Australian Journal of Crop Science AJCS

AJCS 10(6):857-863 (2016) ISSN:1835-2707
DOI: 10.21475/ajcs.2016.10.06.p7520

The first high-density sequence characterized SNP-based linkage map of olive (*Olea europaea* L. subsp. *europaea*) developed using genotyping by sequencing

Annalisa Marchese[1]*, Francesco Paolo Marra[1], Tiziano Caruso[1], Khethani Mhelembe[2,3], Francesca Costa[1], Silvia Fretto[1], Daniel James Sargent[3]

[1]Department of Agricultural and Forest Sciences (SAF), University of Palermo, Viale delle Scienze, Edificio 4 Ingresso H - 90128 Palermo, Italy
[2]ARC Infruitec-Nietvoorbij, Private Bag X5026, Stellenbosch 7599, South Africa
[3]Research and Innovation Centre, Fondazione Edmund Mach, Via E. Mach 1, 38010, San Michele all'Adige, Italy

***Corresponding author: annalisa.marchese@unipa.it**

Abstract

A number of linkage maps have been previously developed in olive; however, these are mostly composed of markers that have not been characterized at the sequence level, supplemented with smaller numbers of microsatellite markers. In this investigation, we sought to develop a saturated linkage mapping resource for olive composed entirely of sequence characterized markers. We employed genotyping by sequencing to develop a map of a F_2 population derived from the selfing of the cultivar Koroneiki. The linkage map contained a total of 23 linkage groups comprised of 1,597 tagged SNP markers in 636 mapping bins spanning a genetic distance of 1189.7 cM. An additional 6,658 segregating SNPs were associated with the 23 linkage groups identified but their marker order was not determined in this investigation. The SNP markers sequences were submitted to NCBI database. The linkage map produced will be an invaluable resource for the study of tree habit and vigour traits segregating in the progeny, and will assist to anchor and orientate sequencing scaffolds from future genome sequencing efforts.

Key words: F_2 progeny, genome anchoring; next-generation sequencing; olive; plant habit; self-compatibility, tree architecture.
Abbreviations: AFLP Amplified Fragment Length Polymorphism; cM centiMorgan; DArT Diversity Array Technology; GBS Genotyping by Sequencing; LG Linkage group; RADseq Restriction site Associated DNA sequencing; RAPD Random Amplified Polymorphic DNA; RFLP Restriction Fragment Length Polymorphism; SNPs Single Nucleotide Polymorphisms; SSR Simple Sequence Repeats.

Introduction

Olive (*Olea europaea* L. subsp. *europaea*) is one of the most extensively cultivated fruit crops in the Mediterranean climatic zones of world, and has been of great importance as the primary source of fats for the populations of those areas for millennia. However, until recently, very few formal olive breeding programs existed and the majority of the existing cultivated varieties were selected empirically within their region of cultivation (Atienza et al., 2014). To date, only a very small number of varieties have been released from formal breeding programs (e.g. 'Chiquitita'; Rallo et al., 2008a). There exists a high degree of natural diversity in olive for agronomically important characteristics, and due to its long juvenility period, and highly heterozygous nature, systematic breeding would benefit greatly from the use of markers to assist in the selection of germplasm with superior fruit and oil quality, as well as yield, disease resistance characteristics, self-fertility, and compact, low vigour tree habit suitable for mechanical harvesting and pruning.

Mature olive trees are usually between 6-8 meters in height, but some cultivars are able to grow up to 26 meters and to reach a crown diameter of 20 meters under certain environmental conditions ('Ottobratica' and 'Sinospolese' in Calabria; Abenavoli and Marcianò 2013). Olive plants also have a large canopy diameter, usually between 5-6 meters. A reduction in tree size is commercially desirable, in order to reduce the cost of production through the mechanization of harvesting and pruning, and thus intensive research in olive is currently being undertaken to understand and control tree vigour and canopy architectural traits (Rallo et al., 2008a, b; Hammami et al., 2011; Ben Sadok et al., 2013). Among international olive germplasm, 'Arbequina' (Rallo et al., 2008a), 'Arbosana' (Del Rio et al., 2005) and 'Koroneiki' (Androulakis and Loupassaki 1990) have close to the commercially ideal habit; they display medium-low ('Koroneiki') and low vigour ('Arbequina' and 'Arbosana'), as well as early and constant bearing, whilst maintaining commercially acceptable fruit yield and oil quality. Olive has a nuclear DNA content ranging from 2.90 ± 0.020 to 3.07 ± 0.018 pg/2C (Loureiro et al., 2007), giving an estimated genome size in base pairs of ~1.5 Gb, based on the genome size of *F. vesca* of 254 Mb for a DNA content of 0.52 pg/2C (Shulaev et al., 2011). Cytological studies have shown that the species is diploid and has a chromosome complement of (2n = 2× =46) (Breviglieri and Battaglia 1954; Minelli et al., 2000). Numerous molecular marker classes have been studied in olive, but to date they have mainly been applied to diversity studies and the characterisation of core collections (e. g. Fabbri et al., 1995, Cipriani et al., 2002, Angiolillo et

