding regions that display the high ω* values observed in the evolutionary EST analysis. Levels of within-species nucleotide diversity at silent sites, π, for these seven putative rapidly-evolving genes ranged from 0.0003 to 0.0140 in *A thaliana*, with a mean of 0.0052 ± 0.0017 (see Table 3). This is comparable to the mean of ~ 0.007 observed for other previously studied *A. thaliana* nuclear genes (Miyashita et al. 1998; Purugganan and Suddith 1999; Aguade 2001; Olsen et al. 2002). The mean value for silent site θ is 0.0061 ± 0.0018.

The mean ω estimates for the regions sequenced in this population genetic survey are all, except for the unknown gene AT2G04410, less than 1 (see Table 3). All of the ω values, however, except for the *MSP1*-like locus, are greater than the mean for *A. thaliana* genes. This suggests that the increase in evolutionary rates observed may be specific to certain gene regions and do not encompass the entire protein-coding region.

Table 3. Nucleotide variation levels of rapidly-evolving *Arabidopsis thaliana* genes

Gene	Gene ID[a]	function[b]	Ka*/Ks*[c]	length (bps)	n	π (silent)	θ (silent)
1	AT3G22060	hypothetical	0.30	917	13	0.0067	0.0086
2	AT4G28000	*MSP1*-like	0.09	1572	11	0.0067	0.0086
3	AT5G04410	*NAC2*	0.43	1574	10	0.0021	0.0024
4	AT4G28470	*p55.11*-like	0.30	363	10	0.0038	0.0047
5	AT1G67140	hypothetical	0.60	1470	11	0.0003	0.0005
6	AT4G15950	putative	0.42	992	12	0.0140	0.0146
7	AT2G04410	unknown	2.62	1011	13	0.0029	0.0036

[a]*Arabidopsis* genome sequence gene reference number, [b]homologies or functional classifications, [c]Ka*/Ks* values estimated for the entire sequenced region in the sequence variation analysis.

Elevated levels of fixed replacement differences among rapidly-evolving *Arabidopsis* genes

The relative levels of within- to between-species polymorphisms in nucleotide sites that encode a gene's products provide information on the selective forces that act in protein-coding regions (McDonald and Kreitman 1991; Bustamante et al. 2002). Levels of within-species replacement and synonymous polymorphisms as well as fixed differences between *A. thaliana* and *A. lyrata* in seven putative rapidly-evolving genes are shown in Table 4. The levels of evolutionary change observed for these seven putative rapidly-evolving loci can be compared with the levels and patterns of nucleotide variation observed among 12 other previously studied *Arabidopsis thaliana* genes (Bustamante et al. 2002). In this study, the latter genes represent a set of typical *Arabidopsis* nuclear loci.

Table 4. Replacement and synonymous changes at rapidly-evolving *A. thaliana* genes

Gene	EST clone ID	Gene ID[a]	Length (codons)	Polymorphisms		Fixed differences	
				R	S	R	S
1	P1WA1-H06	AT3G22060	200	6	4	19	24
2	P1WA2-D03	AT4G28000	316	1	8	6	21
3	P1WB1-C01	AT5G04410	341	1	0	27	22
4	P1WB1-D02	AT4G28470	117	0	2	3	5
5	P2WB1-E05	AT1G67140	319	8	0	23	18
6	P3WA1-G10	AT4G15950	142	7	8	12	8
7	P3WB2-A11	AT2G04410	84	2	0	5	2

[a]*Arabidopsis* genome sequence gene reference number

The posterior distributions of the interspecies divergence time, *t*, between *A. thaliana* and *A. lyrata* are comparable between the two gene classes (for the method, see Appendix A). For the typical gene class, the mean of the posterior distribution for t_1 equaled 8.6 in multiples of twice the effective population size with 95% highest posterior CI of $[6.9 \leq t_1 \leq 11.2]$. Using data for the putative rapidly-evolving genes only, the posterior distribution of t_2 has mean 9.5 and 95% HPCI of $[7.0 \leq t_2 \leq 13.9]$. Using all of the data, we get 95% HPCI of $[7.50 \leq t \leq 11.19]$ for *t* with the posterior distribution having a mean of 9.22. The similarity in interspecies divergence time estimates suggests that the data from both gene classes compare loci of similar divergence times, and are thus likely orthologous, and not paralogous, between *A. thaliana* and *A. lyrata*.

As expected, there is a significant elevation in the levels of fixed replacement and synonymous differences between *A. thaliana* and *A. lyrata* in these seven putative rapidly-evolving genes. Among these loci, a total of 95 of the 195 coding region differences between these two species (~49%) result in amino acid replacements in the encoded proteins. In contrast, only 123 of 373 fixed differences (~33%) in typical *Arabidopsis* nuclear genes are replacement differences. The contrast in relative levels of replacement to synonymous fixed differences between these two gene classes is significant (Fisher's Exact Test, $P < 0.017$).

By comparison, the relative levels of within-species replacement to synonymous nucleotide polymorphisms within *A. thaliana* do not differ significantly between the putative rapidly-evolving loci and typical nuclear genes. Among the

genes in this study, 25 of the 47 intraspecific coding region polymorphisms (~53%) are replacement polymorphisms. Among 12 previously studied *A. thaliana* nuclear genes, 108 of 212 polymorphisms (~51%) are replacement changes. The relative levels of within-species replacement to synonymous site polymorphisms are not significantly different between the two gene classes (Fisher's Exact Test, $P < 0.891$). These results indicate that while the putative rapidly-evolving loci have increased levels of fixed replacement differences this is not accompanied by a significant increase in relative levels of intraspecific replacement polymorphisms.

Putative rapidly-evolving genes display elevated selection intensities in protein-coding regions

Selection in a specific protein-coding gene is conventionally detected in a test of homogeneity (the McDonald-Kreitman test) that examines within- and between species replacement to synonymous nucleotide changes (McDonald and Kreitman 1991). Despite the over-all increase in between-species fixed replacement differences between *A. thaliana* and *A. lyrata* in these seven putative rapidly-evolving genes, none of these individual genes show evidence of positive selection (Fisher's Exact Tests, $P < 0.10$ to 1.00).

Although none of these individual genes show evidence of selection, each gene contains information regarding the selective forces that act on amino acid replacements (Bustamante et al. 2002). Using the cell entries from a conventional McDonald-Kreitman contingency table, it is possible to estimate the four parameters

in a Poisson Random Field model (θ for silent sites, θ for replacement sites, t for interspecies divergence time, and γ for replacement sites selection coefficient) [Bustamante et al. 2002]. For a set of McDonald-Kreitman-type tables from the same species pairs, we can also model variation in selection among genes. This information can be analyzed using a hierarchical Bayesian framework to describe the probability distribution of the selection intensity, γ, for each individual *Arabidopsis* gene (Bustamante et al. 2002; see Appendix A). These selection intensities can be considered as the relative levels of selection on amino acid replacements with respect to synonymous site changes. Variation in selection among genes is modeled as normally distributed with unknown mean, μ, and variance, σ^2, for both the putative rapidly-evolving and typical gene classes.

The Markov chain Monte Carlo (MCMC) sampling scheme described in Appendix A was used to draw from the joint posterior probability distribution of several parameters in the model given the data in the McDonald-Kreitman tables for all 19 genes. These include the mean and variance parameters for typical (μ_1, σ_1) and putative rapidly-evolving loci (μ_2, σ_2), the scaled species divergence parameter (t), the vectors of mutation rates at silent sites (θ^S) and replacement sites (θ^R), and the vector of selection coefficients (γ). At completion of the sampling scheme, we have 10,000 quasi-independent vectors for each parameter in the model drawn from the joint probability distribution of the parameters given the data.

The distribution of the sampled values of γ for each locus ($\gamma_i^{(1)}$, $\gamma_i^{(2)}$,..., $\gamma_i^{(10,000)}$ for $1 \leq i \leq 19$) within and between the two classes of *Arabidopsis* genes provide several striking results. The means of the γ draws for each gene in the class of typical *Arabidopsis* loci ranged from -2.281 to +0.874. Six of these loci have 95% highest posterior probability credibility intervals (HPCI) that are entirely below 0 (see Figure 3). These negative selection intensities suggest that most amino acid replacements are slightly deleterious and persist due to the inbreeding associated with the predominant selfing observed in this species (Bustamante et al. 2002). In contrast, the means of the γ samples for the putative rapidly evolving loci ranged from -0.813 to 0.968. Three of the seven putative rapidly evolving genes (43%) have γ below 0. Only one of these genes, which encodes a protein of unknown function, has a 95% HPCI below 0. Four genes have γ greater than 0, although the 95% HPCI for all of these also encompass 0.

The posterior distribution of μ, the average selective effect of amino acid replacement changes, also shows a significant difference between putative rapidly evolving loci and typical *Arabidopsis* genes (see Figure 4.). The previously studied *Arabidopsis* loci have posterior mean for the average selective effect of amino acid replacements (μ_1) of -0.9733. We find that the posterior probability that μ_1 is greater than 0, $P(\mu_1 \geq 0)$, is 0.03. In contrast, the posterior distribution for μ_2 (the average selective effect of amino acid replacement in the putative rapidly-evolving genes) has a mean of +0.2539 and $P(\mu_2 \geq 0)$ is approximately 0.67. Lastly, the probability that the difference between average selection coefficients of typical and putative rapidly-

evolving nuclear genes (μ_1 - μ_2) is less than zero, P(μ_1 - $\mu_2 \leq 0$) is 0.97. This analysis indicates that amino acid replacement changes in these putative rapidly-evolving genes in *A. thaliana* are more beneficial (or less deleterious) than those found in typical nuclear loci.

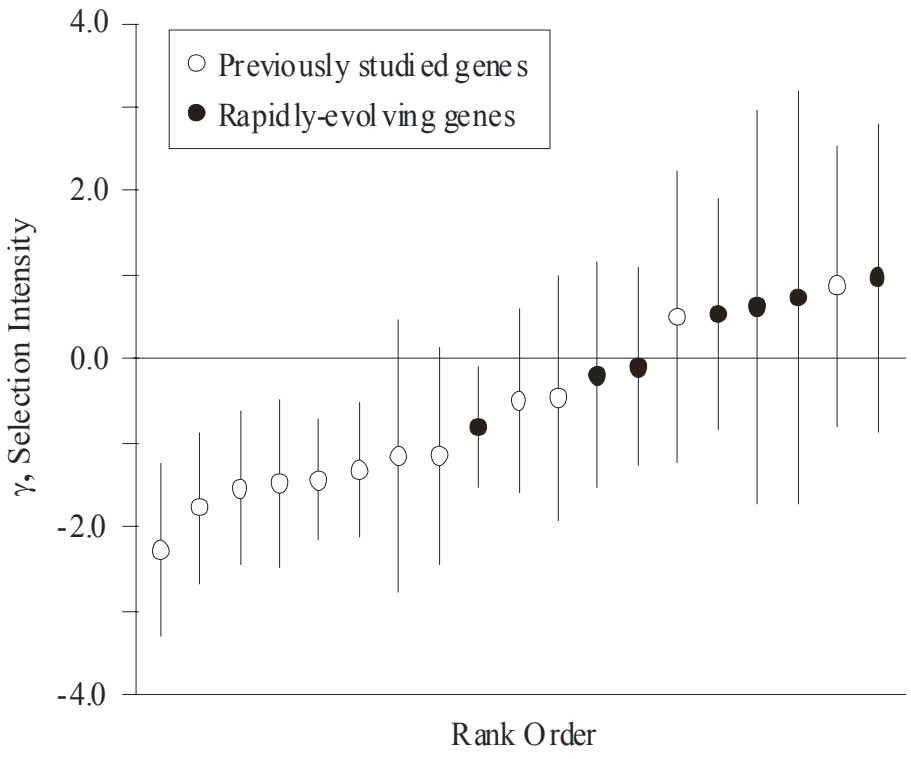

Figure 3. Selection intensities, γ, of *A. thaliana* nuclear genes. The open and closed circles are γ values for typical (24) and rapidly-evolving *A. thaliana* genes, respectively. The bars indicate the 95% highest posterior credible intervals of γ estimates. The genes from left to right are *AP3, PgiC, PI, AP1, ChiA, CAL, TFL1, FAH1,* hypothetical gene (AT1G67140), *Adh1, F3H,* unknown gene (AT2G04410), putative gene (AT4G15950), *LFY,* hypothetical gene (AT3G22060), *MSP1*-like gene (AT4G28000), *p55.11*-like gene (AT4G28470), *CHI,* and *NAC2.*

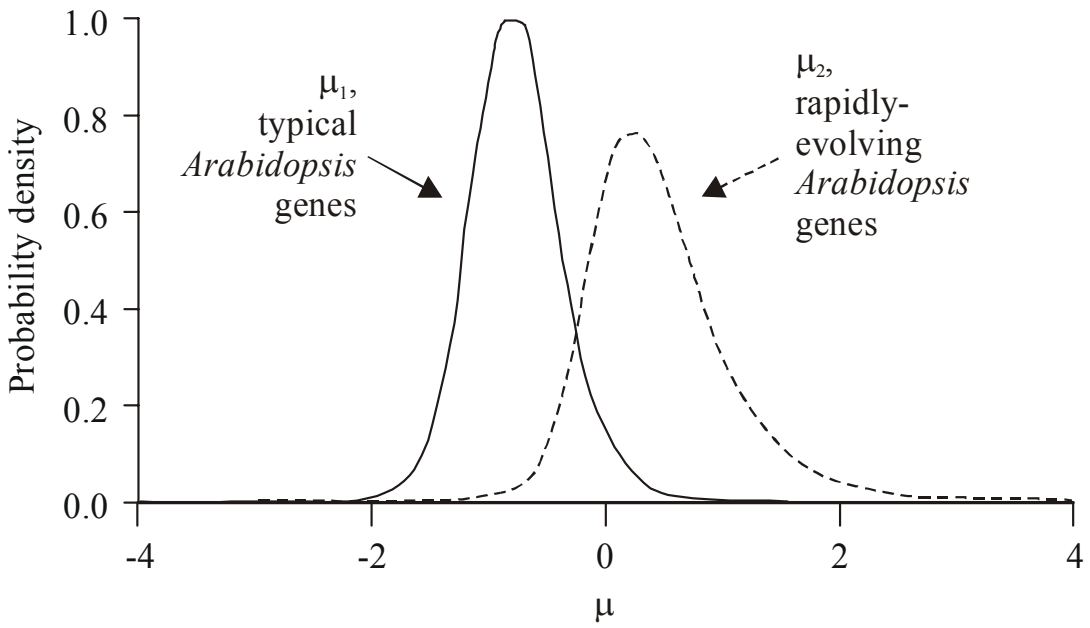

Figure 4. Distribution of selection coefficients for *A. thaliana* nuclear genes. The μ_1 (solid line) and μ_2 (dotted line) are the selection coefficient distributions for typical and rapidly-evolving *A. thaliana* loci.

