Theor Appl Genet
DOI 10.1007/s00122-007-0605-2

ORIGINAL PAPER

Universal markers for comparative mapping and phylogenetic analysis in the Asteraceae (Compositae)

Mark A. Chapman · JianCheng Chang · David Weisman · Rick V. Kesseli · John M. Burke

Received: 3 April 2007 / Accepted: 30 June 2007
© Springer-Verlag 2007

Abstract The development of universal markers that can be assayed across taxa, but which are polymorphic within taxa, can facilitate both comparative map-based studies and phylogenetic analyses. Here we describe the development of such markers for use in the Asteraceae, which includes the crops lettuce, sunflower, and safflower as well as dozens of locally important crop and weed species. Using alignments of a conserved orthologous set (COS) of ESTs from lettuce and sunflower and genomic sequences of *Arabidopsis*, we designed a suite of primer pairs that are conserved across species, but which are predicted to flank introns. We then tested 192 such primer pairs in 8 species from across the family. Of these, 163 produced an amplicon in at least 1 taxon, and 125 amplified in at least half of the taxa surveyed. Thirty-nine amplified in all 8 species. Comparisons amongst sequences within the lettuce and sunflower EST databases indicate that the vast majority of these loci will be polymorphic. As a direct test of the utility of these markers outside the lettuce and sunflower subfamilies, we sequenced a subset of ten loci from a panel of cultivated safflower individuals. All 10 loci proved to be single-locus, and nine of the 10 loci were polymorphic with an average of 12.8 SNPs per kb. Taken together, these loci will provide an initial backbone for comparative genetic analyses within the Asteraceae. Moreover, our results indicate that these loci are phylogenetically informative, and hence can be used to resolve evolutionary relationships between taxa within the family as well as within species.

Communicated by M. Xu.

Mark A. Chapman and JianCheng Chang have contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-007-0605-2) contains supplementary material, which is available to authorized users.

M. A. Chapman (✉) · J. M. Burke
Department of Plant Biology, Miller Plant Sciences Building, University of Georgia, Athens, GA 30602, USA
e-mail: mchapman@plantbio.uga.edu

J. Chang · D. Weisman · R. V. Kesseli
Department of Biology, University of Massachusetts, Boston, 100 Morrissey Blvd, Boston, MA 02125, USA

Introduction

The field of comparative genomics relies upon the identification of orthologous genes and genomic regions between the species of interest. Mapping of these orthologous regions in different species provides insight into the extent of synteny, and can facilitate the map-based cloning of genes of interest (Gale and Devos 1998b; Paterson et al. 2000). Comparative genetic maps have been produced for a number of plant species; the majority of these are for crops with extensive linkage maps often containing thousands of mapped loci. Thus far, the largest efforts have focused on the Solanaceae (Bonierbale et al. 1988; Livingstone et al. 1999; Doganlar et al. 2002), Brassicaceae (Kowalski et al. 1994; Lagercrantz and Lydiate 1996; Lagercrantz et al. 1996; Lagercrantz 1998; Lan and Paterson 2000), Fabaceae (Choi et al. 2006; Kalo et al. 2004), Rosaceae (Dirlewanger et al. 2004), and Poaceae (Lin et al. 1995; Paterson et al. 1995; Maroof et al. 1996; Gale and Devos 1998a; Ming et al. 1998; Wu et al. 2003; Bowers et al. 2005). Recent studies have revealed the potential of extending these analyses to comparisons of taxa from different families; however, such studies are generally restricted to localized regions and/or microsyntenic analyses based on the full

genome sequence of *Arabidopsis thaliana* (e.g., Grant et al. 2000; Ku et al. 2001; Lee et al. 2001; Bowers et al. 2003; Dominguez et al. 2003; Timms et al. 2006).

Early comparative mapping projects relied on hybridization-based assays of probes from one species against the genome of another (Tanksley et al. 1992; Prince et al. 1993; Van Deynze et al. 1998; Fulton et al. 2002). Using this strategy, the pairing gene *Ph* from wheat and flowering time genes in *Brassica* were isolated based on the known genomic locations in other species (Foote et al. 1997; Osborn et al. 1997; Axelsson et al. 2001; Kole et al. 2001; Griffiths et al. 2006). More recently, PCR-based markers have been developed for comparative mapping in a range of species (Fourmann et al. 2002; Choi et al. 2006; Feltus et al. 2006; Fredslund et al. 2006; Wu et al. 2006). Ideally, such markers should utilize primers that anneal to highly conserved regions of genes (e.g., exons) such that the same primers can be used in a range of related species. In addition, these primers should flank a variable region of the genome (e.g., introns) such that polymorphism can be detected within species, allowing the loci to be mapped.

Beyond their utility for comparative genomic analyses, readily-transferable nuclear DNA markers could also be of great value for assessing phylogenetic relationships (Doyle and Gaut 2000; Small et al. 2004). For phylogenetic reconstruction, DNA sequencing is the most widely chosen technique. Universal markers for phylogenetics should be designed with similar goals as those for comparative mapping: the primers must anneal to highly conserved gene regions; however, depending on the scale of the phylogenetic analysis, the internal portion of the amplicon should be more variable. Currently the majority of universal primer pairs available that amplify across families of flowering plants are specific to the chloroplast genome (reviewed in Small et al. 2004).

