Genetic Diversity, Structure and Differentiation in Cultivated Walnut (*Juglans regia* L.)

M. Aradhya[1], K. Woeste[2] and D. Velasco[1]
[1] National Clonal Germplasm Repository, USDA-ARS, University of California, Davis, CA 95616, USA
[2] Hardwood Tree Improvement Center, U.S. Forest Service, Purdue University, West Lafayette, IN 47907, USA

Keywords: fixation index, gene diversity, heterozygosity, *Juglans regia*, molecular characterization, walnut

Abstract

An analysis of genetic structure and differentiation in cultivated walnut (*Juglans regia*) using 15 microsatellite loci revealed a considerable amount of genetic variation with a mild genetic structure indicating five genetic groups corresponding to the centers of diversity within the home range of walnut in Eurasia. Despite the narrow genetic differentiation among groups accounting for only 10 to 15% of the total variation, the groups differed significantly with respect to frequency and composition of alleles for different loci. Moderate to high genetic variability with mild genetic structure is found to be ideal for association genetic analysis.

INTRODUCTION

The thin-shelled cultivated walnut (*Juglans regia*) belongs to the section *Juglans* within the genus *Juglans* of the family Juglandaceae. Its native range extends from the Carpathian Mountains of Eastern Europe to the Southern Caucasus, northern Turkey, Iran, to the Tien Shan province of western China to the Himalayan states of India, Sikkim, and Bhutan (Zohary and Hopf, 1993). The taxonomic placement of the cultivated walnut within the genus *Juglans* was problematic and the earlier studies placed it as sister either to the Asian butternuts section, Cardiocaryon (Stanford et al., 2000) or to the black walnut section, Rhysocaryon (Manos and Stone, 2001). But a recent study based on sequences from the chloroplast non-coding regions strongly supports the section *Juglans* as an independent clade sister to the remaining three sections within the genus *Juglans* (Aradhya et al., 2004). The evolutionary history of the section *Juglans* is riddled with widespread extinctions, geographic isolations, and bottlenecks during the Quaternary glaciations. Subsequent expansion and human selection in the Transcaucasia, Central, West and East Asia greatly influenced the genetic structure within the section *Juglans* (Popov, 1929; Beug, 1975). The modern distribution of Walnut (*J. regia*) extends beyond its native range occurring under cultivation in both the Old and New World. Its sister taxon, *J. sigillata* with a hard shell bearing a black kernel may represent a semi-domesticated or primitive form within the section *Juglans* restricted to southern China.

Knowledge of genetic diversity, structure and differentiation of cultivated walnut is important for effective conservation, management and utilization of germplasm. Several studies have examined the genetic diversity and relationships among walnut cultivars using allozymes (Arulsekar et al., 1985), restriction fragment length polymorphisms (RFLP; Fjellstrom et al., 1994), randomly amplified polymorphic DNA (RAPD; Nicese et al., 1998), and microsatellite markers (Dangl et al., 2005). Genetic analysis of germplasm collections often provides insights into the complex interactions of evolutionary forces such as mutation, gene flow, selection, and drift shaping the ecogeographic structure and domestication history of a crop species. In outcrossing species such as walnut information on genetic structure and co-ancestral relationships among germplasm accessions have proven to be useful in association genetic analysis. The present study represents a first step towards association or disequilibrium mapping of genes. Here we report results of a preliminary analysis of genetic structure and differentiation in a walnut germplasm collection maintained at the USDA Germplasm

Repository at the University of California, Davis, California, USA based on genetic polymorphism at microsatellite loci.

MATERIALS AND METHODS

459 trees representing 203 diverse accessions of walnut (*J. regia*) germplasm conserved at the USDA repository were genotyped using 15 microsatellite (also known as SSR) loci *WGA001*, *WGA004*, *WGA009*, *WGA069*, *WGA089*, *WGA106*, *WGA118*, *WGA178*, *WGA202*, *WGA237*, *WGA318*, *WGA321*, *WGA331*, *WGA338* and *WGA384* using standard PCR protocols with florescent labeled primers (Dangl et al., 2005). The microsatellite loci were originally developed for *J. nigra* at the Hardwood Tree Improvement Center, U.S. Forest Service, Purdue University, Indiana, USA (Woeste et al., 2002) and adapted here to *J. regia*. Amplified products were resolved using capillary electrophoresis in an ABI Prism 3100 genetic analyzer with data collection software, version 1.2 (PE/Applied Biosystems). The fragment data were further analyzed using Genescan, version 3.1 and Genotyper, version 2.5 to assess the size of alleles and data assembled as microsatellite genotypes as well as in binary format.

