Microsatellite Markers in Hazelnut: Isolation, Characterization, and Cross-species Amplification

Nahla V. Bassil
USDA–ARS, NCGR, 33447 Peoria Road, Corvallis, OR 97333

R. Botta
Università degli Studi di Torino, Dipartimento di Colture Arboree, Via Leonardo da Vinci 44, 10095 Grugliasco (TO), Italy

S.A. Mehlenbacher
Oregon State University, Department of Horticulture, 4017 Agricultural and Life Sciences Building, Corvallis, OR 97331

ABSTRACT. Three microsatellite-enriched libraries of the European hazelnut (Corylus avellana L.) were constructed: library A for CA repeats, library B for GA repeats, and library C for GAA repeats. Twenty-five primer pairs amplified easy-to-score single loci and were used to investigate polymorphism among 20 C. avellana genotypes and to evaluate cross-species amplification in seven Corylus L. species. Microsatellite alleles were estimated by fluorescent capillary electrophoresis fragment sizing. The number of alleles per locus ranged from 2 to 12 (average = 7.16) in C. avellana and from 5 to 22 overall (average = 13.32). With the exception of CAC-B110, di-nucleotide SSRs were characterized by a relatively large number of alleles per locus (≥ 5), high average observed and expected heterozygosity (H_e and H_o > 0.6), and a high mean polymorphic information content (PIC ≥ 0.6) in C. avellana. In contrast, tri-nucleotide microsatellites were more homozygous (H_e = 0.4 on average) and less informative than di-nucleotide simple sequence repeats (SSRs) as indicated by a lower mean number of alleles per locus (4.5), H_e (0.59), and PIC (0.54). Cross-species amplification in Corylus was demonstrated. These microsatellite markers were highly heterozygous and polymorphic and differentiated among genotypes of C. avellana irrespective of geographical origin. They will aid in fingerprinting genotypes of the European hazelnut and other Corylus species, genome mapping, and genetic diversity assessments.

Received for publication 22 Nov. 2004. Accepted for publication 29 Jan. 2005.

We acknowledge Isabela Mackey for her dedication in performing PCR, fragment separation and tabulating the data. A special thank you is extended to Kim Hummer for editorial suggestions and to all the technicians who keep the hazelnut trees healthy including Dave Smith, Joe Snead, and Ray Gekosky. We thank the Oregon Hazelnut Commission and the USDA–ARS CRIS 5356-150-033-00D for their financial support.

To whom reprint requests should be addressed. E-mail address: cornb@ars-grin.gov
to characterize 20 genotypes of *C. avellana*, and 3) evaluate their cross-species amplification.

**Materials and Methods**

**Plant material and DNA preparation.** For development of the microsatellite-enriched libraries, dark-germinated seeds of a mixture of *Corylus avellana* genotypes were used. For SSR analysis, DNA was extracted from young leaf tissue (in the spring, April–June) of 20 genotypes of *C. avellana*, two genotypes of *C. colurna*, and a single representative of six other *Corylus* species (Table 1). The Puregene kit (Genta Systems, Minneapolis) was used according to the manufacturer’s instructions.

*Corylus avellana* cultivars were chosen to represent the various countries that grow hazelnut, including Turkey, Italy, Spain, and the United States (Table 1). OSU 252.146 and OSU 414.062 are the parents of a hazelnut mapping population segregating for resistance to eastern filbert blight (Mehlenbacher et al., 2003). Microsatellite-enriched libraries and primer design. Genetic Identification Services (GIS, Chatsworth, Calif.) constructed hazelnut genomic libraries enriched for di-nucleotide repeats CA (library A) and GA (library B) and for tri-nucleotide repeat GAA (library C). Inserts were sequenced by GIS or Central Services Laboratory (CSL) of the Center for Gene Research and Biotechnology at Oregon State Univ. (OSU) using an ABI 377 (Applied Biosystems, Foster City, Calif.). Sequences were compared using ClustalW and identical sequences were eliminated. Primers were designed by GIS or at OSU using Primer 3 software (Rozen and Skalatsky, 2000). The primer design parameters included an optimum annealing temperature of 60 °C, a GC content of 50%, and an amplicon size of 100–350 base pairs (bp). Primers were purchased from Operon Technologies (Qiagen, Valencia, Calif.). Primer pairs are described (Table 2) along with their motifs and the annealing temperatures (Ta) used during PCR amplification.