While comparing chloroplast DNA (cpDNA) gene sequences is useful at higher taxonomic scales (Chase et al. 1993; Kim and Jansen 1995), these genes do not usually provide sufficient resolution for the analysis of more closely related taxa; hence, intergenic chloroplast regions or introns are used (e.g., Taberlet et al. 1991; Hamilton 1999) because of their higher rate of nucleotide substitution (Gielly and Taberlet 1994). There are, however, problems associated with the sole use of organellar markers to resolve phylogenies. For example, the rate of cpDNA sequence evolution is approximately one fourth that of nuclear DNA (Wolfe et al. 1987), and thus even intergenic sequences may not provide sufficient resolution between closely related species. Further, because cpDNA is uniparentally inherited in the majority of species, it cannot resolve relationships in taxa that have evolved via hybridization or allopolyploidy. It is also possible for chloroplast haplotypes to be transferred from one taxon to another via introgressive hybridization (e.g., Soltis and Kuzoff 1995; Okuyama et al. 2005). Finally, the lack of recombination in the chloroplast genome means that a phylogenetic investigation based on multiple cpDNA sequences is little more than an analysis of one 'super-locus'. The use of multiple, unlinked nuclear loci, on the other hand, results in a genome-wide phylogenetic signal that is more likely to reflect true species relationships (Small et al. 2004).

In contrast to a cpDNA-based approach, phylogenetic studies using nuclear DNA sequences have traditionally been hampered by the lack of sequence information available for the design of universal primers and difficulties distinguishing between orthologous and paralogous sequences (Soltis and Soltis 1998; Small et al. 2004). Consequently, nuclear DNA phylogenies have relied heavily on the sequence of the two internal transcribed spacers (ITS) of the nuclear ribosomal DNA (Alvarez and Wendel 2003), with only a handful of other nuclear genes/gene families having been used for phylogenetic inference, although the list is continually growing (e.g., Strand et al. 1997; reviewed in Small et al. 2004).

Previous studies have shown that anchoring primers in conserved orthologs can provide markers for comparative mapping in birds, mammals, and insects (Lyons et al. 1997; Smith et al. 2000; Chambers et al. 2003) and these same markers are suitable for phylogenetic investigations (e.g., Roca et al. 2001; Gaines et al. 2005). The recent availability of the whole genome sequence of *Arabidopsis* and large expressed sequence tag (EST) databases for a growing number of plant species have provided the potential to align DNA sequences from multiple species, and to identify a conserved orthologous set (COS) of low or single-copy genes (Fulton et al. 2002) that could be useful for comparative mapping studies and phylogenetic analyses. Wu et al. (2006) expanded on this concept to identify COS loci from six species of the euasterids, and homology was sufficient within the euasterid I clade (tomato, potato, pepper and coffee) to allow primer design for amplification across species within this clade.

To date, there are no such universal nuclear markers available for use in the Asteraceae (Compositae). This is despite the family comprising one-tenth of all flowering plants, and including a number of major and minor crops such as lettuce, sunflower, safflower, globe artichoke, and chicory, as well as numerous ecologically-important taxa including many weedy and invasive species (Funk et al. 2005; Kesseli and Michelmore 1997). To remedy this situation, we aligned ESTs defined as COS loci from two phylogenetically distant members of the family, sunflower (*Helianthus annuus* L.) and lettuce (*Lactuca sativa* L.), with coding sequences from *Arabidopsis*. Primers were designed to anneal to highly conserved exon regions and to flank putative introns based on their known positions in

Arabidopsis. These primer pairs should be suitable for comparative mapping in crops such as sunflower, lettuce, and safflower or globe artichoke, which represent the three major subfamilies of the Asteraceae, as well as in phylogenetic investigations at lower taxonomic levels.

Materials and methods

Primer design

The publicly available global alignments between sunflower and lettuce ESTs with *Arabidopsis* open reading frames was downloaded from the Compositae Genome Project Database (CGPDB; http://www.cgpdb.ucdavis.edu/COS_Arabidopsis/). The *Arabidopsis* sequence from each alignment was then used in a BLAST search against the full *Arabidopsis* genome database (available from http://www.arabidopsis.org/). Software written for this study in BioPython (Chapman and Chang 2000) was used to map the BLAST output onto the global alignments, thereby allowing for the identification of *Arabidopsis* intron positions. For each tri-species alignment, a FASTA file was generated, with the *Arabidopsis* intron positions annotated per the input specification of the primer design program PriFi (Fredslund et al. 2005). Because intron positions are thought to be highly conserved across species (e.g., Strand et al. 1997; Ku et al. 2000; Roy et al. 2003), the known *Arabidopsis* intron position allowed us to predict intron positions within the lettuce and sunflower sequences. The source code for PriFi was kindly supplied by the authors, and it was used in batch mode (automated, without using the graphical user interface) to process the FASTA files and design conserved primers flanking the putative introns. We modified PriFi parameter settings to generate primer pairs at least 25 bases from the intron/exon boundaries, 15–30 nucleotides in length, with GC contents around 50–60%, and to have annealing temperatures within 10°C of each other. The selected target regions for primers were not allowed to differ between lettuce and sunflower by more than four bases, and the expected amplicon sizes were confined to 200–1,000 bases (assuming that introns were the same as in *Arabidopsis*). Specific information on the source code modifications can be obtained from the authors. Information concerning the primers can be found in the Electronic Supplementary Material Table S1.

