Mapping Unexplored Genomes: A Genetic Linkage Map of the Woody Sonchus Alliance (Asteraceae: Sonchinae) in the Macaronesian Islands

SEUNG-CHUL KIM

From the Department of Botany and Plant Sciences, University of California, Riverside, CA 92521, USA (Kim).

Address correspondence to S.-C. Kim at the address above, or e-mail: sckim@ucr.edu.

Abstract

As an initial step to mapping quantitative trait loci for species differences and adaptive radiation of insular endemics in Macaronesia, a genetic linkage map was constructed from an intergeneric backcross between Lactuconsonchus webbii and Sonchus radicatus, core members of the tree lettuces in the Macaronesian Islands. A total of 152 amplified fragment length polymorphism markers were mapped into 10 major and 3 minor linkage groups for a total map length of 644 cM with an average distance of 4.53 cM for the 10 major groups. The genetic linkage map length is considerably less than the estimated, and this may reflect incomplete genomic coverage in the current study or reduced recombination, which is a common feature of maps for hybrids of divergent taxa. Segregation distortion occurred in 34% of the mapped markers, and they were located primarily in 4 linkage groups. Segregation distortion in the current BC1 intergeneric population is slightly lower than average (40%) for BC1 interspecific populations. This level of segregation distortion implies that unlike what we normally assume no to few reproductive barriers, oceanic island plant taxa do exhibit some degree of postmating reproductive isolation.

Adaptive radiation is often responsible for exceptionally rich biodiversity on oceanic islands. It refers to the diversification of a lineage into species that exploit a variety of different resource types, and that differ in the morphological and physiological traits used to exploit those resources (Schluter 1996). Adaptive radiation involves both the formation of species and the evolution of ecological differences between them. The process of adaptive radiation is hypothesized to start with long-distance dispersal of a continental ancestor to an open habitat on islands. Subsequently, relaxed selection in the new habitat allows novel recombinants to be established through morphological and physiological adaptation to the corresponding habitat (Carlquist 1974). This ultimately leads to speciation, and repeated colonization events of the offspring of the founding population to different habitats form several new species over a short period of time. The result of adaptive radiation is often characterized by extensive divergence in morphological traits and habitat, yet by little divergence in molecular sequences or crossing ability. There are many known examples of adaptive radiation in plants in several archipelagos: the silversword alliance (Carr et al. 1989, 1995, 1996) on Hawaii; Echium (Böhle et al. 1996), Argyranthemum (Francisco-Ortega et al. 1996, 1997), and the woody Sonchus alliance (Kim, Crawford, Francisco-Ortega, and Santos-Guerra 1996; Kim, Crawford, and Jansen 1996; Kim et al. 1999) of Macaronesia; Dendroseris (Crawford et al. 1987; Sang et al. 1994) of the Juan Fernandez Islands. Recently, numerous molecular phylogenetic studies of adaptive radiation in several archipelagos have been carried out to infer the history of organisms evolution and to understand the underlying evolutionary processes (see examples in Baldwin et al. 1998). However, at the present time, only a handful of studies (e.g., Okada et al. 1997; Whitkus 1998; Whitkus et al. 2000) have been directed toward understanding the genetic basis of species differences in relation to adaptive radiation in the Pacific islands.

The Macaronesian Islands in the Atlantic Ocean include 5 archipelagos: Azores (9 islands), Salvagens (2 islands), Canaries (7 islands), Madeiras (3 islands), and Cape Verdes (10 islands). They are unique in 2 aspects compared with the ones in the Pacific Ocean. First, islands, especially the Canary archipelago, are very close to the continent: one of the closest island of the Canaries, Fuerteventura, is only about 100 km from the west coast of Morocco. Second,
the Macaronesian Islands are geologically much older than those in the Pacific: Azores (4–8 Myr), Salavgens (10 Myr), Canaries (0.5–20.7 Myr), Madeiras (30 Myr), and Cape Verde (6–20 Myr). These 2 features may contribute to somewhat unusually high genetic diversity and unusual patterns of plant evolution compared with the endemics in the Pacific archipelagos (Francisco-Ortega et al. 2000).

