

DRAFT: UPDATED VERSIONS WILL BE POSTED ON THE CWP WEBSITE AS INPUT IS RECEIVED AND REFINEMENTS ARE MADE.

White paper: Priorities for research, education and extension in genomics, genetics, and breeding of the Compositae
(Draft 10/11/07)

Executive Summary

The Compositae is the most diverse and largest of angiosperm families, comprising one-tenth of all flowering plant species. Several species in the family are economically important but understudied. There have been multiple parallel domestications. The family contains over 40 crops and six of the top ten noxious weeds in the U.S. Lettuce is one of the top ten crops in the US worth over \$2 billion and sunflower ranks fourth or fifth in production among oilseed crops worldwide with a value over ~\$40 billion.

Extensive genetic, EST and more limited BAC library resources exist for lettuce and sunflower. Sequencing efforts have focused on generating ESTs from over 20 Compositae species; more than 1% of all EST sequences at NCBI are currently from the Compositae, making this one of the best represented plant families. A chip for massively parallel genetic and expression analyses is being developed for lettuce and sunflower. However, other genomic resources are limited. *Leontodon taraxacoides* has been recently identified as a potential small genome model species but has not yet been developed.

In the short term, the high priorities include sequencing of the gene space and massively parallel genetic analysis of lettuce and sunflower as well as comparative functional genomics to determine the genetic and molecular bases of agriculturally and ecologically important as well as evolutionarily significant genes from the diverse economically important species in the Compositae.

In addition, resources need to be developed to facilitate the genomic sequencing of lettuce, sunflower and *Leontodon* as the next generation of highly-efficient sequencing technologies become available. These resources include expansion of the BAC libraries, fingerprinting and development of a minimum tiling path, and BAC-end sequencing as well as application of next generation sequencing technologies. When sequencing costs are further reduced, a minimum of these three species should be sequenced using a combination of approaches. Also, genomics tools developed for lettuce, sunflower and *Leontodon* need to be applied to other Composite species.

Objectives

- To establish tools and resources for genomic investigations in multiple crops and invasive species in the Compositae.
- To develop comprehensive gene catalogs for economically important species of the Compositae.
- To develop detailed genetic maps integrating phenotypic data for agriculturally important and environmentally friendly traits with sequenced genes.
- To enhance the introgression of agriculturally useful alleles from wild species.
- To characterize the evolutionary forces responsible for the origin and diversification of the Compositae family.
- To determine patterns of genome evolution in the Compositae and establish micro-syntenic relationships with *Arabidopsis* and other flowering plants.

- To develop strategies to minimize the transfer of transgenes from cultivated to wild species, particularly weeds.
- To train students at all levels from diverse backgrounds in the technical and analytical aspects of evolutionary genomics.

Background

The Compositae are the focus of at least two somewhat distinct groups of researchers: those interested in the crop and weedy species and those interested in the broad taxonomic aspects of this diverse family. The Compositae is one of the largest, most ecologically successful and in aggregate one of the most economically important plant families. Lettuce and sunflower are the most economically important representatives of this understudied but very large and diverse family that also includes many minor crops, important weedy species, and species with novel biochemistries (Table 1).

(i) The Compositae Family.

The Compositae (Asteraceae) is the largest and most diverse family of flowering plants (Heywood, 1978; Funk *et al.*, 2005), comprising one-tenth of all known Angiosperm species. Plants within this family are characterized by a compound inflorescence that has the appearance of a single "composite" flower. The Compositae is divided into three major subfamilies and one minor subfamily, with 1,100 to 2,000 genera and over 20,000 species (Fig. 1; Heywood, 1978; Cronquist, 1977; Jansen *et al.*, 1991; Funk *et al.*, 2005). Lettuce, sunflower, and safflower are representatives of each of the three major subfamilies. The family has undergone extensive diversification producing a cosmopolitan array of taxa encompassing ephemeral herbs, vines, and trees that thrive in some of the world's most inhospitable habitats (e.g., vertisols, deserts, and salt marshes). In contrast to some other large Angiosperm families, much of the biodiversity of the Compositae is in extreme environments rather than within the tropics. Representatives of this family are present on every continent and nearly all habitats except Antarctica (Funk *et al.*, 2005). While the size and adaptive success of the Compositae have stimulated considerable research into its systematics and evolution, molecular characterization has lagged behind other families (Kesseli & Michelmore, 1997).

Figure 1. Systematic relationships within the Compositae.

Omitted as 4 Mbytes. Can be accessed separately from <http://compgenomics.ucdavis.edu/cwp/draft.php>.

Over 40 economically important species have been domesticated within the Compositae, (Table 1; Kesseli & Michelmore, 1997). These include food (e.g. lettuce, chicory, Jerusalem artichoke), oil (sunflower, safflower), medicinal (*Echinacea*; chamomile) and many ornamental (chrysanthemum, dahlia, zinnia, marigold) crops as well as many semi-domesticated, minor crops. Food and non-food Compositae are grown on over 21 million ha of land per year worldwide (Supplementary Table 1; <http://www.fao.org/>). High quality edible oils are low in saturated and high in mono- and di-unsaturated fatty acids. The Compositae are renowned for their variety of novel secondary chemicals (Caligari & Hind, 1996), including several novel industrial fatty acids in *Dimorphotheca*, *Crepis*, *Vernonia*, and *Stokesia* (Smith, 1985), powerful insecticides and industrial chemicals (e.g., *Chrysanthemum*), and rubber (guayule) (Heywood *et al.*, 1977).

Table 1. Economically Significant Members of the Compositae.

Common Name	Genus and Species	Economic Impact
Lettuce	<i>Lactuca sativa</i> L.	Food
Sunflower	<i>Helianthus annuus</i> L.	Oil, Food, and Ornamental
Safflower	<i>Carthamus tinctorius</i> L.	Oil, Food, and Ornamental
Endive	<i>Cichorium endivia</i> L.	Food
Chicory	<i>Cichorium intybus</i> L.	Food
Artichoke	<i>Cynara scolymus</i> L.	Food
Cardoon	<i>Cynara cardunculus</i> L.	Food
Jerusalem Artichoke	<i>Helianthus tuberosus</i> L.	Food
Spanish Salsify	<i>Scolymus hispanicus</i>	Food
Stevia	<i>Stevia rebaudiana</i>	Food
Scorzonera	<i>Scorzonera hispanica</i>	Food
Yacon	<i>Smallanthus sonchifolius</i>	Food
Chop-Suey Greens	<i>Chrysanthemum coronarium</i>	Food
French Scorzonera	<i>Reichardia picroides</i>	Food
Sumpweed	<i>Iva annua</i>	Food
Jambu	<i>Spilanthes acmella</i>	Food
Balsamorhiza	<i>Balsamorhiza</i> spp.	Food
Yam Daisy	<i>Microseris scapigera</i>	Food
Woolly Wyethia	<i>Wyethia mollis</i>	Food
Food Salsify	<i>Tragopogon porrifolius</i>	Food and Medicinal
Bidens	<i>Bidens</i> spp.	Medicinal
Noug	<i>Guizotia abyssinnica</i> L.	Oil and Food
Calendula	<i>Calendula officinalis</i> L.	Oil, Ornamental, and Herb
Dimorphotheca Daisy	<i>Dimorphotheca pluvialis</i> L.	Oil and Ornamental
Vernonia	<i>Vernonia</i> spp.	Oil
Crepis	<i>Crepis</i> spp.	Oil
Vernonia	<i>Vernonia galamensis</i> Cass.	Oil
Stoke's Aster	<i>Stokesia laevis</i> L.	Oil
Osteospermum	<i>Osteospermum</i> spp.	Oil and Ornamental
Stoke's Aster	<i>Stokesia laevis</i> L.	Oil and Ornamental
Guayule	<i>Parthenium argentatum</i> L.	Rubber
Tau-Saghyz	<i>Scorzonera tau-saghyz</i>	Rubber
Rubber Dandelion	<i>Taraxacum kok-saghyz</i>	Food and Rubber
Pyrethrum Daisy	<i>Chrysanthemum cinerariifolium</i> L.	Pesticide and Ornamental
Coneflowers	<i>Echinacea</i> spp.	Medicinals and Ornamentals
Black-Eyed Susans	<i>Rudbeckia</i> spp.	Ornamentals
Gerbera Daisies	<i>Gerbera</i> spp.	Ornamentals
Marigolds	<i>Tagetes</i> spp.	Ornamentals
Chrysanthemums	<i>Chrysanthemum</i> spp.	Ornamentals
Cosmos	<i>Cosmos</i> spp.	Ornamentals
Zinnias	<i>Zinnia</i> spp.	Ornamentals
Lawn Daisy	<i>Bellis perennis</i> L.	Ornamental and Herb
Tansy	<i>Tanacetum vulgare</i> L.	Ornamental and Herb
Feverfew	<i>Tanacetum parthenium</i> L.	Ornamental and Herb
Cotton Thistle	<i>Onopordum acanthium</i> L.	Ornamental and Herb
Elecampane	<i>Inula helenium</i> L.	Ornamental and Herb
Santolina	<i>Santolina chamaecyparissus</i> L.	Ornamental and Herb
Curry Plant	<i>Helichrysum angustifolium</i> L.	Ornamental and Herb

