Molecular Evolution across the Asteraceae: Micro- and Macroevolutionary Processes

Nolan C. Kane,*[,1] Michael S. Barker,[2] Shing H. Zhan,[1] and Loren H. Rieseberg[1,3]

[1]Department of Botany, The Biodiversity Research Centre, University of British Columbia, Vancouver, Canada

[2]Ecology and Evolutionary Biology Department, University of Arizona

[3]Department of Biology, Indiana University

***Corresponding author:** E-mail: nkane@biodiversity.ubc.ca.

Associate editor: Jody Hey

Abstract

The Asteraceae (Compositae) is a large family of over 20,000 wild, weedy, and domesticated species that comprise approximately 10% of all angiosperms, including annual and perennial herbs, shrubs and trees, and species on every continent except Antarctica. As a result, the Asteraceae provide a unique opportunity to understand the evolutionary genomics of lineage radiation and diversification at numerous phylogenetic scales. Using publicly available expressed sequence tags from 22 species representing four of the major Asteraceae lineages, we assessed neutral and nonneutral evolutionary processes across this diverse plant family. We used bioinformatic tools to identify candidate genes under selection in each species. Evolution at silent and coding sites were assessed for different Gene Ontology functional categories to compare rates of evolution over both short and long evolutionary timescales. Our results indicate that patterns of molecular change across the family are surprisingly consistent on a macroevolutionary timescale and much more so more than would be predicted from the analysis of one (or many) examples of microevolution. These analyses also point to particular classes of genes that may be crucial in shaping the radiation of this diverse plant family. Similar analyses of nuclear and chloroplast genes in six other plant families confirm that many of these patterns are common features of the plant kingdom.

Key words: macroevolution, microevolution, Asteraceae, comparative genomics, adaptive evolution, speciation.

Research article

Introduction

Ever since Darwin (1859) argued that species and lineages could originate and evolve via natural selection, biologists have debated whether larger patterns of evolution (macroevolution) could be explained by many small steps (microevolution). Many subsequent researchers agreed with Darwin and further argued that micro- and macroevolutionary processes are fundamentally similar and that the same processes that generate diversity within and between populations of a species also lead to formation of new species, genera, and even Kingdoms of life, given enough time (e.g., Dobzhansky 1951; Leroi 2000; Reznick and Ricklefs 2009). Although microevolution is well accepted below the species level, such process is not universally embraced as the causal forces behind diversity at broader phylogenetic scales (Gould and Lewontin 1979; Erwin 2000; Carroll 2001). Nonetheless, large discontinuities in the fossil record (Gould and Eldridge 1977), key developmental innovations (Erwin 1999), and higher level processes such as species sorting (Gould and Lewontin 1979; Vrba and Gould 1997) have all been used to argue that macroevolution is not a simple extension of local adaptation and other short-term processes. According to this view, emergent properties occur over long timescales in ways that cannot be predicted based on short-term observations (Erwin 2000).

In contrast, Dobzhansky (1937) argued that "full comprehension of the microevolutionary process observable within the span of a human lifetime" can lead to "an understanding of the mechanisms of macroevolutionary changes." Rapid shifts in morphology due to strong selection forces have been documented in a number of microevolutionary cases (e.g., Boag and Grant 1981; Grant PR and Grant BR 1997; Losos et al. 1997; Reznick et al. 1997; Janick 2003) and even the slowest observable microevolutionary changes would easily qualify as "punctuations" in the fossil record (Charlesworth et al. 1982). Moreover, extreme phenotypes are found segregating within species (Rieseberg et al. 2003), or in some cases in hybrids between species (Rieseberg et al. 1996, 2003). Thus, as predicted by theoretical work, the same processes that occur over microevolutionary timescales can lead to periods of stasis followed by rapid morphological changes, as often seen over macroevolutionary scales (Kirkpatrick 1982; Lynch 1990; Hansen and Martins 1996). Also, key innovations underlying habitat shifts (Filchak et al. 2000) or adaptive radiations (e.g., Hunter and Jernvall 1995) have often been shown to evolve due to incremental shifts in morphology rather than macromutations. However, even when superficially similar morphological changes occur over both short and long timescales, some have argued that the genetic basis is fundamentally different (Carroll 2000). Thus, after decades of debate, the argument remains unresolved as

to whether long-term evolutionary patterns are governed by fundamentally different forces than short-term evolution.

Comparisons of micro- versus macroevolution have been complicated by the fact that these two stages of evolution are usually studied with different tools. Macroevolution is typically the province of paleontology, phylogenetics, and evo-devo (Carroll 2000; Jablonski 2000; Stern 2000), whereas microevolution is amenable to manipulative approaches, such as controlled crosses and selection experiments. However, with the advent of high-throughput sequencing technologies, massive amounts of sequence data are becoming available for numerous taxa (Benson et al. 2007), making it possible to compare changes in a large number of genes across numerous and diverse taxa. These data enable direct comparisons of changes in the same characters (i.e., changes in DNA sequences) over both short and long evolutionary timescales. Moreover, because many genes are examined rather than only a few, light may be shed on genome-wide patterns of evolution.

The Asteraceae (Compositae) is a large family of over 20,000 wild, weedy, and domesticated species, including an estimated 10% of all Angiosperm species (Heywood 1978; Cronquist 1981). Using a database of over 800,000 Sanger expressed sequence tag (EST) sequences from 22 species covering four different tribes and representing several of the major lineages of the Asteraceae (Barker et al. 2008; Heesacker et al. 2008; http://cgpdb.ucdavis.edu/), we performed multiple bioinformatic analyses to assess neutral and nonneutral evolutionary processes across this diverse plant family. Rates of evolution at silent and coding sites were assessed for different Gene Ontology (GO) functional categories (Rhee et al. 2003), comparing rates of evolution over shorter and longer evolutionary timescales. Similar analyses were conducted in six other plant families to assess whether patterns of molecular evolution in the Asteraceae are observed in other plant taxa.

We used these data to test the hypothesis that rates of evolution over longer timescales are predictable due to functional constraints (i.e., purifying selection), but more variable over short timescales, where positive selection is detectable. We also addressed the following questions: What genetic changes accompany divergence over different timescales, such as between domesticated species and their wild progenitors, sister species, and more distant congeners or tribes within a family? Are the same genes evolving more rapidly/slowly in each comparison? Do particular classes of genes always tend to be highly conserved in amino acid sequence over long timescales?

Materials and Methods

Sequence Data

A total of 804,654 EST sequences from plants in the family Asteraceae were available through GenBank (Benson et al. 2007) on 1 June 2009, including substantial EST databases (greater than 15,000 high-quality sequences) for 22 species (fig. 1). More information on the ESTs can be obtained from the Compositae Genome Project (http://cgpdb.ucdavis.edu/), which generated the majority of those available for the Asteraceae (Barker et al. 2008; Heesacker et al. 2008). Five of the major lineages within the Asteraceae—the tribes Heliantheae (eight species of *Helianthus* and one of *Zinnia*), Cichorieae (five species of *Lactuca*, two *Taraxacum*, and two *Cicchorium*), Cardueae (two species of *Centaurea* and one of *Carthamus*), Anthemideae (*Artemisia annuua*), and Mutiseae (*Gerbera* × *hybrida*)—were represented. Additionally, for some species, substantial numbers of sequences were available for multiple genotypes. In particular, a total of 94,111 EST sequences were used from *Helianthus annuus*, including 66,099 sequences from domesticated sunflower varieties and 28,012 sequences from a wild genotype.

