PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

CC BY

Research

Cite this article: Machado RSR *et al*. 2025 The origins and spread of the opium poppy (*Papaver somniferum* L.) revealed by genomics and seed morphometrics. *Phil. Trans. R. Soc. B* **380**: 20240198.
https://doi.org/10.1098/rstb.2024.0198

Received: 13 January 2025
Accepted: 26 March 2025

One contribution of 17 to a theme issue 'Unravelling domestication: multi-disciplinary perspectives on human and non-human relationships in the past, present and future'.

Subject Areas:
plant science, genomics, evolution

Keywords:
domestication, opiates, agrobiodiversity, plant genome, seed morphology

Author for correspondence:
Rui S. R. Machado
e-mail: rsmachado@ualg.pt

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.7742946.

THE ROYAL SOCIETY PUBLISHING

The origins and spread of the opium poppy (*Papaver somniferum* L.) revealed by genomics and seed morphometrics

Rui S. R. Machado[1], Vincent Bonhomme[4], Raül Soteras[5], Angele Jeanty[4,6], Laurent Bouby[4], Allowen Evin[4], M. Joao Fernandes Martins[1], Sandra Gonçalves[2], Ferran Antolín[5,7], Aurélie Salavert[6] and Hugo Rafael Oliveira[1,3]

[1]ICArEHB, Interdisciplinary Center for Archaeology and Evolution Human Behaviour, [2]MED – Mediterranean Institute for Agriculture, Environment and Development & CHANGE – Global Change and Sustainability Institute, Faculdade de Ciências e Tecnologia, and [3]Faculdade de Ciências Humanas e Sociais, Universidade do Algarve, Campus de Gambelas, Faro, Portugal
[4]ISEM, Université Montpellier, CNRS, IRD, EPHE, Montpellier, France
[5]Division of Natural Sciences, German Archaeological Institute (DAI), Berlin, BE, Germany
[6]UMRAASPE/BIOARCH, Muséum National d'Histoire Naturelle (MNHN), Centre National de Recherche Scientifique (CNRS), Alliance Sorbonne Université, Paris, France
[7]Environmental Sciences, Integrative Prehistory and Archaeological Science (IPNA/IPAS), University of Basel, Basel, Switzerland

RSRM, 0000-0002-4080-2353; VB, 0000-0002-2742-6349; AJ, 0000-0002-0673-0704; LB, 0000-0002-3633-9829; AE, 0000-0003-4515-1649; MJFM, 0000-0002-9118-7397; SG, 0000-0002-3038-4434; FA, 0000-0002-0533-5788; AS, 0000-0002-7854-9170; HRO, 0000-0002-5038-073X

The opium poppy (*Papaver somniferum* L.) is one of the most important plants in human history. It is the main source of opiates used as analgesic medicines or psychotropic drugs, the latter related to addiction problems, illegal trafficking and geopolitical issues. Poppyseed is also used in cooking. The prehistoric origins, domestication and cultivation spread of the opium poppy remain unresolved. Traditionally, *Papaver setigerum* has been considered the wild ancestor with early cultivation presumed to have occurred in the Western Mediterranean region, where *setigerum* is autochthonous. Other theories suggest that *somniferum* may have been introduced by Southwest Asian early farmers as a weed. To investigate these hypotheses, we analysed 190 accessions from 15 *Papaver* species using genotype-by-sequencing and geometric morphometric (GMM) techniques. Our analysis revealed that *setigerum* is the only taxa genetically close to *somniferum and can be better described as a subspecies*. The domesticated plants are, however, distinct from *setigerum*. Additionally, GMM analysis of seeds also revealed morphological differences between *setigerum* and *somniferum*. Some phenotypically wild *setigerum* accessions exhibited intermediate genetic features, suggesting introgression events. Two major populations were found in *somniferum* and, to some extent, these correspond to differences in seed form. These two populations may reflect recent attempts to breed varieties rich in opiates, as opposed to varieties used for poppyseed production. This study supports the idea that opium poppy cultivation began in the Western Mediterranean, with *setigerum* as the wild progenitor, although some wild varieties are likely to be feral forms, which can confound domestication studies.

This article is part of the theme issue 'Unravelling domestication: multi-disciplinary perspectives on human and non-human relationships in the past, present and future'.

© 2025 The Author(s). Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

Figure 1. Population structure of 186 accessions of wild and cultivated *Papaver* based on 4102 SNPs (Dataset1) discovered by GBS. (A) PCA: wild accessions are shown as open points and cultivated *somniferum* as full points. (B) STRUCTURE $K = 2$ and $K = 5$ models.