al., 1999, Belaj et al., 2003, Belaj et al., 2012, Marra et al., 2013, Caruso et al., 2014), and only a few studies have employed markers to develop linkage maps for the species.
The first genetic linkage maps of olive were constructed primarily from arbitrarily primed AFLP and RAPD markers, enriched with a smaller number of sequence-characterized markers such as RFLPs and microsatellites (SSRs) (De la Rosa et al., 2003; Wu et al., 2004; Khadari et al., 2010; El Aabidine et al., 2010). The linkage maps were developed from F_1 progenies and spanned between 879 cM and 3,823.2 cM. More recently, Domínguez-García et al. (2012) reported a linkage map composed of diversity array technology (DArT) markers. DArT markers exploit reduced-representation genome enrichment using restriction enzymes, and micro-array technology, to produce genotyping assays that can interrogate large numbers of polymorphisms found in the gene pool from which the genotyping panel was developed (Jaccoud et al., 2001). The linkage map of Domínguez-García et al. (2012) contained a total of 257 markers spanning 1,205.1 cM on the female map, and 392 markers spanning 1,639.3 cM on the male map across 23 linkage groups on each map. The linkage map was subsequently phenotyped for fruit and vigour traits and QTL were identified for a number of economically important characters (Atienza et al., 2014). Recently, studies have used short read sequencing platforms to characterize the olive genome including the sequencing of the olive transcriptome (Alagna et al., 2009, Munõz-Mérida et al., 2013, Kaya et al., 2013), which led to the identification of abundant SNP markers that were successfully used to discriminate Turkish olive cultivars (Kaya et al., 2013). Additionally, Barghini et al. (2014), in the context of a project to develop a whole genome sequence for olive, studied the repetitive portion of the olive genome and determined that ~31% of the olive genome is composed of tandem-repeats. Despite an on-going project to sequence the olive genome (*Olea* project;http://genomes.cribi.unipd.it/olive/wordpress/project/) , there remains a need for a saturated, sequence characterized linkage map for the species, and the SNP data that exists in public databases for olive have so far not been exploited for linkage map development. The emergence of 'next-generation' genotyping tools, such as RADseq (Miller et al., 2007; Baird et al., 2008) and genotyping by sequencing (GBS) (Elshire et al., 2011) has revolutionized the development of molecular markers in experimental and natural populations. By exploiting restriction enzyme recognition sites, reduced representation genomic libraries enriched for the coding portions of the genome have been generated and sequenced in multiplex using the Illumina sequencing platforms to produce datasets containing many thousands of sequence characterized SNP markers. This has led to the development of saturated SNP-based linkage maps of numerous plant species (e.g. Troggio et al., 2007; Ward et al., 2013). Using the GBS protocol of Elshire et al. (2011), we have developed a SNP-based linkage map of a F_2 population derived from the selfing of the self-fertile, low vigour cultivar Koroneiki (Marchese et al., 2016). The population was raised primarily in order to investigate the genetic control of plant habit. The map and its downstream use in QTL analyses of plant vigour traits and genome sequence scaffold anchoring are discussed.

Results

Genotyping by sequencing

A total of 85 progeny established as juvenile plants were available for genetic analysis. From the 85 progeny and nine parental genotype replicates of 'Koroneiki' from which the F_2 progeny was raised, sequence data were recovered for a total of 81 progeny and eight parental replicates; the datasets ranged in size from 1.4 Gbytes to 58 Mbytes. Following analysis with Stacks (Catchen et al., 2011), a total of 25,802 tagged SNPs were called in at least one of the seedling progeny. Of the 81 seedlings, 63 returned data for at least 50% of the segregating SNPs, the remainder of which were removed from the analysis and not considered further. This number of seedlings can be considered statistically sufficient for the construction of a linkage map, since the map of *Fragaria* (Sargent et al., 2011) was successfully built with a similar number of progeny and reduced representation bin map populations of as little as six individuals have provided accurate, if not precise mapping positions for markers (Sargent et al., 2008, Howad et al., 2005).