**PCR conditions.** PCR was performed in a total reaction volume of 10 μL containing: 1x PCR buffer, 2 mM MgCl₂, 0.2 mM of each dNTP, 0.6 μM of each primer, 0.05 U of Biolase enzyme (Bioline USA, Randolph, Mass.), and 2.5 ng of DNA template. For each primer pair, a gradient PCR ranging from 45 to 65 °C was carried out to determine the optimum annealing temperature using an Eppendorf Gradient thermocycler (Brinkmann Instruments, Westbury, N.Y.) or an MJ Research Tetrad thermocycler (MJ Research, Watertown, Mass.). The PCR conditions consisted of an initial denaturation cycle of 94 °C for 3 min, 35 cycles of denaturation at 94 °C for 40 s, annealing temperature for 40 s, and extension at 72 °C for 30 s. A final extension at 72 °C for 30 min was used to maximize nontemplated dA addition to the 5′ ends.

Loci were given an acronym to indicate the species from which they were isolated (*C. avellana*), followed by C for Corvallis, and A, B, or C to indicate the library from which the sequence was isolated. When two markers were isolated from the same sequences, a final letter (a or b) was added.

**SSR marker genotyping.** SSR primers were initially screened for length polymorphism and for cross-species amplification on 3% agarose gels using a random set of 11 hazelnut genotypes (six *C. avellana* cultivars and a single representative of each of five species of *Corylus*). If polymorphism was observed in this set of 11 genotypes, forward primers were fluorescently labeled with 6FAM or HEX purchased from Operon Technologies (Qiagen) and NED purchased from Applied Biosystems. Fluorophores were used to label primers that amplified fragments of different lengths to facilitate post-PCR multiplexing. At the CSL, the ABI 3100 (Applied Biosystems), a capillary electrophoresis instrument, was used to separate 1 μL of a mix of three or four PCR products diluted 80 to 320 times. GeneScan and Genotyper softwares were used for fragment size determination.

**Data analysis.** After determining the allelic profile at each SSR locus, the program PowerMarker (Kejun and Muse, 2003) was used to calculate the number of alleles per locus, observed heterozygosity (Hₓ), gene diversity (synonymous with expected heterozygosity, Hₑ), and the polymorphic information content (PIC). The observed heterozygosity is calculated as the number of heterozygous genotypes at a given locus divided by the number of genotypes present at that locus. Gene diversity is defined as the probability that two randomly chosen alleles from the population are different. PIC is an estimate of the probability that the parental origin of an allele can be determined from the marker locus.

### Table 1. List of 20 *Corylus avellana* genotypes, eight genotypes of *Corylus* species, their accession numbers, and geographical origins.

| Accession | Name       | Species         | Origin          |
|-----------|------------|-----------------|-----------------|
| None      | OSU 252.146| *C. avellana*    | Oregon          |
| None      | OSU 414.062| *C. avellana*    | Oregon          |
| PI 557035 | Tonda Gentile delle Langhe | *C. avellana* | Italy |
| PI 557233 | Römische Nuss | *C. avellana* | Italy |
| PI 271105 | Imperial de Trebizonde | *C. avellana* | Turkey |
| PI 557039 | Cosford | *C. avellana* | United Kingdom |
| PI 557057 | Giresun 54.021 | *C. avellana* | Turkey |
| PI 557125 | B-4 Pelargonia | *C. avellana* | Macedonia |
| PI 557117 | San Giovanni | *C. avellana* | Italy |
| PI 557219 | Bulgaria XI-8 | *C. avellana* | Bulgaria |
| PI 634202 | Ganja | *C. avellana* | Georgia |
| PI 557025 | Tonda Gentile Romana | *C. avellana* | Italy |
| PI 557042 | Gasaway | *C. avellana* | United States |
| PI 557027 | Half's Giant | *C. avellana* | Germany |
| PI 557049 | Contorta | *C. avellana* | United Kingdom |
| PI 557033 | Casino | *C. avellana* | Spain |
| PI 557224 | Da Viega | *C. avellana* | Portugal |
| PI 634203 | Negret | *C. avellana* | Russia |
| PI 270340 | Pellicule | *C. avellana* | Spain |
| PI 271110 | Rouge² | *C. avellana* | France |
| PI 557249 | X-11 | *C. colurna* | United States |
| PI 634200 | W-5 | *C. chinensis*³ | Southern China |
| PI 557323 | Ogyoo | *C. heterophylla* | South Korea |
| PI 557019 | Winkler | *C. americana* | United States |
| None      | B0509 | *C. californica* ³ | United States |
| None      | Paperbark | *C. papyraceae* | Southern China |
| PI 557302 | C. ferox | *C. ferox* | China |
| PI 557251 | N 387 | *C. colurna* | Hungary |

¹Believed by some to be a representative of *C. maxima*.
²Synonymous to *C. colurna* var. chinensis.
³Synonymous to *C. colurna* var. californica.
genotype in any given offspring. The equations for $H_o$, $H_e$, and PIC are given in the PowerMarker software manual. Genetic distance (D) between genotypes was computed as (1 - proportion of shared alleles) (Bowcock et al., 1994). Cluster analysis of distance data used the UPGMA (unweighted pair-group method using arithmetic averages) method and resulted in a dendrogram that depicts the genetic relationships among the genotypes.