Evolution of rapidly-evolving genes in the genome

Approximately 7% of the inflorescence expressed genes examined in this evolutionary study have values of ω* greater than 1 between *A. thaliana* and *A. lyrata* orthologues and are potential candidates for genes associated with adaptive divergence between these two species. This compares favorably with the 11% of loci with $\omega > 1$ observed for the *Drosophila* male accessory gland ESTs (Swanson et al. 2001); together, this suggests that rapidly-evolving protein coding loci may represent a significant portion of genes in eukaryotic genomes.

Molecular population genetic analysis confirms the increased selection intensities associated with genes that display accelerated rates of nonsynonymous evolution. In typical *A. thaliana* nuclear genes, most replacement changes are slightly deleterious and their estimated selection intensities are generally negative (Bustamante et al. 2002). Many *A. thaliana* nuclear loci studied to date possess high levels of within-species replacement polymorphisms (Purugganan and Suddith 1999), few of which go to fixation and contribute to differences between *A. thaliana* and *A. lyrata*. In contrast, the class of putative rapidly-evolving genes that were identified in this evolutionary EST study as having accelerated rates of nonsynonymous evolution generally possesses higher selection intensities on amino acid replacements. This is evident in the shift of the distribution of selection coefficients, μ, towards positive values compared to the distribution of typical *A. thaliana* genes (see Figure 4). The increase in selection intensities for putative rapidly-evolving genes may arise either from (i) neutral evolutionary forces on replacement polymorphisms, leading to

increased fixation of amino acid changes, or (ii) selection that leads to adaptive fixation of protein sequence variants. It is unlikely that neutral evolution alone can explain the increase in selection intensities on these putative rapidly-evolving loci. All the putative rapidly-evolving genes are expressed in both species and there are no premature stop codons in these loci that may be associated with pseudogene formation.

While several loci with $\omega* > 1$ identified in this evolutionary EST study may be associated with adaptive divergence, this underestimates the role of positive or diversifying selection in shaping gene structure and function in the genome. The criteria of $\omega* > 1$ as an indicator of selection can be overly stringent, as it requires that amino acid fixations occur throughout the gene and does not recognize adaptive fixation of small numbers of replacement changes. Moreover, it does not identify genes in which the selective force acts on regulatory regions of the gene, which is believed to be a major factor in adaptive divergence between species (Doebley and Lukens 1998; Sucena and Stern 2000). Nevertheless, evolutionary ESTs provide a general genomic approach to identify at least some of the loci associated with adaptation. It will also be intriguing to examine the evolutionary genetics of these loci not only in the highly selfing *A. thaliana* but also in its outcrossing relative *A. lyrata*

The function of these putative rapidly-evolving genes remains largely unknown. Of the four genes in the molecular population genetic analysis that have selection intensities greater than 0, one encodes a previously unknown protein while the other three are homologous to genes in *A. thaliana* or other eukaryotic organisms.

One gene is *NAC2*, which belongs to a family of transcription factors required in *Arabidopsis* development (Xie et al. 2000). Another locus encodes a protein homologous to the yeast *MSP1* ATPase (Nakai et al. 1993), while the fourth gene encodes a protein homologous to the human *p55.11* protein that binds to the tumor necrosis factor *p55* receptor (Boldin, Mett and Wallach 1995). The precise functions of these genes in *Arabidopsis* remain to be elucidated. Expression studies as well as phenotypic analysis of T-DNA insertion mutants for these genes may permit an assessment of their functions, and may also provide clues as to the traits that are the targets of selection in the divergence between *A. thaliana* and *A. lyrata*.

Acknowledgements

The authors would like to thank the NCSU Genome Research Laboratory for use of their facilities and the Purugganan laboratory for a critical reading of the manuscript. M. B. was funded by an NIH Training Grant on the Genetics of Complex Traits and C. D. B. by a Marshall-Sherfield Fellowship from the Marshall Commemoration Commission. This work was funded in part by grants from the National Science Foundation Integrated Research Challenges in Environmental Biology program M. D. P, J. I. Schmitt and T. F. C. Mackay and the Alfred P. Sloan Foundation to M. D. P.

Literature Cited

Aguade, M. (2001) Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the *FAH1* and *F3H* genes, in *Arabidopsis thaliana. Mol. Biol. Evol.* **18**: 1-9.

Anisimova, M., Bielawski, J. P. and Yang, Z. H. (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* **18**: 1585-1591.

The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana. Nature* **408**: 796-815.

Boldin, M. P., Mett, I. L. and Wallach, D. (1995) A protein related to a proteasomal subunit binds to the intracellular domain of the *p55 TNF* receptor upstream to its 'death domain'. *FEBS Lett.* **367**: 39-44.

Bustamante, C. D., Nielsen, R. and Hartl, D. L. (2002) A maximum likelihood method for analyzing pseudogene evolution: implications for silent site evolution in humans and rodents. *Mol. Biol. Evol.* **19**: 110-117.

Bustamante, C. D., Nielsen, R., Sawyer, S., Olsen, K. M., Purugganan, M. D. and Hartl, D. L. (2002) The cost of inbreeding in *Arabidopsis. Nature* **416**: 531-534.

Doebley, J. and Lukens, L. (1998) Transcriptional regulators and the evolution of plant form. *Plant Cell* **10**: 1075-1082.

Ewing, B., Hillier, L., Wendl, M. C. and Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**: 175-185.

Gelman, A., Carlin, J. S., Stern, H. S. and Rubin, D. B. (1997) *Bayesian Data Analysis* Chapman-Hall, Boca Raton.

Haag, E. S. and True, J. R. (2001) Perspective: From mutants to mechanisms? Assessing the candidate gene paradigm in evolutionary biology. *Evolution* **55**: 1077-1084.

Hughes, A. L. (1991) Circumsporozoite protein genes of malaria parasites (*Plasmodium* ssp.): evidence for positive selection on immunogenic regions. *Genetics* **127**: 345-353.

Hughes, A. L. and Yeager, M. (1998) Natural selection at major histocompatibility complex loci of vertebrates. *Annu. Rev. Genetics* **32**: 415-435.

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

Kusaba, M. et al. (2001) Self-incompatibility in the genus *Arabidopsis*: characterization of the S locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana.* *Plant Cell* **13**: 627-643.

Li, W. -H., Gojobori, T. and Nei, M. (1981) Pseudogenes as a paradigm of neutral evolution. *Nature* **292**: 237-239.

Li, W. -H., Wu, C. -I. and Luo, C. -C. (1985) A new method for estimating synonymous and nonsynonymous rates on nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* **2**: 150-174.

McDonald, J. and Kreitman, M. (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila.* *Nature* **351**: 652-654.

Messier, W. and Stewart, C. B. (1997) Episodic adaptive evolution of primate lysozymes. *Nature* **385**: 151-154.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953) Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**: 1087-1092.

Miyashita, N. T., Kawabe, A., Innan, H. and Terauchi, R. (1998) Intra- and Interspecific DNA Variation and Codon Bias of the *Alcohol Dehydrogenase (Adh)* Locus in *Arabis* and *Arabidopsis* species. *Mol. Biol. Evol.* **15**: 1420-1429.

Nakai, M., Endo, T., Hase, T. and Matsubara, H. (1993) Intramitochondrial protein sorting. Isolation and characterization of the yeast *MSP1* gene which belongs to a novel family of putative ATPases. *J. Biol. Chem.* **268**: 24262-24269.

Nei, M. (1987) *Molecular Evolutionary Genetics*, Columbia University Press, New York.

Nei, M. and Gojobori, T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**: 418-426.

Nielsen, R. (2001) Statistical tests of selective neutrality in the age of genomics. *Heredity* **86**: 641-647.

Olsen, K. M., Womack, A., Garrett, A. R., Suddith, J. I. and Purugganan, M. D. (2002) Contrasting evolutionary forces in the *Arabidopsis thaliana* floral developmental pathway. *Genetics* **160**: 1641-1650.

Orr, H. A. and Coyne, J. A. (1992) The genetics of adaptation: a reassessment. *Am. Nat.* **140**: 725-742.

Purugganan, M. D., Rounsley, S. D., Schmidt, R. J. and Yanofsky, M. F. (1995) Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family. *Genetics* **140**: 345-356.

Purugganan, M. D. and Suddith, J. I. (1999) Molecular population genetics of floral homeotic loci: Departures from the equilibrium-neutral model at the *AP3* and *PI* genes of *Arabidopsis thaliana.* *Genetics* **151**: 839-848.

Rozas, J. and Rozas, R. (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinf.* **15**: 174-175.

Sawyer, S. A. and Hartl, D. L. (1992) Population genetics of polymorphism and divergence. *Genetics* **132**: 1161-1176.

Schmid, K. J. and Aquadro, C. F. (2001) The evolutionary analysis of "orphans" from the *Drosophila* genome identifies rapidly diverging and incorrectly annotated genes. *Genetics* **159**: 589-598.

Sucena, E. and Stern, D. L. (2000) Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by cis-regulatory evolution of ovo/shaven-baby. *Proc. Natl. Acad. Sci. USA* **97**: 4530-4534.

Swanson, W. J., Clark, A. G., Waldrip-Dail, H. M., Wolfner, M. F. and Aquadro, C. F. (2001) Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **98**: 7375-7379.

Swanson, W. J. and Vacquier, V. D. (1995) Extraordinary divergence and positive Darwinian selection in a fusagenic protein coating the acrosomal process of abalone spermatozoa. *Proc. Natl. Acad. Sci. USA* **92**: 4957-4961.

Watterson, G. A. (1975) On the number of segregating sites in genetical models without recombination. *Theor. Pop. Biol.* **7**: 256-276.

Wu, C. -I. (2001) The genic view of the process of speciation. *J. Evol. Biol.* **14**: 851-865.

Xie, Q., Frugis, G., Colgan, D. and Chua, N. H. (2000) *Arabidopsis NAC1* transduces auxin signal downstream of *TIR1* to promote lateral root development. *Genes and Dev.* **14**: 3024-3036.

CHAPTER 5.

Summary

Hawaiian Silversword Alliance Ancestry

Polyploidy was once believed to be a rare event and an evolutionary dead-end, or at least a hindrance to successful evolution in higher plants (Stebbins 1971). More recent research has revealed that polyploidy is not only common in plants, but has occurred at some point in the evolutionary history of most angiosperms (Wendel 2000). Examples of recent and ancient polyploids include many important crop species such as sugarcane, potatoes, soybean, wheat and corn (Wendel 2000).

The nature of the polyploid Hawaiian silversword alliance (Asteraceae) genome was clarified in this dissertation research. Although it had been determined in previous studies that the alliance species were ancient polyploids, it was not clear whether this was due to an allopolyploid or autopolyploid event. Through biosystematic, cytogenetic, and molecular studies, the ancestry of the Hawaiian silversword alliance had been narrowed to a subgroup of the North American tarweeds.

A more specific evolutionary history of the Hawaiian silversword alliance was uncovered in this dissertation research through molecular evolutionary analysis of two regulatory genes. Two distinct copies of the floral homeotic genes, *ASAP3/TM6* and *ASAP1*, were found in the polyploid alliance members while single copies were found in the diploid North American tarweeds. Through phylogenetic analysis of these genes, the duplicate copies in the Hawaiian silversword alliance species were determined to be the result of an interspecific hybridization between two lineages of

the North American tarweeds. This narrowed down the North American lineages of the hybrid progenitors to a specific group of the continental tarweeds containing *Carlquistia muirii* (and perhaps *Harmonia nutans)* and *Anisocarpus scabridus*. Data from both regulatory genes confirm the allopolyploid nature of the alliance ancestor. The relationships are also consistent with previous phylogenetic studies of the Hawaiian silversword alliance and North American tarweed species using cpDNA and rDNA sequences (Baldwin et al. 1991, Baldwin 1992).

Knowledge of this ancestral polyploidization event strengthens previous research and opens up further avenues of study into the phylogenetics of the Madiinae. The evolutionary effects of polyploidy are still unclear, but studies of synthetic allopolyploids in wheat (Feldman et al. 1997; Liu et al. 1998; Liu, Vega and Feldman 1998) and *Arabidopsis* (Song et al. 1995) revealed significant changes in the genome sequence and structure in the first few generations. It has long been thought that the added genomic diversity of allopolyploidization could lead to evolutionary innovation (Wendel 2000). These observations suggest that this added genomic diversity may be associated with the spectacular adaptive radiation of the Hawaiian silversword alliance.