Analysis of SNP segregation

Of the 25,802 tagged SNPs identified, 11,812 loci contained data for at least 50% of the 63 progeny and a chi-squared analysis was performed on these SNPs to identify those loci for which segregation did not deviate significantly from the expected 1:2:1 Mendelian segregation ratio. A total of 1,730 SNPs returned segregation ratios that did not exhibit segregation distortion, and these were used for subsequent linkage map construction.

Linkage map development

Since the phase of the 1,730 segregating SNPs was initially unknown, they were scored as AB×AB markers segregating in a F_1 population (i.e. in a 1:2:1 Mendelian ratio) and were grouped using JOINMAP 4.1. A total of 29 linkage group clusters were recovered containing more than ten markers, but following removal of groups that contained all markers mapping within less than 5 cM of each other within a discrete linkage group, a total of 23 linkage groups remained. Marker ordering was initially determined using the Maximum Likelihood mapping function of JOINMAP 4.1, and imputation was used to 'correct' any unlikely genotype calls. The data were then reanalyzed using regression mapping and once marker order was determined, all redundant markers were added back to the linkage map using a custom Python script provided by D. Michelletti (personal communication). The resultant linkage map contained a total of 1,597 tagged SNP markers in 636 mapping bins distributed throughout 23 discrete linkage groups, spanning a total genetic distance of 1189.7 cM (Fig 1).

SNP marker coverage

Linkage group (LG) 1 was the longest group at 115.2 cM, whilst LG23 was the shortest at 14.7 cM. Linkage group 6 contained the most markers (165) in the most mapping bins (58), whilst LG19 contained the fewest (ten markers in five mapping bins). Overall, marker density on the map averaged one marker every 0.75 cM, and one mapping bin every 1.87 cM. The 1,597 robustly mapped markers were used as seed markers to identify additional SNPs exhibiting segregation distortion that were associated with the 23 identified linkage groups. Screening of all 11,812 markers for which data for at least 50% of seedlings was available revealed an additional 6,658 segregating SNPs associated with the 23 linkage groups identified. The greatest number of additional markers was associated with LG2 (659), whilst LG23 was associated with the fewest additional markers (59).

Table 1. Summary data for the 23 linkage groups of the 'Koroneiki' F_2 linkage map detailing linkage group length in centi-Morgans (cM), the number of markers not displaying segregation distortion mapped on each linkage group, the number of non-redundant mapping bins per linkage groups, and the number of additional SNP markers associated with each group.

Linkage group	LG length (cM)	No. markers	No. mapping bins	Additional markers
G1	115.164	81	50	429
G2	107.531	115	52	659
G3	92.266	164	54	493
G4	91.38	49	29	460
G5	90.657	97	46	323
G6	87.041	165	58	363
G7	63.022	95	36	257
G8	60.286	49	27	417
G9	58.889	151	43	338
G10	55.585	98	33	230
G11	45.521	82	29	404
G12	43.888	28	19	207
G13	36.818	113	30	171
G14	34.652	22	15	133
G15	31.713	16	7	221
G16	27.517	29	17	108
G17	27.26	29	15	97
G18	23.121	93	18	168
G19	21.587	10	5	303
G20	21.05	54	20	94
G21	20.184	11	11	84
G22	19.905	17	10	640
G23	14.647	29	12	59
Total	1,189.68	1,597	636	6,658

Table 1 details the lengths of each of the 23 linkage groups, the numbers of markers and mapping bins they contain and the number of additional markers associated with each linkage group. Supplementary Tables S1 and S2 contain all SNPs mapped or associated with one of the 23 linkage groups recovered, in addition to the sequence tags for each SNP, and the position of SNPs within the tags; NCBI accession numbers of SNP sequences are reported in the supplementary Table S3.