PCR amplification

Of the 232 primer pairs designed by PriFi, 192 were tested for amplification across the Asteraceae. Eight DNA samples were chosen to include members of three major subfamilies. Two species each from the subfamilies Carduoideae (*Carthamus tinctorius* L. and *Centaurea maculosa* Lam.) and Cichorioideae (*Lactuca sativa* L. and *Cichorium intybus* L.) were chosen, plus four members of the largest subfamily, the Asteroideae (*Senecio squalidus* L., *Artemisia vulgaris* L., *Helianthus annuus* L. and *Liatris scariosa* (L.) Willd.). Lettuce DNA was supplied by Dr. R. Michelmore (University of California, Davis, USA). *Centaurea* and *Liatris* DNA was extracted using a modified CTAB-based protocol (Doyle and Doyle 1990) from leaf material collected from wild populations in Massachusetts (USA). Seed of sunflower, safflower, chicory and *Artemisia* were obtained from the USDA (http://www.ars-grin.gov/npgs/), whereas *Senecio* seed was supplied by Prof. R. J. Abbott (University of St. Andrews, Scotland). Seed were germinated on damp filter paper and transferred to soil in the University of Georgia greenhouses. DNA was then extracted from leaf tissue using the Qiagen DNeasy protocol (Qiagen, Valencia, CA, USA) following the manufacturer's recommendations.

PCR was carried out in 15 µl total volume containing 15 ng DNA, 30 mM tricine pH 8.4 KOH, 50 mM $MgCl_2$, 100 µM of each dNTP, 0.2 µM of each primer and 1 unit of *Taq* DNA polymerase. The PCR conditions followed a 'touchdown' regime to reduce spurious amplification as follows: 3 min at 95°C; 10 cycles of 30 s at 94°C, 30 s at 60°C and 45 s at 72°C, annealing temperature decreasing to 50°C by 1°C per cycle, followed by 30 cycles of 30 s at 94°C, 30 s at 50°C, 45 s at 72°C, followed by 20 min at 72°C. PCR products and appropriate size standards were resolved on 1.5% agarose, stained with ethidium bromide, and visualized under UV light.

For each primer pair, we scored amplification success in each of the eight taxa, and approximate size (to the nearest 50 bp) of the amplicon (see Table S1). PCR products were presumed single-copy unless an obvious second product was evident, although this was not explicitly tested in most cases (but see below). In some cases two amplicons were obvious and the sizes of both were noted. In each instance, these two values were then averaged prior to our analyses of amplicon size.

To demonstrate the utility of these markers, we carried out two different analyses. First, for the set of loci that could be successfully amplified in all eight taxa, we searched for evidence of polymorphism within the CGPDB. The lettuce assembly was derived from two *Lactuca* accessions, whereas the sunflower assembly was derived from two *Helianthus* accessions (for details see http://www.cgpdb.ucdavis.edu/Library_Construction/). Thus, in the case of loci that were represented by a contig containing both of the lettuce or both of the sunflower genotypes, we were able to check the EST sequences for putative SNPs distinguishing lines within species. While some fraction of these SNPs are undoubtedly the result of sequencing errors,

this approach is likely to underestimate the true level of variation, as these sequences are derived from cDNAs, and thus provide no information on polymorphisms within introns.

Second, to demonstrate the utility of these markers beyond the two taxa (sunflower and lettuce) from which they were derived, we selected ten loci that amplified in members of all three subfamilies and sequenced them in a panel of eight individuals of cultivated safflower (*Carthamus tinctorius*), which belongs to a subfamily (Carduoideae) different from those of either lettuce or sunflower. DNA polymorphism is known to be relatively low within and between species of *Carthamus* (e.g., Vilatersana et al. 2000; Garcia-Jacas et al. 2001); hence, if these ten loci exhibit variation between safflower cultivars, there is likely to be sufficient variation between species of *Carthamus* to resolve phylogenetic relationships. Moreover, the presence of sequence polymorphism would indicate that these markers are suitable for genetic mapping in cultivated safflower. For comparison, the ITS region of the nuclear rRNA genes was sequenced using primers ITS-5 and ITS-4 (Baldwin 1993). PCR was performed as above on DNA extracted from eight accessions obtained from the USDA National Plant Germplasm System (PI numbers: 193473, 250601, 250606, 253762, 271070, 576995, 603207, 610263). PCR products were incubated with 0.5 units Shrimp Alkaline Phosphatase and 5 units Exonuclease I (USB, Cleveland, OH, USA) at 37°C for 45 min to remove primers and excess dNTPs from the product. Sequencing reactions were then performed using BigDye v3.1 and the same primers as used to generate the initial product following the manufacturer's protocol (Applied Biosystems, Foster City, CA, USA) and resolved on an ABI 3730xl sequencer. Chromatograms were inspected by eye and aligned using Genedoc (http://www.psc.edu/biomed/genedoc). Estimates of nucleotide diversity (Watterson's θ and π) were then calculated using DnaSP ver. 4.10.9 (Rozas et al. 2003). DNA sequences have been deposited in GenBank under accession numbers EF483943-EF484030.

Results

Primer screening across taxa

Of the 1,343 alignments that were previously found to fit the definition of COS loci, 232 met our criteria for primer design. Of these, 192 primer pairs arbitrarily selected for inclusion in our study were tested for amplification in eight diverse members of the Asteraceae (Table S1). Most primer pairs tested (163 of 192, 85%) amplified in at least one taxon, and 125 (65%) amplified in at least half of the taxa tested (Fig. 1; Tables 1, S1).

