ORIGINAL PAPER

Long-term human impacts on genetic structure of Italian walnut inferred by SSR markers

**Paola Pollegioni · Keith Woeste · Irene Olimpieri ·
Danilo Marandola · Francesco Cannata ·
Maria Emilia Malvolti**

Abstract Life history traits, historic factors, and human activities can all shape the genetic diversity of a species. In Italy, walnut (*Juglans regia* L.) has a long history of cultivation both for wood and edible nuts. To better understand the genetic variability of current Italian walnut resources, we analyzed the relationships among the genetic structure of local walnut populations (inferred by SSR markers) and human migrations along ancient routes, using the territory of Royal Tratturo Candela-Pescasseroli (RT) as a case study. Sixteen *J. regia* provenances were collected along RT and compared with 13 Italian provenances and the landrace Sorrento. Although the level of SSR polymorphism we observed was moderately high, AMOVA revealed that most of the diversity was located within individuals (92.58%), and geographical differentiation was low ($D_{est}=$ 0.076). Evidence for human-mediated domestication bottleneck events was detected in about 95% of walnut provenances. A Bayesian approach divided 456 walnut samples into three clusters: (1) Sorrento genotypes, (2) trees from the island of Sicily, and (3) the remaining germplasm. The UPGMA tree based on Nei's distances distinguished northeastern provenances and weakly grouped 12 of 16 prove-nances of RT. The observed genetic differences derived mainly from gradations in allele frequencies. Separation of the Sicilian provenance from the mainland may be explained in terms of founder effects and prolonged geographic isolation. Two contrasting forces, selection, and frequent inter-regional transfer of propagules, appear to drive the patterns of genetic variability for *J. regia*.

Keywords *J. regia* · SSR genetic structure · Royal Tratturo

P. Pollegioni (✉) · I. Olimpieri · D. Marandola · F. Cannata ·
M. Emilia Malvolti
C.N.R. Institute of Agro-environmental and Forest Biology,
viale Marconi 2, 05010,
Porano, Terni, Italy
e-mail: paola.pollegioni@ibaf.cnr.it

K. Woeste
Department of Forestry and Natural Resources, U.S.D.A. Forest Service, Hardwood Tree Improvement and Regeneration Center, Purdue University,
715 West State Street,
West Lafayette 47907-2061 IN, USA

Introduction

A detailed knowledge of genetic diversity and spatial genetic structure is essential for conservation and manage-ment of tree species. Historical events such as habitat fragmentation (e.g., due to climatic changes during glacial and post-glacial periods) (Petit et al. 2005), life history traits, such as reproductive biology, seed dispersal mecha-nisms, ability to be vegetatively propagated (Heuertz et al. 2006), selection and human activities (agriculture, defores-tation, urbanization), play important roles in shaping the genetic diversity within tree species. In particular, habitat fragmentation and a decline in effective population size can affect genetic composition, erode genetic variation, and increase intra-population inbreeding (Cornuet and Luikart 1996). Nevertheless, the response of forest tree species to habitat fragmentation depends in part on their life history characteristics, including long generation time, predominant outcrossing, and a high level of long-distance gene flow via pollen (Dutech et al. 2004). In addition, the distribution of several tree species has been strongly modified by human management during the last 2,000 years, in particular around the Mediterranean basin (Bagnoli et al. 2009). This is especially true for *Juglans regia* (Persian walnut), one of

the most economically important members of the genus *Juglans*.

Persian walnut is an agro-forestry species that is highly valued for its high-quality wood and energy-rich nut-meats. Its nuts are easily stored and transported over long distances. This species, native from South-Eastern Europe to North-Western China through Tibet, Nepal, Northern India, Pakistan, and Iran and probably dispersed along ancient trade routes between China and Greece, grows well in virtually all parts of the world with a temperate climate (Huntley and Birks 1983). The first post-glacial appearances of Persian walnut pollen in Europe occurred around 1,500–2,500 y BP and corresponded with the establishment of Greek and Roman settlements, as demonstrated by the presence of carbonized, unshelled walnuts in archeological excavations in Pompeii and Herculaneum (Meyer 1980). From Italy, the cultivation of walnuts spread to western Europe and northern Africa via trade within the Roman empire (McGranahan and Leslie 2009). Currently, Persian walnut grows in Italy from sea level to 1,000–1,200 m elevation from the Alps to Sicily, and it is adapted to a range of environmental conditions. Because it is useful for wood/fruit production and reforestation of wasted agricultural land, several studies over the last two decades have addressed the genetic variability of Italian walnut resources. Nevertheless, the amount and geographical distribution of nuclear genetic diversity in *J. regia* is not clear. It is widely accepted that no natural population of *J. regia* exists in Italy even if distinctive groups of walnut genotypes can still be found on farms in rural areas. Most of these distinctive types are growth from local seed and mainly planted for family consumption. These walnut groups may be genetically and morphologically variable, and can be considered a source of biodiversity (Malvolti et al. 2010).

The genetic variability of Italian walnut groups, defined as "populations" or "geographic provenances" and named by sampling site, was investigated using isozymes (Malvolti et al. 1993; 1997; Fornari et al. 1999) and RAPD markers (Ferrazzini et al. 2007). These studies revealed low levels of genetic differentiation among Italian populations (F_{ST}= 0.066), and cluster analysis of genotypes did not show a geographic pattern. These data were confirmed by analyzing chloroplast PCR-RFLP markers in 29 Euro–Asiatic populations (Fornari et al. 2001), and a marked erosion of genetic resources was supposed. Recently Woeste et al. (2002) developed a panel of 30 nuclear microsatellites (SSR) for a wide range of genetic investigations in *Juglans*, including clonal identification (Robichaud et al. 2006; Dangl et al. 2005; Foroni et al. 2007), a broad-scale study of the genetic structure of *Juglans nigra* populations in the Central Hardwood Region of the United States of America (Victory et al. 2006), and the identification of

hybridogenic walnut plants (Pollegioni et al. 2009a, b). In a recent study, Gunn et al. (2010) evaluated 220 walnut trees from six Tibetan villages in China using 14 SSR markers. Their data indicated that village environments and familiar relationships were the key factors influencing the genetic variation of Tibetan walnuts. These observations led us to consider the degree of genetic diversity within and among Italian walnut provenances using microsatellites markers. If the observed differentiation is low at the population level, would it be possible to identify genetically homogenous groups of individuals at a higher hierarchical level? We also wondered if the structure of Italian walnut populations was affected by local traditions, the economy of rural communities, and human migrations along ancient routes of transhumance such as *Tratturi*.

Tratturi were the old, grassy tracks formed over centuries by transhumance around the Mediterranean basin. As reported by Avram (2009), the seasonal migration of people and animals from the mountains to the coastal plains occurred along a highly regulated system of wide fixed routes. In Italy, the network of transhumance was extensive (around 3,100 km) and is still partially visible. The longest, most famous and well-conserved Tratturo is the Royal Tratturo Candela-Pescasseroli (RT), which lies mainly in the regions of Abruzzo, Molise, and Campania. The Royal Tratturo path and the transhumance associated with it had an enormous historical, economic, and cultural impact on social structure of these Southern regions (Palasciano 1999). After thousands of years, the pastoral economy of these regions declined and disappeared in the second half of the twentieth century with the advent of railroads and the industrial revolution. Nevertheless, interest in the Tratturi has increased recently, and the Royal Tratturo has been proposed as a World Heritage Site by UNESCO (WWW.http://UNESCO.org). In the framework of the research project "FIMONT" (Italian Ministry of Research), which was devoted to economic development in the Apennine areas, a strong correlation was found between walnut nut production and local culture (e.g., legend of Benevento witches, food), history, and the identity of rural communities along Royal Tratturo Candela-Pescasseroli (Marandola et al. 2008). We postulated there may be walnut provenances tightly connected with the Royal Tratturo, and that they may be genetically distinct from other Italian provenances.

In this study, 16 *J. regia* provenances collected along "Royal Tratturo Candela-Pescasseroli" were compared with 13 other Italian provenances. Nuclear microsatellites (SSR) were used to (1) infer the genetic structure of 29 Italian walnut provenances and (2) provide new insights regarding the possible role of human activities in shaping genetic diversity of *J. regia* in Italy.

