Genet Resour Crop Evol (2013) 60:543–568
DOI 10.1007/s10722-012-9857-z

RESEARCH ARTICLE

Nuclear and chloroplast microsatellite markers to assess genetic diversity and evolution in hazelnut species, hybrids and cultivars

Nahla Bassil · Paolo Boccacci · Roberto Botta · Joseph Postman · Shawn Mehlenbacher

Received: 19 January 2012 / Accepted: 7 May 2012 / Published online: 19 July 2012
© Springer Science+Business Media B.V. (outside the USA) 2012

Abstract The US Department of Agriculture, Agricultural Research Service, National Clonal Germplasm Repository in Corvallis, Oregon, preserves more than 800 accessions of hazelnut (*Corylus*), including *C. avellana* cultivars and representatives of 10 other recognized shrub and tree species. Characterization and study of genetic diversity in this collection require cross-transferable markers, such as trinucleotide microsatellite or simple sequence repeat (SSR) markers and universal chloroplast SSR markers. We developed new SSR markers and evaluated 114 *Corylus* accessions representing 11 species and 44 interspecific hybrids. Eight of 23 SSRs generated easy-to-score alleles in all species and seven were highly polymorphic. For those seven, the average heterozygosity was moderate at 0.49, and mean allele number, genetic diversity and polymorphism information index were high at 11.71, 0.79 and 0.76, respectively. The three most polymorphic SSRs were CaC-C008, CaC-C040 and CaC-C118. Neighbor-joining (NJ) clustering and structure analysis agreed with classical taxonomic analysis and supported inclusion of *C. maxima* within the large polymorphic species, *C. avellana*. Analysis also indicated that *C. californica* is a distinct species rather than a botanical variety of *C. cornuta*. Six universal cpSSRs were polymorphic in *Corylus* and generated 21 distinct chlorotypes with an average of 3 alleles per locus. Diversity at these cpSSRs was high and ranged from 0.33 to 0.64, with an average of 0.54. Incongruence in NJ topologies between the nuclear and chloroplast markers could be attributed to chloroplast capture related to hybridization during the ancestral diversification of the genus, or to homoplasy. The phylogeographical relationships among the 21 chlorotypes in the 11 *Corylus* species support Asia as a refugium where several hazelnut lineages survived during glaciation and from which they continued to evolve after dispersal from Asia through the Mediterranean to Europe, and across the Atlantic and/or the Bering land bridge to North America.

Electronic supplementary material The online version of this article (doi:10.1007/s10722-012-9857-z) contains supplementary material, which is available to authorized users.

N. Bassil (✉) · J. Postman
United States Department of Agriculture (USDA), Agricultural Research Service (ARS), National Clonal Germplasm Repository (NCGR), 33447 Peoria Road, Corvallis, OR 97333, USA
e-mail: nahla.bassil@ars.usda.gov

P. Boccacci
Plant Virology Institute, National Research Council (IVV-CNR), UOS of Grugliasco, Via Leonardo da Vinci, 44, 10095 Grugliasco, TO, Italy

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

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

Keywords *Corylus* · Filbert · Simple sequence repeat (SSR) markers · Universal chloroplast SSRs

Introduction

Hazelnut, *Corylus* L., belongs to the family Betulaceae and subfamily Coryloideae. In addition to *Corylus*, the Coryloideae contains hornbeam (*Carpinus* L.), hop-hornbeam (*Ostrya* Scopoli), and *Ostryopsis* Decne. (Crane 1989; Cronquist 1981). The second subfamily, the Betuloideae, consists of alder (*Alnus* Mill.) and birch (*Betula* L.). The oldest known fossil record attributed to *Corylus* is a fruit involucre from the middle Eocene ($\sim$45 mya) in the Republic Flora of central Washington (Chen et al. 1999; Pigg et al. 2003). Coryloideae is supported as a monophyletic group (Yoo and Wen 2002, 2007) and shares several distinguishing characters including nutlets without lateral wings, vessels without spiral thickenings, absence of tracheids, and pollen without arci. Hazelnuts, like other members of the birch family, are deciduous, wind-pollinated, monoecious shrubs and trees with toothed, simple, ovate to obovate leaves alternately arranged. Morphological synapomorphies that are characteristic of *Corylus* include large animal-dispersed nuts and filaments that are completely divided longitudinally (Chen et al. 1999). The chromosome number of the genus is $2n = 2x = 22$ (Thompson et al. 1996).

The taxonomy of *Corylus* has been investigated since the mid-nineteenth century, with the number of recognized species dependent on the emphasis placed by various authors on certain anatomical and morphological characters (illustrated in Table 1 of Whitcher and Wen 2001). The inclusion of taxa within each section or subgenus of *Corylus* has varied significantly. The division of the genus into two sections, *Acanthochlamys* and *Corylus*, as proposed by De Candolle (1864) and followed by Schneider (1916), and Li and Cheng (1979), agrees with internal transcribed spacer (ITS) phylogeny (Whitcher and Wen 2001). The tree species *C. ferox* Wall., with its distinctive spiny bur-like involucres, has invariably been placed in section or subgenus *Acanthochlamys* Spach. Within section *Corylus*, three subsections are traditionally recognized. Subsection *Colurnae* Schneider consists of the tree species: *C. colurna* L., *C. jacquemontii* Decne., *C. chinensis* Franch. and *C. fargesii* C. K. Schneider. Subsection *Siphonochlamys* contains the bristle-husked shrubs: *C. cornuta* Marshall, *C. californica* Marshall and *C. sieboldiana* Blume. Subsection *Phyllochlamys* includes the shrubs with leafy involucres: *C. avellana* L., *C. americana* Marshall and the *C. heterophylla* Fisch. complex. Based on morphological traits (especially the husk or involucres) and molecular ITS and chloroplast rbcL phylogenetic analyses, *Acanthochlamys* is sister to the remainder of the genus *Corylus,* and subgenera *Siphonochlamys* and *Phyllochlamys* are sister taxa (Erdoğan and Mehlenbacher 2000a; Forest and Bruneau 2000; Forest et al. 2005; Whitcher and Wen 2001).

Corylus contains 11 commonly recognized species disjunctly distributed in the Northern Hemisphere. Of 11 species, two species occur in Europe and Asia Minor (*C. avellana* and *C. colurna*), three in North America (*C. americana* and *C. cornuta* in the east and *C. californica* in the west), and one in the Himalayas (*C. jacquemontii*). The remaining species are endemic to eastern Asia and include the tree hazels: *C. chinensis*, *C. fargesii* Schneid. and *C. ferox*, and the shrub hazels: *C. heterophylla* and *C. sieboldiana* (Whitcher and Wen 2001). Although these 11 species are commonly recognized, other species designations can be found in the literature. *Corylus maxima* Mill., *C. pontica* Koch, and *C. colchica* Alb. have been recognized by some authors (Kasapligil 1972) as distinct species closely related to *C. avellana.* Others consider these three to be variants within that highly polymorphic species. Their morphological traits show continuous distributions, they are easily crossed with each other and give fully fertile offspring, and their geographic distributions overlap (Mehlenbacher 1991; Rovira 1997; Thompson et al. 1996). Within the bristle-husked shrubs (*Siphonochlamys*), *C. californica* is recognized as a distinct species by some authorities, and as a subspecies or botanical variety of *C. cornuta* by others. Within the Asian leafy-husked shrubs,varieties *sutchuensis* Franch. and *yunnanensis* Franch. are adapted to warmer climates than is the typical variety *heterophylla* of *C. heterophylla* (Thompson et al. 1996). They are recognized as botanical varieties of *C. heterophylla* by some authorities, but as separate species, *C. kweichowensis* Hu (Liang and Zhang 1988) and *C. yunnanensis* (Franch.) A. Camus, respectively, by others (Liang and Zhang 1988; Thompson et al. 1996). Further, *C. thibetica*

Table 1 List of *Corylus* accessions used in this study

No.	Accession number	Local inv. (CCOR)	Taxon	Name	Origin	Chlorotype	Name in dendrogram
1	PI 557018	61.001	*C. americana*	C. amer. 61.001	Missouri	Q	C. amer. CCOR61
2	PI 557019	99.001	*C. americana*	'Winkler'	Iowa	R	C. amer. Winkler
3	PI 557020	117.001	*C. americana*	C. amer. 117.001	Minnesota	Q	C. amer. CCOR117
4	PI 495606	180.002	*C. americana*	C. amer. 180.002	Iowa	Q	C. amer. CCOR180.002
5	PI 557021	225.001	*C. americana*	C. amer. 225.001	Iowa	Q	C. amer. CCOR225
6	PI 617169	228.001	*C. americana*	C. amer. 228.001	Missouri	Q	C. amer. CCOR228
7	PI 557022	386.001	*C. americana*	'Rush'	Pennsylvania	Q	C. amer. Rush
8	PI 617242	675.001	*C. americana*	C. amer. 675.001	Illinois	R	C. amer. CCOR675
9	PI 617243	676.001	*C. americana*	C. amer. 676.001	Wisconsin	Q	C. amer. CCOR676
10	PI 617244	677.001	*C. americana*	C. amer. 677.001	North Dakota	Q	C. amer. CCOR677
11	PI 617245	678.001	*C. americana*	C. amer. 678.001	Pennsylvania	S	C. amer. CCOR678
12	PI 617246	679.001	*C. americana*	C. amer. 679.001	West Virginia	B	C. amer. CCOR679
13	PI 617248	681.001	*C. americana*	C. amer. 681.001	Kentucky	Q	C. amer. CCOR681
14	PI 617249	682.001	*C. americana*	C. amer. 682.001	Michigan	T	C. amer. CCOR682
15	PI 617250	683.001	*C. americana*	C. amer. 683.001	Iowa	R	C. amer. CCOR683
16	PI 617251	684.001	*C. americana*	C. amer. 684.001	Iowa	R	C. amer. CCOR684
17	PI 617252	685.001	*C. americana*	C. amer. 685.001	Wisconsin	U	C. amer. CCOR685
18	PI 617253	686.001	*C. americana*	C. amer. 686.001	Pennsylvania	U	C. amer. CCOR686
19	PI 617254	687.001	*C. americana*	C. amer. 687.001	Maryland	S	C. amer. CCOR687
20	PI 617260	693.001	*C. americana*	C. amer. 693.001	New Jersey	Q	C. amer. CCOR693
21	PI 617261	694.001	*C. americana*	C. amer. 694.001	Minnesota	R	C. amer. CCOR694
22	PI 617262	695.001	*C. americana*	C. amer. 695.001	Minnesota	Q	C. amer. CCOR695
23	PI 617263	696.001	*C. americana*	C. amer. 696.001	Michigan	T	C. amer. CCOR696
24	PI 617272	709.001	*C. americana*	C. amer. 709.001	Wisconsin	U	C. amer. CCOR709
25	PI 617275	712.001	*C. americana*	C. amer. 712.001	Massachusetts	Q	C. amer. CCOR712
26	PI 617278	715.001	*C. americana*	C. amer. 715.001	Michigan	U	C. amer. CCOR715
27	PI 270340	8.001	*C. avellana*	'Negret'	Spain	A	C. av. Negret
28	PI 557037	36.001	*C. avellana*	'Barcelona'	Spain	A	C. av. Barcelona
29	PI 557167	344.001	*C. avellana*	'Ratoli'	Spain	A	C. av. Ratoli
30	PI 271110	38.001	*C. maxima*	'Pellicule Rouge'	France	A	C. max. Pellicule Rouge
31	PI 557400	272.001	*C. maxima*	'Istarski duguljasti'	Croatia	A	C. max. Istarski duguljasti
32	PI 557401	357.001	*C. maxima*	'di San Benedetto'	Italy	A	C. max. San Benedetto
33			*C. chinensis*	OSU 567.011	China	I	C. chi. CCOR567.011
34			*C. chinensis*	OSU 567.018	China	I	C. chi. CCOR567.018
35			*C. chinensis*	OSU 529.001	China	K	C. chi. OSU 529.001

Table 1 continued

No.	Accession number	Local inv. (CCOR)	Taxon	Name	Origin	Chlorotype	Name in dendrogram
36			*C. chinensis*	OSU 529.017	China	K	C. chi. OSU 529.017
37			*C. chinensis*	OSU Lagerstedt East	China via Australia	J	C. chi. OSU Lag.East
38			*C. chinensis*	OSU Lagerstedt West	China via Australia	J	C. chi. OSU Lag.West
39			*C. chinensis*	OSU W03	China via Australia	J	C. chi. OSU W3
40			*C. chinensis*	OSU W05	China via Australia	J	C. chi. OSU W5
41	PI 617204	591.001	*C. chinensis*	OSU 91502	China	I	C. chi. CCOR591.001
42			*C. colurna*	C. colurna 97093	Serbia	E	C. col. 97093
43			*C. colurna*	C. colurna 97094	Serbia	E	C. col. 97094
44			*C. colurna*	C. colurna 97095	Serbia	E	C. col. 97095
45			*C. colurna*	C. colurna 97096	Serbia	E	C. col. 97096
46			*C. colurna*	C. colurna 97097	Serbia	E	C. col. 97097
47			*C. colurna*	C. colurna 97098	Serbia	E	C. col. 97098
48			*C. colurna*	C. colurna 97099	Serbia	E	C. col. 97099
49			*C. colurna*	C. colurna 97100	Serbia	E	C. col. 97100
50			*C. colurna*	C. colurna LB1.26	Serbia	E	C. col. LB1_26
51			*C. colurna*	OSU Pole Barn	France	E	C. col. Pole Barn
52	PI 557253	450.001	*C. colurna*	C. colurna N451	Warsaw, Poland	F	C. col. CCOR450
53	PI 557255	452.001	*C. colurna*	C. colurna N504	Slepcany, Czech Rep.	E	C. col. CCOR452
54	PI 557256	453.001	*C. colurna*	C. colurna 550	Geisenheim, Germany	E	C. col. CCOR453
55	PI 557269	109.001	*C. cornuta*	C. cornuta Minnesota	Maine	Q	C. cor. CCOR109
56	PI 637894	814.001	*C. cornuta*	C. cornuta CC2.50 Minnesota	New York	Q	C. cor. CCOR814
57	PI 637895	815.001	*C. cornuta*	C. cornuta CC3.01 New York	Minnesota	Q	C. cor. CCOR815
58	PI 637896	816.001	*C. cornuta*	C. cornuta CC3.47 Wisconsin	New York	Q	C. cor. CCOR816
59	PI 637897	817.001	*C. cornuta*	C. cornuta CC3.58	Wisconsin	Q	C. cor. CCOR817
60	PI 637898	818.001	*C. cornuta*	C. cornuta CC3.113 Quebec	Quebec	Q	C. cor. CCOR818
61	PI 637899	819.001	*C. cornuta*	C. cornuta CC4.46 North Dakota	North Dakota	Q	C. cor. CCOR819
62	PI 637900	820.001	*C. cornuta*	C. cornuta CC4.53 Manitoba	Manitoba	Q	C. cor. CCOR820
63	PI 637901	821.001	*C. cornuta*	C. cornuta OSU 373.032 British Columbia	British Columbia	Q	C. cor. CCOR821
64	PI 637886	801.001	*C. cornuta*	C. cornuta 661.081 Manitoba	Minnesota	Q	C. cor. CCOR801
65	PI 637887	802.001	*C. cornuta*	C. cornuta 662.006 Saskatch	Manitoba	Q	C. cor. CCOR802
66	PI 557280	233.001	*C. californica*	C. californica 61-4 Lewis, WA	Oregon	P	C. cal. CCOR233
67	PI 557281	234.001	*C. californica*	C. californica 27-5 Hood River	Oregon	P	C. cal. CCOR234