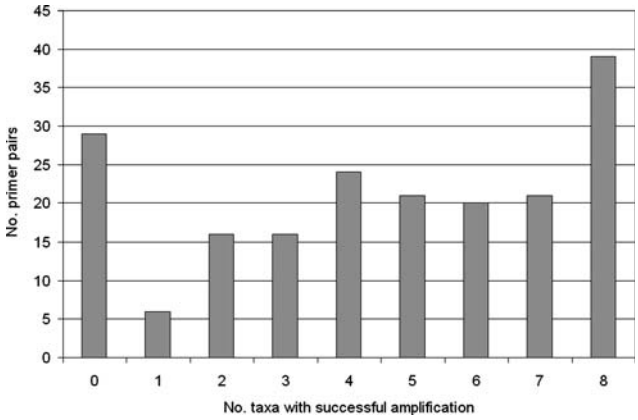

Fig. 1 Portability of 192 'universal' primer pairs in the Asteraceae. For each locus the number of successful amplifications (out of eight) was counted

Table 1 Successful amplifications (out of 192 attempted primer pairs) within each species as well as across each of the three major subfamilies of the Asteraceae

Subfamily/species	*N*
Carduoideae	56
Carthamus tinctorius L. (safflower)	85
Centaurea maculosa Lam. (spotted knapweed)	67
Cichorioideae	122
Lactuca sativa L. (lettuce)	146
Cichorium intybus L. (chicory)	128
Asteroideae	64
Senecio squalidus L. (oxford ragwort)	100
Artemisia vulgaris L. (wormwood/mugwort)	79
Helianthus annuus L. (sunflower)	126
Liatris scariosa (L.) Willd. (Devil's bite)	134

The number of successful amplifications per species ranged from 67 in *Centaurea*, to 146 in *Lactuca* (Fig. 2; Table 1). The lowest number of successful amplifications per subfamily was in the Carduoideae where only 56 primer pairs amplified in both members. This result is not surprising considering that the primers were designed based on EST alignments from *Lactuca* and *Helianthus*, which are members of the two other subfamilies. Thirty-nine primer pairs (20%) amplified fragments in all 8 taxa tested (Fig. 3), 26 of which are putatively single-copy in all individuals. Eight of the 163 loci that could be amplified in at least 1 taxon (including seven of the 39 that worked in all 8 taxa) appeared to lack introns based on amplicon size, despite intron presence in *Arabidopsis*.

Discounting the seven loci for which introns appeared absent, there was good general correspondence between the

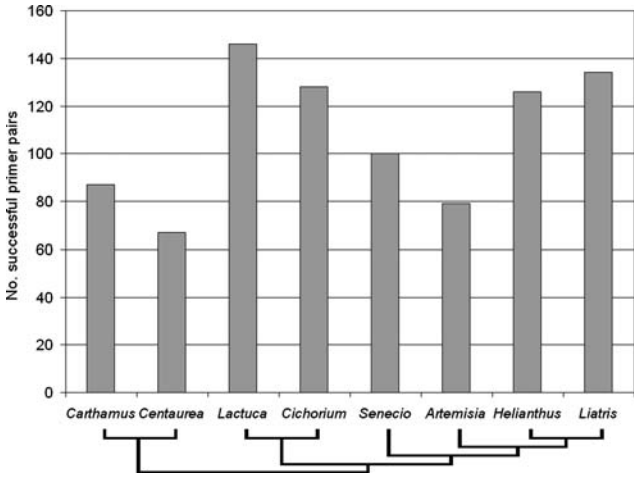

Fig. 2 The *number* of loci that amplify in each of the eight taxa surveyed. An outline phylogenetic tree representing the relationships between the eight taxa is shown below (adapted from Funk et al. 2005)

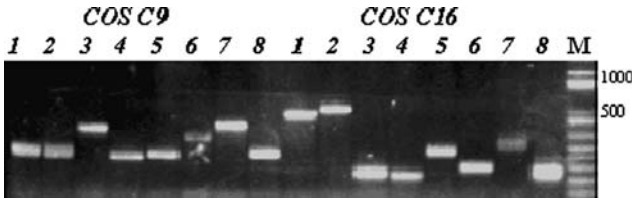

Fig. 3 Representative agarose gel of two 'universal' loci (i.e. loci that amplified across all taxa). Lane designations are as follows: 1 *Carthamus*; 2 *Centaurea*; 3 *Lactuca*; 4 *Cichorium*; 5 *Senecio*; 6 *Artemisia*; 7 *Helianthus*; 8 *Liatris*; M 1 kb Plus DNA ladder (bands of 500 and 1,000 bp are indicated). Note that these taxa are presented in the same order as in Fig. 2

predicted (based on *Arabidopsis* sequence data) and actual amplicon size for loci that produced a single amplicon in all eight taxa ($r^2 = 0.37$; $P = 0.0002$). There was, however, a tendency toward larger amplicon sizes in the Asteraceae as compared to *Arabidopsis* (388 ± 14 bp, mean ± SE vs. 342 ± 17 bp; $P = 0.001$).

Sequence polymorphism

We investigated the degree of polymorphism of these markers in two ways. First, of the 39 markers that amplified in all 8 taxa, we determined whether or not each was represented by a contig containing both of the lettuce or both of the sunflower genotypes in the CGPDB. Of these loci, nearly all (26 of 30 in lettuce and 14 of 15 in sunflower) exhibited at least one SNP between genotypes within species. Second, of the ten loci that were amplified and sequenced from the eight genotypes of safflower, all ten were single-locus, and nine were polymorphic (Table 2). For the nine polymorphic loci the number of SNPs per kb ranged from 2.6 to 26.9 (mean = 12.8). The ITS region, in contrast, only harbored 2.7 SNPs per kb. Nucleotide polymorphism (measured as both Watterson's θ and total nucleotide diversity, π) averaged five times greater than the ITS values (Table 2).