Regulatory Gene Evolution in the Hawaiian Silversword Alliance

Previous studies of rapid morphological evolution failed to find correlated rates of molecular evolution at structural genes (King and Wilson 1975). This dissertation research examined regulatory genes in the adaptively-radiating Hawaiian silversword alliance species, which exhibit significant morphological diversity, in order to identify any correlated increase in the molecular evolutionary rates of regulatory genes.

Two of the genes studied, *ASAP3/TM6* and *ASAP1*, are involved in the regulatory pathways of flower development in *Arabidopsis*. The structural gene, *ASCAB9* was also included for comparison. These three loci were examined in a collection of species from the Hawaiian silversword alliance as well as from members of the ancestral North American tarweeds, which do not display such diverse morphologies.

The regulatory genes did exhibit significant increases in nonsynonymous (Ka) relative to synonymous (Ks) nucleotide substitution rates in members of the Hawaiian silversword alliance compared to rates in the North American tarweeds. This was not due to a general increase in the neutral mutation rates at these loci in the alliance species, suggesting that these genes may be under positive or diversifying selection. The structural gene *ASCAB9* displayed a small increase in Ka relative to Ks in the alliance species, but not to the same extent observed in the regulatory genes.

The analyses suggest that rapid morphological adaptations may be associated more with accelerated rates of evolution of regulatory genes rather than of structural genes. An expanded study with more regulatory and structural genes is now needed to determine if this pattern will hold true for regulatory and structural genes throughout the genome. This can be the basis for a more in-depth look at the function and expression of the fast-evolving regulatory genes and the more placid structural gene in populations of the Hawaiian silversword alliance and North American tarweed species.

Identifying Fast-Evolving Genes in *Arabidopsis* With Evolutionary ESTs

Genes involved in adaptive divergence between species often display evidence of positive or diversifying selection. This dissertation research proposed to identify genes which may be involved with adaptations in *Arabidopsis thaliana* by looking for evidence of strong positive selection.

In order to do this, a cDNA library was created with flowering material from *Arabidopsis lyrata*, a species closely-related to *A. thaliana.* This library contains a collection of the genes expressed in the flowering tissue of this species. ESTs from a subset of these *A. lyrata* genes were sequenced. Homologous sequences from the *A. thaliana* online database were identified and nonsynonymous (Ka) and synonymous (Ks) rates of nucleotide substitution between these protein-coding sequences were estimated.

Examining the ratio of Ka to Ks allowed the identification of potentially fast-evolving genes as those with a Ka/Ks ratio greater than one. From the evolutionary EST approach, a set of 21 putative rapidly-evolving genes in *Arabidopsis* were identified out of the 304 unique genes isolated. According to the descriptions of the *A. thaliana* homologs from the online database, these putatively fast-evolving genes cover a variety of gene functions.

Longer sequence stretches of seven of these putative fast-evolving genes were then isolated in a set of *A. thaliana* ecotypes. This was done in order to verify the divergence between species as well as to examine the polymorphism within *A.*

thaliana. These seven putative rapidly-evolving loci did not exhibit increased levels of replacement polymorphisms within *A. thaliana* despite their high levels of nonsynonymous divergence between *A. thaliana* and *A. lyrata.* The intensity of selection for these loci was measured using a hierarchical Bayesian analysis of protein-coding region evolution within and between species. These selection intensities were elevated in the putative fast-evolving genes when compared to a set of typical *A. thaliana* nuclear genes.

This evolutionary EST method is powerful because it takes advantage of the available complete genome sequences. The potential applications will continue to grow as more organisms are fully sequenced. Comparing closely related species, like *A. thaliana* and *A. lyrata,* in this way can help with gene identification and characterization. This method could also be used for other types of comparisons of expressed genes such as between multiple tissue types, developmental stages, or even between individuals.

Summary

In this dissertation research, the evolution of rapidly-evolving genes undergoing diversifying selection was examined. Identifying the types of genes undergoing rapid evolution due to selection is an important step towards understanding the mechanisms involved in adaptation and speciation. In this research, rapid regulatory-gene evolution was observed in the adaptively radiating Hawaiian silversword alliance. This also led to the discovery of ancestral relationships between specific species of the North American tarweeds and the Hawaiian silversword alliance. Rapidly-evolving genes can sometimes be useful in elucidating evolutionary relationships between species. A group of potential rapidly-evolving genes were identified in *Arabidopsis* using a powerful evolutionary EST approach. This method could be very useful in the identification of adaptively important genes in any of the species whose genomes are fully sequenced. As more genome sequences become available, rapidly-evolving genes will be able to provide increasing insight into the molecular evolution of adaptation.

Literature Cited

Baldwin, B. G. (1992) Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae, *Mol. Phylogenet. Evol.* **1**: 3-16.

Baldwin, B. G., Kyhos, D. W., Dvorak, J. and Carr, G. D. (1991) Chloroplast DNA evidence for a North American origin of the Hawaiian silversword alliance (Asteraceae). *Proc. Natl. Acad. Sci. USA* **88**: 1840-1843.

Cherry, L. M., Case, S. M. and Wilson, A. C. (1978) Frog perspective on morphological difference between humans and chimpanzees. *Science* **200**: 209-211.

Feldman, M., Liu, B., Segal, G., Abbo, S., Levy, A. A. and Vega, J. M. (1997) Rapid elimination of low-copy DNA sequences in polyploidy wheat: a possible mechanism for differentiation of homoeologous chromosomes. *Genetics* **147**: 1381-1387.

King, M. C. and Wilson, A. C. (1975) Evolution at 2 levels in humans and chimpanzees. *Science* **188**: 107-116.

Liu, B., Vega, J. M. and Feldman, M. (1998) Rapid genomic changes in newly synthesized amphiploids of *Trilicum* and *Aegilops*. II. Changes in low-copy coding DNA sequences. *Genome* **41**: 535-542.

Liu, B., Vega, J. M., Segal, G., Abbo, S., Rodova, M. and Feldman, M. (1998) Rapid genomic changes in newly synthesized amphiploids of *Trilicum* and *Aegilops*. I. Changes in low-copy non-coding DNA sequences. *Genome* **41**: 272-277.

Song, K. M., Lu, P., Tang, K. L. and Osborn, T. C. (1995) Rapid genomic change in synthetic polyploids of *Brassica* and its implications for polyploidy evolution. *Proc. Natl. Acad. Sci. USA* **92**: 7719-7723.

Stebbins, G. L. (1971) Chromosome evolution in higher plants. Arnold, London.

Wendel, J. F. (2000) Genome evolution in polyploids. *Plant Mol. Biol.* **42**: 225-249.

APPENDICES

APPENDIX A

Supplemental Information for Chapter 4

Written by Carlos D. Bustamante

Sawyer and Hartl (1992) show that under the assumptions of the Poisson Random Field setting, the sampling distributions for the number of polymorphic sites within species, *S*, and fixed differences between two species, *K*, for samples of size *n* from one and *m* from the other are independent Poisson distributed random variables with rates

$$E(S) = \theta \frac{2\gamma}{1 - e^{-2\gamma}} \left(F(\gamma, m) + F(\gamma, n) \right) \quad (1)$$

$$E(S) = \theta \frac{2\gamma}{1 - e^{-2\gamma}} \left(G(\gamma, m) + G(\gamma, n) + t \right) \quad (2)$$

where γ is the scaled selection coefficient for new mutations ($2N_e \cdot s$), θ is the scaled mutation rate ($4N_e \mu$), t is the scaled time since *species* divergence (# of generations since divergence / $2N_e$),and N_e is the effective population size. The functions $F(\gamma, n)$ and $G(\gamma, n)$ are given in (Bustamante et al. 2002) and (Sawyer and Hartl 1992).

Using the cell entries from a conventional McDonald-Kreitman (1991) table, it is possible to estimate four parameters in such a model: θ^S (mutation rate for silent sites), θ^R (mutation rate for replacement sites), t, and γ (selection intensity for replacement sites) assuming silent sites are neutral (i.e., $\gamma = 0$ for all silent sites). For a set of such tables from the same species pairs, it also possible to model variation in

selection among genes by specifying a distribution for γ and estimating the parameters of this hyper-distribution given the data in all of the tables. A convenient form to use is the normal distribution since selection coefficients can be either positive or negative. It should also be noted that the species divergence time is a shared parameter across all the tables in such an analysis.

The analysis we present is based on a description of the joint and marginal posterior probability distributions of the following model:

Hierarchical model

1. Let γ be the vector of selection coefficients, with $\gamma_1, \ldots, \gamma_{12}$ being the set of typical loci (24) and $\gamma_{13}, \ldots, \gamma_{19}$ representing rapidly-evolving loci. θ^R and θ^S are the corresponding vectors of mutation rates at replacement and silent sites. Denote the mean and variance of the distribution of γ among typical genes as μ_1 and σ_1^2 and use μ_2 and σ_2^2 for the analogous quantities for the rapidly evolving genes.

2. Set a truncated uniform prior distribution for t on (0, T), where T is chosen based on prior information on the upper bound for the species divergence time.[1]

3. Assume a normal conjugate prior probability distribution for the mean and variance parameters for each of the two classes of genes so that (borrowing notation from Gelman, Carlin, Stern and Rubin 1997, p.72).

$$\mu_i \mid \sigma_i^2 \sim N(\mu_0, \frac{\sigma^2}{\kappa_0}) \qquad (3)$$

$$\sigma_i^2 \sim Inv - \chi^2(\nu_0, \sigma_0^2) \qquad (4)$$

[1] We used T = 100, corresponding to an upper bound of between 20 and 200 million years ago.

where μ_0, κ_0, ν_0, σ_0^2, are parameters of the prior distributions for μ_i and σ_i^2, Inv-χ^2 refers to an inverse χ^2 distribution, and $i \in \{1, 2\}$ indexes the two sets of hyper-parameters for both classes of genes.[2]

The joint posterior distribution of interest $p(\gamma, \theta^R, \theta^S, t, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2 \mid data)$ can be approximated using a Markov chain Monte Carlo sampling scheme similar to that implemented in Bustamante et al. (2002) using the following results:

[2] If κ_o and ν_o are chosen to be small and σ_o^2 to be large, the prior distribution will be uninformative. In our runs we used $\mu_o = 0$, $\sigma_o^2 = 100$, $\kappa_o = 0.001$, $\nu_o = 0.001$

Results for conditional posterior distributions

1. The conditional posterior distribution of $\mu_I \mid \sigma_I^2, \gamma$ depends only on the entries in γ that are members of the class i and can be shown to be normally distributed as

$$\mu_i \mid \sigma_i^2, \gamma \sim N\left(\frac{\frac{\kappa_o}{\sigma_i^2}\mu_o + \frac{J_i}{\sigma_i^2}\bar{\gamma}_i}{\frac{\kappa_0}{\sigma_i^2} + \frac{J_i}{\sigma^2}}, \frac{1}{\frac{\kappa_0}{\sigma_i^2} + \frac{J_i}{\sigma_i^2}} \right) \qquad (5)$$

where $\bar{\gamma}_i$ is the arithmetic average of the entries in γ for class i and J_i is the number of genes in the class.

2. The marginal posterior distribution of σ_i^2 conditional on γ, which depends only on the sample variance of the entries in γ that are members of the class i and the parameters of the prior distribution, has an Inverse χ^2 distribution with parameters ν_{J_i} and $\sigma_{J_i}^2$ as given in Gelman et al. (1997, p. 72).

3. Using independent Gamma prior distributions with parameters α_0 and β_0 for each of the mutation rates yields independent Gamma posterior

distributions conditional on t and γ with parameters $\alpha_0 + K + S$ and

$$\beta_0 + \frac{2\gamma}{1-e^{-2\gamma}}\left(F(\gamma,m) + F(\gamma,n) + t + G(\gamma,n) + G(\gamma,m)\right).^{3}$$

4. The posterior distribution $p(t \mid \theta^R, \theta^S, \gamma, data)$ is proportional to the likelihood function at the point $(t, \theta^R, \theta^S, \gamma)$ if $t < T$ and 0 otherwise.

5. The joint conditional posterior distribution $p(\gamma \mid \theta^R, \theta^S, t, \mu, \sigma^2, data)$ factors into the product of the individuals conditional distributions $p(\gamma_j \mid \theta_j^R, \theta_j^S, t, \mu_i, \sigma_i^2, K_j^R, S_j^R, data)$. Furthermore, the conditional posterior distribution $p(\gamma_j \mid \theta_j^R, \theta_j^S, t, \mu_i, \sigma_i^2, K_j^R, S_j^R)$ for a given gene j in class i is proportional to the product of the likelihood for the gene given $\theta_j^R, \theta_j^S, \gamma_j$, and t, and the probability density of a normal distribution with mean μ_i and variance σ_i^2, at the point γ_j.

Given the model as outlined above and the quoted results, it is then possible to sample from $p(\gamma, \theta^R, \theta^S, t, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2 \mid data)$ using the following algorithm for each chain:

[3] The mean of this distribution is α/β and the variance is α/β^2. As such, if α_0 and β_0 are chosen to be small, the prior will be uninformative. For all our analysis, we used $\alpha = \beta = 0.001$.

Markov chain Monte Carlo algorithm

1. Initialize γ by drawing a value for γ_j for $1 \leq j \leq J_1 + J_2$ independently from a normal distribution with mean near 0 and a reasonably large variance.[4]

2. Using the values in γ, update σ_i^2 for $i \in \{1, 2\}$ by sampling from the conditional distribution of $\sigma_i^2 \mid \gamma$ which is Inverse-χ^2 distributed as detailed above.

3. Using the values in γ and σ_i^2 for $i \in \{1, 2\}$, update μ_i by sampling a new value from a normal distribution with the updated parameters in Result 1 above.