Sweet Joe Pye	<i>Eupatorium purpurea</i> L.	Ornamental and Herb
Yarrow	<i>Achillea millefolium</i> L.	Ornamental and Herb
Artemesias	<i>Artemesia</i> spp.	Ornamentals and Herbs
Tarragon	<i>Artemesia dracunculus</i> L.	Herb
Costmary	<i>Chrysanthemum balsamita</i> L.	Herb
Chamomile	<i>Anthemis nobilis</i> L.	Herb
Coltsfoot	<i>Tussilago farfara</i>	Herb
Dandelion	<i>Taraxacum officinale</i> L.	Food and Weed
Ragweeds	<i>Ambrosia</i> spp.	Weeds
Hawkweeds	<i>Hieracium</i> spp.	Weeds
Thistles	<i>Cirsium</i> spp.	Weeds
Thistles	<i>Sonchus</i> spp.	Weeds
Groundsels	<i>Senecio</i> spp.	Weeds

Also, Composite species such as thistles, knapweeds, and dandelions are among the world's most noxious weeds, the control of which costs \$25 billion to \$130 billion annually in the U.S. (Pimentel *et al.*, 2000; Pennisi, 2003). Indeed, eight of the 20 worst weeds in the U.S. are Composites (Anon. 2003), as are 36 of 181 new, potentially invasive American species in Europe (Forman, 2003). *Lactuca* and *Helianthus* are particularly interesting and complementary genera with regard to their reciprocal histories of domestication and evolution of invasiveness. Sunflower was domesticated in North America, yet today *H. annuus* and 21 other taxa in the genus *Helianthus* are considered naturalized or invasive in Europe (Rehorek, 1997; Forman & Kesseli, unpublished). Also, due to high levels of gene flow between cultivated and weed sunflower (Arias & Rieseberg 1994; Linder *et al.* 1998), sunflower has been featured in debates about the role of crop-wild gene flow and transgene escape in the evolution of "super weeds" (e.g., Burke & Rieseberg, 2003; Ellstrand, 2003; Snow *et al.*, 2003). Conversely, lettuce was domesticated in the Mediterranean region, yet today *L. serriola* (the progenitor of cultivated lettuce) and 22 other taxa of *Lactuca* are currently defined as weeds in the U.S. The potential for transgene flow from lettuce crops to weeds is the focus of a major research initiative in Europe (<http://www.plant.wageningen-ur.nl/projects/angel/>).

Crops and weeds each have distinct sets of traits commonly associated with them, but they also share some traits, presumably due to parallel adaptation to disturbed habitats created by humans (Baker, 1974; Harlan, 1975). For example, germination in many environments, rapid growth through vegetative phase to flowering, self-compatibility, and high seed output in favorable environments are common to both weeds and cultivated species. On the other hand, discontinuous germination due to seed dormancy is crucial for weed success, while rapid synchronized germination is desirable in crops. Extended seed production during a growing season benefits weeds, but harvesting promotes synchronized maturation in crops. Adaptations for short- and long-distance seed dispersal are critical for weeds, yet non-shattering is the hallmark of crop species.

***Lactuca* spp. (lettuce).**

Lettuce (*Lactuca sativa* L.) is a diploid (2n = 18) species within the Lactuoideae subfamily of the Compositae (Koopman & De Jong, 1996). There are four well-established species within section *Lactuca*, cultivated *L. sativa* and three wild species, *L. serriola*, *L. saligna*, and *L. virosa*. *L. serriola* is probably the progenitor of *L. sativa* (Kesseli *et al.*, 1991; de Vries, 1997) and also a cosmopolitan weed found in North and South America, Europe, Asia, Australia, and Africa. The tribe also possesses such notorious weeds as *Taraxacum officinale* (dandelion), *Hieracium* spp. (hawkweeds), *Sonchus* spp. (sow thistle), and *Cichorium intybus* (chicory).

Lettuce is an important vegetable crop species and ranks as one of the top ten most valuable crops in the U.S. (Anon., 2001) with an annual value of over \$2 billion. Genetic improvement programs are focused on morphology, horticultural performance, physiological disorders, and disease resistance (Ryder, 1986). Wild species, particularly *L. serriola*, have been sources of several disease resistance genes, but have not been accessed systematically (Crute, 1988).

Lettuce, which has an estimated genome size of ca. 2.5 Gb (Michaelson *et al.*, 1991; Table 2), is amenable to classical and molecular genetic analyses. Cultivars of *L. sativa* are highly inbred. Crosses can be made readily and multiple generations can be produced each year. Our consensus genetic map now includes over 2,700 markers and 9 linkage groups (Landry *et al.*, 1987; Kesseli *et al.*, 1994; R. Michelmore *et al.*, unpublished). Our core mapping population is based on an interspecific cross between *L. sativa* and *L. serriola* and segregates for most traits associated with domestication or invasiveness. This population has been adopted by the European ANGEL Project (<http://www.plant.wageningen-ur.nl/projects/angel/>) and RILs have been distributed to several groups for additional mapping of markers and phenotypic traits. QTL analyses have been conducted for many horticultural and morphological traits, including bolting, root architecture, shattering, and seed dormancy, oil content, and size (Johnson *et al.* 2000 & unpublished). Many genes for resistance to several diseases have and are being characterized (e.g. Kesseli *et al.*, 1993, 1994; Maisonneuve *et al.*, 1994; Robbins *et al.*, 1994).

***Helianthus* spp. (sunflower and Jerusalem artichoke).**

The genus *Helianthus* belongs to the Asteroideae, a second major subfamily of the Compositae. The genus is native to temperate North America and contains 12 annual and 36 perennial species (Schilling & Heiser 1981). Cultivated sunflower (*H. annuus* L.) is an annual diploid ($n = 17$), whereas the Jerusalem artichoke (*H. tuberosus*) is a perennial hexaploid ($n = 51$). Both crops originated in the continental U.S. (Heiser & Smith 1955; Harter *et al.* 2004) and are conspecific with their progenitors. As with *Lactuca*, the wild progenitors of domesticated *Helianthus* are also two of the worst weeds in the genus. *H. annuus* is a major weed in corn, soybean, wheat, and sugar beets (Al-Khatib *et al.* 1998), whereas *H. tuberosus* is an aggressive perennial that spreads primarily via rhizomes.

Cultivated sunflower is a globally important oilseed, food, and ornamental crop. Sunflower oil is the world's third most important vegetable oil accounting for about 13% of the total world's edible oil production. Oilseed sunflower was produced on 19.6 million hectares in 70 countries in 2002 and ranks fourth or fifth in production among oilseed crops with a value over ~\$40 billion (Supplementary Table 3. <http://www.fao.org/>). Cultivated sunflower is primarily grown from single-cross hybrid seed, which was valued at \$640 million in 2002, second only to maize. By contrast, the Jerusalem artichoke is a minor crop that produces tubers for food and livestock feed. Wild *Helianthus* species have been important sources of genes for disease resistance, cytoplasmic male-sterility (CMS), abiotic tolerances, and other traits (Jain *et al.* 1993; Jan 2000; Seiler & Rieseberg 1997).

The whole sunflower seed contains ~ 47% oil and ~ 17% protein making it a good animal feedstock. However, the major portion of sunflower production is devoted to oil extraction (Dorrell and Vick 1997) and the greatest breeding efforts are dedicated to this trait. There are three types of sunflower oil, some of which are specially suited for baby foods and cooking since they do not need hydrogenation (mid-oleic and high oleic oils). Sunflower oil is considered a premium oil due to its high unsaturated fatty acid composition and low linolenic acid content (Supplementary Table 3). It is used in soaps and detergents in Eastern Europe. It is also used as

“carrier” in agrochemicals, particularly hydrophobic pesticides, surfactants, adhesives, plastics, softeners, and lubricants.