Table 2 Polymorphism data for ten universal loci amplified in eight accessions of safflower

Locus	Length (bp)	S[a]	S per kb	Indels	θ_W[b]	π[c]
A19	365	7	19.2	0	0.0063	0.0071
A25	486	9	18.5	0	0.0076	0.0059
A27	395	3	7.6	0	0.0025	0.0079
A28	548	7	12.8	2	0.0040	0.0066
A39	578	0	0.0	0	0.0000	0.0000
B7	386	1	2.6	1	0.0009	0.0007
B12	408	9	22.1	1	0.0068	0.0073
B27	621	4	6.4	1	0.0022	0.0017
C32	802	24	29.9	6	0.0096	0.0128
D22	355	3	8.5	1	0.0040	0.0058
Average	494.4	6.7	12.8	1.2	0.0044	0.0050
ITS	733	2	2.7	0	0.0009	0.0009

For comparative purposes the same accessions were sequenced for the ITS region if the rRNA genes

[a] Number of segregating sites

[b] Watterson's θ

[c] Total nucleotide diversity

Discussion

The markers developed here provide a strong foundation for comparative map-based analyses of the Asteraceae as well as for phylogenetic reconstructions of the group. Although only 39 primer pairs appeared universal (i.e., worked across the whole family), there was some phylogenetic structuring apparent in the data, with primers that only worked in a subset of taxa typically performing well across more closely-related species. Thus, primers that amplified in all samples of the Carduoideae, Cichorioideae, or Asteroideae may indeed turn out to be useful within an entire subfamily even if they are not more broadly useful across subfamilies. Similarly, even primers that amplified in only one taxon are potentially informative across an array of more closely related species than what was analyzed here. The failure of 29 primer pairs to amplify DNA fragments from any taxon may have been caused by mismatch of the 3′ end of the primer sequence due to nucleotide divergence in lettuce or sunflower. Redesigning these primers to anchor the 3′ end to the conserved second base position in a codon of the targeted genes could facilitate their amplification.

Although these markers were designed to span introns, it is clear that intron loss or gain since the last common ancestor of *Arabidopsis* and the Asteraceae has occurred in some cases. The relatively low rate of intron loss/gain is, however, fully consistent with the notion that intron locations are for the most part conserved across taxa (e.g., Strand et al. 1997; Ku et al. 2000; Roy et al. 2003). An additional observation is that intron size in general has been increasing in the lineage leading to the Asteraceae or decreasing in the lineage leading to *Arabidopsis*. There were no clear trends within the Asteraceae, although intron sizes were somewhat smaller in the Cichorioideae as compared to the Carduoideae and the Asteroideae.

The majority of loci were found to be polymorphic within species, as shown by the analysis of the CGPDB contigs. Although some of the apparent polymorphisms may be due to sequencing errors, it is worth noting that an estimated 72% of the sunflower SNPs identified from within the CGPDB have ultimately been validated by genetic mapping (Lai et al. 2005). Thus, it appears that the majority of the loci developed herein will be appropriate for genetic map-based analyses in lettuce and sunflower. In addition, the analysis of ten loci in safflower indicates that these loci will likely be useful for answering phylogenetic questions even amongst relatively closely related taxa.

Utility of the markers

The markers described in this paper hold great promise for a wide range of applications within the Asteraceae. For example, because a number of these loci can be amplified across the family, they are likely to be useful tools for comparative genetic mapping. While a substantial amount of effort has been devoted to comparative mapping within the genus *Helianthus* (e.g., Rieseberg et al. 1995; Burke et al. 2004; Lai et al. 2005), relatively little has been done with regard to making comparisons across the family, primarily due to a lack of suitable markers (but see Timms et al. 2006).

Given the potentially high level of intraspecific polymorphism revealed by the markers described herein, it appears that they will be valuable tools for anchoring comparisons between existing genetic maps of various species within the family, as well as for aiding in the construction of new maps. These markers will thus aid efforts aimed at transferring information amongst well-characterized crops, such as lettuce and sunflower, as well as efforts to transfer data from such species to other members of the family with less well-developed genomic resources. Even though only 20% of all loci amplified across all eight of the species included in our survey, a much larger fraction have the potential to inform pairwise comparisons amongst taxa (e.g., lettuce–sunflower, lettuce–safflower, and safflower–sunflower). Moreover, this work was carried out using a standardized PCR protocol; it is therefore likely that a much larger fractions of these markers could be utilized in each taxon with additional primer modifications and/or PCR optimizations.

In addition to their utility for comparative map-based analyses, these markers are likely to be useful for phylogenetic analyses within the Asteraceae. Unfortunately, phylogenetic reconstructions based on only one or a few genes may not reflect accurate species relationships (Kopp and True 2002; Rokas et al. 2003). Although incomplete taxon sampling is likely to result in incongruence between datasets, it has become clear that the sampling of more genes per taxon increases the likelihood of resolving the true topology (Rokas et al. 2003; Rokas and Carroll 2005). The main limitation to date has been access to a sufficiently large number of suitable genes for analysis, particularly in less well-characterized taxa. Although the underlying sequences vary in length, with a number of the loci developed here being relatively short, a substantial number of these markers spanned 500 bp or more (ranging from a low of 15 such loci in *Centaurea* to a high of 54 loci in *Liatris*), and thus show great promise for phylogenetic applications.