Discussion

Recent advances in 'next-generation' genotyping techniques that exploit short-read sequencing technology have permitted the rapid characterization of many thousands of SNP markers in segregating progenies (Elshire et al., 2011) and the development of dense, saturated linkage maps (e.g. Ward et al., 2013), such technology has shown clear advantages in other species and as such, will enable the rapid development of genomics resources in olive, that have to date lagged behind other plant crop species. In this investigation, we have presented the first comprehensive, saturated linkage map for olive composed of sequence-characterized SNP markers, using the GBS approach. The F_2 'Koroneiki' map spans the expected 23 chromosomes associated with the base chromosome number for the species, and covers a total genetic distance of 1,189.7 cM (Fig 1). Previous linkage maps for olive described in the literature have been reported with varying lengths. The first linkage map of De la Rosa et al. (2003), derived from a F_1 progeny, spanned a total of 2,765 cM across 39 linkage groups in one parent and 2,445 cM across 30 linkage groups in the other. The RAPD-based map of Wu et al. (2004) contained 23 and 27 linkage groups in the female and male linkage maps, covering 798 cM and 759 cM respectively. More recent linkage maps of olive have spanned 36 and 31 linkage groups and 2,210 cM and 1,966 cM for the female and male maps respectively in the study of Khadari et al. (2010), and 40 and 38 linkage groups, covering 1,547.4 cM and 1,428 cM for the female and male maps respectively in the study of El Aabidine et al. (2010). The most comprehensive linkage map published to date prior to the map presented here was the DArT-based linkage map of Domínguez-García et al. (2012). In that study, the authors used a genome enrichment approach similar to that employed by GBS; however, two enzymes were used. The resultant DArTs were scored in the 91 seedlings and parents of the cross and the data were supplemented with some SSR markers. The resultant linkage map comprised 23 significant linkage groups for each parent, containing 257 markers (24 SSRs and 237 DArT markers) in 125 mapping bins on the female map, and 392 markers (23 SSRs and 369 DArT markers) in 204 mapping bins on the male map. The maps covered a genetic distance of 1,205.1 cM and 1,639.3 cM for the female and male maps respectively. The SNP-based F_2 linkage map presented in this investigation is somewhat shorter than the initial AFLP and RAPD based linkage maps for the species, but comparable to the more recent maps, and almost identical in length to the DArT-based linkage map constructed using similar marker technology. The large differences in linkage map lengths reported for initial maps for the species is perplexing, but speculatively, may be the result of errors in the scoring of arbitrarily-primed markers such as RAPDs, which would have introduced additional spurious recombination into the segregation dataset, and inflated genetic distances. Whilst 1,189.7 cM likely underestimates the total map length, due to our initial criterion for only selecting markers that did not exhibit significant segregation distortion for map construction, it contains linkage groups with a similar length distribution as the DArT map of Domínguez-García et al. (2012) and so can be considered to cover the majority of the olive genome. Although the DArT linkage map approach affords many advantages, it is limited in as much as it is based on micro-array technology and probe hybridization. As such, whilst mapping DArT markers does not require *a priori* information about informative SNPs for array development, it