Table 1 continued

No.	Accession number	Local inv. (CCOR)	Taxon	Name	Origin	Chlorotype	Name in dendrogram
68	PI 557282	235.001	*C. californica*	C. californica 49-2 Clatsop	Oregon	P	C. cal. CCOR235
69	PI 557283	236.001	*C. californica*	C. californica 58-5 Columbia	Oregon	P	C. cal. CCOR236
70	PI 557284	237.001	*C. californica*	C. californica 52-5 Multnomah	Oregon	P	C. cal. CCOR237
71	PI 557285	238.001	*C. californica*	C. californica 51-3 Multnomah	Oregon	P	C. cal. CCOR238
72	PI 557286	239.001	*C. californica*	C. californica 59-1 Douglas	Oregon	P	C. cal. CCOR239
73	PI 557287	240.001	*C. californica*	C. californica 23-6 Wash. Co.	Oregon	P	C. cal. CCOR240
–	PI 557288	241.001	*C. californica*	C. californica 21-5 Lincoln	Oregon	P	C. cal. CCOR241
74	PI 557290	243.001	*C. californica*	C. californica 10-6 Benton	Oregon	P	C. cal. CCOR243
75	PI 557291	244.001	*C. californica*	C. californica 45-6 Lane	Oregon	P	C. cal. CCOR244
76	PI 557293	428.001	*C. californica*	C. californica 13-3 Oregon	Oregon	P	C. cal. CCOR428
77	PI 557294	429.001	*C. californica*	C. californica 3-6 Oregon	Oregon	P	C. cal. CCOR429
78	PI 557295	430.001	*C. californica*	C. californica 25-5 Oregon	Oregon	P	C. cal. CCOR430
79	PI 557297	432.001	*C. californica*	C. californica 20-6 Oregon	Oregon	P	C. cal. CCOR432
–	PI 557298	433.001	*C. californica*	C. californica 41-2 Oregon	Oregon	P	C. cal. CCOR433
80	PI 557299	434.001	*C. californica*	C. californica 53-4 Oregon	Oregon	P	C. cal. CCOR434
81	PI 557300	435.001	*C. californica*	C. californica 13-5 Oregon	Oregon	P	C. cal. CCOR435
82	PI 557273	470.001	*C. californica*	C. californica #8	Oregon	P	C. cal. CCOR470
83	PI 557274	497.001	*C. californica*	C. californica # 8/D	Oregon	P	C. cal. CCOR497
–	PI 557275	498.001	*C. californica*	C. californica # 2/S	Oregon	P	C. cal. CCOR498
–	PI 557276	503.001	*C. californica*	C. californica #3	Oregon	P	C. cal. CCOR503
84	PI 557277	504.001	*C. californica*	C. californica #15	Oregon	P	C. cal. CCOR504
–	PI 557278	506.001	*C. californica*	C. californica #16	Oregon	P	C. cal. CCOR506
85	PI 617197	583.001	*C. californica*	C. californica 4-6	Oregon	P	C. cal. CCOR583
86	PI 617198	584.001	*C. californica*	C. californica 13-3	Oregon	P	C. cal. CCOR584
87	PI 617199	585.001	*C. californica*	C. californica 25-3	Oregon	P	C. cal. CCOR585
88	PI 617200	586.001	*C. californica*	C. californica 53-6	Oregon	P	C. cal. CCOR586
89	PI 617201	588.001	*C. californica*	C. californica 66-5	Oregon	P	C. cal. CCOR588
90	PI 617202	589.001	*C. californica*	C. californica 19-4	Oregon	P	C. cal. CCOR589
91	OSU	Mehlenb	*C. fargesii*	C. fargesii 1[a]	China	L	C. fargesii 1
92		Mehlenb	*C. fargesii*	Paperbark C-3 Farris		M	C. fargesii C-3
93	PI 557302	185.001	*C. ferox*	C. ferox 185.001	China	H	C. ferox CCOR185
94	OSU	Mehlenb	*C. ferox*	C. ferox WS		H	C. ferox WS
95	PI 557309	67.001	*C. heterophylla*	C. heterophylla Korea-10	Korea	N	C. het. CCOR67
96	PI 557310	124.001	*C. heterophylla*	C. heterophylla Jilin	China	N	C. het. CCOR124

Table 1 continued

No.	Accession number	Local inv. (CCOR)	Taxon	Name	Origin	Chlorotype	Name in dendrogram
97	PI 557311	146.001	*C. heterophylla*	C. heterophylla 'Nanking'		N	C. het. CCOR146
98	PI 557311	147.001	*C. heterophylla*	C. heterophylla 'Nanking'		N	C. het. CCOR147
99	PI 557315	351.001	*C. heterophylla*	C. heterophylla seedling A		N	C. het. CCOR351
100	PI 557328	64.001	*C. heterophylla* var. *thunbergii*	C. heterophylla var. thunbergii Korea-66	Korea	O	C. het. thunbergii CCOR64
101	PI 557330	286.001	*C. heterophylla* var. *yunnanensis*	C. heterophylla var. yunnanensis China	China	N	C. het. yun. CCOR286
102	OSU	Mehlenb	*C. jacquemontii*	OSU 397.027	Pakistan	G	C. jacqu. OSU397.027
103	OSU	Mehlenb	*C. jacquemontii*	OSU 397.050	Pakistan	G	C. jacqu. OSU397.050
104	OSU	Mehlenb	*C. jacquemontii*	OSU 397.024	Pakistan	G	C. jacqu. OSU397.024
105	PI 557268	311.001	*C. jacquemontii*	C. jacquemontii 880430 Pakistan	Pakistan	G	C. jacqu. CCOR311
106	PI 617206	593.001	*C. jacquemontii*	C. jacquemontii OSU 88501	India	G	C. jacqu. CCOR593
107	PI 557404	348.001	*C. sieboldiana*	C. sieboldiana	Korea	N	C. sieb. CCOR348
108	PI 557409	347.001	*C. sieboldiana* var. *brevirostris*	C. sieboldiana var. brevirostris seedling	Korea	N	C. sieb. brevirostris CCOR347
109	PI 557415	349.001	*C. sieboldiana* var. *mandshurica*	C. sieboldiana var. mandshurica	Korea	N	C. sieb. mand. CCOR349
	PI 557337	100.001	*Corylus* hybrid (Ame–Ave)	C. americana hybrid NY 104	New York	Q	C. amer. hybrid NY 104
	PI 557338	101.001	*Corylus* hybrid (Ame–Ave)	C. americana hybrid NY 110	New York	Q	C. amer. hybrid NY 110
	PI 557339	102.001	*Corylus* hybrid (Ame–Ave)	C. americana hybrid NY F-45	New York	Q	C. amer. hybrid NY F-45
	PI 557340	103.001	*Corylus* hybrid (Ame–Ave)	C. americana hybrid NY 200	New York	P	C. amer. hybrid NY 200
	PI 557341	104.001	*Corylus* hybrid (Ame–Ave)	C. americana hybrid NY 616	New York	Q	C. amer. hybrid NY 616
	PI 557379	189.001	*Corylus* hybrid (Ame–Ave)	C. americana hybrid NY F-20	New York	Q	C. amer. hybrid NY F-20
	PI 557383	194.001	*Corylus* hybrid (Ame–Ave)	C. americana hybrid NY 1464	New York	Q	C. amer. hybrid NY 1464
	PI 557391	377.001	*Corylus* hybrid (Ame–Ave)	'Potomac'	Maryland	Q	Potomac
	PI 557334	378.001	*Corylus* hybrid (Ame–Ave)	'Buchanan'	Pennsylvania	Q	Buchanan
	PI 557392	383.001	*Corylus* hybrid (Ame–Ave)	'Reed'	Maryland	Q	Reed
	PI 617214	638.001	*Corylus* hybrid (Ame–Ave)	Corylus americana hybrid	Oregon	Q	C. amer. hybrid CCOR638
	OSU	G081S	*Corylus* hybrid (Ame–Ave)	Rutter G081S	Minnesota	R	Rutter G081S
	OSU	G227S	*Corylus* hybrid (Ame–Ave)	Rutter G227S	Minnesota	Q	Rutter G227S
	OSU		*Corylus* hybrid (Ame–Ave)	Weschcke TP2	Wisconsin	Q	Weschcke TP2
	PI 617187	561.001	*Corylus* hybrid (Ame–Ave)	Weschcke TP3	Wisconsin	Q	Weschcke TP3
	PI 641155	853.001	*Corylus* hybrid (Ame–Ave)	'Yoder 5'	Ohio	Q	Yoder 5
	OSU		*Corylus* hybrid (Ame–Ave)	Weschcke TP1	Wisconsin	Q	Weschcke TP1
	PI 557331	33.001	*Corylus* hybrid (Col–Ave)	'Morrisoka'	British Columbia	E	Morrisoka
	PI 557332	53.001	*Corylus* hybrid (Col–Ave)	'Filcorn'	Oregon	A	Filcorn

Table 1 continued

No.	Accession number	Local inv. (CCOR)	Taxon	Name	Origin	Chlorotype	Name in dendrogram
	PI 557333	57.001	*Corylus* hybrid (Col–Ave)	'Laroka'	British Columbia	E	Laroka
	PI 557349	137.001	*Corylus* hybrid (Col–Ave)	'Moturk-D'	Michigan	E	Moturk-D
	PI 557261	138.001	*Corylus* hybrid (Col–Ave)	Chinese Trazel Gellatly No. 6	British Columbia	E	Chinese Trazel Gellatly No. 6
	PI 557357	148.001	*Corylus* hybrid (Col–Ave)	'Eastoka'	British Columbia	E	Eastoka
	PI 557359	150.001	*Corylus* hybrid (Col–Ave)	'Moturk-B'	Michigan	E	Moturk-B
	PI 557362	154.001	*Corylus* hybrid (Col–Ave)	'Freeoka'	British Columbia	E	Freeoka
	PI 557369	165.001	*Corylus* hybrid (Col–Ave)	'Dundee'	Oregon	E	Dundee
	PI 557372	168.001	*Corylus* hybrid (Col–Ave)	'Newburg'	Oregon	E	Newburg
	PI 557263	170.001	*Corylus* hybrid (Col–Ave)	Chinese Trazel J-1	Oregon	E	Chinese Trazel J-1
	PI 557374	171.001	*Corylus* hybrid (Col–Ave)	USOR 13-71	Oregon	E	USOR 13–71
	PI 557264	173.001	*Corylus* hybrid (Col–Ave)	Chinese Trazel Gellatly No. 11	British Columbia	E	Chinese Trazel Gellatly No. 11
	PI 557387	199.001	*Corylus* hybrid (Col–Ave)	'Chinoka'	British Columbia	B	Chinoka
	PI 557389	201.001	*Corylus* hybrid (Col–Ave)	'Erioka'	British Columbia	B	Erioka
	PI 557390	202.001	*Corylus* hybrid (Col–Ave)	'Ruby'	Oregon	B	Ruby
	PI 557393	405.002	*Corylus* hybrid (Col–Ave)	'Faroka'	British Columbia	E	Faroka
	PI 557394	406.001	*Corylus* hybrid (Col–Ave)	'Karloka'	British Columbia	E	Karloka
	PI 557396	408.001	*Corylus* hybrid (Col–Ave)	Turktrazel Gellatly No. 15	British Columbia	E	Turktrazel Gellatly No. 15
	PI 617185	559.001	*Corylus* hybrid (Col–Ave)	'Grand Traverse'	Michigan	E	Grand Traverse
	PI 617191	574.001	*Corylus* hybrid (Col–Ave)	Farris 88 BS	Michigan	E	Farris 88 BS
	OSU		*Corylus* hybrid (Col–Ave)	'Lisa'	Michigan	E	Lisa
	PI 557429	9.001	*Corylus* × *colurnoides* C. K. Schneider (Col × Ave)	C. × colurnoides L-1		E	C. × colurnoides L-1
	PI 557350	139.001	*Corylus* hybrid (Het Sut–Ave)	Estrella No. 1	Michigan	N	Estrella No. 1
	PI 557351	140.001	*Corylus* hybrid (Het Sut–Ave)	Estrella No. 2	Michigan	N	Estrella No. 2
	PI 557430	14.001	*Corylus* × *vilmorinii* Rehder (Chi x Ave)	C. × vilmorinii Arnold Arboretum	Massachusetts	B	C. × vilmorinii CCOR14
	PI 617265	701.001	*Corylus* hybrid (or avellana?)	18-32 EFB-resistant	New York	A	Medium long
			C. avellana L.	'Culplà'	Spain	A	C. av_Culplà
			C. avellana L.	'Gironell'	Spain	A	C. av_Gironell
			C. avellana L.	'Grifoll'	Spain	A	C. av_Grifoll
			C. avellana L.	'Morell'	Spain	A	C. av_Morell
			C. avellana L.	'Pauetet'	Spain	A	C. av_Pauetet
			C. avellana L.	'Ribet'	Spain	A	C. av_Ribet
			C. avellana L.	'Trenet'	Spain	A	C. av_Trenet