### Table 2. Microsatellite loci isolated from library A, library B, and library C. The repeat motif, sequence of the fluorescent forward primer (FAM, NED, HEX) and the reverse primer (R), and the optimum annealing temperature (Ta) are included.

| SSR Locus | Motif | Primers (5’ – 3’) | Ta |
|-----------|-------|-------------------|----|
| From (AC)$_b$ library A | | | |
| CAC-A014a (CA)$_b$ | (CA)$_b$ | FAM-GGTGGTACAGTAGGTAAGTCTAGCT | 60 ºC |
| CAC-A014b (AG)$_b$ | (AG)$_b$ | TGTGGTAGGTTCTTCGAGTTCTG | 60 ºC |
| CAC-A024b (GA)$_b$ (AT)$_2$ | (GA)$_b$ (AT)$_2$ | NED-GAATAGCTGGATGGTACAGATCT | 62 ºC |
| CAC-A036 (CA)$_b$ | (CA)$_b$ | AGGTTAGGCAAGAAGGAAAAC | 64 ºC |
| CAC-A102 (AG)$_b$ (AC)$_2$ | (AG)$_b$ (AC)$_2$ | TGTGGTAGGTTCTTCGAGTTCTG | 62 ºC |
| CAC-A105 (CA)$_b$ (AT)$_2$ | (CA)$_b$ (AT)$_2$ | NED-GAATAGCTGGATGGTACAGATCT | 58 ºC |
| CAC-A108 (CA)$_b$ (AT)$_2$ | (CA)$_b$ (AT)$_2$ | AGGTTAGGCAAGAAGGAAAAC | 58 ºC |
| From (AG)$_b$ library B | | | |
| CAC-B001 (GA)$_b$ | (GA)$_b$ | NED-CCCAATCAAACACCAACCAAC | 62 ºC |
| CAC-B010 (GA)$_b$ | (GA)$_b$ | FCCGTCGCTCAGTACACCAATGTA | 62 ºC |
| CAC-B020 (GA)$_b$ | (GA)$_b$ | HGAGAACCAACACACACACACACACACAC | 60 ºC |
| CAC-B028 (AG)$_b$ | (AG)$_b$ | TCACCAGAACCAACACACACACACACACAC | 55 ºC |
| CAC-B029b (GA)$_b$ | (GA)$_b$ | NED-GAATAGCTGGATGGTACAGATCT | 58 ºC |
| CAC-B101 (AG)$_b$ | (AG)$_b$ | HGAGAACCAACACACACACACACACACAC | 62 ºC |
| CAC-B108 (GA)$_b$ | (GA)$_b$ | GCCGTCGCTCAGTACACCAATGTA | 55 ºC |
| CAC-B109 (GA)$_b$ | (GA)$_b$ | HGAGAACCAACACACACACACACACACAC | 58 ºC |
| CAC-B110 (AG)$_b$ | (AG)$_b$ | TCACCAGAACCAACACACACACACACACAC | 64 ºC |
| CAC-B111 (AG)$_b$ | (AG)$_b$ | AGGCATGTGCCACTTAGG | 60 ºC |
| CAC-B113 (GA)$_b$ | (GA)$_b$ | CCCGCTAAGCTGGTAGGTAGT | 64 ºC |
| CAC-B114 (GA)$_b$ | (GA)$_b$ | GCAGACCAAGAGTAGGATGCA | 64 ºC |
| CAC-C003 (AG)$_b$ | (AG)$_b$ | GCCGTCGCTCAGTACACCAATGTA | 62 ºC |
| CAC-C005 (GA)$_b$ (AT)$_2$ | (GA)$_b$ (AT)$_2$ | NED-GAATAGCTGGATGGTACAGATCT | 58 ºC |
| CAC-C010 (GA)$_b$ | (GA)$_b$ | GCCGTCGCTCAGTACACCAATGTA | 58 ºC |
| CAC-C028 (GA)$_b$ | (GA)$_b$ | NED-GAATAGCTGGATGGTACAGATCT | 60 ºC |
| CAC-C111 (AG)$_b$ | (AG)$_b$ | CCCGCTAAGCTGGTAGGTAGT | 62 ºC |
| CAC-C114 (AG)$_b$ | (AG)$_b$ | GCAGACCAAGAGTAGGATGCA | 60 ºC |