The woody Sonchus alliance (Asteraceae: Sonchinae) includes 6 genera and ~31 species and is a premier example of adaptive radiation among angiosperms in the Macaronesian Islands (Aldridge 1975, 1979). The members of the alliance display extensive morphological, ecological, and anatomical diversity (Aldridge 1976). Several phylogenetic studies revealed the monophyly of the entire alliance despite its extensive morphological and ecological diversity and geographical proximity of the islands to a large continental source area (Kim, Crawford, Francisco-Ortega, and Santos-Guerra 1996; Kim, Crawford, and Jansen 1996; Lee et al. 2005). The low average nuclear and Chloroplast DNA sequence divergence and unresolved phylogenetic tree also suggest a rapid radiation of major lineages during the late Miocene or early Pliocene. Beyond little known phylogenetic relationships among the alliance, nothing is known about genetic basis of species differences and adaptive radiation of the alliance. In fact, the genetics of plant species differences and adaptive radiation have never been attempted or determined in Macaronesia despite its unique phytogeographic features. As an initial step to understand the genetics of adaptive radiation in the Macaronesian endemics, I report the first molecular marker-based genetic linkage map in an intergeneric cross in the woody Sonchus alliance.

**Materials and Methods**

**Study System and Mapping Population**

*Lactucononchus webbii* is a monotypic endemic to the island of La Palma in the Canaries. It is an herbaceous perennial with long tuberous roots and occurs in the north-facing under story of pine forest (*Pinus canariensis*). In contrast, *Sonchus radicatus*, endemic to Tenerife, occurs in north- and east-facing coastal ancient basalt cliffs and is a caudex perennial forming a distinctively thick woody stem with terminal rosette leaves after the first flowering season. *Lactucononchus* occurs in relatively mesic habitat at high elevation (between 400 and 600 m), whereas *S. radicatus* primarily occurs in lower elevation near coastal cliffs, including coastal desert.

Seeds for the 2 species of the woody *Sonchus* alliance were collected from natural populations in the Canary Islands, Spain, and subsequently germinated and grown in greenhouses at the University of California at Riverside. These 2 species show extensive morphological and ecological divergence; yet, due to their genetic similarity, they can be easily crossed, and F1 hybrids are vigorous and fertile. Both nuclear and chloroplast genomes are not too divergent between the 2 species (less than 3% in nuclear ribosomal DNA internal transcribed spacer (ITS) and 0.15% in cpDNA). In addition, an enzyme electrophoretic study found high genetic identity between these 2 species (Nei’s genetic identity value of >0.96), clustering them together in the dendrogram (Kim et al. 1999). The 2 species are diploids and have the same chromosome numbers (2n = 18, n = 9; Aldridge 1975).

A BC1 mapping strategy was taken for several reasons. First, the mapping will be accomplished with dominant markers, which can be more accurately genotyped in BC1 than F2 populations, in which the dominant phenotypes could be genotypically heterozygous or homozygous. Second, a BC1 population will reduce the complexity of predicted segregation patterns in progeny from crosses between heterozygous parental individuals. Finally, all members of the alliance have sprophytic self-incompatibility, so selfing is not possible in this system. The herbaceous *L. webbii* served as the maternal parent, and the woody perennial *S. radicatus* was the paternal parent. Several F1 hybrids were germinated and grown, and a single F1 hybrid chosen as the paternal parent was backcrossed to the recurrent maternal *L. webbii* parent, generating a BC1 mapping population. Almost 200 BC1 seeds were collected and a total of 123 individuals were successfully grown and used for genotyping and linkage analysis.