Table 1 continued

No.	Accession number	Local inv. (CCOR)	Taxon	Name	Origin	Chlorotype	Name in dendrogram
			C. avellana L.	'Camponica'	Italy	A	C. av_Camponica
			C. avellana L.	'Mortarella'	Italy	A	C. av_Mortarella
			C. avellana L.	'Nocchione'	Italy	A	C. av_Noccione
			C. avellana L.	'Riccia di Talanico'	Italy	A	C. av_Riccia di Talanico
			C. avellana L.	'San Giovanni'	Italy	A	C. av_San Giovanni
			C. avellana L.	'Tonda bianca'	Italy	D	C. av_T.Bianca
			C. avellana L.	'Tonda di Giffoni'	Italy	A	C. av_T.Giffoni
			C. avellana L.	'Tonda Gentile Langhe'	Italy	A	C. av_T.G.Langhe
			C. avellana L.	'Tonda Gentile Romana'	Italy	A	C. av_T.G.Romana
			C. avellana L.	'Tonda rossa'	Italy	D	C. av_T.Rossa
			C. avellana L.	'Badem'	Turkey	A	C. av_Badem
			C. avellana L.	'Extra Ghiagli'	Turkey	A	C. av_Extra Ghiagli
			C. avellana L.	'Imperiale di Trebisonda'	Turkey	B	C. av_I.Trebizonde
			C. avellana L.	'Incekara'	Turkey	B	C. av_Incekara
			C. avellana L.	'Kalinkara'	Turkey	B	C. av_Kalinkara
			C. avellana L.	'Palaz'	Turkey	B	C. av_Palaz
			C. avellana L.	'Sivri'	Turkey	A	C. av_Sivri
			C. avellana L.	'Sivri Ghiaghli'	Turkey	B	C. av_Sivri Ghiagli
			C. avellana L.	'Tombul'	Turkey	A	C. av_Tombul
			C. avellana L.	'Tombul Ghiaghli'	Turkey	B	C. av_Tombul Ghiagli
			C. avellana L.	'Asle Gharebag'	Iran	C	C. av_Asle Gharebag
			C. avellana L.	'Dobooseh'	Iran	A	C. av_Dobooseh
			C. avellana L.	'Jorow Gharebag'	Iran	C	C. av_Jorow Gharebag
			C. avellana L.	'Mish-pestan'	Iran	C	C. av_Mish-pestan
			C. avellana L.	'Nakhoni Rood'	Iran	C	C. av_Nakhoni Rood
			C. avellana L.	'Pashmineh'	Iran	C	C. av_Pashmineh
			C. avellana L.	'Rasmi'	Iran	C	C. av_Rasmi
			C. avellana L.	'Shastak-2'	Iran	C	C. av_Shastak-2
			C. avellana L.	'Shirvani'	Iran	C	C. av_Shirvani
			C. avellana L.	'Tabari Rood'	Iran	B	C. av_Tabari Rood

Their plant introduction (PI) number, Local inventory number (prefix CCOR for Corvallis *Corylus*), taxon, origin and chlorotype are listed. O.P. indicates open pollinated. The number listed for each accession corresponds to the numbers in Fig. 3 and the dash (–) indicates *C. californica* samples that were not included in NJ clustering or structure analyses because they amplified 3 alleles at CAC-C040 while empty cells refer to the 37 *C. avellana* samples previously characterized by Boccacci and Botta 2009

[a] Indicates single accession of *C. fargesii* included in assessing amplification and polymorphism of the 15 SSRs described in Suppl. Table 1

Batalin is sometimes listed as a morphological variant of *C. ferox* (Liang and Zhang 1988), and *C. mandshurica* Maxim. (The Plant List 2010, Thompson et al. 1996) and *C. hallaisanensis* Nakai (The Plant List 2010) have been noted as synonyms or variants of *C. sieboldiana* and *C. wangii* Hu has been considered a form of *C. chinensis* (Liang and Zhang 1988). In this paper, we follow the consensus recognition of six shrub species (*C. avellana*, *C. americana*, *C. heterophylla*, *C. cornuta*, *C. californica*, and *C. sieboldiana*) and five tree species (*C. colurna*, *C. jacquemontii*, *C. chinensis*, *C. fargesii* and *C. ferox*) (Mehlenbacher 2009).

The US Department of Agriculture (USDA), Agricultural Research Service (ARS), National Clonal Germplasm Repository (NCGR), in Corvallis, Oregon, conserves more than 800 hazelnut accessions representing cultivars and representatives of each of these 11 species (Bassil et al. 2009).

Microsatellite or simple sequence repeat (SSR) markers have become valuable molecular tools for fingerprinting accessions, assessment of genetic diversity in collections and linkage mapping, due to their abundance, high degree of polymorphism, co-dominance and suitability for automation. For such a diverse germplasm collection, markers that are transferable across species are needed. Trinucleotide SSRs seem to be better candidates than dinucleotide SSRs for cross-transferability (Kutil and Williams 2001; Morgante et al. 2002; Scotti et al. 2000; Wang et al. 1994; Young et al. 2000). They are often clustered in regulatory genes (Young et al. 2000) and are more likely than dinucleotide SSRs to be found within expressed regions (Morgante et al. 2002; Wang et al. 1994). Trinucleotide repeats were three times more frequent in transcribed than in non-transcribed regions of the *Arabidopsis thaliana* L. and *Zea mays* L. genomes (Morgante et al. 2002). They are more likely to be conserved across taxa, but tend to be less polymorphic than are dinucleotide SSRs (Kutil and Williams 2001; Rajora et al. 2001; Shepherd et al. 2002). Alleles at trinucleotide SSRs are easier to score due to a lower frequency and extent of the characteristic stuttering that plagues most dinucleotide alleles. Trinucleotide and tetranucleotide repeats have become the markers of choice for population, linkage and forensic studies in humans and other animal species (Gastier et al. 1995; Sheffield et al. 1995; Tozaki et al. 2000) and are recommended as universal markers in plants (Testolin and Cipriani 2010). SSR markers were developed in *C. avellana* (Bassil et al. 2005a, b; Boccacci et al. 2005; Gürcan and Mehlenbacher 2010a, b; Gürcan et al. 2010a) and used for linkage mapping (Mehlenbacher et al. 2006; Gürcan et al. 2010a), to assess genetic relationships among cultivars (Boccacci and Botta 2010; Boccacci et al. 2006, 2008; Ghanbari et al. 2005; Gökirmak et al. 2009, Gürcan et al. 2010b) and to fingerprint cultivars in collections, identify synonyms, and determine parentage (Botta et al. 2005; Gökirmak et al. 2009; Sathuvalli and Mehlenbacher 2011). Cross-species transference of SSRs was demonstrated in *Corylus* (Bassil et al. 2005a; Boccacci et al. 2005) and, more broadly, within the Betulaceae (Gürcan and Mehlenbacher 2010b).

The chloroplast genome has a lower evolutionary rate than does the nuclear genome. It is non-recombining and shows a uniparental mode of inheritance, usually maternal in angiosperms and paternal in gymnosperms (Provan et al. 2001). In *Corylus*, interspecific hybrids have the maternal allele (Malusà 1994), indicating maternal inheritance. Thus, in hazelnut the chloroplast genome can only be disseminated by seeds or cuttings, and chloroplast DNA markers provide information on past changes in species distribution that are mostly unaffected by subsequent pollen exchange or dispersal. Despite its conserved gene order and relative lack of recombination, the chloroplast genome shows length polymorphisms associated with mononucleotide repeats. Non-coding intron and intergenic spacers are particularly variable and contain microsatellite and non-microsatellite polymorphisms even between closely related individuals and taxa in a range of plant groups (Provan et al. 2001). In recent years, universal primer pairs have been developed for the analysis of chloroplast SSRs (cpSSRs) in different species (Provan et al. 2001). In several studies, cpSSRs provided insights into intraspecific phylogeographic variability (e.g., Petit et al. 2003) and allowed investigation of the origins and domestication of different crop species (e.g., Arroyo-García et al. 2006). Their application to hazelnut is recent and to date has only been applied to *C. avellana* for investigating the post-glacial migration of wild populations in Europe (Palmé and Vendramin 2002) and studying the origin and diffusion of hazelnut cultivars in the Mediterranean basin (Boccacci and Botta 2009).

The aim of this study was to determine cross-transferability of nuclear (n) SSRs isolated from a *C. avellana* library enriched for trinucleotide repeats to the 11 *Corylus* species preserved at the NCGR, to identify the nuclear and chloroplast SSR markers most suitable for future studies of *Corylus*, to fingerprint representative accessions from each species, and to assess diversity, structure and evolution within the genus.

Materials and methods

Plant material and DNA extraction

The hazelnut accessions evaluated in this study were in the collection at USDA-ARS-NCGR and the Oregon State University's Smith Horticultural Research Farm in Corvallis, OR (Table 1). We evaluated 158 accessions, including 6 *C. avellana* (which include 3 previously assigned to *C. maxima*), 26 *C. americana*, 30 *C. californica*, 9 *C. chinensis*, 13 *C. colurna*, 11 *C. cornuta*, 2 *C. fargesii*, 2 *C. ferox*, 7 *C. heterophylla*, 5 *C. jacquemontii*, 3 *C. sieboldiana* and 44 interspecific hybrids. DNA was extracted from actively growing leaves collected from the NCGR field in the spring by using a modified PUREGENE® kit (Gentra Systems Inc., Minneapolis, MN) protocol. Proteinase K and RNAse A treatments were added, and the protein-precipitation step was repeated twice.

Cross-species amplification

GAA-enriched library 'C' construction and primer design were previously described (Bassil et al. 2005a; Gürcan et al. 2010a, b). Twenty-three primer pairs were designed from 22 SSR-containing sequences and were tested for amplification in each of the accessions. Amplification success was indicated by the presence of a PCR product after ethidium bromide staining of 3 % agarose gels. The 15 unique SSR primer pairs (Supplementary Table 1) that generated a product in all 11 species were investigated further, with sizing by capillary electrophoresis.

Microsatellite marker analysis

Fluorescently-labeled forward primers for the 15 SSR products were used for PCR amplification (Suppl. Table 1). PCR reactions were carried out separately for each primer pair, and up to three PCR products (one per SSR primer set) were multiplexed and separated with an ABI 3100 capillary electrophoresis instrument (Applied Biosystems, Foster City, CA) at the Core Labs of the Center for Genome Research and Biocomputing at Oregon State University. PCR reactions were carried out in 10 μL volumes by using forward primers fluorescently labeled with 6-FAM, 5-HEX, or NED and unlabeled reverse primers (Operon Biotechnologies, Huntsville, AL). The PCR reactions were diluted with water by a factor ranging from 1:80 (FAM-labeled amplicons) and 1:160 (HEX-labeled products) to 1:320 (NED-labeled amplicons), and 0.5 μL was injected into the instrument. GeneScan version 2.1 (Applied Biosystems) was used for automated data collection and Genotyper version 2.0 (Applied Biosystems) for allele-size estimation.

PCR reactions were performed in a 10 μL volume containing 1 × reaction buffer, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.3 μM of each primer, 0.25 units of Biolase *Taq* DNA polymerase (Bioline USA Inc., Randolph, MA), and 2.5 ng genomic DNA. The PCR protocol consisted of one cycle of initial denaturation at 94 °C for 3 min, followed by 35 cycles of denaturation at 93 °C for 40 s, annealing at optimum T_a (Suppl. Table 1) for 40 s, and extension at 72 °C for 40 s. A final extension cycle at 72 °C for 30 min followed. DNA was amplified in an Eppendorf Gradient thermocycler (Brinkmann Instruments, Inc., Westbury, NY) or an MJ Research Tetrad thermocycler (MJ Research Inc., Watertown, MA). The success of the PCR reaction was verified by 2 % agarose gel electrophoresis prior to capillary electrophoresis.

Diversity and clustering

Of the 15 primer pairs from 23 tested (see Suppl. Table 1) that generated a product in all 11 species, CaC-C114 uniquely generated up to four PCR products, indicating its presence in more than one location in the hazelnut genome. Because of this, data for CaC-C114 were not included in further analyses. PowerMarker (Version 3.25) (Liu and Muse 2005) was used to calculate genetic diversity parameters for the 11 species at the remaining 14 SSR loci (Table 2) using all except for five *C. californica* accessions that generated 3 alleles with CAC-C040 (Table 1). These five *C. californica* accessions were excluded from

Table 2 Diversity parameters of 14 hazelnut loci in each of the 11 species evaluated in this study

Species	*C. americana*					*C. avellana*					*C. californica*				
Marker	H_e	H_o	*PIC*	*A*	A_u	H_e	H_o	*PIC*	*A*	A_u	H_e	H_o	*PIC*	*A*	A_u
CaC-C001b	0.75	0.81	0.72	8	127	0.74	0.17	0.70	5	–	0.22	0.24	0.21	4	95, 112
CaC-C003	0.39	0.35	0.36	4	–	0.40	0.17	0.36	3	–	0.69	0.72	0.64	5	–
CaC-C005	0.30	0.31	0.28	4	115, 121	0	0	0	1	–	0.68	0.76	0.61	4	–
CaC-C008	0.72	0.38	0.69	7	236	0.71	1	0.65	4	–	0.81	0.80	0.79	8	189
CaC-C028	0.75	0.35	0.71	8	–	0.64	1	0.57	4	–	0.40	0.32	0.37	5	–
CaC-C036	0	0	0	1	–	0	0	0	1	–	0	0	0	1	–
CaC-C040	0.58	0.69	0.49	3	–	0.50	0.33	0.45	3	–	0.50	1	0.38	2	–
CaC-C108	0.15	0.12	0.14	5	178	0.44	0	0.35	2	–	0.11	0.12	0.11	2	–
CaC-C111	0.61	0.62	0.53	3	–	0.44	0.33	0.35	2	–	0	0	0	1	–
CaC-C112	0.59	0.15	0.51	4	256, 276	0	0	0	1	–	0.63	0.44	0.59	4	266
CaC-C118	0.64	0.69	0.58	5	–	0.28	0.33	0.24	2	–	0.53	0.40	0.49	7	–
CaC-C119	0.41	0.35	0.39	4	–	0.42	0.50	0.39	4	–	0	0	0	1	–
CaT-C501	0.79	0.50	0.76	10	188	0.72	0.83	0.68	5	212, 213	0.80	0.72	0.78	10	183, 191, 192
CaT-C504	0.81	0.62	0.78	8	–	0.61	0	0.54	3	–	0.34	0.32	0.32	4	–
Mean	0.53	0.42	0.50	5.29		0.42	0.33	0.38	2.86		0.41	0.42	0.38	4.14	