$q$(GA)$_b$, q(GA)$_b$ (GA)$_b$ | (GA)$_b$ (AT)$_2$ (GAA)$_b$ | NED-GAATAGCTGGATGGTACAGATCT | 62 ºC |
| $s$(AAG)$_b$, (GA)$_b$, (AAG)$_b$ (NAG)$_b$ | (AAG)$_b$ (NAG)$_b$ | NED-GAATAGCTGGATGGTACAGATCT | 62 ºC |

Results

Initial sequencing of 30 inserts each of the CA- and GA-enriched libraries (A and B, respectively) and of 36 clones of the GAA-enriched libraries indicated an enrichment of 90% and 69.4% of di-nucleotide- and tri-nucleotide-containing sequences, respectively. Microsatellite repeats are classified as perfect, imperfect, and compound according to Weber (1990). Only 35% of the CA microsatellite motifs isolated from library A were perfect and 65% were compound, with a run of a different motif adjacent to the expected CA motif. The 12 compound motifs isolated from library A contained the CA motif, an adjacent AT motif in eight sequences, and an adjacent GA motif in the remaining four. However, 88% of GA repeats of library B were perfect. In the tri-nucleotide enriched library C, the three classes of repeat motifs were represented: 47.7% of the GA repeats were perfect, 38% were imperfect, and 14.3% were compound. This library contained the largest number of nontarget SSR motifs (4 out of 25), which consisted of AGG (2), GA (1), and GTAA (1) motifs.

**Characteristics of SSRs isolated from CA-enriched library.** Seven SSR loci were scored in 20 C. avellana cultivars and seven Corylus species (Table 3). Three of these loci contained perfect repeats (CAC-A014a, CAC-A014b, and CAC-A036) while the remaining four were compound repeats. The number of alleles per locus was high. In the C. avellana cultivars, the average number of alleles was 8.42 and the range was 5 to 11. When the eight species genotypes were included, the number of alleles per locus was much higher: the average was 16.71 and the range was 12 to 22. Microsatellites isolated from the CA-enriched library were heterozygous in C. avellana and in all the species tested (H$_o$ = 0.72 and 0.67, respectively) and highly informative as indicated by gene diversity H$_e$ of 0.76 and a PIC of 0.73 in C. avellana and an H$_o$ of 0.85 and a PIC of 0.84 over all species. At two loci, CAC-A102 and CAC-A108, an H$_o$ lower than H$_e$ indicates that alleles scored as homozygous could actually be heterozygous for a null allele. The amplified fragment sizes (in base pairs) in general ranged around the size fragment of the cloned hazelnut plasmid insert. At locus CAC-A014a, however, a much larger fragment (455 bp) than expected (222 bp) was amplified in C. ferox.

**Characteristics of GA-containing SSRs.** The 12 GA-containing loci evaluated were perfect and as heterozygous as SSRs isolated from the CA-enriched library (H$_o$ = 0.67 in C. avellana and 0.65 overall). However, three SSR loci were less...
CAC-B110 was nearly monomorphic in *C. avel-

**Table 3.** Seven loci from the CA-enriched library (A), 12 loci from the GA-enriched library (B), and 6 loci from the GAA library (C), allele size ranges, observed heterozygosity (H_o), expected heterozygosity (H_e), and polymorphic information content (PIC) in *Corylus avellana* cultivars and in eight accessions representing seven *Corylus* species.