Conclusions and future directions

The recent identification of a large number of PCR-based COS markers for the Solanaceous crops (pepper, tomato, potato) by Wu et al. (2006) reveals the potential for EST databases to provide markers useful for comparative mapping and/or phylogenetic investigations. Unfortunately, resources of this sort have not previously been available to individuals studying members of the Asteraceae, despite the availability of extensive EST resources within the family. Our results help to remedy this situation by generating an initial set of universal markers that will facilitate comparative mapping and phylogenetic analyses within the family. Going forward, these markers can be utilized in studies focused on elucidating patterns of genome evolution, map-based cloning, marker-assisted selection, and phylogenetic reconstruction with the Asteraceae.

Acknowledgments EST sequence data were obtained from the Compositae Genome Project website. This work was supported by grants from the National Science Foundation (DBI-0421630 to RVK and DBI-0332411 to JMB), the NSF Research Experience for Undergraduates Program (DBI-0354125 to UMass Boston) and the United States Department of Agriculture (03-35300-13104 to JMB). Comments from members of our lab groups improved an earlier version of this manuscript.

References

Alvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. Mol Phylogenet Evol 29:417–434

Axelsson T, Shavorskaya O, Lagercrantz U (2001) Multiple flowering time QTLs within several *Brassica* species could be the result of duplicated copies of one ancestral gene. Genome 44:856–864

Baldwin BG (1993) Molecular Phylogenetics of *Calycadenia* (Compositae) based on ITS sequences of nuclear ribosomal DNA—chromosomal and morphological evolution reexamined. Am J Bot 80:222–238

Bonierbale MW, Plaisted RL, Tanksley SD (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120:1095–1103

Bowers JE, Abbey C, Anderson S, Chang C, Draye X, Hoppe AH, Jessup R, Lemke C, Lennington J, Li ZK, Lin YR, Liu SC, Luo LJ, Marler BS, Ming RG, Mitchell SE, Qiang D, Reischmann K, Schulze SR, Skinner DN, Wang YW, Kresovich S, Schertz KF, Paterson AH (2003) A high-density genetic recombination map of sequence-tagged sites for Sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. Genetics 165:367–386

Bowers JE, Arias MA, Asher R, Avise JA, Ball RT, Brewer GA, Buss RW, Chen AH, Edwards TM, Estill JC, Exum HE, Goff VH, Herrick KL, Steele CLJ, Karunakaran S, Lafayette GK, Lemke C, Marler BS, Masters SL, McMillan JM, Nelson LK, Newsome GA, Nwakanma CC, Odeh RN, Phelps CA, Rarick EA, Rogers CJ, Ryan SP, Slaughter KA, Soderlund CA, Tang HB, Wing RA, Paterson AH (2005) Comparative physical mapping links conservation of microsynteny to chromosome structure and recombination in grasses. Proc Natl Acad Sci USA 102:13206–13211

Burke JM, Lai Z, Salmaso M, Nakazato T, Tang S, Heesacker A, Knapp SJ, Rieseberg LH (2004) Comparative mapping and rapid karyotypic evolution in the genus *Helianthus*. Genetics 167:449–457

Chambers EW, Lovin DD, Severson DW (2003) Utility of comparative anchor-tagged sequences as physical anchors for comparative genome analysis among the Culicidae. Am J Trop Med Hyg 69:98–104

Chapman B, Chang J (2000) Biopython: python tools for computational biology. ACM SIGBIO Newsl 20:15–19

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu YL, Kron KA, Rettig JH, Conti E, Palmer JD, Manhart JR, Sytsma KJ, Michaels HJ, Kress WJ, Karol KG, Clark WD, Hedren M, Gaut BS, Jansen RK, Kim KJ, Wimpee CF, Smith JF, Furnier GR, Strauss SH, Xiang QY, Plunkett GM, Soltis PS, Swensen SM, Williams SE, Gadek PA, Quinn CJ, Eguiarte LE, Golenberg E, Learn GH, Graham SW, Barrett SCH, Dayanandan S, Albert VA (1993) Phylogenetics of seed plants—an analysis of nucleotide sequences from the plastid gene *Rbc*L. Ann Mo Bot Gard 80:528–580

Choi HK, Luckow MA, Doyle J, Cook DR (2006) Development of nuclear gene-derived molecular markers linked to legume genetic maps. Mol Genet Genomics 276:56–70

Dirlewanger E, Graziano E, Joobeur T, Garriga-Caldere F, Cosson P, Howad W, Arus P (2004) Comparative mapping and marker-assisted selection in Rosaceae fruit crops. Proc Natl Acad Sci USA 101:9891–9896

Doganlar S, Frary A, Daunay MC, Lester RN, Tanksley SD (2002) A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the Solanaceae. Genetics 161:1697–1711

Dominguez I, Graziano E, Gebhardt C, Barakat A, Berry S, Arus P, Delseny M, Barnes S (2003) Plant genome archaeology: evidence for conserved ancestral chromosome segments in dicotyledonous plant species. Plant Biotechnol J 1:91–99

Doyle JJ, Gaut BS (2000) Evolution of genes and taxa: a primer. Plant Mol Biol 42:1–23