Sunflower is a widely adapted summer annual that crosses readily and produces abundant seed (~1,000 seeds/plant). Although naturally outcrossing, highly selfing germplasm has been developed that is the core for hybrid breeding and genetic analysis. Several detailed genetic maps have been developed using variety of F₂ and RIL populations with over 2,000 simple sequence repeat (SSR) and sequence-tagged-site (STS) markers by groups in the US, Europe, and Argentina (Flores Berrios *et al.* 2000; Tang *et al.* 2002, 2003b; Burke *et al.* 2002, 2004; Tang & Knapp 2003; Rieseberg *et al.* 2003; Yu *et al.* 2003; Al-Chaarani *et al.* 2004). The first sunflower cross-referenced map was generated by integrating data for 657 SSR and RFLP loci from three populations derived from crosses between contrasting germplasm (Yu *et al.* 2003). This composite linkage map compromised 1423 cM and allowed the selection of SSR markers for genotyping sunflower germplasm (Zhang *et al.* 2005). An international cooperation has recently resulted in further integration of independently developed linkage maps. This map is based public SSR markers with a common linkage group nomenclature, includes 304 AFLP and 191 SSR with a total length of 1824.6 cM and mean density of 3.7 cM per locus (S. Poormohammad Kiani *et al.* 2006, unpublished).

The genetics of a broad spectrum of traits, including seed oil concentration, root morphology, salt and drought stress, branching, seed dormancy, heterosis, male sterility, fertility restoration, flowering time, seed shattering, self-incompatibility, fatty acid and tocopherol composition and concentration, and disease resistance have been and are being analyzed (e.g. post 2000: Burke *et al.* 2002; Burke *et al.* 2005; Slabaugh *et al.* 2003; Tang *et al.* 2003a; Rieseberg *et al.* 2003; León *et al.* 2001, 2003; Bert *et al.* 2001; Hervé *et al.* 2001; Al-Chaarani *et al.* 2002; Mokrani *et al.* 2002; Pérez-Vich *et al.* 2002). These traits are being mapped relative to candidate genes mined from the EST databases to provide maps based on transcribed sequences. The sunflower genome has been estimated to be ~3.5 Gb (Baack *et al.*, 2006; Table 2).

In situ hybridization techniques involving GISH, FISH and BAC-FISH are being optimized for diversity and evolutionary studies between species of the genus *Helianthus* and development of a physical sunflower map allowing a cross reference to the genetic map (Paniego *et al.* 2006). The hexaploid *H. resinosus* is currently being investigated using molecular cytogenetics and markers in order to elucidate species origin through polyploidy (Carrera *et al.* 2004).

Small Genome Model Species for the Compositae: *Leontodon taraxacoides*

Despite the large number of species in the Compositae, few have small genome sizes. Most species in the family have genome sizes in excess of 1 Gb. The rapid cycling *Senecio* species seem to have increased chromosome number rather reduced genome size. However, *Leontodon taraxacoides* has a genome size only twice that of Arabidopsis. This has only recently been discovered and efforts are now underway to develop this species as a small-genome model for the Compositae.

Leontodon (2N = 2X = 8) is in the subfamily Cichorioideae and Lactuceae tribe. It is native to Europe; however, it is found as a weed in 23 states in the U.S. including the West Coast, Midwest, Central East Coast, TX, AL, TN, and KY. There are two subspecies with annual and perennial growth forms which can hybridize.

Table 2. Relative Genome Sizes.

Taxon	Genome size 1C (Gb and pg)
-------	----------------------------

<i>Arabidopsis thaliana</i> ¹	0.16 / 0.16
<i>Leontodon taraxacoides</i> spp. <i>taraxacoides</i> ²	0.29 / 0.30
<i>Leontodon taraxacoides</i> spp. <i>longirostris</i> ²	0.33 / 0.34
<i>Oryza sativa</i> ³	0.49 / .50
<i>Lactuca sativa</i> ⁴	2.6 / 2.7
<i>Helianthus annuus</i> ⁵	3.5 / 3.6

¹Bennett *et al.* (2003), ²E. Baack & L. Rieseberg, unpublished; ³Uozu *et al.*, (1997), ⁴Michaelson *et al.*, (1991), ⁵Price *et al.*, (2000).

No estimate of genome size is currently available for safflower (2N = 24).

Current Resources

Genetic stocks

The International *Lactuca* database reports over 12,028 accessions worldwide; however, there is considerable overlap among the collections (www.cgn.wur.nl/pgr/collections/ildb/). The Center for Genetic Resources, Wageningen, NL includes 2,429 lettuce accessions, including 1168 cultivars, 896 wild species, and 201 landraces. Smaller collections are housed in Prague, Czech Republic (1,328 accessions), Pullman, WA (1,315), Davis, CA (1,200), Salinas, CA (1,221), Gatersleben, Germany (870), Wellesbourne, UK (770), St. Petersburg, Russia (709), Brion, France (534), Zaragoza, Spain (472).

For sunflower, a fairly extensive collection of germplasm is available at the Northern Crop Science Laboratory in Ames, IA. The USDA Germplasm Resources Information Network (GRIN; <http://www.ars-grin.gov/>) lists 2,919 *Helianthus* accessions, with 2,384 of these coming from *H. annuus*. In Argentina, the INTA Germplasm Resource Program comprises of 829 cultivated accessions at the Active Bank in Manfredi Experimental Station and a lower number of wild species at Balcarce Experimental Station. A fairly large collection is also maintained in South Africa (xxxx ref?).

The USDA-GRIN lists 2,223 accessions of *Carthamus*, with 2,144 accessions of these coming from *C. tinctorius* (safflower).

There are 250 accessions for chicory and endive germplasm at the North Central Regional Plant Introduction Station, Ames, Iowa.

Germplasm for other species.....??

Collections for weedy species..??

Genetic Maps and Mapping Populations

Lettuce: The core mapping population comprises of 110 RILs from *L. sativa* cv. Salinas x *L. serriola* (Agryris *et al.*, 2005; Truco *et al.*, 2007). Over 1,600 markers have been mapped on this population (<http://cgp.ucdavis.edu>). This population has recently been expanded to 300 RILs to provide increased genetic resolution. Detailed genetic maps have been constructed using six inter- and intra-specific crosses of *Lactuca* spp. An integrated map of all the current segregation data comprises over 2,700 markers and spans 1,505 cM (Landry *et al.*, 1987; Kesseli *et al.*, 1994; Truco *et al.*, 2007). In addition, a set of 29 Backcross Inbred Lines from the interspecific cross *L. saligna* x *L. sativa* introgression lines have been created and used for mapping (Jeuken & Lindhout 2004, Jeuken *et al.* 2001).

Sunflower: Mapping resources include high-resolution linkage maps for the cultivated sunflower and six wild species with over 2,000 simple sequence repeat (SSR) and sequence-tagged-site (STS) markers (Tang *et al.* 2002, 2003b; Burke *et al.* 2002, 2004, 2005; Tang & Knapp 2003; Rieseberg *et al.* 2003; Lai *et al.* 2005). The composite linkage map of sunflower based on these markers comprises 657 loci and spans 1423 cM with a mean density of 2.2 cM per locus. The integrated map based on AFLP and SSR markers includes 495 markers, spans 1824.6 cM and mean density of 3.7 cM per locus (Al-Chaarani *et al.* 2004, Poormohammad Kiani *et al.* unpubl. data).

Artichoke *Cynara cardunculus* L. var. *scolymus*: A genetic linkage map is being developed using a pseudo testcross approach at the University of Bari (S. Pavan, G. Sonnante, M. Ippedico, A. De Paolis). *More details...*

Summarize details (generation, size etc.) in table??

Molecular Resources

A ~3x BAC library exists for lettuce with an average insert size of 111 kb (Frijters *et al.*, 1997); additional libraries may exist in the private sector. A ~8x BAC library with average insert size of ~100 kb has been constructed for sunflower (xxxx). Neither of these libraries have been fingerprinted. BAC libraries do not exist for other Compositae species.

Sequence Information, EST databases.

The Compositae Genome Project (CGP) was initiated with funding from the USDA IFAFS program and continued with support from the NSF Plant Genome Program. The CGP has developed extensive EST data that have allowed both lettuce and sunflower to partly catch up with other more intensively studied species.

The first phase of the CGP generated over 132,000 ESTs from lettuce and sunflower. EST libraries were made from ten pools of RNA from different tissues/developmental stages/environmental conditions of each of the two genotypes that had been used as parents for the core mapping populations for each species. Over 68,000 ESTs were generated from *L. sativa* and *L. serriola*. Likewise, more than 67,000 ESTs were generated for sunflower, including 44,000 from cultivated confectionary and oilseed sunflower lines and 23,000 from drought- and salt-tolerant wild sunflowers, *H. argophyllus* and *H. paradoxus*, respectively.

The second phase of the CGP has generated over 600,000 ESTs from several tissues of 17 species within the Compositae (Table 3). EST data as well as the trace files are displayed at the CGP web site (<http://cgpdb.ucdavis.edu/>), and have been released to GenBank. Clones generated by the CGP can be obtained on a recharge basis from the Arizona Genomics Institute (AGI) (<http://www.genome.arizona.edu/>).