4. Update t by using Metropolis sampling.

   (a) Sample a proposal value t' from a $U(t - \delta_t, t + \delta_t)$ distribution.

   (b) If $p(t' \mid \theta^R, \theta^S, \gamma, data) > p(t \mid \theta^R, \theta^S, \gamma \mid data)$, set $t = t'$. Otherwise, set $t = t'$ with probability proportional to the ratio of these two quantities.

5. Update each γ_j in γ by using $J_1 + J_2$ independent Metropolis steps as follows:

[4] We used several starting values for the mean in the range [-5, 5] and the variance in [1, 100] with no discernible effect on the resulting posterior distribution at the point we retained samples.

(a) Sample a proposal value γ'_j from a $U(\gamma_j, -\delta, \gamma_j + \delta)$.

(b) If $p(\gamma'_j \mid \theta_j^R, \theta_j^S, t, \mu_i, \sigma_i^2, S_j^R, K_j^R) > p(\gamma_j \mid \theta_j^R, \theta_j^S, t, \mu_i, \sigma_i^2, S_j^R, K_j^R)$, set $\gamma_j = \gamma'_j$ Otherwise, set $\gamma_j = \gamma'_j$ with probability proportional to their ratio.

6. For each gene, draw a value for θ_j^R and θ_j^S using the result that the posterior distribution for $\theta_j \mid \gamma_j, t$ is a Gamma distribution with parameters as described in Result 3 above.

7. Repeat steps 2 through 6.

We used the above algorithm to approximate the joint posterior distributions using 10 different starting points (i.e., 10 different chains) run for 10,000 steps each after an initial 2,000 step burn-in and retention of draws every 10 steps (for a total of 10,000 draws for each parameter in the model). For the Metropolis step for updating t, we used a proposal distribution with $\delta_t = 1.0$ which gave a rejection rate of 26.19% for the 100,000 draws retained after the initial burn-in. To measure convergence we used $\sqrt{\hat{R}}$ statistic which was below 1.02 for all parameters in the model before samples were retained (conventionally one retains after 1.2 or less), illustrating that the 10 chains had converged well before we retained our samples.

Literature Cited

Bustamante, C. D., Nielsen, R., Sawyer, S., Olsen, K. M., Purugganan, M. D. and Hartl, D. L. (2002) The cost of inbreeding in *Arabidopsis.* *Nature* **416**: 531-534.

Gelman, A., Carlin, J. S., Stern, H. S. and Rubin, D. B. (1997) *Bayesian Data Analysis* Chapman-Hall, Boca Raton.

McDonald, J. and Kreitman, M. (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila.* *Nature* **351**: 652-654.

Sawyer, S. A. and Hartl, D. L. (1992) Population genetics of polymorphism and divergence. *Genetics* **132**: 1161-1176.

APPENDIX B

Interspecific Evolution in Plant Microsatellite Structure

Marianne Barrier, Elizabeth Friar, Robert Robichaux and Michael D. Purugganan

GENE 241 (1): 101-105 (2000)

Note: Species name changes since publication are listed here old name -> new name. *Madia/Raillardiopsis* group -> "*Madia*" group; *Madia bolanderi -> Kyhosia bolanderi*; *Madia nutans -> Harmonia nutans*; *Raillardiopsis muirii -> Carlquistia muirii*; *Raillardiopsis scabrida -> Anisocarpus scabridus*

Abstract

Several intragenically linked microsatellites have been identified in the floral regulatory genes *A. sandwicense APETALA1* (*ASAP1*) and *A. sandwicense APETALA3/TM6* (*ASAP3/TM6*) in 17 species of the Hawaiian and North American Madiinae (Asteraceae). Thirty-nine microsatellite loci were observed in the introns of these two genes, suggesting that they are hotspots for microsatellite formation. The sequences of four of these microsatellites were mapped onto the phylogenies of these floral regulatory genes, and the structural evolution of these repeat loci was traced. Both nucleotide substitutions and insertion/deletion mutations may be responsible for the formation of perfect microsatellites from imperfect repeat regions (and vice versa).

Introduction

Microsatellites, or simple sequence repeats (SSRs), are short sequence elements composed of tandem repeat units one to seven base pairs (bp) in length (Tautz 1989). Microsatellites are present in high numbers in mammals and are randomly distributed with a density of approximately one microsatellite present every 10–15 kb of sequence (Tautz 1989). These simple sequence repeats are also present in plant genomes, but appear to be less abundant than in mammalian or insect systems (Van Treuren et al. 1997). Microsatellite sequences possess high mutation rates, estimated at 10^{-2}–10^{-3} per locus per gamete per generation (Tautz 1989). These repeat sequences have been shown to be highly polymorphic within and between species, a property that has permitted their application as molecular markers in population genetics (Goldstein et al. 1999), systematics (Goldstein and Pollock 1997), and genome mapping (Weissenbach et al. 1992). Microsatellite instability is also implicated in several human diseases (Kunkel 1993) and may be associated with variation in gene regulation (Meloni et al. 1998).

Relatively little is known about the origins and interspecific evolution of microsatellite loci. Documented examples of microsatellite formation suggest that a variety of mutations, including repeat number amplification (Schlotterer and Tautz 1992), simple nucleotide substitutions, and insertion/deletion events, can contribute to the formation of simple sequence repeats (Estoup et al. 1995; Messier et al. 1996). Studies on the origins and fates of sequence repeats may provide insights into

evolutionary trends in microsatellite behavior and assist investigators in identifying polymorphic markers that can be utilized across different taxa. One approach to investigating patterns of microsatellite evolution focuses on the analysis of between-species diversification of sequence repeat structures among a large number of closely related taxa.

We have isolated the floral regulatory genes *A. sandwicense APETALA1* (*ASAP1*) and *A. sandwicense APETALA3/TM6* (*ASAP3/TM6*) from members of the Hawaiian silversword alliance and their nearest North American tarweed relatives (Asteraceae: Heliantheae- Madiinae). These floral regulatory genes contain a large number of sequence repeat regions, many of which can be considered microsatellite sequences. Phylogenetic analyses of the gene sequences and mapping of the structure of the associated microsatellites provide a framework to examine interspecific evolution in plant microsatellite structure.

Materials and Methods

Gene sampling

The floral regulatory genes *ASAP3/TM6* and *ASAP1* were isolated from 10 species of the Hawaiian silversword alliance and seven species of North American tarweeds as previously described (Barrier et al. 1999). The Hawaiian species were selected to represent each of the four major lineages in the silversword alliance as previously identified from rDNA ITS trees (Baldwin and Robichaux 1995; Baldwin 1996). Four North American species (*Madia bolanderi*, *M. nutans*, *Raillardiopsis scabrida*, and *R. muirii*) were chosen to represent each of the four major lineages in the *Madia/Raillardiopsis* group as also identified from rDNA ITS trees (Baldwin 1996). Three other North American tarweed species (*Adenothamnus validus*, *Raillardella pringlei*, and *Osmadenia tenella*) are known to fall outside the clade comprising *Madia/Raillardiopsis* and the silversword alliance (Baldwin 1996) and were included to serve as the outgroups in the analyses.

The PCR primers ASAP3-2 and ASAP3-3R were designed to allow amplification of a region spanning exons 1–4 of the *ASAP3/TM6* gene; the two copies of this gene in the Hawaiian species were discriminated from one another by size (Barrier et al. 1999). For *ASAP1*, gene-specific primer pairs (ASAP1-F1A/ASAP1-RB and ASAP1-AF/ASAP1-R) were constructed to amplify sequences from exons 3 to 8 of different duplicate copies of the gene in the Hawaiian species. Only single copies of the gene were identified in the North American species. The primers were

used in PCR amplifications using the error-correcting rTth polymerase formulation (Perkin-Elmer) in a standard buffer with 40 cycles of 95°C for 1 min, 55°C for 1 min, and 72°C for 4 min. The nucleotide error rate for this formulation is less than 1 bp in 7 kb of sequence (unpublished observations). PCR-amplified DNA was cloned using the TA cloning kit (Invitrogen) and sequenced using automated sequencers (Iowa State University Sequencing Facility, NCSU DNA Sequencing Facility). Sequencing was carried out using nested primers, with multiple sequencing reactions conducted for divergent sequences. All sequence changes were rechecked visually against sequencing chromatograms, and are deposited in GenBank (Accession Nos. AF147210 to AF147258).

Evolutionary analyses

Nucleotide sequences were aligned visually. Phylogenetic analyses were conducted using maximum parsimony techniques implemented in PAUP* 4.0d54 (Swofford 1998). Both substitution and insertion/deletion (indel) differences were used and weighted equally in the analyses, with the indels separately coded (as additional characters) to reflect non-independence of continuous gaps. Identified microsatellite sequences were excluded from the analyses. Parsimony analyses were conducted using the heuristic search procedure, with random taxon addition (10 replicates), tree–bisection–reconnection branch swapping, and MULPARS in effect. Clade support was estimated by parsimony analysis of 500 bootstrap replicates of the data set using the search procedures outlined above. Microsatellite sequences were

mapped onto the phylogenies under maximum parsimony, using MacClade under ACCTRAN conditions (Maddison and Maddison 1992).

Results and Discussion

The floral regulatory genes *ASAP1* and *ASAP3/TM6* were isolated from 17 members of the Hawaiian and North American Madiinae [Asteraceae] (Barrier et al. 1999). There are two copies of *ASAP1* and *ASAP3/TM6* (*A* and *B* copies) in species of the Hawaiian silversword alliance, while only a single copy of these loci is found within members of the North American *Madia/Raillardiopsis* group. Phylogenetic analyses lead us to conclude that the duplicate copies of these two loci in the Hawaiian silversword alliance arose from an interspecific hybridization event between species within the North American *R. scabrida* and *R. muirii* lineages (Barrier et al. 1999). Likelihood estimates based on data from the rDNA internal transcribed spacer (ITS) locus suggest that the most recent common ancestor of the Hawaiian silversword alliance existed 5.2±0.8 million years ago (Ma), contemporaneous with the origin of the Island of Kaua'i (Baldwin and Sanderson 1998). In contrast, the earliest date for the diversification of the North American Madiinae appears to be in the mid-Miocene 15 Ma (Baldwin and Sanderson 1998).

Introns of the *ASAP1* and *ASAP3/TM6* genes appear to harbor a number of sequence repeat motifs containing four or more mono-, di- or trinucleotide repeats. There does not appear to be a consensus regarding the minimum number of repeats that defines a microsatellite sequence; for our purposes, we recognize a sequence motif as a microsatellite locus if it contains at least six repeat units. In more than 3 kb of common aligned sequence for the *ASAP1* locus in both Hawaiian and North

American species, at least 102 repeat regions have been identified. Twenty-five of these have at least six tandem repeats and are therefore defined as microsatellite loci (see Figure 1). The 2.7 kb of common aligned sequence for *ASAP3/TM6* yields at least 46 repeat regions, 14 of which are microsatellites (see Figure 1). Twenty-three of 25 microsatellites found in *ASAP1* and 11 of the 14 microsatellites found in *ASAP3/TM6* consist of $(A)_n$ or $(T)_n$ mononucleotide repeats. The other three microsatellites from *ASAP3/TM6* contain $(AT)_n$ or $(TC)_n$ dinucleotide repeats. The large number of repeat regions in the introns of these two loci suggests that these genes may be hotspots for microsatellite formation.

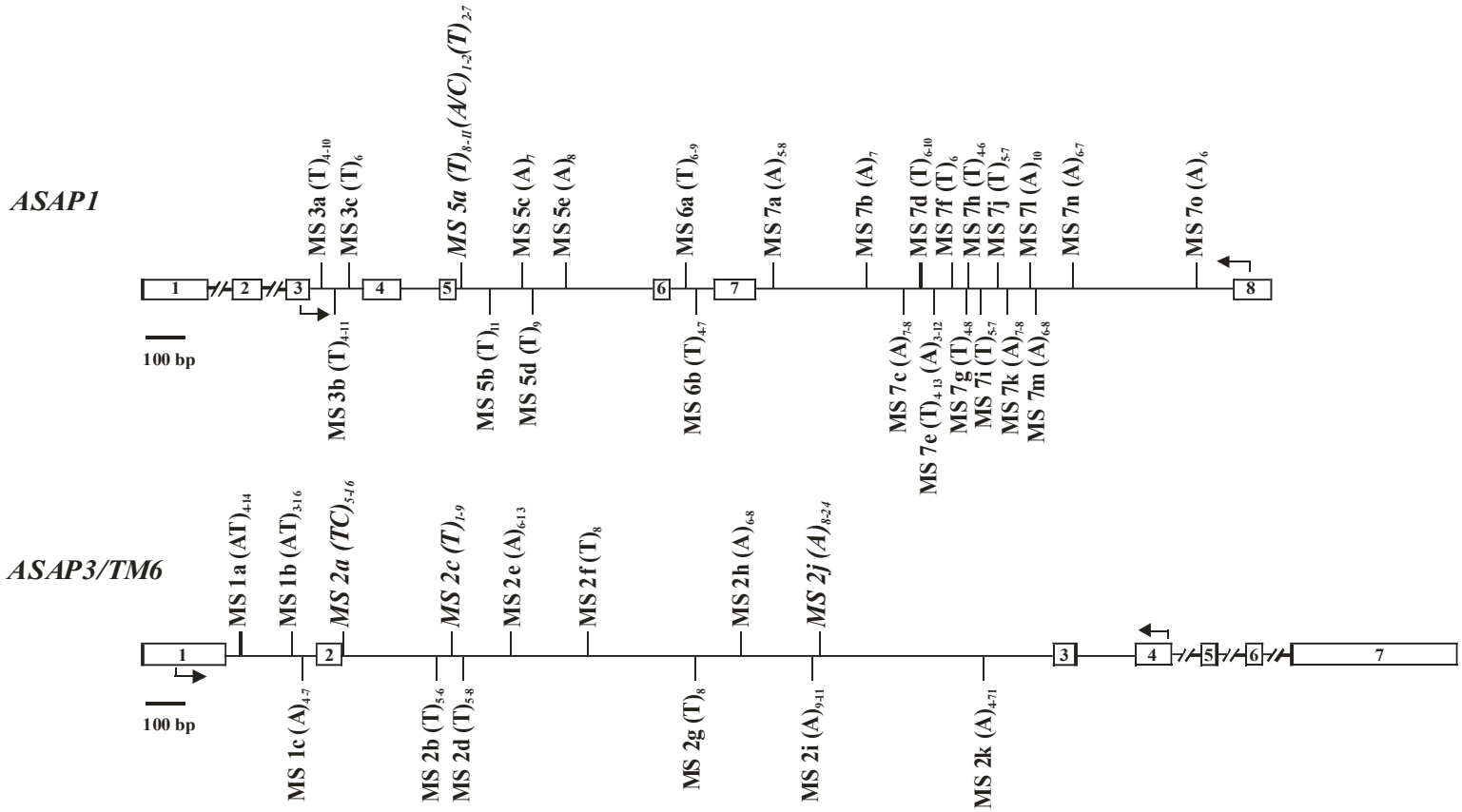

Figure 1. Structures of the *ASAP1* and *ASAP3/TM6* genes. Exons are shown as numbered boxes. The relative positions of identified microsatellites are indicated. A microsatellite is identified if it contains at least six repeat units in one species. The microsatellites discussed in this paper are shown in italics. The positions of the PCR primers used in sequence amplification are marked by arrows (Barrier et al. 1999).