When only *somniferum* and *setigerum* accessions are analysed (Dataset2), most wild accessions are separated from the domesticated ones (figure 2). Three Italian and one Spanish *setigerum* accessions are genetically closer to the domesticated group (PAP 843, PAP 373, PAP 371 and BGE 032578) in the PCA plot (figure 2A). The latter accession is identical to a cultivated *somniferum* also from Spain (PAP 400). As these came from different seed banks, it is unlikely this similarity results from mislabelling or gene flow during regeneration of collections. Likewise, *setigerum* PAP 371 and *somniferum* PAP 760 are genetically similar. ΔK indicates $K = 3$ and $K = 5$ as the most likely models in this dataset. In the $K = 3$ model of STRUCTURE, the four abovementioned *setigerum* accessions are seen as sharing the same gene pools as *somniferum* (blue and red in figure 2B) with only a small contribution of *setigerum* alleles (dark red). Similarly, the *somniferum* PAP 400 and PAP 760 have a small contribution of the *setigerum* gene pool. All other *setigerum* are quite distinct from the cultivated forms, with the PCA suggesting a separation between plants from France, Italy and Spain in one hand, and plants from Portugal. There is also a *somniferum* accession from Romania (PAP 770) that resembles a hybrid between *setigerum* and *somniferum*. The *somniferum* accessions are divided into two distinct groups (blue and red) along the principal component 1 (PC1) axis (electronic supplementary material, S5) with STRUCTURE suggesting a high degree of allele sharing between these two groups in some accessions (figure 2B). The phylogeny produced by MrBayes confirms the monophyly of cultivated *somniferum* and an early separation between Pop1 (red) and Pop2 (blue) (electronic supplementary materials, S5 and S6).

Pop3 (dark red), which includes all the bona fide *setigerum* accessions, has the highest genetic diversity measured by H_E and H_O, whereas the two cultivated Pop1 and Pop2 have similar levels of diversity (table 1).

The geographical distribution of the $K = 3$ populations shows that accessions throughout Europe receive their alleles from Pop1 (red), particularly the Central and Northern European ones, and the same is true of Chinese accessions (figure 3). Accessions clustered mostly in Pop2 (blue) are in smaller number and are more common in North Africa, south Europe, Turkey, Afghanistan, India and Japan. The geographical distribution of alleles derived from Pop3 (dark red) reflects the distribution of *setigerum* accessions and *somniferum* in their vicinity. The exception is a *somniferum* from Romania with a high proportion of alleles derived from Pop3. The geographical distribution of the sub-populations identified in the $K = 5$ model shows Pop2B (purple) encompassing accessions from India; Pop2A (blue) encompassing accessions from Turkey, North Africa and Italy, plus one accession from Scandinavia; Pop1B being predominant in East European and East Asian accessions and Pop1A being widely distributed in Europe (electronic supplementary material, S7).

(c) Introgression and selection

The ABBA–BABA test detected introgression between *somniferum* Pop1 (red) and *setigerum* (D statistic: 0.2506; Z score: 4.4461). *Papaver orientale* was used as an outgroup (electronic supplementary material, S8). Those accessions considered hybrid or