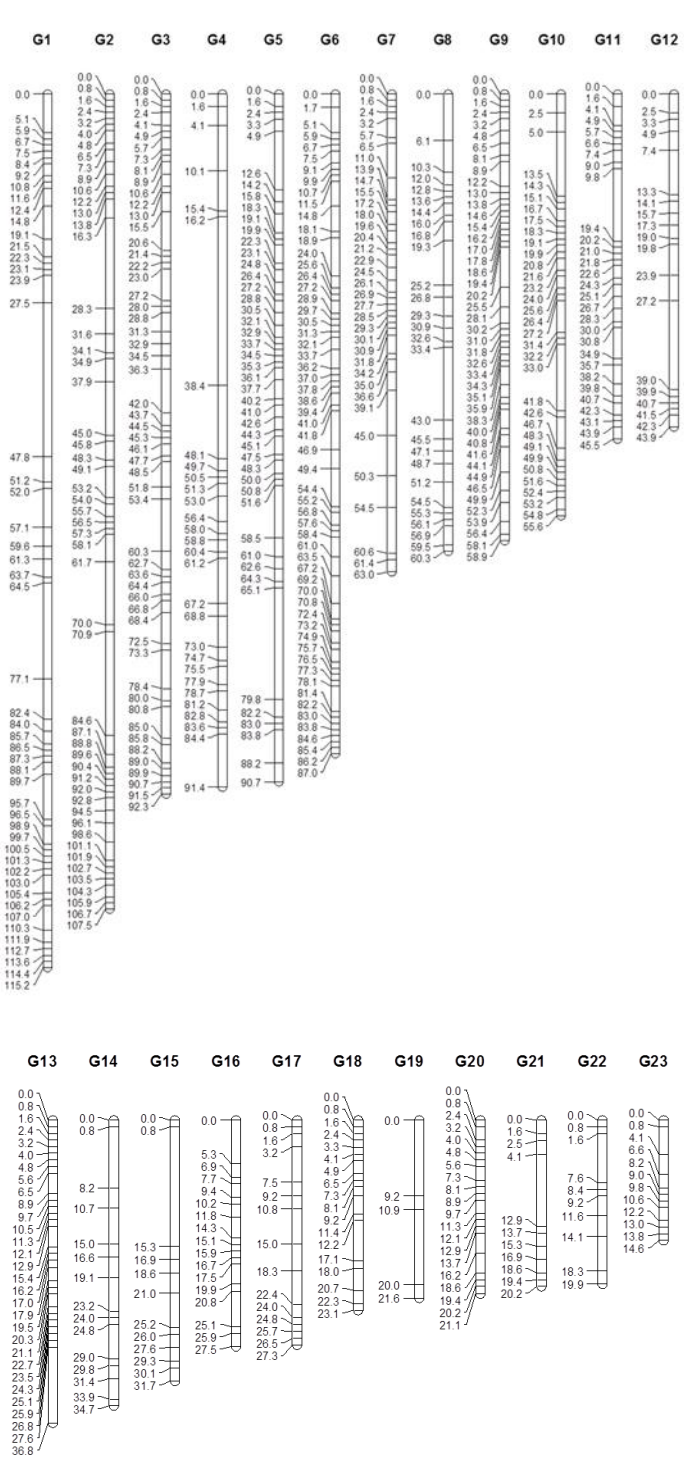

Fig 1. A genetic linkage map of the 'Koroneiki' F_2 mapping progeny composed of 1,579 sequence characterized SNP markers in 636 mapping bins. Marker loci are represented by horizontal lines and linkage group lengths are shown to scale in centi-Morgans.

requires the array-bound SNPs to be present in the genome under interrogation. Whilst the use of GBS to score segregating SNPs results in a high percentage of missing values in the dataset produced, following an imputation strategy as has been done here, permits the development of a reliable linkage map framework containing a large numbers of markers relative to other genotyping methodologies. Indeed, the 'Koroneiki' map presented herein contains over twice as many imputed segregating anchor loci in almost twice as many mapping bins in a comparable progeny size to the DArT map of Domínguez-García et al. (2012), and ten times as many potentially useful genetic markers in the full dataset, with the advantage that all markers placed on the linkage map have a sequence-characterized tag associated with them (Supplementary Tables S1 and S2). Thus, the 'Koroneiki' linkage map is likely a representation of the vast majority of the olive genome and the research community working on the development of olive maps would benefit from our sequence-characterized SNP markers, available in the Tables S1 and S2 and submitted to NCBI database (Table S3). The *Olea* project aims to sequence the olive genome using a strategy of shot-gun sequencing and assembly of

short-read sequence data from the Illumina and 454 sequencing platforms (http://genomes.cribi.unipd.it/olive/wordpress/project/). Recently, Barghini et al. (2014) reported that ~31% of the olive genome is composed of tandem-repeats from six main families, two of which were not previously characterized, and it has been estimated that in excess of 70% of the olive genome is composed of repetitive DNA sequence (Barghini et al., 2015). Thus the central challenge of whole-genome shotgun sequence assembly using short-read technologies is resolving repetitive sequences (Schatz et al., 2010), since often these sequences are longer than the read length of the technologies used for sequencing. The extent of repetitive DNA sequence within the olive genome has resulted in initial assemblies that are highly fragmented (Barghini et al., 2015). As such, a genetic linkage map resource, enriched for the coding portion of the genome (Elshire et al., 2011) and containing thousands of markers, would be a valuable resource for scaffold anchoring and orientation. The 'Koroneiki' linkage map resource containing some 6,658 sequence-characterized SNP markers associated with the expected 23 linkage groups/chromosomes of olive, will be a valuable resource for assisting in making sense of the complexity of the olive genome. The genetic variability in vegetative habit has been a focus of study in olive since genotypes with a compact and upright habit could be used in high density planting systems. The traits in olive that affect the tree growth habit and vigour are complex and believed to be under polygenic control. Although a F_1 linkage map of the cross 'Oliviere' × 'Arbequina' (Khadari et al., 2010) was reported to segregate with a high degree of variability for characters related to the tree form and vigour (Ben Sadok et al., 2013), so far no character has been mapped regarding the tree architecture and/or vigour in olive. Ben Sadok et al. (2013) reported that some 'Oliviere' × 'Arbequina' progenies were less vigorous than the low vigour parent 'Arbequina', and thus useful for future selection programs, and that tree form variables presented medium broad sense heritability values. The progeny for which we present a genetic linkage map in this report segregates for a high range of plant habits and for vigour, and 20% of offspring seems to have a dwarf or "brachitic habit", maybe due to the effect of transgressive segregation, not previously reported in literature, since this is the first F_2 progeny so far obtained in olive.

Though at the present time, the plants are still in a juvenile phase, it is expected that once they reach maturity, the progeny will be useful for the identification of QTL associated with plant habit and vigour characteristics.