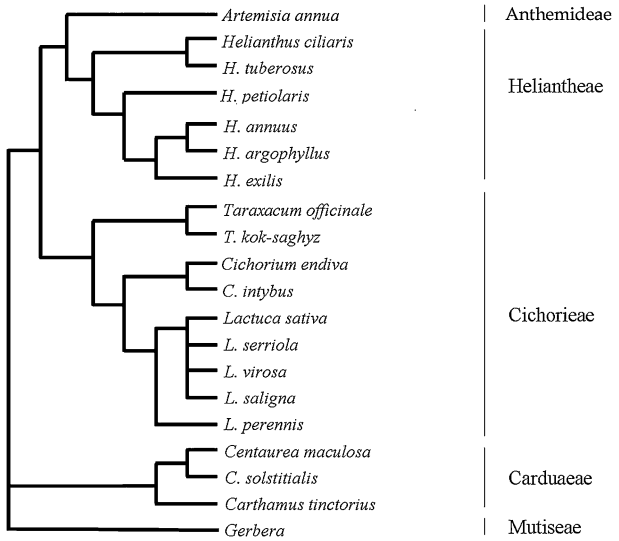

FIG. 1. Phylogenetic relationships among the species used, adapted from Rieseberg (1991), Funk et al. (2005), and Timme et al. (2007).

Other ESTs were downloaded from GenBank, including all large (greater than 15,000 sequences) EST libraries from the Brassicaceae, Pinaceae, Fabaceae, Poaceae, and Solanaceae, and the predicted coding sequences from all whole chloroplast genomes available from five clades: Campanulids, Brassicaceae, Fabaceae, Poaceae, and Solanaceae. These taxa were chosen for the diversity of species and genera with substantial EST resources represented within each group, enabling both micro- and macroevolutionary comparisons within each family.

Ka, Ks, and Likelihood Ratio Tests

A bioinformatic pipeline was developed to identify all reciprocal best hits found in a given set of species (Barker et al. 2010). First, EST sequences were cleaned and clustered into contigs as in Barker et al. (2008), using Seqclean (http://compbio.dfci.harvard.edu/tgi/software/) and the UniVec database (http://www.ncbi.nlm.nih.gov/VecScreen/UniVec.html) to remove vector contamination and using CAP3 as implemented by the TGICL pipeline with default settings (Huang 1996) to assemble these cleaned reads into contigs for each species. These contigs were used in reciprocal all versus all megablast searches in

pairwise comparisons between the species (Altschul et al. 1990, 1997) to identify significant similarities in sequences found. Orthologs were defined conservatively as reciprocal best hits between all pairwise combinations within each comparison made, with a minimum cutoff of 90% sequence similarity over 300 bp. The best protein hit from the Viridiplantae protein database (Wheeler et al. 2005; Barker et al. 2008) was identified using BlastX (Altschul et al. 1997), with a minimum of 80% identity of 100 bp. Because we relied on protein-guided alignments, orthologs without significant protein hits were discarded. Predicted protein sequences for the remaining orthologs were made using the GeneWise hidden Markov models (Birney et al. 2004) implemented in the program Wise2.2 and aligned using MUSCLE 3.6 (Edgar 2004). The corresponding coding DNA sequence alignment was found using the program RevTrans1.4 (Wernersson and Pedersen 2003) and formatted for PAML (Yang 1997, 2007) using the program SEALS (Walker and Koonin 1997). The codeml program from PAML3.15 (http://abacus.gene.ucl.ac.uk/software/paml.html) was used to calculate divergence at synonymous (Ks) and nonsynonymous (Ka) sites for each ortholog set as well as the ratio between these statistics, Ka/Ks.

The diversity of available resources in this family enables a number of interesting comparisons. Evolutionary rates (Ka and Ks) and rapidly evolving genes possibly experiencing positive selection (defined as Ka/Ks $>$ 1 or by significant likelihood ratio test [LRT], see below) were obtained for short time frames (domestication or speciation), intermediate time frames (divergence between sections within a genus or divergence between genera within tribes), and longer time frames (divergence between tribes). Short timescale comparisons included the domestication of sunflowers (*H. annuus* wild vs. domesticated) and lettuce (*Lactuca serriola* vs. *L. sativa*) and the divergence of sister or closely related species (*H. annuus* vs. *H. argophyllus*, *H. petiolaris* vs. *H. exilis*, *H. ciliaris* vs. *H. tuberosus*, *Cichorium intybus* vs. *C. endivia*, *Taraxacum officinale* vs. *T. kok-saghyz*, *Centaurea maculosa* vs. *C. solstitialis*). Intermediate timescales included *H. petiolaris* versus *H. tuberosus*, *L. perennis* versus *L. saligna*, *Cichorium* versus *Taraxacum*, and *Centaurea* versus *Carthamus*. The longest timescales included comparisons between the tribes, including *Helianthus* versus *Artemisia* and *Lactuca* versus *Centaurea*.

Patterns identified in the Asteraceae data set were confirmed in EST data sets from five other taxa: the Brassicaceae, Pinaceae, Fabaceae, Poaceae, and Solanaceae, and for choroplast coding sequences from five taxa: Campanulids, Brassicaceae, Fabaceae, Poaceae, and Solanaceae. Short microevolutionary timescales included the divergence of congeneric species: between hundreds of thousands and a few millions of years, for instance, for *Helianthus* (Rieseberg et al. 1991; Strasburg and Rieseberg 2008). Macroevolutionary timescales included the divergence of tribes within major families, or the divergence of families themselves, representing tens of millions to over 100 My (Kellogg 1998; Wikström et al. 2001).

Genes Under Selection

Candidate genes likely to be under positive selection were identified using several different approaches. The codeml program in PAML3.15 was used to perform LRTs, based on the methodology of Yang (1998) and Yang and Nielsen (1998), to detect significant variation in evolutionary rates along different taxonomic lineages. Codons showing evidence of strong positive selection were detected using the sites model of PAML (Nielsen and Yang 1998, Yang et al. 2000), by comparing the likelihood calculated under the codeml null model M7, which does not allow Ka/Ks to be greater than one, to the likelihood calculated under the codeml model M8, which allows the Ka/Ks ratio to be greater than 1.0 (Yang 2007). These tests were run on several different sets of taxa. Lineage-specific selection was examined during the radiation of the annual sunflowers using the comparison of the four annual sunflowers *H. annuus*, *H. argophyllus*, *H. petiolaris*, and *H. exilis*. The radiation of annual lettuce species was examined using *L. sativa*, *L. serriola*, *L. saligna*, and *L. virosa*. The radiation of the Cichorieae was examined using two members of each genus, with *L. sativa*, *L. serriola*, *C. intybus*, *C. endivia*, *T. officinale*, and *T. kok-saghyz*. The whole family was examined using one member of each tribe, *H. annuus*, *L. sativa*, *C. maculosa*, *A. annua*, and *G. hybrida*. Significance was assigned for each test by calculating the likelihood ratio statistic (twice the log-likelihood difference between the two compared models), which yields a χ^2 statistic with critical values of 5.99 and 9.21 at 5% and 1% significance levels, respectively (Yang 2007). To statistically minimize the false discovery rate (FDR), *q* values, rather than *P* values were used, with an FDR cutoff of 0.05, so that fewer than 5% of the genes identified as significant are likely to be statistical false positives (Storey and Tibshirani 2003). In addition to the LRTs, codeml was used to calculate Ka and Ks values for each ortholog set. Significance of the association between rates of evolution in different tribes was assessed using correlation analysis in R 2.4.1 (R Development Core Team 2006).