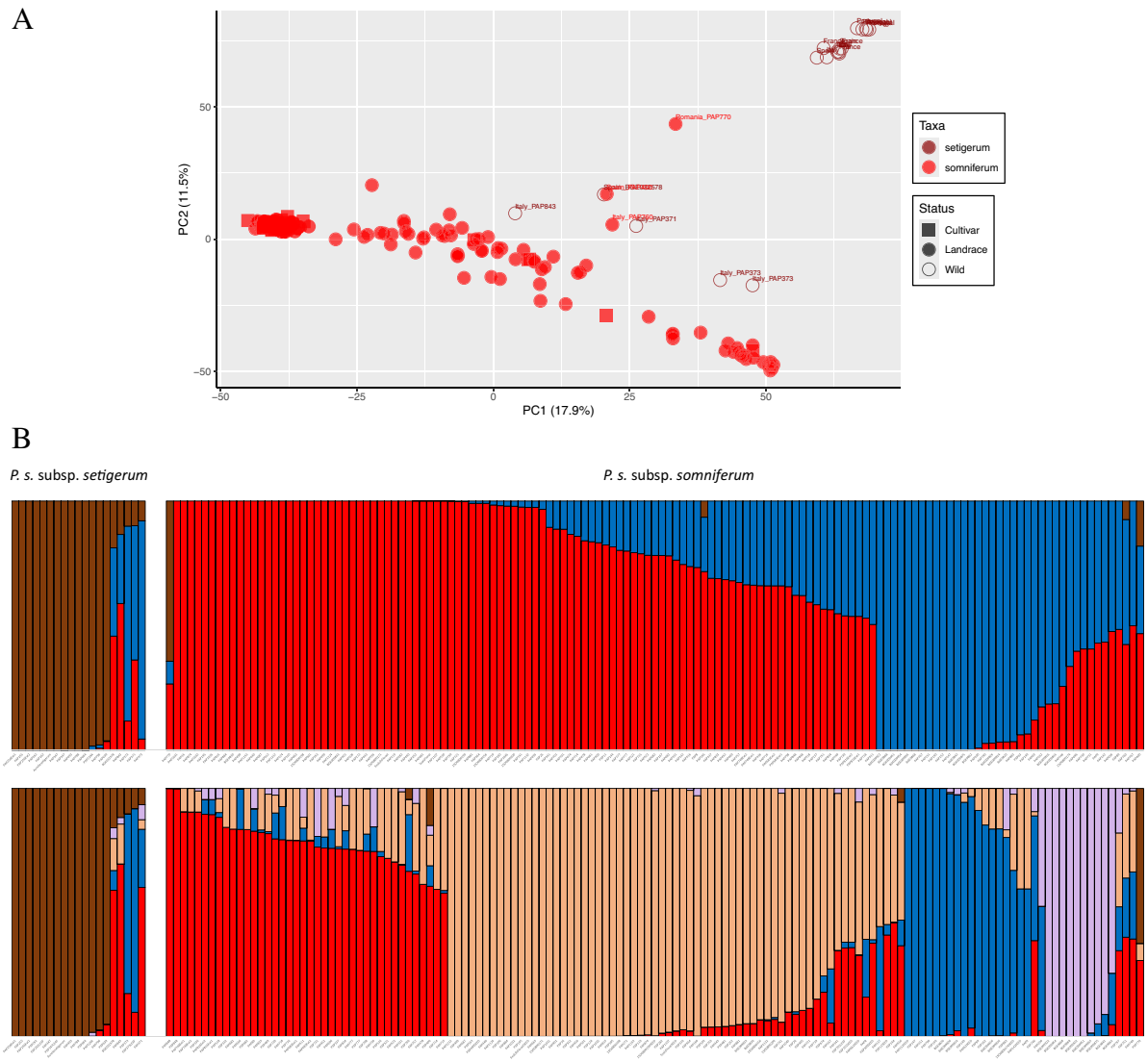

Figure 2. Population structure of 158 accessions of *Papaver somniferum* and *Papaver setigerum* based on 7048 SNPs (Dataset2) discovered by GBS. (A) PCA: accessions described as wild in the genebank passport data are shown as open points and those referred to as domesticated *somniferum* as full points; names of accessions referred to in the text are noted. (B) STRUCTURE $K = 3$ and $K = 5$ models.

not clearly clustered in the population structure analysis were excluded. The Patterson's *D* statistic allows the detection of introgression but not the quantification of how much of the genome is affected by it [94].

Both the GLM and MLM GWAS for the 'domesticated' trait in *P. somniferum* produced inconclusive results, with every chromosome having a series of SNPs associated with the trait (electronic supplementary material, S9). In the GLM GWAS, however, some outlier markers were observed in chromosomes 1, 3, 4, 7, 9, 10 and 11. BLASTING the sequences of the 20 SNPs with highest −log[10] *p*-values yielded an association with seven gene ontology annotated genes (electronic supplementary material, S10). These included, from the *P. somniferum* genome, an agmatine coumaroyltransferase-2-like protein gene (defence against pathogens in *Arabidopsis*), a deoxyloganetic acid glucosyltransferase-like (involved in the synthesis of monoterpene indole alkaloids accumulation) [95], a sodium/hydrogen exchanger 7-like gene (salt tolerance in soybean [96,97]), a chloroplast oxidoreductase, an AT-hook motif nuclear localized promotion factor gene (involved in a wide range of plant development processes [98,99]), as well as a methylenetetrahydrofolate reductase (lignin synthesis [100]) in *Brassica rapa*.

The GLM and MLM GWAS for seed colour detected SNPs associated with this trait in chromosomes 6 and 2, respectively, (electronic supplementary material, S11). One of these, in chromosome 6 of the *P. somniferum* genome is a 2-oxoglutarate dehydrogenase, shown to be involved in seed production, weight and number, as well as plant growth in *Arabidopsis* [101,102] (electronic supplementary material, S10). Another SNP detected by GWAS in chromosome 6 is associated with the phosphomethylpyrimidine synthase gene, involved in thiamine (vitamin B) biosynthesis, a compound that affects plant health, yield and immunity [103,104]. Moreover, an outlier SNP is located in the gene of the trifunctional UDP-glucose 4,6-dehydratase/UDP-4-keto-6-deoxy-D-glucose 3,5-epimerase/UDP-4-keto-L-rhamnose-reductase RHM1-like enzyme, which catalyses reactions in the cell wall synthesis and in flavonol biosynthesis [105,106]. Genetic diversity (π per base-pair) and Tajima's *D* along the chromosomes was different between *setigerum* and *somniferum* for two of the zones of chromosome 6, between positions 54 and 62 000 000 and around position 190 000 000, where the markers detected by GWAS were positioned (electronic supplementary materials, S10 and S12). This suggests selection acting in these regions.