**DNA Extraction and Amplified Fragment Length Polymorphism Genotyping**

Total genomic DNA was isolated from 100 mg of fresh leaf tissue using DNeasy Plant Mini kits (Qiagen, Chatsworth, CA). This method is very efficient and produces highly purified DNA from 100 mg of fresh leaf tissue (Kim and Rieseberg 1999). All 123 BC1 progeny were genotyped using amplified fragment length polymorphisms (AFLPs). The AFLP procedure was the same as that described by Vos et al. (1995) with minor modifications (Kim and Rieseberg 1999; Kim et al. 2005). Genomic DNA (300 ng) was digested for 1 h at 37 °C with 3 units of EcoRI and MsiI, 4 mg bovine serum albumin (BSA), 4 μl 10× buffer 2 (New England Biolabs, Beverly, MA), and ddH2O to a final volume of 40 μl. Adapters were ligated to the digested fragments by adding 15 pmol EcoRI adapters, 150 pmol MsiI adapters, 0.5 μl T4 polynucleotide DNA ligase, 1 μl 10 nm ATP, 1 mg BSA, 2.0 μl 10× buffer 2, and ddH2O to a total volume of 50 μl and incubated for 3 h at 37 °C. Preamplification reactions were performed with 2 AFLP primers having a single selective nucleotide (A). The preamplification reactions contained 1 μl of template DNA from the ligation reaction, 187.5 ng EcoRI+I and 187.5 ng MsiI+I, 0.4 μl dNTPs (each at 25 nm), 2.5 μl polymerase chain reaction (PCR) buffer (500 mM KCl, 15 mM MgCl2, and 200 mM Tris–HCl), 0.4 units *Taq* polymerase, and ddH2O in a total volume of 25 μl. The reactions were placed in a thermal cycler (PTC-100, MJ Research, Watertown, MA) programmed for 20 cycles, each consisting of 30 s at 94 °C, 30 s at 60 °C, and 1 min at 72 °C.

Once the preamplifications were complete, selective amplifications were performed using 2.5 μl of 1:20 diluted preamplification reaction as a template, 5 ng of the EcoRI+3 nucleotide selective primer (Vos et al. 1995), 15 ng of MsiI+3 and +4 nucleotide selective primer, 0.16 μl dNTPs, 1 μl PCR buffer, *Taq* polymerase, and ddH2O to a final volume of
Amplifications were conducted in a MJ Research thermal cycler programmed for 36 cycles, each consisting of 30 s at 94 °C, 90 s at 65 °C (see below), and 90 s at 72 °C. The 65 °C annealing temperature of the first cycle was subsequently reduced by 1 °C for the next 10 cycles and then continued at 54 °C for the remaining 26 cycles. Likewise, the extension time of 90 s was reduced to 1 min for the last 26 cycles. All enzymes and buffers were purchased from New England Biolabs. Adapters and primer sequences are the same as those described in Vos et al. (1995). The EcoRI+3 primers were labeled with IRD 700 and IRD 800 fluorescence dyes (Li-Cor, Lincoln, NE). Following selective amplification, reaction products were mixed with an equal volume (10 μl) of formamide dye (98% formamide, 10 mM ethylenediaminetetraacetic acid pH 8.0, and bromphenol blue). The resulting mixtures were heated for 3 min at 90 °C, quickly cooled on ice, and 1.2 μl was immediately loaded on KBplus 6.5% Gel Matrix (Li-Cor) polyacrylamide gels. The size standard 50–700 bp (Li-Cor) was run with the samples to estimate the size of fragments. Electrophoresis was performed at constant voltage (1500 V) for 3 h at 43 °C using an automated DNA sequencer (Li-Cor IR).

