Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

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

Roberto Mariotti
National Research Council, Institute of Biosciences and Bioresources, 06128
Simone Pizzi
National Research Council, Institute of Biosciences and Bioresources, 06128
Lorenzo Cruciani
National Research Council, Institute of Biosciences and Bioresources, 06128
Claudio Cantini
National Research Council, Institute of BioEconomy, 58022
Jalal Kassout
Regional Center of Agricultural Research of Marrakech, National Institute of Agricultural Research, Avenue Ennasr, BP 415, Rabat Principale, 10090
Pedro Valverde
University of Cordoba (UCO)
Sara Oulbi
National Institute of Agricultural Research
Chiara Piccini
National Research Council, Institute of BioEconomy, 58022
Omar Abou-Saaid
National Institute of Agricultural Research
Carlos Trapero
University of Cordoba (UCO)
José Alberto Pereira
CIMO, LA SusTEC, Instituto Politécnico de Bragança
Nuno Rodrigues
CIMO, LA SusTEC, Instituto Politécnico de Bragança
Khalid Alhudaib
King Faisal University
Bendjama Abdallah
University of Skikda
Nursen Ustun

Plant Protection Research Institute, 35040, Bornova,

Murat Ozaltas

Olive Research Institute, Universite Cad. No:43, Bornova,

Hatem Elhadi Abukraa

Libyan Biotechnology Research Center,

Roya Nikbakht

National Research Council, Institute of Biosciences and Bioresources, 06128

Silvia Casciani

National Research Council, Institute of Biosciences and Bioresources, 06128

Muhammad Ajmal Bashir

National Research Council, Institute of Biosciences and Bioresources, 06128

Raied Abou Kubaa

Institute for Sustainable Plant Protection (IPSP)—National Research Council

Maria Saponari

Institute for Sustainable Plant Protection (IPSP)—National Research Council

Soraya Mousavi

soraya.mousavi@cnr.it

National Research Council, Institute of Biosciences and Bioresources, 06128

Research Article

Keywords: allelic profiling, diallelic self-incompatibility, floral phenology, germplasm diversity, Olea europaea

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11757927/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

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*]

[1] National Research Council, Institute of Biosciences and Bioresources, Perugia 06128, Italy; roberto.mariotti@cnr.it; simonepizzi@cnr.it; soraya.mousavi@cnr.it; rnikbakht2@gmail.com; silviacasciani@cnr.it; muhammadajmalbashir@cnr.it

[2] National Research Council, Institute of BioEconomy, Follonica 58022, Italy; claudio.cantini@cnr.it; chiarapiccini@cnr.it

[3] Regional Center of Agricultural Research of Marrakech, National Institute of Agricultural Research, Avenue Ennasr, BP 415, Rabat Principale, Rabat 10090, Morocco; jalal.kassout@inra.ma; sara.oulbi@inra.ma; omar.abousaaid@ced.uca.ma

[4] Agronomy Department, University of Cordoba (UCO), 14005 Cordoba, Spain; g82vacap@uco.es; g32trrac@uco.es

[5] CIMO, LA SusTEC, Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal; jpereira@ipb.pt; nunorodrigues@ipb.pt

[6] Plant Protection Dept. King Faisal University, P.O. 380, Al-Ahsa 31982, Saudi Arabia; kalhudaib@kfu.edu.sa; alhudaib@hotmail.com

[7] Department of Natural and Life Sciences, University of Skikda, Algeria; ab.bendjama@univ-skikda.dz

[8] Plant Protection Research Institute, 35040, Bornova, Izmir, Türkiye; nursen.ustun@tarimorman.gov.tr

[9] Olive Research Institute, Universite Cad. No:43, Bornova, Izmir Türkiye; murat.ozaltas@tarimorman.gov.tr

[10] Libyan Biotechnology Research Center, Tripoli, Libya; hatem.abukraa@btc.org.ly

[11] National Research Council, Institute for Sustainable Plant Protection (IPSP), Via Amendola 122/D, Bari, 70126, Italy;
maria.saponari@cnr.it
[++]Present address: Umbria Region, Phytosanitary service and production defense, Perugia 06128, Italy; lorenzocruciani@regione.umbria.it;
§Present address: Department of Plant Pathology, UC Davis, CA, USA; raboukubaa@ucdavis.edu

[*]Corresponding author:
E-mail address: soraya.mousavi@cnr.it; Tel: +39 075 5014812
https://orcid.org/0000-0002-4433-2512

Abstract

Background: Diallelic self-incompatibility (DSI) in olive (*Olea europaea* L.) enforces cross-pollination between two mating groups: Group 1 (dominant *S*) and Group 2 (recessive *s*). Commercial orchards regularly experience yield losses from incompatible pairings. Previously, assigning compatibility depended on laborious, season-constrained phenotypic assays (stigma tests and pollination trials) prone to environmental artifacts. The lack of molecular markers has hindered breeding, precluded seedling selection, and obscured whether the ~700 kb hemizygous *S*-locus supergene harbors recessive embryonic lethality. Furthermore, fruit-set requires overcoming a compound reproductive barrier: genetic compatibility and temporal flowering synchrony, both vulnerable to climate-driven phenological shifts.

Results: To establish a definitive diagnostic framework, we developed a dominant marker (DomRooM, targeting *GA2ox-S*) and a codominant marker (CoRooM, targeting the *AAA-ATPase* locus) within the *S*-locus supergene. Screening 905 accessions across 25 countries, including cultivars, monumental trees, wild forms (*var. sylvestris*), and three diploid subspecies (*guanchica*, *laperrinei*, and *cuspidata*), resolved five distinct alleles (*S1*, *S2*, s_1, s_2, s_3). While DomRooM provided binary G1/G2 separation, CoRooM distinguished heterozygous (*Ss*) from homozygous (*SS*) genotypes and eliminated false-negative G2 assignments. In self-pollinated progenies ('Frantoio' and 'Koroneiki'), CoRooM confirmed Mendelian 3:1 inheritance and identified viable adult *SS* individuals for the first time, demonstrating that the ~700 kb insertion lacks embryonic lethality. Subspecies exhibited geographic lineage sorting: *cuspidata* exclusively carried *S2* and s_2, *guanchica* possessed *S1* and s_1, and *laperrinei* harbored the private s_3 allele. Multi-year phenological modeling across seven seasons showed that while flowering onset is strongly heritable ($H^2 = 0.84$), flowering duration is governed by environmental variance (43.7%) and genotype-by-environment plasticity (43.2%), causing a severe shortage of synchronized G1 pollen donors for mid-to-late blooming Mediterranean cultivars.

Conclusions: This study replaces empirical phenotypic testing with the first definitive molecular framework for olive self-incompatibility. By proving the viability of homozygous (*SS*) genotypes and resolving geographic lineage sorting across the *Olea* complex, our findings redefine the evolutionary genetics of sporophytic DSI. Integrating codominant *S*-allele genotyping with flowering duration stability provides an essential foundation for early nursery selection, targeted breeding of late-blooming G1 cultivars, and climate-resilient orchard design.

Keywords: allelic profiling, diallelic self-incompatibility, floral phenology, germplasm diversity, *Olea europaea*

Introduction

The olive tree (*Olea europaea* L.) exhibits an enigmatic reproductive strategy characterized by a staggering overproduction of flowers. In a typical "on" year, a mature tree can produce up to 500,000 flowers, yet 98–99% of these ultimately abscise as unfertilized flowers or juvenile fruitlets within the first two months of blooming [1]. Rather than a mere consequence of resource limitation or simple pollination deficits, this massive abscission reflects a deeply ingrained evolutionary strategy [2]. While the transition from self-incompatibility (SI) to self-compatibility is a common and largely irreversible evolutionary shift in many flowering plants [3], SI systems are successfully maintained over long timescales in certain taxa [4]. This persistence is driven by species-level selection; lineages that enforce outcrossing through SI tend to exhibit higher net diversification rates and greater long-term resilience. Consequently, the olive's reproductive strategy, while seemingly inefficient, serves as a powerful mechanism to prevent inbreeding and promote the genetic diversity necessary for adaptation to environmental fluctuations and disease pressures [5, 6].

Recent genetic research has illuminated the precise mechanics of this system, identifying it as a homomorphic diallelic self-incompatibility (DSI) mechanism [6-8]. This system divides all cultivars into two distinct incompatibility groups (G1 and G2), dictating that successful fertilization can only occur through cross-pollination between members of opposite groups [9-10]. At the genetic level, mating group specificity is governed by two principal alleles: a dominant allele (S), which confers the G1 phenotype, and a recessive allele (s), which determines the G2 phenotype and, consequently, genotypes carrying the dominant S allele (Ss or SS) belong to Group 1, whereas genotypes homozygous for the recessive s allele (ss) belong to Group 2. The DSI system is an ancient, evolutionarily conserved mechanism present across all *Olea europaea* subspecies and wild relatives, a status corroborated by paternity tests in the Saharan olive (*O. e*. subsp. *laperrinei*), which show a complete absence of self-fertilization [5].