Conclusion

We have presented the first olive linkage map based on a F_2 population derived from the selfing of the low vigour cultivar Koroneiki, developed using GBS, without the use of a reference genome. The map contains 1,597 tagged SNP markers, covering a total genetic distance of 1189.7 cM over 23 linkage groups, the expected number for *Olea europea* (2n = 2x = 46) and can be considered the most informative olive map so far published. As the sequencing tags and marker positions are made available in the NCBI database, information can be used for aiding the complex olive genome sequencing and assembly and to update other less saturated olive maps reported in the literature. For example, transforming these SNPs in cleaved amplified polymorphic sequence (CAPS) markers could be a cost effective method to create anchoring markers between this and other olive maps. This map will also form the starting point for a study of tree habit, vigour traits and self-compatibility that are expected to segregate in the 'Koroneiki' progeny. Since this is a F_2 progeny composed of individuals homozygous for different loci it would be possible to study how this affects certain phenotypes and the phenomenon of transgressive segregation.

Materials and Methods

Plant material

A F_2 progeny was raised for linkage map construction derived from the self-fertilization of the olive cultivar Koroneiki, suitable for the study of the genetic control of tree habit. The cultivar was chosen for its medium-low vigour, required for developing high density planting systems, self-compatibility and for good quality oil. Genomic DNA was extracted from young fresh leaves of the progeny and nine replicates of the parental genotype using the DNeasy mini kit (Qiagen) according to the manufacturer's recommendations. DNA was quantified using a Nanodrop spectrophotometer and Qubit 2.0 fluorometer according to the manufacturers' recommendations, and diluted to 20 ng / ul for library preparation. The genetic correspondence of offspring was confirmed using SSR markers and any out-crosses were discarded (data not shown).

Library construction, sequencing and SNP identification

Genotyping-by-sequencing (GBS) was performed following the protocol described in Elshire et al. (2011) using the *Ape*KI restriction enzyme and adaptor dilutions as described by Ward et al. (2013). A total of 100 ng of DNA from eight parental replicates derived from independent DNA extractions and 85 progeny were digested with *Ape*KI and ligated to 1.8 ng of barcoded adaptor using T4 DNA ligase (manufacturer). Ligation reactions for each progeny genotype were performed separately, samples were pooled and PCR amplification was performed on the pooled library. The library was purified using the QiaQuick PCR purification system (Qiagen) following which the library was sequenced using single-end sequencing on the HiSeq2000 sequencing platform (Illumina, San Diego, USA) with 101 cycles. Sequenced samples were de-multiplexed as described by Elshire et al. (2011) and data were analyzed using Stacks (Catchen et al., 2011) running Stacks de novo with default settings. The SNP segregation data recovered from Stacks were filtered for progeny individuals containing less than 50% missing data, and SNP loci containing less than 50% missing data.

Marker segregation, imputation and linkage map construction

Since segregating SNP data were un-phased, markers were scored as segregating AB×AB in a F_1 progeny and linkage mapping was performed using JOINMAP 4.1 (Kyazma). Initially only marker data for which segregation data did not deviate significantly from a 1:2:1 Mendelian ratio were used for linkage map construction. Following an initial round of grouping with a minimum LOD threshold of 4.0, markers were ordered in linkage groups using the Maximum Likelihood mapping function, and imputation was performed following the methods of Ward et al. (2013). The imputed dataset was used to construct a linkage map using regression mapping with the Kosambi mapping function, following standard mapping criteria, i.e. marker placement was determined using a minimum LOD score threshold of 3.0, a

recombination fraction threshold of 0.35, ripple value of 1.0, jump threshold of 3.0, and a triplet threshold of 5.0. Data for redundant marker loci removed automatically by JOINMAP 4.1 were replaced using a custom Python script provided by D. Michelletti (personal communication), and the linkage map presented was visualised using MapChart 2.0 (Voorrips 2002). Subsequently, imputed markers were used as seed markers to determine the number of markers exhibiting segregation distortion could be associated with a linkage group on the map. Putative linkage between mapped markers and additional segregating SNPs was determined using the grouping function of JOINMAP 4.1 (Kyasma).

Acknowledgements

We acknowledge financial support from PON02_ 00451_3361785 project "Dieta Mediterranea e Salute (DI.ME.SA)".

References

Abenavoli LM, Marcianò C (2013) Technical and economic analysis of alternative pruning systems in high dimensions olive trees in Calabria. Agron Res. 11:7-12.

Alagna F, D'Agostino N, Torchia L, Servili M, Rao R, Pietrella M, Giuliano G, Chiusano ML, Baldoni L, Perrotta G (2009) Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. BMC Genomics. 10:399.