Species	*C. chinensis*					*C. colurna*					*C. cornuta*				
Marker	H_e	H_o	*PIC*	*A*	A_u	H_e	H_o	*PIC*	*A*	A_u	H_e	H_o	*PIC*	*A*	A_u
CaC-C001b	0.45	0.44	0.42	4	107	0.78	0.85	0.74	8	101	0.79	0.73	0.76	6	118, 122
CaC-C003	0.55	0.44	0.49	3	–	0.46	0.00	0.40	3	–	0.67	0.64	0.63	5	–
CaC-C005	0.44	0.22	0.41	4	94	0.07	0.08	0.07	2	–	0.46	0.36	0.36	2	–
CaC-C008	0.84	0.78	0.82	8	–	0.60	0.62	0.54	3	–	0.64	0.64	0.58	6	187
CaC-C028	0.69	0.67	0.63	4	–	0.67	0.38	0.62	5	–	0	0	0	1	–
CaC-C036	0	0	0	1	–	0	0	0	1	–	0	0	0	1	–
CaC-C040	0.38	0.44	0.35	4	–	0.71	0.54	0.65	5	170	0.60	0.45	0.57	6	206
CaC-C108	0	0	0	1	–	0.51	0.46	0.45	3	–	0.62	0.64	0.54	3	–
CaC-C111	0.54	0.78	0.47	3	206	0	0	0	1	–	0.17	0.18	0.15	2	–
CaC-C112	0	0	0	1	–	0.07	0.08	0.07	2	–	0.09	0.09	0.08	2	–
CaC-C118	0.59	0.56	0.57	6	–	0.73	0.54	0.68	5	165	0.68	0.45	0.64	6	200
CaC-C119	0.64	0.67	0.58	4	–	0.21	0.23	0.20	3	–	0.09	0.09	0.08	2	–
CaT-C501	0.71	0.33	0.66	5	206	0.47	0	0.36	2	–	0.79	0.91	0.75	6	–

Table 2 continued

Species	*C. chinensis*					*C. colurna*					*C. cornuta*				
Marker	H_e	H_o	*PIC*	*A*	A_u	H_e	H_o	*PIC*	*A*	A_u	H_e	H_o	*PIC*	*A*	A_u
CaT-C504	0.73	0.67	0.69	6	156, 162, 175	0.82	0.92	0.79	7	–	0.71	0.73	0.67	5	147
Mean	0.47	0.43	0.44	3.86		0.44	0.34	0.40	3.57		0.45	0.42	0.42	3.79	
Species	*C. fargesii*					*C. ferox*					*C. heterophylla*				
Marker	H_e	H_o	*PIC*	*A*	A_u	H_e	H_o	*PIC*	*A*	A_u	H_e	H_o	*PIC*	*A*	A_u
CaC-C001b	0.5	0	0.38	2	–	0.50	0	0.38	2	98	0.76	0.71	0.73	7	–
CaC-C003	0	0	0	1	–	0.38	0.50	0.30	2	–	0.52	0.43	0.46	3	–
CaC-C005	0	0	0	1	–	1	1	0.38	2	–	0.52	0.57	0.46	3	118, 124
CaC-C008	0	0	0	1	–	0.63	1	0.55	3	–	0.78	0.86	0.74	6	230, 245
CaC-C028	0	0	0	1	–	0	0	0	1	138	0.65	0.43	0.60	5	–
CaC-C036	0	0	0	1	–	0	0	0	1	–	0	0	0	1	–
CaC-C040	0.38	0.50	0.30	2	–	0.63	1	0.55	3	218	0.53	0.57	0.48	4	–
CaC-C108	0	0	0	1	–	0	0	0	1	–	0.54	0.57	0.50	4	–
CaC-C111	0	0	0	1	–	0	0	0	1	–	0.64	1	0.57	3	–
CaC-C112	0	0	0	1	–	0.38	0.50	0.30	2	259, 280	0	0	0	1	–
CaC-C118	0.38	0.5	0.30	2	–	0.63	1	0.55	3	–	0.46	0.43	0.43	4	–
CaC-C119	0.5	1	0.38	2	–	0.63	0.50	0.55	3	–	0.69	0.57	0.63	4	–
CaT-C501	0.38	0.50	0.30	2	211	0.50	0	0.38	2	–	0.74	0.71	0.72	7	221, 224, 225
CaT-C504	0	0	0	1	–	0	0	0	1	–	0.55	0.57	0.52	5	–
Mean	0.15	0.18	0.12	1.36		0.34	0.39	0.28	1.93		0.53	0.53	0.49	4.07	
Species	*C. jacquemontii*					*C. sieboldiana*					Overall				
Marker	H_e	H_o	*PIC*	*A*	A_u	H_e	H_o	*PIC*	*A*	A_u	H_e	H_o	*PIC*	*A*	A_u
CaC-C001b	0.50	0.60	0.38	2	–	0.67	0.67	0.59	3	106					–
CaC-C003	0	0	0	2	100, 103	0.50	0.33	0.38	2	–	0.78	0.39	0.74	8	–
CaC-C005	0	0	0	1	–	0.28	0.33	0.24	2	103	0.68	0.34	0.66	12	–
CaC-C008	0.50	0.60	0.38	2	–	0.78	1	0.74	5	–	0.92	0.66	0.91	21	–
CaC-C028	0.48	0.40	0.36	2	–	0.50	0.33	0.45	3	–	0.80	0.44	0.78	9	–
CaC-C036	0	0	0	1	–	0	0	0	1	–					–
CaC-C040	0.54	0.80	0.47	3	173	0.28	0.33	0.24	2	–	0.83	0.65	0.80	14	–
CaC-C108	0.48	0.40	0.36	2	–	0	0	0	1	182					–

Table 2 continued

Species	*C. fargesii*					*C. ferox*					*C. heterophylla*				
Marker	H_e	H_o	*PIC*	*A*	A_u	H_e	H_o	*PIC*	*A*	A_u	H_e	H_o	*PIC*	*A*	A_u
CaC-C111	0	0	0	1	–	0.44	0.67	0.35	2	–	0.68	0.38	0.64	6	–
CaC-C112	0	0	0	1	–	0.44	0	0.35	2	–					–
CaC-C118	0.42	0.60	0.33	2	–	0.50	0.67	0.45	3	–	0.82	0.54	0.80	12	–
CaC-C119	0	0	0	1	–	0.28	0.33	0.24	2	–					–
CaT-C501	0.34	0.40	0.31	3	187	0.72	1	0.67	4	–					–
CaT-C504	0.62	0.20	0.55	3	177	0.72	1	0.67	4	–					–
Mean	0.31	0.29	0.25	1.86		0.44	0.48	0.38	2.57		0.79	0.49	0.76	11.7	

Allele number (*A*), observed heterozygosity (H_o), expected heterozygosity (H_e), and polymorphism information index (*PIC*) were calculated for each species with PowerMarker. Number of unique alleles (A_u) is also listed. Overall *A*, H_o, and *PIC* were calculated only for the eight SSR loci that amplified in all species and were used for cluster and structure analysis

further downstream nuclear SSR analyses resulting in 109 of the 114 *Corylus* species representatives and 44 hybrid accessions. These diversity measures consisted of: number of alleles (*A*); observed heterozygosity (H_o) or the number of heterozygous individuals in that population; gene diversity, often referred to as expected heterozygosity (H_e) and defined as the probability that two randomly chosen alleles from the population are different; and polymorphism information content (*PIC*) (Botstein et al. 1980). Species-specific or unique alleles (A_u) observed in only one species were also noted (Table 2).

Eight of the 14 SSR loci characterized in each species were easy to score in all species and generated allele sizes expected on the basis of repeat motif (Suppl. Table 1). Genetic distance matrices were computed with PowerMarker from data for these eight SSRs by calculating the proportion of shared allele distance (D_{sa}):

$$D_{sa} = \frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{a_j}\min(p_{ij}, q_{ij})$$

where p_{ij} and q_{ij} are the frequencies of the *i*th allele at the *j*th locus, *m* is the number of loci examined, and a_j is the number of alleles at the *j*th locus. Neighbor-joining (NJ) cluster analysis was used to group all the accessions except for the 5 *C. californica* samples that had 3 alleles at CAC-C040 based on these eight SSR loci (Fig. 2).

Structure analysis

The software program Structure 2.3.3 (Pritchard et al. 2000) was used to infer population structure and assign individuals to modeled populations based on their SSR genotypes. Structure uses a Bayesian approach to model-based clustering. Multiple runs were performed by setting the number of populations, k, from 5 to 12. The burn-in length was set to 200,000 with runs of 100,000 steps, and each run was replicated three times.

Chloroplast haplotype determination and data analysis

Ten cpSSR loci were analyzed: ccmp1, ccmp2, ccmp3, ccmp4, ccmp5, ccmp6, ccmp7, ccmp8, ccmp9, and ccmp10. The corresponding primer pairs were

designed by Weising and Gardner (1999) for *Nicotiana tabacum* L., and loci were initially tested in 40 accessions representing 11 *Corylus* species. Then, polymorphic cpSSR were used to determine the chloroplast haplotypes of 158 accessions, of which 114 represented *Corylus* species and 44 were labeled as interspecific hybrids. PCR amplification was carried out by using a reaction mixture (15 µl) consisting of 40 ng DNA template, 0.5 µM of each primer, 200 µM dNTPs, 2 mM $MgCl_2$, 1.5 µl 10 × NH_4 buffer [160 mM $(NH_4)_2SO_4$, 670 mM Tris–HCl (pH 8.8 at 25 °C), 0.1 % Tween-20], and 0.5 U BioTaq DNA polymerase (Bioline, London, UK). A thermocycler (MJ Research Inc., Watertown, MA) was used with the following temperature profile: 3 min of denaturation at 95 °C, then 28 cycles of 30 s of denaturation at 95 °C, 45 s of annealing at 54 °C, and 90 s of extension at 72 °C, with 10 min at 72 °C as the final extension step. Amplified fragments were loaded on a capillary sequencer ABI-PRISM 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA). Results of the run were processed with Genemapper v. 4.0 software and allele sizes estimated from GeneScan-500 LIZ size standards (Applied Biosystems).

In order to characterize allelic diversity and informativeness of polymorphic cpSSRs in *Corylus* species, the number of alleles (*A*) and the gene diversity (H_e) were calculated for 114 *Corylus* accessions (excluding hybrids) and 37 additional *C. avellana* cultivars previously characterized by Boccacci and Botta (2009), who also employed the aforesaid methods (PCR amplification and SSR analysis). *A* was directly estimated, while H_e was calculated as:

$$H_e = 1 - \Sigma p_i^2$$

where p_i is the frequency of the *i*th allele (Nei 1987).

Pairwise genetic distances (1,000 bootstraps) between 151 *Corylus* accessions were computed as:

$$D = [1 - (\text{proportion of shared alleles})]$$

with Microsat software (Minch 1997). A NJ tree was constructed with Mega v. 5 software (Tamura et al. 2011), including an individual of *Carpinus betulus* L. as an outgroup taxon. To reconstruct a chloroplast DNA genealogy, a reduced median (RM) network was built based on the length multi-state of microsatellites. This maximum-parsimony analysis was performed by using Network software (Bandelt et al. 1999), selecting the reduced median algorithm and the maximum parsimony (MP) option.

Results

SSR amplification and polymorphism

Nuclear SSRs developed from a GAA-enriched library contained GA/CT, GAA/CTT, AGG/TCC, and GTAA motifs (Suppl. Table 1). Only CaC-C001b and CaC-C119 contained dinucleotide motifs, while CaC-C001a uniquely contained a hepta-nucleotide motif, CACAGAG. Amplification of 23 SSR primer pairs was assessed first after 3 % agarose gel electrophoresis (Suppl. Table 1). Polymorphism in *C. fargesii* could not be properly evaluated, since only a single accession (Table 1) of this species was available. Amplification rates across species were high, ranging from 74 to 100 %. All 23 primer pairs amplified in *C. avellana* as well as in *C. americana*. In fact, CaC-C103 only amplified in these two species but failed to amplify in any accessions of the other nine species. Based on the SSR primer pairs that generated amplification products for all the species, the polymorphism rate ranged from 41 % in *C. jacquemontii* to 90 % in *C. heterophylla*. The results (Suppl. Table 1) indicate that a variety of options are available for researchers interested in using SSRs for *Corylus* diversity assessments, even in those taxa that are disjunctly distributed (Fig. 1).