**Linkage Analysis**

AFLP data were obtained by manually scoring the TIFF images generated by the Li-Cor IR sequencer. Only unambiguous bands specific to S. radicatus were scored (present = 1, absent = 0), and this was done twice by different people. Marker segregations were checked for deviation from the expected Mendelian segregation (1:1) by chi-square analysis. The map was constructed with MAPMAKER/Exp 3.0 (Lander et al. 1987). Markers were first divided into linkage groups using the “group” command (parameters logarithm of odds (LOD) >5.0; θ < 2.0). For linkage groups with 9 or fewer markers, an exhaustive analysis (i.e., the “compare” command) was used to order markers. For linkage groups with 10 or more markers, the “order” command was employed to obtain the order of markers with unique placement followed by the “try” command to find the most likely placement of the remaining markers, and subsequent orders were tested using the “ripple” command. Haldane’s (1919) mapping function was then used to transform the recombination frequency between linked loci into centimorgan distances.

**Genome Length**

Three approaches were used to roughly estimate genome length (L). I calculate the average marker spacing (j) by dividing the summed length of all linkage groups by the number of intervals. The first estimate was made by adding 2r to the length of each linkage group to account for chromosome ends beyond the terminal markers. The second method estimates the genome coverage. The first approach to estimating genome length is the sum of recombination fractions across the linkage map to the expectation given a single crossover per chromosome arm (Sybenga 1996; Whitkus 1998; Ott 1999; Fishman et al. 2001). Lastly, I used Method 4 of Chakravarti et al. (1991), which multiplies the length of each linkage group by the factor \( (m + 1)/(m - 1) \), where m is the number of markers on each linkage group.

**Results**

From a survey of over 100 AFLP primer combinations, 29 selective primer pairs were selected to genotype 123 BC1 progeny. Those 29 selective primer pairs showed banding patterns with very high reproducibility and clear banding resolution and a large number of polymorphic loci. A total of 161 S. radicatus specific markers (present in S. radicatus and F1 hybrids and absent in L. webbii) were scored. Of these 161 markers, 55 (34%) exhibit significant segregation distortion (P ≤ 0.05). Of the 55 distorted markers, 34 (62%) were significantly overrepresented (i.e., 34 S. radicatus markers are significantly overrepresented in L. webbii genetic background), whereas 21 markers (38%) were significantly underrepresented.

**Linkage analysis** of the 161 markers × 123 progeny data set revealed 10 major (4 or more markers) linkage groups and 3 minor (with 2 markers) ones (Figure 1). Nine markers could not be placed into any linkage group with group linkage criteria of LOD 5.0 and recombination frequency of 0.25 (25.5 cM). The number of markers per linkage group ranged from 4 (LG8) to 27 (LG1). Linkage groups 2 (LG2) and 10 (LG10) are composed entirely of markers significantly distorted toward S. radicatus, whereas 9 (LG9) contained only underrepresented S. radicatus alleles. Nearly all of linkage group 7 (LG7) are overrepresented toward S. radicatus. The map distance covered by the 10 major and 3 minor linkage groups is 644 cM, with an average distance between markers of 4.53 cM for 10 major linkage groups.

If I assume a random distribution of markers and simply add twice the average interval length (j = 9.06 cM) to each group to account for chromosome ends extending beyond the terminal markers, I estimate the map length to be 684.6 and 761.8 cM for 10 major linkage groups only and 10 major plus 3 minor groups, respectively. Thus, the genome covered in this map for 10 major groups only and 10 major and 3 minor ones is 86.7% and 84.5%, respectively. Based on the Sybenga et al. (1996), Whitkus (1998), and Ott (1999) method, which provides a minimum estimate of the genome-wide recombination fraction, the 18 chromosome arms in the woody Sonchus alliance give a total expected recombination fraction of 900 (18 crossovers). The total recombination fraction on the map of 573.6 is 64% of the expected 900. Finally, using Method 4 of Chakravarti et al. (1991), the estimated genome length to be 696.5 cM for 10 major groups. The observed total map length for 10 major groups is 85% of this value.