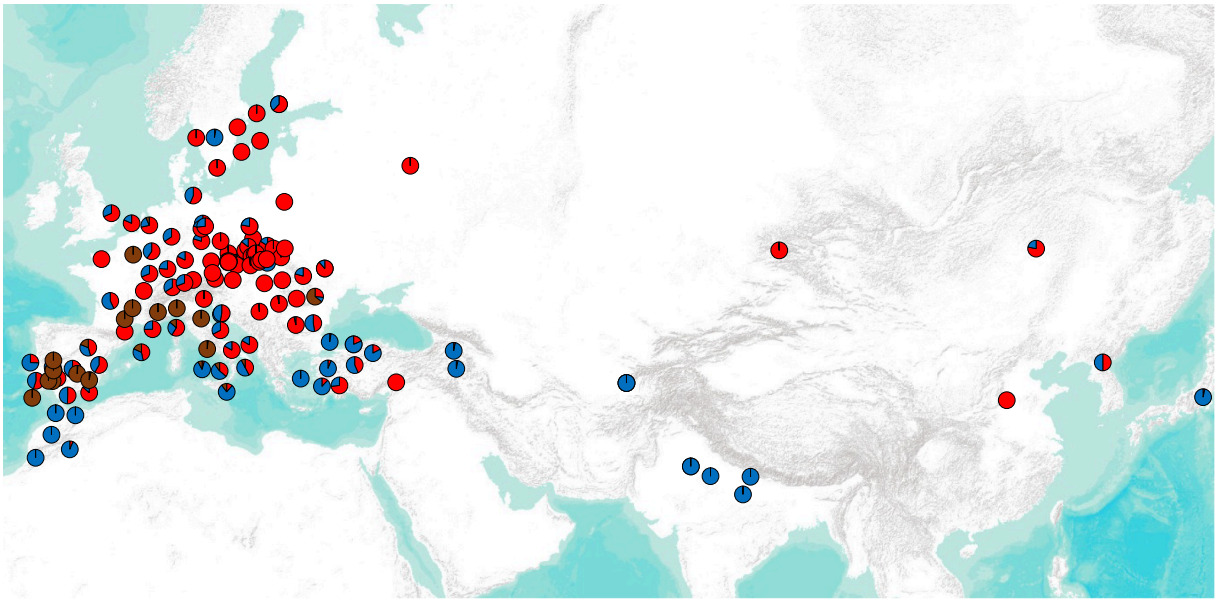

Figure 3. Geographical distribution of the three clusters identified in the STRUCTURE model $K = 3$ for *setigerum* and *somniferum* accessions. Each pie chart represents an individual accession and the proportional memberships to each of the three clusters.

To investigate associations between allelic frequencies and environmental plus geographical variables (for those accessions with known provenance), a gradient forest method was used. To validate our analyses, we compared these results with those run from randomized matches between environmental data (while maintaining the geographical distance). The run with the original data performed better than the random counterpart (considering the accuracy importance data; electronic supplementary material, S13). However, no strong associations with the variables tested were found; the highest associations were with the components of the principal coordinates of neighbourhood matrix (PCNM—based on the Euclidean distance between geographical provenance of the accessions) (weighed R^2 values approx. 0.01 in both runs). Likewise, the Mantel tests did not find strong associations between pairwise genetic distances and pairwise differences in geographical distances and environmental differences (electronic supplementary material, S14). The best correlation was with maximum temperature of the warmest month ($r = 0.5379$, $p = 0.0001$).

To further investigate environmental associations with genetic differences, we repeated these analyses considering only the accessions in Pop1 (red) and Pop2 (blue), as defined by STRUCTURE and PCA (figure 2B and electronic supplementary material, S5; $K = 3$). For Pop1 (red), the Mantel test found very week correlations with insignificant p values (electronic supplementary material, S14), the highest of which ($r = 0.1457$) was with precipitation of driest quarter. Accordingly, the performance of both gradient forest runs (random and real data) was similar, with the highest association detected by the gradient forest between genetic data and PCNM1 and elevation having a weighed R^2 values approximately 0.01 (figure 4A). For Pop2 (Blue), the Mantel test detected a significant high association with geographical distances ($r = 0.7029$) and the mean temperature of the coldest quarter ($r = 0.549$). The forest gradient test found a very weak association with PCNM1 (weighed R^2 values < 0.01; figure 4B).

(d) Seed geometric morphometrics

The first PC of the PCA to investigate the overall shape of the seeds accounts for more than two thirds of the total shape variability (67.2%) and captures the depth of the hilum (figure 5A). The second PC accounts for 11.9% of the total shape variability and captures the skewness of the poppy seed (the asymmetry between the two parts above and below the hilum on the lateral view). The two taxa roughly occupy the same morphological space on the PC1 × PC2 factorial plane (figure 5B), and *somniferum* appears to have a slightly more variable shape occupancy of the global morphological space. This is confirmed by the reconstructed mean shapes (figure 5B).