Materials and methods

Plant material

Over the last 3 years, The National Research Council—Institute of Agro-environmental and Forest Biology (CNR-IBAF, Porano, Italy) has extensively monitored and sampled walnut germplasm along Royal Tratturo Candela-Pescasseroli (RT). This Tratturo has a total length of 221 km and a width of up to 111 m. It connected Candela (Apulia) and Pescasseroli (National Park of Abruzzo) with the ancient Samnite cities of *Auphidena* (Alfedena), *Aesernia* (Isernia), and *Saepinum* (Sepino) (Fig. 1). A total of 288 adult walnut trees, presumably grown from local seed, were sampled in 16
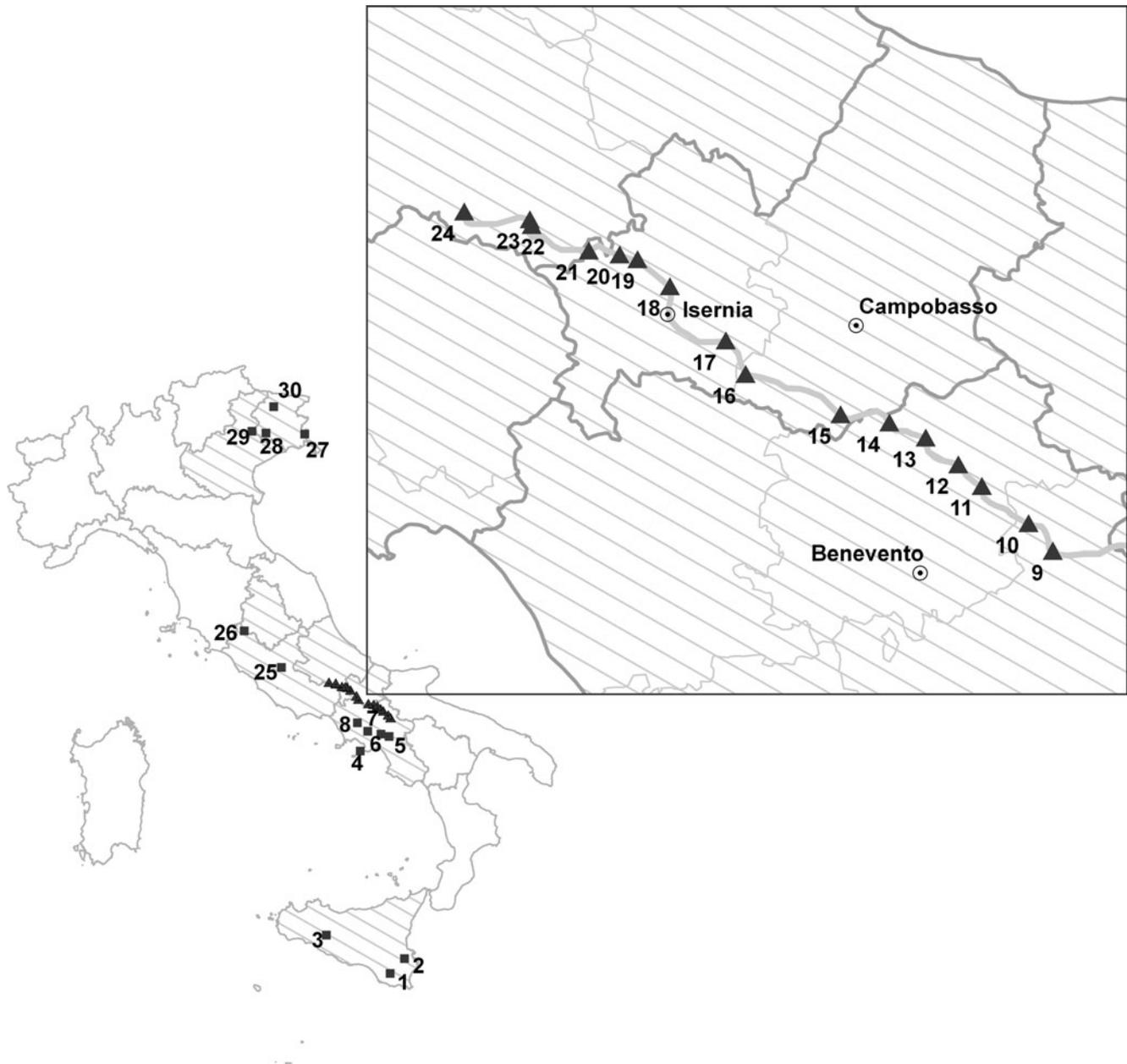
Fig. 1 Location of 16 provenances (*black triangles*) of *Juglans regia* collected along Royal Tratturo Pescasseroli-Candela (9-Ariano Irpino, 10-Montecalvo Irpino, 11-San Giorgio la Molara, 12-San Marco dei Cavoti, 13-Circello, 14-Santa Croce del Sannio, 15-Sepino, 16-San Massimo, 17-S. Maria del Molise, 18-Miranda, 19-Forlì del Sannio, 20-Rionero Sannitico, 21-Montenero Val Cocchiaro, 22-Civitella Alfedena, 23-Villetta Barrea, 24-Pescasseroli), 13 Italian walnut provenances (*black squares*) sampled in Sicily (1-Ragusa, 2-Anapo Valley, 3-Bivona), Campania (5-Montella, 6-San Michele Serino, 7-Tufino, 8-Casolla), Lazio (25-Palombara Sabina), Umbria (26-Castel Giorgio), Veneto (29-Osigo), and Friuli–Venezia Giulia region (27-Gabria, 28-Pordenone, 30-Preone), and 20 genotypes of the variety Sorrento (4). All details of provenances are reported in Table 1. This map was kindly provided by Francesca Chiocchini (CNR-IBAF, Porano, Italy)

Table 1 Number of samples (*N*), geographic coordinates (*Lat*, *Long*), and elevation above sea level (*Elev.*) for 29 Italian provenances of Persian walnut and the Italian non-clonal variety Sorrento

Region	Province	Provenance	Abbreviation	Lat.	Long.	Elev. (*m*)	No. of samples
Sicily	Ragusa	Ragusa	RAGUSA	36°55′45″N	14°43′4″E	502	10
	Siracusa	Anapo Valley	ANAPO	37°9′27″N	15°1′36″E	438	10
	Agrigento	Bivona	BIVONA	37°37′13″N	13°27′21″E	503	10
Campania	Avellino	Ariano Irpino	ARIANO[a]	41°8′59″N	15°5′3″E	788	23
		Montecalvo Irpino	MONTEC[a]	41°11′47″N	15°2′3″E	623	26
		San Giorgio la Molara	MOLARA[a]	41°16′26″N	14°55′52″E	667	8
		Montella	MONT	40°50′40″N	15°1′7″E	670	20
		San Michele Serino	SERINO	40°53′35″N	14°51′19″E	364	10
	Benevento	San Marco dei Cavoti	CAVOTI[a]	41°18′39″N	14°52′50″E	695	29
		Circello	CIRCE[a]	41°21′21″N	14°48′35″E	650	20
		Santa Croce del Sannio	CROCE[a]	41°23′21″N	14°43′59″E	478	11
	Caserta	Casolla	CASOLLA	41°5′43″N	14°21′24″E	200	10
	Napoli	Tufino	TUFINO	40°57′18″N	14°34′0″E	306	10
	Sorrento	Sorrento[b]	SORRENTO	40°37′33″N	14°22′32″E	50	20
Molise	Campobasso	Sepino[c]	ALTILIA[a]	41°24′32″N	14°37′9″E	698	20
		San Massimo	MAS[a]	41°29′38″N	14°24′39″E	630	20
		S. Maria del Molise[d]	FONT[a]	41°33′13″N	14°22′8″E	650	20
	Isernia	Miranda	MIRA[a]	41°38′45″N	14°14′49″E	860	14
		Forlì del Sannio	SANNIO[a]	41°41′48″N	14°10′49″E	610	20
		Rionero Sannitico	RIONERO[a]	41°42′46″N	14°8′22″E	1051	8
		Montenero Val Cocchiaro	VALCO[a]	41°43′4″N	14°4′13″E	950	10
Abruzzo	Aquila	Pescasseroli	PESC[a]	41°48′18″N	13°47′23″E	1200	20
		Civitella Alfedena	ALF[a]	41°45′58″N	13°56′36″E	1123	18
		Villetta Barrea	BARREA[a]	41°46′36″N	13°56′21″E	1000	21
Lazio	Roma	Palombara Sabina	SABINA	42°4′3″N	12°46′8″E	372	8
Umbria	Terni	Castel Giorgio	GIORGIO	42°42′23″N	11°58′38″E	559	15
Friuli-Venezia Giulia	Udine	Preone	PREONE	46°23′43″N	12°52′3″E	460	12
	Pordenone	Pordenone	PORD	45°57′45″N	12°39′22″E	24	10
	Gorizia	Gabria	GABRIA	45°54′26″N	13°34′32″E	49	10
Veneto	Treviso	Osigo	OSIGO	46°0′0″N	12°20′00″E	327	13
Total	–	–				–	456

[a] Walnut provenances sampled along Royal Tratturo Candela-Pescasseroli crossing three Italian regions (Abruzzo, Molise, Campania)

[b] Sorrento non-clonal variety (landrace)

[c] Archeological site of the Roman town *Saepinum/Altilia*

[d] Walnut fountain site

different geographic sites of RT: six sites in Campania, seven in Molise, and three in Abruzzo (Table 1). These samples were compared with 148 adult walnut genotypes collected since 1999 in 13 different Italian sites. The samples were from three sites on the island of Sicily, six sites in south-central Italy and four sites in north-eastern Italy. Mature leaves from each plant were sampled, immediately frozen in liquid nitrogen and stored at −80°C. The number of plants collected, the geographic coordinates and altitude for each walnut provenance are reported in Table 1; the spatial location of the 29 sampled sites are displayed in Fig. 1.

Plants with stem diameter (at 1.30 m above ground) more than 40 cm were sampled and classified as not less than 40-year-old adult trees. As reported by Di Vaio and Minotta (2005), in favorable conditions a mean diameter increment of ~1 cm/year has been observed in walnut plantations of Campania. Thus, we supposed that the sampled trees were growing before the application of European Union-Regulation Directive 2080/92 for afforestation/reforestation of arable lands and the subsequent introduction of germ-plasm from the French and U.S breeding programs. The objective was to collect a minimum of 20 samples per site,

but adult walnut trees with these features were scattered in farmlands at low density, so although all available adult trees per farm were collected, the number of samples taken at each site was not consistent.

For comparison, 20 genotypes of the non-clonal variety (landrace) Sorrento, previously sampled and conserved at the CNR-IBAF repository (Porano, TR) were also included in this study. Foroni et al. (2007) pointed out that Italian walnut varieties are usually composed of genotypes that are similar but not identical. For this reason, they are defined as "landraces" and labeled using the name of their region of origin (e.g., Sorrento).

DNA extraction

Genomic DNA was extracted from all 456 samples by grinding 100 mg of leaf tissue in a 2-ml microcentrifuge tube containing a 5-mm steel bead. The leaf tissue was homogenized in a Mixer Mill 300 (QIAGEN) cooled with liquid nitrogen. Genomic DNA was extracted and purified using the DNeasy96 Plant Kit (QIAGEN) according to the manufacturer's instructions (http://www.qiagen.com), and stored at −20°C. DNA quantity was assessed by comparing all samples against six standardized solutions of λphage DNA (15 ng/μL, 31 ng/μL, 63 ng/μL, 125 ng/μL, 250 ng/μL, 500 ng/μL; Life Technologies) in a 1% agarose gel stained with ethidium bromide and visualized with UV light. The DNA in the samples was brought to a working concentration of 5 ng/μL.

Microsatellite analysis

Ten unlinked microsatellite loci (WGA1, WGA4, WGA9, WGA69, WGA89, WGA118, WGA202, WGA276, WGA321, WGA331) already sequenced and used for retrospective identification of hybridogenic plants (genotypes with a spontaneous crossing ability to produce hybrids) in *Juglans* spp. (Pollegioni et al. 2009b) and for characterization of *J. × intermedia* trees (Pollegioni et al. 2009a) were used to characterize the samples. Polymerase chain reaction (PCR) was performed using 20 ng of DNA template, 10 mM Tris–HCl (pH=8.0), 50 mM KCl, 1.5 mM MgCl₂ reaction buffer, 200 μM dNTP (each), 0.2 μM primer (both), 0.008 μg BSA, and 0.4 U of Taq polymerase (Roche Applied Science). Total reaction volume was 20 μl. Reactions were performed in a GENEAmp 9700 Thermocycler according to the following procedure: an initial denaturation at 94°C for 5 min, followed by 35 cycles of 45 s at 94°C, 45 s at the optimum annealing temperature for each primer pair, and 1 min at 72°C; then a final extension step at 72°C for 7 min. A 5-μl aliquot of the amplified fragment was checked by electrophoresis in 1.8% agarose in 0.5× TBE buffer, and stained with ethidium bromide. To

determine the exact size of the amplified microsatellite fragments, samples were diluted up to 1:10 in water, and 1 μL of the diluted PCR product was mixed with 0.3 μL of a 500-bp internal-lane size standard (Gene Scan™ −500 ROX, Applied Biosystems) and 9.7 μL of pure deionized formamide, denatured in a thermocycler at 95°C for 5 min and immediately chilled on ice. PCR amplification fragments were resolved by capillary electrophoresis with an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems). The resulting data were collected applying Gene Scan Analysis version 3.7 software and genotype profiles were assigned with Genotyper version 3.7 NT software (Applied Biosystems) using six *J. regia* genotypes already characterized by ten SSR markers (Pollegioni et al. 2009a) as a standards across multiple plates.