Protein Function

The best protein hit from *Arabidopsis thaliana* was identified for each ortholog group using BlastX, with a minimum *e* value of 10^{-20} and at least 100 amino acid alignment length (Altschul et al. 1997) allowing annotation for possible functions categorized according to biological function and the cellular localization of protein products based on GO functional groups (Rhee et al. 2003). Differences in the distribution of genes into the various GO functional categories were evaluated statistically using chi-squared tests, with *P* values computed from 100,000 Monte Carlo simulations in R (R Development Core Team 2005). GO categories with residuals of magnitude greater than 2.0 were considered to be major contributors to significant ($P < 0.05$) chi-square statistics.

Radical Amino Acid Substitutions

The four-species ortholog sets, which consist of predicted protein sequences from *H. annuus*, *L. sativa*, *C. endivia*,

and *A. annua*, were used for BLOcks of Amino Acid SUbstitution Matrix (BLOSUM) analysis. Pairwise alignments between the orthologs of *H. annuus* and *A. annua* and between those of *C. endivia* and *L. sativa* were performed using MUSCLE 3.6 (Edgar 2004). For each ortholog pair, the BLOSUM80 scores for the amino acid substitutions (AAS) indicated by alignment (excluding the gap positions) were summed up and averaged over the total number of AASs to produce a measure that reflects the per-site likelihood of AAS for that ortholog. Relationships between BLOSUM80 scores and divergence as measured by Ka/Ks ratios were assessed by regressing the average BLOSUM80 scores against Ka/Ks for each phylogenetically independent comparison.

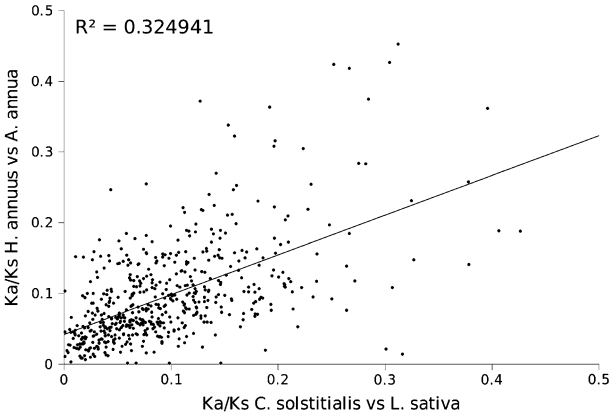

FIG. 2. Regression of Ka/Ks values for pairwise comparisons, macroevolution in the Asteraceae: Heliantheae versus Cichorieae (*x* axis), Cardueae versus Mutiseae (*y* axis). $P < 10^{-15}$, $R^2 = 0.33$.

Results

Ka, Ks, and LRTs

Over long evolutionary timescales, the vast majority of protein-coding genes appeared to be under strong purifying selection. A total of 1650 orthologous protein-coding genes were identified in one species from each of the three tribes represented in the Compositate Genome Project Database, including sunflower *H. annuus* (from the Heliantheae), lettuce *L. sativa* (from the Cichorieae), and safflower *Carthamus tinctorius* (from the Cardueae). Low Ka/Ks ratios in most of these genes suggest strong purifying selection. Of these 1,650 genes, only 548 had an identifiable ortholog in *Artemisia*, the only member of the Anthemideae in our data set. Nonetheless, adding this taxon provides an interesting data set with many of the same characteristics of the larger three-species ortholog set. Nearly, all orthologs in this four-species set showed strong evidence of purifying selection, with average Ka/Ks ratios of 0.10 over all genes. Only one gene (with homology to AT1G12410, a gene with no known function which was also found to be evolving rapidly over shorter time periods in *Lactuca* species) showed Ka/Ks greater than one in one lineage based on LRT. As in the comparison of only three tribes, genes that showed relatively high Ka/Ks in one pairwise comparison showed similarly high levels in other pairwise comparisons (fig. 2). The correlation between Ka/Ks in the *Lactuca* versus *Carthamus* and Ka/Ks in the *Artemisa* versus *Helianthus* was significant ($P < 10^{-16}$, Pearson's $R^2 = 0.32$). There were relatively few ESTs available for *Gerbera*, and all were from floral tissue (Laitinen et al. 2005), so the Mutiseae provided far fewer orthologs. Nevertheless, very similar patterns were found when examining the 249 putative orthologs identified between *Helianthus*, *Lactuca*, *Centaurea*, and *Gerbera* (supplementary fig. S1, Supplementary Material online, $P < 10^{-15}$, $R^2 = 0.37$). Thus, selection appears to have been similar in all four lineages over these longer evolutionary time periods, and it is predominantly purifying.

However, when examining evolution over shorter timescales, we see evidence of many more genes with high Ka/Ks ratios (roughly 4%; see below), and nearly, all of this putatively positive selection appears to be lineage specific. For instance, when looking at pairwise comparisons between the domesticated species and their wild relatives in the two genera, there is no relationship between the rate of ortholog evolution across multiple lineages, based on either Ka (fig. 3*a*, $R^2 = 0.0036$, Not significant, NS) or Ka/Ks ($R^2 = 0.0032$, NS).

Likewise, we see little overlap during divergence of wild species. There is no significant correlation between Ka for *H. annuus* versus *H. argophyllus*, two sister species of annual sunflower (Timme et al. 2007), and Ka for *H. ciliaris* versus *H. tuberosus*, two perennial sunflowers (fig. 3*b*, NS, $R^2 = 0.004$). There is also no correlation between two different speciation events in the Cichorieae, *L. sativa*, and *L. serriola* as compared with *C. endivia* (fig. 3*c*, NS, $R^2 = 0.0025$). However, when we move out to the genus level, there is a more significant relationship, although not as strong as between tribes (supplementary fig. S1, Supplementary Material online, $P < 10^{-16}$, $R^2 = 0.085$). There are only two tribes with more than one genus in the current data set, so this is the only comparison that can be made. Nonetheless, it is consistent with increasing concordance in the evolutionary rates of genes at greater taxonomic distances.

These patterns were also found in the other taxonomic groups examined. In every macroevolutionary comparison but one (Solanaceae), rates of evolution were correlated such that genes evolving rapidly in one lineage also evolved rapidly in the other lineage examined for both nuclear ESTs and chloroplast-encoded coding sequence (fig. 4, nuclear; supplementary fig. S1, Supplementary Material online, chloroplast). In contrast, there was little evidence for concordance over shorter periods of time. Of the ten microevolutionary comparisons made with ESTs and six microevolutionary comparisons made with chloroplast coding sequence, only comparison one was significant, and with a very low R^2 (Solanaceae ESTs, $R^2 = 0.06$, $P < 0.00001$, fig. 5, nuclear; supplementary fig. S4, Supplementary Material online, chloroplast). The Pinaceae was removed from this analysis as most of the orthologs showed no divergence at nonsynonymous sites in at least one comparison, making calculation of Ka/Ks impossible.