There are several other EST projects in progress in addition to the CGP. These include 6,000 normalized ESTs for *Centaurea maculosa* (J. Vivanco, Colorado State University), 9,777 floral ESTs for *Senecio* spp. (S. Hiscock, University of Bristol, UK), 40,000 ESTs for *Senecio* spp. (A. Lowe, University of Queensland, Australia), xxxx for *Taraxacum* xxxx (D. Shinati, U Nevada, Reno), *French and Argentinean sunflower efforts??*.

Other EST projects??

There are a total of nearly 800,000 ESTs from over 30 species available from NCBI. These represent over 1.5% of all eukaryote ESTs and 6% of all plant ESTs. Because many of the clones available from the AGI are full-length, these represent a tremendous resource for functional and comparative studies.

Table 3. EST Resources, February 2007.

Species	Common name	Total # ESTs	% CGP	# Unigenes
<i>Helianthus annuus</i>	Common sunflower	94110	76.58	31605
<i>Lactuca sativa</i>	Cultivated lettuce	80779	99.25	26720
<i>Lactuca serriola</i>	Prickly lettuce	55490	100.00	19877
<i>Cichorium intybus</i>	Chicory	41747	91.80	22038
<i>Taraxacum officinale</i>	Dandelion	41296	100.00	15761
<i>Carthamus tinctorius</i>	Safflower	40878	100.00	19395
<i>Centaurea solstitialis</i>	Yellow starthistle	40407	100.00	22917
<i>Helianthus tuberosus</i>	Jerusalem artichoke	40361	100.00	21994
<i>Centaurea maculosa</i>	Spotted knapweed	39958	99.99	20922
<i>Helianthus argophyllus</i>	Silverleaf sunflower	35720	100.00	17839
<i>Helianthus exilis</i>	Serpentine sunflower	33960	100.00	19784
<i>Lactuca saligna</i>	Willowleaf lettuce	30696	100.00	11491
<i>Helianthus paradoxus</i>	Puzzle Sunflower	30517	100.00	17874
<i>Cichorium endivia</i>	Endive	30171	100.00	18951
<i>Lactuca virosa</i>	Bitter lettuce	30068	100.00	11865
<i>Lactuca perennis</i>	Blue lettuce	29125	100.00	11472
<i>Helianthus petiolaris</i>	Prairie sunflower	27484	100.00	12982
<i>Helianthus ciliaris</i>	Texas blueweed	21590	100.00	14857
<i>Zinnia elegans</i>		17975	0.00	
<i>Gerbera hybrid cv. Terra Regina</i>	Gerbera	15851	0.00	8665
<i>Stevia rebaudiana</i>	Stevia (sugarleaf)	5548	0.00	3345
<i>Taraxacum kok-saghyz</i>		4731	0.00	3363
<i>Senecio cambrensis</i>		2019	0.00	
<i>Senecio chrysanthemifolius</i>		2019	0.00	
<i>Senecio vulgaris subsp. vulgaris</i>		1956	0.00	
<i>Senecio squalidus subsp. squalidus</i>		1925	0.00	
<i>Senecio aethnensis</i>		1884	0.00	
<i>Gerbera hybrid cv. Terra Nero</i>	Gerbera	1146	0.00	
<i>Cirsium arvense</i>		173	0.00	
<i>Ageratum conyzoides</i>		20	0.00	
<i>Tagetes erecta</i>		2	0.00	
<i>Artemisia apiacea</i>		1	0.00	
All Asteraceae		799607	89.83	

Other initiatives have generated differential cDNA libraries using suppressed subtractive hybridization (Diatchenko *et al.* 1996) to identify low-copy mRNA and differentially expressed sunflower transcripts from several developmental stages (Fernández *et al.* 2003). This included 319 unique sunflower unigenes related to responses to abiotic and biotic stresses as well as low abundant transcripts with high similarity to homeobox genes, transcription factors, and signaling components. Functional SNPs and IN/DELS markers are being characterized in a set of 20 inbred lines including parental genotypes of three mapping populations. A total of 52 SNPs and 15 indels were preliminarily obtained with an overall frequency of 1 SNP per 91 bp and 1 indel per 326 bp in sunflower genome.

Phenotypic Information

QTLs have been or are being identified for numerous traits including root architecture, bolting, branching, spines, leaf development, leaf succulence, shoot biomass, flowering, seed size, seed oil concentration and composition, seed germination, mineral ion uptake traits, photosynthetic rate, water use efficiency, nutrient uptake, disease resistance, senescence, tolerance to water deficit, etc. These are being mapped relative to candidate genes (Table 4) for

nutrient ion transporters, enzymes associated with seed dormancy and germination, disease resistance and response, drought, salt and stress tolerance, glycerolipid, tocopherol, and flavonoid synthesis, and floral, leaf and root development.

Table 4. (Numbers need to be updated)

Traits	# candidates to date		# so far mapped	
	Lettuce	Sunflower	Lettuce	Sunflower
Disease resistance	724	140	273	60
Developmental	250	114	68	59
Abiotic stress	49	102	6	68
Physiological	199	211	17	105
Other ^a	302	1103	302	433
Total	1524	1655	666	725

^a includes genes mapped due to obvious indel polymorphisms.

Databases

A series of public and private databases provide access to Compositae genomic data: Compositdb contains genetic map and marker information; Lettvcv contains information on over 4,500 cultivars; Sundb contains information for 214 markers on 10 genotypes; the CGPdb contains extensive EST and genetic information (<http://compgenomics.ucdavis.edu/>).

Resources Under Development

Genetic stocks

Several new RIL populations are being generated for lettuce, in particular 300 RILs from an intra-specific *L. sativa* cv. Salinas x cv. Valmaine that will allow the dissection of the determinants of plant type. (M. Truco, R. Michelmore, unpublished).

For sunflower, Cms HA89 x Ann1238 (elite x wild progenitor) population consisting of 192 RILs is being generated in the US. Also, the Hopi x Ann1238 (primitive domesticate x wild progenitor) population currently consisting of 384 F₂s, could be advanced to RILs depending on resources. Argentinean mapping strategies include the development of advanced population for resistance/tolerance to biotic and abiotic stress: a) two advanced (F₅) mapping populations derived from RHA801 x HA89 and RK416 x HA89 (EEA INTA Balcarce, Argentina) for resistance to *Sclerotinia sclerotiorum* and additional mapping populations for resistance to *Verticillium dahliae*, and b) drought and middle drought resistance populations derived from R 417 x R 419, A 59 x R423 and HA 64 x HA 89 (F₃) were developed at EEA INTA Manfredi, Argentina.

Mapping populations and RILs involving annual x annual and annual x perennial forms of *Leontodon* are being developed (L. Rieseberg, unpublished).

Others?

Genomic Resources

An Affymetrix chip has been developed for massively parallel genetic mapping and sensitive expression analysis of lettuce. This chip has a total of 6.6 million features representing ~29,000 unigenes, each with ~170 oligonucleotides with a 2 bp stagger (A. van Deynze, R. Michelmore *et al.* unpublished). This chip is available and protocols for using and analyzing it

are under development. An Agilent array (60mer) will also be developed for lettuce (~1000 annotated genes from the EST database, M. Jeuken).

The development of an oligonucleotide array for sunflower is also under development. This will be likely be similar, although less dense Affymetrix chip with 2.6 million features and 80-100 probes per unigene.

The utility of these chips for the analysis of other Compositae species remains to be determined.

Genomic Sequence Information

There is only limited genomic sequence information for any species in the Compositae. There has yet to be a coordinated effort to sequence any genome completely.

Needs

These are displayed by subject area and in approximate priority within in each area.

Research

Genetics and Genetic Stocks

(1) Genetic analysis of agriculturally important traits and those involved in domestication and invasiveness. This genetic characterization provides the foundation for the functional and comparative genomics described below. Traits of importance include:

- (a) Disease resistance. There are numerous sources of disease resistance with varying degrees of genetic characterization.
- (b) Seed oil composition. There have been multiple domestications oil seed crops within the Compositae.
- (c) Flowering time and bolting. There is wide variation within and between species for these agriculturally important traits.
- (d) Seed dormancy, physiology, size, vigor and longevity. There are large amounts of phenotypic variation that are beginning to be characterized.
- (e) Vegetative development, leaf morphology and branching. There are many examples of convergent and divergent evolution within and between species of the Compositae
- (f) Floral development. The Compositae have a rich diversity in floral architectures that are unique to this family.
- (g) Abiotic tolerances (particularly drought and salt). The Compositae offer a rich source of genetic diversity as they are adapted to many extreme environments. This has yet to be exploited.
- (h) Domestication traits, self-incompatibility, shattering. The multiple parallel domestications in the Compositae provide the opportunity to identify the genes critical to domestication and the effects of domestication on their evolution.
- (i) Post-harvest quality. Little is known about the genetic basis of this complex phenotype.
- (j) Secondary metabolism. The Compositae provide an unparalleled diversity of secondary metabolites that await further characterization.
- (k) Secondary products (particularly rubber). There are several Compositae spp. that make high quality rubber.
- (l) Genetic basis of invasiveness traits. Very little is known of the genetic basis of weediness. Analyses of the spread of herbicide tolerance and inter-specific hybrid fitness are needed.
- (m) Genetic basis of phenotypic diversification in wild model spp.