Of the 15 primer pairs that were evaluated by capillary electrophoresis in the 158 accessions, six proved less than reliable for inclusion in our analyses. CaC-114 generated one or two PCR products ranging in size from 260 to 279 bp in *C. avellana*, the bristle–husked species, *C. californica*, *C. cornuta* and *C. sieboldiana,* and the tree hazels, *C. fargesii* and *C. chinensis*, where it can be used for genetic studies. However, it generated up to four PCR products in the remaining species, indicating a possible genomic duplication. Of the two dinucleotide-containing SSRs identified in this library, CaC-C001b was highly diverse, as estimated from *A*, H_o, H_e and *PIC* in each of the species, but CaC-C119 was less polymorphic (*A,* 2–4; *PIC*, 0.22–0.63) and amplified a single product in *C. californica, C. jacquemontii* and most of the *C. cornuta* accessions (Table 2). CaC-C001b also generated a large number (9) of species-specific

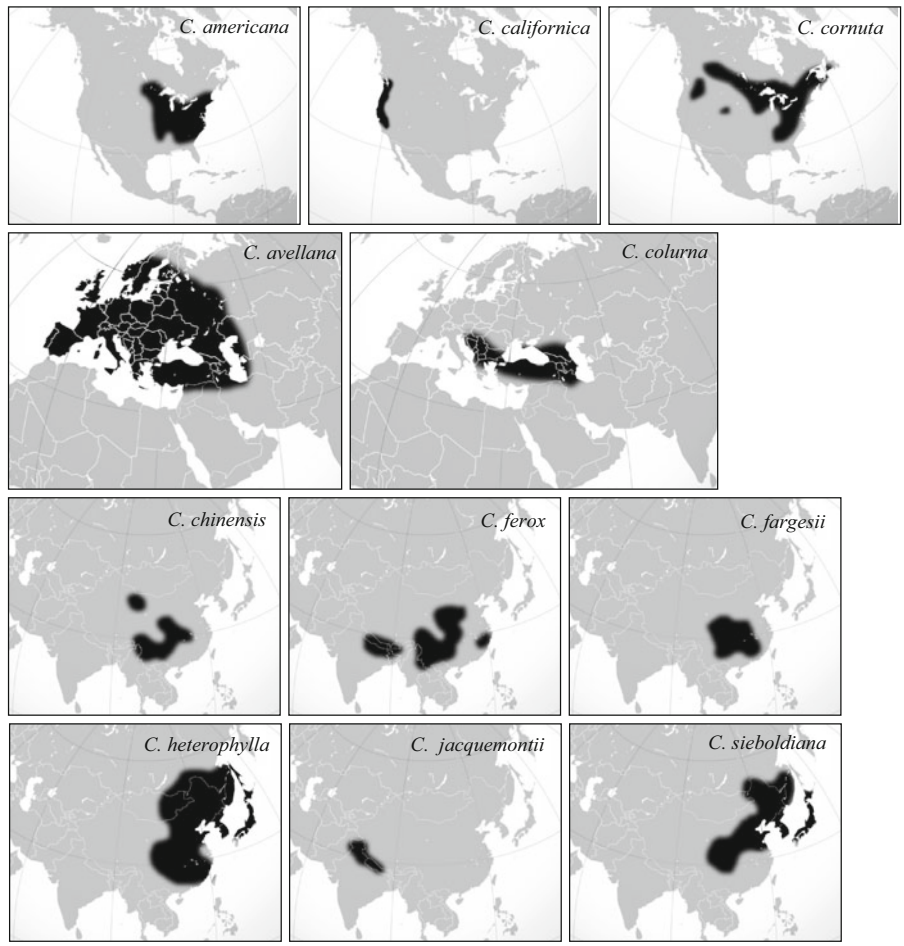

Fig. 1 Geographic distribution of *Corylus* species

alleles (Table 2). Four of the trinucleotide containing SSRs (CaC-C108, CaC-C112, CaT-C501 and CaT-C504) generated many alleles that differed by 1 or 2 bp, possibly indicating sequence differences in the sequence flanking the repeat and other than in repeat number. The resulting alleles generated by these four primer pairs were also difficult to score and were thus excluded in cluster or structure analyses. The above-mentioned 7 SSRs were excluded from further analysis.

Among the remaining 8 SSRs that generated easy-to-score alleles in all species, CaC-C036 contained a tetra-nucleotide motif and amplified the same allele (163) in all species except for *C. californica* and *C. jacquemontii*, where it generated a 155 bp long fragment. In the other 7 SSRs, the average heterozygosity was moderate at 0.49, while mean allele number, genetic diversity and *PIC* were high at 11.71, 0.79 and 0.76, respectively. A single allele (128) was in common between *C. cornuta* and *C. fargesii* accessions at CaC-C028 which generated another single unique allele (138) in *C. ferox*. CaC-C028 was polymorphic in the remaining species. The three most polymorphic trinucleotide SSR primer pairs, as based on the largest number of alleles (A) and a relatively high number of unique alleles (A_u) as compared to the others, were CaC-C008, CaC-C040 and CaC-C118 (Table 2). The largest number of alleles ($A = 21$) was observed at CaC-C008; this included five species-specific alleles. At CaC-C040, A was 15 and A_u was 4 (Table 2). At CaC-C118, A was 4 and A_u was 2 (Table 2).

Nuclear microsatellite-based clustering

NJ cluster analysis based on the shared allele distance (D) is depicted in Fig. 2. The hazelnut accessions were grouped into six groups: a '*Species*' group that contained eight of the species, but not *C. americana*, *C. avellana* or *C. fargesii*; two small hybrid groups (Hyb1 and Hyb2); two *C. americana* groups (Americana-Winkler and Americana-Rush); and a *C. avellana* group.

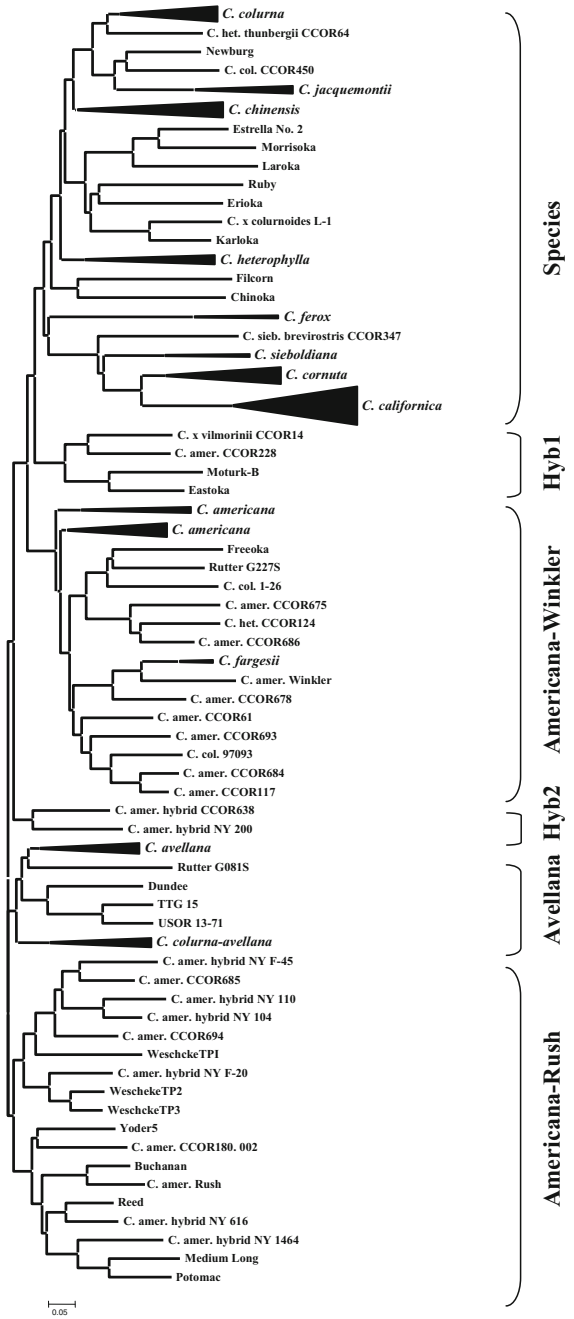

Fig. 2 NJ cluster analysis of hazelnut accessions based on the proportion of shared allele distance for 8 trinucleotide-containing SSRs (except for CAC-C036 which contains a tetranucleotide repeat)

Species cluster

In the 'Species' group, accessions of the tree species, *C. colurna*, *C. jacquemontii* and *C. chinensis*, grouped together, as did accessions of the bristle-husked species, *C. sieboldiana*, *C. cornuta* and *C. californica*. Five of the seven *C. heterophylla* accessions formed a *C. heterophylla* group, which also included one *C. heterophylla* × *C. avellana* hybrid (Estrella #1). *Corylus heterophylla* CCOR124 was in a mixed subgroup within the Americana group, and the sole C. *heterophylla* var. *thunbergii* accession (CCOR64) was sister to the *C. colurna* group. The two *C. ferox* accessions grouped together and were sister to the *C. cornuta* complex. Three groups of *C. colurna* × *C. avellana* accessions were also found in this large group: 'Newberg' (CCOR168) grouped with *C colurna* accession CCOR450 in the tree species group; five *C. colurna* × *C. avellana* hybrids, mostly from Gellatly's work in British Columbia, grouped together with the *C. heterophylla* × *C. avellana* hybrid Estrella #2 and *C.* × *colurnoides* Schneid. CCOR9; and a third group was composed of two hybrid accessions, 'Filcorn' and 'Chinoka'.

Hybrid groups

The first hybrid group (Hyb1) contained the only *C.* × *vilmorinii* Rehder accession (CCOR14), which grouped with a *C. americana* accession from Missouri (CCOR228). These two accessions were adjacent to the *C. colurna* × *C. avellana* hybrids, 'Moturk-B' from Michigan and 'Eastoka' from British Columbia. The second hybrid group (Hyb 2) was formed by the *C. americana* × *C. avellana* hybrids CCOR638 and NY 200.

Americana groups

Two large groups contained the majority of the *C. americana* accessions. The first group included 'Winkler', and the second included 'Rush'. The

Americana-Winkler group contained the largest number of *C. americana* accessions and was divided into three subgroups. The first two subgroups consisted of *C. americana* accessions from West Virginia, North Dakota, Kentucky, Wisconsin, Michigan, Iowa, Maryland, Massachusetts and Minnesota. The third subgroup included *C. americana* accessions CCOR675 from Illinois and CCOR686 from Pennsylvania, and *C. heterophylla* CCOR124 from China. Also in this subgroup were *C. americana* × *C. avellana* hybrid 'Rutter G227S', *C. colurna* LB01.26 from Serbia and a *C. colurna* × *C. avellana* hybrid, 'Freeoka' from British Columbia. The second subgroup contained the two *C. fargesii* accessions which grouped together, in addition to a *C. colurna* accession (97093) from Serbia and a group of *C. americana* accessions from Iowa ('Winkler' and CCOR684), Pennsylvania, Missouri, New Jersey and Minnesota.

The Americana-Rush group contained the selections of *C. americana* × *C. avellana* hybrids of the early breeders, John F. Jones (Lancaster, PA), Clarence A. Reed (Washington, DC), George L. Slate (Geneva, NY), and Carl Weschcke (St. Paul, MN). This group was subdivided into two subgroups. The first one contained three of Weschcke's hybrids (TP1, TP2 and TP3), Slate's New York selections (NY F-45, NY 110, NY 104, and NY F-20), and two *C. americana* accessions, CCOR685 from Wisconsin and CCOR694 from Minnesota. The second subgroup contained the Jones hybrid 'Buchanan', which grouped with its parent 'Rush', the hybrid selections of Reed ('Reed' and 'Potomac'), Yoder #5, *C. americana* accession CCOR386 from Missouri, the Slate selections, NY 616 and NY 1464, and 'Medium Long', whose origin is unknown but was maintained and described by Slate.

Avellana group

The Avellana group contained a single *C. americana* × *C. avellana* hybrid accession, 'Rutter G081S' and three subgroups. Subgroup 1 contained the 3 accessions obtained as *C. maxima* and 3 *C. avellana* accessions in addition to the *C. colurna* × *C. avellana* hybrid Chinese Trazel J-1 from Oregon. Subgroup 2 was close to Subgroup 1 and contained three *C. colurna* × *C. avellana* hybrids: 'Dundee' and USOR 13-71 from Oregon, and 'Turkish Trazel Gellatly #15' from British Columbia. Subgroup 3 contained the remaining *C. colurna* × *C. avellana* hybrids from British Columbia (Chinese Trazels Gellatly #6 and #11, and 'Faroka', and three selections of Cecil Farris ('Grand Traverse', 88BS and 'Lisa'), which are descended from 'Faroka'.

Structure analysis

We evaluated population structure and differentiation in 109 *Corylus* accessions chosen to represent distinct species and 44 hybrid accessions (153 in total) with a Bayesian Markov Chain Monte Carlo approach implemented in Structure 2.1 (Pritchard et al. 2000). This approach is well-suited for outcrossing taxa like hazelnuts and minimizes deviations from Hardy–Weinberg equilibrium within an inferred population. The analyses using Structure with the species-only dataset produced a clear 'plateau' in the estimated log probability of data Pr(X/K) between $k = 9$ (−1,756.43 on average) and $k = 10$ (−1,741.23 on average) and increased after $k = 11$ (−1,766.13 on average). Therefore we chose $k = 9$ (Fig. 3) based on the ad hoc ln Pr(X|K) method (Pritchard et al. 2000), which recommends picking the smallest value of *K* that captures the major structure of the data. However, when the hybrid accessions were included in the dataset, log probability of data Pr(X/K) did not reach a plateau even at $k = 11$, so we elected to describe population differentiation in the data only from distinct species. However, it is interesting to note that in the Structure analysis of the full data set, unlike the species-only data set, *C. colurna* × *C. avellana* hybrids formed a distinct group at $k = 9$, before *C. ferox* accessions which were differentiated at $k = 10$. In the species-only data set, at $k = 2$, the hazelnut accessions split into two groups, the *C. cornuta* complex + *C. ferox* group versus all other *Corylus* species. At $k = 3$, *C. americana* accessions separated from the mixed species group. At $k = 4$, *C. avellana* accessions formed a distinct group. At $k = 5$, *C. californica* accessions differentiated into a distinct group. At $k = 6$, *C. jacquemontii* accessions formed a distinct group, while at $k = 7$, *C. chinensis* formed a distinct cluster. At $k = 8$, *C. colurna* accessions and *C. heterophylla* accessions were clearly differentiated. Finally, at $k = 9$, the two *C. ferox* accessions were differentiated into a single cluster. The *C. fargesii* accessions had the highest average ancestry coefficient (defined as the inferred proportion of

