Supplementary Material

A Codominant Molecular Diagnostic Toolkit for High-Throughput S-Locus Genotyping and Phenological Synchrony Assessment across the Olea europaea Complex

Roberto Mariotti[1], Simone Pizzi[1], Lorenzo Cruciani[1++], Claudio Cantini[2], Jalal Kassout[3], Pedro Valverde[4], Sara Oulbi[3], Chiara Piccini[2], Omar Abou-Saaid[3], Carlos Trapero[4], José Alberto Pereira[5], Nuno Rodrigues[5], Khalid Alhudaib[6], Bendjama Abdallah[7], Nursen Ustun[8], Murat Ozaltas[8], Hatem Elhadi Abukraa[10], Roya Nikbakht[1], Silvia Casciani[1], Muhammad Ajmal Bashir[1], Raied Abou Kubaa[11§], Maria Saponari[11], Soraya Mousavi[1*]

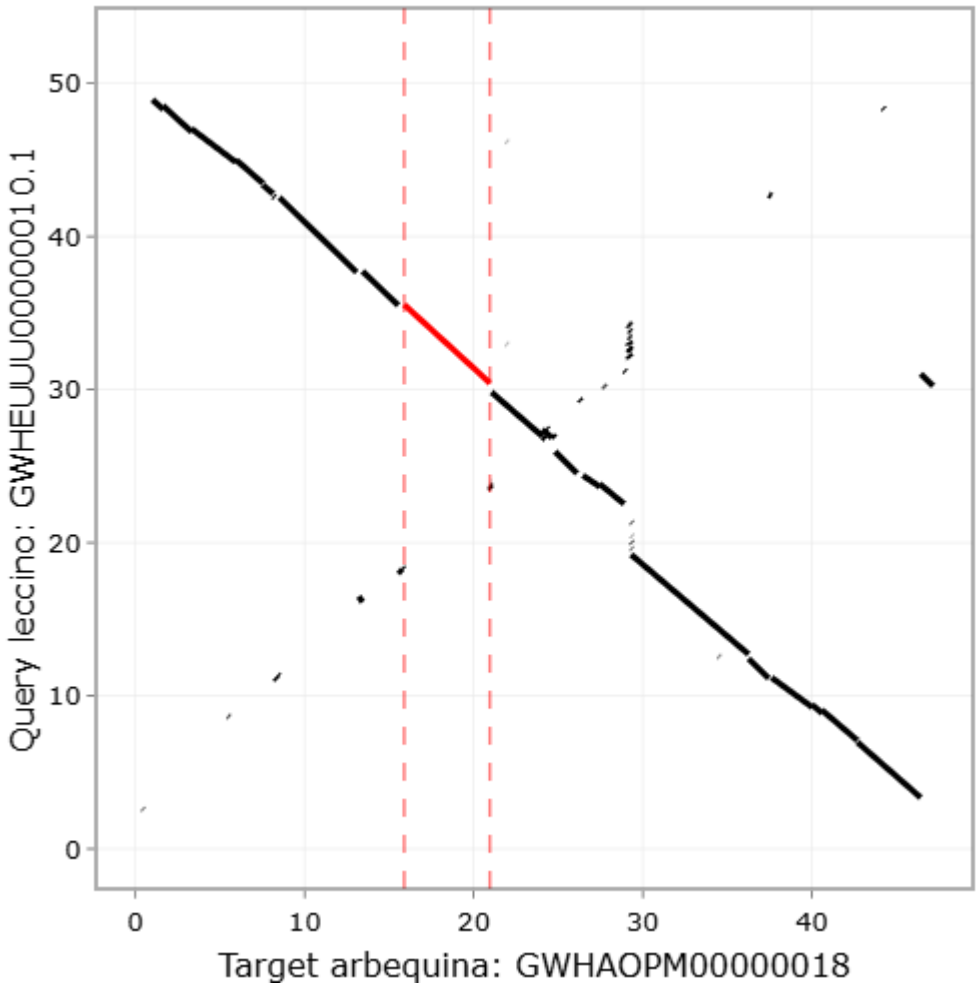


Fig. S1 Dot plot based on the .paf output file from minimap2, generated by aligning scaffold GWHEUUU00000010.1 from the Leccino assembly to Arbequina chromosome 18. The S locus is located within a large, 5.35 Mb region (red) with no alignment gaps (MAPQ = 60).

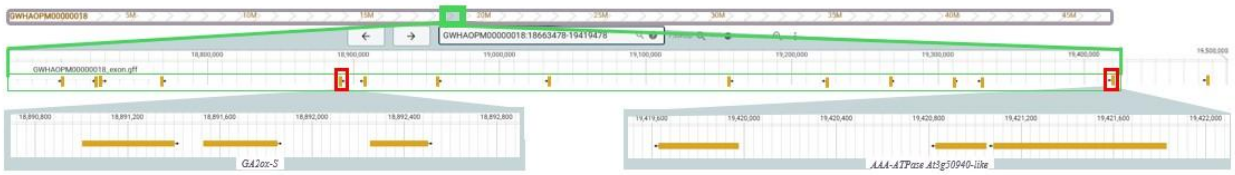


Fig. S2 Graphic representation, based on the Arbequina reference assembly, of the genomic insertion found in the dominant allele (S1) which defines the G1 reproductive group in olive. To better visualize the place where the new markers were developed (both DomRooM and CoRooM), the two regions that include the Ga2ox and the AAA-ATPase genes are bordered in red. Modified image from the JBrowse genome browser.