Data analysis

Genetic diversity of SSR loci

Descriptive gene diversity parameters, number of observed alleles (A), effective number of alleles (N_e), observed (H_o), and expected (H_E) heterozygosity, polymorphic information content (PIC), and the estimated null allele frequency were calculated at each locus and over all loci using POPGENE version 1.32 software (Yeh et al. 1997) and CERVUS version 2.0 software (Marshall et al. 1998). Departures from Hardy–Weinberg expectations at each locus were tested by the likelihood ratio (G test) procedure provided by POPGENE version 1.32 software. In addition, the unbiased estimators of Wright's F-statistics (Weir and Cockerham 1984), within-population inbreeding coefficient f (F_{IS}), total-population inbreeding coefficient F (F_{IT}), and among-population genetic differentiation coefficient θ (F_{ST}) were computed for each locus across all populations and over all loci using hierarchical locus-by-locus AMOVA as implemented in Arlequin version 3.11 software (Excoffier et al. 2005). Statistical significance of f (F_{IS}), F (F_{IT}), and θ (F_{ST}) were tested using a non-parametric approach described in Excoffier et al. (1992) with 1,000 permutations. We used SMOGD 1.2.5 software (Crawford 2009) to measure the actual differentiation coefficient (D_{est}) for each locus across all populations and over all loci among provenances according to Jost (2008).

Identification of SSR outlier loci was carried out following the approach proposed by Beaumont and Nichols (1996), further developed by Beaumont and Balding (2004) and implemented in the FDIST2 software (http://www.rubic.rdg.ac.uk/~mab/software.html). Atypical behavior of a locus was detected by comparison between the observed F_{ST} estimates (Weir and Cockerham 1984) and the expected neutral distribution of F_{ST} conditioned on heterozygosity (H_E). In the present study, the stepwise mutation model of

alleles (SMM, Kimura and Otha 1978) was used. In a first step, the theoretical distribution of F_{ST} conditioned on heterozygosity was computed by 20,000 coalescent simulations based on the overall mean value of F_{ST} calculated from all markers ($F_{ST}=0.054$), 30 populations, and 27 individuals as a sample size per population. FDIST2 software provided the expected confidence intervals of F_{ST} vs. H_E by estimating the 0.05, 0.50, and 0.95 quantiles of the F_{ST} distribution. Loci that were outside the 95% confidence intervals were removed, and a new analysis was performed with a recalculated mean value of F_{ST} (0.049). Markers with F_{ST} values that fell outside the 0.95 limits after this second analysis were considered as outlier loci. This procedure reduces bias in the estimation of the mean neutral F_{ST} by removing the most extreme loci from the estimation.

Genetic diversity of the provenances

Descriptive gene diversity statistics, mean number of alleles per locus (A), observed (H_o) expected (H_E), and unbiased expected heterozygosity (UH_E) were calculated for each geographic provenance using the GenAlEx software 6.3 (Peakall and Smouse 2005). The estimation of mean number of alleles per locus as a measure of allelic richness can be affected by differences in sample size. For this reason, allelic richness (Rs) and private allele richness (PR), which are independent of sample size (El Mousadik and Petit 1996), were computed by the rarefaction method with HP-Rare software (Kalinowski 2004). This approach uses the frequency of alleles at a locus to estimate the expected number of alleles and/or private alleles in a sub-sample of n individuals selected at random from a sample of N individuals in each population. In this study, the estimates of Rs and PR were based on minimum sample size of seven individuals. The within-population inbreeding coefficient F_{IS} (Weir and Cockerham 1984) per provenance was calculated using hierarchical locus-by-locus AMOVA as implemented in Arlequin software. The statistical significance of F_{IS} was tested using a non-parametric approach described in Excoffier et al. (1992) with 1,000 permutations. In addition, all provenances were used in a locus-by-locus AMOVA to examine the distribution of molecular variance at three hierarchical levels: among provenances, among individuals within provenances, and within individuals. In order to explore the power of SSR markers to identify individuals, we also computed the probability of identity (PI_{unb} and PI_{sib}) for each provenance over all loci and determined private alleles per provenance and per individual (allele present in only one individual). These calculations were performed using GenAlEx software 6.3.

The possibility of founder effects due to a recent colonization or/and the occurrence of human-mediated domestication bottleneck events was tested for each provenance using the BOTTLENECK software (Piry et al. 1999). This approach is based on the observation that populations that have experienced a recent reduction in effective population size exhibit a more rapid reduction of allelic diversity than heterozygosity at polymorphic loci. Hence, in recently bottlenecked populations, the observed heterozygosity is higher than the expected heterozygosity estimated from the observed allele numbers under the assumption of mutation-drift equilibrium (Cornuet and Luikart 1996). Significance was assessed using the "Wilcoxon's signed-rank" test, which provides relatively high power and can be used with as few as four polymorphic loci and any number of individuals. Three models of evolution have been proposed for microsatellite loci in BOTTLENECK software: Stepwise Mutation Model (SMM), Infinite Alleles Mutation Model of loci (IAM), and Two-Phase Model (TPM). Pollegioni et al. (2009a) found that TPM most accurately reflected the mutational mechanism of the ten microsatellite loci used in this study. As recommended by Piry et al. (1999), we used the TPM with 95% SMM and 5% multistep mutations. Because some of our sample sizes were small, the bottleneck analysis was supplemented with the "M-ratio test" of Gaza and Williamson (2001) implemented in Arlequin software. This method computes the mean ratio (M) of the total number of alleles (k) to the range in allele size (r). In populations that experienced large reductions in effective size, the allele number is expected to be reduced more quickly than the range of allele size, leading to a decrement of M values. As proposed by Gaza and Williamson (2001), the ratio M estimated for each locus, averaged over loci, was compared to a critical value: any data set with a value of $M<0.68$ can be assumed to have gone through a recent reduction in size.

Population structure analysis

The population structure and proportion of membership (Q value) for each predefined population and each individual sample in each of the predicted clusters were inferred using the Markov Chain Monte Carlo (MCMC) and Bayesian clustering algorithms implemented in STRUCTURE software 2.3.3 (Pritchard et al. 2000). This method attempts to assign individuals to several genetic groups in order to minimize within-group linkage disequilibrium and deviation from Hardy–Weinberg equilibrium. As suggested by Falush et al. (2007), STRUCTURE analysis was performed using the admixture model on the whole dataset with no previous population information and the correlated allele frequencies between population options. In this study, the range of possible number of clusters (K) tested was from 1 to 33 (the putative number of provenances plus 3). Based on the initial results, a series

of six independent runs were performed for K between 1 and 16 with a burn-in period of 10,000 steps followed by 10^5 MCMC replicates. Furthermore, the ad hoc statistic ΔK defined by Evanno et al. (2005) was used to detect the most likely number of populations. The ΔK-statistics is based on the second order rate of change of L(K) (the posterior probability of the data among given K) between successive K values over six replicates. As demonstrated by Evanno et al. (2005), it is possible to identify the number of clusters corresponding to the uppermost hierarchical level of genetic partitioning between populations. Therefore, the groups inferred by the first STRUCTURE analysis were subsequently processed separately in order to identify possible substructure. The six runs from the most probable number of clusters were averaged applying *FullSearch* algorithm provided by CLUMPP software 1.1.2 (Jakobsson and Rosenberg 2007). The corresponding Q matrices were graphically displayed by DISTRUCT software (Rosenberg 2004).

Two assignment tests, the Paetkau et al. (1995) frequency method and Rannala and Mountain (1997) partial Bayesian method implemented in GENECLASS software 2.0 (http://montpellier.inra.fr/CBGA/softwares/), were applied to identify genotypes that were unlikely to be encountered assuming the most likely number of K clusters as determined using the software STRUCTURE. Both approaches removed the individual being assigned (leave-one-out procedure), computed the allelic frequencies in all candidate clusters (assuming HWE), calculated the likelihoods of the individual's multilocus genotypes occurring in each cluster (independence of loci), and assigned the individual to the cluster with the highest likelihood. Missing alleles were assigned an arbitrary nonzero frequency (0.01). The Rannala and Mountain (1997) method is comparable with the frequency method, but uses a Bayesian approach to estimate the allele frequencies of the population. In this study, we applied these two tests incorporating, in addition, the exclusion–simulation approach of Cournet et al. (1999) to obtain a confidence level for each individual assignment (p value=0.01). The statistical threshold was calculated by simulating 1,000 genotypes with the novel Monte Carlo resampling method (Paetkau et al. 2004).

Finally, in order to visualize the relationships between provenances, a UPGMA (Unweighted Pair-Group Method with Arithmetic mean) tree was constructed based on Nei's (1972) genetic distance. Bootstrap support for this tree was determined by resampling loci 1,000 times using POP-TREE2 software (Takezaki et al. 2010).

Isolation by distance

Genetic differentiation between provenance pairs was measured by Wright's F_{ST} coefficient (Weir and Cockerham 1984) using Arlequin software. Although the R_{ST} coefficient (Slatkin 1995) is based on the SMM, which can

reflect more accurately the mutation pattern of microsatellites, we excluded R_{ST} computation from this study because, as reported by Balloux and Lugon-Moulin (2002), F_{ST} appears to be a more sensitive measure of intra-specific variation. Pair-wise differentiation based on Jost's D coefficient (D_{est}) was calculated using SMOGD (V.1.2.5) software. Assuming a non-linear distribution of sampling sites, a non-parametric pairwise correlation test between the matrices of $F_{ST}/(1-F_{ST})$ or $D_{est}/(1-D_{est})$, and the natural logarithm of geographic distances was applied to test for isolation by distance (Mantel 1967; Rousset 1997). The p value for the Z score of the Mantel association parameter was inferred using 1,000 permutations. These calculations were performed using GenAlEx software 6.3.

