Ultra-high Density Genetic Map of Sunflower

Christopher J. Grassa[1], Nolan C. Kane[1], John Bowers[2], Steve Knapp[3], John M. Burke[2], Loren H. Rieseberg[1,4]

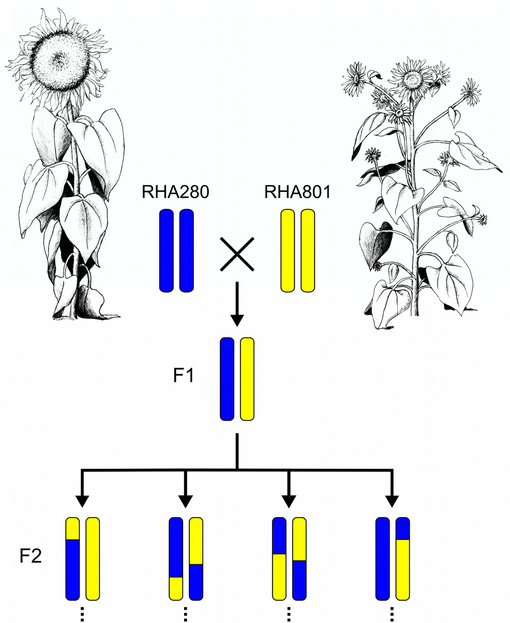
[1]Botany Department, University of British Columbia;[2]Department of Plant Biology, University of Georgia;[3]Monsanto, Co.;[4]Department of Biology, Indiana University

Ultra-high density genetic maps based on whole genome shotgun sequencing provide a useful scaffold for the assembly of large and highly repetitive plant genomes. We have constructed an ultra-high density genetic map for sunflower, which is the target on an ongoing sequencing effort. We sequenced *Helianthus annuus* cultivars RHA801 and RHA280, which are the parental lines of our core recombinant inbred line (RIL) mapping population, to an average depth of 10x. Alignment to the contiguous sequences of the draft sunflower genome assembly revealed over two million high quality SNPs fixed between the two parental cultivar lines. Ninety-six eighth generation RILs were sequenced to a mean depth of 1.0x. In the RILs, genomic regions were called as descended from one or the other parent based on the presence of at least ten SNP calls at informative sites. Perfectly correlated regions not exhibiting significant segregation distortion were binned for use as genetic markers. A genetic map of seventeen linkage groups corresponding to the seventeen H. annuus chromosomes was constructed de novo using a Minimum Spanning Tree algorithm as implemented in the MSTmap software. This map allowed us to confirm the composition of genomic scaffolds, assign them to linkage groups, and place them in linear order.

Interior radiuses show the segregation pattern in 93 RIL individuals. Black regions indicate chromosomal blocks descended from RHA280 and white regions from RHA801. Color transitions indicate putative locations of chromosomal crossover.

The genetic map is drawn along the outer black and white chromosomes.

The ray length is proportional to the sum of base pairs placed in 1 cM bins. Each linkage group was divided into 1 cM bins. For each bin, the lengths of all the contigs with an estimated placement within the bin were summed. The lengths of the bars making up the rays are proportional to those sums. They are scaled as bp^(1/2).



Our core mapping population was derived from two highly homozygous *H. annuus* cultivars: RHA280, a confectionary line with single apical florescence, and RHA801, a branching oil-seed line. Eighth generation RILs of single seed descent were used for mapping.

Whole genome shotgun sequencing was carried out with 100 base pair paired-end Illumina reads. One lane of Illumina sequence was generated for each parent. Eight lanes were each multiplexed with twelve RILs, producing about 1x of coverage for each barcoded sample.

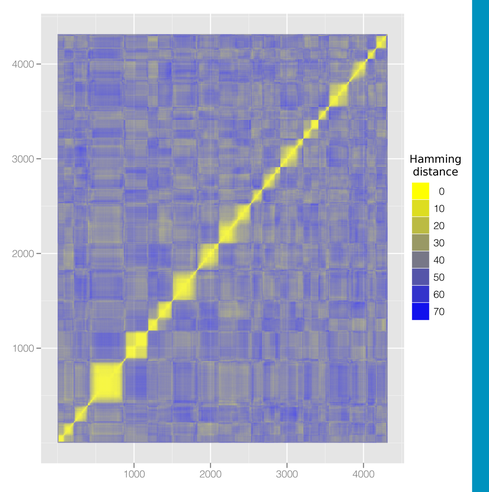
Parental reads were aligned to our draft reference assembly using the Burrows-Wheeler Aligner. Genotypes were called using SAMtools mpileup. Fixed SNPs with a genotype quality more than 20 and a mapping quality more than 30 on the Phred scale were used as candidate sites for calling genotype blocks in the RILs. The RIL reads were aligned to our draft reference assembly using BWA. SAMtools was used to convert the alignments to the pileup format. Each candidate site was 'softly called' as inherited from either or both parents based on the presence of no more than three aligned reads. In each individual, genomic contigs were called as descended from one or the other parent based on the presence of least 9 soft calls at candidate sites and at least 90% called as descended from one of the two parents. Contigs meeting these requirements in at least 75% of the RILs and with a minor allele frequency greater than 30% were used as map markers.

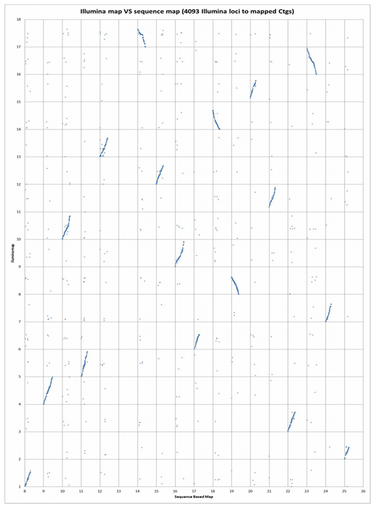
Contigs containing segregating SNPs were compared to the template map in forward and reverse order and the best match was stored for each direction. A contig was placed with an upper distance of the best forward match and a lower distance of the best reverse match if both were found on the same linkage group. 243,048 contigs were placed within 5 cM and 261,999 contigs were placed within 9 recombinations.

The probe sequences from a previous Illumina map were matched by BLAST to the contigs in the sunflower assembly. The centimorgan positions of the positions on the two maps were compared (Fig. 5). The two maps agreed very well in terms of synteny and ordering even though they were completely independently constructed. The micro and macro orders seem very well conserved, considering the Illumina map was in part constructed on a different mapping population. The chromosomes from the sequence based maps can now be named and oriented relative to the previous literature.

Single copy Illumina mapped probes Blasted to sequence contigs and plotted by map position on Illumina map vs. map position on sequence map. 4093 Contigs mapped. 6984 single copy illumina probes mapped to 6298 contigs. Of the 6298 contigs 4093 had been mapped based on sequence position. 90.4% of the hits are in 17 syntenic blocks. The roughly 10% non syntenic hits can be explained by blast picking the second best hit if the true homologue is fragmented into several contigs, or if the sequence is multi copy.

MSTmap was used to place the markers in linear order. MSTmap groups markers based on the minimum sum of recombination events (Hamming distance) between their segregation patterns and divides them into linkage groups if the sum is significantly different than observed across all markers. Markers on each linkage group are then ordered using a recursive minimum spanning tree algorithm. Kosambi's mapping function was used to calculate the map distance between adjacent pairs of markers ordered by MSTmap.





GenomeCanada
NSF
INRA
USDA
The University of Georgia