| Marker    | Size range | Allele no. | H_o  | H_e  | PIC  | Size range | Allele no. | H_o  | H_e  | PIC  |
|-----------|------------|------------|------|------|------|------------|------------|------|------|------|
| From (AC)_A, library A |            |            |      |      |      |            |            |      |      |      |
| CAC-A014a | 211-221    | 6          | 0.80 | 0.75 | 0.70 | 205-258    | 15         | 0.75 | 0.85 | 0.84 |
| CAC-A014b | 159-187    | 11         | 0.75 | 0.86 | 0.84 | 155-173    | 16         | 0.68 | 0.89 | 0.88 |
| CAC-A024b | 120-138    | 7          | 0.80 | 0.75 | 0.71 | 111-155    | 16         | 0.79 | 0.85 | 0.83 |
| CAC-A036  | 124-143    | 5          | 0.60 | 0.61 | 0.56 | 122-149    | 12         | 0.43 | 0.79 | 0.77 |
| CAC-A102  | 282-303    | 11         | 0.65 | 0.79 | 0.78 | 269-316    | 19         | 0.61 | 0.85 | 0.84 |
| CAC-A105  | 335-358    | 4          | 0.67 | 0.57 | 0.53 | 269-316    | 19         | 0.61 | 0.85 | 0.84 |
| CAC-A106  | 216-243    | 4          | 0.66 | 0.70 | 0.62 | 198-234    | 13         | 0.49 | 0.77 | 0.76 |
| Mean      | 8.42       | 0.72       | 0.76 | 0.73 |      | 16.71      | 0.67       | 0.85 | 0.84 |      |
| From (AG)_B, library B |            |            |      |      |      |            |            |      |      |      |
| CAC-B001  | 100-119    | 8          | 0.70 | 0.71 | 0.68 | 98-110     | 11         | 0.64 | 0.77 | 0.75 |
| CAC-B023  | 208-223    | 6          | 0.85 | 0.74 | 0.70 | 202-212    | 14         | 0.68 | 0.84 | 0.82 |
| CAC-C020  | 273-289    | 8          | 0.75 | 0.77 | 0.73 | 238-286    | 18         | 0.75 | 0.87 | 0.86 |
| CAC-C028  | 254-278    | 10         | 0.95 | 0.85 | 0.84 | 252-288    | 17         | 0.86 | 0.88 | 0.87 |
| CAC-C029b | 134-163    | 10         | 0.80 | 0.82 | 0.80 | 152-185    | 17         | 0.80 | 0.88 | 0.87 |
| CAC-C108  | 311-328    | 7          | 0.52 | 0.48 | 0.46 | 310-343    | 15         | 0.36 | 0.69 | 0.67 |
| CAC-C109  | 149-174    | 12         | 0.85 | 0.83 | 0.81 | 137-159    | 18         | 0.75 | 0.87 | 0.86 |
| CAC-C110  | 195-199    | 2          | 0.05 | 0.05 | 0.05 | 197-211    | 7          | 0.21 | 0.50 | 0.48 |
| CAC-B111  | 167-189    | 6          | 0.40 | 0.66 | 0.61 | 176-186    | 9          | 0.41 | 0.72 | 0.68 |
| CAC-B113  | 158-181    | 9          | 0.85 | 0.79 | 0.75 | 160-174    | 14         | 0.81 | 0.84 | 0.87 |
| CAC-B114  | 137-149    | 6          | 0.60 | 0.64 | 0.60 | 131-159    | 13         | 0.57 | 0.78 | 0.77 |
| Mean      | 7.75       | 0.67       | 0.67 | 0.64 |      | 13.9       | 0.65       | 0.78 | 0.77 |      |
| From (GAA)_C, library C |            |            |      |      |      |            |            |      |      |      |
| CAC-C003  | 109-124    | 6          | 0.50 | 0.73 | 0.69 | 112-124    | 6          | 0.43 | 0.77 | 0.74 |
| CAC-C005  | 110-126    | 6          | 0.10 | 0.76 | 0.73 | 97-119     | 11         | 0.18 | 0.72 | 0.70 |
| CAC-C010  | 272-287    | 3          | 0.21 | 0.53 | 0.47 | 258-294    | 9          | 0.19 | 0.68 | 0.66 |
| CAC-C028  | 131-144    | 5          | 0.60 | 0.68 | 0.62 | 128-149    | 8          | 0.54 | 0.77 | 0.74 |
| CAC-C111  | 200-203    | 2          | 0.55 | 0.50 | 0.37 | 192-200    | 5          | 0.46 | 0.59 | 0.51 |
| CAC-C118  | 162-185    | 5          | 0.45 | 0.58 | 0.36 | 159-206    | 10         | 0.46 | 0.65 | 0.63 |
| Mean      | 4.3        | 0.40       | 0.59 | 0.54 |      | 8.16       | 0.37       | 0.69 | 0.66 |      |
| Overall Mean | 7.16     | 0.62       | 0.68 | 0.64 |      | 13.32      | 0.59       | 0.78 | 0.76 |      |

*Indicates the presence of null alleles.

polymorphic. CAC-B110 was nearly monomorphic in *C. avellana* genotypes. The prevalent allele was 195 bp and only ‘Imperial de Trebizonde’ was heterozygous with a 199-bp allele. A 199-bp allele was also present in *C. heterophylia* ‘Ogyoo’ and in *C. americana* ‘Winkler’. Although nearly monomorphic in *C. avellana*, CAC-B110 was heterozygous and polymorphic in the seven other *Corylus* species, which resulted in an increase in H_e from 0.05 to 0.21 over all species. Loci CAC-B108 and CAC-B111 were mostly homozygous (H_o < 0.5) and there was a large discrepancy between the H_e and H_o indicating the possible presence of null alleles. Primers for the CAC-B111 locus failed to amplify in the presence of null alleles. Primers for the CAC-B111 locus failed to amplify in *C. californica* ‘Negret’. Genetic diversity was calculated using two indices. Gene diversity (H_e) estimates the probability that two alleles at any locus are different from each other while PIC is a measure of the probability that two randomly sampled genotypes are different by their allelic profiles. In hazelnut cultivars, both H_e and PIC were high, at 0.68 and 0.64, respectively.