The binary data were used to compute a distance matrix using Nei and Li distance (Nei and Li, 1979) based on the proportion of alleles shared between two accessions for all possible pair-wise combinations. The resultant distance matrix was subjected to a cluster analyses (CA) using the neighbor-joining method to produce a phenogram. The multilocus SSR genotype data were pooled into groups based on the results of CA and analyzed for various within-group genetic variability measures such as mean number of alleles per locus and observed and expected levels of heterozygosities. Contingency χ^2 analysis was performed to determine the genetic heterogeneity among groups. Genetic differentiation within and among groups was computed using the Nei's gene diversity analysis (Nei, 1973). Total gene diversity (H_T) was partitioned into gene diversity within groups (H_G) and gene diversity between groups (D_{GT}), where $H_T = H_G + D_{GT}$. Genetic differentiation between groups is calculated as $G_{GT} = D_{GT}/H_T$, where G_{GT} varies from zero (when $H_G = H_T$) and unity (when $H_G = 0$), i.e., groups fixed for different alleles. Analysis of molecular variance (AMOVA; Excoffier et al., 1992) was performed on group-wise genotypic data to partition the total variance into variance within and among groups.

RESULTS AND DISCUSSION

Although the origin of walnut is obscure, it is considered to be native to the region extending from the Carpathian Mountains to Transcaucasia and parts of West Asia, East Asia into the Himalayan regions comprising Jammu and Kashmir, Himachal Pradesh, and North Eastern regions of India, Sikkim and Bhutan (Dode, 1909; McGranahan and Leslie, 1991). The species went through a series of bottlenecks during the Quaternary glaciations in isolated cryptic refugia in Carpathian, Ponto-Caspian and other central Asian regions rapidly eroding the diversity. Climatic deterioration and human activity during the postglacial range expansion and colonization of new areas by small founder populations have rapidly modified the genetic structure of the species. The germplasm collection assayed in this study of cultivated walnut is somewhat reminiscent of the evolutionary and domestication history of walnut.

Genetic diversity and the patterns of distribution within a species germplasm collection determine the potential for improving the species through breeding programs. The walnut collection assayed showed considerable variability with the observed number of alleles per locus ranging from 5 for *WGA384* to 19 for *WGA202* with an average of 11 alleles per locus (Table 1). There was a significant deficiency of heterozygotes compared to Hardy-Weinberg proportions for 14 out of the 15 loci assayed suggesting substructuring within the collection and obviously indicating some level of inbreeding with restricted gene flow among subpopulations. The CA identified 5 broad groups corresponding to the major areas of distribution of walnut in its native central, west and East Asia (Fig. 1). Further analysis of the affinities among the groups indicated that the West Asia walnuts are the most diverse within which there is a subgroup closely allied

with East Asian walnuts and a second one with the derived group containing breeders' selections and some of the elite breeding lines, whereas the Carpathian walnuts are somewhat intermediate between the East and West Asian groups. Again there was deficiency of heterozygotes in all regions except for the Carpathian group indicating existence of further subpopulations within each of these areas. However, the groups did not differ with respect to the mean number of alleles per locus, but did differ for levels of observed and expected levels of heterozygosity (Table 2). Surprisingly, the derived group representing the breeders' selections and other elite germplasm also showed significant deficiency of heterozygotes. The overall within and among group genetic variability measures observed in the present study correspond well with the out crossing mode of pollination of the species (Loveless and Hamrick, 1984), except for significant deficiency of heterozygotes for most loci at both within group and total collection levels.