Feltus FA, Singh HP, Lohithaswa HC, Schulze SR, Silva TD, Paterson AH (2006) A comparative genomics strategy for targeted discovery of single-nucleotide polymorphisms and conserved-noncoding sequences in orphan crops. Plant Physiol 140:1183–1191

Foote T, Roberts M, Kurata N, Sasaki T, Moore G (1997) Detailed comparative mapping of cereal chromosome regions corresponding to the *Ph1* locus in wheat. Genetics 147:801–807

Fourmann M, Barret P, Froger N, Baron C, Charlot F, Delourme R, Brunel D (2002) From *Arabidopsis thaliana* to *Brassica napus*: development of amplified consensus genetic markers (ACGM) for construction of a gene map. Theor Appl Genet 105:1196–1206

Fredslund J, Schauser L, Madsen LH, Sandal N, Stougaard J (2005) PriFi: using a multiple alignment of related sequences to find primers for amplification of homologs. Nucleic Acids Res 33:W516–W520

Fredslund J, Madsen LH, Hougaard BK, Nielsen AM, Bertioli D, Sandal N, Stougaard J, Schauser L (2006) A general pipeline for the development of anchor markers for comparative genomics in plants. BMC Genomics 7:207

Fulton TM, Van der Hoeven R, Eannetta NT, Tanksley SD (2002) Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14:1457–1467

Funk VA, Bayer RJ, Keeley S, Chan R, Watson L, Gemeinholzer B, Schilling EE, Panero JL, Baldwin BG, Garcia-Jacas N, Susanna A, Jansen R (2005) Everywhere but Antarctica: using a supertree to understand the diversity and distribution of the Compositae. Biol Skr 55:343–374

Gaines CA, Hare MP, Beck SE, Rosenbaum HC (2005) Nuclear markers confirm taxonomic status and relationships among highly endangered and closely related right whale species. Proc R Soc Lond B 272:533–542

Gale MD, Devos KM (1998a) Comparative genetics in grasses. Proc Natl Acad Sci USA 95:1971–1974

Gale MD, Devos KM (1998b) Plant comparative genetics after ten years. Science 282:656–658

Garcia-Jacas N, Susanna A, Garnatje T, Vilatersana R (2001) Generic delimitation and phylogeny of the subtribe Centaureinae (Asteraceae): a combined nuclear and chloroplast DNA analysis. Ann Bot 87:503–515

Gielly L, Taberlet P (1994) The use of chloroplast DNA to resolve plant phylogenies - noncoding versus *Rbc*L sequences. Mol Biol Evol 11:769–777

Grant D, Cregan P, Shoemaker RC (2000) Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. Proc Natl Acad Sci USA 97:4168–4173

Griffiths S, Sharp R, Foote TN, Bertin I, Wanous M, Reader S, Colas I, Moore G (2006) Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat. Nature 439:749–752

Hamilton MB (1999) Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. Mol Ecol 8:521–523

Kalo P, Seres A, Taylor SA, Jakab J, Kevei Z, Kereszt A, Endre G, Ellis THN, Kiss GB (2004) Comparative mapping between *Medicago sativa* and *Pisum sativum*. Mol Genet Genomics 272:235–246

Kesseli R, Michelmore RM (1997) The Compositae: systematically fascinating but specifically neglected. In: Paterson AH (ed) Genome mapping of Plants. R.G. Landes Co., Georgetown, pp 179–191

Kim KJ, Jansen RK (1995) *Ndh*F sequence evolution and the major clades in the sunflower family. Proc Natl Acad Sci USA 92:10379–10383

Kole C, Quijada P, Michaels SD, Amasino RM, Osborn TC (2001) Evidence for homology of flowering-time genes VFR2 from *Brassica rapa* and FLC from *Arabidopsis thaliana*. Theor Appl Genet 102:425–430

Kopp A, True JR (2002) Phylogeny of the oriental *Drosophila melanogaster* species group: a multilocus reconstruction. Syst Biol 51:786–805

Kowalski SP, Lan TH, Feldmann KA, Paterson AH (1994) Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization. Genetics 138:499–510

Ku HM, Vision T, Liu JP, Tanksley SD (2000) Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny. Proc Natl Acad Sci USA 97:9121–9126

Ku HM, Liu JP, Doganlar S, Tanksley SD (2001) Exploitation of *Arabidopsis*-tomato synteny to construct a high-resolution map of the ovate-containing region in tomato chromosome 2. Genome 44:470–475

Lagercrantz U (1998) Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. Genetics 150:1217–1228

Lagercrantz U, Lydiate DJ (1996) Comparative genome mapping in *Brassica*. Genetics 144:1903–1910

Lagercrantz U, Putterill J, Coupland G, Lydiate D (1996) Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome collinearity and congruence of genes controlling flowering time. Plant J 9:13–20

Lai Z, Livingstone K, Zou Y, Church SA, Knapp SJ, Andrews J, Rieseberg LH (2005) Identification and mapping of SNPs from ESTs in sunflower. Theor Appl Genet 111:1532–1544

Lan T-H, Paterson AH (2000) Comparative mapping of quantiative trait loci sculpting the curd of *Brassica oleracea*. Genetics 155:1927–1954

Lee JM, Grant D, Vallejos CE, Shoemaker RC (2001) Genome organization in dicots. II. *Arabidopsis* as a 'bridging species' to resolve genome evolution events among legumes. Theor Appl Genet 103:765–773

Lin Y-R, Schertz KF, Paterson AH (1995) Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum mapping population. Genetics 141:391–411