- (2) More RIL populations, particularly RIL populations with inter-mated progenitors for lettuce and sunflower. This is needed to provide enhanced genetic resolution of mapping.
- (3) Development of NILs differing for individual QTLs, for functional analyses and sexual or transgenic introgression of traits of agricultural or evolutionary importance.
- (4) Development of mapping and RIL populations for minor Compositae crops (safflower) and model weed species (*Leontodon*, dandelion, and spotted knapweed).
- (5) Development of mapping populations and RILs for species important for evolutionary studies, silverswords, wild sunflowers, *Senecio* spp., and *Stephanomeria*. Hexaploid-diploid crosses for evolutionary studies of polyploidy in *Helianthus* species.

Genomic resources

- (1) *Massively parallel genetic analysis of transcribed genes.*

The high density oligonucleotide chips should be exploited to analyze allelic variation in multiple populations. This will locate many unigenes (10,000s?) in bins. The size of the bins will be determined by the sizes of the segregating populations analyzed and therefore the numbers of cross-overs. There is a need to increase the sizes of the populations analyzed to reduce the bin size. In the long term, this will provide a genetically-defined framework to anchor genomic shotgun sequences.

- (2) *Tiled oligonucleotide arrays for sunflower and other Compositae spp.*

It is currently unclear how useful Affymetrix arrays designed from *Lactuca* and sunflower sequences will be for other species. However, it is possible that chips will at least be useful for species within individual tribes. It will be particularly beneficial if the lettuce array can be used to analyze *Leontodon*. The novel *Lactuca* array will be a learning experience that will allow the design of arrays with enhanced utility for other Compositae species. The oligonucleotide arrays will also be useful for identifying sequences that may be useful as PCR primers for amplifying sequences from little-studied exotic Compositae species.

- (3) *Greater depth in BAC libraries.*

Current BAC libraries are ~3x for lettuce and ~8x for sunflower. At least 10 to 15x coverage is desirable.

- (4) *BAC fingerprints, contigs, minimum tiling path, and BAC end sequences.*

As a foundation for sequencing, it would be useful to develop BAC fingerprints, generate a minimum tiling path, and sequence their ends. This would provide a framework for assemblies of random shotgun sequences and complement the genetically-defined bins of unigenes.

- (5) *Development of tools for Leontodon as a model.*

Considerable resources will be required to develop *Leontodon* as a useful model. These include EST sequencing, BAC library construction, and physical mapping.

- (6) *Integration of genetic and cytological maps.*

FISH, GISH, BAC-FISH are being developed for sunflower at the University of Buenos Aires-Instituto de Biotecnología INTA, Argentina. The hexaploid *H. resinosus* is currently investigated using molecular cytogenetics and markers in order to elucidate species origin through polyploidy (Carrera *et al.* 2004).

Gene Discovery and Sequencing

- (1) *Sequencing of the gene space for lettuce and sunflower.*

Sequencing the gene space of lettuce and sunflower is a high priority as it will enable proteomic approaches and provide alleles and regulatory sequences for transgenic manipulation. Gene discovery through conventional EST sequencing is probably approaching saturation. Based on data from other species, sequencing of the high C₀t and methyl filtered fractions will provide over 90% of the genic sequences in these two crops. Alternative strategies for accessing the gene space of these species should be applied as they become available. Such sequencing would supply the fraction of coding regions that is currently missing and as well as sequences of introns, promoters, and 5' and 3' UTRs.

(2) *Analysis of allelic and orthologous series of agriculturally important and evolutionarily significant genes.*

Resequencing projects are needed to access allelic variation across numerous genotypes within and between species.

(3) *Exploratory sequencing of Leontodon.*

Sequencing of low numbers of random genomic sequences as well as the gene space of *Leontodon* will provide data on the desirability of whole genome sequencing of *Leontodon*.

Whole Genome Sequencing.

The new generation of sequencing technologies is revolutionizing the opportunities to sequence large genomes. Individually, however, they are still less than ideal. We should prepare for the whole genome sequencing of several Compositae species in the next ~five years using a combination of sequencing approaches.

We envision three phases:

Phase one, immediate priorities:

Sequencing the gene space of lettuce and sunflower using a combination of sequencing approaches.

Sequencing of a small number of BACs from high interest areas in lettuce and sunflower.

Developing resources and information for lettuce and sunflower: high density maps of transcribed sequences, minimal tiling path of BACs, and xxxx.....

Developing of *Leontodon* as a small genome model.

Developing of bioinformatics capabilities for acquiring, curating, and querying large amounts of sequence data.

Organizing the community to share the work load and generate the funding.

Phase two, intermediate term:

Draft whole genome, random shotgun sequencing of lettuce and sunflower using several technologies: e.g. 1 – 2x skim sequencing with Sanger, 5 – 10⁺x with one or more high-throughput, paired-end technologies (454 or Solexa).

BAC end sequencing of the minimum tiling paths of lettuce and sunflower.

Draft 1 – 2x genome sequencing of *Leontodon*.

Selective sequencing of BACs from high interest areas from multiple Compositae species.

Phase three, longer term:

High-quality sequencing of lettuce, sunflower, and *Leontodon*.

Draft sequencing of key nodal species across the Compositae.

Re-sequencing of close relatives of lettuce and sunflower.

The exact timescale and the amount of overlap between of each of these three phases will be determined by the availability of funding and the advancement of sequencing technologies.

Functional Genomics

(1) *Understanding the molecular basis of agriculturally important traits as well as those involved in domestication and invasiveness.*

As detailed above, disease resistance, dormancy, flowering time, seed oil traits, salt and drought tolerance, and seed size, vigor and longevity are all traits that require detailed analysis at the molecular level.

Bulked segregant analysis using Affymetrix chips will rapidly and efficiently correlate phenotypes with candidate genes to provide functional maps of sunflower, lettuce, and *Leontodon*. The rate-limiting step will be the generation of high quality, phenotypic data on large populations and accessions.

(2) *Improved methods for gene silencing.*

Current projects will generate many correlations between phenotypes and candidate genes. A major challenge will be to move from gene/trait correlations to functional assays. The development of VIGS and efficient RNAi are needed for silencing of individual genes and gene families.

(3) *Comparative functional genomics.*

The functional significance of allelic variation for traits of agricultural and/or evolutionary importance should be determined.

(4) *Massively-parallel genetic and expression analyses across other crops, weeds, and adaptive radiation/speciation models in the Compositae.*

Development and utilization of microarrays or tag sequencing for global transcriptomic analyses of specific developmental stages associated with genetic and phenotypic variation for traits of interest. Use of the tiling arrays to obtain differential expression data among genotypes will link genetic variation to expression and phenotypic variation.

Database Development, Data Visualization

(1) *Acquisition, curation, visualization of increasingly complex datasets from multiple sources.* It will be particularly challenging to assimilate and exploit the large amount of sequence data that will be generated by the next generation of sequencing technologies.

(2) *Tools for data visualization and enhanced queries across heterogeneous datasets.*

(3) *Universal ontology for phenotypic information.* This is required for integration of datasets and queries across different species within the Compositae as well as with species in other families.

Translation (Breeding)

(1) *High-throughput, inexpensive marker systems and robust molecular markers for agriculturally important traits.*

Current and future studies have/will identify numerous trait-marker associations. However, many of the current marker technologies are too expensive and laborious for routine analysis of breeding materials. Methods for are required to identify and exploit molecular markers in a breeding context. This includes the development of inexpensive automated DNA extraction methods and efficient SNP discovery as well as subsequent analysis. Total costs should be below a few cents per datapoint. Assay systems should be flexible and allow variable numbers (in the few hundred range) and compositions of markers. Disease resistance should be an earlier application of marker-assisted selection.

(3) *Methods for stable, regulated transgene expression.*

Agrobacterium-mediated transformation is routine for lettuce but not sunflower. However, for both species introduced transgenes become silenced and transgenic lines with

adequate levels are rare. Research is needed to increase transformation efficiencies in recalcitrant Compositae species and to provide a variety of predictable transgene expression patterns.

(4) *Analysis and access of broader germplasm.*

The sequences generated, particularly the candidate genes, will be useful for introgression of favorable alleles from wild germplasm using marker-assisted selection.

(5) *Development of stress tolerant plants.*

There is great intra- and inter-specific for resistance to abiotic stresses that has yet to be exploited but could greatly reduce the impact of agriculture as well as improve the efficiency and predictability of several Compositae crops.