The *S*-locus controlling this trait has been mapped [10] and is now understood to be a classic supergene, a cluster of tightly linked genes inherited as a single unit [11-12]. This large (~700 kb), ancient genomic structure originated at least 51 million years ago and is characterized by a hemizygous indel, a structural variation that suppresses recombination and preserves the integrity of the incompatibility alleles. Importantly, although G2 genotypes (*ss*) lack this (~700 kb) genomic region on both homologous chromosomes, are fully viable, morphologically normal, and agronomically productive. This vegetative viability occurs because the (~700 kb) insertion specifically harbors reproductive mating-type regulators rather than essential housekeeping or vegetative genes [10]. Within this insertion, two key discoveries are paramount: first is the identification of *gibberellin-2-oxidase* (*GA2ox-S*), a gene involved in gibberellin hormone catabolism

whose expression is tightly associated with the SI phenotype and experimental modulation of gibberellin levels in floral buds can switch SI specificities, confirming this gene's central role in the incompatibility reaction [11]; second, the fact that brassinazole-resistant 1 (*BZR1*-S) is an inactivated pseudogene in the homomorphic olive [12], is absent from the DSI supergene of *Phillyrea angustifolia* [11], and has an active homolog potentially regulating floral morphology in the distylous relative *Chrysojasminum fruticans*, provides compelling molecular evidence that the homomorphic olive DSI system evolved from a heteromorphic ancestor by decoupling the incompatibility mechanism from floral morphology [12].

Despite the rigidity of the DSI system, variability in SI strength exists. Cultivars such as 'Frantoio' and 'Koroneiki' have historically been considered functionally self-fertile [13-15]. However, this phenomenon is better understood as Pseudo-Self-Compatibility (PSC), not the true absence of an incompatibility barrier, but a physiological "leak" in the sporophytic SI system [14, 16]. In recent research it was hypothesized that elevated temperatures can suppress rejection mechanisms, allowing incompatible pollen to fertilize ovules [2]. Molecularly, this instability could be linked to *GA2ox* and gibberellin fluctuations that mimic a compatible signature, a mechanism also demonstrated in *Ligustrum vulgare* [11]. Consequently, under open pollination, 'Frantoio' yields less than 10% self-fertilized embryos [9, 13], while, in the self-sterile 'Arbequina' cultivar, selfing often results in seed abortion and pseudo-drupes [17].

Beyond physiological leakiness, the DSI system is profoundly influenced by ecological factors. Successful gene flow strictly requires temporal overlap in flowering between compatible partners [18]. While genotype significantly influences flower quality [19] and flowering time exhibits high broad-sense heritability (H^2=0.84) across Mediterranean genetic clusters [20], phenology remains heavily environment-driven. Climate change, specifically milder winters that disrupt chilling requirements, causes earlier, prolonged, and increasingly asynchronous flowering [2]. This phenological asynchrony disrupts crucial cross-pollination dynamics [21], acting as a severe reproductive barrier given that synchrony often dictates pollen donor success more strongly than physical proximity [22-23]. With predictive models forecasting bloom advancements of 13–20 days by the end of the century [1], understanding these reproductive mechanisms is essential to overcome breeding bottlenecks and ensure population sustainability [24-25].

To cross these evolutionary and ecological barriers, integrating Marker-Assisted Selection (MAS) is vital [10, 12]. While a previously described dominant marker [11] effectively detects the presence or absence of specific *S*-locus fragments, it carries inherent risks of false negatives (due to primer polymorphisms), false positives (from contamination), the impossibility of verifying *SS* genotypes and crucially, lacks validation across wild subspecies. To address this gap, the present study aimed

to develop and validate novel codominant markers applicable to *Olea europaea* and its related diploid subspecies. This codominant approach is mandatory for distinguishing homozygous (*SS*) from heterozygous (*Ss*) genotypes, a crucial capability for tracking G1 intra-group crosses, verifying self-fertilized populations, and accurately characterizing the evolutionary conservation of the *S*-locus. By validating these markers across hundreds of genotypes spanning cultivated varieties, wild types, and three ancient subspecies, this work provides a high-resolution molecular toolkit to trace the evolutionary dynamics of the DSI system and facilitate the design of genetically compatible and phenologically synchronized plant populations.

Materials and Methods

Plant material and DNA extraction

This study investigated a comprehensive set of 905 genotypes. The sampling strategy was designed to achieve two main goals: to represent the widest possible range of currently cultivated global genotypes, and to expand the potential for future breeding programs by characterizing the reproductive groups of wild types and into other three related diploid subspecies. The final dataset includes: 491 cultivated genotypes, representing at least 607 distinct cultivars when known synonyms are considered [26], 36 self-crossed progenies ('Koroneiki' and 'Frantoio'), 112 progenies from a controlled cross between two compatible cultivars ('Leccino' and 'Dolce Agogia'), 73 monumental olive trees (defined by a trunk diameter >1.5 meters at 1 meter height), 106 wild olive trees (*Olea europaea* subsp. *europaea* var. *sylvestris*), 34 *Olea europaea* subsp. *guanchica*, 35 *Olea europaea* subsp. *laperrinei*, and 18 *Olea europaea* subsp. *cuspidata*. Plant material was obtained from three major international repositories: the World Olive Germplasm Bank (WOGB) at INRA (Marrakech, Morocco); the International Olive Germplasm Collection (IOGC) in Lugnano in Teverina (Terni, Italy); and the WOGB in Follonica (Grosseto, Italy). The latter two collections are managed by the National Research Council (CNR, Italy). Additionally, samples from ancient trees, wild types, and the *guanchica*, *laperrinei*, and *cuspidata* subspecies were collected through prospecting surveys in different countries (Table S1) performed by some of the co-authors of the present study.

The geographic scope of the study covers 25 countries, ranging from major global producers, such as Spain, Italy, Tunisia, Greece, Türkiye, Morocco, and Portugal, to regions rarely reported in olive biodiversity literature, including Libya, Saudi Arabia, Nepal, Ethiopia, and China.

Genomic DNA was extracted from leaf tissue using either the Zymo Research Mini Plant/Seed DNA Extraction Kit (following manufacturer instructions) or the CTAB method [27], with slight modifications [28]. DNA quantity and quality were assessed using a NanoPhotometer® NP80 (IMPLEN, DE).

Detecting chloroplast polymorphisms

All accessions were analysed by chloroplast markers [29], including SNPs, INDELs and length (mononucleotide microsatellites) polymorphisms able to distinguish the three main maternal chloroplast lineages [29]: E1, predominant in cultivated olives and Eastern Mediterranean wild types including the Iranian plateau; E2, typical of Central and Western Mediterranean wild olives; and E3, predominantly associated with Western Mediterranean wild olives; and they are also able to discriminate the difference between the subspecies included in the present study (*guanchica*, *laperrinei* and *cuspidata*) following the protocols already published (Mariotti *et al.*, 2023; Hosseini-Mazinani *et al.*, 2014; Mariotti *et al.*, 2010). Negative (no template DNA) and positive (cv. Frantoio DNA) controls were included in each amplification run, to detect non-specific products and verify the success of PCR reaction, respectively. To visualize the alleles the fluorescent fragments were resolved by capillary electrophoresis in an SeqStudio Genetic Analyzer (Thermo Fisher Scientific, US) using the internal GeneScan™ 500 LIZ Size Standard (Thermo Fisher Scientific, US). Sample analyses were performed using GeneMapper™ Software 6 (Thermo Fisher Scientific, US). To detect the single nucleotide polymorphism, the purified BigDye PCR products were sequenced on a SeqStudio Genetic Analyzer (Thermo Fisher Scientific). The SNP polymorphisms were observed using BioEdit v7.1.1 software (Ibis Biosciences, Carlsbad, CA).

Polymorphisms and primers not reported here were previously fully explained and published [10, 30-31].

Genome browser and sequence analysis

To identify the indel region distinguishing the two reproductive groups, we selected cvs Leccino (G1) [32] and Picual (G2) [33] as representative cultivars. The S locus was localized on chromosome 18 (Chr18) in the 'Arbequina' genome [34] and chromosome 10 (Chr10) in cv Leccino (Table S2; Figure S1). Comparative analysis indicates extensive large-scale homology between the assemblies; approximately 67% of cv Leccino Chr10 aligns to cv Arbequina Chr18, covering roughly 76% of the latter's total length. The specific region analysed in this study maps unambiguously to the 29,985,469–35,098,437 bp interval on cv Leccino Chr10. This 5.35 Mb alignment constitutes the third-highest scoring hit in the Minimap2 analysis and is supported by a maximum mapping quality score (MAPQ = 60). Within the cv Leccino Chr10 locus, we identified the *GA2ox* and *ATPase* genes at positions 31,955,359 and 31,585,397, respectively.

Primer Development and Amplification

Targeting the G1-Specific Fragment (*GA2ox-S*). Based on the recent findings of Castric *et al.* [11], a new primer pair designated as the **Dom**inant **R**epr**o**ductive **o**live **M**arker (DomRooM) targeting the *GA2ox-S* gene (Table S3). As demonstrated in the referenced study, this marker identifies a gene fragment related only to reproductive group G1; consequently, amplification indicates the presence of the G1 determinant, while absence suggests group G2. PCR amplifications were performed in a 12.5 µL reaction volume containing 25 ng of genomic DNA, 10× PCR buffer, 200 µM of each dNTP, 10 pmol of each forward and reverse primer, and 1 U of Q5 High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA). Each run included a negative control (no template) and a positive control (cv. Leccino, G1) to monitor non-specific products and verify reaction success. Thermal cycling conditions were 95°C for 5 min; 25 cycles of 95°C for 25 seconds, 60°C for 30 seconds, and 72°C for 25 seconds; followed by a final elongation at 72°C for 10 min. Amplicons were initially visualized by electrophoresis on 2% agarose gels. Subsequently, fluorescently labeled fragments were resolved by capillary electrophoresis on a SeqStudio Genetic Analyzer using the GeneScan™ 500 LIZ internal size standard. Data were analysed using GeneMapper™ Software 6 (Thermo Fisher Scientific, US).