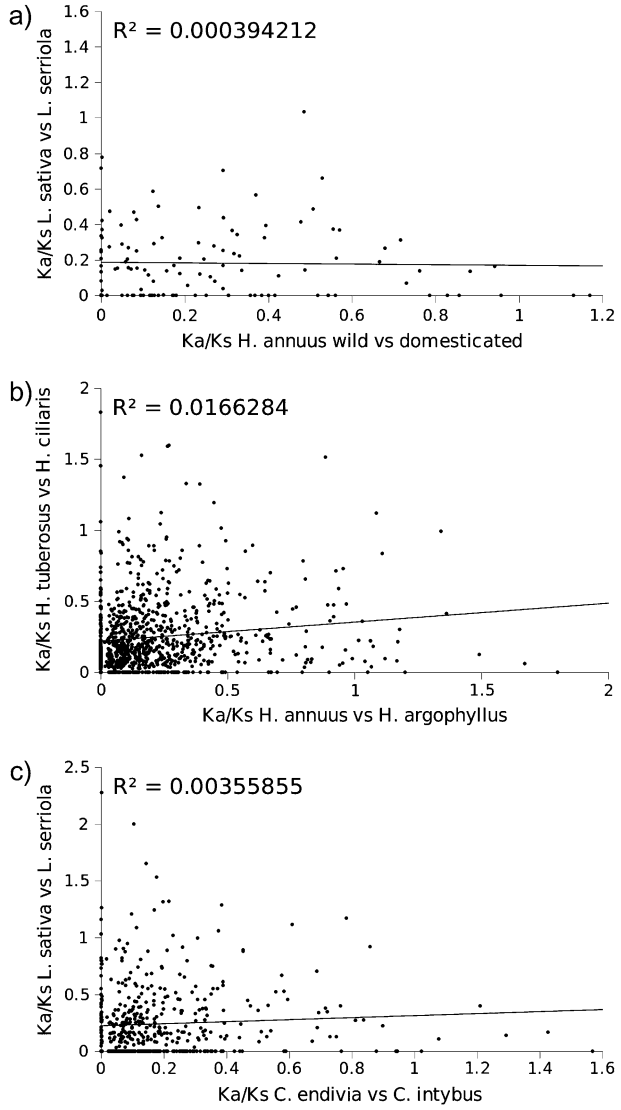

Fig. 3. Regression of Ka/Ks values for pairwise comparisons, microevolution in the Asteraceae: (*a*) *Helianthus annuus* wild versus domesticated (*x* axis) and *Lactuca sativa* versus *L. serriola* (*y* axis), NS, $R^2 = 0.004$. (*b*) *H. annuus* versus *H. argophyllus* (*x* axis) and *H. ciliaris* versus *H. tuberosus* (*y* axis), NS, $R^2 = 0.017$. (*c*) Regression of Ka values for pairwise comparisons: *Cichorium endivia* versus *C. intybus* (*x* axis) and *L. sativa* versus *L. serriola* (*y* axis), NS, $R^2 = 0.004$.

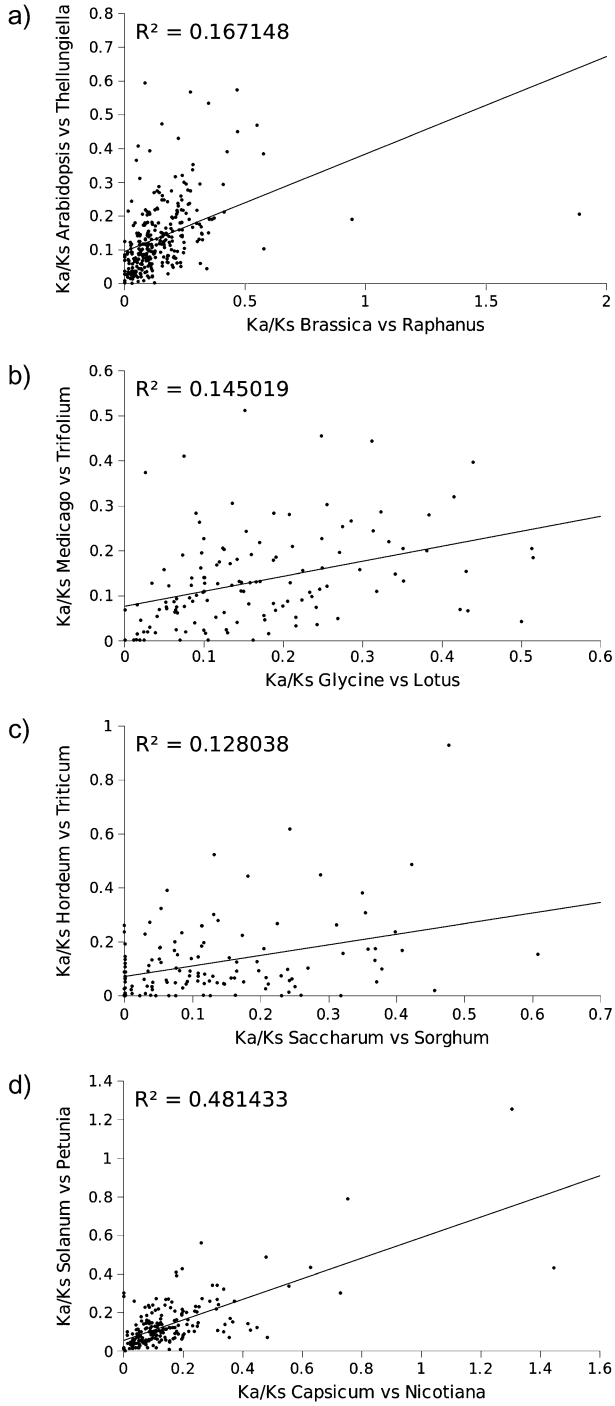

Fig. 4. Regression of Ka/Ks values for pairwise comparisons, macroevolution in the Brassicaceae (*a*), Fabaceae (*b*), Poaceae (*c*), and Solanaceae(*d*).

The LRTs identified highly significant genes in every Asteraceae comparison made (supplementary table S1, Supplementary Material online). The patterns are largely similar to those seen in the Ka/Ks comparisons: very little overlap among rapidly evolving genes across different speciation events. Of 928 orthologs identified among the four annual sunflowers examined (*H. annuus*, *H. argophyllus*, *H. exilis*, and *H. petiolaris*), 37 (4.0%) showed significant evidence for rapid protein evolution in the domesticated lineage based on the LRTs ($q < 0.05$). In the four annual lettuce lineages (*L. sativa*, *L. serriola*, *L. virosa*, and *L. saligna*), 68 (4.2%) of 1,589 orthologs identified show significant evidence for positive selection during domestication based on LRTs ($q < 0.05$), and none of them are shared with *Helianthus*. Of the 1,853 orthologs identified across the three genera from the Cichorieae, *Lactuca*, *Cichorium*, and *Taraxacum*, 52 showed significant ($q < 0.05$) LRT-based tests for selection.