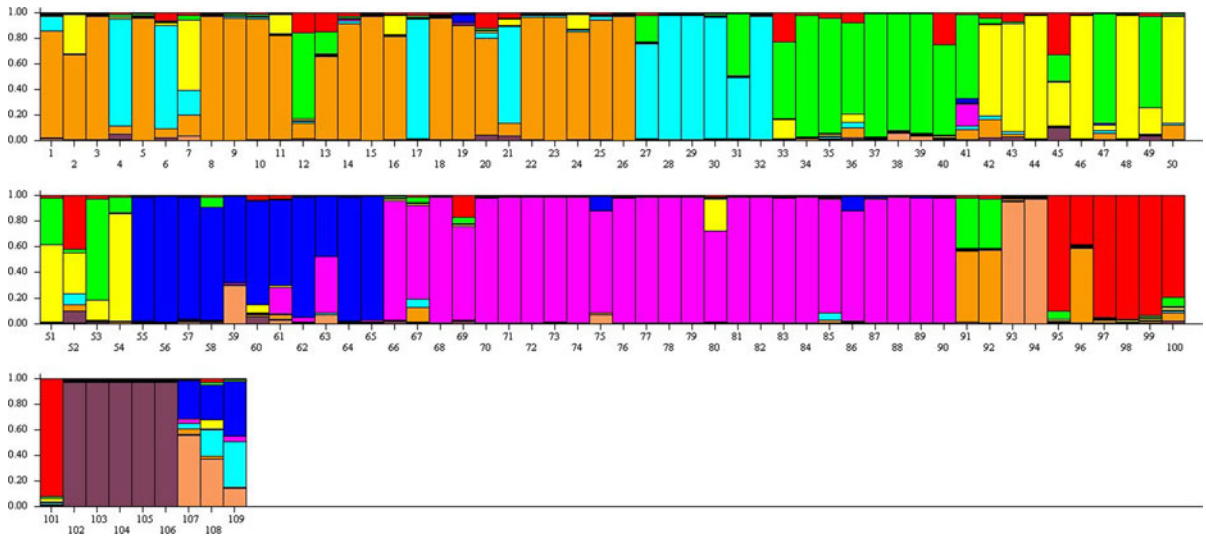

Fig. 3 Assignment of 109 *Corylus* accessions to 9 populations by Structure version 2.3.3. Each *individual bar* represents an accession (see Table 1 for accession information) Numbers 1–26 = *C. americana*, 27–32 = *C. avellana*, 33–41 = *C. chinensis*, 42–54 = *C. colurna*, 55–65 = *C. cornuta*, 66–90 = *C. californica*, 91–92 = *C. fargesii*, 93–94 = *C. ferox*, 95–101 = *C. heterophylla*, 102–106 = *C. jacquemontii*, 107–109 = *C. sieboldiana*. The *Y-axis* displays the estimated membership of each individual in a particular cluster or population

membership in the hazelnut gene pool) from the *C. americana* population (0.56) followed by that from the *C. chinensis* population (0.39) (Fig. 3). *Corylus sieboldiana* accessions had average ancestry coefficients of 0.35 and 0.34 from *C. ferox* and *C. cornuta*, respectively. As K increased, accessions from these two species, *C. fargesii* and *C. sieboldiana*, never differentiated into their respective species populations.

In each of the species groups differentiated by Structure, the highest ancestry coefficient for each accession was from its identified taxon, except for some accessions of *C. americana* and *C. colurna* and one accession of *C. heterophylla*. *Corylus americana* accessions CCOR180, CCOR685, CCOR694 (4, 17 and 21, respectively in Fig. 3) had the highest average ancestry coefficient from *C. avellana*. These results agree with those obtained from NJ cluster analysis, where these three *C. americana* accessions, along with 'Rush' (7 in Fig. 3), whose highest ancestry coefficient was from the *C. colurna* gene pool (0.567), followed by *C. avellana* (0.226), were found in the Americana-Rush cluster (Fig. 2). CCOR228 (6 in Fig. 3) also had the highest ancestry coefficient from *C. avellana* (0.8) and was not found in the major *C. americana* only clusters of the NJ dendogram. Instead, it grouped with *C. avellana* hybrid accessions in the Hyb 1 cluster. The highest ancestry coefficient in CCOR679 (12 in Fig. 3), the only accession from West Virginia, was from *C. chinensis* (0.675) indicating its divergence from other tested representatives of the *C. americana* gene pool. One (*C. colurna* 97098, 47 in Fig. 3) out of the three *C. colurna* accessions (97100, CCOR452 = 49 and 53, respectively in Fig. 3) that had the highest ancestry coefficient from the *C. chinensis* pool grouped with *C. chinensis* accession in the NJ cluster dendrogram (Fig. 2). Both of the *C. colurna* accessions that had the second highest ancestry coefficient from the *C. americana* pool (97093 and LB1_26, 42, and 50, respectively, in Fig. 3) grouped with *C. americana* accessions in the Americana cluster (Fig. 2), as did the sole *C. heterophylla* accession (CCOR124, 96 in Fig. 3) that had the highest ancestry coefficient from the *C. americana* population.

Chloroplast haplotype determination

Preliminary analysis of 40 *Corylus* accessions at 10 cpSSR loci identified polymorphism in six loci. Locus ccmp10 showed four size variants. Three variants were found at loci ccmp2, ccmp3, ccmp4, and ccmp5, while two variants were observed at locus ccmp6. Alleles differed by increments of 1 bp, varying in their

Table 3 Chlorotypes and allelic diversity at 6 cpSSR loci in 114 *Corylus* species individuals and 37 additional *C. avellana* accessions previously characterized by Boccacci and Botta (2009)

Chlorotype	ccmp2	ccmp3	ccmp4	ccmp5	ccmp6	ccmp10	N. individuals[a]
A	212	118	116	107	98	107	26
B	212	117	116	107	98	107	8
C	213	117	116	107	98	107	8
D	214	118	115	107	98	106	2
E	214	117	115	107	98	106	12
F	213	117	115	106	98	106	1
G	212	117	116	106	98	109	5
H	212	117	116	106	98	107	2
I	213	117	116	108	98	106	3
J	213	117	116	106	98	106	4
K	213	117	117	108	98	106	2
L	213	118	116	107	98	107	1
M	213	118	116	108	98	107	1
N	213	117	116	107	98	106	9
O	212	117	115	108	98	106	1
P	213	116	116	106	99	108	30
Q	212	116	116	106	99	108	23
R	212	116	115	106	99	109	5
S	212	116	115	106	99	108	2
T	212	116	115	107	99	108	2
U	213	116	115	106	99	108	4
Number of alleles	3	3	3	3	2	4	
Gene diversity	0.576	0.636	0.330	0.542	0.493	0.688	

[a] N. individuals did not include any of the hybrids

number of A or T residues within mononucleotide repeats. Ccmp2, ccmp3, ccmp4, and ccmp10 loci were previously found to be polymorphic in 26 European natural hazelnut populations (Palmé and Vendramin 2002) and 75 *C. avellana* cultivars (Boccacci and Botta 2009), but ccmp5 and ccmp6 revealed polymorphism only in this work and in other species. This set of 6 cpSSR loci was then used to assess genetic variability in the *Corylus* complex. Of the remaining four loci, ccmp1 (129 bp) and ccmp7 (153 bp) were monomorphic, ccmp8 showed a very low PCR amplification level, and ccmp9 gave no amplification products. Since the chloroplast genome is inherited maternally in hazelnut (Malusà 1994), results were used to verify which *Corylus* species (known or hypothesized) was the female parent of each hybrid or to identify possible mistakes (Table 1).

Allelic diversity and informativeness of polymorphic chloroplast microsatellites were determined by using the number of alleles (*A*) and the diversity values (H_e) in 114 *Corylus* accessions and 37 cultivars of *C. avellana* previously analyzed by Boccacci and Botta (2009) but excluding the hybrids. *Corylus avellana* is economically the most important species of the genus and is the source of the most important cultivars. This species is very polymorphic based on morphology (Mehlenbacher 1991) and genetic studies (Boccacci and Botta 2010; Gökirmak et al. 2009). Four chlorotypes were observed by Boccacci and Botta (2009) in a previously reported study of 75 *C. avellana* genotypes. Thus, a representative set of hazelnut cultivars from Spain, Italy, Turkey, and Iran (Table 1) were included in our study to help reveal polymorphisms in cpSSR loci and to investigate relationships among the *Corylus* species. Eighteen chlorotypes were observed in the 114 *Corylus* accessions and 44 hybrids (Table 1) based on 6 polymorphic cpSSR loci (ccmp2, ccmp3, ccmp4, ccmp5, ccmp6, and ccmp10). The number of alleles per locus ranged from 2 to 4, with an average of 3. Diversity values ranged from 0.33 to 0.64, with an

average of 0.54 (Table 3). This average value is higher than those reported in rice (Ishii and McCouch 2000) and wheat (Ishii et al. 2001).

After including 37 previously analyzed *C. avellana* cultivars (Boccacci and Botta 2009), the number of detected chlorotypes increased to 21 (Table 3), and most *Corylus* species showed a unique, most frequent haplotype (Table 1). Chlorotypes A, B, C, and D were reported in *C. avellana* by Boccacci and Botta (2009). Of these, chlorotype A was the most frequent and present in all geographical groups. All accessions of *C. colurna* showed chlorotype E with the exception of one individual (CCOR451) that had chlorotype F. A single chlorotype was found in *C. ferox* (H), *C. californica* (P), *C. jacquemontii* (G), and *C. sieboldiana* (N). All but one accession of *C. cornuta* had chlorotype Q. Chlorotype N was observed both in *C. heterophylla* and *C. sieboldiana*, but one individual of *C. heterophylla* showed chlorotype O. Three chlorotypes were observed in *C. chinensis* (I, J, and K) and two in *C. fargesii* (L and M). The most frequent chlorotype (Q) in *C. americana* was also most frequent in *C. cornuta*. However, the *C. americana* accession CCOR679 from West Virginia had a *C. avellana* chlorotype (B). Furthermore, four additional chlorotypes were specific to *C. americana*: S (mostly in Iowa accessions), T, U (only in two Michigan accessions), and V (Table 1).

The phylogenetic relationships among *Corylus* species using cpSSRs were examined in a NJ phylogram (Fig. 4) and an RM network diagram (Fig. 5). In the phylogram, 151 *Corylus* accessions were placed in five main clusters (Fig. 4). The accessions of *C. colurna* were placed in the first cluster with two *C. avellana* cultivars ('Tonda Bianca' and 'Tonda Rossa') from southern Italy. The accessions of *C. chinensis* were placed separately in two subgroups in the second cluster with the *C. heterophylla* and *C. sieboldiana* accessions. The third group included almost all of the *C. avellana* cultivars and the two *C. fargesii* samples. The fourth group consisted of the North American species and the fifth cluster included all accessions of *C. ferox* and *C. jacquemontii* placed in two main clades.

In the reduced median network (Fig. 5), the 21 chlorotypes found in 11 *Corylus* species were placed in three main groups. The first group included the haplotypes observed in *C. heterophylla* and *C. sieboldiana* (N and O) and *C. chinensis* (I, J, and K) from eastern Asia and *C. colurna* (E and F). Moreover, chlorotype E was related to the rare chlorotype D observed in two *C. avellana* cultivars ('Tonda Bianca' and 'Tonda Rossa'). The second cluster included the chlorotypes reported in *C. avellana* (A, B, and C) that were related to the chlorotypes obtained in *C. fargesii*. Chlorotypes H (*C. ferox*) and G (*C. jacquemontii*) were placed in an intermediate position between the second and the third group. The third group comprised the 6 haplotypes observed in the North American species (*C. californica*, *C. cornuta*, and *C. americana*) (Fig. 5).

Discussion

The high cross-amplification of hazelnut microsatellite markers in this study (74–100 %) agrees with previous reports in *Corylus* (Bassil et al. 2005a; Boccacci et al. 2005; Gürcan and Mehlenbacher 2010a). Based on seven trinucleotide SSRs, the average heterozygosity was moderate at 0.49 while allele number, genetic diversity and *PIC* were high (means of 11.71, 0.79 and 0.76, respectively). The diversity parameters were higher than those previously observed for 6 trinucleotide SSRs evaluated in 28 accessions that included seven *Corylus* species (Bassil et al. 2005a). The higher values were expected, as this study included a larger number of species representatives. In fact, for five of the SSRs in common between the two studies (CaC-C003, CaC-C005, CaC-C028, CaC-C111 and CaC-C118) (Bassil et al. 2005a), all of the diversity parameters were higher in this study (Table 2). Based on diversity parameters, trinucleotide motifs have been reported as less informative than the dinucleotide types (Bassil et al. 2005a; Liewlaksaneeyanawin et al. 2004; Stàgel et al. 2008) and are typically associated with a low level of variability. When compared in hazelnut (Bassil et al. 2005a), the number of alleles as well as heterozygosity were lower for trinucleotide SSRs. The moderate heterozygosity and high number of alleles of the seven best trinucleotide SSRs chosen for this study must be viewed as biased, because we chose the best performing trinucleotide SSRs from a larger group.