**Discussion**

This study represents the first genetic linkage study conducted in insular plant endemics, which underwent rapid radiation, in the Atlantic Ocean: the only other such study was Tetramolopium (Asteraceae), endemics to the Hawaiian and Cook Islands (Whitkus 1998; Whitkus et al. 2000). Although there are some experimentally attractive features of insular
plants, many groups that have undergone adaptive radiation are not tractable for genetic analysis because they are either long-lived perennials, of polyploid origin, do not represent a recent radiation, or are not interfertile. In contrast, the treelettces of the Macaronesian Islands (i.e., the woody \textit{Sonchus} alliance) provide a unique set of biogeographic and biological attributes and include diverse interfertile species ideal for determining the genetic basis of species differences and adaptive radiation. Strikingly, the \textit{Sonchus} alliance displays extensive morphological, ecological, and anatomical diversity; yet, this diversity is due to a single colonization event followed by recent adaptive radiation (Kim, Crawford, Francisco-Ortega, and Santos-Guerra 1996; Kim, Crawford, and Jansen 1996; Kim et al. 1999; Lee et al. 2005). Furthermore, although the majority of the alliance consists of long-lived perennials, some species are either herbaceous perennials or early flowering woody perennials, thus making it feasible to do a genetic analysis of the characters of interest.

Linkage analysis placed 152 AFLP markers into 10 major and 3 minor linkage groups with an average distance of 4.53 cM for 10 major groups. The woody \textit{Sonchus} alliance has a haploid chromosome number of 9, and it is likely that major linkage groups correspond to 9 chromosomes. The additional linkage groups of one major and 3 minor ones suggest that map coverage is incomplete with some of the groups representing separate segments of the same chromosome.

Figure 1. A genetic linkage map of the woody \textit{Sonchus} alliance. A total of 10 major and 3 minor linkage groups (in insert) are arranged on the basis of their length. Haldane map distances in centimorgans are listed to the left and loci to the right of each linkage group. The first 2 characters represent the \textit{Eco}RI primer and second 2 or 3 characters represent the \textit{Mse}I primer followed by estimated fragment sizes. Loci with square and broken lined square represent over- and underrepresented loci, respectively.
The map length estimates (64–86%) also suggest incomplete coverage of the genome, and additional markers are probably needed to cover the entire genome. It is hard to predict how many more markers would be needed to coalesce these linkage groups and unlinked loci into the 9 chromosomal linkage groups. Kesseli et al. (1994) showed that nearly doubling the number of markers did not reduce the number of linkage groups in lettuce. In order to coalesce the linkage map, bulk segregant analysis can be employed to efficiently increase the marker density in a specific region, especially near the ends of linkage groups (Michelmore et al. 1991; Reiter et al. 1992; Kesseli et al. 1994).

Segregation distortion appears to be correlated with increasing genetic divergence between parental lines (Zamir and Tadmor 1986; Patterson et al. 1991; Quillet et al. 1995; Grandillo and Tanksley 1996; Jenczewski et al. 1997; Bradshaw et al. 1998; Whitkus 1998). In general, F2 interspecific populations show more skewing (average 70%) than BC1 interspecific populations (average 40%) (Bernacchi and Tanksley 1997). The proportion of distorted markers in this current intergeneric BC1 population (34%) is slightly lower than average for interspecific BC1 populations. The distorted markers are concentrated in particular linkage groups (LG 2, 7, 9, and 10; Figure 1), and these groups are largely/entirely unidirectional in bias (mostly having an excess of S. radicatus alleles in LG 2, 7, and 10 and a deficit in LG 9). This is consistent with the common nature of unequal transmission of alleles at nuclear loci for wide interspecific and interspecific crosses (e.g., Zamir and Tadmor 1986; Jenczewski et al. 1997; Whitkus 1998; Fishman et al. 2001).

This pattern suggests that biological mechanisms, rather than chance or error, underlie most of the observed transmission distortion. Several possible reasons for transmission ratio distortion were postulated, including genetic factors operating in pre- and postzygotic phases of reproduction, structural rearrangements, or gametic selection (Whitkus 1998). No single genetic mechanism can account for the patterns of transmission ratio distortion observed in this study, and additional study may reveal underlying mechanisms.

