Microsatellite Development, Characterization and Mapping in Hazelnut

Gehendra Bhattarai and **Shawn A. Mehlenbacher**

Department of Horticulture, Oregon State University, 4017 ALS Building, Corvallis, Oregon 97331 (USA)

ABSTRACT

The genome sequences of 'Jefferson' and seven other hazelnut (*Corylus avellana* L.) cultivars allowed efficient *in silico* comparisons and identification of polymorphic SSRs. The 'Jefferson' genome was searched using MISA and 17,588 SSRs >15bp were identified. Removal of duplicates, short fragments, repeats at ends, and repeats containing only A's and T's reduced the number of unique fragments to 2,069. The 'Jefferson' sequences were aligned with reads from the seven other cultivars using MAQ and the results visualized in Tablet. 489 SSRs showed variation in repeat number and primers were designed for them. PCR amplification and separation on agarose gels confirmed that 368 were polymorphic. Fluorescent forward primers were ordered and were used to amplify 48 diverse accessions plus the parents of the mapping population. After post-PCR multiplexing, samples were submitted for sizing using capillary electrophoresis. Allele sizes were determined using GeneMapper v4.1 software, entered in a spreadsheet and analyzed using PowerMarker and Cervus software. 366 new polymorphic microsatellite loci were developed in hazelnut. A total of 1,880 alleles were present at the 366 polymorphic loci and 50 genotypes. The number of alleles per locus ranged from 2 to 17 with an average of 5.14. The mean observed heterozygosity, expected heterozygosity, polymorphism information content, and the frequency of null alleles were 0.52, 0.53, 0.483, and 0.026 respectively. SSR markers segregating in the mapping population of Mehlenbacher et al., 2006 (OSU252.146 x OSU 414.062) were mapped using JoinMap 4.1 and the two-way pseudo-testcross mapping approach.

INTRODUCTION

European hazelnut (*Corylus avellana* L.), one of the major nut crop, is widely grown in the Willamette Valley of Oregon. Microsatellites or Simple Sequence Repeats (SSRs) are DNA segments containing tandem repeat motifs of 1-6 nucleotides in length. Microsatellites are the markers of choice as they are PCR based, co-dominant, highly reproducible, multi-allelic, and abundant. Genetic diversity and evolutionary studies, genetic mapping, fingerprinting, MAS, and QTL analysis are some applications of microsatellites. Also, SSR markers serve as anchor loci and are useful for aligning the linkage map, physical map and genome sequences. Around 350 polymorphic SSRs were previously developed from enriched libraries, ISSR fragments, BAC and transcriptome sequences. The genome of 'Jefferson' hazelnut was sequenced using Illumina technology, resulting in 46.1 Gb of sequence data equivalent to 115X genome coverage. Genomic sequences were assembled to form 36,642 contigs with N50 = 21,548 bp using the programs Velvet and MIRA. The genomes of seven additional accessions were also sequenced using Illumina. We developed 366 new polymorphic SSRs from these sequences.

MATERIALS AND METHODS

Paired-end Illumina sequences of 7 cultivars aligned with 'Jefferson' sequences as the references

↓

Visualized with Tablet software

↓

Primers designed for 489 SSRs polymorphic in Tablet

↓

PCR amplification of 24 diverse accessions and gel electrophoresis verified polymorphism of 368 SSRs

↓

PCR amplification of 48 accessions and parents of mapping population with fluorescent forward primers

↓

Multiplexing, capillary electrophoresis, and allele sizing with GeneMapper v4.1 software

↓

Characterization of 366 new polymorphic marker loci (Ho, He, PIC, r): PowerMarker and Cervus Software

↓

Loci segregating in mapping population (MP) are mapped using 138 seedlings

↓

15 parents of alternate mapping populations are fingerprinted at 366 new SSR loci

↓

Using alternate MP, linkage groups will be assigned for SSRs not segregating in the primary MP

REFERENCES

Mehlenbacher, S.A., Brown, R.N., Nouhra, E.R., Gökirmak, T., Bassil, N.V., and Kubisiak, T.L. 2006. A genetic linkage map for hazelnut (*Corylus avellana* L.) based on RAPD and SSR markers. Genome 49: 122-133

Peterschmidt, B.C. 2013. DNA markers and characterization of novel sources of eastern filbert blight resistance in European hazelnut (*Corylus avellana* L.). MS Thesis, Oregon State Univ., Corvallis.