The amplification and polymorphism rates were not correlated to the distance of each species from *C. avellana* but were definitely limited by the number of accessions representing each species. For example,

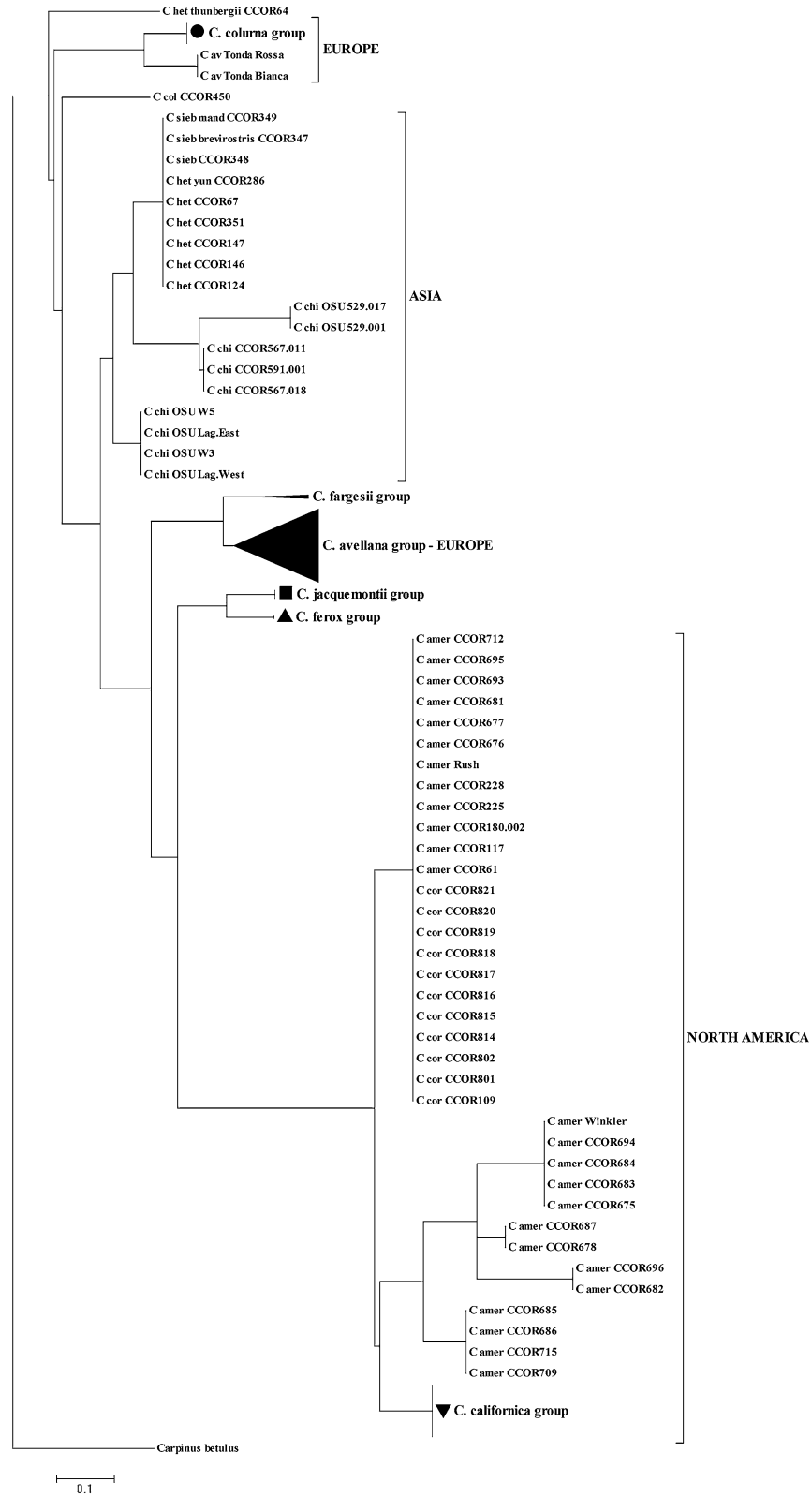

Fig. 4 A NJ tree showing phylogenetic relationships among *Corylus* accessions revealed by 6 cpSSR loci

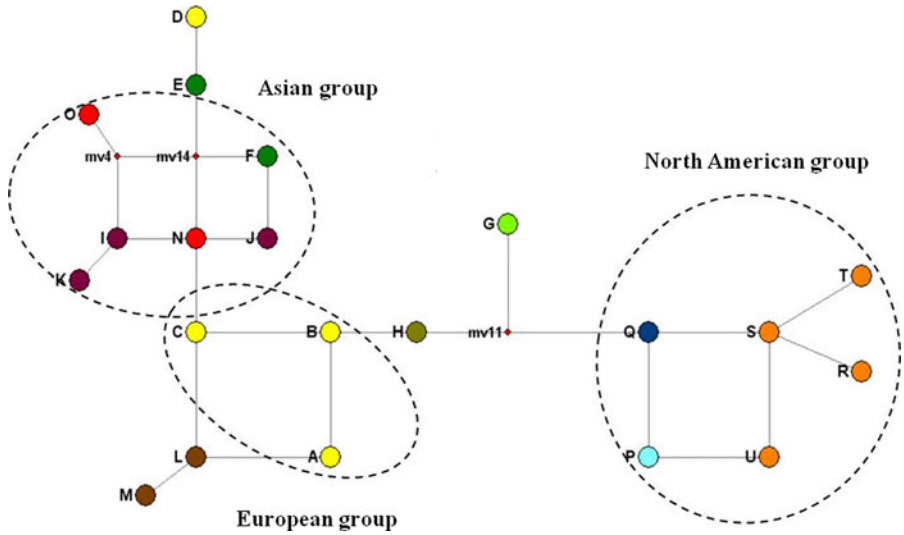

Fig. 5 Reduced median network representing relations of 21 chlorotypes in the *Corylus* complex. Legend: A–D—*C. avellana*; E and F—*C. colurna*; G—*C. jacquemontii*; H—*C. ferox*; I–K—*C. chinensis*; L and M—*C. fargesii*; N and O—*C. heterophylla* and *C. sieboldiana*; P—*C. californica*; Q—*C. cornuta* and *C. americana*; R–U—*C. americana*

a lower rate of amplification (78 %) in *C. ferox* and the lowest rate of polymorphism (41 %) in *C. sieboldiana* are likely the result of the use of few accessions of these species (2 and 3, respectively). Additional examples of east Asian *Corylus* would benefit future studies. Furthermore, our reported levels of polymorphism may be underestimated since polymorphism in all species was initially assessed with the relatively lower resolution 3 % agarose gel electrophoresis technique rather than by capillary electrophoresis. In fact, by using capillary electrophoresis, we found that CaC-C028 and CaC-C003 were polymorphic in *C. avellana* and *C. jacquemontii*, respectively, while four SSR loci (CaC-C005, Cac-C112, CaC-C119 and CaC-C501) were polymorphic in *C. colurna* (Suppl. Table 1).

Despite the small number of nuclear SSRs used in this study (8), nuclear SSR-based clustering mostly agreed with recent taxonomic classifications in hazelnut (Erdoğan and Mehlenbacher 2000a; Forest and Bruneau 2000; Forest et al. 2005; Whitcher and Wen 2001). The bristle-husked shrub species of subsection *Siphonochlamys* (*C. californica*, *C. cornuta* and *C. sieboldiana*) grouped together in the Species clade; as did the *Colurnae* subsection tree species, *C. jacquemontii* (all 5 accessions), most of the *C. colurna* (8 of 13 accessions) and *C. chinensis* (all 9 accessions). However, the two accessions of *C. fargesii* grouped together but were placed in the Americana-Winkler clade. Accessions of other species formed distinct and separate groups: *C. ferox* (n = 2) and *C. heterophylla* (5 of 7). Accessions of *C. avellana* (n = 3) and *C. maxima* (n = 3), grouped together in the dendrogram, supporting their placement in one large, polymorphic species designated *C. avellana*. The sample sizes for each species in this study may be small, but still, our study agrees with previous results (Erdoğan and Mehlenbacher 2000a) and does not support *C. maxima* as a separate taxon. However, our data clearly indicate that *C. californica* is a separate species rather than a botanical variety of *C. cornuta* (Erdoğan and Mehlenbacher 2000a).

The leafy-husked shrub species of the subsection *Phyllochlamys* did not group together, most likely due to the large number of hybrid accessions between *C. americana* and *C. avellana*, or that contained *C. avellana*, included in this study. This is illustrated by clade Americana-Rush, where 'Rush', the *C. americana* selection used in early efforts to breed hazelnuts adapted to the eastern US, grouped with its hybrid offspring 'Buchanan', 'Reed', 'Potomac', and several of the New York selections made by Slate (1947). The diversity among accessions of *C. colurna*, *C. americana*, *americana* × *avellana* hybrids, and *colurna* × *avellana* hybrids is striking, as illustrated by their presence in multiple clades in the dendrogram (Fig. 2). The diversity displayed among *C. americana* accessions and *C. americana* × *C. avellana* hybrids agrees with previous findings (Sathuvalli and Mehlenbacher 2011). Hybrids between *C. colurna* and *C. avellana* were found in the Species, Hybrid1, Americana-Winkler and Avellana clades. Hybrids between

C. americana and *C. avellana* were found in all except the Species clade. *Corylus americana* accessions were found in the many groups of the Americana-Winkler clade and in the Hybrid1 and Americana-Rush clades. Such diversity in *C. americana* and its hybrids may prove useful in the breeding of new hazelnut cultivars adapted to the eastern US (Molnar et al. 2005).

Structure, a Bayesian clustering approach that probabilistically assigns individuals to populations based on genotype, differentiated all species into groups except for *C. fargesii* (n = 2) and *C. sieboldiana* (n = 3). These two species never differentiated into individual populations, which is not surprising given the small number of accessions available for these two species. Assignment of some individuals from *C. americana* and *C. colurna* to multiple populations (Fig. 3) agreed with their placement in the distance-based NJ dendrogram (Fig. 2) and further supports the high diversity of accessions in these species. Still, unexpected clustering of some of the accessions (e.g., *C. americana* CCOR679 from West Virginia, *C. colurna* 97098, 97093 and LB1_26; and *C. heterophylla* CCOR124) is not surprising and resulted from high level of polymorphism within *Corylus* species and the low number of DNA markers used in this study.

The NJ phylogenetic trees produced from nuclear and chloroplast SSR loci did not give congruent topologies (Figs. 2 and 4, respectively). The phylogeny obtained with nSSR markers corresponded fairly well with those based on morphological characteristics or ITS sequences (Erdoğan and Mehlenbacher 2000a; Whitcher and Wen 2001) and on nontranscribed spacer of the 5S rRNA genes (Whitcher and Wen 2001). The classification based on cpSSR markers is not in agreement with the results of commonly accepted taxonomic classifications, as discussed earlier, but closely resembled the findings of Erdoğan and Mehlenbacher (2000a) who compared chloroplast *matK* gene sequences. The cpSSR-based tree separated American, European, and Asian species, in spite of intercontinental morphological similarities among some of these species.

The incongruence between nuclear and chloroplast phylogenetic topologies is typically explained either by lineage sorting or hybridization (Wendel and Doyle 1998). Lineage sorting assumes that there was notable ancestral polymorphism that was rapidly fixed, so that little remains detectable today. The discrepancy in the two topologies could also result from ancient hybridization and subsequent chloroplast capture, so that chloroplast topologies do not accurately reflect organismal relationships. The cpSSR results suggested possible hybridizations among some *Corylus* species that shared the same chlorotype profile: chlorotype N was observed in almost all *C. heterophylla* accessions and in all *C. sieboldiana* individuals; and 12 *C. americana* accessions shared chlorotype Q with *C. cornuta*. Sharing of chlorotypes between two potentially hybridizing species only in areas where they are sympatric would lend support to the local hybridization hypothesis. As reported in Fig. 1, each of these species pairs are sympatric: *C. heterophylla* and *C sieboldiana* are from eastern Asia, and *C. americana* and *C. cornuta* are native to eastern North America. In contrast, we should note that controlled hybridizations among *Corylus* species showed that crosses between *C. heterophylla* and *C. sieboldiana*, and between *C. americana* and *C. cornuta* are very difficult (Erdoğan and Mehlenbacher 2000b). However, chloroplast capture may not be recent and most likely occurred during the ancestral diversification of the genus (Whitcher and Wen 2001). Alternatively the same cpSSR profile observed in these pairs of species could be a consequence of homoplasy (occurrence of alleles identical in state but not identical by descent). We are not aware of reports that evaluated homoplasy in any genus in the Fagales that may allow us to estimate likelihood of homoplasy in *Corylus*. Estimates based on simulations (Navascués and Emerson 2005) were done under specific conditions and tested on *Pinus resinosa* Ait., but cannot be directly transferred to other plant species. Authors have generally considered the level of homoplasy to be low enough to permit plant population genetic analysis (Terrab et al. 2006). Even when homoplasy was identified, it has been considered moderate and its potential for confounding results disregarded (Cuenca et al. 2003). Although the possibility of homoplasy yielding by chance the same haplotype in the mentioned *Corylus* species cannot be excluded without further studies, the combined use of cpSSR and nSSR in this paper can strengthen results and conclusions of the genetic analyses. For *C. maxima* and *C. avellana*, cpSSR data agree with nSSR results, and indicate that *C. maxima* is not a separate taxon.

The RM network based on cpSSR polymorphism enabled the identification of three main chlorotype

lineages (Fig. 5). General distribution of plastid lineages was not fully congruent with present-day taxonomy, but was very similar to the topology of the cpSSR-based NJ tree (Fig. 4). The clear geographical distribution of lineages supported an early differentiation among *Corylus* species from Asia, Europe, and North America with a few exceptions. *Corylus fargesii* (chlorotypes L and M) and *C. jacquemontii* (chlorotype G) did not cluster with other Asian species, while two *C. avellana* accessions (chlorotype D) were closely related to *C. colurna* (chlorotype E) in the Asian lineage. Divergence between the Himalayan *C. jacquemontii* and the other Asian species, particularly the tree species of subsection *Colurnae*, was probably due to the rise of the Himalaya mountains (Whitcher and Wen 2001). *Corylus fargesii* from China, called the paperbark tree hazel, is morphologically distinct from the other tree species in that its bark exfoliates like river birch (*Betula nigra* L.) (Erdoğan and Mehlenbacher 2000a). The PCR–RFLP and SSR data from cpDNA obtained by Palmé and Vendramin (2002) suggested that hybridization could have occurred between *C. colurna* and several wild *C. avellana* individuals. The close relationship between *C. colurna* and two *C. avellana* accessions ('Tonda Bianca' and 'Tonda Rossa') supports this hypothesis. Nevertheless, *C. colurna* is presently found from the Balkans to Asia Minor, while 'Tonda Bianca' and 'Tonda Rossa' are only located in southern Italy. This might seem to argue against hybridization, but chloroplast capture might not have taken place directly and transfer could have occurred via wild and cultivated forms of *C. avellana*, during migrations in the Mediterranean Basin (Boccacci and Botta 2009).

The phylogeographical relationships among the 21 chlorotypes found in 11 *Corylus* species support several biogeographic observations reported in the literature (Chen et al. 1999; Whitcher and Wen 2001). Asia may have served as a refugium where several hazelnut lineages survived during the glaciations and from which they continued to evolve after their dispersal from Asia through the Mediterranean to Europe, and across the Atlantic and/or the Bering land bridge to North America (Whitcher and Wen 2001). The high number of cpSSR haplotypes observed among the Asian species supports this hypothesis, already demonstrated on the basis of morphological, fossil and molecular data (Chen et al. 1999; Whitcher and Wen 2001). In the RM network, the intermediate position of Asian chlorotypes I, J, and K (*C. chinensis*), and N and O (*C. heterophylla* and *C. sieboldiana*) between the European chlorotypes A, B, and C (*C. avellana*), which were associated with the Chinese chlorotypes L and M (*C. fargesii*), also support the migration hypothesis from Asia to the Mediterranean Basin and Europe from local common ancestors (Whitcher and Wen 2001). Moreover, the position of chlorotype Q in the American group, observed both in *C. cornuta* and in several accessions of *C. americana*, supports the hypothesis that long distance migration to North America may have occurred during the late Tertiary both from Asia via the Bering land bridge (*C. cornuta* and *C. californica*) and from Europe via the Atlantic (*C. americana*) (Whitcher and Wen 2001).