A single and balanced ($n = 150$) LDA was calculated on the SNP ($K = 5$) population attributions provided by STRUCTURE (figure 6). Class accuracies, the proportion of the 150 seeds correctly classified in their actual group, as obtained with leave-one-out cross-validation, are good for the K5_4 and K5_5 groups (i.e. Pop 2A and 2B, respectively) (>0.70) but below 0.5 for the three others. (electronic supplementary material, S15). This is reflected in the distribution of the scores of these groups onto the four linear discriminant axes of this single LDA (figure 6), which shows that LD1 clearly separates K5_5 (Pop2B) and, in a lesser extent K5_4 (Pop2A), from the other groups.

4. Discussion

(a) Genotype-by-sequencing in *Papaver* accessions

The GBS method proved effective for inferring evolutionary relationships between cultivated *somniferum* and other species of the *Papaver* genus. The method was applied before to a panel of opium poppy accessions although this only included three

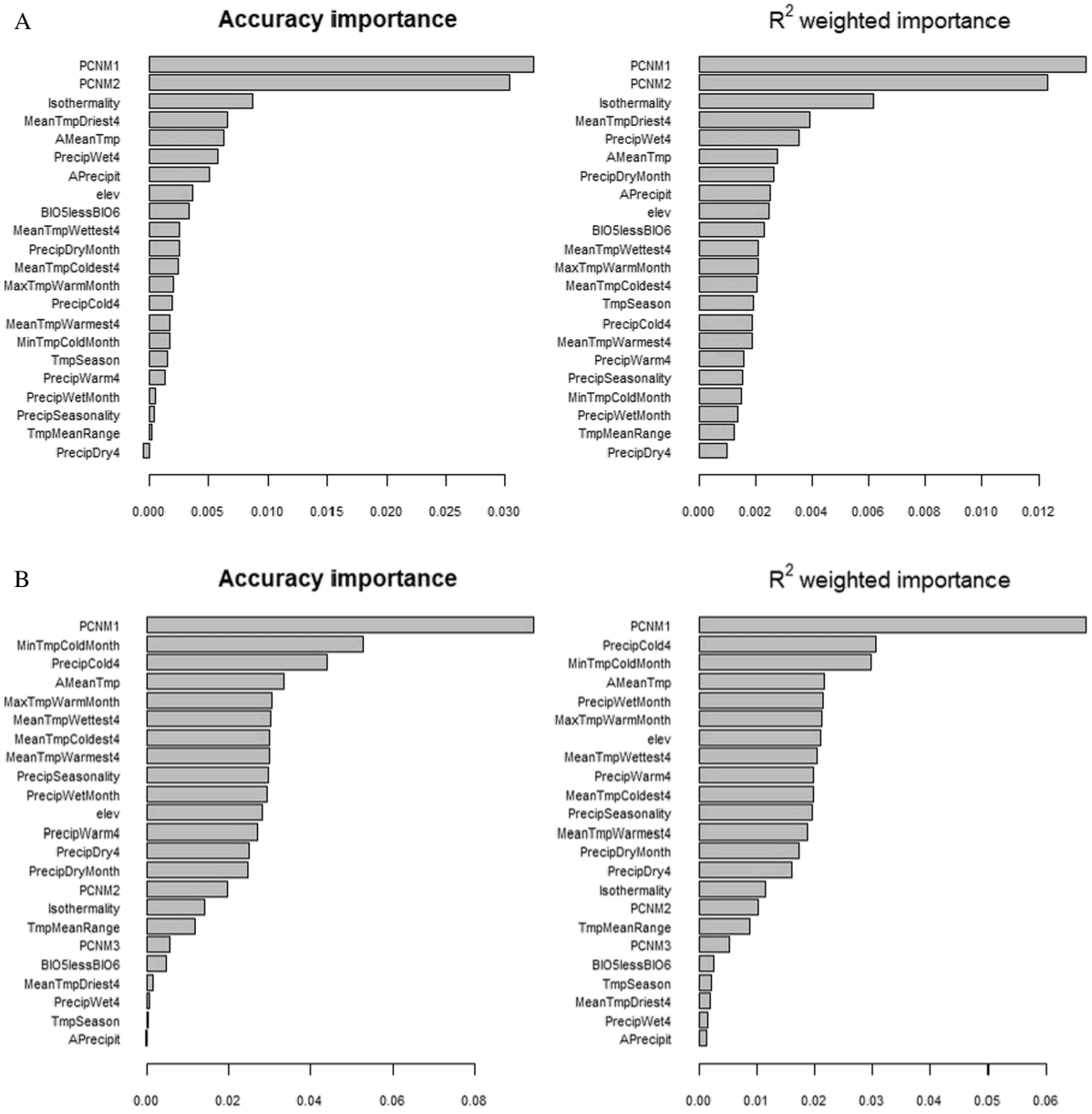

Figure 4. Association between environmental variables and allelic frequencies as well as between geographical and genetic distances, for the $K = 3$ model defined Pop1 and Pop2 (*setigerum* and *somniferum* accessions). (A,B) Random forest test for association between environmental variables and allele frequencies for Pop1 and Pop2, respectively.