Protein Function

Given the high level of correlation between the rates of evolution of specific genes in different macroevolutionary comparisons, it was not surprising to see a high degree

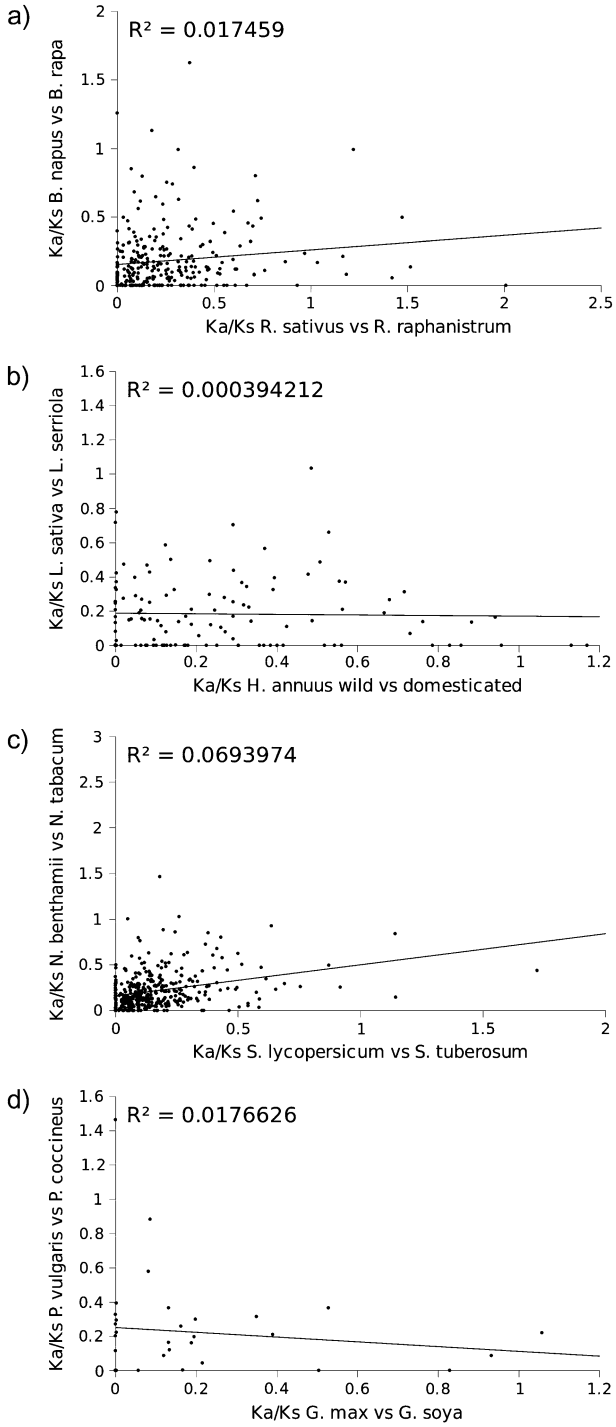

Fig. 5. Regression of Ka/Ks values for pairwise comparisons, microevolution in the Brassicaceae (*a*), Fabaceae (*b*), Poaceae (*c*), and Solanaceae (*d*).

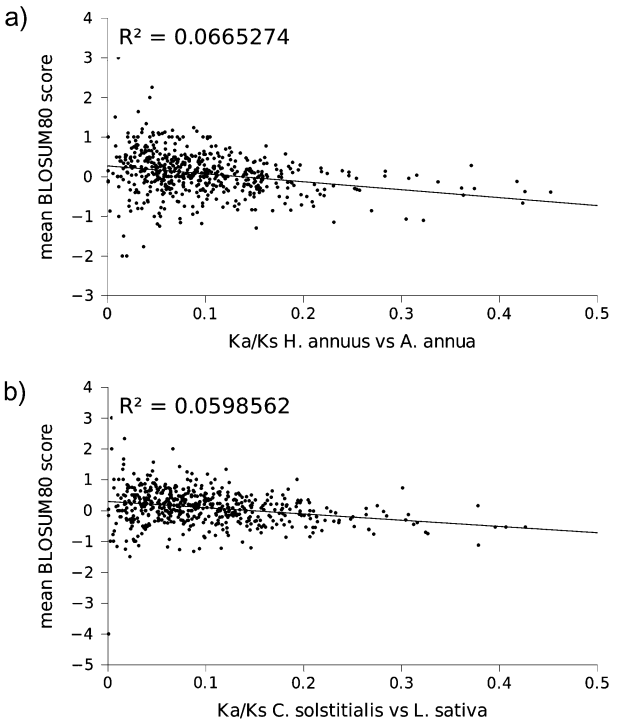

Fig. 6. Regression of Ka/Ks values between Asteraceae tribes and average BLOSUM scores (see text). (*a*) Heliantheae versus Anthemidieae. (*b*) Cichorieae versus Cardueae.

of overlap in the categories of genes evolving most rapidly or most slowly, based on GO annotations (supplementary figs. S2 and S3, Supplementary Material online). There were significant differences between the gene distributions in the highest 5% Ka/Ks fraction and the lowest 5% fraction in both comparisons, Helianthae versus Cichorieae ($\chi^2 = 59.2$, 15 degrees of freedom [df], $P < 10^{-7}$) and Cardueae versus Mutiseae ($\chi^2 = 59.2$, 14 df, $P < 0.01$). The genes evolving most slowly included a high proportion of genes that contribute to molecular structures (particularly ribosomal proteins), and few in membrane-bound proteins, nuclear genes targeted to the chloroplast, and genes with unknown function (supplementary figs. S2 and S3, Supplementary Material online). The genes evolving most rapidly included an overrepresentation of nuclear genes targeted to the chloroplast as well as membrane components (supplementary fig. S2, Supplementary Material online). At the microevolutionary level, there was much less consistency in which genes had many AASs and which had the fewest. Nonetheless, it should be noted that similar classes of genes show the highest levels of constraint based on GO annotations, including a consistently high proportion of very slowly evolving ribosomal proteins, just as in the macroevolutionary comparisons (supplementary fig. S3, Supplementary Material online). Similarly, the genes showing significant evidence for positive selection (supplementary fig. S4, Supplementary Material online) show functional similarities to the rapidly evolving genes in the macroevolutionary analysis, with few genes that contribute to molecular structures, many membrane-bound genes, and a relatively large fraction targeted to the organelles.

Radical AASs

The results of the BLOSUM analysis (fig. 6*a* and *b*) indicate that genes that more rapidly accumulate AASs (higher Ka/Ks ratios) over macroevolutionary time also tend to have more unlikely substitutions (negative BLOSUM scores), which are more likely to bear significant effects on protein

structure and function (Henikoff S and Henikoff JG 1992). In contrast, the more slowly evolving genes (lower Ka/Ks ratios) tended to have less radical substitutions (more positive BLOSUM scores). This is consistent with other evidence for purifying selection at these loci, allowing only conserved substitutions to occur.

Discussion

Micro- Versus Macroevolutionary Change

We used a variety of bioinformatic tools to examine relationship between the identity and putative function of genes and evolutionary rate over shorter and longer time periods. Over macroevolutionary time (during the divergence of genera and tribes), the genes evolving most rapidly or most slowly were highly predictable and quite consistent (figs. 2 and 4). There was little evidence for positive selection over these longer periods; purifying selection appeared to predominate. Moreover, genes evolving more rapidly also tended to have more drastic AASs (fig. 6), again consistent with weaker purifying selection. Genes with more selective constraints such as ribosomal proteins and other genes that contribute to molecular structures consistently evolved more slowly, as expected (Hirsh and Fraser 2001; Degnan et al. 2005).

At a microevolutionary timescale, purifying selection also was the norm, and genes with structural molecular activity exhibited the strongest constraints. However, there was always some evidence for strong positive selection. Candidate genes under selection were identified in every speciation or domestication event examined, but there was generally very little overlap between any two speciation or domestication events in the identity of these genes. Nonetheless, particular classes of genes did recur, such as nuclear-encoded organellar proteins, membrane proteins, and hydrolases (supplementary table S1, Supplementary Material online). The lack of consistency in rates of evolution of orthologous genes on a microevolutionary timescale appears to be a consequence of both the very low levels of divergence in constrained genes (hence considerable noise) and lack of overlap in positively selected genes in different lineages.