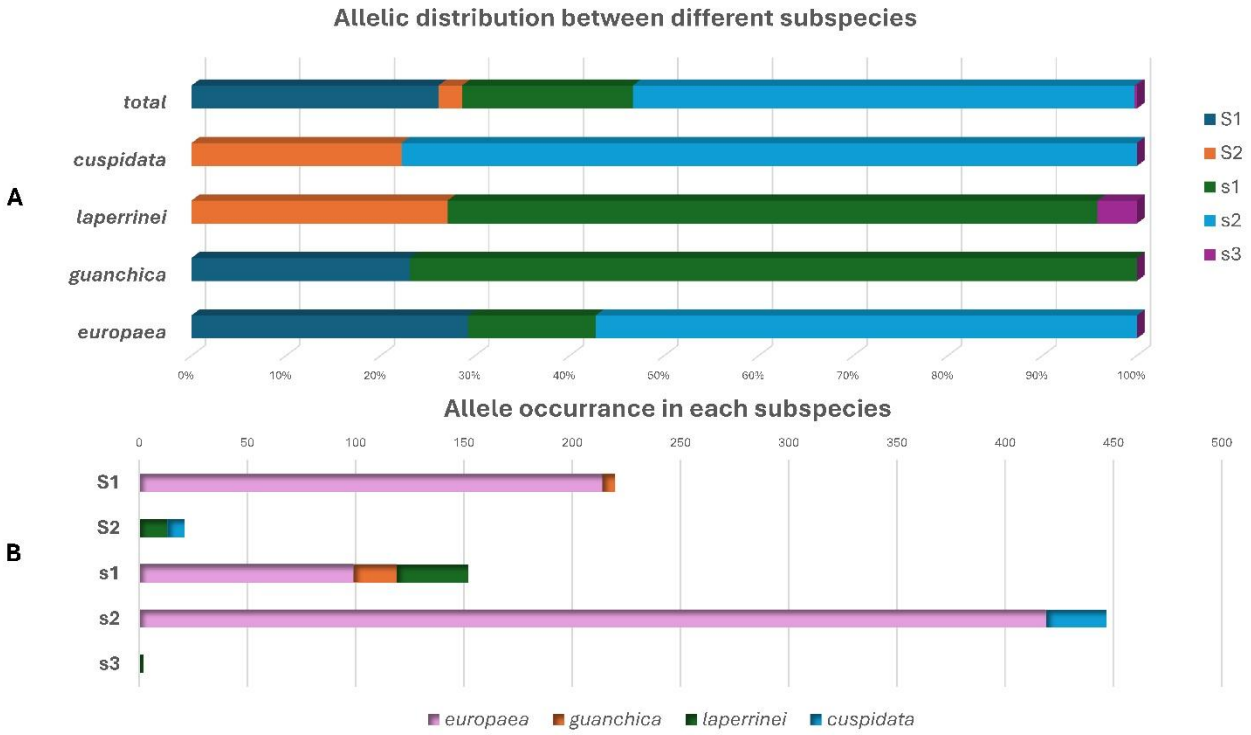


Fig. S3 Clustered bar charts showing (A) genotypic distribution across analysed subspecies, and (B) the frequency of occurrence for each allele within those subspecies.