Livingstone KD, Lackney VK, Blauth JR, van Wijk R, Jahn MK (1999) Genome mapping in *Capsicum* and the evolution of genome structure in the Solanaceae. Genetics 152:1183–1202

Lyons LA, Laughlin TF, Copeland NG, Jenkins NA, Womack JE, Obrien SJ (1997) Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. Nat Genet 15:47–56

Maroof MAS, Yang GP, Biyashev RM, Maughan PJ, Zhang Q (1996) Analysis of the barley and rice genomes by comparative RFLP linkage mapping. Theor Appl Genet 92:541–551

Ming R, Liu SC, Lin YR, da Silva J, Wilson W, Braga D, van Deynze A, Wenslaff TF, Wu KK, Moore PH, Burnquist W, Sorrells ME, Irvine JE, Paterson AH (1998) Detailed alignment of *Saccharum* and *Sorghum* chromosomes: comparative organization of closely related diploid and polyploid genomes. Genetics 150:1663–1682

Okuyama Y, Fujii N, Wakabayashi M, Kawakita A, Ito M, Watanabe M, Murakami N, Kato M (2005) Nonuniform concerted evolution and chloroplast capture: heterogeneity of observed introgression patterns in three molecular data partition phylogenies of Asian *Mitella* (Saxifragaceae). Mol Biol Evol 22:285–296

Osborn TC, Kole C, Parkin IAP, Sharpe AG, Kuiper M, Lydiate DJ, Trick M (1997) Comparison of flowering time genes in *Brassica rapa*, *B. napus* and *Arabidopsis thaliana*. Genetics 146:1123–1129

Paterson AH, Lin Y-R, Li Z, Schertz KF, Doebley JF, Pinson SRM, Liu S-C, Stansel JW, Irvine JE (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269:1714–1718

Paterson AH, Bowers JE, Burow MD, Draye X, Elsik CG, Jiang CX, Katsar CS, Lan TH, Lin YR, Ming RG, Wright RJ (2000) Comparative genomics of plant chromosomes. Plant Cell 12:1523–1539

Prince JP, Pochard E, Tanksley SD (1993) Construction of a molecular linkage map of pepper and a comparison of synteny with tomato. Genome 36:404–417

Rieseberg LH, Van Fossen C, Desrochers AM (1995) Hybrid speciation accompanied by genomic reorganization in wild sunflowers. Nature 375:313–316

Roca AL, Georgiadis N, Pecon-Slattery J, O'Brien SJ (2001) Genetic evidence for two species of elephant in Africa. Science 293:1473–1477

Rokas A, Carroll SB (2005) More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. Mol Biol Evol 22:1337–1344

Rokas A, Williams BL, King N, Carroll SB (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425:798–804

Roy SW, Fedorov A, Gilbert W (2003) Large-scale comparison of intron positions in mammalian genes shows intron loss but no gain. Proc Natl Acad Sci USA 100:7158–7162

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Small RL, Cronn RC, Wendel JF (2004) Use of nuclear genes for phylogeny reconstruction in plants. Aust Syst Bot 17:145–170

Smith E, Shi L, Drummond P, Rodriguez L, Hamilton R, Powell E, Nahashon S, Ramlal S, Smith G, Foster J (2000) Development and characterization of expressed sequence tags for the turkey (*Meleagris gallopavo*) genome and comparative sequence analysis with other birds. Anim Genet 31:62–67

Soltis DE, Kuzoff RK (1995) Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). Evolution 49:727–742

Soltis DE, Soltis PS (1998) Choosing an approach and an appropriate gene for phylogenetic analysis. In: Soltis DE, Soltis PM, Doyle JJ (eds) Molecular systematics of plants. II. DNA sequencing. Kluwer Academic Publishers, Boston, pp 1–42

Strand AE, Leebens-Mack J, Milligan BG (1997) Nuclear DNA-based markers for plant evolutionary biology. Mol Ecol 6:113–118

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of 3 noncoding regions of chloroplast DNA. Plant Mol Biol 17:1105–1109

Tanksley SD, Ganal MW, Prince JP, Devicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Roder MS, Wing RA, Wu W, Young ND (1992) High-density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141–1160

Timms L, Jimenez R, Chase M, Lavelle D, McHale L, Kozik A, Lai Z, Heesacker A, Knapp S, Rieseberg L, Michelmore R, Kesseli R (2006) Analyses of synteny between *Arabidopsis thaliana* and species in the Asteraceae reveal a complex network of small syntenic segments and major chromosomal rearrangements. Genetics 173:2227–2235

Van Deynze AE, Sorrells ME, Park WD, Ayres NM, Fu H, Cartinhour SW, Paul E, McCouch SR (1998) Anchor probes for comparative mapping of grass genera. Theor Appl Genet 97:356–369

Vilatersana R, Susanna A, Garcia-Jacas N, Garnatje T (2000) Generic delimitation and phylogeny of the *Carduncellus-Carthamus* complex (Asteraceae) based on ITS sequences. Plant Syst Evol 221:89–105

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84:9054–9058

Wu XL, Larson SR, Hu ZM, Palazzo AJ, Jones TA, Wang RRC, Jensen KB, Chatterton NJ (2003) Molecular genetic linkage maps for allotetraploid *Leymus* wildryes (Gramineae: Triticeae). Genome 46:627–646

Wu FN, Mueller LA, Crouzillat D, Petiard V, Tanksley SD (2006) Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: A test case in the euasterid plant clade. Genetics 174:1407–1420