When eight genotypes representing seven *Corylus* species were included in the analysis, 25 SSRs generated 144 alleles for an average of 13.32 alleles per locus. The number of unique alleles observed in *Corylus* species ranged from 14 in *C. colurna* to 455 in *C. heterophylia* ‘Ogyoo’.
X-11 and in *C. heterophylla* ‘Ogyoo’ to 22 alleles in *C. ferox*. The indices of genetic diversity were also high; $H_e$ was 0.78 and PIC was 0.76.

A dendrogram depicting the relationships among the *Corylus* genotypes was generated from an UPGMA cluster analysis of genetic distances (Fig. 1). Genotypes were separated into three large groups: the other species, the European *C. avellana* cultivars and a cultivar group that contained the three genotypes from Turkey or the South Caucasus (*ʻImperial de Trebizondeʼ*, *ʻGiresun 54.021ʼ*, and *ʻGanjaʼ*), the German cultivar Hall’s Giant, the Russian seedling 681.078, and the EFB-resistant ‘Gasaway’ (which originated in Washington state).

**Cross-species amplification in *Corylus*.** Cross-species amplification was successful when up to two sharp bands of the expected size range were obtained. Cross-species transferability was high in *Corylus*. Twenty-two primer pairs amplified expected size fragments in all eight genotypes tested. CAC-A014 amplified a 455-bp fragment in *C. ferox* and CAC-B010 amplified alleles of 460 and 475 bp in *C. heterophylla* ‘Ogyoo’. Amplification failure, indicative of null genotypes, was observed at two loci. CAC-C010 failed to amplify in one species (*C. papyraceae*) while CAC-B111 amplified in only *C. heterophylla* ‘Ogyoo’ and *C. americana* ‘Winkler’, representatives of the two species that are taxonomically closely related to *C. avellana*.

**Discussion**

The enrichment protocol employed by GIS resulted in a high frequency of microsatellite-containing sequences. The larger number and frequency of microsatellites obtained from di-nucleotide-enriched libraries A and B as opposed to that obtained in tri-nucleotide library C could be the result of the predominance of di-nucleotide motifs in many plants including hazelnut. Motifs enriched in the three different libraries had different characteristics. GA motifs (library B) were exclusively perfect while CA motifs (library A) were mostly compound. This was previously observed in other plants including *Actinidia* Lindl. (Weising et al., 1996), *Malus* Mill. (Guilford et al., 1997), *Prunus* L. (Cipriani et al., 1996), and *Corylus* (Cav.)
The high level of amplification and polymorphism in the 25 microsatellite loci was expected, as it was one of the criteria used during the initial screening of the primer pairs on agarose gels using a set of 11 hazelnut genotypes. Fluorescent labeling and fragment separation of the PCR products by capillary electrophoresis were carried out only with primers that gave a high level of amplification and polymorphism in this initial set. Primer pairs that failed to amplify in some genotypes or appeared to be less polymorphic will be evaluated at a later date. Still, null alleles were likely in 8 of 25 SSR loci developed in this study as indicated by a lower than expected $H_E$. The presence of null alleles at these loci will be conclusively demonstrated by segregation analysis.

In *C. avellana* cultivars, only one GA-containing locus, CAC-B110, was nearly homozygous, detecting polymorphism in a single cultivar (‘Imperial de Trebizondé’). The remaining 18 di-nucleotide SSRs were characterized by a relatively large number of alleles per locus (≥6), high average observed and expected heterozygosity ($H_E > 0.7$, $H_O > 0.75$), and a high mean polymorphic information content (PIC ≥ 0.7). Even though the number of tri-nucleotide-containing microsatellites evaluated in this study was only 6 as opposed to 19 di-nucleotide SSRs, differences in polymorphism were apparent. Tri-SSRs were more homozygous ($H_E = 0.4$ on average) and less informative than di-nucleotide SSRs in cultivars of *C. avellana*, as indicated by a lower mean number of alleles per locus (4.5), $H_E (0.59)$, and PIC (0.54). Similar differences in level of polymorphism between the two classes of microsatellites (di-SSRs vs. tri-SSRs) were reported in other plant species (Huttel et al., 1999; Rossetto et al., 1999; Scotti et al., 2002). The high informativeness of CA- and GA-containing microsatellites as opposed to that obtained from tri-nucleotide SSRs explains the predominant enrichment and use of di-nucleotide SSRs for cultivar identification in plants. Still, trinucleotide markers are useful for cultivar discrimination in *Glycine max* (L.) Merr. for example (Song et al., 1999) and we showed them to be similarly useful in hazelnut.