Microsatellite loci can be classed as perfect or imperfect (Estoup et al. 1995). Imperfect microsatellites contain nucleotide repeats that are interrupted by one or more non-repeat nucleotides. Structural interruptions decrease the number of tandem repeats and may stabilize microsatellite loci, rendering them less prone to slippage mutations (Van Treuren et al. 1997).We observe several cases of structural evolution of imperfect to perfect microsatellites (and vice versa). Mutational steps that govern the interspecific dynamics of microsatellite loci were inferred by reconstructing the phylogenies of the genes that contain microsatellite sequences (Barrier et al. 1999) and mapping the structures of the embedded simple sequence repeats and their associated flanking regions onto the phylogenies. For four simple sequence repeat regions (see Figure 2), we can document insertions/deletions or substitutions of nucleotide interruptions in microsatellite loci.

Perfect microsatellite loci may arise from imperfect repeat sequences by the removal of nucleotide interruptions. One example is the *ASAP3/TM6* MS (microsatellite) 2a locus, which is a perfect dinucleotide $(TC)_n$ microsatellite in the *ASAP3/TM6-B* gene in the Hawaiian species (see Figure 2B). This microsatellite displays a greater degree of between-species variability (n= 6–16) than its imperfect microsatellite orthologue in the *ASAP3/TM6-A* gene in the Hawaiian species, all of which share the same $(TC)_4C(TC)_1$ structure (see Figure 2B). In this case, an insertion of a T nucleotide upstream of the C interruption appears to have led to the formation of the perfect, uninterrupted microsatellite locus.

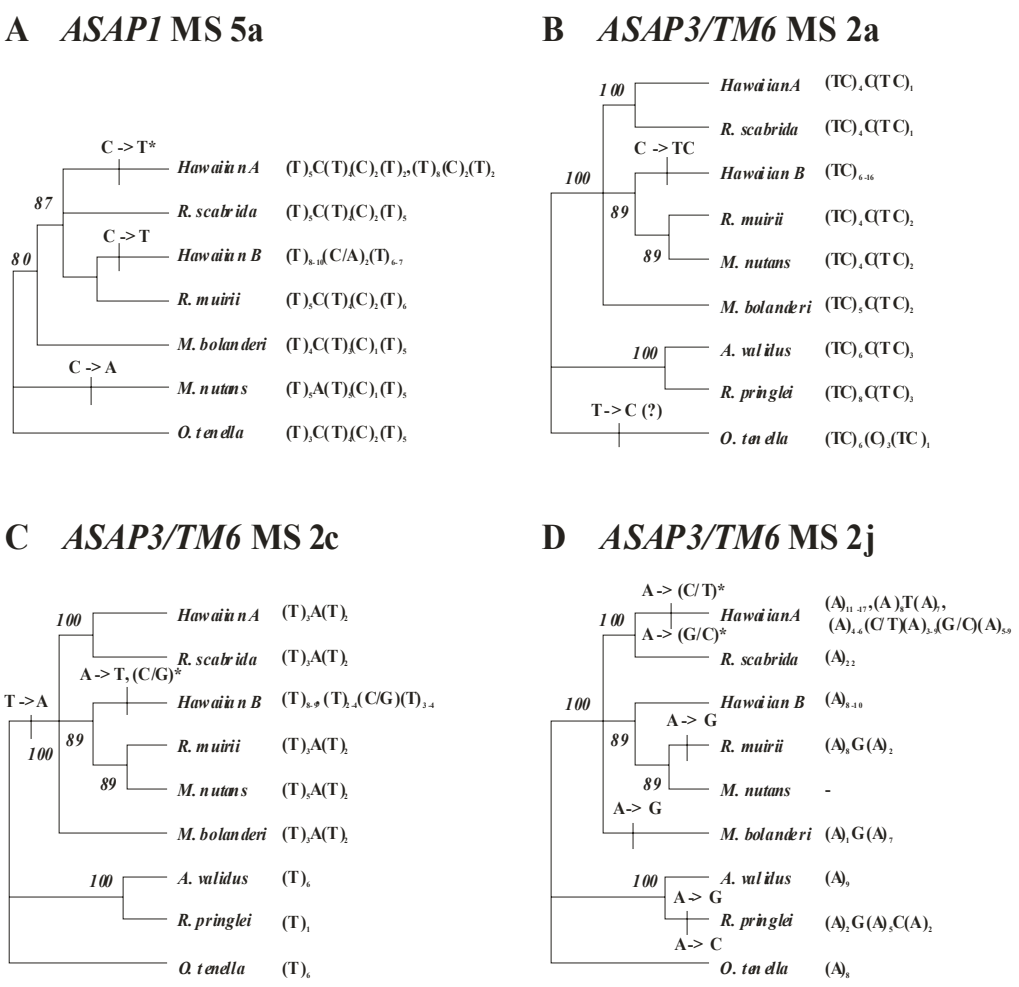

Figure 2. Evolution of four *ASAP1* and *ASAP3/TM6* microsatellite loci. The structures of the microsatellite loci are mapped onto the respective *ASAP1* and *ASAP3/TM6* gene phylogenies (Barrier et al. 1999). The variable microsatellite sequences were excluded when the gene phylogenies were reconstructed. The Hawaiian genes are indicated as either *A* or *B* copy. The inferred mutations represent the most parsimonious changes for the structures of these loci. Mutations with an asterisk indicate that the changes are observed in only some members of the clade. A dash indicates that a deletion event has removed that microsatellite locus. In the *ASAP1* MS 5a locus, we assume that the first C interruption is the ancestral state. In the *ASAP3/TM6* MS 2j locus, we assume that the perfect $(A)_n$ microsatellite is the ancestral state. Bootstrap percentage values from 500 replicates of the data are given in italics next to the nodes.

A similar pattern is observed in the *ASAP1* MS 5a locus, where removal of nucleotide bases that interrupt the contiguity of repeat regions results in the formation of longer repeat tracts (see Figure 2A). MS 5a is a compound microsatellite and consists of a $(T)_n(C/A)_1(T)_m(C)_{1–2}(T)_p$ interrupted repeat region in the *ASAP1* gene of the North American species. The *ASAP1-B* gene of the Hawaiian species, however, contains a longer, variable $(T)_n$ tract at the 5' end of this complex microsatellite. From the phylogeny, we infer that the imperfect $(T)_{3–5}C(T)_{3–5}(C)_{1–2}(T)_{5–6}$ was the founder sequence, and that loss of the first C interruption in the ancestral copy of the *ASAP1-B* locus resulted in a longer $(T)_n$ microsatellite tract in the Hawaiian species. This has occurred at least twice in the Hawaiian taxa; at least one Hawaiian species has an *ASAP1-A* gene that also exhibits a loss of the first C interruption.

Several clear instances of evolution of imperfect, interrupted microsatellites from perfect loci are also observed. In the *ASAP3/TM6* MS 2c locus, the members of the *Madia/Raillardiopsis* group and the Hawaiian silversword alliance contain an A interruption that is absent in the $(T)_n$ repeat sequences in the outgroup species (see Figure 2C). Interestingly, there is a secondary loss of this interruption in the *ASAP3/TM6-B* gene in some of the Hawaiian species. The evolution of structural interruptions may occur independently several times in species groups. There is evidence, for example, of multiple gains (or losses) of interruptions in the *ASAP3/TM6* MS 2j microsatellite locus (see Figure 2D).

It is unclear in most of these cases (as in most other cases examined) whether the gain or loss of the interrupted base(s) occurs via insertion/deletion events or simple

point mutations. Previous studies on microsatellite loci in primates (Messier et al. 1996) and bees (Estoup et al. 1995) have highlighted the role that both simple nucleotide substitutions and insertion/deletion events play in the formation and evolution of perfect microsatellite loci. Our results also suggest that the structural evolution of plant loci may proceed via multiple molecular mechanisms.

Acknowledgements

The authors thank Bruce Baldwin for providing us with some DNA samples for this study, and J. Mark Porter for helpful discussions. This work was supported by a grant from the National Science Foundation to M. D. P. and R. H. R. and from the Alfred P. Sloan Foundation to M. D. P.

Literature Cited

Baldwin, B. G. (1996) Phylogenetics of the California tarweeds and the Hawaiian silversword alliance (Madiinae: Heliantheae sensu lato). In: Compositae: Systematics. Hind, D. J. N. and Beentje, H. J. (eds.) Royal Botanic Gardens, Kew, England. Pp. 377–391.

Baldwin, B. G. and Robichaux, R. H. (1995) Historical biogeography and ecology of the Hawaiian silversword alliance: New molecular phylogenetic perspectives. In: Hawaiian Biogeography: Evolution on a Hot Spot Archipelago. Wagner, W. L., Funk, V. A. (eds.), Smithsonian Institution Press, Washington, D.C. pp. 259–287.

Baldwin, B. G. and Sanderson, M. J. (1998) Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proc. Natl. Acad. Sci. USA* **95**: 9402–9406.

Barrier, M., Baldwin, B. G., Robichaux, R. H. and Purugganan, M. D. (1999) Interspecific hybrid ancestry of a plant adaptive radiation: Allopolyploidy of the Hawaiian silversword alliance (Asteraceae) inferred from floral homeotic gene duplications. *Mol. Biol. Evol.* **16**: 1105–1113.

Estoup, A., Tailliez, C., Cornuet, J. M. and Solignac, M. (1995) Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). *Mol. Biol. Evol.* **12**: 1074–1084.

Goldstein, D. B. and Pollock, D. D. (1997) Launching microsatellites: A review of mutation processes and methods of phylogenetic inference. *J. Hered.* **88**: 335–342.

Goldstein, D. B., Roemer, G. W., Smith, D. A., Reich, D. E. and Wayne, R. K. (1999) The use of microsatellite variation to infer population structure and demographic history in a natural model system. *Genetics* **151**: 797–801.

Kunkel, T.A. (1993) Slippery DNA and human diseases. *Nature* **365**: 207–208.

Maddison, W. and Maddison, D. (1992) MACCLADE 3.0: Analysis of phylogeny and character evolution. Sinauer Press, Sunderland, MA.

Meloni, R., Albanese, V., Ravassard, P., Treilhou, F. and Mallet, J. (1998) A tetranucleotide polymorphic microsatellite, located in the first intron of the tyrosine hydroxylase gene, acts as a transcription regulatory element in vitro. *Hum. Mol. Genet.* **7**: 423–428.

Messier, W., Li, S.H. and Stewart, C.B. (1996) The birth of microsatellites. *Nature* **381**: 483.

Schlotterer, C. and Tautz, D. (1992) Slippage synthesis of simple sequence DNA. *Nuc. Acids Res.* **20**: 211–215.

Swofford, D. L. (1998) PAUP* 4.0 Beta. Sinauer Associates, Sunderland, MA.

Tautz, D. (1989) Hypervariability of simple sequence repeats as a general source for polymorphic DNA markers. *Nuc. Acids Res.* **17**: 6463–6471.

Van Treuren, R., Kuittinen, H., Karkkainen, K., Baena-Gonzalez, E. and Savolainen, O. (1997). Evolution of microsatellites in *Arabis petraea* and *Arabis lyrata*, outcrossing relatives of *Arabidopsis thaliana*. *Mol. Biol. Evol.* **14**: 220–229.

Weissenbach, X., Gyapay, G., Dib, C., Vignal, A., Morisset, J. and Millasse, P. (1992) A second-generation linkage map of the human genome. *Nature* **359**: 794–801.