When all comparisons are looked at together, a clear pattern emerges: short-term comparisons (with low K_S) show no conserved constraints, whereas deeper comparisons show very significant constraints in the rates of evolution (fig. 7). This association is highly significant ($R^2 = 0.64$; $P < 0.00001$), strongly supporting our hypothesis that short-term patterns are dominated by less predictable forces, whereas long-term evolution is more affected by relatively constant purifying selection.

Our results indicate that both the strength and targets of purifying selection remain consistent and predictable over long timescales. This further implies that for many genes, their evolutionary rates on a macroevolutionary scale are controlled primarily by internal genetic selection pressures and constraints, which appear to be remarkably uniform despite large changes in genetic background and external environment. In contrast, the identity of positively selected genes appears to be idiosyncratic, and any given gene is rarely targeted by positive selection for very long. One explanation for this pattern is that positive selection may result more from external and highly variable environmental pressures than does purifying selection. Other explanations include the possibility that positive selection ceases once a new evolutionary optimum is reached and/or that there are abundant means (i.e., genetic pathways) by which the same evolutionary optimum can be attained. It is also possible that part of the pattern may be due to ongoing resolution of duplicate pairs in closely related taxa. Compared with distantly related organisms, the closely related organisms will generally share more recently duplicated genes, which are known to evolve more rapidly than single-copy genes (Gu et al. 2004).

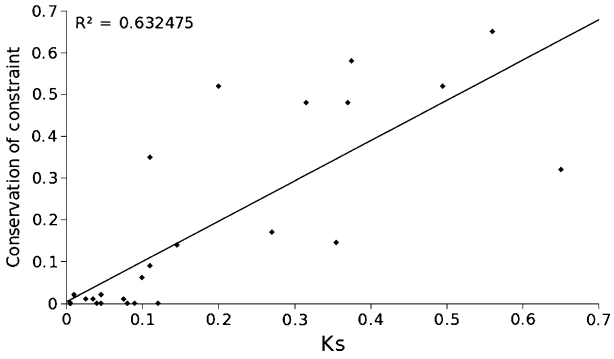

FIG. 7. Regression showing conservation of constraints, (i.e., regression coefficient, R), against mean Ks for the 25 comparisons analyzed. $R^2 = 0.64$, $P < 0.00001$.

So does macroevolution represent an extension of microevolution? Yes, but not necessarily in the manner envisioned by Darwin or by the architects of the neo-Darwinian synthesis, in which large-scale change was thought to be accomplished through the slow and steady accumulation of many small changes. Instead, our data imply that macroevolution consists of the continuous accumulation of many, largely independent episodes of microevolutionary change, each driven by strong positive selection and mostly acting on different genes. However, this occurs against a background of constant change in the vast majority of genes, dictated by mutation rates and internal selection pressures.

One caveat is that the ESTs used only represent a subset of the genes in the genome. Thus, some genes may have a long history of positive selection but were not included in our study. Indeed, this has been reported for reproductive proteins in plants (Richman and Kohn 2000) and animals (Metz et al. 1998; Swanson et al. 2001). Also, the "unknown function" category often contained the largest fraction of rapidly evolving genes, perhaps implying that long-term rapid evolution of these genes has obscured their identities. However, these problems would not affect our chloroplast comparisons, which are based on well-annotated fully-sequenced chloroplast genomes. Orthology is easy to ascertain for these genes. The chloroplast data sets show the same patterns as the ESTs, indicating that these

findings are not due solely to difficulties in working with EST data.

Genes Under Selection in the Asteraceae

The number of loci potentially under selection in the Asteraceae is much higher than found in most other taxa (Gossman et al. 2010), probably because the comparisons made here involve many species with relatively high effective population sizes (Strasburg and Rieseberg 2008). The functions of genes under positive selection included numerous interesting candidate genes. In particular, quite a few positively selected genes showed homology to *Arabidopsis* loci involved in lipid biosynthesis (see supplementary table S1, Supplementary Material online), including AT3G60340, AT2G19010, AT1G24360, AT2G43710, AT3G12120, AT5G03610, and AT3G04290 (Slabas et al. 1992). The oil content and composition of seeds from cultivated sunflower (Burke et al. 2005), lettuce (Vries 1997), and several other crops within the Asteraceae (Chapman and Burke 2007) have been important aspects of their domestication and also have crucial ecological functions in the wild (Linder 2000). Another interesting class of candidate genes is those with homology to genes involved in interactions between the nucleus and either the mitochondrion or the chloroplast, a set of 31 genes including AT2G22250 and AT1G26460, which can cause maternal-effect embryonic lethality (Pagnussat et al. 2005). Other genes may underlie species-specific characteristics, such as a gene under divergent selection in the Cichorieae genera showing homology to AT1G13180, which affects trichome morphology (Szymanski 2005). Additionally, putative transcription factors were well represented, including a gene with homology to NGA1 (AT2G46870), a regulator of floral and leaf development (Alvarez et al. 2006). Finally, a number of genes with evidence of positive selection showed homology to genes involved in disease resistance: AT3G56400 (Li et al. 2004), AT4G22300 (Cunnac et al. 2007), or stress responses: AT1G75280 (Babiychuk et al. 1995).

The Value of Genomic Data

One approach to understanding the genetic basis of evolution and adaptation is to identify genes under selection. Many recent studies have used population genetic approaches, such as hitchhiking mapping (Schlotterer 2002; Storz 2005; Vasemägi et al. 2005; Kane and Rieseberg 2007, 2008; Yatabe et al. 2007; Kane et al. 2009; Nosil et al. 2009; Tonteri et al. 2010) or genomic scans (Chapman et al. 2008) to identify candidate genes experiencing positive selection. However, this approach involves genotyping dozens, hundreds or even thousands of individuals, and can only be used to examine a small fraction of the genome. In taxa whose entire genomes have been sequenced, comparative genomics has led to substantial insights into the genes underlying reproductive isolation (Orr 2005) and adaptation (International Mouse Genome Sequencing Consortium 2002; Drosophila 12 Genomes Consortium 2007) as well as leading to a better understanding of the genetic material under selective constraint (Oeltjen et al. 1997; McGuire et al. 2000; Kellis et al. 2003; Stein et al. 2003; Stark et al. 2007). However, little has been done to exploit the potential of EST databases, which are much less expensive than whole genomes and are available for a wide variety of organisms (Benson et al. 2007). The present study illustrates how these EST databases can be used to make inferences about genome-wide rates of evolutionary change, thereby expanding the range of organisms available for comparative genomic studies. The fact that we see the same patterns in analysis using whole chloroplast genomes, which are very well annotated, demonstrates that the patterns are robust and not due to artifacts related to EST sequencing, our transcriptome assemblies or ortholog detection.