Development of a Codominant Marker. Although the *GA2ox-S* marker showed a perfect match with phenotypes determined by the stigma test [6], its dominant nature (presence/absence) posed a risk of confounding G2 genotypes with PCR failures (false negatives) and it does not allow differentiation between Ss and SS, this latter, if exist, would be a sign of G1 cultivars intra-fertilization or G1 self-fertilization. To clarify the haplotype assignment, it is important to emphasize that the S allele is exclusive to the G1 group. In contrast, the s allele is present in both groups: as a single copy in G1 (heterozygous, *Ss*) and as two copies in G2 (homozygous, *ss*). To address this, we aimed to develop a marker capable of amplifying alleles specific to groups and S haplotypes. We screened boundaries at both sides of the hemizygote region specifically of the S haplotype and compared the flanking regions with their homologues in s haplotype [10-11] to identify conserved flanking regions that encompassed reproducible polymorphisms. A suitable target was identified in a region which includes the *AAA-ATPase* gene and a neighboring region partially described by Raimondeau *et al* [12]. Due to high sequence divergence between the alleles, a three-primer system was designed consisting of a single conserved forward primer (located in exon 3 of *AAA-ATPase*, 3'-5' direction) and two allele-specific reverse primers corresponding to S and s haplotypes respectively (Figure S2; Table S4). PCR was performed in a final volume of 12.5 µL containing 25 ng DNA, 10× PCR buffer, 200 µM each dNTP, 5 pmol of forward primer and 5 pmol of reverse primer specific for G1 allele and 5 pmol of the reverse primer related to G2 allele, together with 1U of Q5 High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA). Two positive controls (cv Leccino

for G1 and cv Dolce Agogia for G2) were included in each amplification. Cycling conditions: 95°C for 5 min; 30 cycles of 95°C for 25 seconds, 52°C for 30 seconds, and 72°C for 25 seconds; followed by a final elongation at 72°C for 30 min. The forward primer was labeled with FAM fluorescence. Amplicons were checked on 2% agarose gels and analysed via capillary electrophoresis on a SeqStudio Genetic Analyzer using the GeneScan™ 500 LIZ standard and GeneMapper™ Software 6.

Characterization of sequence variants detected on S and s haplotypes through Sanger Sequencing

The regions amplified by the newly developed primers were validated via Sanger sequencing to confirm base composition and amplicon length. A total of 11 fragments were selected for sequencing, representing five distinct allele lengths and including representatives from cultivars, wild types, and subspecies groups (Table S5).

PCR amplifications used 10 pmol of the common forward primer and the appropriate group-specific reverse primer (G1 or G2), utilizing the Q5 High-Fidelity DNA Polymerase protocol described above (annealing at 52°C; final elongation of 10 min). Fragments were sequenced in both directions using the ABI Prism BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific). The sequencing reaction, cycling conditions, sequencing and visualization were already described in the previous paragraph.

Moreover, to assess the specificity of the newly designed "CoRoom" primers and the PCR products corresponding to the alleles, we performed BLASTN searches (BLAST v2.16.0+) of these sequences against publicly available reference assemblies, specifically those of the 'Leccino' (PRJCA015700, [32]), and 'Picual' [35]' cultivars and the '*cuspidata*' (CNP0002655, [36]; GCA_050576095.1, GCA_023089605.1, [37]) and '*laperrinei*' (ASM4528209v1, ASM4528221v1, [11]) subspecies. The "-task blastn-short" option was used for primer sequences. In some reference assemblies, the deposited scaffolds are not labelled with chromosome numbers; however, their alignment to the reference 'Arbequina' chromosome 18 was verified (Table S6).

Statistical Analysis of Phenological Traits and Genotype-by-Environment (G × E) Interactions

Phenological records were collected at the World Olive Germplasm Collection (WOGB) in Marrakech across seven evaluation years (2014, 2015, 2016, 2017, 2018, 2019, and 2021). From the 229 cultivars evaluated in the germplasm collection, a subset of 154 cultivars for which complete multi-year phenological tracking information (both Stage 60 flowering onset and Stage 69 end of flowering) was available, has been selected [38-39] to evaluate flowering duration ($\Delta T = T_{Stage\ 69} - T_{Stage\ 60}$) and disentangle the effects of genotype (G), environment (E, represented different evaluation years), and their interaction (G × E).

Full flowering dates (*FFD*) were fitted into a mixed linear model according to the equation:

$FFDijk = \mu + G_i + Y_j + (GY)_{ij} + \varepsilon_{ijk}$; where FFD_{ijk} is the recorded date for tree $_k$ of genotype $_i$ in year $_j$; μ represents the overall mean; G_i is the random effect of genotype $_i$; Y_j is the fixed effect of year $_j$; $(GY)_{ij}$ is the random interaction effect between genotype $_i$ and year $_j$; and ε_{ijk} represents the random residual error. Best Linear Unbiased Predictions (BLUPs) were extracted from this model to establish a standardized, temperature-adjusted ranking of cultivar flowering onset.

To quantify variance components and assess G × E interactions specifically for flowering duration (ΔT), a Two-Way Analysis of Variance (ANOVA) and Restricted Maximum Likelihood (REML) linear mixed effects model were conducted using statsmodels v0.14 in Python 3.12. Variance components were partitioned into proportion of total sum of squares (η^2). Broad-sense heritability (H^2) was estimated at both the individual observation level [$H^2_{ind} = \sigma^2_G / (\sigma^2_G + \sigma^2_{GE} + \sigma^2_e)$] and on a multi-year cultivar-mean basis [$H^2_{mean} = \sigma^2_G / (\sigma^2_G + \sigma^2_{GE}/e + \sigma^2_e/re)$]. where σ^2_G is the genotypic variance, σ^2_{GE} is the genotype-by-environment interaction variance, σ^2_e is the residual error variance, $_e$ is the number of evaluation years ($_e = 7$), and $_r$ is the average number of tree replicates per cultivar. Mean comparisons were performed using One-Way ANOVA followed by Tukey's honest significant difference (HSD) test ($p < 0.05$).

Data were also analysed using One-way ANOVA followed by Tukey's multiple comparisons test of GraphPad Prism version 9.0.0 (GraphPad Software, San Diego, CA, USA, www.graphpad.com) at a significance threshold of $p < 0.05$, Tukey test was used to compare mean values. Box and whisker and Violin plots were made by the same software.

Segregation ratios of incompatibility groups (G1 and G2) across germplasm collections, wild accessions, and controlled progenies (1:1 testcrosses and 3:1 selfed populations) were evaluated against expected Mendelian segregation using Chi-square (χ^2) goodness-of-fit tests ($df = 1$). To evaluate statistical power in wild subspecies with modest sample sizes (subsp. *cuspidata* and subsp. *laperrinei*), exact 95% binomial confidence intervals (and upper bounds for unobserved alleles) were calculated using the Clopper–Pearson method. Pairwise differences in allele frequency distributions across subspecies and maternal chloroplast lineages were evaluated using Fisher's exact test ($p < 0.05$).

Results

Amplification of the 905 genotypes with the **DomRooM** marker (Table S1; Table S7), within the *GA2ox gene* provided a clear separation of the two reproductive groups. The validity of the obtained results was confirmed by observations from stigma test phenotypic analyses conducted in previous published studies [6, 9]. Specifically, in addition to dozens of phenotyped cultivars, 112 progenies from the cross between Leccino and Dolce Agogia ('Le x DA'), previously screened by stigma test, were included (Table S7). To map the distribution of reproductive groups across *Olea europaea*, we first screened a global panel of established cultivars and wild germplasm. A total of 450 genotypes

were assigned to Group 1 (G1; *Ss* or *SS*) and 455 to Group 2 (G2; *ss*) (Chi-square goodness-of-fit test against 1:1 ratio: $\chi^2 = 0.028, df = 1, p = 0.867$). Among the 491 screened cultivars (cultivated varieties), 221 were identified as Group 1 (G1; *Ss* or *SS*) and 270 as Group 2 (G2; *ss*). A comparable near 1:1 distribution of reproductive groups was also detected in *Olea europaea* subsp. *europaea* var. *sylvestris* (62 G1 *vs*. 44 G2; $\chi^2 = 3.06, df = 1, p = 0.080$) and across all the other analysed subspecies the reproductive group assignation was equally distributed for both groups although sample sizes were smaller compared to others. To validate the Mendelian inheritance of the DomRooM marker and DSI locus, we analysed progeny derived from controlled crosses and selfing events. Balanced 1:1 segregation between the two reproductive groups was observed in the F_1 progeny from the controlled 'Leccino' × 'Dolce Agogia' testcross (61 G1 *vs*. 51 G2; $\chi^2 = 0.89, df = 1, p = 0.345$), confirming single locus diallelic inheritance. Furthermore, in the two self-pollinated populations ('Frantoio' × 'Frantoio' and 'Koroneiki' × 'Koroneiki'), the G1/G2 segregation followed the expected 3:1 Mendelian ratio (¾ G1 : ¼ G2) for the self-fertilization of a heterozygous locus with a dominant allele $\chi^2 = 0.35, df = 1, p = 0.554$ for 'Frantoio'; $\chi^2 = 0.61, df = 1, p = 0.435$ for 'Koroneiki'), results that strongly support the hypothesis of the existence of homozygous dominant (*SS*) genotypes, which at the moment are indistinguishable.

Reproductive group distribution within the cultivated genotypes

The presence of both reproductive groups is essential for successful fertilization and subsequent crop production. By examining the most widespread cultivars across the five primary olive-growing countries, we observed the following distribution of these reproductive groups (Fig. 1).

In Spain, the varietal composition of olive orchards is traditionally dominated by a few cultivars with high local adaptation. 87 olive cultivars were analysed in the present study, 34 of them are belonging to G1 while all the rest to the G2. Olive groves are never planted with numerous cultivars together but are often made up of one, two or a maximum of three main cultivars. The cultivar Picual (G2) remains the most widespread cultivar, covering over 50% of the national surface, particularly in Andalusia. Other important cultivars include cvs Arbequina (G1), Hojiblanca (G1), Cornicabra (G1), and Manzanilla de Sevilla (G2). New commercial orchards, especially those adapted to high-density or super-intensive systems, are constituted by a single or two cultivars, Arbequina (G1) and Arbosana (G2).