Based on our microsatellite analysis, the hazelnut is a genetically diverse heterozygous plant species. The 13.32 alleles per locus generated by these 25 primer pairs is high and is comparable to that obtained from other out-crossed woody perennials, such as the 9.5 alleles per locus reported for *Persea* (Lavi et al., 1994), 14.3 for *Quercus* L. (Dow et al., 1995), 10.7 for *Prunus cerasus* L. (Cantini et al., 2001), and 12.1 for *Malus* (Hokanson et al., 1998). The average values for $H_E (0.59)$, $H_O (0.78)$, and PIC (0.76) indicate a high level of genetic diversity in hazelnut. The three most informative SSRs (CAC-A014b, CAC-A108, and CAC-B109) contained di-nucleotide motifs and distinguished between the 20 cultivars of *C. avellana* as well as the remaining seven species (data not shown). Both classes of microsatellite markers developed in this study will be used for fingerprinting and cultivar identification in hazelnut.

Transportability of microsatellites in different species confirmed by sequence analysis was reported in *Prunus* (Zhebenty-ayeva et al., 2003), in *Vitis* L. (Di Gasparo et al., 2000), between species of *Malus* and *Pyrus* L. in the subfamily Pomoideae of the Rosaceae (Yamamoto et al., 2001), and between *Quercus* and *Castanea* Mill. (Barreneche et al., 2004). The small number of *Corylus* species genotypes used in this study is not sufficient to estimate the extent of polymorphism of microsatellite markers that were developed in *C. avellana* in various hazelnut species. However, most of the 25 loci isolated from *C. avellana* in this study also amplified in seven other species of *Corylus*. All 25 primer pairs gave successful amplification in the shrub species *C. heterophylla* and *C. americana*, two species known to be taxonomically closest to *C. avellana*. Amplification in the remaining five species failed at only one locus, CAC-B111. CAC-C010 primers failed to amplify only in *C. californica*. The presence of insertions could have caused the larger than expected fragments amplified by CAC-B010 and CAC-A014a in *C. heterophylla* and *C. ferox*, respectively, and will be confirmed by sequencing. Microsatellite markers isolated from *C. avellana* will be evaluated for the ability to amplify the same locus in other hazelnut species (by sequencing) and for taxon identification using a large group of genotypes of each species.

Tri-nucleotide SSRs may be more likely to generate conserved loci useful for cross-species transportability. High cross-species transportability is associated with their presence in open reading frames (ORFs) involved in eukaryotic transcription, replication, gene expression, and regulation (Kashi and Seller, 1999; Sinden, 1999; Young et al., 2000). In pine, a set of *Pinus strobus* L. di-nucleotide microsatellites had no (Echt et al., 1999) or limited (0.54) similarity to *C. avellana*. In addition, tri-nucleotide SSRs explain the predominant enrichment and use of di-nucleotide SSRs for cultivar identification in plants. Still, trinucleotide markers are useful for cultivar discrimination in *Glycine max* (L.) Merr. for example (Song et al., 1999) and we showed them to be similarly useful in hazelnut.

In this study, we clearly demonstrated cross-species amplification in *Corylus*. However, no differences in trans-specific amplification were observed between di-nucleotide and tri-nucleotide SSRs, probably because the loci were selected for their ability to amplify an initial set of genotypes that included five *Corylus* species. Differences in the ability of these two classes of SSRs to cross-amplify in species and the orthology of the microsatellite loci in other *Corylus* species will be evaluated in the future.

This initial study of microsatellite polymorphism indicates that microsatellites will be extremely useful for curation of the hazelnut collection at the NCGR. Microsatellites will be used to estimate overall genetic diversity in the hazelnut collection, assign unique genetic fingerprints to core cultivars of *C. avellana*, and identify duplicates. The potential for using microsatellites developed in the cultivated hazelnut for fingerprinting *Corylus* species remains to be determined.

**Literature Cited**

Barreneche, T., M. Casasoli, K. Russell, A. Akkak, H. Meddour, C. Plomion, F. Villani, and A. Kremer. 2004. Comparative mapping between *Quercus* and *Castanea* using simple-sequence repeats (SSRs). Theor. Appl. Genet. 108:558–566.

Bassil, N.V., R. Botta, and S.A. Mehlenbacher. 2003. Microsatellite markers of the European hazelnut. HortScience 38:740–741. (Abstr.)

Bassil, N.V. and A.N. Azarenko. 2001. Random amplified polymorphic DNA (RAPD) markers for self-incompatibility in *Corylus avellana*. Acta Hort. 556:537–543.

Bowcock, A.M., A. Ruis-Linares, J. Tomofohrde, E. Minch, J.R. Kidd, and L.L. Cavalli-Sforza. 1994. High resolution of human evolutionary trees from polymorphic microsatellites. Nature 368:455–457.