Results

Microsatellite polymorphism and genetic diversity of walnut provenances

All ten SSR loci used in the present study were highly polymorphic in the sampled populations (Table 2). A total of 62 alleles were detected in the 456 walnut trees genotyped. The number of alleles per locus ranged from three at locus WGA4 and WGA331 to a maximum of 14 at locus WGA276, with an average of 6.2. Except WGA4 (0.355) and WGA331 (0.382), all markers were highly informative (PIC>0.50) and useful for genetic diversity studies. The observed heterozygosity and gene diversity greatly varied across the ten SSR loci. The average observed (H_O) and expected (H_E) heterozygosity were 0.597 (SE=0.114) and 0.644 (SE=0.103), respectively. This large variance mainly resulted from the large variation in the number of alleles per locus and allele frequency distribution detected among provenances. As expected, loci with smaller number of alleles tended to have lower heterozygosity and vice versa. Highly significant ($p<0.01$) departures from Hardy–Weinberg expectations across all samples were found for WGA4 and WGA69 using a likelihood ratio (G test) procedure. The within-population inbreeding coefficient $f(F_{IS})$ was negative for five loci (WGA89, WGA1, WGA202, WGA276, WGA321) and positive for the remaining SSR loci (WGA4, WGA118, WGA69, WGA9, WGA331). Nevertheless, as reported in Table 2, only for WGA69 was $f(F_{IS})$ significantly greater than zero, indicating a high level of heterozygote deficiency at this locus, probably as a consequence of the allelic dropout including presence of null alleles (Soulsbury et al. 2007). The null allele frequency (F_{nul}) estimated in this study ranged from 0.0094 for WGA276 to 0.0358 for WGA9. WGA69 was determined to have a high frequency of putative null alleles (0.2019) not associated with missing

Table 2 Genetic characterization of ten microsatellite loci (Pollegioni et al. 2009a) for 29 Italian walnut provenances and the Sorrento non-clonal variety

| Locus	Size range (bp)	A	Ne	PIC	H_o	H_E	P value[a]	F_{nul}	$f(F_{IS})$ [b]	F (F_{IT}) [b]	θ (F_{ST}) [b]	D_{est}
WGA89	211–223	4	2.863	0.576	0.6162	0.6514	0.101	0.0294	−0.02102	0.05672	0.07614**	0.116262
WGA1	176–192	7	2.980	0.597	0.6601	0.6652	0.214	0.0025	−0.03886	0.00931	0.04637**	0.068085
WGA202	251–299	12	4.140	0.719	0.7303	0.7595	0.995	0.0181	−0.01575	0.04050	0.05538**	0.137026
WGA4	231–235	3	1.805	0.355	0.4232	0.4467	**0.000**	0.0230	0.00160	0.05441	0.05289**	0.038176
WGA276	165–199	14	3.964	0.708	0.7588	0.7486	0.090	0.0097	−0.06913	−0.01164	0.05378**	0.139536
WGA118	183–206	4	2.441	0.505	0.5504	0.5911	0.060	0.0329	0.04423	0.06970*	0.02665	0.058328
WGA69	159–179	6	3.632	0.677	0.4781	0.7255	**0.000**	0.2019	0.29086**	0.34304**	0.07357**	0.195663
WGA9	239–247	4	2.950	0.588	0.6140	0.6618	0.179	0.0358	0.03097	0.07378*	0.04417**	0.082762
WGA321	226–245	5	3.175	0.624	0.6732	0.6859	0.016	0.0074	−0.02064	0.01983	0.03965**	0.071648
WGA331	274–278	3	2.011	0.382	0.4693	0.5033	0.019	0.0334	0.00312	0.06987	0.06696**	0.055643
Mean (SE)	–	6.2 (3.82)	2.996 (0.77)	0.573 (0.13)	0.597 (0.114)	0.644 (0.103)		–	0.02161*	0.07414**	0.05369**	0.076

Twenty-nine Italian walnut provenances and the Sorrento non-clonal variety: for each locus total number of alleles (*A*), effective number of alleles (*Ne*), observed (*H_o*), and expected heterozygosity (*H_E*), polymorphic information content (*PIC*), null allele frequency estimated (*F_{nul}*), and unbiased estimate of Wright's fixation indices, within-population inbreeding coefficient $f(F_{IS})$, total-population inbreeding coefficient F(F_{IT}), and among-population genetic differentiation coefficient $\theta(F_{ST})$ and estimator of actual differentiation D_{est} (Jost, 2008) are shown

[a] *P* value associated with likelihood ratio test (*G* test) for departure from Hardy–Weinberg equilibrium: significant *p* values are in *bold* (=*p*<0.01).

[b] Level of significance of unbiased estimate of Wright's fixation indices were tested using a non-parametric approach described in Excoffier et al. (1992) with 1,000 permutations: **p*<0.05, ***p*<0.01

data. Application of the Beaumont and Nichols (1996) to the ten SSR markers identified WGA69 as an outlier locus. The estimated F_{ST} for WGA69 was found to lie outside the 95% confidence region of the conditional joint distribution of F_{ST} and mean heterozygosity based on analysis using FDIST2 (Fig. 2).

Genetic diversity estimates showed that the provenances maintained high levels of genetic diversity, and a large number of private alleles (Table 3). The mean expected and observed heterozygosity across all loci ranged from 0.633 (MONT) to 0.487 (SORRENTO) and from 0.691
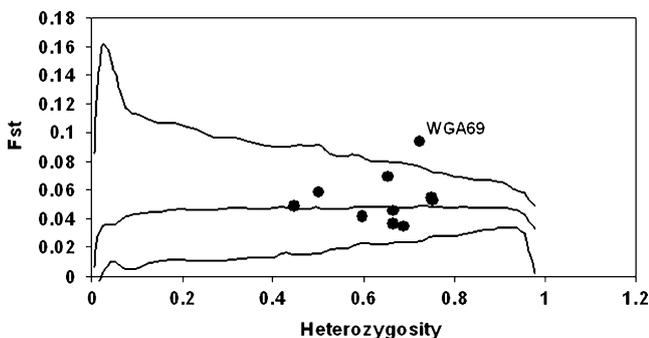
Fig. 2 Expected heterozygosity and F_{st} distribution for ten SSR loci based on differentiation among 29 walnut populations and Sorrento non-clonal variety. The simulated confidence area for neutral loci (with F_{st}=0.049 and mean size=27) was computed using FDIST2 software (Beaumont and Nichols, 1996; Beaumont and Balding, 2004), with a confidence level set to 95%. Loci from the observed dataset are represented as dots; outlier loci are labeled (see Table 2)

(CROCE*) to 0.425 (SORRENTO), respectively. The overall F_{IS} (inbreeding level within population) varied from 0.152 (SORRENTO) to −0.105 (CROCE) and was negative for 11 out of the 29 walnut provenances, indicating a slight surplus of heterozygotes in those provenances. Where F_{IS} values were large and positive, in Civitella Alfedena site (ALF) and SORRENTO, the observed heterozygosity deficit was significantly different from Hardy–Weinberg expectation (Table 3). The allelic richness (R_s) computed by the rarefaction method did not differ greatly among provenances, varying from 3.53 (PORD) to 2.92 (MOLARA*) with a minimum value (2.67) in SORRENTO variety. Of 62 alleles detected across the loci, ten were unique to a single geographic provenance. Two of these alleles, WGA1 (176) and WGA202 (279), were found at low frequency (−0.10) among trees in CIRCELLO and RAGUSA provenance, the remaining eight were private to single individuals. The chance of finding two individuals with the same genotypes in each group was almost nil; the probability that two unrelated individuals would share the same genotypes (PI$_{unb}$) and the probability that two full-sibs will have identical genotypes (PI$_{sib}$) were extremely low in most cases. Across all ten loci, PI$_{unb}$ and PI$_{sib}$ values ranged from 1.4×10^{-5} (SORRENTO) to 9.8×10^{-8} (MONT) and from 1.0×10^{-3} (CIRCE*) to 9.8×10^{-4} (PORD), respectively.

Although population differentiation was significant at nine SSR loci (*p*<0.05; Table 2), the average multilocus estimate of F_{ST} was low, 0.05369. In addition, the mean

Table 3 Genetic diversity of Persian walnut within 29 Italian provenances and the Sorrento non-clonal variety assessed with ten microsatellite loci

Population	A	Rs	H_o	H_E	UH_E	F_{IS}	Probability of identity		Private alleles (provenance)	Private alleles (individual)	PR
							PI_{unb}	PI_{sib}	Locus (bp)	Locus (bp)	
ALTILIA[b]	3.6	3.19	0.590	0.600	0.616	0.042	2.8×10^{-07}	1.1×10^{-03}		WGA202(299)	0.04
MAS[b]	3.5	3.02	0.640	0.577	0.592	−0.083	7.5×10^{-07}	1.5×10^{-03}		WGA1(182); WGA276(195)	0.07
FONT[b]	3.6	3.17	0.600	0.596	0.611	0.019	3.2×10^{-07}	1.1×10^{-03}			0.02
MIRA[b]	3.7	3.32	0.586	0.593	0.615	0.049	2.7×10^{-07}	1.1×10^{-03}		WGA202(252)	0.05
SANNIO[b]	3.6	3.2	0.600	0.593	0.608	0.014	2.4×10^{-07}	1.1×10^{-03}			
RIONERO[b]	3.1	3.04	0.575	0.571	0.609	0.060	1.0×10^{-06}	1.7×10^{-03}		WGA276(199)	0.15
VALCO[b]	3.3	3.11	0.570	0.551	0.580	0.019	1.2×10^{-06}	2.0×10^{-03}			
ALF[b]	3.8	3.39	0.578	0.631	0.649	0.112[b]	7.2×10^{-08}	6.9×10^{-04}			
BARREA[b]	3.7	3.17	0.619	0.589	0.604	−0.026	3.4×10^{-07}	1.2×10^{-03}			
MOLARA[b]	3.0	2.92	0.538	0.498	0.531	−0.013	7.6×10^{-06}	4.0×10^{-03}			0.03
ARIANO[b]	3.6	3.02	0.670	0.596	0.609	−0.102	5.6×10^{-07}	1.2×10^{-03}			
MONTEC[b]	3.7	3.23	0.635	0.630	0.642	0.012	1.0×10^{-07}	7.3×10^{-04}			
CROCE[b]	3.6	3.3	0.691	0.600	0.628	−0.105	3.4×10^{-07}	1.1×10^{-03}		WGA276(165)	0.06
CIRCE[b]	3.8	3.34	0.585	0.600	0.615	−0.105	2.0×10^{-07}	1.0×10^{-03}	WGA1(176)		0.11
CAVOTI[b]	3.5	3.09	0.559	0.580	0.590	0.054	5.6×10^{-07}	1.4×10^{-03}			
MONT	3.5	3.19	0.660	0.633	0.649	−0.017	9.8×10^{-08}	7.0×10^{-04}			
PESC[b]	3.8	3.29	0.595	0.621	0.637	0.067	1.4×10^{-07}	8.1×10^{-04}		WGA1(186)	0.05
SORRENTO	3.1	2.67	0.425	0.487	0.500	0.152[b]	1.4×10^{-05}	4.7×10^{-03}			0.02
OSIGO	3.8	3.35	0.608	0.587	0.610	0.004	3.2×10^{-07}	1.2×10^{-03}			0.01
PORD	3.8	3.53	0.660	0.601	0.632	−0.047	1.5×10^{-07}	9.8×10^{-04}			0.01
PREONE	3.7	3.36	0.583	0.610	0.637	0.087	1.7×10^{-07}	9.2×10^{-04}			0
GABRIA	3.2	3.11	0.650	0.598	0.629	−0.035	3.7×10^{-07}	1.1×10^{-03}			0
SERINO	3.4	3.19	0.600	0.569	0.599	−0.002	9.0×10^{-07}	1.7×10^{-03}			0.01
CASOLLA	3.2	3.1	0.570	0.563	0.592	0.039	1.2×10^{-06}	1.8×10^{-03}			0
TUFINO	3.3	3.19	0.570	0.578	0.608	0.066	4.3×10^{-07}	1.4×10^{-03}			0.01
SABINA	3.0	2.95	0.538	0.558	0.595	0.103	2.0×10^{-06}	2.0×10^{-03}			0
GIORGIO	3.6	3.22	0.580	0.598	0.619	0.065	2.5×10^{-07}	1.1×10^{-03}		WGA202(271)	0.08
RAGUSA	3.2	3.14	0.650	0.602	0.634	−0.027	2.7×10^{-07}	1.1×10^{-03}	WGA202(279)		0.13
ANAPO	3.4	3.24	0.600	0.600	0.632	0.053	3.2×10^{-07}	1.1×10^{-03}			0.02
BIVONA	3.2	3.08	0.570	0.587	0.617	0.081	4.8×10^{-07}	1.3×10^{-03}			0.03