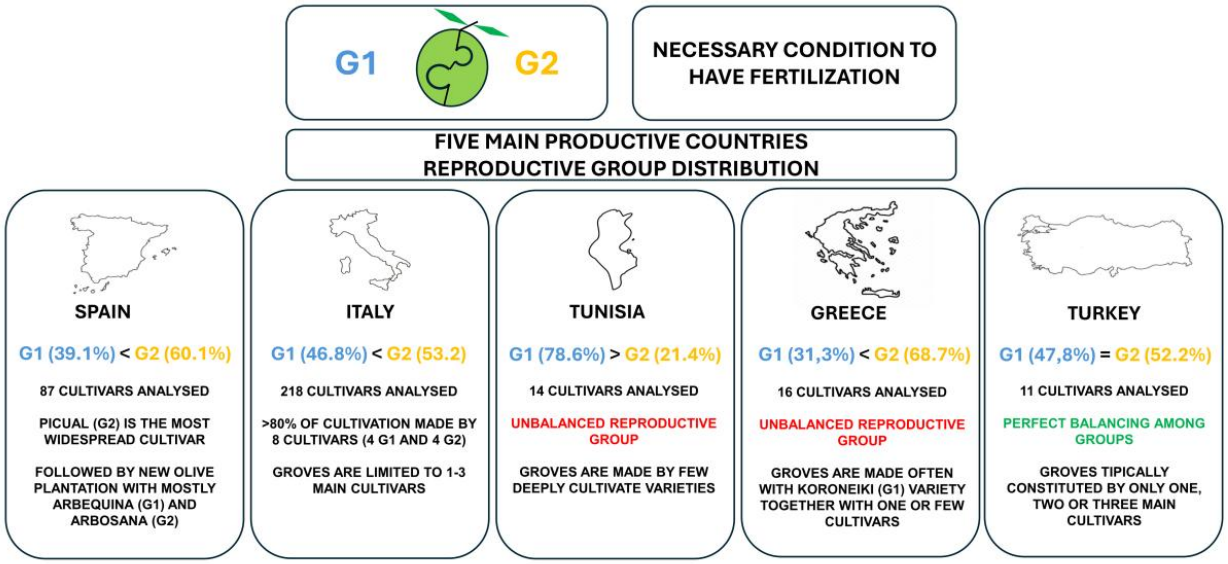


Fig. 1 The composition of reproductive groups within the five main olive producing countries and their potential impact on cross-compatibility in orchards.

In Italy, 102 were belonging to G1 and 116 to G2. Nevertheless, even if these numbers seem to be well balanced in this country, the olive groves are mainly made up of a few cultivars among those reported in this work as the G1 cultivars Leccino, Frantoio, Moraiolo, Cerasuola, and from G2 group Itrana, Coratina, Bosana and Carolea, together with their synonymy cases specially for cvs Frantoio and Bosana (Table S7).

In other countries, important for oil production but less represented in the present study as number of cultivars respect Italy and Spain, such as Tunisia, we have analysed 14 cultivars and among them eleven belonging to G1 and only three to G2, Chemchali, Ouslati and Sayali. In Greece, we considered, for the reproductive group detection, sixteen of the most cultivated cultivars. Conversely of what observed in Tunisia, here only five cultivars are G1, Karydolia, Koroneiki, Kothreiki, Megaritiki and Pikrolia, while 11 are placed in the G2. In Türkiye, a perfect balancing among the two reproductive groups was found with 11 cultivars G1 and 12 G2.

Reproductive groups under natural conditions

The monumental olive trees, mostly collected from Italy, which are randomly distributed in the territory, often traditionally regarded by growers as candidate pollen donors (putative pollenizers) or as remaining from ancient cultivation, demonstrated 50% of both reproductive groups (36 G1 *vs*. 37 G2). From what concern the 106 wild types, verified by chloroplast markers (Table S8) and collected in three different countries including Morocco, Spain and Italy, 62 are belonging to G1 and the rest to G2.

Out of the subspecies *europaea*, the subsp. *guanchica* samples were for 41.2% G1 and 58.8% G2, subspecies *laperrinei* 19 G1 and 16 G2, in *cuspidata* subspecies samples originated from four different countries 44.4% G1 and 55.6% of G2.

In general, it was observed that, growing strictly under natural conditions, the reproductive groups are both represented and often well balanced between G1 and G2 genotypes here analysed (Table S7), even if the number of some subspecies sample is limited.

Application of new developed codominant marker to individuate reproductive group in olive

During our experiments, we noted the absolute need to avoid minimal contamination by extraneous olive DNA, since the amplifiability of the dominant fragment is so high that a small percentage of contamination could produce false positives. To ensure the simultaneous presence of both alleles of the locus relating to the olive reproductive group, we developed a new codominant marker (Fig. 2).

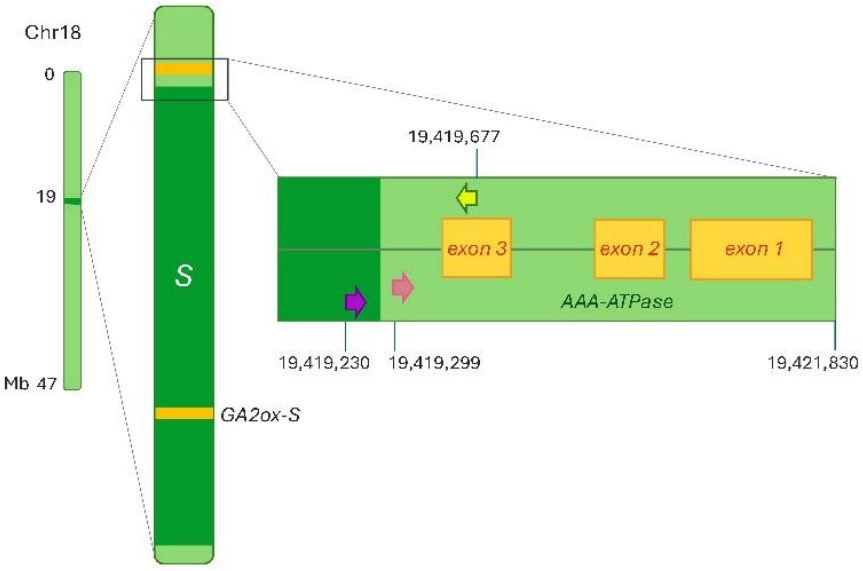


Fig. 2 Schematic, not-to-scale representation of the genomic insertion identified in the dominant S1 allele, which defines the G1 reproductive group in olive. The *Ga2ox* and *AAA-ATPase* genes are shown in yellow; the exons of the latter are shown in the magnified view. The CoRooM forward primer is indicated by the bright yellow arrow, the CoRooM G1 reverse primer by the dark magenta arrow, and the CoRooM G2 reverse primer by the pink arrow. The figure is based on the 'Arbequina' reference assembly; see also Figure S2 for the exact representation, drawn to scale.

Considering what was described in the methods section about the olive reproductive group locus, three different primers were designed, one common to both alleles and the other two, specific for each allele corresponding to incompatibility groups G1 and G2, respectively. A selection from the main panel of 422 genotypes, including several genotypes already phenotyped by stigma test and

previously identified by the dominant marker, were amplified by using the mix of these three developed primers in a single reaction discovering a high correspondence between the three datasets (phenotyped by stigma test; analysed through DomRooM; and by Codominant Reproductive olive Marker, CoRooM).

We found five different alleles length (Table S9; Fig. S3), two were related to the dominant (*S*) allele specific of G1 here after called *S1* and *S2*, while three for the (*s*) allele (denominated: s_1, s_2, s_3) present in G2 and G1 (Fig. 3).

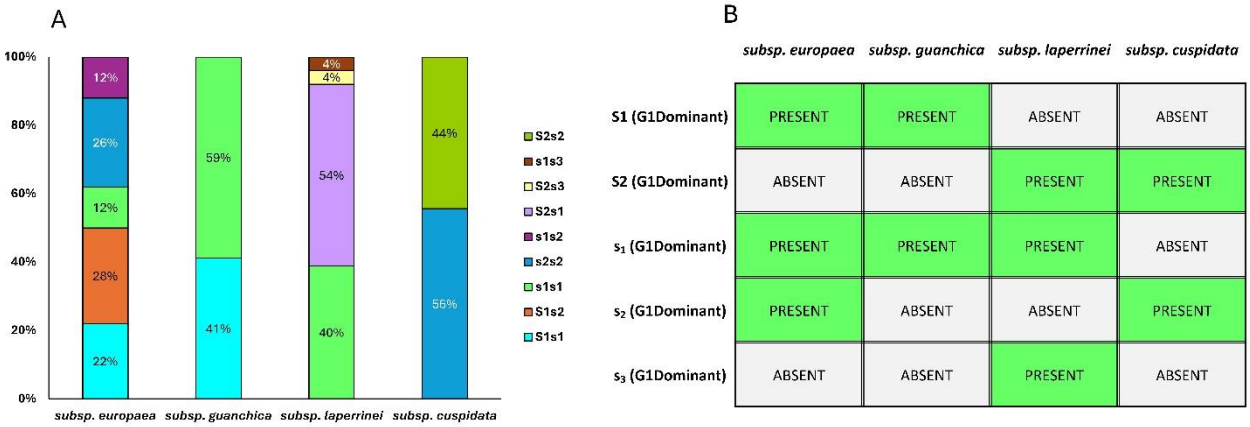


Fig. 3 Allelic composition determined using the newly developed Codominant Reproductive olive Marker (CoRooM).

