DISCOVERY OF THE CAUSATIVE MUTATION OF THE LATERAL BEARING PHENOTYPE IN WALNUT

Jan Dvorak, Malli Aradhya, Charles Leslie, and Ming-Cheng Luo

ABSTRACT

The first phase of map-based isolation of the *LB1*gene was successfully completed. A total of 750 F_2 'Chandler' progeny with recombined chromosomes in the 8.1cM JM079D23-JH036K04 region were identified by genotyping 5,060 nuts with the two markers. The two markers widely bracketed the location of the *LB1* gene on the Jr11 chromosome. A total of 520 of these trees have been transplanted into the field for the ultimate evaluation of their nut bearing (lateral vs terminal) phenotype. The rest will be transplanted early in 2016. The bacterial artificial chromosome (BAC) clone contig ctg199, which we anticipate harbors the *LB1* gene, has been sequenced and sequence has been assembled. Homology search against the walnut gene database indicated the presence of at least 514 genes in the BAC contig.

OBJECTIVES

Goal. The goal of the project is to develop an assay for the causative mutation of the lateral bearing trait and deploy it in the California walnut breeding program. The prerequisite for achieving this goal is the isolation of the *LB1* gene, which will facilitate the comparison of the sequences of the *LB1* (lateral bearing) and *lb1* (terminal bearing) alleles. We have previously developed genetic and physical maps for Chandler, and, therefore, the best strategy available to us for the isolation of the *LB1* gene is the map-based gene isolation approach. Our previous work showed that Chandler is heterozygous for the lateral bearing gene, and we located the gene in the centromeric region of chromosome Jr11. Chandler is a clonally propagated hybrid and is therefore an F_1 generation. Nuts from random pollination among Chandler trees in an isolated orchard will produce F_2 progeny. During gamete development, the two homologous chromosomes bearing the *LB1* and *lb1* alleles will recombine with each other, generating a new gene combinations in the *LB1* region in F_2 trees. We will use the locations of crossover points relative to markers surrounding the gene and the *LB1/lb1* phenotype to infer the location of the *LB1* candidate gene in the nucleotide sequence of the chromosome.

To achieve the overall goal of the project and deploy this gene isolation approach, we will accomplish four objectives stretched over a four-year period.

Objective 1. Select F_2 Chandler progeny harboring recombined chromosomes in the *LB1* region. (year 1)

Objective 2. Identify all genes in the *LB1* region so that they can be used as markers in high resolution mapping of *LB1*. (year 3)

Objective 3. Identify the candidate gene for *LB1*. (year 3 and 4)

Objective 4. Develop an SNP-based assay for *LB1* and implement it in the California walnut breeding program. (year 4)

SIGNIFICANT FINDINGS

The most important outcome of this phase of the project is the development of a large mapping population of trees harboring chromosomes recombined in the *LB1*region. The recombination level in the *LB1* region in the F_2 population was very close to what we expected on the basis of the Chandler genetic map.

PROCEDURES

Objective 1 Plan: About 10,000 'Chandler' nuts will be acquired from a grower with a spatially isolated 'Chandler' orchard and stored in a cold room. Nuts will be germinated in batches of 200 to 500, DNA will be isolated from the roots, and genotyped with markers flanking the *LB1* region.

Actual: Chandler nuts were provided by Castle Farms in Merced County from an isolated Chandler-only orchard. Because of poor germination of the nuts we processed a total of 23,000 nuts. We developed single nucleotide polymorphism (SNP) assays for markers JH066M23, JM001H03, JM079D23, and JH036K04, which span greater distance on the chromosome than originally planned to make sure that the gene is between the markers. Otherwise, the planned method was followed.

Objective 2 Plan: The *LB1* region spans 10 Mb (million base pairs) and includes bacterial artificial chromosome (BAC) contigs (contiguous sequences of overlapping BAC clones) ctg31, ctg199, ctg74, and ctg279. A nanomap (optical map of molecular markers along a DNA molecule) for the 'Chandler' genome and the four contigs will be constructed. Nanomaps of individual BAC clones will be compared with the entire genome nanomap, and each BAC will be assigned to either *LB1* or *lb1* haplotype. Minimal tiling path across the two haplotypes will be constructed from the fingerprinted BAC clones. The clones will be sequenced to the depth of 100X walnut genome equivalent using the MiSeq next generation sequencing platforms and assembled into sequence scaffolds. The sequence will be assembled in collaboration with the Walnut Genome Sequencing Project and scaffolds will be validated by comparison with the nanomap as we have done previously. Genes will be annotated with the annotation pipeline MAKER (http:// gmod.org/wiki/MAKER) and verified manually.