Mean number of alleles per locus (A), allelic richness (Rs), and private allelic richness (PR) standardized to seven individuals from the original number of trees per provenance, observed (H_o), expected (H_E), and unbiased expected heterozygosity (UH_E), inbreeding coefficient (F_{IS}), unbiased probability of identity (PI_{unb}), probability of identity between two random full sibs (PI_{sib})

[a] Significance of inbreeding coefficient F_{IS} was tested using a non-parametric approach described in Excoffier et al. (1992) with 1,000 permutations: *$p<0.05$

[b] Walnut provenances sampled along Royal Tratturo Candela-Pescasseroli

actual differentiation D_{est} coefficient (Jost, 2008), an alternative measure of genetic differentiation, was 0.076, ranging from 0.0381 (WAG4) to 0.1956 (WGA69). This indicates that genetic differentiation among walnut provenances was relative low. The hierarchical locus-by-locus AMOVA revealed that the majority of molecular variance (92.58%) was partitioned within individuals, while the 5.37% was distributed among provenances, and 2% among individuals within provenance (Table 4). Nevertheless, 79.3% of the pairwise F_{ST} comparisons between provenances were statistically significant ($p<0.05$). The highest F_{ST} and D_{est} values

were detected between SORRENTO (F_{ST}=0.053–0.156; D_{est}=0.100–0.217), Sicilian provenances (F_{ST}=0.089–0.029; D_{est}=0.01–0.124), and the remaining walnut collection (data not shown). The Mantel correlations between the pairwise linearized genetic differentiation values [F_{ST}/(1-F_{ST}) or D_{est}/(1-D_{est})], and the natural logarithm of geographic distances connected with the sampling sites were not significant. This indicates an absence of isolation by distance among the sampled walnut provenances.

Evidence for human-mediated domestication bottleneck events was detected using two methods. Wilcoxon's signed-

Table 4 The hierarchical locus-by-locus AMOVA (Excoffier et al. 2005) of 456 walnut samples using ten SSR loci

Degrees of freedom (*d.f.*) and the significance (*p* value) of the variance components are shown

Source of variation	d.f.	Variance components	% variation	p value
Among populations	29	0.17320	5.37	0.0410
Among individuals within populations	426	0.06598	2.05	<0.0001
Within individuals	456	2.98684	92.59	<0.0001
Total	911	3.22602		

rank test revealed a significant excess of heterozygosity for all walnut provenances ($p<0.05$) except for the two Royal-Tratturo sites VALCO and MOLARA, and for the provenances OSIGO, PORD, SERINO, SABINA. A marginal deficiency in heterozygosity was observed for SOR-RENTO. Using the "M-ratio test" of Gaza and Williamson (2001), we observed a genetic signature consistent with a bottleneck in all the provenances; G–W values ranged from 0.582 (RIONERO) to 0.491 (TUFINO), lower than the critical threshold (0.68).

Hierarchical cluster analysis of walnut provenances

Higher-level hierarchical genetic structure of the 29 Italian walnut provenances and SORRENTO was evaluated by the Bayesian cluster analysis implemented in STRUCTURE (Pritchard et al. 2000). For the admixture and correlated frequency model, the log-likelihood value ($L(K)$) as a function of K (number of clusters) averaged over six replicates increased almost linearly from $K=1$ up to $K=3$ ($-10,483$) and declined from $K=4$ to $K=6$ ($-11,890$) (Fig 3a). The approach of Evanno et al. (2005) strongly supported $K=3$ as the most likely number of clusters (Fig. 3b), because the highest second order of change of the log-likelihood of the data (ΔK), as a function of K, was detected at $K=3$. The first cluster was comprised of all 20 genotypes of SORRENTO, based on each sample's estimated membership percentage (admixture proportion or Q value) (Fig. 4). The second cluster grouped *J. regia* plants collected from three distinct geographic sites on the island of Sicily, ANAPO, BIVONA, and RAGUSA. The third

cluster assembled the remaining 15 provenances (396 trees), including all samples from northeastern (PREONE, PORD, GABRIA; OSIGO), central (SABINA, GIORGIO) and southern Italy (MONT, SERINO, CASOLLA) except TUFINO, and along Royal Tratturo Candela-Pescasseroli (ARIANO, MONTEC, MOLARA, CAVOTI, CIRCE, CROCE, ALTILIA, MAS; FONT, MIRA, SANNIO, RIONERO, VALCO, PESC, ALF, BARREA). The samples in the third cluster showed a remarkable similarity with those from the second cluster (from Sicily), with a provenance mean membership coefficient (Q_3) only slightly higher than their coefficient for Q_2 (Table 5). The provenances with the highest mean membership coefficient in the third cluster (Q_3) were ALTILIA ($Q_3=0.497$), BARREA ($Q_3=0.561$), and RIONERO ($Q_3=0.540$), whereas the walnut genotypes sampled in TUFINO (Campania region) were only slightly more genetically related to cluster 2 (SICILY) ($Q_2=0.550$) than cluster 3 (Table 5). The subsequent Bayesian clustering analysis within each inferred cluster did not reveal any genetic substructure. The assignment tests, the Paetkau et al. (1995) frequency method, and Rannala and Mountain's (1997) partial Bayesian method, combined with the exclusion simulation significance test of Cournet et al. (1999) based on individuals, confirmed the above results and revealed that about 80% of the genotypes of SORRENTO were assigned to cluster 1, and 77% of trees collected on Sicily were incorporated in cluster 2 (Table 5). Walnut trees located in central, northeastern, and southern Italy, including sites along Royal Tratturo Candela-Pescasseroli route, were mainly assigned to cluster 3. Only 11 (Bayesian approach)
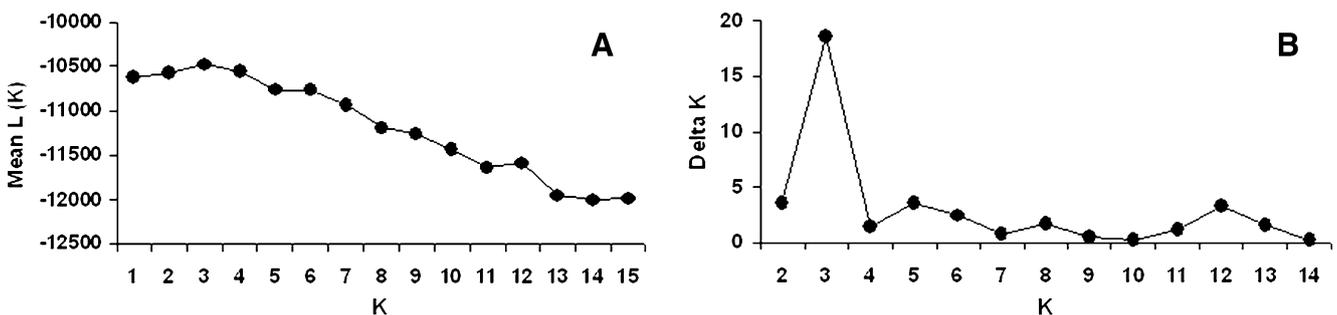
Fig. 3 Inference of *K*, the most probable number of clusters, using STRUCTURE software, based on microsatellite analysis of 456 total walnut samples (*circle*). **a** Log-likelihood value of data *L(K)* as a function of *K* averaged over six replicates. **b** Second order of change of the log-likelihood of the data (ΔK) as a function of *K*, calculated over six replicates
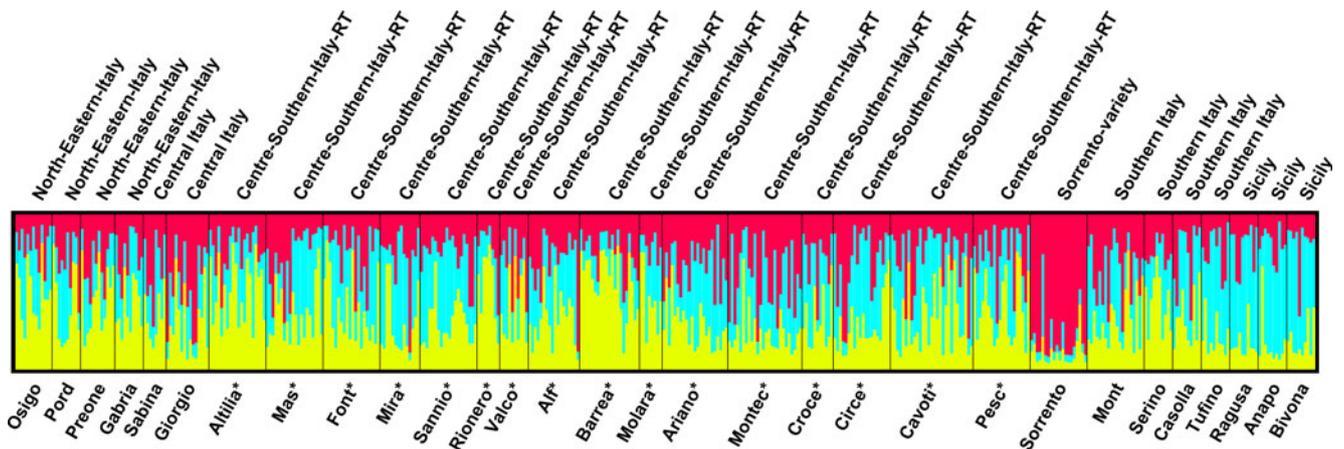
Fig. 4 Population structure inference for 456 Italian walnut samples by Bayesian assignment using STRUCTURE software (K=3 clusters). Each individual is represented by a vertical line and populations are separated by a *vertical black line*. *Different colors* in the same line indicate the individual's estimated membership percentage in K clusters (admixture proportion or Q value): red=cluster 1, light blue=cluster 2, yellow=cluster 3. For details of provenance abbreviations and locations see Table 1 and Fig. 1

or 7 (frequency approach) samples were not assigned, exceeding the critical threshold ($p<0.01$) (Table 5).