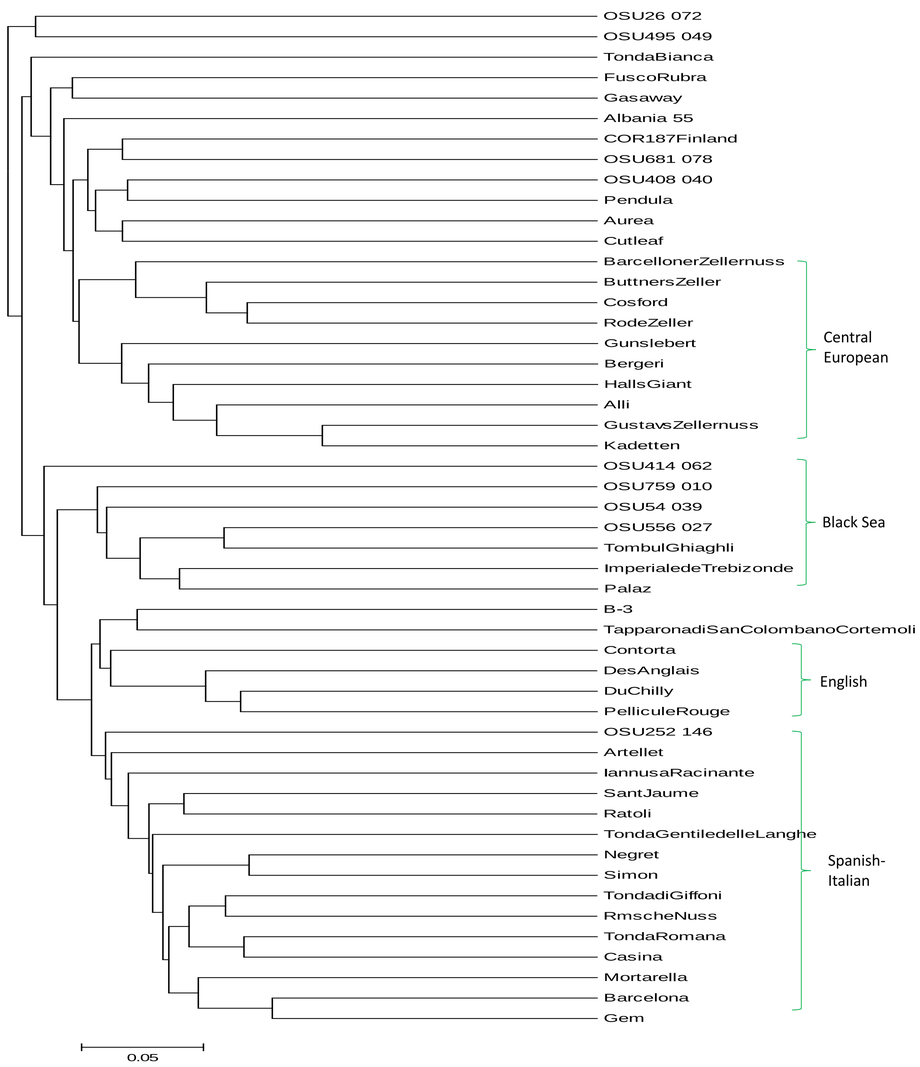

Fig. 1: UPGMA dendrogram of 50 hazelnut accessions

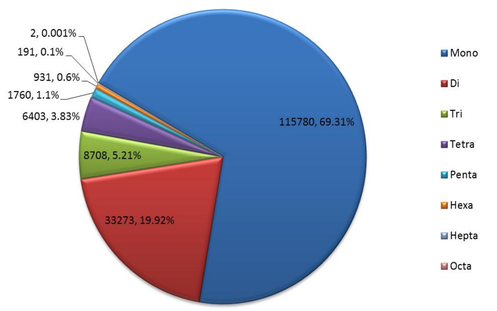

Fig 2: SSR loci distribution in 'Jefferson' genome

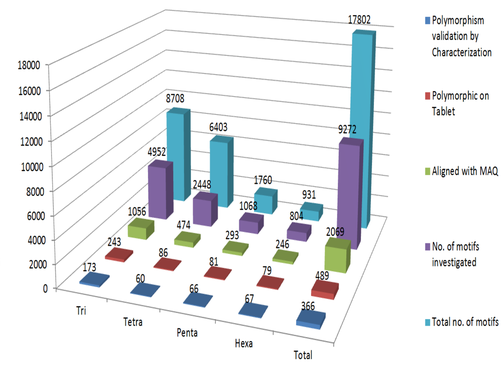

Fig 3: *In silico* SSR development approach in hazelnut

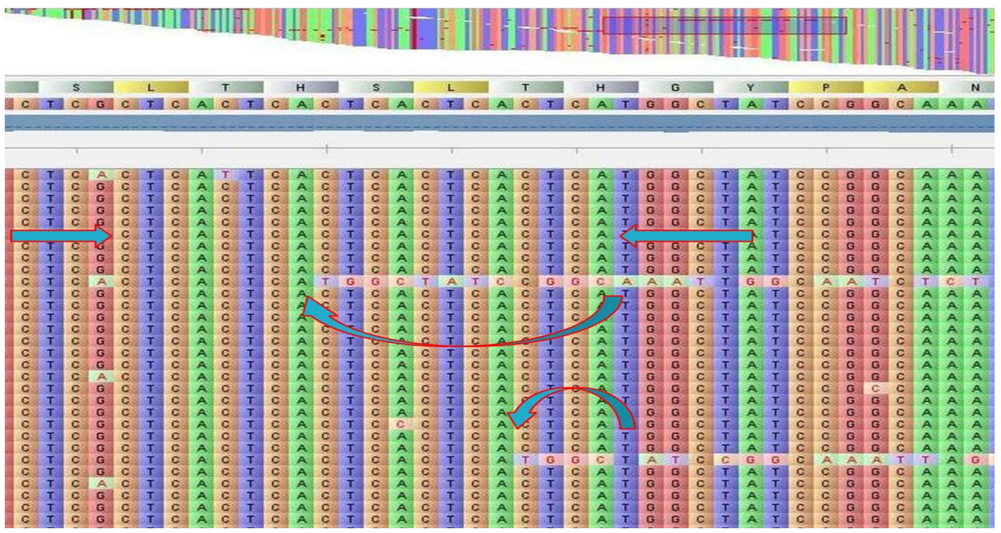

Fig. 4: Visualization in Tablet of Illumina reads aligned with the 'Jefferson' reference using MAQ. Inspection allows *in silico* detection of polymorphism, expressed as variation in the number of tetra-nucleotide $(CTCA)_5$ repeats among 8 cultivars but conserved flanking regions.

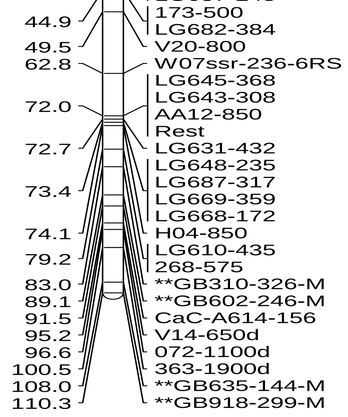

Fig. 5: Map of LG 6 in reference mapping population (OSU 252.146 x OSU 414.062). The map for the EFB resistant male parent was created using BC1 function of JoinMap v 4.1. New SSR markers are shown with asterisks.

RESULTS AND DISCUSSION

Next generation sequencing technology allows efficient SSR marker development at low cost. A total of 167,048 microsatellites (with repeat motifs of 1-8 bp) were identified from 333,492 sequences of 'Jefferson' hazelnut. The average number of repeat units as wells as the number of loci decreased as repeat motif length increased. 2,069 unique sequences were pursued after removal of repeats containing only A's and T's, repeats at the fragment ends, and short fragments (<400 bp). Visual inspection of these unique fragments in Tablet software for variation in the number of repeats among cultivars but with conserved flanking regions identified 489 sequences for further evaluation. The *in silico* visualization allowed rejection of 1,213 monomorphic sequences and reads that failed to align well with the reference sequences. 366 new polymorphic microsatellite markers were developed, validated by genotyping using fluorescent forward primers, and used to fingerprint 48 diverse accessions and 15 parents. These 366 markers were characterized using 48 diverse accessions plus the parents of the mapping population. A total of 1880 alleles were present. The number of alleles per locus ranged from 2 to 17 with an average of 5.14. The mean observed heterozygosity, expected heterozygosity, polymorphism information content, and the frequency of null alleles were 0.52, 0.53, 0.483, and 0.026 respectively. 185 of the 255 loci segregating in the MP were mapped. A dendrogram created using UPGMA and a frequency based distance matrix shows accessions clustered according to their geographic origin. Markers segregating in other mapping populations will be assigned to LG.