Codominant allele distribution within subspecies europaea

Within *Olea europaea* subsp. *europaea*, the s_2 allele accounted for 69% of the recessive alleles (calculated from 219 genotypes, excluding the dominant S alleles), while the *S2* and s_3 alleles were completely absent. Since the frequency of the s_2 allele differed significantly between *var. europaea* and *var. sylvestris* (Fig. S4A), we analysed the distribution of alleles across maternal lineages (Fig. S4B).

We correlated s_1 and s_2 with their corresponding chloroplast lineages (E1, E2, and E3), excluding experimental progenies (Table S9). In genotypes belonging to the E1 lineage, the s_1 allele was nearly four times less frequent than s_2 (22% and 78%, respectively). Conversely, in the other chloroplast lineages (E2 and E3), the frequency of the s_1 allele increased to 51%.

In the cultivars analysed using this newly developed Codominant Reproductive olive Marker (CoRooM), three alleles were detected: one strictly related to all G1 cultivars (*S1*) and two *s* (s_1 and s_2) related to the reproductive group G2 (Fig. S3; Fig. S4), and all the five possible allelic compositions, $S1s_1$ and $S1s_2$ for G1 cultivars, while s_1s_1, s_2s_2 and s_1s_2 in G2 cultivars were detected (Table S9).

About the self-crossing 'Frantoio' and 'Koroneiki' progenies it was observed for the first time the G1 phenotype in homozygosity (two copies of the *S* allele) this allele composition was exclusively found in these self-fertilized populations and never observed in all the other genotypes here analysed. Visual inspection of these adult SS individuals confirmed their basic vegetative and reproductive viability, demonstrating that they are able to growth, flowering, and fruit-bearing (Fig. S5) and ruling out embryonic or early developmental lethality of the S-locus supergene. However, whether subtle inbreeding depression affects fine-scale physiological traits, such as pollen viability, fruit set percentage, or seed germination, relative to their Ss siblings remains to be quantified, and these progenies are currently under multi-trait evaluation.

Codominant allele distribution in subspecies guanchica, laperrinei and cuspidata

The samples of subspecies *guanchica* from Canary Islands showed only the presence of *S1* and s_1 alleles, both identical in length to what observed in the subspecies *europaea* groups (Fig. S3). This subspecies was a genetic outlier since is overwhelmingly dominated by the s_1 allele (77%). It shares the *S1* allele with *europaea*, suggesting a common ancestry or ancient link as hypothesized above. It completely lacks s_2, the most common allele in cultivated olives. This, from one side could be related to its isolation on the islands but also confirm the possible relationship among the *guanchica* subspecies and the E2 and E3 lineages, which dominated the var. *sylvestris* from central and western Mediterranean Basin.

After the analysis of subspecies *laperrinei,* three alleles were detected: the *S2* related to all G1 samples; the s_1 highly represented as already observed in subsp. *guanchica*; the s_3 detected only in 6% of G2 genotypes (Fig. 3; Fig. S3). It was one of the most distinct group, since is the only subspecies carrying the s_3 allele and showing the newly discovered allele *S2* dominant allele.

The subspecies *cuspidata* genotypes analysed here (N = 18 individuals; 36 alleles) exhibited an exclusive presence of *S2* and s_2 alleles (44.4% $S2s_2$ and 55.6% s_2s_2), completely lacking *S1*, s_1, and s_3 alleles (95% Clopper–Pearson exact confidence upper bound for unobserved alleles: $\leq$ 9.7% at N = 36 alleles).

Notably, the analysis reveals a clear geographic distribution of incompatibility alleles across Olea subspecies, reinforcing their genetic isolation while strongly confirming that the DSI system is preserved at the subspecies level. Specifically: subspecies *cuspidata* (including samples from China, Ethiopia, Nepal, and Saudi Arabia) exclusively carries the *S2* and s_2 alleles. Subspecies *guanchica* (including samples from Tenerife and Gran Canaria) possesses the *S1* and s_1 alleles. Subspecies *laperrinei* (from Tamanrasset, Algeria) shares the *S2* allele with subsp. *cuspidata* and the s_1 allele with subsp. *guanchica*. This distribution underscores the distinct genetic identities of these groups while highlighting the conservation of the underlying incompatibility mechanism. The observed

lengths were confirmed by direct Sanger sequences using the same set of newly developed markers without fluorescence (Table S5), moreover all primers and their corresponding amplicons showed significant matches at the predicted positions, and the amplicon lengths were as expected when blast against the published olive genomes (Table S6).

The choice of the pollen donor to ensure effective pollination of olive trees

To evaluate temporal requirements for effective cross-pollination, Best Linear Unbiased Predictions of full flowering date (BLUP-FFD) were cross-analysed with *S*-allele reproductive group assignments (G1 *vs*. G2) across 228 cultivars originating from 14 countries (Table S10). Geographic origin significantly influenced flowering onset: cultivars from France and Spain displayed the earliest bloom initiation (Mean BLUP-FFD ~114–115 Julian days), whereas central and eastern Mediterranean germplasm pools, particularly from Italy and Syria, exhibited significantly later bloom onset (~117–118 Julian days, $p < 0.0001$; Fig. 4).

Crucially, correlating flowering timing with reproductive groups revealed a severe structural skew across the bloom season. While early blooming cultivars are evenly distributed between G1 (52%) and G2 (48%), mid-to-late blooming varieties are overwhelmingly dominated by G2 genotypes. This creates a critical deficit of late-blooming G1 pollen donors in regions such as Italy and Syria. Consequently, growers selecting mid-to-late G2 cultivars face an elevated risk of reproductive failure due to a scarcity of synchronized G1 donors.

These findings demonstrate that successful orchard cross-pollination requires navigating a two-step reproductive barrier: a phenological barrier (temporal flowering synchrony) that physically enables pollen transfer, and a genetic barrier (diallelic self-incompatibility, G1 × G2) that molecularly permits fertilization. For instance, in Greece, the earliest-blooming cultivar ('Karydolia', G1) is temporally isolated from the subsequent eight G2 cultivars, while all remaining Greek G1 cultivars bloom significantly later (Table S10). Integrating *S*-allele profiling with multi-year phenological timing is therefore essential to prevent reproductive isolation in commercial orchard design.

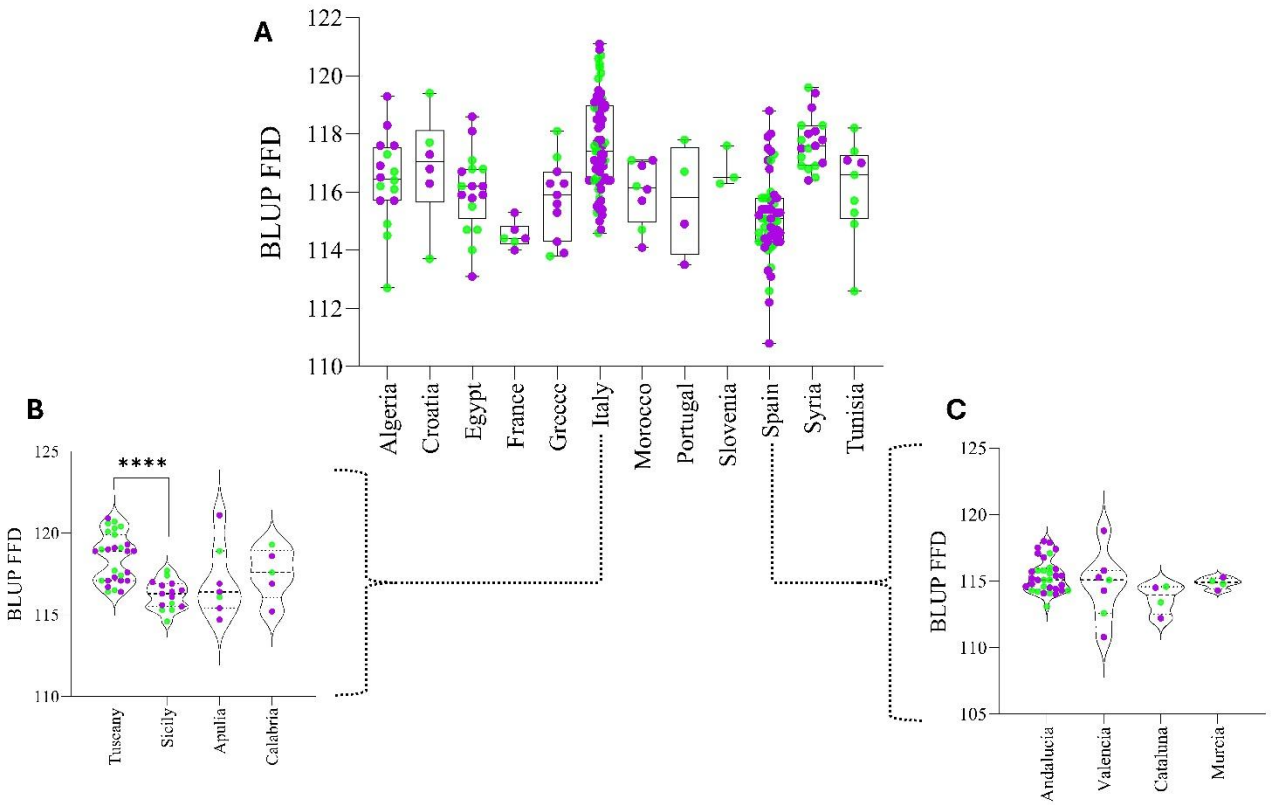


Fig. 5 Boxplot analysis of flowering time (BLUP FFD) and reproductive groups. (A) Cultivars categorized by country of origin. Tukey's multiple comparisons test revealed significant differences: Italy vs. Spain and Spain vs. Syria (**** $p < 0.0001$); France vs. Italy and France vs. Syria (*** $p < 0.001$); Egypt vs. Italy and Greece vs. Italy (** $p < 0.01$); and Egypt vs. Syria and Greece vs. Syria (* $p < 0.05$). (B) Analysis of Italian cultivars across four major olive oil-producing regions. (C) Association between BLUP FFD and reproductive group for Spanish cultivars within four different provinces, without significant differences among them. G1 is shown by green and G2 by violet color.