Androulakis I, Loupassaki MH (1990) Studies on self-fertility of some olive cultivars in the area of Crete. Acta Hortic. 286:159-162.

Angiolillo A, Mencuccini M, Baldoni L (1999) Olive genetic diversity assessed using amplified fragment length polymorphisms. Theor Appl Genet. 98:411-421.

Atienza SG, De la Rosa R, Leon L, Martin A, Belaj A (2014) Identification of QTL for agronomic traits of importance for olive breeding. Mol Breeding. 34:725-737.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE. 3(10):e3376.

Barghini L, Natali L, Cossu RM, Giordani T, Pindo M, Cattonaro F, Scalabrin S, Velasco R, Morgante M, Cavallini A (2014) The peculiar landscape of repetitive sequences in the olive (*Olea europaea* L.) genome. Genome Biol Evol. 6:776-791.

Barghini E, Natali L, Giordani T, Cossu RM, Scalabrin S, Cattonaro F, Šimková H, Vrána J, Doležel J, Morgante M, Cavallini A (2015) LTR retrotransposon dynamics in the evolution of the olive (*Olea europaea*) genome. DNA Res. 22:91-100.

Belaj A, Satovic Z, Cipriani G, Baldoni L, Testolin R, Rallo L, Trujillo I (2003) Comparative study of the discriminating capacity of RAPD, AFLP and SSR markers and their effectiveness in establishing genetic relationships in olive. Theor Appl Genet. 107:736-744.

Belaj A, Dominguez-García MC, Atienza SC, Urdíroz NM, De la Rosa R, Satovic Z, Martín A, Kilian A, Trujillo I, Valpuesta V, Del Río C (2012) Developing a core collection of olive (*Olea europaea* L.) based on molecular markers (DArTs, SSRs, SNPs) and agronomic traits. Tree Genet Genomes. 8:365-378.

Ben Sadok I, Moutier N, Garcia G, Dosba F, Grati-Kamoun N, Rebai A, Khadari B, Costes E (2013) Genetic determinism of the vegetative and reproductive traits in a F1 olive tree progeny. Tree Genet Genomes. 9:205-221.

Breviglieri N, Battaglia E (1954) Ricerche cariologiche in *O. europaea* L. Caryologia. 61:271-283.

Caruso T, Marra FP, Costa F, Campisi G, Macaluso L, Marchese A (2014) Genetic diversity and clonal variation within the main Sicilian olive cultivars based on morphological traits and microsatellite markers. Sci Hortic-Amsterdam. 180:130-138.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping Loci de novo from short-read sequences G3: Genes, Genomes, Genetics (Bethesda). 1:171-182.

Cipriani G, Marrazzo MT, Marconi R, Cimato A, Testolin R (2002) Microsatellite markers isolated in olive are suitable for individual fingerprinting and reveal polymorphism within ancient cultivars (*Olea europaea* L.). Theor Appl Genet. 104:223-228.

De la Rosa R, Angiolillo A, Guerrero C, Pellegrini M, Rallo L, Besnard G, Bervillé A, Martin A, Baldoni L (2003) A first linkage map of olive (*Olea europaea* L.) cultivars using RAPD, AFLP, RFLP and SSR markers. Theor Appl Genet. 106:1273-1282.

Del Río C, Caballero JM, García-Fernández MD (2005) Vigor (Banco de germoplasma de Córdoba). In: Rallo L et al. (eds) Variedades de olivo en España, Libro II: Variabilidad y selección. Coed. Junta de Andalucía/MAPA/Mundi-Prensa, Madrid.

Dominguez-Garcia MC, Belaj A, De la Rosa R, Satovic Z, Heller-Uszynska K, Kilian A, Martin A, Atienza SG (2011) Development of DArT markers in olive (*Olea europaea* L.) and usefulness in variability studies and genome mapping. Sci Hortic-Amsterdam. 136:50-60.

El Aabidine AZE, Charafi J, Grout C, ì Doligez A, Santoni S, Moukhli A, Jay-Allemand C, El Modafar C, Khadari B (2010) Construction of a genetic linkage map for the olive based on AFLP and SSR markers. Crop Sci. 50:2291-2302.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE. 6(5) e19379.

Fabbri A, Hormaza JI, Polito VS (1995) Random amplified polymorphic DNA analysis of olive (*Olea europea* L.) cultivars. J Am Soc Hortic Sci. 120:538-542.

Hammami SBM, León L, Rapoport HF, De la Rosa R (2011) Early growth habit and vigour parameters in olive seedlings. Sci Hortic-Amsterdam. 129:761-768.