Out crossing species are generally composed of many local populations, the genetic integrity of which is maintained by complex interactions of evolutionary forces spatially and temporally within the range of the species. Comprehensive germplasm collections of such species should permit us to assess the amount and pattern of distribution of genetic variation and estimate the role of evolutionary forces that shape the overall genetic structure of a species. Little genetic differentiation has occurred at the molecular level among the 5 groups identified based on CA with nearly 90% of the total genetic variation residing within groups. However, the contingency chi-square analysis suggested that the groups differed significantly for frequency and composition of alleles indicating significant differentiation among groups (Table 3). The orthogonal partitioning of molecular variation using AMOVA confirms the gene diversity analysis with 86% of the total variation accounted for within and 14% among, groups (Table 4). The above results suggest that although the gene diversity analysis partitions allele frequency variation for different loci within and among groups, it does not reflect the differences among populations with respect to allelic composition for different loci.

Overall, the walnut gene pool representing the native range of distribution assayed in the present study contains moderate to high variability for the microsatellite loci examined. 5 genetic groups were recognized within the collection based on neighbor-joining cluster analysis representing distinct walnut centers of diversity in Eurasia. There was evidence for marginal differentiation among the 5 groups identified based on the CA, but they differed significantly with respect to frequency and composition of alleles. Nevertheless, the cultivated walnut germplasm with a mild genetic structure will probably permit an efficient association genetic analysis.

ACKNOWLEDGEMENTS

This study was funded by the California Walnut Board and the University of California Discovery Grants.

Literature Cited

Aradhya, M.K., Potter, D., Gao, F. and Simon, C.J. 2007. Molecular phylogeny of *Juglans* (Juglandaceae): A biogeographic perspective. Tree Genetics and Genomes 3:363-378.

Arulsekar, S., Parfitt, D.E. and McGranahan, G.H. 1985. Isozyme gene markers in *Juglans* species. J. Hered. 76: 103-106.

Beug, H.J. 1975. Man as a factor in the vegetational history of the Balkan Peninsula. p.72-78. In D. Jordanov, I. Bondev, S. Kozuharov, B. Kuzmanov and E. Palamarev (eds.), Problems of Balkan Flora and Vegetation. Proc. First Int. Symp. On Balken Flora and Vegetation, Varna, June 7-14, 1973. Publishing House of the Bulgarian Academy of Sciences. Sofia, Bulgaria.

Dangl, G.S., Woeste, K., Aradhya, M.K., Koehmstedt, A., Simon, C., Potter, D., Leslie, C.A. and McGranahan, G. 2005. Characterization of 14 microsatellite markers for genetic analysis and cultivar identification of walnut. J. Amer. Soc. Hort. Sci. 130:348-354.

Dode, L.A. 1909. Contribution to the study of the genus *Juglans* (English translation by R.E. Cuendett). Bulletin of the Society of Dendrology, France 11:22-90.
Excoffier, L., Smouse, P.E. and Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479-491.
Fjellstrom, R.G. and Parfitt, D.E. 1994. Walnut (*Juglans* spp.) genetic diversity determined by restriction fragment length polymorphism. Genome 37:690-700.
Loveless, M.D. and Hamrick, J.L. 1984. Ecological determinants of genetic structure in plant populations. Ann. Rev. Ecol. Syst. 15:65-95.
Manos, P.S. and Stone, D.E. 2001. Evolution, phylogeny, and systematics of the Juglandaceae. Ann. Mo. Bot. Gard. 88:231-269.
McGranahan, G. and Leslie, C. 1991. Walnuts (*Juglans*). p.907-951. In: J.N. Moore and J.R. Ballington Jr. (eds.), Genetic Resources of Temperate Fruit and Nut Crops. International Society for Horticultural Science, Wageningen.
Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci., USA 70:3321-3323.
Nei, M. and Li, W. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci., USA. 76:5269-5273.
Nicese, F.P., Harmoza, J.I. and McGranahan, G.H. 1998. Molecular characterization and genetic relatedness among walnut (*Juglans regia*) genotypes based on RAPD markers. Euphytica 101:199-206.
Popov, M.G. 1929. Wild growing fruit trees and shrubs of Asia Minor (in Russian). Bull. Appl. Bot. Pl. Breed. 22:241-483.
Stanford, A.M., Harden, R. and Parks, C.R. 2000. Phylogeny and biogeography of *Juglans* (Juglandaceae) based on *mat*K and ITS sequence data. Am. J. Bot. 87:872-882.
Woeste, K., Bruns, R., Rhodes, O. and Michler, C. 2002. Thirty polymorphic nuclear microsatellite loci from black walnut. J. Hered. 93:58-60.
Zohary, D. and Hopf, M. 1993. Domestication of plants in the Old World. Clarendon Press, Oxford.