Genotype-by-Environment (G × E) Interaction and Plasticity of Flowering Duration

Multi-year variance partitioning across 679 cultivar–year observations representing 154 selected cultivars revealed a fundamental physiological divergence between the timing of flowering onset (Stage 60) and total flowering duration ($\Delta T = T_{Stage\ 69} - T_{Stage\ 60}$). While flowering onset (Stage 60) is strongly anchored to cultivar genetics ($H^2 = 0.84$; Environment/year = 80.17%, Genotype = 8.24%, $p = 2.93 \times 10^{-13}$), flowering duration (ΔT) is predominantly driven by environmental variation (43.67% of total variance, $p = 2.21 \times 10^{-75}$) and G × E interaction plasticity (43.18% of total variance) (Table S10). Main genetic effects accounted for 13.15% of duration variance ($p = 0.417$), yielding a low single-year heritability ($H^2_{ind} = 0.13$) that rises to moderate levels when averaged across multiple evaluation years ($H^2_{mean} = 0.65$).

Inter-annual environmental variation caused macro-shifts in population mean duration across the germplasm collection, ranging from brief population-wide bloom windows of 11.26 ± 3.49 days in 2021 and 13.07 ± 1.22 days in 2015 to extended periods of 19.45 ± 2.99 days in 2018 and 17.00 ± 3.37 days in 2016. Furthermore, cultivars exhibited contrasting G × E duration stability across evaluation seasons. Targeted analysis of widely diffused Mediterranean cultivars revealed marked differences in flowering duration stability (Table 1). Cultivars such as 'Coratina' (G2, Italy; std ±2.78 days), 'Leccino' (G1, Italy; ±3.05 days), and 'Arbequina' (G1, Spain; ±3.55 days) maintained highly predictable bloom windows across evaluation seasons, whereas 'Moraiolo' (G1) and 'Bosana' (G2) exhibited high environmental plasticity (std ±5.97 and ±8.14 days, respectively). Evaluating inter-compatible bloom window overlaps (G1 × G2) demonstrated strong synchrony in premier commercial pairings such as 'Arbequina' (G1) × 'Coratina' (G2) (12.5 days overlap, 73.2% synchrony) and 'Cerasuola' (G1) × 'Bosana' (G2) (13.6 days, 98.9%) (Table S11A). Conversely, severe temporal mismatches occurred in 'Arbequina' (G1) × 'Picual' (G2) (5.0 days, 34.5%) and 'Pikrolia' (G1) × 'Manzanilla de Sevilla' (G2) (4.7 days, 32.7%) (Table S11B). These findings confirm that genetic compatibility (G1 × G2) must be combined with phenological duration stability to ensure reliable cross-pollination in commercial orchards.

Table 1. Phenological metrics and duration stability of widely diffused cultivars.

Cultivar	Country	Group (DomRooM)	Start Day (Stage 60)	End Day (Stage 69)	Mean Duration (Days)	Std Dev (Days)
Coratina	Italy	**G2**	114.5 (Apr 24)	128.9 (May 08)	**17.1**	**±2.78**
Leccino	Italy	**G1**	120.8 (Apr 30)	128.0 (May 08)	**13.4**	**±3.05**
Arbequina	Spain	**G1**	112.1 (Apr 22)	127.0 (May 07)	**18.3**	**±3.55**
Pikrolia	Greece	**G1**	122.9 (May 02)	134.4 (May 14)	**17.4**	**±3.85**
Manzanilla de Sevilla	Spain	**G2**	115.0 (Apr 25)	127.6 (May 07)	**14.4**	**±4.01**
Hojiblanca	Spain	**G1**	119.2 (Apr 29)	132.0 (May 12)	**16.0**	**±4.12**
Frantoio	Italy	**G1**	116.7 (Apr 27)	127.9 (May 07)	**14.8**	**±4.51**
Chemchali	Tunisia	**G2**	119.0 (Apr 29)	129.8 (May 09)	**14.7**	**±4.72**
Cerasuola	Italy	**G1**	120.1 (Apr 30)	134.0 (May 14)	**17.8**	**±4.87**
Itrana	Italy	**G2**	118.2 (Apr 28)	131.4 (May 11)	**15.0**	**±4.95**
Picual	Spain	**G2**	122.0 (May 02)	131.0 (May 11)	**14.5**	**±4.97**
Cornicabra	Spain	**G1**	117.9 (Apr 28)	132.0 (May 12)	**15.9**	**±5.14**
Carolea	Italy	**G2**	114.9 (Apr 25)	130.9 (May 11)	**17.5**	**±5.20**
Moraiolo	Italy	**G1**	118.8 (Apr 29)	130.2 (May 10)	**14.7**	**±5.97**
Bosana	Italy	**G2**	120.4 (Apr 30)	135.2 (May 15)	**13.8**	**±8.14**

Discussion

Molecular Dissection and the Precision of the DSI System

This study successfully developed and validated a robust molecular toolkit for the rapid identification of the two diallelic self-incompatibility (DSI) groups in *Olea europaea*. Initial screening using the dominant marker (DomRooM), targeting the *GA2ox* locus, provided a clear separation of the reproductive groups. This binary segregation (G1 vs. G2) aligned perfectly with expectations for a single-locus system, a finding further supported by controlled cross-segregations. However, the development of the codominant CoRooM marker represents a critical methodological advancement by moving beyond simple presence/absence assays to precise allelic composition analysis.

The analytical power and biological relevance of this codominant marker are most clearly demonstrated through our experimental self-pollinated populations ('Frantoio' and 'Koroneiki'). The classic Mendelian segregation observed in these progenies is a pivotal finding: it definitively validates the single-locus, dominant inheritance model of the olive DSI system and proves our marker's ability to accurately identify heterozygous (*Ss*) individuals.

More importantly, these controlled self-pollinations provide the vital proof of concept that homozygous dominant (*SS*) genotypes are vegetatively and reproductively viable and can be generated, at least experimentally, confirming that the S-locus insertion does not harbor a complete embryonic lethal mutation. The complete absence of *SS* individuals across all 869 accessions analyzed that include ancient trees, cultivated varieties, wild and subspecies is a direct structural consequence of the DSI mechanism rather than a geographic sampling artifact or rare mutation threshold. Because DSI strictly prevents G1 (*Ss*) × G1 (*Ss*) crosses and natural self-fertilization, every open-pollinated fertilization onto a G1 mother is exclusively sired by a G2 (*ss*) pollen donor. Because a G2 donor can only supply an s allele, open-pollinated progenies are mathematically restricted to *Ss* and *ss* genotypes in a 1:1 ratio. Consequently, *SS* genotypes cannot form in open-pollinated natural populations regardless of sample size or geographic origin and can only be recovered through forced experimental self-pollination.

Balancing Selection and the Evolutionary History of the Olea Genus

Our investigation into wild olives (var. *sylvestris*), ancient trees, and related subspecies provides strong evidence for the long-term stability of the DSI system. Across these natural populations, the G1 and G2 groups were found in a near-perfect 1:1 ratio. This equilibrium is a classic signature of balancing selection driven by strong negative frequency-dependent selection, a mechanism that maximizes mating opportunities and has maintained these two groups over tens of millions of years within the Oleaceae family [4-5].

The distribution of these incompatibility alleles offers novel insights into the evolutionary history and geographic spread of the species. The genetic singularity of the subspecies reflects their geographic isolation among each other: *subsp. guanchica* (Canary Islands) and *subsp. cuspidata* (Asia/Africa) display non-overlapping dominant alleles (*S1* versus *S2*, respectively), indicating divergent evolutionary paths. Meanwhile, *subsp. laperrinei* acts as a genetic bridge across the Sahara/Asia axis, lacking *S1* but sharing the *S2* allele with *cuspidata*, while also harboring a novel s_3 allele. While sample sizes for wild relict taxa such as subsp. *cuspidata* (N = 18) and subsp. *laperrinei* (N = 35) are inherently modest due to the extreme geographic isolation and endangered status of these trees, exact binomial probability calculations (Clopper–Pearson 95% CI) establish that undetected alternative allele in *cuspidata* would exist at a population frequency below 9.7% ($p < 0.05$). Although expanded future sampling across remote Asian and East African populations might uncover rare low-frequency alleles, the overwhelming statistical contrast between *cuspidata* (dominated by $S2/s_2$) and Mediterranean *europaea* (dominated by $S1/s_1/s_2$; Fisher's exact test $p < 0.0001$) reflects deep lineage sorting and prolonged evolutionary isolation rather than a sampling artifact. The separation of *guanchica* and *laperrinei* significantly reframes the evolutionary picture, demonstrating that the SI locus has continued to diversify following geographic separation.