Conclusions

From this analysis, it does appear that, at least in some respects, macroevolution looks very much like "repeated rounds of microevolution" (Erwin 2000). However, macroevolutionary change cannot be easily predicted by analysis of any one microevolutionary event. The genes evolving most rapidly over the short term are a subset of the genes evolving rapidly over the long term, and different short-term evolutionary changes overlap only partially. The variation between different microevolutionary events is most likely at least partially due to external environmental factors as well as neutral processes. In contrast, long-term macroevolutionary patterns may be largely driven by internal genomic selection pressures and constraints, such as functional constraints, multigene interactions, and intergenomic conflicts. Nevertheless, there is a substantial amount of variation not explained by these constraints even in very distant comparisons, indicating that purifying selection is not entirely predictable.

Supplementary Material

Supplementary table S1 and figures S1–S4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

Thanks to J. Bever, I. Mayrose, M. Lynch, J. Palmer, J. Hey, and two anonymous reviewers for helpful comments. This work was supported by National Science Foundation Plant Genome Awards 0421630 and 0820451 to L.H.R. M.S.B. was supported by a NSERC-BRITE postdoctoral fellowship.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*. 215:403–410.

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipam DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 25:3389–3402.

Alvarez JP, Pekker I, Goldschidt A, Blum E, Amsellem Z, Eshed Y. 2006. Endogenous and synthetic microRNAs stimulate simultaneous, efficient, and localized regulation of multiple targets in diverse species. *Plant Cell* 18:1134–1151.

Babiychuk E, Kushnir S, Belles-Boix E, Van Montagu M, Inzé D. 1995. *Arabidopsis thaliana* NADPH oxidoreductase homologs confer tolerance of yeasts toward the thiol-oxidizing drug diamide. *J Biol Chem*. 270:26224–26231.

Barker MS, Dlugosch KM, Dinh L, Challa S, Kane NC, King MG, Rieseberg LH. 2010. EvoPipes.net: bioinformatic pipelines and forums for ecological and evolutionary genomics. *Evol Bioinform Online*. 6:143–149.

Barker MS, Kane NC, Kozik A, Michelmore RW, Knapp SJ, Kesseli RK, Still DW, Bradford KJ, Rieseberg LH. 2008. Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. *Mol Biol Evol*. 25:2445–2455.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. 2007. Genbank. *Nucleic Acids Res*. 35:D21–D25.

Birney E, Clamp M, Durbin R. 2004. GeneWise and Genomewise. *Genome Res*. 14:988–995.

Boag PT, Grant PR. 1981. Intense natural selection in a population of Darwin's Finches. Geospizinae. in the Galápagos. *Science* 214:82–85.

Burke JM, Knapp SJ, Rieseberg LH. 2005. Genetic consequences of selection during the evolution of cultivated sunflower. *Genetics* 171:1933–1940.

Carroll RL. 2000. Towards a new evolutionary synthesis. *Trends Ecol Evol*. 15:27–32.

Carroll SB. 2001. The big picture. *Nature* 409:669.

Chapman MA, Burke JM. 2007. DNA sequence diversity and the origin of cultivated safflower(*Carthamus tinctorius* L.; Asteraceae). *BMC Plant Biol*. 7:60.

Chapman MA, Pashley CH, Wenzler J, Hvala J, Tang S, Knapp SJ, Burke JM. 2008. A genomic scan for selection reveals candidates for genes involved in the evolution of cultivated sunflower (*Helianthus annuus*). *Plant Cell*. 20:2931–2945.

Charlesworth B, Lande R, Slatkin M. 1982. A neo-Darwinian commentary on macroevolution. *Evolution*. 36:474–498.

Cronquist A. 1981. An integrated system of classification of flowering plants. New York: Columbia University Press.

Cunnac S, Wilson A, Nuwer J, Kirik A, Baranage G, Mudgett MB. 2007. A conserved carboxylesterase is a SUPPRESSOR OF AVRBST-ELICITED RESISTANCE in *Arabidopsis*. *Plant Cell* 19:688–705.

Darwin CC. 1859. The origin of the species by means of natural selection or the preservation of favoured races in the struggle for life. London: John Murray.

Degnan PH, Lazarus AB, Wernegreen JJ. 2005. Genome sequence of *Blochmannia pennsylvanicus* indicates parallel evolutionary trends among bacterial mutualists of insects. *Genome Res*. 15:1023–1033.

Dobzhansky T. 1937. Genetics and the origin of species. New York: Columbia University Press.

Dobzhansky T. 1951. Genetics and the origin of species, 3rd ed. New York: Columbia University Press.

Drosophila 12 Genomes Consortium. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203–218.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32:1792–1797.

Erwin DH. 1999. The origin of bodyplans. *Am Zool*. 39:617–629.

Erwin DH. 2000. Macroevolution is more than repeated rounds of microevolution. *Evol Dev*. 2:78–84.

Filchak KE, Roethele JB, Feder JL. 2000. Natural selection and sympatric divergence in the apple maggot *Rhagoletis pomonella*. *Nature* 407:739–742.

Funk VA, Bayer RJ, Keeley S, et al. (12 co-authors). 2005. Everywhere but Antarctica: using a supertree to understand the diversity and distribution of the Compositae. *Biol Skr*. 55:343–374.

Gossman TI, Song B, Windsor AJ, Mitchell-Olds T, Dixon CJ, Kapralov MV, Filatov DA, Eyre-Walker A. 2010. Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Mol Biol Evol*. 27:1822–1832.

Gould SJ, Eldridge N. 1977. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* 3:115–151.

Gould SJ, Lewontin RC. 1979. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc R Soc Lond B Biol Sci*. 205:581–598.

Grant PR, Grant BR. 1997. Genetics and the origin of bird species. *Proc Natl Acad Sci U S A*. 94:7768–7775.

Gu Z, Rifkin SA, White KP, Li W. 2004. Duplicate genes increase gene expression diversity within and between species. *Nat Genet*. 36:577–579.

Hansen TF, Martins EP. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* 50:1404–1417.

Heesacker A, Kishore VK, Gao W, et al. (12 co-authors). 2008. SSRs and INDELs mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility. *Theor Appl Genet*. 117:1021–1029.

Henikoff S, Henikoff JG. 1992. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A*. 89:10915–10919.

Heywood VH. 1978. Flowering plants of the world. New York: Mayflower Books.

Hirsh AE, Fraser HB. 2001. Protein dispensability and rate of evolution. *Nature* 411:1046.

Huang X. 1996. An improved sequence assembly program. *Genomics* 33:21–31.

Hunter JP, Jernvall J. 1995. The hypocone as a key innovation in mammalian evolution. *Proc Natl Acad Sci U S A*. 92:10718–10722.

International Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–562.

Jablonski D. 2000. Micro- and macroevolution: scale and hierarchy in evolutionary biology and paleobiology. *Paleobiology* 26:S15–S52.

Janick J. 2003. Plant breeding reviews, Volume 24:Part1: long-term selection: maize. Part 2: long-term selection: crops, animals, and bacteria. Hoboken (NJ): Wiley.

Kane NC, King M, Barker MS, Raduski A, Karrenberg S, Yatabe Y, Knapp SJ, Rieseberg LH. 2009. Comparative genomic and population genetic analyses indicate highly porous genomes and high levels of gene flow between divergent *Helianthus* species. *Evolution* 63:2061–2075.

Kane NC, Rieseberg LH. 2007. Selective sweeps reveal candidate genes for adaptation to drought and salt tolerance in common sunflower, *Helianthus annuus*. *Genetics* 175: 1823–1824.

Kane NC, Rieseberg LH. 2008. Genetics and the evolution of weediness in *Helianthus annuus*. *Mol Ecol*. 17:384–394.

Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423:241–254.