The UPGMA tree based on Nei's (1972) genetic distances provided additional insights into the relationships between walnut provenances (Fig. 5). As expected, three major groups were detected with bootstrap values more than 50%, corresponding to the three clusters inferred via STRUCTURE analysis (Table 4). SORRENTO and Sicilian provenances formed two distinct groups, clearly separated from the remaining walnuts collected at the mainland sites. The TUFINO plant collection appeared to represent a transition zone between Sicily and mainland provenances. Although the Bayesian clustering approach did not allow us to clearly identify a sub-genetic structure within cluster 3, all provenances from northeastern Italy clustered together in the UPGMA tree. In addition, despite the low-estimated F_{ST}, 12 of 16 provenances sampled along Royal Tratturo Candela-Pescasseroli showed a tendency to group together, especially ALTILIA, BARREA, RIONERO, and VALCO.

The genetic difference between the 29 Italian walnut provenances and SORRENTO derived mainly from gradations in allele frequencies rather than from distinctive private alleles. In particular, a clear gradient in allele frequency distribution was observed among groups at locus WGA69. As showed in Fig. 6, Sicilian provenances and TUFINO (cluster 2) exhibited high frequency of the 159-bp allele (71.3%) and low frequency for the 161 bp (3.8%), 169 bp (0.0%), 175 bp (5.0%), 177 bp (1.3%), and 179 pb (18.8%) alleles, whereas for SORRENTO (cluster 1) the 179 bp (60.0%) and 175 bp (22.5%) alleles predominated and the 159 bp (7.5%), 161 bp (10%), 169 bp (0.0%) and 177 bp (0.0%) alleles were uncommon. The average allele frequency for the remaining walnut germplasm (cluster 3)

ranged from 0.2% (177 bp) to 35.9% (159 bp), with the 169-bp allele private to northeastern Italy.

Discussion

This study represents the first large-scale analysis of Persian walnut germplasm in Italy. The analysis was based on ten nuclear microsatellite markers originally developed in *Juglans nigra* (Woeste et al. 2002), but successfully used in related species for a wide range of genetic applications (Pollegioni et al. 2009a). In this study, mean number of alleles per locus (6.2) and effective number of alleles (2.99) computed over all 456 walnut samples were relatively high compared to the levels of variability detected in 48 *J. regia* cultivars (Dangl et al. 2005), SORRENTO (Foroni et al. 2007) and five *J. regia* populations from central and south-western China (Wang and Pei 2008). Conversely, the ranges of allelic richness and observed heterozygosity (0.597) observed across the same subset of SSR loci were lower than the corresponding values observed in autochthonous populations of *Juglans mandshurica* (Bai et al. 2007), in 39 open-pollinated *J. nigra* families (Robichaud et al. 2006), and 43 indigenous populations of *J. nigra* collected in the Central Hardwood Region of the United States (Victory et al. 2006). This result was not surprising because of the large sample sizes and extensive ranges sampled in the studies of black walnut. Nevertheless, the genetic diversity parameters we observed for the 29 Italian provenances and Sorrento variety were similar to values measured for other domesticated tree species such as black poplar (Van der Schoot et al. 2000), *Cupressus sempervirens* (Bagnoli et al. 2009), *Fraxinus mandshurica* (Hu et al. 2008), and

Table 5 Number of walnut genotypes from each provenance correctly assigned to each of three clusters inferred by STRUCTURE analysis

| Provenance	No.	Cluster 1			Cluster 2			Cluster 3			Not assigned	
		Q_1	Bayesian	Frequ.	Q_2	Bayesian	Frequ.	Q_3	Bayesian	Frequ.	Bayesian	Frequ.
OSIGO	13	0.197	–	–	0.322	–	–	0.481	11	12	2	1
PORD	10	0.279	–	–	0.320	–	–	0.400	10	10	–	–
PREONE	12	0.288	–	–	0.249	1	–	0.463	11	12	–	–
GABRIA	10	0.257	–	–	0.344	–	–	0.399	10	9	–	1
SABINA	8	0.338	–	–	0.317	–	–	0.345	8	8	–	–
GIORGIO	15	0.444	3	2	0.268	2	1	0.288	9	12	1	–
ALF[a]	18	0.305	–	–	0.361	3	3	0.334	15	15	–	–
BARREA[a]	22	0.250	–	–	0.189	1	1	0.561	21	20	–	1
PESC[a]	20	0.259	–	–	0.337	4	4	0.404	14	16	2	–
ALTILIA[a]	20	0.226	–	–	0.277	2	1	0.497	18	19	–	–
MAS[a]	20	0.266	–	–	0.377	3	1	0.357	17	19	–	–
FONT[a]	20	0.228	–	–	0.385	4	4	0.387	16	16	–	–
MIRA[a]	14	0.271	1	–	0.439	2	2	0.290	10	12	1	–
SANNIO[a]	20	0.248	–	–	0.441	4	4	0.311	16	16	–	–
RIONERO[a]	8	0.226	–	–	0.234	–	–	0.540	7	7	1	1
VALCO[a]	10	0.313	–	–	0.284	1	1	0.403	9	9	–	–
MOLARA[a]	8	0.208	–	–	0.331	–	1	0.461	8	7	–	–
ARIANO[a]	23	0.328	–	–	0.378	1	1	0.293	22	22	–	–
MONTEC[a]	26	0.395	1	–	0.367	7	8	0.239	18	18	–	–
CROCE[a]	11	0.282	–	–	0.387	3	4	0.331	8	7	–	–
CIRCE[a]	20	0.308	–	–	0.403	4	4	0.289	14	14	2	2
CAVOTI[a]	29	0.273	–	–	0.379	–	3	0.348	29	26	–	–
SERINO	10	0.254	–	–	0.310	–	1	0.437	10	9	–	–
CASOLLA	10	0.254	–	–	0.407	1	–	0.339	9	10	–	–
MONT	20	0.351	–	–	0.335	4	5	0.314	16	15	–	–
SORRENTO	20	0.777	16	15	0.105	–	1	0.118	3	4	1	–
TUFINO	10	0.231	–	–	0.550	3	5	0.219	7	5	–	–
RAGUSA	10	0.215	–	–	0.608	9	8	0.177	–	1	1	1
ANAPO	10	0.198	–	–	0.627	8	9	0.176	2	1	–	–
BIVONA	10	0.192	–	–	0.607	6	6	0.201	4	4	–	–

Two different assignment tests were used: the Paetkau et al. (1995) frequency method, and Rannala and Mountain's (1997) partial Bayesian method incorporating the exclusion–simulation approach of Cournet et al. (1999) to obtain a confidence level for each individual assignment (*p* value <0.01). Percentage of membership (*Q*) of each predefined provenance in each of three clusters computed by Bayesian approach (Pritchard et al. 2000)

[a] Walnut provenances sampled along Royal Tratturo Candela-Pescasseroli

Castanea sativa (Martin et al. 2010). As reported for these species, the high levels of allelic richness (R_s) typical of microsatellite loci positively influenced the identity probabilities (PI_{unb}, PI_{sib}), which were very low for each walnut provenance. These positive features make the SSR loci used in this work as suitable tools for genetic structure analysis.

AMOVA revealed that the distribution of molecular variance of Italian walnut germplasm was similar to that observed in many long-lived woody plant species. In contrast with most herbaceous species, forest trees generally display high within-population diversity and low differentiation among populations (Hamrick et al. 1992; Müller-Starck et al. 1992). The observed genetic differentiation among the 29 Italian provenances and Sorrento variety (F_{ST}=0.05369) was similar to F_{ST} values estimated using isozymes (Malvolti et al. 1993; 1997; Fornari et al. 1999) and RAPDs markers (Ferrazzini et al. 2007). These previous investigations led to the single conclusion that there is no clear bio-geographic pattern among Italian walnut provenances. Nevertheless, there are reasons to be skeptical of this finding. Jost (2008) has seriously questioned the role of F_{ST} and its derivates as measures of genetic differentiation between subpopulations. Like many
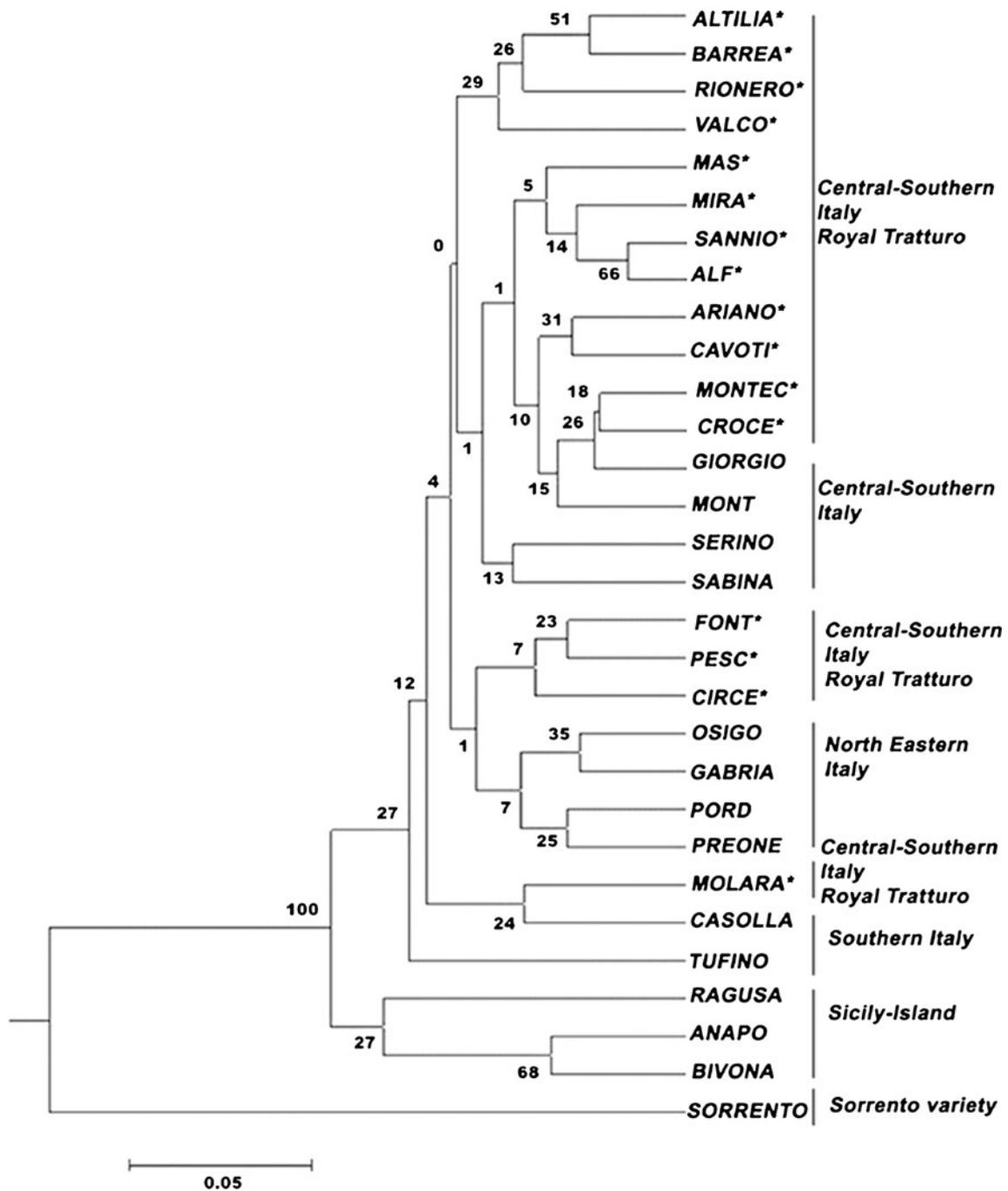
Fig. 5 Unweighted pair-group method with arithmetic mean (UPGMA) cluster analysis based on Nei's (1972) genetic distance and 1,000 bootstraps for 29 Italian Persian walnut populations and Sorrento non-clonal variety. The number near each node represents the percentage of times when the node occurred among 1,000 boothstraps. (*asterisks*) Walnut populations sampled along Royal Tratturo Candela-Pescasseroli

other authors, he noticed that estimates of F_{ST} approached zero when gene diversity is high, which is often the case for microsatellite data, even if subpopulations are completely differentiated. The dependence of F_{ST} (or its relatives) on within-population heterozygosity can lead to an underestimation of the true level of genetic differentiation. Jost