(6) *Improvement of post-harvest quality and shelf life.*

Several Compositae crops are subject to major post-harvest losses. Application of the developing knowledge of the physiological and pathological bases of these losses provides opportunities to mitigate these losses.

(7) *Development of reliable strategies for transgene containment (e.g., transgenic mitigation; Gressel, 1999; Stewart, 2004).*

Some Compositae crops, including lettuce and sunflower, are grown in the same geographic region as sexually compatible wild relatives. This raises the possibility of transgene escape into the wild population. Research and implementation of strategies to minimize transgene flow from crops to wild relatives will assist the acceptance of transgenics for these and other crops.

Education

(1) Training of students in plant breeding methods to take advantage of the large amounts of genomics data available for Compositae species.

(2) Training of students in functional, comparative, and evolutionary genomics using the opportunities afforded by the Compositae.

(3) Training of the next generation of weed scientists with a particular focus on the many noxious weeds in the Compositae.

(4) Training of students in physiological genetics to exploit the diversity of abiotic stress phenotypes exhibited by the Compositae.

(5) Exposure of high school students and high school science teachers to current issues and technology used to sustain and increase food production while minimizing environmental impact.

(6) Developing alternative approaches to train plant breeders to supplement the reduced numbers of academic plant breeding programs, particularly in specialty crops. Professional education programs can be targeted to personnel currently employed in the agricultural or seed industry to enhance their theoretical and practical knowledge of genetics and breeding, as for example in the UC Davis Plant Breeding Academy

(http://sbc.ucdavis.edu/Events/Plant_Breeding_Academy.htm).

Extension

Workshops targeted at plant breeders and weed scientists.

Workshops are needed to publicize the molecular markers and genetic maps generated in the Compositae and to educate breeders in their application in these crops. A model for such workshops is the Breeding with Molecular Markers extension course presented by the UC Davis

Seed Biotechnology Center (<http://sbc.ucdavis.edu/Events/Courses.htm>). Such workshops should emphasize hands-on sessions illustrating how to access genomic and marker information and apply it in practical breeding programs.

Development of new courses such as the one developed in 2004 at U. Mass, Boston that apply new understandings of invasive species biology to land management strategies (<http://bio697.blogspot.com/>).

Stakeholders involved in this draft

This draft reflects the input from the PIs of the Compositae Genome Project: Richard Michelmore (UC Davis), Loren Rieseberg (U. Indiana), K. Bradford (UC Davis), J. Burke (Vanderbilt U), D. Still (Cal Poly, Pomona), R. Kesseli (U. Mass, Boston), and S. Knapp (U. Georgia). There has also been significant input from M. Jeuken (U. Wageningen), and Ruth Heinz (INTA, Buenos Aires) as well as from participants in the Compositae White Paper meeting.

The Compositae White Paper meeting was held at UC Davis in August 2006 to get input from the broad community of researchers working on Compositae species (<http://compgenomics.ucdavis.edu/cwp2006/>). This was attended by ~50 researchers from diverse scientific backgrounds and several different countries. The program consisted of scientific presentations in the morning and break-out groups to discuss priorities followed by an integrating session in the afternoon. Input from this meeting has been incorporated into this draft.

We are now seeking additional input and refinement from the broad community of researchers working on Compositae species.

Stakeholder groups that have or will be consulted in the continued development of this white paper (partial list; please suggest others)

Large community of plant systematists, evolutionary biologists, and weed biologists study who Compositae species.

PIs of the Compositae Genome Project.

The NSF Floral Genome Project.

Breeders of Compositae crops.

USDA Leafy Vegetable Germplasm Committee.

California Lettuce Research Board <http://www.calettuceresearchboard.org/>

Seed companies breeding Compositae species.

USDA facilities involved in maintenance and study of Compositae species

Land managers and environmental protection agencies interested in managing invasive or endangered Composite species

Others???

References: Only partial listing. Not checked.

- Aguirrezabal L, Bouchier-Combaud S, Radziejwoski A, Dauzat M, Cookson SJ, Granier C. 2006. Plasticity to soil water deficit in *Arabidopsis thaliana*: dissection of leaf development into underlying growth dynamic and cellular variables reveals invisible phenotypes. Plant, Cell and Environment, accepted for pub
- Aguirrezabal, L.A.N., Lavaud, Y., Dosio, G.A.A. Izquierdo, N.G. Andrade F.H. and González. L.M. 2003. Intercepted Solar Radiation Effect During Seed Filling Determines Sunflower Weight per Seed and Oil Concentration. Crop Science 43: (1) 152-161

- Al-Chaarani GR, Gentzbittel L, Huang XQ, Sarrafi A (2004) Genotypic variation and identification of QTLs for agronomic traits, using AFLP and SSR markers in RILs of sunflower (*Helianthus annuus* L.). *Theor Appl Genet* 109:1353-1360
- Al-Khatib K, Baumgartner JR, Peterson DE, Currie RS. (1998). Imazethapyr resistance in common sunflower (*Helianthus annuus*). *Weed Sci.* **46**:403-407.
- Anon. (2003) http://invader.dbs.umt.edu/noxious_weeds/noxlist.asp.
- Argyris J., Truco M.J., Ochoa O., Knapp S.J., Still D.W., Lenssen G.M., Schut J.W., Michelmore R.W., Bradford K.J. (2005). Quantitative trait loci associated with seed and seedling traits in *Lactuca*. *Theor. Appl. Genet.* **111**:1365-1376.
- Arias, D.M., Rieseberg, L.H. (1994). Gene flow between cultivated and wild sunflowers. *Theor. Appl. Genet.* **89**:655-660.
- Baker, H.G. (1974). The evolution of weeds. *Annu. Rev. Ecol. Syst.* **5**:1-24.
- Bennett MD, Leitch IJ, Price HJ and Johnston JS. 2003. Comparisons with *Caenorhabditis* (~100 Mb) and *Drosophila* (~ 175 Mb) using flow cytometry show genome size in *Arabidopsis* to be ~157 Mb and thus ~25 % larger than the Arabidopsis Genome Initiative estimate of ~125 MB. *Annals Bot.* **91**: 547-557.
- Burke, J.M. Rieseberg, L.H. (2003). Fitness effects of transgenic disease resistance in sunflowers. *Science* **300**: 1250.
- Burke, J.M., S. Tang, S.J. Knapp, Rieseberg, L.H. (2002). Genetic analysis of sunflower domestication. *Genetics* **161**: 1257-1267.
- Burke, J.M., S.J. Knapp, Rieseberg, L.H. (2005). Genetic consequences of selection during the evolution of cultivated sunflower. *Genetics* **171**: 1933-1940.
- Burke, J.M., Z. Lai, M. Salmaso, T. Nakazato, S. Tang, A. Heesacker, S.J. Knapp, and L.H. Rieseberg. 2004. Comparative mapping and rapid karyotypic evolution in the genus *Helianthus*. *Genetics* **167**: 449-457.
- Carrera A, Poverene M, Rodríguez R. (2004) Isozyme and cytogenetic analysis in *Helianthus resinusus* Small. 16th International Sunflower Conference. Fargo, North Dakota. USA. August 29- September 2, 2004. 685-691.
- Chan RL, Gago GM, Palena CM, Gonzalez DH (1998) "Homeoboxes in Plant Development" *Biochimica et Biophysica Acta* 1442, 1-19
- Covacevich, F. Echeverría, H.E. y L.A.N. Aguirrezábal. Soil available phosphorus status determines indigenous mycorrhizal colonization into field and glasshouse-grown spring wheat in Argentina.. *Applied Soil Ecology*. MS 05/145. (In press)
- Cronquist, A. (1977). The Compositae revisited. *Brittonia* **29**:137-153.
- Crute, I.R. (1988). The impact of breeding on pest and disease control in lettuce. *Aspects Appl. Biol.* **17**:305-312.
- deVries, I.M. (1997). Origin and domestication of *Lactuca sativa* L *Genet. Res. Crop Evol.* **44**:165-174
- Dezar CA, Fedrigo GV, Chan RL (2005) "The promoter of the sunflower HD-Zip protein gene *Hahb4* directs tissue-specific expression and is inducible by water stress, high salt concentrations and ABA" *Plant Science* 169, 447-459
- Dezar CA, Gago GM, González DH, Chan RL (2005) "*Hahb-4*, a sunflower homeobox-leucine zipper gene, confers drought tolerance to *Arabidopsis thaliana* plants" *Transgenic Research* 14, 429-440.
- Diatchenko L, Lau YF, Campbell AP, Chenchick A, Moqaddam F, Huang B, Lukyanov S, Konstantin L, Gurskaya N, Sverdlov E, Siebert PD (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci USA* 93:6025-6030