APPENDIX C

Supplemental Tables of Chapter 3

Table 2. Nonsynonymous (Ka) and synonymous (Ks) substitutions in Hawaiian *ASAP1-A*

	Argyroxiphium caliginis	Argyroxiphium sandwicense	Dubautia laevigata	Dubautia latifolia	Dubautia laxa	Dubautia plantaginea	Dubautia raillardioides	Dubautia scabra	Dubautia sherffiana	Wilkesia gymnoxiphium
Argyroxiphium caliginis		0.00736	0.01855	0.01115	0.00735	0.0185	0.01475	0.01851	0.01513	0
Argyroxiphium sandwicense	0.0096		0.02615	0.01871	0.01479	0.02608	0.02228	0.02609	0.02285	0.00736
Dubautia laevigata	0.00957	0.01918		0.02254	0.01108	0.02232	0.01855	0.02993	0.02676	0.01856
Dubautia latifolia	0.01948	0.02928	0.0096		0.01116	0.02248	0.01867	0.00742	0.00379	0.01116
Dubautia laxa	0.01941	0.02916	0.00956	0.01947		0.01105	0.00734	0.01851	0.01513	0.00735
Dubautia plantaginea	0.01941	0.02916	0.00956	0.01947	0.01939		0.01849	0.02984	0.02474	0.01851
Dubautia raillardioides	0.01944	0.02921	0.00958	0.0195	0.01942	0.01942		0.02603	0.0228	0.01476
Dubautia scabra	0.01939	0.02913	0.00956	0	0.01937	0.01937	0.01941		0.01132	0.01851
Dubautia sherffiana	0.0199	0.02989	0.0098	0	0.01988	0.02493	0.01991	0		0.01513
Wilkesia gymnoxiphium	0.02931	0.03914	0.01925	0.0294	0.02928	0.02928	0.02933	0.02925	0.03002	

Ka above diagonal, Ks below diagonal

Table 3. Nonsynonymous (Ka) and synonymous (Ks) substitutions in Hawaiian *ASAP1-B*

	Argyroxiphium caliginis	Argyroxiphium sandwicense	Dubautia plantaginea	Dubautia raillardioides	Dubautia scabra	Dubautia sherffiana	Wilkesia gymnoxiphium
Argyroxiphium caliginis		0.00366	0.01104	0.0074	0.01479	0.02293	0.01103
Argyroxiphium sandwicense	0		0.00734	0.00369	0.01106	0.01906	0.00734
Dubautia plantaginea	0.01944	0.01944		0.00369	0.01107	0.01907	0.00734
Dubautia raillardioides	0.00974	0.00974	0.00973		0.00742	0.01534	0.00369
Dubautia scabra	0.0096	0.0096	0.0096	0		0.02299	0.01106
Dubautia sherffiana	0.0202	0.0202	0.02018	0.01011	0.00996		0.01906
Wilkesia gymnoxiphium	0	0	0.01946	0.00975	0.00961	0.02021	

Ka above diagonal, Ks below diagonal

Table 4. Nonsynonymous (Ka) and synonymous (Ks) substitutions in North American *ASAP1*

	Calycadenia multiglandulosa	Centromadia pungens	Deinandra lobbii	Kyhosia bolanderi	Harmonia nutans	Madia sativa	Osmadenia tenella	Carlquistia muirii	Anisocarpus scabridus
Calycadenia multiglandulosa		0.09223	0.08431	0.089	0.03274	0.05383	0.03166	0.03352	0.03361
Centromadia pungens	0.29455		0.01045	0.01047	0.07585	0.07859	0.1037	0.06969	0.0618
Deinandra lobbii	0.29397	0.0893		0.014	0.06795	0.0706	0.09165	0.06166	0.05382
Kyhosia bolanderi	0.26646	0.08886	0.03814		0.06808	0.07073	0.09584	0.06178	0.05392
Harmonia nutans	0.09006	0.24918	0.23542	0.20838		0.02599	0.04399	0.01475	0.01479
Madia sativa	0.17659	0.29957	0.25696	0.22891	0.13681		0.06159	0.01844	0.01849
Osmadenia tenella	0.08772	0.26879	0.24151	0.2272	0.05878	0.14111		0.04119	0.04131
Carlquistia muirii	0.13751	0.29852	0.25526	0.22681	0.0919	0.08155	0.09238		0.01476
Anisocarpus scabridus	0.15981	0.29583	0.28091	0.25158	0.12425	0.13633	0.13631	0.09159	

Ka above diagonal, Ks below diagonal

Table 5. Nonsynonymous (Ka) and synonymous (Ks) substitutions in Hawaiian *ASAP3/TM6-A*

	Argyroxiphium sandwicense	Dubautia laevigata	Dubautia latifolia	Dubautia laxa	Dubautia plantaginea	Dubautia raillardioides	Dubautia scabra	Dubautia sherffiana	Wilkesia gymnoxiphium
Argyroxiphium sandwicense		0.01481	0.00739	0.01482	0.01113	0.01109	0.00737	0.00368	0.00368
Dubautia laevigata	0.02108		0.01484	0.02231	0.01862	0.01856	0.01481	0.01108	0.01108
Dubautia latifolia	0.01038	0.01041		0.01485	0.01115	0.01112	0.00739	0.00369	0.00369
Dubautia laxa	0.01045	0.01048	0		0.01864	0.01857	0.01482	0.01108	0.01108
Dubautia plantaginea	0.02081	0.02087	0.01027	0.01034		0.01488	0.01113	0.0074	0.0074
Dubautia raillardioides	0.01043	0.01046	0	0	0.01033		0.01109	0.00738	0.00738
Dubautia scabra	0.03177	0.03186	0.0209	0.02105	0.03144	0.02101		0.00368	0.00368
Dubautia sherffiana	0.02103	0.02109	0.01038	0.01045	0.02081	0.01043	0.03177		0
Wilkesia gymnoxiphium	0.01044	0.01047	0	0	0.01033	0	0.02103	0.01044	

Ka above diagonal, Ks below diagonal

Table 6. Nonsynonymous (Ka) and synonymous (Ks) substitutions in Hawaiian *ASAP3/TM6-B*

	Argyroxiphium sandwicense	Dubautia laevigata	Dubautia latifolia	Dubautia laxa	Dubautia raillardioides	Dubautia scabra	Dubautia sherffiana	Wilkesia gymnoxiphium
Argyroxiphium sandwicense		0.01111	0.01486	0.01112	0.0111	0.00738	0.0074	0.00368
Dubautia laevigata	0.01039		0.01861	0.01485	0.01483	0.01109	0.01112	0.01483
Dubautia latifolia	0.02089	0.01039		0.01862	0.0186	0.01484	0.01487	0.0186
Dubautia laxa	0	0.01039	0.02089		0.01484	0.0111	0.01112	0.01484
Dubautia raillardioides	0.0104	0	0.0104	0.0104		0.01109	0.01111	0.01483
Dubautia scabra	0.03167	0.021	0.03167	0.03167	0.02102		0.00739	0.01109
Dubautia sherffiana	0.01035	0	0.01035	0.01035	0	0.02093		0.01111
Wilkesia gymnoxiphium	0.03163	0.02098	0.03163	0.03163	0.021	0.04264	0.0209	

Ka above diagonal, Ks below diagonal

Table 7. Nonsynonymous (Ka) and synonymous (Ks) substitutions in North American *ASAP3/TM6*

	Adenothamnus validus	Calycadenia multiglandulosa	Centromadia pungens	Kyhosia bolanderi	Harmonia nutans	Madia sativa	Osmadenia tenella	Carlquistia muirii	Raillardella pringlei	Anisocarpus scabridus
Adenothamnus validus		0.00368	0.00738	0.00751	0.01481	0.00738	0.00368	0.01107	0.00368	0.01108
Calycadenia multiglandulosa	0.07614		0.00369	0.00375	0.00737	0.00368	0	0.00367	0	0.00368
Centromadia pungens	0.01041	0.06454		0.00752	0.01109	0.00738	0.00369	0.00737	0.00368	0.00738
Kyhosia bolanderi	0.05437	0.08875	0.04302		0.01128	0.00751	0.00375	0.0075	0.00375	0.0075
Harmonia nutans	0.09994	0.16135	0.09958	0.13819		0.01108	0.00738	0.01106	0.00737	0.01107
Madia sativa	0.05369	0.09932	0.04249	0.07722	0.0765		0.00369	0.00736	0.00368	0.00737
Osmadenia tenella	0.04245	0.05331	0.0315	0.05409	0.09941	0.06457		0.00368	0	0.00368
Carlquistia muirii	0.06531	0.12409	0.06507	0.10147	0.07698	0.04287	0.08793		0.00367	0.00736
Raillardella pringlei	0.04259	0.08751	0.03161	0.06562	0.11171	0.05359	0.05336	0.07663		0.00368
Anisocarpus scabridus	0.03183	0.08784	0.03172	0.06586	0.06525	0.03181	0.05355	0.032	0.04267	

Ka above diagonal, Ks below diagonal

Table 8. Nonsynonymous (Ka) and synonymous (Ks) substitutions in Hawaiian *ASCAB9-A*

	Argyroxiphium caliginis	Argyroxiphium sandwicense	Dubautia laevigata	Dubautia latifolia	Dubautia laxa	Dubautia plantaginea	Dubautia raillardioides	Dubautia sherffiana	Wilkesia gymnoxiphium
Argyroxiphium caliginis		0.00561	0	0.00561	0.0028	0	0.0028	0	0
Argyroxiphium sandwicense	0.00646		0.00561	0.01126	0.0028	0.00561	0.00843	0.00561	0.00561
Dubautia laevigata	0.00646	0		0.00561	0.0028	0	0.0028	0	0
Dubautia latifolia	0.00647	0	0		0.00843	0.00561	0.0028	0.00561	0.00561
Dubautia laxa	0.00646	0	0	0		0.0028	0.00561	0.0028	0.0028
Dubautia plantaginea	0.00646	0	0	0	0		0.0028	0	0
Dubautia raillardioides	0.00647	0	0	0	0	0		0.0028	0.0028
Dubautia sherffiana	0.00646	0	0	0	0	0	0		0
Wilkesia gymnoxiphium	0.01298	0.00647	0.00646	0.00648	0.00646	0.00646	0.00647	0.00646	

Ka above diagonal, Ks below diagonal

Table 9. Nonsynonymous (Ka) and synonymous (Ks) substitutions in Hawaiian *ASCAB9-B*

	Argyroxiphium caliginis	Argyroxiphium sandwicense	Dubautia laevigata	Dubautia plantaginea	Wilkesia gymnoxiphium
Argyroxiphium caliginis		0	0	0	0
Argyroxiphium sandwicense	0		0	0	0
Dubautia laevigata	0	0		0	0
Dubautia plantaginea	0.00646	0.00646	0.00646		0
Wilkesia gymnoxiphium	0.00646	0.00646	0.00646	0.01298	

Ka above diagonal, Ks below diagonal

Table 10. Nonsynonymous (Ka) and synonymous (Ks) substitutions in Hawaiian *ASCAB9-C*

	Argyroxiphium sandwicense	Dubautia latifolia	Dubautia laxa	Dubautia scabra	Wilkesia gymnoxiphium
Argyroxiphium sandwicense		0.0028	0	0.0028	0.00561
Dubautia latifolia	0.00646		0.0028	0	0.0028
Dubautia laxa	0.01298	0.00646		0.0028	0.00561
Dubautia scabra	0.01298	0.00646	0.01298		0.0028
Wilkesia gymnoxiphium	0.01298	0.00646	0.01298	0	

Ka above diagonal, Ks below diagonal

Table 11. Nonsynonymous (Ka) and synonymous (Ks) substitutions in North American *ASCAB9*

	Centromadia pungens	Deinandra lobbii	Harmonia nutans	Osmadenia tenella	Carlquistia muirii	Anisocarpus scabridus
Centromadia pungens		0.00562	0.00281	0.00281	0.00562	0.00281
Deinandra lobbii	0.20546		0.0028	0.0028	0.00561	0.0028
Harmonia nutans	0.21379	0.00646		0	0.0028	0
Osmadenia tenella	0.21379	0.00646	0.01298		0.0028	0
Carlquistia muirii	0.2053	0.01298	0.01955	0.01955		0.0028
Anisocarpus scabridus	0.2053	0	0.00646	0.00646	0.01298	

Ka above diagonal, Ks below diagonal

APPENDIX D

Supplemental Tables for Chapter 4

Table 1. Primers for amplifying population dataset

Gene #	Primer Orientation	Primer sequence
1	Forward	5'-GCATCGTTTGGTTCCCTATCTT-3'
	Internal Reverse	5'-TCGGACATCTCTGACGTA-3'
	Reverse	5'-AGTTACAGCTCCCACCCACAA-3'
2	Forward	5'-GGTCAAGGATGCTTTGAGGA-3'
	Internal Forward	5'-CAAATGAGATTGGTGTGACT-3'
	Internal Reverse	5'-TTGCTTTATCAGCTCCCTCACT-3'
	Reverse	5'-CTGGAGCCTCCTTCTCCATA-3'
3	Forward	5'-GAGAAAAGTGGTACGGGTCCTA-3'
	Internal Forward	5'-CAAATGCCAAAGAGTGATAAATG-3'
	Internal Reverse	5'-CATCTGAGTGTTACCTGAATCA-3'
	Reverse	5'-GACACCATGGTTACTATTTTGAAGC-3'
4	Forward	5'-GGAAAGTGTGGAGGCTACTGC-3'
	Internal Forward	5'-CAGGTGACTGTTATGGACAC-3'
	Reverse	5'-GATTTGGAAACTTATCCAGCTTTTAGA-3'
5	Forward	5'-GTACAGCGATACAACAACCCAGA-3'
	Internal Forward	5'-TCTTGTGGTGAAACATGG-3'
	Internal Reverse	5'-TTCTCTAAGACTTCAGATGG-3'
	Reverse	5'-CGTTACAATAACATCAGCTCCAACAT-3'
6	Forward	5'-ATGTCAGAGAAAGGAGGCAAGG-3'
	Internal Forward	5'-GACAAACCTACCAAGTCTAGCA-3'
	Internal Reverse	5'-GATACTCAAGGACTTTTCTGA-3'
	Reverse	5'-GCGATGGAATAAAGGCAAGAAC-3'
7	Forward	5'-GGCTCCTACATTTTCGCCTAA-3'
	Internal Forward	5'-TATTTTGACTTGGTTTTTGCTTGT-3'
	Internal Reverse	5'-AAAGTATTGAAATCCAGAAGATGA-3'
	Reverse	5'-TCTGCGCGTATACAGCAAAAC-3'