(2008) quantified genetic diversity in terms of effective number of alleles rather than heterozygosity, and proposed an alternative measure. As suggested by Heller and Siegismund (2009), we also computed the unbiased estimator of Jost's (D_{est}) for each locus across all populations and over all loci. As expected, D_{est} was always higher

Fig. 6 Allele frequency distribution at locus WGA69 for three clusters inferred by STRUCTURE analysis
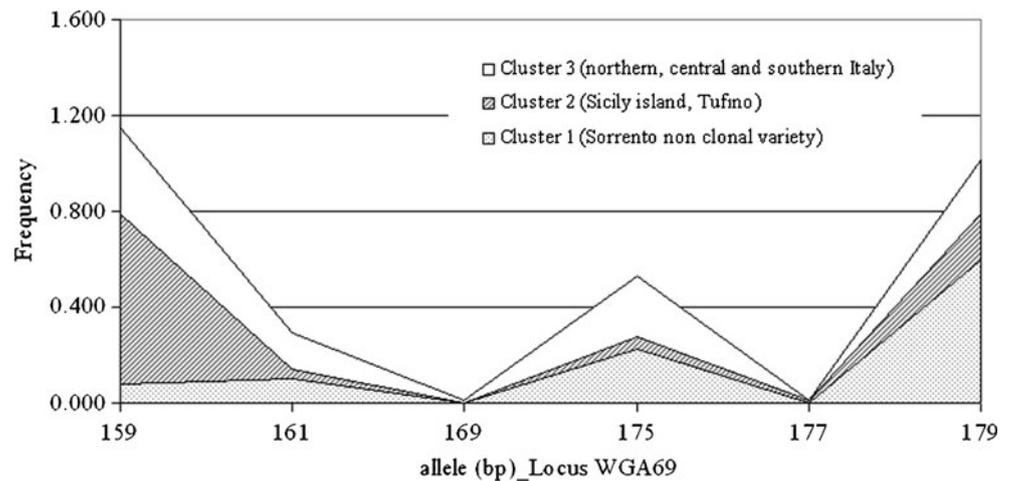
than F_{ST}, with a mean D_{est} value of 0.076 among walnut provenances. A low-to-moderate level of genetic differentiation among populations is commonly observed in forest tree species, probably as a result of long-distance zygotic and gametic gene dispersal (Krutovsky et al. 2009). In addition, domesticated plant species might also show low levels of molecular differentiation because of their long exposure to human selection and human-mediated dispersal of selected genotypes. As suggested by Fornari et al. (1999), humans rapidly spread Persian walnut in Europe and caused a considerable erosion of genetic resources through domestication bottlenecks. In multipurpose species as *J. regia*, the progressive selection of valuable genotypes for nut production and removal of vigorous trees with high wood quality negatively affected tree density, natural regeneration (dysgenic effect), and genetic diversity. Our data seem to confirm this point, with a statistically significant Wilcoxon's signed-rank test for a recent reduction of effective population size in about 95% of walnut provenances. Their expected heterozygosity (H_E) was significantly greater than the expected equilibrium gene diversity (H_{EQ}) under the TPM mutation model. Similarly, the M-ratio test of Gaza and Williamson (2001) detected signatures of bottlenecks in all walnut groups, although this result should be considered with caution, as the sample size was often less than the 25 individuals per provenance recommended for this analysis.

In this study, Bayesian analysis divided our walnut samples into three main clusters although significant isolation by distance was not observed. This result indicated that the genetic discontinuity we observed in Italian *J. regia* germplasm did not fit spatial boundaries defined a priori. As shown in Fig. 4, 20 genotypes of SORRENTO (cluster 1) diverged from the remaining trees. The SORRENTO trees showed the lowest allelic richness (Rs=2.67), an excess of homozygotes (F_{IS}=0.152), and significant deviation from Hardy–Weinberg expectations

for allele frequency. All these factors indicated that inbreeding was occurring in SORRENTO. Using the Wilcoxon's signed-rank test, we also found that the SORRENTO genotypes showed a slight deficiency of heterozygosity under the TPM model. As explained by Cornuet and Luikart (1996), a population that has recently suffered a severe reduction in size and has subsequently expanded for several generations without immigration is characterized by a heterozygosity deficiency across all loci. These data are consistent with the putative origin of the Sorrento (non-clonal) variety. Among the Italian walnut varieties, Sorrento is the oldest and most famous. It originated from the Sorrento peninsula (Campania region) but is cultivated in the entire Italian peninsula. According to Foroni et al. (2007), in the last century, "some farmers appear to have performed a strong selection in favor of a small group of walnut trees to improve yield and nut quality on Sorrento peninsula".

The second cluster contained *J. regia* samples from three distinct geographic sites on the island of Sicily. Similar results were observed for other tree species, including white oak species such as *Quercus robur* L., *Quercus petraea* (Matt.) Liebl., *Quercus pubescens* Willd. s.l., and *Quercus frainetto* Ten. (Fineschi et al. 2002), *Cupressus sempervirens* (Bagnoli et al. 2009) and *Olea europea* (Belaj et al. 2007). In these species, the application of chloroplast and nuclear microsatellite markers revealed genetic uniqueness in Sicilian populations (fixed haplotypes). Their separation from the mainland Italian populations may be the result of a bottleneck, possibly associated with founder effects, and prolonged geographic isolation. Fineschi et al. (2002) showed that for oak populations, gene flow by pollen between Sicily and Southern Italy may have been prevented by the natural barrier represented by the Ionic sea. As suggested by Dupanloup et al. (2002), at a fine spatial scale, the most prevalent cause of genetic differentiation is the isolation-by-distance process (IBD; Wright 1943). IBD leads

to the formation of local pedigree structures as a result of limited gene dispersal (caused, for example, by loss of a seed dispersal mechanism or geographic barriers) and local random genetic drift. Episodes of human transportation of seeds and the lack of gene flow by pollen with mainland populations might affect the genetic structure of Sicilian walnut germplasm, causing the shifts in allele frequencies observed in this study (Fig. 6). As reported by Gaza and Williamson (2001), the random sampling process of genetic drift changes allele frequency and causes some alleles to be lost and others to become fixed. At locus WGA69, Sicilian walnut trees exhibited high frequency for the 159-bp allele (71.3%), which was almost fixed (80%) in plants from the Anapo Valley (ANAPO), near the ancient Rocky Necropolis of Pantalica. These data also affected the null allele frequency estimated at locus WGA69 over all provenances (F_{nul}= 0.2064). The inbreeding coefficient (F_{IS}) at this locus was neither systematically positive nor exceptionally high for any Italian provenance, leading us to question the presence of null alleles at WGA69. Removing the Sicilian provenances from the calculation, the null allele frequency estimator declined from 0.2064 to 0.160, similar to F_{nul} value computed by Dangl (2005) in 46 walnut accessions at locus WGA69. Locus WGA69 did not fit neutral expectations when analyzed using Beaumont and Nichols (1996), however, and it was identified as an outlier locus by FDIST2 software. Lewontin and Krakauer (1973) and Luikart et al. (2003) observed that selection and mutation have locus-specific effects while genetic drift and gene flow act at a genome-wide scale. By analyzing the number of alleles per locus and the large number of alleles in common between *J. nigra* and *J. regia*, Pollegioni et al. (2009a) postulated a low mutation rate at locus WGA69. Several studies report that the interruption of perfect microsatellites is related to DNA stability in the region (Taylor et al. 1999). These authors suggested that the purity of a repeat region influences its mutation rate and, consequently, the level of polymorphism in SSR loci. Interrupted microsatellites, such as WGA69, appear to have lower mutation rates than pure microsatellites. As reported by Cornuet and Luikart (1996), this feature makes WGA69 a useful marker for detecting a bottleneck. Storz (2005) also indicated that the risk of detecting false positives is high using Beaumont and Nichols (1996) because bottlenecks can produce effects similar to natural selection. In this study, the atypical behavior of WGA69 may be a consequence of its low rate of mutation and a human-mediated domestication bottleneck.

Finally, the third main cluster inferred by STRUCTURE grouped 13 walnut provenances located in northeastern, central, and southern Italy, as well as the 16 provenances sampled along Royal Tratturo Candela–Pescasseroli. Further substructures could not be identified within cluster 3, and no strong correlation was found between the genetic variation of neutral SSR markers and rural history or human migrations along ancient Royal Tratturo Candela–Pescasseroli. These results lead to the conclusion that on the Italian mainland, walnut is represented by a single, dominant lineage. Nevertheless, the UPGMA dendrogram based on Nei's genetic distances distinguished the northeastern provenances and slightly clustered 12 of 16 provenances of Royal Tratturo Candela–Pescasseroli. As postulated by Victory et al. (2006) in *J. nigra* populations, the anthropogenic bottleneck associated with rural activity was probably not severe enough to have much impact upon genetic differentiation. On the other hand, human activities like the large scale transfer of walnuts across the peninsula might be sufficient to counteract the genetic drift that would be expected in demographically reduced *J. regia* populations. The genetic distinctiveness of ALTILIA germplasm (relative high Q_3 value) may confirm this point; 20 trees that were centuries old were sampled along Royal Tratturo Candela–Pescasseroli, which crosses the archeological site of the Roman town of *Saepinum/Altilia*. This area, surrounded by ancient walls, was preserved by the Italian Ministry of Heritage, and in the last century, human management of walnut was not permitted there.