- Dosio GA, Rey H, Lecoeur J, Izquierdo NG, Aguirrezabal LA, Tardieu F, Turc O. (2003) A whole-plant analysis of the dynamics of expansion of individual leaves of two sunflower hybrids. *J Exp Bot.*;54(392):2541-52.
- Dosio GA, Tardieu F, Turc O (2006) How does the meristem of sunflower capitulum cope with tissue expansion and floret initiation? A quantitative analysis. *New Phytol.* 2006;170(4):711-22
- Ellstrand, N. (2003). *Dangerous Liaisons? When Cultivated Plants Mate With Their Wild Relatives*. Johns Hopkins, Baltimore.
- Fernández P, Paniego N, Lew S, Hopp HE, Heinz RA (2003) Differential representation of sunflower ESTs in enriched organ-specific cDNA libraries in a small scale sequencing project. *BMC Genome* 4:40
- Forman, J. (2003). The introduction of American plant species into Europe: issues and Consequences. In: *Plant Invasions: Ecological Threats and Management Solutions*. Child, L.E., Brock, J.H., Brundu, G., Prach, K., Pysek, P., Wade, P.M., Williamson, M., Eds. Backhuys Publishers, Leiden, The Netherlands, pp. 17-39.
- Fritjers, A.C.J., Zhang, Z., van Damme, M., Wang, G.-L., Ronald, P.C. and Michelmore, R.W. (1997). Construction of a bacterial artificial chromosome library containing large *EcoRI* and *HindIII* genomic fragments of lettuce. *Theor. Appl. Genet.* **94**:390-399.
- Funk, V.A., Bayer, R.J., Keeley, S., Chan, R., Watson, L., Gemeinholzer, B., Schilling, E., Panero, J.L., Baldwin, B.G., Garcia-Jagas, N., Susanna, A., Jansen, R.K. 2005. Everywhere but Antarctica: Using a supertree to understand the diversity and distribution of the Compositae. *Biol. Skr.* **55**: 343-374.
- Gago GM, Almoguera C, Jordano J, González DH, Chan RL (2002) "Hahb-4, a homeobox-leucine zipper gene potentially involved in ABA-dependent responses to water stress in sunflower" *Plant Cell and Environment* 25, 633-640.
- Gonorazky AG, Regente MC, de la Canal L Stress induction and antimicrobial properties of a lipid transfer protein in germinating sunflower seeds (2005). *J Plant Physiol.* 162(6):618-24.
- González DH, Valle EM, Gago G, Chan RL (1997) "Interaction between proteins containing homeodomains associated to leucine zippers from sunflower" *Biochimica et Biophysica Acta.* 1351, 137-149.
- Gressel, J. (1999). Tandem constructs: preventing the rise of superweeds. *Trends Biotech.*
- Granier, C, L. Aguirrezabal, K. Chenu, S.J. Cookson, M. Dauzat, P. Hamard, J.J. Thioux, G. Rolland, S. Bouchier-Combaud, A. Lebaudy, B. Muller, T. Simonneau and F. Tardieu. 2006 PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in *Arabidopsis thaliana* permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytologist*, 169 (3), 623-635
- Harlan, J.R. (1975). *Crops & Man*. Amer. Soc. Agron., Madison, WI.
- Harter, A.V., Gardner, K.A., Lentz, D.L., Bye, R., Rieseberg L.H. (2004). Single origin of extant domesticated sunflowers in eastern North America. *Nature*, in review.
- Heiser, C.B. and Smith, D.M. (1955). New chromosome numbers in *Helianthus* and related genera. *Proc. Ind. Acad. Sci.* **64**:250-253.
- Heywood, V.H. (1978). *Flowering Plants of the World*. Prentice Hall, Englewood Cliffs, NJ.
- Jain, S., Kesseli, R., Olivieri, A. (1993). Biosystematic status of serpentine Sunflower, *Helianthus exilis* Gray. In: *The Vegetation of Ultramafic (Serpentine) Soils*. Proctor, J., Baker, A., Reeves, R., Eds. Intercept Ltd, Andover UK.
- Jan, C.C. (2000). Cytoplasmic male sterility in two wild *Helianthus annuus* L. accessions and their fertility restoration. *Crop Science* **40**:1535-1538.

- Jansen R.K., Michaels H.J., Palmer J.D. (1991). Phylogeny and character evolution in the Asteraceae based on chloroplast DNA restriction site mapping. *Syst. Bot.* **16**:98-115.
- Jeuken M, van Wijk R, Peleman P and Lindhout P (2001) An integrated interspecific AFLP map of lettuce (*Lactuca*) based on two *L. sativa* × *L. saligna* F₂ populations. *Theor Appl Genet* **103**:638-647.
- Jeuken MJW and Lindhout P (2004) The development of lettuce backcross inbred lines (BILs) for exploitation of the *Lactuca saligna* (wild lettuce) germplasm. *Theor Appl Genet* **109**: 394-401
- Johnson, W.C., Jackson, L.E., Ochoa, O., Peleman, J., van Wijk, R., St.Clair, D.A., Michelmore, R.W. (2000). A shallow-rooted crop and its wild progenitor differ at loci determining root architecture and deep soil water exploitation. *Theor. Appl. Genet.* **101**:1066-1073.
- Kesseli, R.V. and Michelmore, R.W. (1997). The Compositae: systematically fascinating but specifically neglected. In: *Genome Mapping in Plants*. A.H. Paterson, ed. R.G. Landes Co., Georgetown, TX. pp. 179-191.
- Kesseli, R.V., Ochoa, O., Michelmore, R.W. (1991). Variation at RFLP loci in *Lactuca* spp. and origin of cultivated lettuce. *Genome* **34**:430-436.
- Kesseli, R.V., Paran, I., Michelmore, R.W. (1994). Analysis of a detailed genetic linkage map of *Lactuca sativa* (lettuce) constructed from RFLP and RAPD markers. *Genetics* **136**:1435-1446.
- Kesseli, R.V., Witsenboer, H., Stanghellini, M., Vandermark, G., Michelmore, R.W. (1993). Recessive resistance to *Plasmopara lactucae-radicis* maps by bulked segregant analysis to a cluster of dominant disease resistance genes in lettuce. *Molec. Pl. Microbe Interact.* **6**:722-728.
- Koopman, W.J.M., De Jong, J.H. (1996). A numerical analysis of karyotypes and DNA amounts in lettuce cultivars and species (*Lactuca* subsect *Lactuca*, Compositae). *Acta Bot. Neerl.* **45**:211-222.
- Landry, B.S., Kesseli, R.V., Farrara, B., Michelmore, R.W. (1987). A genetic map of lettuce (*Lactuca sativa* L.) with restriction fragment length polymorphism, isozyme, disease resistance, and morphological markers. *Genetics* **116**:331-337.
- Linder, C.R., Taha, I., Seiler, G.J., Snow A.A., Rieseberg, L.H. (1998). Long-term introgression of crop genes into wild sunflower populations. *Theor. Appl. Genet.* **96**:339-347.
- Lindstrom, L.I., Pellegrini, C.L., Aguirrezábal, L.A.N. y L.F. Hernández. 2006 "Growth and Development of Sunflower Fruits under shade During Pre and Early Post-anthesis Period. *Field Crops Research* 96 (151-156)
- Maisonneuve, B., Bellec, Y., Anderson, P., Michelmore, R.W. (1994). Rapid mapping of two genes for resistance to downy mildew from *Lactuca serriola* to existing clusters of resistance genes. *Theor. Appl. Genet.* **89**:96-104.
- Manavella PA, Arce AL, Dezar CA, Bitton F, Renou JP, Crespi M, Chan RL (2006) Analysis of transcriptional networks regulated by the sunflower *Hahb-4* transcription factor in *Arabidopsis* revealed a conserved role in ethylene signaling. *The Plant Journal* 48(1) 125-137
- Michaelson M.J., Price H.J., Ellison J.R., Johnston J.S. (1991). Comparison of plant DNA contents determined by feulgen microspectrophotometry and laser flow cytometry. *American J.I Botany* **78**: 183-188.
- Paniego N, Echaide M, Muñoz M, Fernández L, Torales S, Faccio P, Fuxan I, Carrera M, Zandomeni R, Suárez I, Hopp EH (2002) Microsatellite isolation and characterization in sunflower (*Helianthus annuus* L.). *Genome* 45:34-43