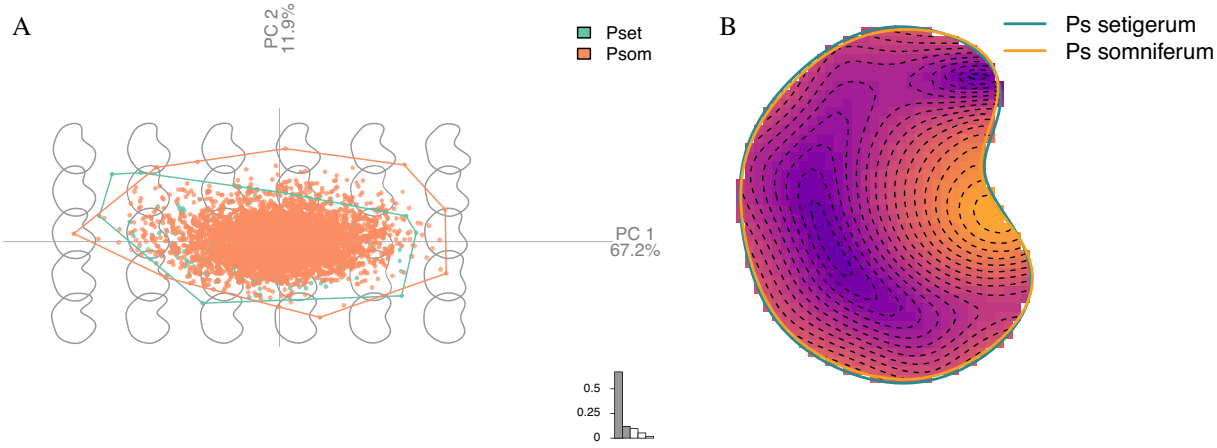

Figure 5. (A) GMM PC1 × PC2 factorial plane of the PCA on *Papaver* sp. shape variables. (B) Reconstructed mean shapes for the two poppy taxa shown with the intensity of differences between the two theoretical mean outlines.

setigerum [42]. We analysed a panel composed of *somniferum* accessions, a set of *setigerum* from this taxon's distribution range and accessions of other wild taxa in the *Papaver* genus, to understand the domestication process. As a variant calling method,

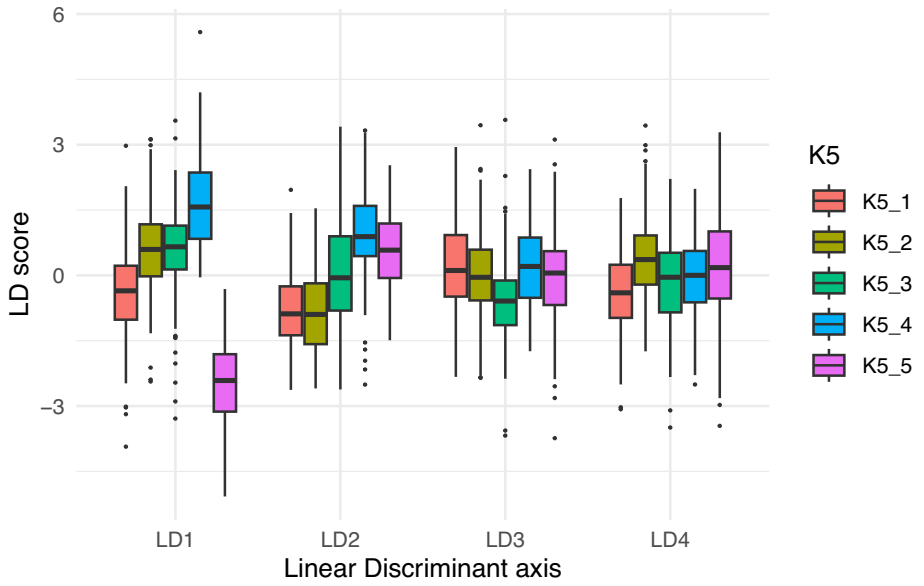

Figure 6. Scores on the four linear discriminant axes on a single and balanced ($n = 150$) LDA on the $K = 5$ grouping provided by STRUCTURE.

mapping of reads against the *P. somniferum* reference genome with FreeBayes generated a large amount of missing data in the final VCF file (table 1), particularly in individuals of taxa other than *P. somniferum* (44%). There was also three times more missing data in *setigerum* (1.2%) than in *somniferum* (0.4%) accessions, reflecting both the similarity and slight genomic distinctiveness of these two taxa. Despite the differences between species within *Papaver,* and the lack of a reference genome for species other than *P. somniferum,* we opt not to use other variant-call pipelines that work without a reference genome, such as STACKS [107]. Some literature suggests it is unlikely to improve the number of SNPs detected [108], and the VCF file generated has less annotations [109]. Moreover, the method used generated a high number of SNPs across all taxa, sufficient for the determination of population structure and other type of analysis. It must be noted that this reference genome bias might result in the underestimation of genetic diversity in the wild taxa or a distortion of divergence times in phylogenetic trees.