Acknowledgments We acknowledge Barbara Gilmore, Christine Neou-Anderson, and April Nyberg for technical assistance in microsatellite marker separation. Funding for this study was provided by the USDA-ARS CRIS 5358-21000-033-00D, a USDA-ARS National Plant Germplasm System Evaluation Grant, and by the Fondazione Cassa di Risparmio di Torino (Italy).

References

Arroyo-García R, Ruiz-García L, Bolling L, Ocete R, López MA, Arnold C, Ergul A, Söylemezoğlu G, Uzun HI, Cabello F, Ibáñez J, Aradhya MK, Atanassov A, Atanassov I, Balint S, Cenis JL, Costantini L, Gorislavets S, Grando MS, Klein BY, McGovern PE, Merdinoglu D, Pejic I, Pelsy F, Primikirios N, Risovannaya V, Roubelakis-Angelakis KA, Snoussi H, Sotiri P, Tamhankar S, This P, Troshin L, Malpica JM, Lefort F, Martínez-Zapater JM (2006) Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *sativa*) based on chloroplast DNA polymorphisms. Mol Ecol 15:3707–3714

Bandelt HJ, Foster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

Bassil NV, Botta R, Mehlenbacher SA (2005a) Microsatellite markers in hazelnut: isolation, characterization and cross-species amplification. J Am Soc Hort Sci 130:543–549

Bassil NV, Botta R, Mehlenbacher SA (2005b) Additional microsatellite markers of the European hazelnut. Acta Hort 686:105–110

Bassil NV, Postman J, Hummer K, Botu M, Sezer A (2009) SSR fingerprinting panel verifies identities of clones in backup hazelnut collection at USDA genebank. Acta Hort 845:95–102

Boccacci P, Botta R (2009) Investigating the origin of hazelnut (*Corylus avellana* L.) cultivars using chloroplast microsatellites. Genet Resour Crop Evol 56:851–859

Boccacci P, Botta R (2010) Microsatellite variability and genetic structure in hazelnut (*Corylus avellana* L.) cultivars from different growing regions. Sci Hortic 124:128–133

Boccacci P, Akkak A, Bassil NV, Mehlenbacher SA, Botta R (2005) Characterization and evaluation of microsatellite loci in european hazelnut (*Corylus avellana* L.) and their transferability to other *Corylus* species. Mol Ecol Notes 5:934–937

Boccacci P, Akkak A, Botta R (2006) DNA-typing and genetic relationships among European hazelnut (*Corylus avellana* L.) cultivars using microsatellite markers. Genome 49: 598–611

Boccacci P, Botta R, Rovira M (2008) Genetic diversity of hazelnut (*Corylus avellana* L.) germplasm in northeastern Spain. HortSci 43:667–672

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Botta R, Akkak A, Boccacci P (2005) DNA-typing of hazelnut: a universal methodology for describing cultivars and evaluating genetic relatedness. Acta Hort 686:117–124

Chen ZD, Manchester SR, Sun HY (1999) Phylogeny and evolution of the Betulaceae as inferred from DNA sequences, morphology, and paleobotany. Am J Bot 86:1168–1181

Crane PR (1989) Early fossil history and evolution of the Betulaceae. In: Crane PR, Blackmore S (eds) Evolution, systematics and fossil history of the Hamamelidae, vol 2, 'Higher' Hamame- lidae. Clarendon Press, Oxford, pp 87–116

Cronquist A (1981) An integrated system of classification of flowering plants. Columbia University Press, London

Cuenca A, Escalante AE, Piñero D (2003) Long-distance colonization, isolation by distance, and historical demography in a relictual Mexican pinyon pine (*Pinus nelsonii* Shaw) as revealed by paternally inherited genetic markers (cpSSRs) M. Mol Ecol 12:2087–2097

De Candolle A (1864) *Corylus*. In: Prodromus systemnatis naturalis regni vegetabilis, vol 16, part 2. Treuttel & Wurtz, Paris, pp 128–133

Erdoğan V, Mehlenbacher SA (2000a) Phylogenetic relationships of *Corylus* species (Betulaceae) based on nuclear ribosomal DNA ITS region and chloroplast matK gene sequences. Syst Bot 25:727

Erdoğan V, Mehlenbacher SA (2000b) Interspecific hybridization in hazelnut. J Am Soc Hort Sci 125(4):489–497

Forest F, Bruneau A (2000) Phylogenetic analysis, organization and molecular evolution of the nontranscribed spacer of 5S ribosomal RNA genes in *Corylus* (Betulaceae). Int J Plant Sci 161:793–806

Forest F, Savolainen V, Chase MW, Lupia R, Bruneau A, Crane PR (2005) Teasing apart molecular- versus fossil-based error estimates when dating phylogenetic trees: a case study in the Birch family (Betulaceae). Syst Bot 30:118–133

Gastier JM, Pulido JC, Sunden S, Brody T, Buetow KH, Murray JC, Weber JL, Hudson TJ, Sheffield VC, Duyk GM (1995) Survey of trinucleotide repeats in the human genome: assessment of their utility as genetic markers. Hum Mol Genet 4:1829–1836

Ghanbari A, Akkak A, Boccacci P, Talaie A, Vezvaie A, Botta A (2005) Characterization of hazelnut (*Corylus avellana* L.) cultivars using microsatellite markers. Acta Hort 686:111–115

Gökirmak T, Mehlenbacher SA, Bassil NV (2009) Characterization of European hazelnut (*Corylus avellana*) cultivars using SSR markers. Genet Resour Crop Evol 56:147–172

Gürcan K, Mehlenbacher SA (2010a) Transferability of microsatellite markers in the Betulaceae. J Am Soc Hort Sci 135(2):159–173

Gürcan K, Mehlenbacher SA (2010b) Development of microsatellite marker loci for European hazelnut (*Corylus avellana* L.) from ISSR fragments. Mol Breed 26:551–559

Gürcan K, Mehlenbacher SA, Bassil NV, Boccacci P, Akkak A, Botta R (2010a) New microsatellite markers for *Corylus avellana* from enriched libraries. Tree Genet Gen 6:513–531

Gürcan K, Mehlenbacher SA, Erdoğan V (2010b) Genetic diversity in hazelnut cultivars from Black Sea countries assessed using SSR markers. Plant Breed 129:422–434. doi:10.1111/j.1439-0523.2009.01753.x

Ishii T, McCouch SR (2000) Microsatellites and microsynteny in the chloroplast genomes of *Oryza* and eight other Graminae species. Theor Appl Genet 100:1257–1266

Ishii T, Mori N, Ogihara Y (2001) Evaluation of allelic diversity at chloroplast microsatellite loci among common wheat and its ancestral species. Theor Appl Genet 103:896–904

Kasapligil B (1972) A bibliography on *Corylus* (Betulaceae) with annotations. Annu Rpt Northern Nut Growers Assn 63:107–162

Kutil BL, Williams CJ (2001) Triplet repeat microsatellites shared among hard and soft pines. J Heredity 92:327–332

Li PC, Cheng SX (1979) Betulaceae. In: Kuang K-Z, Li P-C (eds) Flora republicae popularis sinicae, vol 21. Science Press, Beijing, pp 44–137 (In Chinese)

Liang WJ, Zhang YM (1988) Investigation and study of filbert resources in China. In: Proceedings of the international symposium on horticultural germplasm, Cultivated and Wild. Beijing, China. 5–9 Sept. 1988

Liewlaksaneeyanawin C, Ritland CE, El-Kassaby YA, Ritland K (2004) Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. Theor Appl Genet 109:361–369

Liu K, Muse SV (2005) Powermarker: integrated analysis environment for genetic marker data. Bioinformatics 21:2128–2129

Malusà E (1994) Interspecific relationships among *Corylus* species. Acta Hort 51:335–340

Mehlenbacher SA (1991) Hazelnuts (*Corylus*). Genetic resources of temperate fruit and nut crops. Acta Hort 290:791–836

Mehlenbacher SA (2009) Genetic resources for hazelnut: state of the art and future perspectives. Acta Hort 845:33–38

Mehlenbacher SA, Brown RN, Nouhra ER, Gökirmak T, Bassil NV, Kubisiak TL (2006) A genetic linkage map for hazelnut (*Corylus avellana* L.) based on RAPD and SSR markers. Genome 49:122–133

Minch E (1997) MICROSAT version 1.5b. Stanford University Medical Center, Stanford, CA. http://hpgl.stanford.edu/projects/microsat/ Accessed 02 January 2012

Molnar TJ, Goffreda JC, Funk CR (2005) Developing hazelnuts for the eastern United States. Acta Hort 68:609–617

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30:194–200

Navascués M, Emerson BC (2005) Chloroplast microsatellites: measures of genetic diversity and the effect of homoplasy. Mol Ecol 14:1333–1341

Nei M (1987) Molecular evolutionay genetics. Columbia University Press, New York

Palmé AE, Vendramin GG (2002) Chloroplast DNA variation, postglacial recolonization and hybridization in hazel, *Corylus avellana*. Mol Ecol 11:1769–1779

Petit RJ, Aguinagalde I, de Beaulieu JL, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Müller-Starck G, Demesure-Musch B, Palmé A, Martín JP, Rendell S, Vendramin GG (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. Science 300:1563–1565

Pigg KB, Manchester SR, Wehr WC (2003) *Corylus*, *Carpinus*, and *Palaeocarpinus* (Betulaceae) from the middle Eocene Klondike Mountain and Allenby Formations of northwestern North America. Int J Plant Sci 164:807–822

Pritchard J, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Provan J, Powell W, Hollingsworth PH (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends Ecol Evol 16:142–147

Rajora OP, Rahman MH, Dayanandan S, Mosseler A (2001) Isolation, characterization, inheritance and linkage of microsatellite DNA markers in white spruce (*Picea glauca*) and their usefulness in other spruce species. Mol Gen Genet 264:871–882

Rovira M (1997) Genetic variability among hazelnut (*C. avellana* L.) cultivars. Acta Hort 445:45–50

Sathuvalli SR, Mehlenbacher SA (2011) Characterization of American hazelnut (*Corylus americana*) accessions and *Corylus americana* × *Corylus avellana* hybrids using microsatellite markers. Genet Resour Crop Evol. doi: 10.1007/s10722-011-9743-0

Schneider C (1916) Betulaceae. In: ed., Sargent CS (ed) Plantae wilsonianae: an enumeration of the woody plants collected in western China for the Arnold Arboretum of Harvard University during the years 1907, 1908, and 1910, vol. 2. Publications of the Arnold Arboretum, no. 4, pp 423–508

Scotti I, Magni F, Fink R, Powell W, Binnelli G, Hedley PE (2000) Microsatellite repeats are not randomly distributed within Norway spruce (*Picea abies* L.) expressed sequences. Genome 43:41–46

Sheffield VC, Weber JL, Buetow KH, Murray JC, Even DA, Wiles K, Gastier JM, Pulido JC, Yandava C, Sunden SL et al (1995) A collection of tri- and tetra-nucleotide repeat markers used to generate high quality, high resolution human genome-wide linkage maps. Hum Mol Genet 4:1837–1844

Shepherd M, Cross M, Maguire TL, Dieters MJ, Williams CG, Henry RJ (2002) Transpecific microsatellites for hard pines. Theor Appl Genet 104:819–827

Slate GL (1947) Some results with filbert breeding at Geneva, New York. Annu Rep North Nut Grow Assoc 38:94–100

Stàgel A, Portis E, Toppino L, Rotino GL, Lanteri S (2008) Gene-based microsatellite development for mapping and phylogeny studies in eggplant. BMC Genomics 9:357–370

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. doi: 10.1093/molbev/msr121

Terrab A, Paun O, Talavera S, Tremetsberger K, Arista MF, Stuessy TF (2006) Genetic diversity and population structure in natural populations of Moroccan Atlas cedar (*Cedrus atlantica*, *Pinaceae*) determined with cpSSR markers. Am J Bot 93(9):1274–1280

Testolin R, Cipriani G (2010) Molecular markers for germplasm identification and characterization. Acta Hort 859:59–72

The Plant List (2010) Version 1. Published on the Internet; http://www.theplantlist.org/. Accessed April 23, 2012

Thompson MM, Lagerstedt HB, Mehlenbacher SA (1996) Hazelnuts. In: Janick J, Moore JN (eds) Fruit breeding: nuts, vol 3. Wiley, New York, pp 125–184

Tozaki T, Inoue S, Mashima S, Ohta M, Miura N, Tomita M (2000) Sequence analysis of trinucleotide repeat microsatellites from an enrichment library of the equine genome. Genome 43:354–365

Wang Z, Weber JL, Zhong G, Tanksley SD (1994) Survey of plant short tandem DNA repeats. Theor Appl Genet 88:1–6

Weising K, Gardner R (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome 42:9–19

Wendel JF, Doyle JJ (1998) Phylogenetic incongruence: window into genome history and molecular evolution. In: Soltis D, Soltis P, Doyle J (eds) Molecular systematics of plants, 2nd edn. Chapman & Hall, New York

Whitcher IN, Wen J (2001) Phylogeny and biogeography of *Corylus* (Betulaceae): inference from ITS sequences. Syst Bot 26:283–298

Yoo K-O, Wen J (2002) Phylogeny and biogeography of *Carpinus* and subfamily Coryloideae (Betulaceae). Int J Plant Sci 163:641–650

Yoo K-O, Wen J (2007) Phylogeny of *Carpinus* and subfamily Coryloideae (Betulaceae) based on chloroplast and nuclear ribosomal sequence data. Plant Syst Evol 267:25–35

Young ET, Sloan JS, Van Riper K (2000) Trinucleotide repeats are clustered in regulatory genes in *Saccaromyces cerevisae*. Genetics 154:1053–1068