Cantini, C., A.F. Iezzoni, W.F. Lamboy, M. Boritzki, and D. Struss. 2001. DNA fingerprinting of tetraploid cherry germplasm using simple sequence repeats. J. Amer. Soc. Hort. Sci. 126:205–209.

Cardle, L., L. Ramsay, D. Milbourne, M. Macaulay, D. Marshall, and et al., 1999; Testolin et al., 2000), *Persea* Mill. (Sharon et al., 1997), and *Picea A. Dietr.* (Pfeiffer et al., 1997). The compound repeats contained AT and GA repeats, which are the most common di-nucleotide repeats in plants (Cardle et al., 2000; Lagercrantz et al., 1993; Morgante and Olivieri, 1993; Temnykh et al., 2001; Wang et al., 1994).
R. Waugh. 2000. Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156:847–854.

Cipriani, G., G. Lot, W.-G. Huang, M.T. Marrazzo, E. Peterlunger, and R. Testolin. 1999. AC/GT and AG/CT microsatellite repeats in peach [Prunus persica (L.) Batsch.]: Isolation, characterization and cross-species amplification in Prunus. Theor. Appl. Genet. 99:65–72.

Davis, J. and S.A. Mehlenbacher. 1997. Identification of RAPD markers linked to eastern filbert blight resistance in hazelnut. Acta Hort. 445:553–556.

Di Gaspero, M.C., F. Peterlunger, R. Testolin, K.J. Edwards, and G. Pfeiffer, A., A.M. Olivieri, and M. Morgante. 1997. Identification and characterization of microsatellites in Norway spruce (Picea abies K.). Genome 40:411–419.

Pomper, K.W., A.N. Azarenko, N. Bassil, J.W. Davis, and S.A. Mehlenbacher. 1998. Identification of random amplified polymorphic DNA (RAPD) markers for self-incompatibility alleles in Corylus avellana L. Theor. Appl. Genet. 97:479–487.

Rossetto, M., A. MaLauchlan, F.C.L. Harriss, R.J. Henry, P.R. Baverstock, L.S. Lee, T.L. Maguire, and K.J. Edwards. 1999. Abundance and polymorphism of microsatellite markers in the tea tree (Melaleuca alternifolia, Myrtaceae). Theor. Appl. Genet. 98:1091–1098.

Rozen, S. and H.J. Skaltsky. 2000. Primer3 on the WWW for general users and for biologist programmers, p. 365–386. In: S. Krawetz and S. Misener (eds.). Bioinformatics methods and protocols: Methods in molecular biology. Humana Press, Totowa, N.J.

Scotti, I., G.P. Paglia, F. Magni, and M. Morgante. 2002. Efficient development of dinucleotide microsatellite markers in norway spruce (Picea abies Karst.) through dot-blot selection. Theor. Appl. Genet. 104:1035–1041.

Sharon, D., P.B. Cregan, S. Khameed, M. Kusharska, J. Hillel, E. Lahav, and U. Lavi. 1997. An integrated genetic linkage map of avocado. Theor. Appl. Genet. 95:911–921.

Sinden, R.R. 1999. Biological implications of the DNA structures associated with disease causing triplet repeats. Amer. J. Human Genet. 64:346–353.

Song, Q.J., C.V. Quigley, R.L. Nelson, T.R. Carter, H.R. Boerma, J.L. Strachan, and P.B. Cregan. 1999. A selected set of trinucleotide simple sequence repeat markers for soybean cultivar identification. Plant Var. Seeds 12:207–220.

Temnykh, S., G. DeClerck, A. Lukashova, L. Lipovich, S. Cartinhour, and S. McCouch. 2001. Computational and experimental analysis of microsatellites in rice (Oryza sativa L.): Frequency, length variation, transposon associations, and genetic potential. Genome Res. 11:1441–1452.

Testolin, R., T. Marrazzo, G. Cipriani, R. Quarta, I. Verde, M.T. Dettori, M. Panicaldi, and S. Sansavini. 2000. Microsatellite DNA in peach (Prunus persica L. Batsch.) and its use in fingerprinting and testing the genetic origin of cultivars. Genome 43:512–520.

Thompson, M.M., H.B. Lagerstedt, and S.A. Mehlenbacher. 1996. Hazelnuts, p. 125–184. In: J. Janick and J.N. Moore (eds.). Fruit breeding, seeds 12:207–220.

Young, E.T., J.S. Sloan, and K. Van Riper. 2000. Trinucleotide repeats are clustered in regulatory genes in Saccharomyces cerevisiae. Genet. 154:1053–1068.

Zhebentyayeva, T.N., G.L. Reighard, V.M. Gorina, and A.G. Abbott. 2003. Simple sequence repeat (SSR) analysis for assessment of genetic variability in apricot germplasm. Theor. Appl. Genet. 106:435-444.