The excessive amount of missing data for four accessions (PI304971, C42, D52 and PI688271) can be explained by the taxonomic distance between *P. somniferum* and their species (*P. hybridum, P. occidentale* and *P. nudicale*) hindering mapping of reads to the *somniferum* reference genome or by technical issues during library preparation. A PCA with all the 190 accessions (based on a dataset containing no missing data with only 36 SNPs) hinted that these three taxa are in fact quite distinct from all the others (electronic supplementary material, S16).

PCAs computed on Dataset2 (assuming diploidy) and Dataset3 (assuming tetraploidy) for the *setigerum* and *somniferum* accessions did not show significant differences in population structure (electronic supplementary materials, S17A and S17B, respectively). This was a simple *in silico* way to address the fact that some *setigerum* accessions are described in the literature as tetraploid [7]. These data show that regardless of ploidy level, population structure and domestication history can be inferred from SNP data and are unaffected by differences in ploidy level. However, this might not be the case for other types of evolutionary analysis. The only noteworthy difference was *somniferum* accession PAP 770 (Romania) which was closer to the *setigerum* population than to *somniferum* in the diploid set but clustered with *somniferum* in the tetraploid set. It might be that this accession is a tetraploid. In the future, this will have to be confirmed by flow cytometry, as the accessions studied before [7] were not part of our panel. Replicates included in the panel (both from the same DNA extraction in different plates, and different plants within the same accession) displayed the expected genetic similarity that validates the method.

(b) Wild progenitor and place of domestication of the opium poppy

In our panel, we included all *setigerum* accessions available in germplasm banks, and accessions of 13 other *Papaver* species, to see their relationships with the cultivated *somniferum*. PCAs, STRUCTURE and phylogenies all indicate that none of the 13 species tested is related to *somniferum* (figure 1A,B, electronic supplementary materials, S3 and S4). STRUCTURE also shows that there is no gene flow between these species and *somniferum,* although shared proportional membership to the modelled Pop3 (dark red) population is found between *setigerum* and *P. dubium*, *P. fugax* and *P. armeniacum* (figure 1B). These two latter taxa are also the ones suggested to be genetically closer to *P. somniferum* (including *setigerum*). In a recent study of phylogenetic relationships within the Papaveraceae family, based on two nuclear and seven plastid DNA regions, *P. armeniacum* is placed in a sub-clade within the *Papaver* 3 clade [93]. Likewise, *P. dubium* is also the phylogenetically closer taxa to *P. somniferum* in other studies [51,110–112], although the available literature is not consistent in this regard [49]. Our data are based on nuclear genome-wide markers, and it confirms these phylogenetic studies based mostly on chloroplast genomes and ribossomal DNA sequences. Importantly for the *Papaver* genus taxonomy, *P. orientale* and *P. pseudo-orientale* are genetically close, as well as *P. rhoeas* and *P. umbonatum*, and *P. fugax* and *P. armeniacum* (figure 1, electronic supplementary material, S4).

Our results argue against the hypothesis of *setigerum*/*somniferum* having originally evolved in the Fertile Crescent area, being introduced in Europe as weeds brought by the first farmers. They are genetically distinct from any of the *Papaver* species tested and, at least presently, *setigerum* is not found in the Fertile Crescent. Moreover, archaeobotanical evidence for *somniferum*

seed remains appear in the Fertile Crescent only in later medieval contexts [9]. Of course, we cannot rule out the possibility that the wild progenitor of *somniferum* is indeed some other *Papaver* taxon native to the Fertile Crescent region that was not included in this study. However, based on the current archaeobotanical evidence we consider the 'Western Mediterranean origin' hypothesis as the most likely.