Introgression and Allelic Shifts in Cultivated Olives

In cultivated olives (subsp. *europaea*), the allelic distribution challenges the assumption of simple, linear domestication. We observed a sharp contrast in the prevalence of the recessive s_1 allele between wild lineages and modern cultivars. When measuring the ratio among the three detected alleles, s_1 was poorly represented in the E1 maternal lineage but increased considerably in the E2 and E3 lineages. This suggests that the s_1 allele likely originated from a wild progenitor in the central/western Mediterranean Basin and was introgressed into cultivated genotypes by crossing eastern and central-western olive germplasm. Conversely, the s_2 allele, likely originating from eastern Mediterranean germplasm, appears to have been swept into the western gene pool during a relatively recent domestication event, where few cultivars have been originated more than 60% of the main cultivated varieties [29]. This complex history of introgression explains the reduced allelic diversity observed in modern orchards compared to their wild progenitors, while highlighting the newly identified alleles in *laperrinei* and *cuspidata* as an untapped reservoir for future breeding.

Two-Step Reproductive Barrier in Olive: Phenological Synchrony and Molecular Compatibility

An essential biological outcome of this broad screening is the conceptualization of olive reproduction as a two-step reproductive barrier: the phenological barrier (temporal flowering synchrony) that physically enables pollen transfer; and a genetic barrier (diallelic self-incompatibility, G1 × G2) that

585 molecularly permits fertilization. Our analysis of flowering times (BLUP-FFD) and flowering
586 duration (ΔT) demonstrates that genetic compatibility is functionally void without phenological
587 overlap.
588 Our multi-year variance component analysis reveals that the phenological barrier is governed by two
589 distinct biological dimensions: flowering onset rank (Stage 60) and flowering duration elasticity (ΔT).
590 While flowering onset is strongly anchored to cultivar genetics, flowering duration is overwhelmingly
591 driven by environmental variation and $G \times E$ interaction plasticity. This high duration elasticity
592 explains why $\mathrm{G1} \times \mathrm{G2}$ pollen donor synchrony can collapse unpredictably across seasons: during
593 compressed bloom years, the flowering window contracts to under 10 days, severely narrowing the
594 temporal opportunity for cross-pollination. Crucially, the global germplasm exhibits a severe
595 structural skew that exacerbates this phenological vulnerability. While early blooming Spanish and
596 French cultivars display a balanced $\mathrm{G1/G2}$ ratio (52% G1 *vs*. 48% G2), mid-to-late blooming
597 populations, critical for frost avoidance in Italy and Syria, are overwhelmingly dominated by G2
598 genotypes. Introducing an early blooming G1 pollen donor into a late blooming G2 orchard creates
599 severe temporal mismatches. For example, cross-analyzing major commercial pairings revealed that
600 while 'Arbequina' (G1) and 'Coratina' (G2) achieve an optimal 12.5 days of overlap (73.2%
601 synchrony), pairing 'Arbequina' (G1) with 'Picual' (G2) leaves a narrow 5.0-day overlap window
602 (34.5% synchrony) because 'Arbequina' finishes blooming shortly after 'Picual' opens.
603 Similarly, the Tunisian germplasm exhibits a dangerous skew, with seven out of nine investigated
604 cultivars belonging to G1, alongside early flowering traits that isolate them from potential G2
605 pollinators. As climate change disrupts chilling accumulation and drives earlier, more erratic
606 flowering phenology [1-2, 21], the risk of asynchrony will escalate. Furthermore, cultivars differ
607 significantly in $G \times E$ duration stability across seasons: highly stable cultivars, such as 'Coratina' (std
608 $\pm$2.78 days), 'Leccino' ($\pm$3.05days), and 'Arbequina' ($\pm$3.55 days), maintain predictable bloom
609 windows across variable evaluation years, whereas highly plastic cultivars, such as 'Moraiolo'
610 ($\pm$5.97 days) and 'Bosana' ($\pm$8.14 days), undergo extreme duration expansion or contraction.
611 Importantly, while elevated temperatures during bloom can induce Pseudo-Self-Compatibility (PSC)
612 or physiological "leakiness" in certain cultivars (e.g., 'Frantoio' and 'Koroneiki') by modulating
613 gibberellin pathway dynamics (*GA2ox*) [11, 14], PSC remains highly erratic, thermally stressful, and
614 frequently accompanied by high rates of embryo abortion and pseudo-drupe formation [17].
615 Consequently, growers should not rely on temperature-induced SI breakdown as an agronomic
616 substitute for cross-pollination. Instead, under future climate warming scenarios, growers are advised
617 to adopt multi-pollenizer orchard layouts, combining molecularly verified G1 × G2 inter-
618 compatibility with multiple phenologically stable pollenizers, using PSC merely as a secondary

physiological safety buffer rather than a primary design parameter. Therefore, securing the resilience of this crop requires a paradigm shift: integrating molecular DSI screening to ensure genetic inter-compatibility, while strictly selecting pollenizers with both overlapping mean BLUP-FFD ranks and high flowering duration stability across variable environments [18].

Conclusion

This study provides a molecular resolution to the diallelic self-incompatibility (DSI) system in *Olea europaea*, offering a validated codominant toolkit that fundamentally advances our understanding of olive reproductive biology. By mapping the genetic architecture of this, evolutionarily conserved barrier across a global germplasm, we demonstrate that the massive loss of reproductive potential often observed in this species is a direct consequence of strict genetic constraints conflicting with phenological asynchrony.

The widespread allelic skew identified in this study, particularly the severe global deficit of late-blooming G1 genotypes, exposes a significant structural vulnerability. As climate change continues to disrupt chilling accumulation and exacerbate erratic flowering, this two-step reproductive barrier faces increasing strain: shifting bloom dates and environmentally driven duration contractions narrow the temporal window between compatible G1 and G2 cultivars, while regional allelic skews leave late-blooming orchards with few compatible partners.

Remarkably, these fundamental biological insights translate directly into high-impact agronomic applications. The codominant marker system developed here provides the industry with a rapid, reliable tool to overcome widespread varietal imbalances. By adopting this screening in nurseries and breeding programs, we can enable Marker-Assisted Selection (MAS) and the precise, scientifically proven assignment of compatible pollenizers. Moving forward, integrating this accurate genotyping with robust phenological modeling and $G \times E$ duration stability metrics will allow growers to transition orchard design into a precise science. By utilizing these molecular tools to harness the genetic plasticity of untapped populations, specifically targeting the development of late-blooming G1 cultivars with high phenological stability, we can correct global germplasm skews and secure the long-term yield and climate resilience of this globally vital crop.

Acknowledgments

This research is the result of activities carried out in the following project OMIBREED [CUP-B93C22001440005, 2022]. The authors acknowledge both, Luciana Baldoni, CNR IBBR Italy, who made available the progeny ‘Le x DA’ and Saverio Pandolfi who proposed the first hypotheses of inter-compatibility within olive during his collaboration with the CNR IBBR research group. We are very grateful to Pierre Saumitou-Laprade and Philippe Vernet for reviewing this work suggesting

highlighting some of the main findings, demonstrating once again the close collaboration between the research groups.

Author Contributions

R.M. and S.M. conceived and designed the project. R.M., S.M., L.C., C.C., J.K., P.V., C.P., C.T., J.P., N.R., K.A., B.A., N.U., M.O., H.A., R.N., R.K., M.A.B. and M.S., performed the prospective survey *in situ* and *ex situ* (National and International collections). R.M., S.M., S.P., L.C., O.A., S.C. and M.S. performed the experiments. R.M., S.M., P.V., C.P., O.A., and C.T. contributed to validation of experiments. R.M., S.M., S.P., L.C., C.C., J.K., P.V., C.P., O.A., S.C. and C.T. discussed and analysed the data. R.M., S.M., S.P., L.C., C.C., J.K., P.V., S.O., C.P., O.A., C.T., J.P., N.R., K.A., B.A., N.U., M.O., H.A., R.K., S.C., M.A.B. and M.S., drafted the manuscript, and participated in its revision. All authors read and approved of the final manuscript.

Availability of data and materials

All data generated or analysed during this study are included in this article and its supplementary information files.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Cuevas J. Why olive produces many more flowers than fruit—A critical analysis. Horticulturae. 2025;11(1):26.
2. Bashir MA, Alagna F, Celletti S, Cinosi N, Famiani F, Silvestri C, Cristofori V. Olive reproductive biology: implications for yield, compatibility conundrum, and environmental constraints. J Exp Bot. 2024;75(14):4300–13.
3. Barrett SC, Fairnie AL. The neglected floral polymorphism: mirror-image flowers emerge from the shadow of heterostyly. Evol J Linn Soc. 2024;3(1):kzae004.
4. Goldberg EE, Kohn JR, Lande R, Robertson KA, Smith SA, Igić B. Species selection maintains self-incompatibility. Science. 2010;330(6003):493–5.
5. Besnard G, Dupuy J, Lhuillier E, Baali-Cherif D. Paternity tests support a diallelic self-incompatibility system in a wild olive (*Olea europaea* subsp. *laperrinei*, Oleaceae). Ecol Evol. 2020;10(4):1876–88. [suspicious link removed]
6. Saumitou-Laprade P, Vernet P, Vekemans X, Billiard S, Gallina S, Essalouh L, Khadari B, Guitton B, Delph LF. Elucidation of the genetic architecture of self-incompatibility in olive: Evolutionary consequences and perspectives for orchard management. Evol Appl. 2017;10(9):867–80.