In conclusion, a detailed knowledge of spatial population structure of the existing genetic resources of Italian walnut may be a crucial guide to correct conservation management decisions. In spite of low levels of molecular differentiation among populations, this study showed that it is possible to identify genetically homogenous groups of walnut individuals at higher hierarchical levels using SSR markers. In addition, two contrasting driving forces for genetic differentiation of Italian walnut were identified: selection (domestication bottlenecks), and intense inter-regional transfer of plant material. These forces might play a relevant role in shaping the genetic diversity of Persian walnut in Italy. Finally, it is essential to extend the SSR-based analysis of walnut germplasm to populations in the native range of *J. regia*, especially those lying within important ancient trade routes between Asia and Europe (e.g., China, Pakistan, Uzbekistan, Georgia, Greece) in order to infer the pattern of spatial genetic structure connected with putative geographic and cultural barriers in this wide area.

names is for the information and convenience of the reader and does not imply official endorsement or approval by the United States Department of Agriculture or the Forest Service of any product to the exclusion of others that may be suitable.

References

Avram M (2009) The legacy of transhumance in national park of Abruzzo, Lazio and Molise (PNALM): rediscovery and exploitation. Geo J Tour Geosites 4:153–159

Bagnoli F, Vendramin GG, Buonamici A, Doulis G, Gonzàlez-Martìnez C, La Porta N, Magri D, Raddi P, Sebastiani F, Fineschi S (2009) Is *Cupressus sempervirens* native in Italy? An answer from genetic and paleobotanical data. Mol Ecol 18:2276–2286

Bai WN, Zeng YF, Zhang DY (2007) Mating patterns and pollen dispersal in a heterodichogamous tree, *Juglans mandshurica* (*Juglandaceae*). New Phytol 176:699–707

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. Mol Ecol 11:155–165

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol 13:969–980

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proc Roy Soc Lond B Biol Sci 263:1619–1626

Belaj A, Munoz-Diez C, Baldoni L, Porceddu A, Barranco D, Satovic Z (2007) Genetic diversity and population structure of wild olives from the North-Western Mediterranean assessed by SSR markers. Ann Bot 100:449–458

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001–2014

Cournet JM, Piry S, Luikart G, Estoup A, Soligna M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genetics 153:1989–2000

Crawford NG (2009) SMOGD: software for the measurement of genetic diversity. Mol Ecol Resources, doi:10.1111/j.1755-0998.2009.02801.x

Dangl GS, Woeste K, Ardhya MK, Koehmstedt A, Simon C (2005) Characterization of 14 microsatellite markers for genetic analysis and cultivar identification of walnut. J Am Soc Hortic Sci 130:348–354

Di Vaio C, Minotta G (2005) Indagine sulla coltivazione del noce da legno in Campania. Forest@ 2:185-197

Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. Mol Ecol 11:2571–2581

Dutech C, Joly HI, Jarne P (2004) Gene flow, historical population dynamics and genetic diversity within French Guianan populations of a rainforest tree species, *Vouacapoua americana*. Heredity 92:69–77

El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree (*Argania spinosa* (L.) Skeels) endemic to Morocco. Theor Appl Genet 92:832–839

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47–50

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecol Notes 7:574–578

Ferrazzini D, Monteleone I, Lecce F, Belletti P (2007) Variabilità genetica del noce comune (*Juglans regia*) in Piemonte. Forest 4:386–394

Fineschi S, Taurchini D, Grossoni P, Petit RJ, Vendramin G (2002) Chloroplastic DNA variation of white oaks in Italy. Forest Ecol Manag 156:103–114

Fornari B, Cannata F, Spada M, Malvolti ME (1999) Allozyme analysis of genetic diversity and differentiation in European and Asiatic walnut (*Juglans regia* L.) populations. Forest Genet 6:115–127

Fornari B, Malvolti ME, Taurchini D, Fineschi S, Beritognolo I, Maccaglia E, Cannata F (2001) Isozyme ad organellar DNA analysis of genetic diversity in natural/naturalised European and Asiatic walnut (*Juglans regia* L.) populations. Acta Hortic 544:167–178

Foroni I, Woeste K, Monti LM, Rao R (2007) Identification of "Sorrento" walnut using simple sequence repeats (SSRs). Genet Resour Crop Evol 54:1081–1094

Gaza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. Mol Ecol 10:305–318

Gunn BF, Aradhya M, Salick JM, Miller AJ, Yongping Y, Lin L, Xian H (2010) Genetic variation in walnuts (*Juglans regia* and *J. sigillata; Juglandaceae*): Species distinctions, human impacts, and the conservation of agrobiodiversity in Yunnan, China. Am J Bot. doi:10.3732/ajb.0900114

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. New Forest 6:95–124

Heller R, Siegismund HR (2009) Relationship between three measures of genetic differentiation G_{ST}, D_{EST} and G_{SR} how wrong have we been? Mol Ecol 18:2080–2083

Heuertz M, Carnevale S, Fineschi S, Sebastiani F, Hausman JF, Paule L, Vendramin GG (2006) Chloroplast DNA phylogeography of European ashes, *Fraxinus* sp. (*Oleaceae*): roles of hybridization and life history traits. Mol Ecol 15:2131–2140

Hu LJ, Uchiyama K, Shen HL, Saito Y, Tsuda Y, Ide Y (2008) Nuclear DNA microsatellites reveal genetic variation but a lack of phylogeographical structure in an endangered species, *Fraxinus mandshurica*, across North-East China. Ann Bot 102:195–205

Huntley B, Birks HJB (1983) An atlas of past and present pollen maps for Europe: 0-13000 Years Ago. Cambridge University Press, New York

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodaly in analysis of population structure. Bioinformatics 14:1801–1806

Jost L (2008) G_{ST} and its relatives do not measure differentiation. Mol Ecol 17:4015–4026

Kalinowski ST (2004) Counting alleles with rarefaction: private alleles and hierarchical sampling designs. Conserv Genet 5:539–543

Kimura M, Otha T (1978) Stepwise mutation model and distribution of allelic frequencies in finite populations. PNAS 75:2868–2872

Krutovsky KV, Clair JBS, Saich R, Hipkins VD, Neale DB (2009) Estimation of population structure in coastal Douglas-fir [*Pseudotsuga menziesii* (Mirb.) Franco var. *menziesii*] using allozyme and microsatellite markers. Tree Genet Genomes 5:641–658

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymphisms. Genetics 74:175–195

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2009) The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet 4:981–994

Malvolti ME, Paciucci M, Cannata F, Fineschi S (1993) Genetic variation in Italian populations of *Juglans regia*. Acta Hortic 311:86–91

Malvolti ME, Beritognolo I, Spada M, Cannata F (1997) Ricerche sulle risorse genetiche e sulla biologia riproduttiva in *Juglans regia* L. in Italia mediante marcatori molecolari. Ann. Ist. Sper. Selv. XXV-XXVI: 9–34.

Malvolti ME, Pollegioni P, Mapelli S, Cannata F (2010) Research of *Juglans regia* provenances by molecular, morphological and biochemical markers: a case study in Italy. Bioremediation, Biodiversity and Bioavailability 4:84–92

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

Marandola D, Malvolti ME, Farina ME, Tognetti R (2008) Biodiversity and rural development: the case-study of the "Shepherd's walnut" or "Tratturo's walnut". An action model for sustainable rural development shaped on the peculiar feature of a rural area. EFI 2008 Annual Conference week scientific seminar, 15–20 September, Orvieto, Terni, Italy.

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihhod-based paternity inference in natural populations. Mol Ecol 7:639–655

Martin MA, Mattioni C, Cherubini M, Turchini D, Villani F (2010) Genetic diversità in European chestnut populations by means of genomic and genic microsatellite markers. Tree Genet Genomes 6:735–744

McGranahan G, Leslie C (2009) Breeding walnuts (*Juglans regia*). In: Jain SM, Pryadarshan PM (eds) Breeding Plantation Tree Crop: Temperate Species. Springer Science + Business Media, LLC, pp 249–273

Meyer FG (1980) Carbonized food plants of *Pompeii*, *Herculaneum*, and the Villa at Torre Annunziata. Econ Bot 34:401–437

Müller-Starck G, Baradat P, Bergmann F (1992) Genetic variation within European tree species. New Forest 6:23–47

Nei M (1972) Genetic distance between populations. Am Nat 106:283–292

Paetkau D, Calvert W, Stirling W, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears. Mol Ecol 4:347–354

Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. Mol Ecol 13:55–65

Palasciano I (1999) The long grass routes (Tratturi and pastors of the South) Capone (Ed), pp 1–88.

Peakall R, Smouse PE (2005) GenAlEx V6: Genetic Analysis in Excel. Population genetic software for teaching and research. The Australian National University, Canberra, Australia. http://www.anu.edu.au/BoZo/genAlEx/.

Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Mol Ecol 14:689–701

Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer program for detecting recent reductions in effective population size from allele frequency data. J Hered 90:502–503

Pollegioni P, Woeste K, Major A, Scarascia Mugnozza G, Malvolti ME (2009a) Characterization of *Juglans nigra* (L.), *Juglans regia* (L.) and *Juglans x intermedia* (Carr.) by SSR markers: a case study in Italy. Silvae Genet 58:68–78

Pollegioni P, Woeste K, Scarascia Mugnozza G, Malvolti ME (2009b) Retrospective identification of hybrodogenic walnut plants by SSR fingerprinting and parentage analysis. Mol Breed 24:321–335

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

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. PNAS 94:9197–9201

Robichaud RL, Glaubitz JC, Rhodes OE Jr, Woeste K (2006) A robust set of black walnut microsatellites for parentage and clonal identification. New Forest 32:179–196

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145:1219–1228

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462

Soulsbury CD, Iossa G, Edwards KJ, Baker PJ, Harris S (2007) Allelic dropout from a high-quality DNA source. Conservat Genet 8:733–738

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptative population divergence. Mol Ecol 14:671–688

Takezaki N, Nei M, Tamura K (2010) POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. Mol Biol Evol 27:747–752

Taylor JS, Durkin JMH, Breden F (1999) The death of a microsatellite: a phylogenetic perspective on microsatellite interruptions. Mol Biol Evol 16:567–572

Van der Schoot J, Pospíšková M, Vosman B, Smulders MJM (2000) Development and characterization of microsatellite markers in black poplar (*Populus nigra* L.). Theor Appl Genet 101:317–322

Victory E, Glaubitz JC, Rhodes OE Jr, Woeste K (2006) Genetic homogeneity in *Juglans nigra* (*Juglandaceae*) at nuclear microsatellites. Am J Bot 93:118–126

Wang H, Pei D (2008) Genetic diversity and structure of walnut populations in Central and Southwestern China revealed by microsatellite markers. J Am Soc Hortic Sci 133:197–203

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370

Woeste K, Burns R, Rhodes O, Michler C (2002) Thirty polymorphic nuclear microsatellite loci from black walnut. J Hered 93:58–60

Wright S (1943) Isolation by distance. Genetics 28:114–138

Yeh FC, Yang RC, Boyle TJB, Ye ZH, Mao JX (1997) POPGENE, the user-friendly shareware for population genetic analysis. Edmonton, Molecular Biology and Biotechnology Centre, University of Alberta, Canada