Kellogg EA. 1998. Relationships of cereal crops and other grasses. *Proc Natl Acad Sci U S A*. 95:2005–2010.

Kirkpatrick M. 1982. Quantum evolution and punctuated equilibria in continuous genetic characters. *Am Nat*. 119:833–848.

Laitinen RA, Immanen J, Auvinen P, et al. (11 co-authors). 2005. Analysis of the floral transcriptome uncovers new regulators of organ determination and gene families related to flower organ differentiation in *Gerbera hybrida* (Asteraceae). *Genome Res*. 15:475–486.

Leroi AM. 2000. The scale independence of evolution. *Evol Dev*. 2:67–77.

Li J, Brader G, Palva ET. 2004. The WRKY70 transcription factor: a node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense. *Plant Cell* 16:319–331.

Linder CR. 2000. Adaptive evolution of seed oils in plants: accounting for the biogeographic distribution of saturated and unsaturated fatty acids in seed oils. *Am Nat*. 156:442–458.

Losos JB, Warheitt KI, Schoener TW. 1997. Adaptive differentiation following experimental island colonization in *Anolis* lizards. *Nature* 387:70.

Lynch M. 1990. The rate of morphological evolution in mammals from the standpoint of neutral expectation. *Am Nat*. 136:727–741.

McGuire AM, Hughes JD, Church GM. 2000. Conservation of DNA regulatory motifs and discovery of new motifs in microbial genomes. *Genome Res*. 10:744–757.

Metz EC, Robles-Sikisaka R, Vacquier VD. 1998. Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. *Proc Natl Acad Sci U S A*. 95:10676–10681.

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

Nosil P, Funk DJ, Ortiz-Barrientos D. 2009. Divergent selection and heterogeneous genomic divergence. *Mol Ecol*. 18:375–402.

Oeltjen JC, Malley TM, Muzny DM, Miler W, Gibbs RA, Belmont JW. 1997. Large-scale comparative sequence analysis of the human and murine Bruton's tyrosine kinase loci reveals conserved regulatory domains. *Genome Res*. 7:315–329.

Orr HA. 2005. The genetic basis of reproductive isolation: insights from *Drosophila*. *Proc Natl Acad Sci U S A*. 102:6522–6526.

Pagnussat GC, Yu HJ, Ngo QA, Rajani S, Mayalagu S, Johnson CS, Capron A, Xie LF, Ye D, Sundaresan V. 2005. Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis*. *Development* 132:603–614.

R Development Core Team. 2006. R: a language and environment for statistical computing [Internet]. Vienna (Austria): R Foundation for Statistical Computing.ISBN 3-900051-07-0. Available from: http://www.R-project.org.

Reznick DE, Ricklefs RE. 2009. Darwin's bridge between microevolution and macroevolution. *Nature* 457:837–842.

Reznick DN, Shaw FH, Rodd FH, Shaw RG. 1997. Evaluation of the rate of evolution in natural populations of guppies (*Poecilia reticulata*). *Science* 275:1934–1937.

Rhee SY, Beavis W, Berardini TZ, Chen G, et al. 2003. The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucleic Acids Res*. 31:224.

Richman AD, Kohn JR. 2000. Evolutionary genetics of self incompatibility in the Solanaceae. *Plant Mol Biol*. 42:169–179.

Rieseberg LH. 1991. Homoploid reticulate evolution in *Helianthus*. Asteraceae.—evidence from ribosomal genes. *Am J Bot*. 78:1218–1237.

Rieseberg LH, Beckstrom-Sternberg SM, Liston A, Arias DM. 1991. Phylogenetic and systematic inferences from chloroplast DNA and isozyme variation in *Helianthus* Sect *Helianthus* (Asteraceae). *Syst Bot*. 16:50–76.

Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, Durphy JL, Schwarzbach AE, Donovan LA, Lexer C. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301:1211–1216.

Rieseberg LH, Sinervo B, Linder CR, Ungerer MC, Arias DM. 1996. Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. *Science* 272:741–745.

Rieseberg LH, Widmer A, Arntz AM, Burke JM. 2003. The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. *Philos Trans R Soc B Biol Sci*. 358:1141–1147.

Schlotterer C. 2002. A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics* 160:753–763.

Slabas AR, Chase D, Nishida I, Murata N, Sidebottom C, Safford R, Sheldon PS, Kekwick RG, Hardie DG, Mackintosh RW. 1992. Molecular cloning of higher-plant 3-oxoacyl-(acyl carrier protein) reductase. Sequence identities with the *nodG*-gene product of the nitrogen-fixing soil bacterium *Rhizobium meliloti*. *Biochem J*. 15:321–326.

Stark A, Lin MF, Kheradpour P, Pedersen JS, et al. 2007. Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* 450:219–232.

Stein LD, Bao Z, Blasiar D, et al. 2003. The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. *PLoS Biol*. 1:e45.

Stern DL. 2000. Evolutionary developmental biology and the problem of variation. *Evolution* 54:1079–1091.

Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A*. 100:9440–9445.

Storz JF. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol*. 14:671–688.

Strasburg JL, Rieseberg LH. 2008. Molecular demographic history of the annual sunflowers *Helianthus annuus* and *H. petiolaris*—large effective population sizes and rates of long-term gene flow. *Evolution* 62:1936–1950.

Swanson WJ, Yang Z, Wolfner MF, Aquadro CF. 2001. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc Natl Acad Sci U S A*. 98:2509–2514.

Szymanski DB. 2005. Breaking the WAVE complex: the point of *Arabidopsis* trichomes. *Curr Opin Plant Biol*. 8:103–112.

Timme RE, Simpson BB, Linder CR. 2007. High-resolution phylogeny for *Helianthus* (Asteraceae) using the 18S-26S ribosomal DNA external transcribed spacer. *Am J Bot*. 94:1837–1852.

Tonteri A, Vasemägi A, Lumme J, Primmer CR. 2010. Beyond MHC: signals of elevated selection pressure on Atlantic salmon (Salmo salar) immune-relevant loci. *Mol Ecol*. 19:1273–1282.

Vasemägi A, Nilsson J, Primmer CR. 2005. Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Mol Biol Evol*. 22:1067–1076.

Vrba ES, Gould SJ. 1997. The hierarchical expansion of sorting and selection: sorting and selection cannot be equated. *Paleobiology* 12:217–228.

Vries IM de. 1997. Origin and domestication of Lactuca sativa L. *Genet Resour Crop Evol*. 44:165–174.

Walker DR, Koonin EV. 1997. SEALS: a system for easy analysis of lots of sequences. *Proc Int Conf Intell Syst Mol Biol*. 5:333–339.

Wernersson R, Pedersen AG. 2003. RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res*. 31:3537–3539.

Wheeler DL, Smith-White B, Chetvernin V, Resenchuk S, Dombrowski SM, Pechous SW, Tatusova T, Ostell J. 2005. Plant genome resources at the National Center for Biotechnology Information. *Plant Phys*. 138:1280–1288.

Wikström N, Savolainen V, Chase MW. 2001. Evolution of the angiosperms: calibrating the family tree. *Proc R Soc Lond B Biol Sci*. 268:2211–2220.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*. 13:555–556.

Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol*. 15:568–573.

Yang Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.

Yang Z, Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J Mol Evol*. 46:409–418.

Yang Z, Nielsen R, Goldman N, Pedersen A-MK. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449.

Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH. 2007. Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics* 175:1883–1893.