- Paniego N, Heinz R y Hopp HE (2006) Sunflower. En: "Genome Mapping & Molecular Breeding" Vol. II (Oilseeds), Ed. Springer-Verlag- Germany VoII . En prensa
- Pennisi, E. (2003). Garden flower evolves into weed. *Science* **301**:458-459.
- Pimentel, D. Lach, L. Zuniga, R., Morrison, D. (1999). Environmental and economic costs associated with non-indigenous species in the United States.
http://www.news.cornell.edu/releases/Jan99/species_costs.html
- Poormohammad Kiani; P. Talia ; V Nishinakamasu; P; Maury; P Grieu; A. Perrault ; E Hopp; R. Heinz ;L. Gentzbittel ;. N. Paniego; A. Sarrafi. (2006) Development of a new AFLP and SSR genetic linkage map and identification of QTLs for photosynthesis related traits under well-watered and water-stressed conditions in sunflower (*Helianthus annuus* L.) submitted for publication
- Poverene M, Carrera A., Ureta S., Cantamutto M. (2004a) Wild *Helianthus* species and wild-sunflower hybridization in Argentina. *Helia*, 27(40): 133-142.
- Poverene M., A. Carrera, M. Cantamutto, S. Ureta, C. Delucchi, V. Alonso Roldan, J. Basualdo. (2004b) *Helianthus petiolaris* in Argentina and spontaneous hybridization with cultivated sunflower. 16th. International Sunflower Conference, Fargo ND, USA, Vol. II: 741-746.
- Price H.J., Hodnett G., Johnston J.S. (2000). Sunflower (*Helianthus annuus*) leaves contain compounds that reduce nuclear propidium iodide fluorescence. *Annals Bot.* **86**: 929-934.
- Radonic LM, Zimmermann JM, Zavallo D, López N, López Bilbao MG (2006) Rooting in Km selective media as efficient in vitro selection method for sunflower genetic transformation. *Electronic J. of Biotechnology*, Vol 9 n° 3.
- Regente MC, Giudici AM, Villalain J, de la Canal L. The cytotoxic properties of a plant lipid transfer protein involve membrane permeabilization of target cells. *Lett Appl Microbiol.* 40(3):183-9.
- Rehorek, V. (1997). Cultivated and escaped perennial *Helianthus* species in Europe. *Preslia-Prague* **69**:59-70.
- Rieseberg, L.H., Raymond, O., Rosenthal, D.M., Lai, Z., Livingstone, K., Nakazato, T., Durphy, J.L., Schwarzbach, A.E., Donovan L.A., Lexer C. (2003). Major ecological transitions in annual sunflowers facilitated by hybridization. *Science* **301**:1211-1216.
- Robbins, M.A., Witsenboer, H. Michelmore, R.W., Laliberte, J.F., Fortin, M. (1994). Genetic mapping of turnip mosaic virus resistance in *Lactuca sativa*. *Theor. Appl. Genet.* **89**:583-589.
- Rueda EC, Dezar CA, Gonzalez DH, Chan RL (2005) "*Hahb-10*, a sunflower homeobox-leucine zipper gene, is involved in the response to dark/light conditions and promotes a reduction of the life cycle when expressed in *Arabidopsis*" *Plant and Cell Physiology* 46(12), 1954-1963.
- Ryder, E.J. (1986). Lettuce breeding. In: *Breeding Vegetable Crops*. AVI Pub., Westport, CT., pp 433-474.
- Schilling, E.E., Heiser, C.B. (1981). Infrageneric classification of *Helianthus* (Compositae). *Taxon* **30**:393-403.
- Seiler, G.J. and Rieseberg, L.H. (1997). Systematics, origin and germplasm resources of the wild and domesticated sunflower. In: *Sunflower Technology and Production*. Schneiter, A.A., ed. Amer. Soc. Agronomy, Madison, WI., pp. 21-66.
- Smith, C.R. (1985). Unusual seed oils and their fatty acids. In: *Fatty acids*. Pryde, E.H., eds. Amer. Oil Chemists Soc., Champaign-Urbana, IL, pp. 29-47.
- Snow, A.A., Pilson, D., Rieseberg, L.H., Paulsen, M., Pleskac, N., Reagon, M.R., and Wolfe, D.E. (2003). A Bt transgene reduces herbivory and enhances fecundity in wild sunflowers *Ecol. Appl.* **13**:279-286.

- Stewart, C.N.Jr. (2004). *Genetically Modified Planet. Environmental Impacts of Genetically Engineered Plants*. Oxford University Press, New York.
- Tang, S., Kishore, V.K., Knapp, S.J. (2003b). PCR-multiplexes for a genome-wide framework of simple sequence repeat marker loci in cultivated sunflower. *Theor. Appl. Genet.* **107**: 6-19.
- Tang, S., Knapp S.J. (2003). Microsatellites uncover extraordinary molecular genetic diversity in Native American land races and wild populations of cultivated sunflower. *Theor Appl Genet.* **106**:990-1003.
- Tang, S., Yu, J.K., Slabaugh, M.B., Shintani, D.K., Knapp, S.J. (2002). Simple sequence repeat map of the sunflower genome. *Theor Appl Genet* **105**:1124-1136.
- Uozu S, Ikehashi H, Ohmido N, Ohtsubo H, Ohtsubo E, Fukui K. (1997). Repetitive sequences: cause for variation in genome size and chromosome morphology in the genus *Oryza*. *Pl. Molec. Biol.* **35**: 791-799.

What other supplementary tables would be useful?

Supplementary Table 1. Economics of Major Compositae Crops.

Production Quantity | World

commodity	year					
	1990		2000		2005	
Artichokes	1,323,471.00	tonnes	1,330,299.00	tonnes	1,262,982.00	tonnes
Lettuce and chicory	11,520,342.00	tonnes	18,279,329.00	tonnes	22,399,319.00	tonnes
Safflower oil	215,593.22	tonnes	170,397.72	tonnes	126,581.90	tonnes
Safflower seed	834,539.00	tonnes	623,356.00	tonnes	717,778.00	tonnes
Sunflower oil	8,097,652.95	tonnes	9,975,013.13	tonnes	10,244,873.62	tonnes
Sunflower seed	22,707,244.00	tonnes	26,510,666.00	tonnes	30,595,462.00	tonnes

FAOSTAT | © FAO Statistics Division 2007 | 12 February 2007

Area Harvested | World

commodity	year					
	1990		2000		2005	
Artichokes	117,030.00	Ha	123,056.00	Ha	125,128.00	Ha
Lettuce and chicory	542,103.00	Ha	839,485.00	Ha	1,024,018.00	Ha
Safflower seed	1,207,019.00	Ha	822,270.00	Ha	916,443.00	Ha
Sunflower seed	17,046,747.00	Ha	21,161,673.00	Ha	22,823,330.00	Ha

FAOSTAT | © FAO Statistics Division 2007 | 12 February 2007

Supplementary Table 2. Sunflower seed production (in 1,000 metric tons)

Item	Sunflower seed					
	World supply & disappearance (in 1,000 metric tons)					
	2000/01	2001/02	2002/03	2003/04	2004/05	2005/06
Area harvested (1,000 HA)	19,540	18,485	19,892	22,918	21,262	22,791
yield (MT/HEC)	1.18	1.18	1.2	1.17	1.23	1.29
Seed Production						
Argentina	2,950	3,720	3,340	2,990	3,650	3,800
Eastern Europe	1,657	1,861	1,648	2,295	2,270	1,950
European Union	3,333	3,030	3,718	4,078	4,133	3,765
China (Peoples Republic of)	1,954	1,750	1,946	1,820	1,750	1,850
Russia/Ukraine	7,368	4,936	7,194	9,348	8,001	10,450
United States	1,608	1,551	1,112	1,209	930	1,824
India	730	870	1,060	1,160	1,300	1,250
Turkey	630	530	830	560	640	790
Other	2,880	3,551	3,108	3,467	3,505	3,665
TOTAL	23,110	21,799	23,956	26,927	26,179	29,344
Seed Import						

Mexico	23	10	104	38	11	23
European Union	1,999	1,155	1,007	1,473	763	1,000
Other	704	467	812	1,249	813	801
TOTAL	2,726	1,632	1,923	2,760	1,587	1,824
Oilseed crushed	21,116	18,514	21,149	23,442	23,115	25,510
Seed export						
Argentina	94	342	232	44	97	121
United States	153	176	122	136	116	225
Russia/Ukraine	1,768	100	524	1,271	73	560
Other	711	1,084	1,112	1,277	1,257	957
TOTAL	2,726	1,702	1,990	2,728	1,543	1,863

Source:

<http://www.sunflowernsa.com/stats/table.asp?contentID=109&htmlID=74&submit170=View&submit.x=57&submit.y=12>.

Supplementary Table 3. Oil composition of sunflower and comparison to other oilseeds

Source	Contents (%)		
	Oleic	Linoleic	Saturated
Conventional sunflower oil	20	69	11
Mid oleic sunflower oil	65	26	9
High oleic sunflower oil	82	9	9
Olive oil	77	8	14
Canola oil	62	21	16
Cottonseed oil	18	54	27

Source: US Sunflower Crop Quality Report 2003 <http://www.sunflowernsa.com/uploads/cqr/cqr2003.pdf>