Oceanic island plant species are generally highly cross compatible with few to no internal barriers to crossing if no obvious chromosomal structural differences exist (Carlquist 1966; Gillett and Lim 1970; Carr and Kyhos 1981; Borgen 1984; Ganders 1989). However, several studies suggested some degree of interspecific postdating reproductive isolation (e.g., Gillett 1966; Gillett and Lim 1970; Humphries 1975; Lowrey 1986; Carr and Baker 1988; Whitkus 1998). One exceptional example is Tetramolopium in Hawaii (Whitkus 1998). Whitkus (1998) found a very high level of segregation distortion (49%) in an interspecific BC1 cross and suggested that the 2 species have genomes that are as genetically divergent as many continental plant species, which was not obvious from molecular-marker diversity studies (Okada et al. 1997). The situation in the woody Sonchus alliance in Macaronesian Islands in the Atlantic Ocean is similar to the case of Tetramolopium. 34% of the makers show segregation distortion in this intergeneric cross. This implies that 34% of the genome is potentially influenced by reproductive barriers, that is, postzygotic reproductive isolation. This is somewhat unexpected based on high crossability among the woody Sonchus alliance species and low levels of molecular divergence between the 2 species (Kim, Crawford, Francisco-Ortega, and Santos-Guerra 1996; Kim et al. 1999; Lee et al. 2005). Although we assume few to no reproductive barriers in island taxa due to their recent diversification, this study suggests that oceanic island plant taxa do exhibit some degree of postmating reproductive isolation. Premating mechanisms related to adaptation to new habitats and genetic drift play predominant roles in the origin and isolation of island plant species, but the accumulation of postmating factors responsible for reproductive isolation will still occur (Whitkus 1998).

Linkage maps with moderate marker density have applications in several areas of evolutionary biology. One such application is the examination of quantitative trait loci (QTL) for quantitative traits of interest. The presence of QTLs and their individual contribution (small vs. large effect) to genetic variation of quantitative traits can be determined by various methods (e.g., multiple interval mapping, composite interval mapping, and single marker methods; see review Mauricio 2001). Although numerous QTL studies were conducted using cultivated/domesticated and natural populations, only one study system of insular endemic plant in the Pacific was utilized to understand the genetic basis of morphological differences, adaptation, and reproductive isolation. Whitkus (1998) found a high degree of distorted segregation, suggesting the buildup of internal crossing barriers in Hawaiian Tetramolopium. Whitkus et al. (2000) also found that the breeding system and several related reproductive traits in Hawaiian species of Tetramolopium are controlled by few genetic loci, with no apparent evidence of epistasis. These findings, that is, small number of genes with large effects on most phenotypic and reproductive traits, generally agree with classical crossing and QTL studies (e.g., Grant 1975; Templeton 1981; Turner 1981; Gottlieb 1984; Doebley and Stec 1991, 1993; Tanksley 1993; Bradshaw et al. 1995, 1998) and theoretical work of Orr (1998). It remains unknown if, in general, few genes of large effects are responsible for species differences and adaptation in other divergent island taxa, such as the woody Sonchus alliance in the Canary Islands. The genetic linkage map I present here will be an invaluable guide for QTL analyses of the phenotypic differences associated with adaptive radiation and speciation of the woody Sonchus alliance in the Macaronesian Islands in the Atlantic Ocean.

**Funding**

University of California-Riverside Agricultural Experimental Station funds, Regent’s Faculty Fellowship, and Academic Senate Research funds.

**Acknowledgments**

I thank Vincent Weng for maintaining plants in the greenhouses. Special thanks are due to Mike Roose for careful reading and comments on an earlier version of the manuscript. Comments and/or suggestions from 2 anonymous reviewers improved the earlier version of the manuscript as well.
References


Haldane JBS. 1919. The combination of linkage values and the calculation of distances between the loci or linked factors. J Genet. 8:299–309.