7. Vernet P, Le Péchon T, Carré P, Billiard S, Lepers-Andrzejewski S, Duputié A, Saumitou-Laprade P. Evidence for the long-term maintenance of a rare self-incompatibility system in Oleaceae. New Phytol. 2016;210(4):1408–17.
8. Saumitou-Laprade P, Vassiliadis C, Baudry E, Arzembouck J, Vernet P. A self-incompatibility system explains high male frequencies in an androdioecious plant. Science. 2010;327(5973):1648–50.
9. Mariotti R, Pandolfi S, De Cauwer I, Saumitou-Laprade P, Vernet P, Rossi M, Fabiani A, Baldoni L. Diallelic self-incompatibility is the main determinant of fertilization patterns in olive orchards. Evol Appl. 2021;14(4):983–95.
10. Mariotti R, Pandolfi S, Baldoni L, Saumitou-Laprade P, Vernet P, Rossi M, Alagna F. Genetic mapping of the incompatibility locus in olive and development of a linked sequence-tagged site marker. Front Plant Sci. 2020;10:1760.
11. Castric V, Raimondeau P, Vernet P, Poux C, Vaudey V, Dufay M, Saumitou-Laprade P. The homomorphic self-incompatibility system in Oleaceae is controlled by a hemizygous genomic region expressing a gibberellin pathway gene. Curr Biol. 2024;34(9):1967–76.
12. Raimondeau P, Poux C, Carré P, Gallina S, Vaudey V, Vernet P, Saumitou-Laprade P, Castric V. A hemizygous supergene controls homomorphic and heteromorphic self-incompatibility systems in Oleaceae. Curr Biol. 2024;34(9):1977–86.
13. Moreno-Sanz P, Lourenço S, Arnan X, Alagna F, Mariotti R, Baldoni L, Bertoni M. Genetic resources of *Olea europaea* L. in the Garda Trentino olive groves revealed by ancient trees genotyping and parentage analysis of drupe embryos. Genes. 2020;11(10):1171.
14. Alagna F, Ciacciulli A, Cirilli M, Zelasco S, Muleo R. The paradox of self-fertile varieties in the context of self-incompatible genotypes in olive. Front Plant Sci. 2019;10:725.
15. Marchese A, Marra FP, Caruso T, Mesejo C, Martínez-Fuentes A, Agustí M. 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. 2016;10(1):82–93.
16. Maldera F, Vivaldi GA, Camposeo S. Flowering phenograms and genetic sterilities of ten olive cultivars grown in a super-high-density orchard. Horticulturae. 2026;12(1):110.
17. Sánchez-Estrada A, Cuevas J. 'Arbequina' olive is self-incompatible. Sci Hortic. 2018;230:50–5.
18. Medina-Alonso MG, Navas-López JF, Cabello D, Del Río C, Caballero JM, Santos C, Rapoport HF, De la Rosa R, León L. Flowering phenology of olive cultivars in two climate zones with contrasting temperatures (subtropical and Mediterranean). Agriculture. 2023;13(7):1312.
19. Navas-Lopez JF, León L, Rapoport HF, Moreno-Alías I, Santos C, De la Rosa R. Genotype, environment and their interaction effects on olive tree flowering phenology and flower quality. Euphytica. 2019;215(11):184.
20. Aqbouch L, El Bakkali A, Moukhli A, Essalouh L, Udupa SM, Gaboun N, Khadari B. Characterization of an olive population relevant for genome-wide association studies on flowering time. Acta Hortic. 2023;1412:129–36.
21. Zhang S, Huang S, Wang X, Gao J. Floral mechanisms promote pollination success and reduce the incidence of self-pollination in a fly-pollinated self-incompatible orchid. Ecol Evol. 2024;14(6):e11295.
22. Lombardo L, Farinelli D, Tombesi S, Famiani F. Uncovering olive biodiversity through analysis of floral and fruiting biology and assessment of genetic diversity of 120 Italian cultivars with minor or marginal diffusion. Biology. 2019;8(3):62.
23. Selak GV, Perica S, Ban SG, Radunić M, Poljak M. Reproductive success after self-pollination and cross-pollination of olive cultivars in Croatia. HortScience. 2011;46(2):186–91.
24. Miho H, Moral J, López-González MA, Prieto-Linde ML, Díez CM. Parental effect on agronomic and olive oil traits in olive progenies from reciprocal crosses. Plants. 2024;13(17):2467.
25. Torres M, Pierantozzi P, Searles P, Rousseaux MC, Maestri D, Gheza M, Contreras C, Trentacoste E. Olive cultivation in the southern hemisphere: Flowering, water requirements and oil quality responses to new crop environments. Front Plant Sci. 2017;8:1830.

26. Belaj A, De la Rosa R, Lorite IJ, Mariotti R, Cultrera NG, Baldoni L, León L. Utility of EST-SNP markers for improving management and use of olive genetic resources: A case study at the Worldwide Olive Germplasm Bank of Córdoba. Plants. 2022;11(7):921.
27. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987;19:11–5.
28. Hmmam I, Mariotti R, Ruperti B, Cultrera N, Baldoni L, Barcaccia G. Venetian olive (*Olea europaea*) germplasm: disclosing the genetic identity of locally grown cultivars suited for typical extra virgin oil productions. Genet Resour Crop Evol. 2018;65(6):1733–50.
29. Mariotti R, Belaj A, De la Rosa R, Leòn L, Valeri MC, Mousavi S, Cultrera NG, El Bakkali A, Essalouh L, Khadari B, Pandolfi S, Baldoni L. Genealogical tracing of *Olea europaea* species and pedigree relationships of var. europaea using chloroplast and nuclear markers. BMC Plant Biol. 2023;23(1):452.
30. Hosseini-Mazinani M, Mariotti R, Torkzaban B, Sheikh-Hassani DV, Mousavi S, Mariotti M, Cultrera NG, Pandolfi S, Baldoni L. High genetic diversity detected in olives beyond the boundaries of the Mediterranean Sea. PLoS One. 2014;9(4):e93146.
31. Mariotti R, Cultrera NG, Muñoz-Díez C, Baldoni L. Identification of new polymorphic regions and differentiation of cultivated olives (*Olea europaea* L.) through plastome sequence comparison. BMC Plant Biol. 2010;10(1):211.
32. Lv J, Rao G, Zhu S, Li C, Zhang X, Zhou C, Liu D, Chen Y, Wang X. The gapless genome assembly and multi-omics analyses unveil a pivotal regulatory mechanism of oil biosynthesis in the olive tree. Hortic Res. 2024;11(8):uhae168.
33. Cruz F, Julca I, Gómez-Garrido J, Loska D, Marcet-Houben M, Cano E, Vargas P, Ribeca P, Alioto T, Cow D, Gut I, Gabaldón T, Alioto TS. Genome sequence of the olive tree, Olea europaea. GigaScience. 2016;5(1):s13742-016-0134-4.
34. Rao G, Zhang J, Liu X, Lin C, Sun P, Hu T, Wang B, Lv J, Chen Y, Wang X. De novo assembly of a new Olea europaea genome accession using Nanopore sequencing. Hortic Res. 2021;8(1):45.
35. Ramírez-Tejero JA, Serrano A, Cantos M, Liñán J, Morales-Sillero A, Luque F. Gene expression pattern in olive tree organs (*Olea europaea* L.). Genes. 2020;11(5):544.
36. Wang L, Sun P, Lv J, Zhang X, Rao G, Wang X. High-quality genome assembly of *Olea europaea* subsp. cuspidata provides insights into its resistance to fungal diseases in the summer rain belt in East Asia. Front Plant Sci. 2022;13:879822.
37. Wu T, Zhou C, Liu D, Chen Y, Wang X. The de novo genome assembly of *Olea europaea* subsp. *cuspidata*, a widely distributed olive close relative. Front Genet. 2022;13:868540.
38. Sanz-Cortés F, Martínez-Calvo J, Badenes ML, Bleiholder H, Hack H, Llácer G. Phenological growth stages of olive trees (*Olea europaea*). Ann Appl Biol. 2002;140(2):151–7.
39. Abou-Saaid O, El Bakkali A, Charafi J, Moukhli A, Khadari B, Delalande M, El Modafar C. Statistical approach to assess chill and heat requirements of olive tree based on flowering date and temperatures data: towards selection of adapted cultivars to global warming. Agronomy. 2022;12(12):2975.

Figure Legends

Figure 1. The composition of reproductive groups within the five main olive producing countries and their potential impact on cross-compatibility in orchards.

Figure 2. Schematic, not-to-scale representation of the genomic insertion identified in the dominant S1 allele, which defines the G1 reproductive group in olive. The *Ga2ox* and *AAA-ATPase* genes are shown in yellow; the exons of the latter are shown in the magnified view. The CoRooM forward

primer is indicated by the bright yellow arrow, the CoRooM G1 reverse primer by the dark magenta arrow, and the CoRooM G2 reverse primer by the pink arrow. The figure is based on the 'Arbequina' reference assembly; see also Figure S2 for the exact representation, drawn to scale.

Figure 3. Allelic composition determined using the newly developed Codominant Reproductive olive Marker (CoRooM).

Figure 4. Boxplot analysis of flowering time (BLUP FFD) and reproductive groups. (A) Cultivars categorized by country of origin. Tukey's multiple comparisons test revealed significant differences: Italy vs. Spain and Spain vs. Syria (**** $p < 0.0001$); France vs. Italy and France vs. Syria (*** $p < 0.001$); Egypt vs. Italy and Greece vs. Italy (** $p < 0.01$); and Egypt vs. Syria and Greece vs. Syria (* $p < 0.05$). (B) Analysis of Italian cultivars across four major olive oil-producing regions. (C) Association between BLUP FFD and reproductive group for Spanish cultivars within four different provinces, without significant differences among them. G1 showed by green and G2 by violet color.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.pdf
- FigureS2.jpg
- FigureS1.png
- FigureS3.jpg
- FligureS4.jpg
- FigureS5.jpg
- TableS3.docx
- TableS6.xlsx
- TableS11.xlsx
- TableS8.xlsx
- TableS5.txt
- TableS1.xlsx
- TableS4.docx
- TableS7.xlsx
- TableS2.xlsx
- TableS9.xlsx
- TableS10.xlsx