Table 2. Summary of 304 unique expressed sequence tags (ESTs)

GeneID[a]	Description	Cds length[b]	Ka*	Ks*	ω*
At1g01050	putative - soluble inorganic pyrophosphatase	340	0.0000	0.1266	0.0000
At1g01920	hypothetical - similarity to ribulose-1,5 bisphosphate carboxylase	200	0.0217	0.0000	n.a.[d]
At1g02130	GTP-binding protein, ara-5 ras homolog	451	0.0312	0.5523	0.0565
At1g02560	ATP-dependent clp protease proteolytic subunit (nClpP1)	440	0.0000	0.0782	0.0000
At1g05620	unknown	497	0.0446	0.0561	0.7950
At1g05830	unknown	316	0.0271	0.1116	0.2428
At1g06570	4-hydroxyphenylpyruvate dioxygenase (HPD)	317	0.0089	0.0871	0.1022
At1g07230	unknown	215	0.0134	0.0703	0.1906
At1g07770	ribosomal protein S15	393	0.0000	0.1191	0.0000
At1g08180	hypothetical	235	0.0187	0.0930	0.2011
At1g08840	hypothetical	238	0.0092	0.1152	0.0799
At1g09210	putative calcium-binding protein, calreticulin	464	0.0365	0.1437	0.2540
At1g09280	unknown	439	0.0228	0.0953	0.2392
At1g09430	unknown- similar to ATP-citrate-lyase	359	0.0081	0.0891	0.0909
At1g09980	hypothetical	246	0.0113	0.0638	0.1771
At1g11840	lactoylglutathione lyase-like protein	184	0.0078	0.0196	0.3980
At1g13440	putative- glyceraldehyde-3-phosphate dehydrogenase	336	0.0042	0.1031	0.0407
At1g14400	ubiquitin carrier protein, putative	457	0.0000	0.0763	0.0000
At1g14690	hypothetical	328	0.0309	0.1036	0.2983
At1g17140	hypothetical	273	0.0259	0.1436	0.1804
At1g18070	guanine nucleotide regulatory protein	274	0.0105	0.0959	0.1095
At1g18540	putative - 60S ribosomal protein L6	411	0.0069	0.0925	0.0746
At1g18740	unknown	420	0.0445	0.0915	0.4863
At1g20260	vacuolar H+-ATPase subunit B	376	0.0076	0.1662	0.0457
At1g20440	hypothetical	475	0.0628	0.1477	0.4252
At1g20450	hypothetical	505	0.0944	0.1423	0.6634
At1g20690	high mobility group protein (HMG1)	271	0.0363	0.3655	0.0993
At1g21400	putative- branched-chain alpha keto-acid dehydrogenase	252	0.0055	0.0947	0.0581
At1g22780	putative 40S ribosomal protein S18	459	0.0061	0.1850	0.0330
At1g23290	60s ribosomal protein l27a	363	0.0039	0.0621	0.0628
At1g24620	putative calmodulin	546	0.0101	0.0566	0.1784
At1g25280	unknown- similar to putative Tub family	425	0.0683	0.0439	1.5558
At1g26360	hypothetical - similar to putative alpha/beta hydrolase	411	0.0292	0.0863	0.3384
At1g27370	putative *SQUAMOSA*-promoter binding protein 2	153	0.0191	0.0469	0.4072
At1g27390	putative- import receptor	205	0.0545	0.1554	0.3507
At1g28330	putative- dormancy-associated protein	332	0.0000	0.1294	0.0000
At1g29320	hypothetical	163	0.0174	0.0799	0.2178
At1g29850	unknown- similarity to apoptosis-related	355	0.0039	0.0762	0.0512

GeneID[a]	Description	Cds length[b]	Ka*	Ks*	ω*
At1g32200	glycerol-3-phosphate acyltransferase	336	0.0128	0.1365	0.0938
At1g32640	putative- protein kinase	363	0.0118	0.0916	0.1288
At1g34420	hypothetical - Eukaryotic protein kinase domain	204	0.0138	0.0183	0.7541
At1g36730	Eukaryotic translation initiation factor 5	341	0.0919	0.0582	1.5790
At1g41880	ribosomal	336	0.0000	0.0668	0.0000
At1g47260	Unknown	259	0.0163	0.0725	0.2248
At1g49320	unknown	550	0.0076	0.1129	0.0673
At1g49750	unknown- similarity to extensin-like	323	0.0450	0.0000	n.a.[d]
At1g50010	tubulin alpha-2/alpha-4 chain, putative	225	0.0000	0.0715	0.0000
At1g51200	unknown	198	0.0145	0.1157	0.1253
At1g52400	beta-glucosidase	306	0.0369	0.2301	0.1604
At1g52760	putative lipase similar to monoglyceride lipase	245	0.0329	0.2074	0.1586
At1g53200	hypothetical	615	0.0544	0.1199	0.4537
At1g53580	glyoxalase II, putative	538	0.0118	0.1009	0.1169
At1g54010	myrosinase-associated	503	0.0203	0.1215	0.1671
At1g54410	water stress-induced	297	0.0000	0.1031	0.0000
At1g54610	CRK1 protein	371	0.0038	0.0814	0.0467
At1g55870	putative- poly(A)-specific ribonuclease	453	0.0091	0.0774	0.1176
At1g56220	unknown	281	0.0211	0.1586	0.1330
At1g58290	putative- glutamyl-tRNA reductase	496	0.0113	0.0869	0.1300
At1g60870	unknown	204	0.0068	0.1170	0.0581
At1g60900	putative U2 snRNP auxiliary factor	473	0.0190	0.0762	0.2493
At1g62330	unknown	460	0.0031	0.1591	0.0195
At1g64370	unknown	438	0.0363	0.1108	0.3276
At1g67140	hypothetical	386	0.0697	0.0543	1.2836
At1g69500	putative, cytochrome P450	392	0.0253	0.0676	0.3743
At1g69730	putative- kinase	581	0.0446	0.1217	0.3665
At1g69830	putative alpha-amylase	217	0.0064	0.0885	0.0723
At1g70830	hypothetical- similar to putative ripening-related	603	0.0328	0.2679	0.1224
At1g71710	putative- similar to 30S Ribosomal	319	0.0044	0.0930	0.0473
At1g71860	protein tyrosine phosphatase	170	0.0428	0.0885	0.4836
At1g73030	hypothetical - similarity to developmental protein	157	0.0000	0.1802	0.0000
At1g73540	unknown- similarity to diphosphoinositol polyphosphate phosphohydrolase	358	0.0158	0.1176	0.1344
At1g74250	putative heat shock	469	0.0509	0.0609	0.8358
At1g74540	putative cytochrome P450	220	0.0197	0.0502	0.3924
At1g75330	ornithine carbamoyltransferase precursor	286	0.0049	0.1749	0.0280
At1g75350	chloroplast 50S ribosomal protein L31	287	0.0731	0.1650	0.4430
At1g75950	skp1-related	483	0.0293	0.1929	0.1519
At1g76180	dehydrin	305	0.0283	0.2293	0.1234
At1g80920	J8-like	451	0.0031	0.1049	0.0296
At2g02760	putative ubiquitin-conjugating enzyme E2	276	0.0050	0.0846	0.0591
At2g03120	unknown	377	0.0115	0.1192	0.0965
At2g03350	unknown	384	0.0037	0.2251	0.0164
At2g04410	unknown	182	0.0322	0.0216	1.4907
At2g05910	unknown	464	0.0387	0.1233	0.3139

GeneID[a]	Description	Cds length[b]	Ka*	Ks*	ω*
At2g13350	hypothetical	265	0.0132	0.0940	0.1404
At2g15890	unknown	321	0.0221	0.1231	0.1795
At2g16600	cytosolic cyclophilin (ROC3)	249	0.0085	0.1191	0.0714
At2g16700	actin depolymerizing factor 5	151	0.0000	0.1026	0.0000
At2g16800	unknown	159	0.0000	0.1180	0.0000
At2g17240	unknown	154	0.0501	0.0000	n.a.[d]
At2g18330	putative AAA-type ATPase	215	0.0329	0.0728	0.4519
At2g20790	unknown	405	0.0281	0.0953	0.2949
At2g21190	putative ER lumen protein retaining receptor	159	0.0793	0.0978	0.8108
At2g22780	putative glyoxysomal malate dehydrogenase precursor	369	0.0118	0.0473	0.2495
At2g23170	unknown	489	0.0117	0.0000	n.a.[d]
At2g24200	putative leucine aminopeptidase	431	0.0131	0.1004	0.1305
At2g25160	putative cytochrome P450	578	0.1312	0.4328	0.3031
At2g25170	putative chromodomain-helicase-DNA-binding	348	0.0661	0.1047	0.6313
At2g25590	hypothetical	347	0.0596	0.1340	0.4448
At2g26510	putative membrane transporter	287	0.0252	0.1116	0.2258
At2g27050	ethylene-insensitive3-like1	385	0.0469	0.2243	0.2091
At2g27510	putative ferredoxin	337	0.0088	0.2128	0.0414
At2g27760	putative tRNA isopentenylpyrophosphate transferase	211	0.0280	0.0369	0.7588
At2g28350	unknown	453	0.0316	0.1225	0.2580
At2g28900	putative membrane channel	307	0.0046	0.0837	0.0550
At2g29180	unknown	321	0.0621	0.1297	0.4788
At2g29310	putative tropinone reductase	192	0.1046	0.2461	0.4250
At2g30570	photosystem II reaction center 6.1KD	402	0.0107	0.0813	0.1316
At2g30860	glutathione S-transferase	375	0.0000	0.2255	0.0000
At2g33220	unknown	432	0.0000	0.0811	0.0000
At2g33770	putative ubiquitin-conjugating enzyme E2	312	0.0468	0.1307	0.3581
At2g33830	putative auxin-regulated	176	0.0331	0.0896	0.3694
At2g34600	hypothetical	242	0.0774	0.0830	0.9325
At2g35320	similar to eyes absent protein	300	0.0094	0.0374	0.2513
At2g36680	unknown	290	0.0152	0.2189	0.0694
At2g38530	putative nonspecific lipid-transfer	357	0.0669	0.1483	0.4511
At2g39080	unknown	288	0.0251	0.0866	0.2898
At2g39460	60S ribosomal protein L23A	465	0.0121	0.0910	0.1330
At2g39570	unknown	310	0.0047	0.1050	0.0448
At2g39900	putative LIM-domain protein	563	0.0076	0.0718	0.1058
At2g40000	putative nematode-resistance	260	0.0217	0.2193	0.0990
At2g40060	unknown	386	0.0293	0.0907	0.3230
At2g40410	hypothetical	511	0.0227	0.1124	0.2020
At2g40780	putative translation initiation factor	209	0.0325	0.1045	0.3110
At2g41010	unknown	351	0.0869	0.2514	0.3457
At2g41429	ERD15	327	0.0215	0.1137	0.1891
At2g41710	putative AP2 domain transcription factor	245	0.0182	0.0424	0.4292
At2g41820	putative receptor-like protein kinase	485	0.0000	0.0837	0.0000
At2g43900	putative inositol polyphosphate 5'-phosphatase	620	0.0550	0.1817	0.3027

GeneID[a]	Description	Cds length[b]	Ka*	Ks*	ω*
At2g44360	unknown	274	0.0207	0.2720	0.0761
At2g45510	putative cytochrome P450	603	0.0371	0.1158	0.3204
At2g46600	putative caltractin	374	0.0148	0.1086	0.1363
At2g46640	hypothetical	459	0.0410	0.1706	0.2403
At2g46700	putative calcium-dependent protein kinase	297	0.0097	0.1125	0.0862
At2g47710	unknown	264	0.0054	0.1331	0.0406
AT3g01340	transport protein SEC13	235	0.0059	0.0839	0.0703
AT3g03770	hypothetical	173	0.0085	0.0412	0.2063
AT3g05520	alpha subunit of F-actin capping protein alternative splicing isoform 1	569	0.0122	0.0400	0.3050
AT3g05590	putative 60S ribosomal protein L18	163	0.0179	0.0447	0.4004
AT3g05600	putative epoxide hydrolase	556	0.0279	0.0972	0.2870
AT3g06790	DAG protein	255	0.0697	0.0887	0.7858
AT3g07930	hypothetical	295	0.0594	0.1206	0.4925
AT3g08580	adenylate translocator	170	0.0179	0.0815	0.2196
AT3g09360	putative transcription factor	220	0.0538	0.0574	0.9373
AT3g09390	metallothionein-like	239	0.0059	0.0798	0.0739
AT3g12260	unknown	402	0.0070	0.0953	0.0735
AT3g12570	unknown	549	0.0156	0.1018	0.1532
AT3g14230	DNA-binding protein contains AP2 domain	155	0.0278	0.0517	0.5377
AT3g14600	putative 60S ribosomal protein L18A	342	0.0082	0.1802	0.0455
AT3g15810	unknown	208	0.0138	0.0355	0.3887
AT3g15840	unknown	390	0.0035	0.0898	0.0390
AT3g16840	RNA helicase	545	0.0604	0.0493	1.2252
AT3g18960	hypothetical	545	0.0685	0.1131	0.6057
AT3g19240	dem-like	198	0.0072	0.0750	0.0960
AT3g19370	unknown	202	0.0358	0.0373	0.9598
AT3g20310	ethylene responsive element binding factor	310	0.0139	0.1094	0.1271
AT3g22060	hypothetical	177	0.0578	0.0436	1.3257
AT3g23150	ethylene receptor	438	0.0198	0.1748	0.1133
AT3g23920	beta-amylase	453	0.0125	0.1870	0.0668
AT3g25220	immunophilin (FKBP15-1)	441	0.0196	0.1001	0.1958
AT3g27420	hypothetical	162	0.0956	0.1637	0.5840
AT3g29390	unknown	209	0.0643	0.1113	0.5777
AT3g42790	nucleic acid binding	338	0.0082	0.1111	0.0738
AT3g44110	dnaJ protein homolog atj3	436	0.0097	0.1679	0.0578
AT3g45310	cysteine protease-like	194	0.0075	0.0731	0.1026
AT3g46040	cytoplasmic ribosomal protein S15a -like	393	0.0071	0.0681	0.1043
AT3g46430	putative- mitochondrial ATP Synthase 6 KD Subunit	168	0.0000	0.1232	0.0000
AT3g46440	dTDP-glucose 4-6-dehydratases-like	256	0.0110	0.0931	0.1182
AT3g48530	putative- probable transcription regulator	196	0.0301	0.2502	0.1203
AT3g49470	alpha NAC-like protein Nascent polypeptide associated complex	162	0.0174	0.1006	0.1730
AT3g49640	putative- Nitrogen Regulation Protein NIFR3	502	0.0000	0.0828	0.0000
AT3g49720	hypothetical	153	0.0277	0.0520	0.5327
AT3g50440	putative- pir7a	222	0.0345	0.1041	0.3314
AT3g51300	rac-like GTP binding protein Arac11	264	0.0053	0.0549	0.0965
AT3g51520	putative	352	0.0080	0.1195	0.0669