Howad W, Yamamoto T, Dirlewanger E, Testolin R, Cosson P, Cipriani G, Monforte AJ, Georgi L, Abbott AG, Arús P (2005). Mapping with a few plants: using selective mapping for microsatellite saturation of the *Prunus* reference map. Genetics.171:1305-1309.

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res. 29 (4):e25

Kaya HB, Cetin O, Kaya H, Sahin M, Sefer F, Kahraman A, Tanyolac B (2013) SNP discovery by Illumina-based transcriptome sequencing of the olive and the genetic characterization of Turkish olive genotypes revealed by AFLP, SSR and SNP markers. PLoS ONE. 8(9):e73674.

Khadari B, El Aabidine AZ, Grout C (2010) A genetic linkage map of olive based on amplified fragment length polymorphism, inter-simple sequence repeat and simple sequence repeat markers. J Am Soc Hortic Sci. 135:548-555.

Loureiro J, Rodriguez E, Costa A, Santos C (2007) Nuclear DNA content estimations in wild olive (*Olea europaea* L. ssp. *europaea* var. *sylvestris* Brot.) and Portuguese cultivars of *O. europaea* using flow cytometry. Genet Resour Crop Ev. 54:21-25.

Marchese A, Marra FP, Costa F, Quartararo A, Fretto S, Caruso T (2016) An investigation of the self- and inter-incompatibility of the olive cultivars 'Arbequina' and 'Koroneiki' in the Mediterranean climate of Sicily. Aust J Crop Sci. 10:88-93.

Marra FP, Caruso T, Costa F, Di Vaio C, Mafrica R, Marchese A (2013) Genetic relationships, structure and parentage simulation among the olive tree (*Olea europaea* L. subsp. *europaea*) cultivated in Southern Italy revealed by SSR markers. Tree Genet Genomes. 9:961-973.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. 17:240-248.

Minelli S, Maggini F, Gelati MT, Angiolillo A, Cionini PG (2000) The chromosome complement of *Olea europaea* L.: characterization by differential staining of the chromatin and *in-situ* hybridization of highly repeated DNA sequences. Chromosome Res. 8:615-619.

Munoz-Merida A, Gonzalez-Plaza JJ, Canada A, Blanco AM, Garcia-Lopez Mdel C, Rodríguez JM, Pedrola L, Sicardo MD, Hernández ML, De la Rosa et al. (2013) De novo assembly and functional annotation of the olive (*Olea europaea*) transcriptome. DNA Res. 20:93-108.

Rallo P, Barranco D, De la Rosa R, Leon L (2008a) 'Chiquitita' olive. HortScience. 43:529-531.

Rallo P, Jimenez R, Ordovas J, Suarez MP (2008b) Possible early selection of short juvenile period olive plants based on seedling traits. Aust J Agr Res. 59:933-940.

Sargent DJ, Cipriani G, Vilanova S, Gil-Ariza D, Arús P, Simpson DW, Tobutt KR, Monfort A (2008) The development of a bin mapping population and the selective mapping of 103 markers in the diploid *Fragaria* reference map. Genome. 51:120-7.

Sargent DJ, Kuchta P, Lopez Girona E, Zhang H, Davis TM, Celton J-M, Marchese A, Korbin M, Folta KM, Shulaev V, Simpson DW (2011) Simple sequence repeat marker development and mapping targeted to previously unmapped regions of the strawberry genome sequence. The Plant Genome. 4:165-177.

Schatz MC, Delcher AL, Salzberg SL (2010) Assembly of large genomes using second-generation sequencing. Genome Res. 20:1165-1173.

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP et al. (2011) The genome of woodland strawberry (*Fragaria vesca*). Nat Genet. 43:109-116.

Troggio M., Malacarne G, Coppola G, Segala C, Cartwright DA, Pindo M, Stefanini M, Mank R, Moroldo M, Morgante M, Grando MS, Velasco R (2007) A dense single-nucleotide polymorphism-based genetic linkage map of grapevine (*Vitis vinifera* L.) anchoring Pinot Noir bacterial artificial chromosome contigs. Genetics. 176:2637-2650.

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered. 93:77-78.

Ward JA, Bhangoo J, Fernández-Fernández F, Moore P, Swanson JD, Viola R, Velasco R, Bassil N, Weber CA, Sargent DJ (2013) Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. BMC genomics. 14(1):2.

Wu SB, Collins G, Sedgley M (2004) A molecular linkage map of olive (*Olea europaea* L.) based on RAPD, microsatellite, and SCAR markers. Genome. 47:26-35.