Tables

Table 1. Locus-wise genetic variability in the walnut gene pool assayed.

Locus	*A*	*N*	$H_{(O)}$	$H_{(E)}$	Mean (H)	*F*	Sig.
WGA001	12	840	0.612	0.808	0.740	0.241	***
WGA202	19	828	0.614	0.836	0.800	0.265	***
WGA384	5	810	0.467	0.579	0.482	0.193	***
WGA321	12	842	0.594	0.723	0.659	0.177	***
WGA331	6	834	0.590	0.659	0.628	0.104	***
WGA009	11	794	0.655	0.778	0.724	0.158	***
WGA118	14	836	0.644	0.795	0.717	0.190	***
WGA004	12	832	0.596	0.685	0.630	0.128	***
WGA069	13	840	0.562	0.817	0.758	0.312	***
WGA089	8	842	0.489	0.666	0.587	0.265	***
WGA338	7	842	0.518	0.557	0.522	0.069	NS
WGA178	11	840	0.676	0.747	0.696	0.094	*
WGA318	14	752	0.354	0.822	0.689	0.569	***
WGA106	7	840	0.331	0.423	0.417	0.216	***
WGA237	8	838	0.339	0.593	0.515	0.428	***
Mean	11	827	0.536	0.699	0.638	0.227	***

A = Number of alleles; N = Number of individuals; H(o) & H(E) = Observed and expected levels of heterozygosity; F = Fixation index; *** = P<0.001

Table 2. Within group genetic variability estimates (±Standard Error).

Group	*N*	*A*	$H_{(O)}$	$H_{(E)}$	*F*
East Asia	131	5.4	0.486±0.7	0.571±0.037	0.149±0.046***
W/S Asia 1	119	9.6	0.591±1	0.708±0.04	0.165±0.035***
Carpathian	26	5.1	0.634±0.5	0.668±0.035	0.051±0.035
W/S/ Asia 2	30	6.9	0.625±0.7	0.708±0.036	0.117±0.031***
Derived	108	6.4	0.488±0.6	0.566±0.046	0.138±0.042***

A = Number of alleles; *N* = Number of individuals; $H_{(o)}$ & $H_{(E)}$ = Observed and expected levels of heterozygosity; W/S = West and South Asia; *F* = Fixation index; *** = $P < 0.001$

Table 3. Genetic diversity and differentiation in walnut.

Locus	H_S	H_T	D_{ST}	G_{ST}
WGA001	0.740	0.811	0.071	0.087
WGA202	0.800	0.857	0.057	0.067
WGA384	0.482	0.580	0.098	0.168
WGA321	0.659	0.756	0.097	0.129
WGA331	0.628	0.674	0.046	0.068
WGA009	0.724	0.787	0.063	0.080
WGA118	0.717	0.810	0.094	0.116
WGA004	0.630	0.670	0.040	0.060
WGA069	0.758	0.807	0.049	0.061
WGA089	0.587	0.687	0.100	0.146
WGA338	0.522	0.548	0.026	0.047
WGA178	0.696	0.737	0.041	0.056
WGA318	0.689	0.824	0.135	0.164
WGA106	0.417	0.457	0.039	0.086
WGA237	0.515	0.593	0.078	0.131
Total	0.638	0.707	0.069	0.098

Hs, H_T and D_{ST}= Gene diversity within, total, and between groups, respectively; G_{ST} = proportion of gene diversity due to genetic differentiation among groups.

Table 4. Partitioning of molecular variation within and among groups.

Source	SS	MS	% variation	F_{ST}
Among groups	476.65	0.756	13.904	0.139***
Within groups	3850.30	4.684	86.095	
Total	4326.90	5.441		

F_{ST} = Genetic differentiation among groups.