GeneID[a]	Description	Cds length[b]	Ka*	Ks*	ω*
AT3g51850	calcium-dependent protein kinase	239	0.0000	0.0665	0.0000
AT3g52180	putative	450	0.0200	0.1646	0.1215
AT3g52300	putative - similar to ATP synthase D chain (mitochondrial)	296	0.0095	0.0939	0.1012
AT3g52880	monodehydroascorbate reductase (NADH) - like	239	0.0000	0.1682	0.0000
AT3g54090	fructokinase - like	289	0.0096	0.0944	0.1017
AT3g54560	histone H2A.F/Z	360	0.0000	0.4279	0.0000
AT3g55280	ribosomal L23a - like	338	0.0190	0.1091	0.1742
AT3g55520	putative- P59 Protein (HSP Binding Immunophilin)	250	0.0058	0.0889	0.0652
AT3g56090	putative- ferritin 2 precursor	292	0.0098	0.1038	0.0944
AT3g56660	transcription factor-like	432	0.0099	0.0708	0.1398
AT3g61110	ribosomal protein S27	151	0.0096	0.0244	0.3934
AT4g02510	putative chloroplast outer envelope 86-like	262	0.0166	0.1474	0.1126
AT4g02520	Atpm24.1 glutathione S transferase	403	0.0432	0.0942	0.4586
AT4g02940	hypothetical	218	0.0067	0.0319	0.2100
AT4g02980	auxin-binding protein 1 precursor	332	0.0042	0.1453	0.0289
AT4g03210	putative xyloglucan endotransglycosylase	483	0.0143	0.1298	0.1102
AT4g05050	unknown	174	0.0083	0.1325	0.0626
AT4g09890	putative	178	0.0449	0.0000	n.a.[d]
AT4g10920	putative- RNA polymerase II transcription cofactor p15	394	0.0183	0.1527	0.1198
AT4g11310	drought-inducible cysteine proteinase RD21A precursor -like	325	0.0044	0.1971	0.0223
AT4g11320	drought-inducible cysteine proteinase RD21A precursor -like	517	0.0220	0.1792	0.1228
AT4g14270	hypothetical	374	0.0376	0.1077	0.3491
AT4g14880	cytosolic O-acetylserine(thiol)lyase	484	0.0029	0.1082	0.0268
AT4g15030	putative	251	0.0625	0.1026	0.6092
AT4g15470	hypothetical	210	0.0140	0.1209	0.1158
AT4g15950	putative	422	0.0636	0.0390	1.6308
AT4g16190	cysteine proteinase like protein	535	0.0078	0.1272	0.0613
AT4g16310	hypothetical	364	0.0442	0.0736	0.6005
AT4g16720	ribosomal	447	0.0063	0.0917	0.0687
AT4g18040	translation initiation factor eIF4E	463	0.0117	0.1809	0.0647
AT4g19410	putative pectinacetylesterase	196	0.0000	0.1898	0.0000
AT4g20360	translation elongation factor EF-Tu precursor, chloroplast	238	0.0000	0.1078	0.0000
AT4g20410	putative- gamma-SNAP	262	0.0164	0.0870	0.1885
AT4g21960	peroxidase prxr1	291	0.0000	0.1056	0.0000
AT4g22820	predicted protein PVPR3 protein	298	0.0563	0.1454	0.3872
AT4g23100	gamma-glutamylcysteine synthetase	226	0.0000	0.1491	0.0000
AT4g23250	protein kinase - like	195	0.0954	0.1588	0.6008
AT4g24190	HSP90-like protein HSP90 homolog	398	0.0215	0.1179	0.1824
AT4g24420	hypothetical	200	0.1536	0.0694	2.2133
AT4g25340	putative- immunophilin FKBP46	337	0.0259	0.0904	0.2865
AT4g25890	putative acidic ribosomal	160	0.0287	0.0858	0.3345
AT4g27320	putative	150	0.0195	0.1073	0.1817
AT4g28000	putative- *MSP1*	221	0.0283	0.0178	1.5899

GeneID[a]	Description	Cds length[b]	Ka*	Ks*	ω*
AT4g28400	protein phosphatase 2C-like	191	0.0072	0.1313	0.0548
AT4g28470	putative	302	0.0616	0.0411	1.4988
AT4g28730	putative- glutaredoxin	330	0.0287	0.0710	0.4042
AT4g29200	hypothetical	177	0.1585	0.0883	1.7950
AT4g30220	snRNP Sm protein F - like	251	0.0000	0.1554	0.0000
AT4g31900	putative- zinc-finger helicase	165	0.1268	0.1042	1.2169
AT4g33060	putative- peptidyl-prolyl cis-trans isomerase	369	0.0304	0.0842	0.3610
AT4g34710	arginine decarboxylase *SPE2*	231	0.0244	0.0861	0.2834
AT4g35020	Rho1Ps homolog/ Rac-like protein	459	0.0296	0.0000	n.a.[d]
AT4g35910	putative	241	0.0362	0.1293	0.2800
AT4g36540	putative	487	0.0200	0.1178	0.1698
AT4g36760	aminopeptidase-like	322	0.0000	0.1233	0.0000
AT4g37270	Cu2+-transporting ATPase-like	152	0.0590	0.0710	0.8310
AT4g37870	phosphoenolpyruvate carboxykinase (ATP) -like	419	0.0135	0.1420	0.0951
AT4g37930	glycine hydroxymethyltransferase like	330	0.0064	0.1254	0.0510
AT4g39090	drought-inducible cysteine proteinase RD19A precursor	349	0.0184	0.1276	0.1442
AT4g40040	Histone H3	411	0.0035	0.1480	0.0236
AT5g01410	pyridoxine biosynthesis protein - like	349	0.0040	0.0646	0.0619
AT5g01600	ferritin 1 precursor	410	0.0369	0.0582	0.6340
AT5g02080	putative	356	0.0242	0.0829	0.2919
AT5g02380	metallothionein 2b	234	0.0121	0.1346	0.0899
AT5g02450	60S ribosomal	327	0.0174	0.1174	0.1482
AT5g02960	putative- ribosomal protein S23	320	0.0000	0.0711	0.0000
AT5g03850	Ribosomal Protein S28	195	0.0000	0.1416	0.0000
AT5g04410	putative- *NAC2*	169	0.0437	0.0215	2.0326
AT5g05080	putative- similarity to ubiquitin conjugating enzyme	188	0.0292	0.0757	0.3857
AT5g05220	unknown	437	0.0434	0.1276	0.3401
AT5g07320	peroxisomal Ca-dependent solute carrier-like	235	0.0182	0.0821	0.2217
AT5g07410	pectin methyl-esterase-like	539	0.0000	0.1007	0.0000
AT5g09600	putative	187	0.0000	0.1391	0.0000
AT5g12110	elongation factor 1B alpha-subunit	473	0.0295	0.1616	0.1825
AT5g13190	putative	327	0.0000	0.0463	0.0000
AT5g14320	ribosomal protein precursor	265	0.0000	0.1376	0.0000
AT5g14460	tRNA synthase - like	509	0.0083	0.0739	0.1123
AT5g15970	cold-regulated	200	0.0523	0.0945	0.5534
AT5g16590	receptor-like protein kinase	205	0.0282	0.0745	0.3785
AT5g17070	putative	273	0.0771	0.0803	0.9601
AT5g18540	hypothetical	371	0.0077	0.0874	0.0881
AT5g19390	putative	457	0.0364	0.0620	0.5871
AT5g19510	elongation factor 1B alpha-subunit	308	0.0206	0.2032	0.1014
AT5g19900	putative	299	0.0384	0.1325	0.2898
AT5g20880	putative	375	0.0150	0.1265	0.1186
AT5g23500	putative	271	0.0886	0.1094	0.8099
AT5g23660	MtN3-like	257	0.0312	0.1882	0.1658
AT5g23820	putative	367	0.1147	0.2428	0.4724
AT5g24780	vegetative storage protein *Vsp1*	312	0.0914	0.1503	0.6081

GeneID[a]	Description	Cds length[b]	Ka*	Ks*	ω*
AT5g24970	putative- Ubiquinone Biosynthesis Protein	371	0.0099	0.1021	0.0970
AT5g25800	exonuclease - like protein ribonuclease H	403	0.0592	0.0237	2.4979
AT5g33300	putative- chromokinesin KIF4	471	0.0329	0.1202	0.2737
AT5g33370	putative- proline-rich protein APG	347	0.0080	0.0801	0.0999
AT5g36250	protein phosphatase-2C PP2C-like	279	0.0103	0.1516	0.0679
AT5g36940	cationic amino acid transporter -like	392	0.0111	0.1102	0.1007
AT5g37300	putative	357	0.0404	0.1312	0.3079
AT5g37780	calmodulin 1 (CAM1)	342	0.0000	0.1342	0.0000
AT5g38650	putative	200	0.0523	0.0945	0.5534
AT5g38800	bZIP transcription factor - like	483	0.0405	0.1603	0.2527
AT5g39730	avirulence induced gene (AIG) - like	511	0.0330	0.1197	0.2757
AT5g39740	ribosomal protein L5	272	0.0150	0.1118	0.1342
AT5g42480	putative	492	0.0263	0.1064	0.2472
AT5g44520	putative- similarity to ribose 5-phosphate isomerase	160	0.0436	0.0789	0.5526
AT5g46580	putative- similarity to salt-inducible protein	366	0.0079	0.0579	0.1364
AT5g47120	Bax inhibitor-1 like	440	0.0315	0.5146	0.0612
AT5g47700	60S acidic ribosomal protein P1-like	345	0.0042	0.1594	0.0263
AT5g48490	putative	243	0.0972	0.5281	0.1841
AT5g48630	cyclin C-like	510	0.0082	0.1049	0.0782
AT5g49450	putative - similarity to bZIP transcription factor	276	0.0358	0.1778	0.2013
AT5g49540	unknown	345	0.0081	0.1016	0.0797
AT5g51750	serine protease-like	342	0.0168	0.1453	0.1156
AT5g54160	O-methyltransferase	453	0.0123	0.0697	0.1765
AT5g54270	Lhcb3 chlorophyll a/b binding	404	0.0000	0.0920	0.0000
AT5g54670	kinesin-like	475	0.0030	0.1081	0.0278
AT5g54900	putative - similarity to polyadenylate-binding protein 5	293	0.0208	0.0648	0.3210
AT5g58030	SPP30 - like protein SPP30	341	0.0084	0.1612	0.0521
AT5g58240	bis(5'-adenosyl)-triphosphatase-like	247	0.0119	0.1185	0.1004
AT5g59870	histone H2A - like	451	0.0392	0.3142	0.1248
AT5g59910	histone H2B - like	384	0.0036	0.0926	0.0389
AT5g60470	putative zinc finger	206	0.1381	0.0534	2.5861
AT5g62200	putative- embryo-specific protein 3	231	0.0382	0.0957	0.3992
AT5g62390	KED - like	346	0.0081	0.1488	0.0544
AT5g62540	ubiquitin-conjugating enzyme E2-17 kd 3	405	0.0174	0.1172	0.1485
AT5g62580	putative	289	0.0198	0.0786	0.2519
AT5g63530	putative- similarity to ATFP3	318	0.0181	0.1180	0.1534
AT5g64480	unknown	505	0.0698	0.1055	0.6616
AT5g65360	histone H3	411	0.0000	0.2501	0.0000
AT5g66220	chalcone isomerase-like	322	0.0358	0.0831	0.4308
AC009398[c]	unknown	423	0.0271	0.1308	0.2072
AC079677[c]	unknown- similarity to *MFP1* attachment factor 1	498	0.0516	0.1187	0.4347

[a]Gene ID number based on the *Arabidopsis* Information Resource gene identification number; [b]length of coding region sequence compared between *A. thaliana* and *A. lyrata*; [c]GenBank accession numbers; [d]no synonymous substitutions, so ω is undefined (but greater than 1).