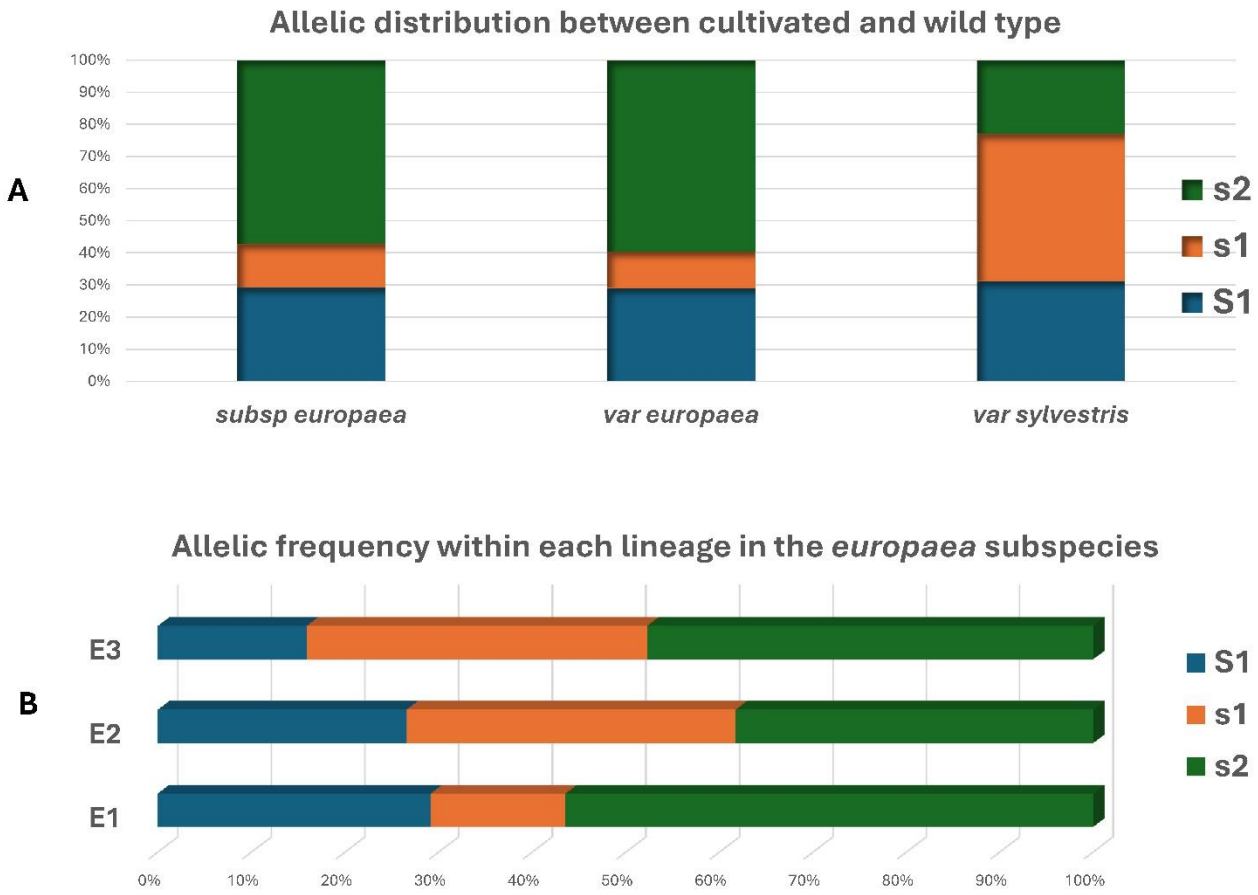


Fig. S4 Clustered bar showing A) allelic distribution of subspecies europaea genotypes analysed by CoRooM marker, including all samples and separated into var. europaea (cultivated varieties) and var. sylvestris (wild types); and B) the allelic frequency of each allele within the three-chloroplast lineage related to this subspecies.



KOR SELF 1



KOR SELF 7

Fig. S5 Two progenies derived from self-pollination of cv. Koroneiki that, after the analysis of the newly developed CoRooM marker, were homozygous for the dominant S1 allele.

Table S1. List of all analysed samples. Cultivar name as well as single code have been reported for each genotypes together with place of origin, synonym cases and DNA provenience**.**

Table S2. Comparative analysis between chromosome 18 (Chr18) of the cv Arbequina genome and chromosome 10 (Chr10) in cv Leccino.

Table S3. (A) Partial sequence of gibberellin-2-oxidase- S gene of *Olea europaea* species, which included the primer pair proposed in Castric et al., 2024 (yellow) and the new pair of primer developed in the present study highlighted in green. (B) Primer sequence (5'-3') of forward and reverse dominant marker. (C) Genomic position of the two amplicons based on the Leccino genome Lv et al., 2024.

Table S4. A) Genomic amplicons that include the codominant developed markers and the second exon of AAA-ATPase gene of *Olea europaea*. The primers are highlighted in green the common forward of both reproductive groups; in yellow the reverse specific for group 1; in light blue the reverse specific of group 2. B) The amplicon position inside two reference genomes, Leccino for G1 and Picual for G2 cultivars, and three more resequencing data mapped under the reference ones Frantoio and Farga for G1 and Chemlal de Kabylie for G2. C) Sequence of specific primers to amplify with reproductive group in olive (5'-3').

Table S5. Multiple alignment of nucleotide sequences of each allele detected by the amplification of the newly developed CoRooM marker for each subspecies and within the subsp. *europaea* ones in each different group (cultivar, wild type, subspecies).

Table S6. BLASTN alignment results for the primers and PCR amplicons against publicly available reference genomes. All primers showed 100% identity at the predicted genomic positions together with the amplicon lengths which were fully confirmed.

Table S7. Group assignment of each genotype amplified with dominant marker DomRooM.

Table S8. Chloroplast lineage assignment of each genotype.

Table S9. List of analyzed genotypes by using codominant marker, except for progenies. The columns of the results derived by dominant marker and the maternal inheritance are reported (Table S7 and S8)

Table S10. List of olive cultivar with their place of origin, reproductive group assignment and best linear unbiased prediction (BLUP) of full flowering date (FFD).