Actual: The plan was followed with the following modifications. The MiSeq reads were assembled into contigs with the SOAPde novo software and annotated genes were identified on the basis of sequence homology between the sequence of BAC contig ctg199 and the walnut whole genome shotgun (WGS) assembly (Pedro Martinez-Garcia *et al*., unpublished), rather than being annotated *de novo* by the MAKER software.

RESULTS AND DISCUSSION

Objective 1. The probability of obtaining a recombined chromosome in a given genomic region is inversely proportional to the region genetic length. Since the *LB1* locus is in the low recombination region of chromosome Jr11 we considered it prudent to genotype as many Chandler nuts as possible. We therefore more than doubled the number of nuts processed, from 10,000 to 23,000 (**Table 1**). Of these we genotyped 5,060 saplings using four SNPs at gene loci JH066M23, JM001H03, JM079D23, and JH036K04. These genes were discovered in the end sequences (Wu, et al., 2012; You, et al., 2012) of BAC clones with the same names and were mapped in the *LB1* region on chromosome Jr11 (Luo et al., 2015). After several genotyping runs we realized that the former two markers were unreliable because of the presence of paralogous genes in the walnut genome and were dropped. All 5,060 saplings were genotyped with the latter two markers. The distance between JM079D23 (located at 17.723 cM on the Jr11 genetic map) and JH036K04 (located at 26.563 cM on the Jr11 genetic map) was 8.84 cM on the Chandler genetic map. This interval was 8.106 in the Chandler F_2 population, which closely agrees with the amount of recombination in the region predicted by the genetic map. We selected 750 saplings with a crossover in the JM079D23-JH036K04 region, grew them in the greenhouse, and transplanted 520 of them into the field in September, 2015. The remaining 180 trees will be transplanted early in 2016. Assuming that the probability of a crossover is even across the JM079D23-JH036K04 region and that only a single chromosome was recombined in each of the 750 trees, we should anticipate the average distance of 0.01 % recombination between crossover breakpoints in the interval in our population, which provides a sufficient resolution to identify the *LB1* candidate gene.

Table 1. Selection of Chandler F_2 progeny with recombined chromosomes in the JM079D23-JH036K04 region

Nuts processed	Nuts sowed	Nuts germinated	Genotyped	Recombined in the region	Transplanted into the field	Remaining in the greenhouse
23,000	13,700	6,500	5,060	750	520	180

Objective 2. The previous quantitative trait locus (QTL) mapping of *LB1* placed it in BAC contig ctg199 (Luo, et al., 2015). We sequenced with the MiSeq next generation sequencing platform a minimum tiling path (MTP) cross the BAC contig. MTP consisted of 30 BAC clones. The reads were assembled into sequence contigs using SOAPdenovo software. The assembled region was 1.9 Mb long. The assembled sequences were archived at the FTP site ftp://ftp.ccb.jhu.edu/pub/dpuiu/Walnut/English_walnut/BAC1/. The assembly contains 128 contigs longer than 1kb with an N50 of 47,370bp (ftp://ftp.ccb.jhu.edu/pub/dpuiu/Walnut/English_walnut/BAC1/README). The primary goal of this objective was to find and order all of the genes in ctg199. This was accomplished by homology search against the walnut WGS sequence. There were 514 complete gene models, 249 partial gene models, 204 low-quality gene models, and 295 ESTs in the contig sequence.

REFERENCES

Luo MC, You FM, Li PC, Wang JR, Zhu TT, Dandekar AM, Leslie CA, Aradhya M, McGuire PE, Dvorak J. 2015. Synteny analysis in Rosids with a walnut physical map reveals slow genome evolution in long-lived woody perennials, BMC Genomics, 16.

Wu JJ, Gu YQ, Hu YQ, You FM, Dandekar AM, Leslie CA, Aradhya M, Dvorak J, Luo MC. 2012. Characterizing the walnut genome through analyses of BAC end sequences, Plant Molecular Biology, 78:95-107.

You FM, Deal KR, Wang JR, Britton MT, Fass JN, Lin DW, Dandekar AM, Leslie CA, Aradhya M, Luo MC, Dvorak J. 2012. Genome-wide SNP discovery in walnut with an AGSNP pipeline updated for SNP discovery in allogamous organisms, BMC Genomics, 13:354.