Figures

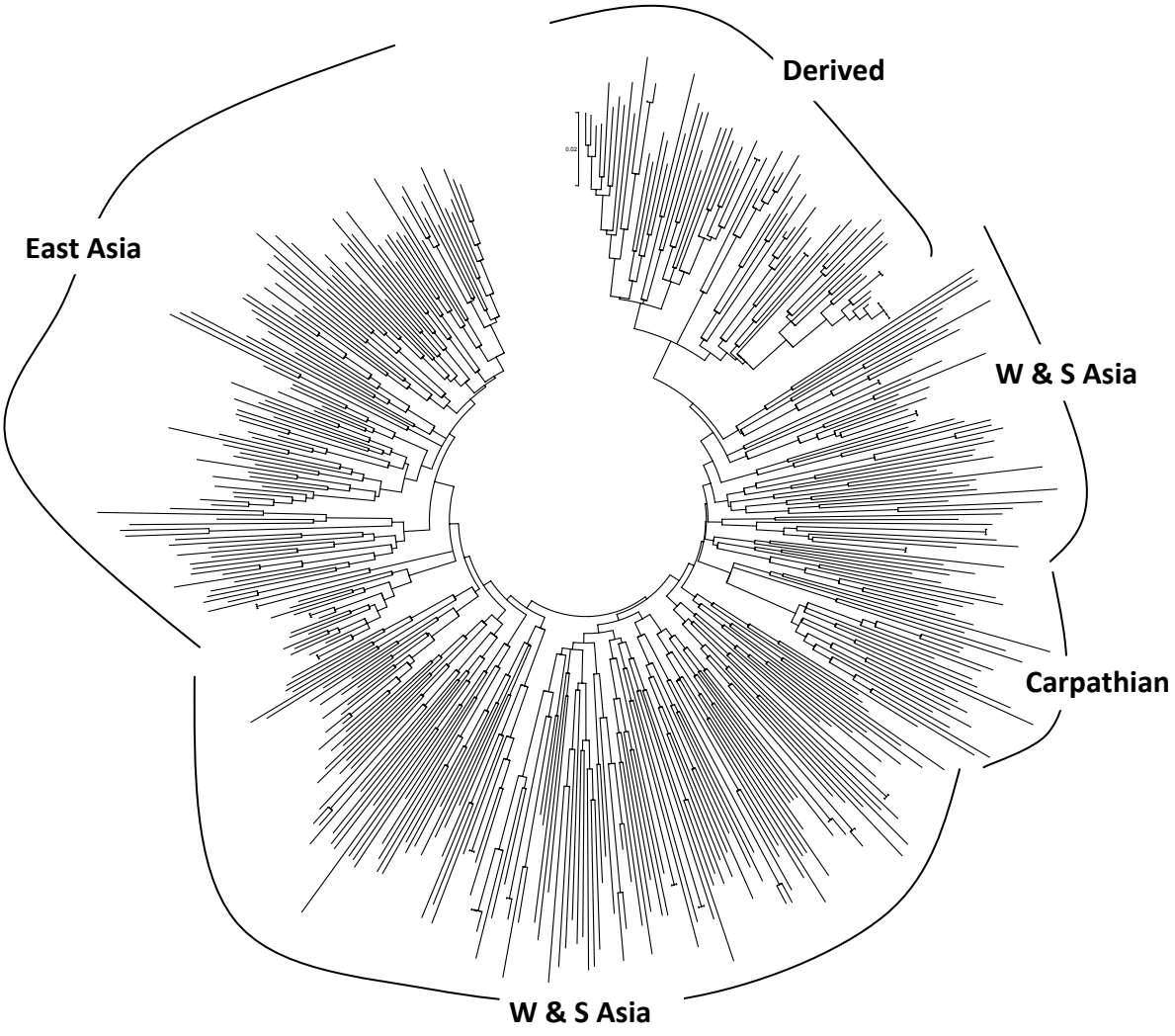

Fig. 1. Genetic relationships among walnut accessions depicting different genetic groups based on a cluster analysis using neighbor-joining method.