Somniferum and *setigerum* are clearly distinct from all the other *Papaver* species, but what is the relationship between them? Population structure analysis on Dataset2 shows they have very distinct gene pools, with the PCA showing a clear separation between *setigerum* accessions (PC1 and PC2 values above 50) relative to the *somniferum* accessions (figure 2). In this sense, and considering the whole panel population structure (where *somniferum* and *setigerum* form a cluster apart from other species), these two taxa are better described as subspecies of the same *P. somniferum* species. The exceptions to the distinct gene pool of these two taxa are the eight *setigerum* and *somniferum* accessions (named in figure 2) that stand intermediate to both clusters. Likewise, STRUCTURE models show these eight accessions share alleles from both the *setigerum* (dark red in figure 2B) and *somniferum* population (blue and red). Human error during library preparation is unlikely, as these samples were either in different plates or sections of the same plate. Replication in germplasm banks is also implausible. For example, according to the passport data, accession PAP 400 (*somniferum*) was collected by IPK in 1978 in Aranjuéz, near Toledo (Spain), whereas BGE032578 (*setigerum*) was collected by CRF on a roadside in Alcala de Henares. On the other hand, these two locations are only 50 km apart, so it is likely that these two accessions belong to the same population. Likewise, accession PAP 760 (*somniferum*) was collected in Calabria, Italy, in 1987, whereas accession PAP 371 (*setigerum*) was collected in 1979 near Palermo, Sicily. Could the seed banks have made a mistake in their taxonomical classification? Although not impossible, the two taxa have distinguishing features that should set them apart by a trained plant taxonomist. Previous comparisons of *setigerum* and *somniferum* based on genomic data by Hong *et al.* [42] did not find any outliers although they only included three *setigerum* accessions in their panel.

If we exclude errors, how do these eight accessions relate to domestication narratives? One possibility is that *setigerum* is indeed the sole wild progenitor of *somniferum*. The fact that this is the only taxa genetically closer to *somniferum* argues for this view. In this case, the outlier *setigerum* accessions can be descendants from the original wild population whose cultivation would lead to the domesticated *somniferum*. The fact they come from Italy and Spain, where archaeobotanical data suggest the earliest *somniferum* was cultivated, would support this. It is also noteworthy that two clusters of *setigerum* are evident in the PCA: one that includes accessions from Italy, France and Spain (closer to *somniferum*) and another (more distant from *somniferum*) with accessions from Portugal. The coincidence of the distribution range of *setigerum* with the earliest seed remains of *somniferum* in sites like La Marmotta, Le Taï or La Draga (all dated to 5500–4800 cal BCE) has long been pointed as an argument for the Western Mediterranean being the place of *somniferum* domestication [10]. The genetic proximity of *setigerum* accessions from these same regions to *somniferum* further strengthens the 'Western Mediterranean origin' hypothesis. Also, ABBA–BABA indicates a high level of introgression between Pop1 and *setigerum* (electronic supplementary material, S8), which could be the result of sympatry between the wild and domesticated forms, before the expansion of the cultivation range of *somniferum*. In this scenario, accessions in Pop1 probably represent the earliest domesticated forms and the area of distribution of *setigerum* acted as a hybrid zone where gene flow between wild and cultivated forms occurred. It is possible that this gene flow hindered the fixation of domestication traits in *somniferum* until landraces were established away from the natural distribution of *setigerum*, namely in Central and Northern Europe or in the Eastern Mediterranean. In these regions, full domestication and selection for specific traits would have occurred. Around Italy, Spain and south France, it would have been cultivated but the presence and introgression of wild *setigerum* would have delayed full domestication. This period of pre-domestication cultivation of wild and semi-domesticated forms, prior to the emergence of fully domesticated forms, has been proposed for other crops like wheat and barley [113].

The phylogeny produced with MrBayes (electronic supplementary material, S6) does not allow a clear picture of the relative divergence of Pop1 and Pop2. Phylogenetic methods to infer times of divergence for recent events, like domestication, are too imprecise in the absence of whole-genomes and other computational methods such as Globetrotter [114,115]. Seed morphometry indicating that *setigerum* is the only taxa within *Papaver* with seeds like *somniferum* [9] also supports the hypothesis of *setigerum* as wild progenitor. Although *setigerum* has much lower content of opiates than *somniferum*, the antispasmodic papaverine can reach high levels in *setigerum* [116]. *Setigerum* is the only taxon—other than *somniferum*—known to have the necessary metabolic steps for the biosynthesis of morphine and codeine [46,117]. It is possible that *setigerum* started to be cultivated as a medicinal plant, with a selection for varieties with higher levels of morphine and codeine occurring at a later stage. It is known that levels of BIAs can vary in *somniferum* depending on water stress, pathogenic agents or soil concentrations of zinc and cadmium [118–120]. Variations in these environmental factors may have originated *setigerum* populations that were richer in alkaloids, their growth and reproduction promoted by the first farmers in Western Europe.

An alternative hypothesis is that *setigerum* is not the progenitor of *somniferum* and that the latter was never domesticated, merely cultivated. As such, it never had a wild progenitor, and we cannot refer to a 'domestication syndrome'. In this case, all *setigerum* would be feral forms of *somniferum* in different stages of ferality and genetic drift. Much literature reports that neither *setigerum* nor other *Papaver* species analysed have the high levels of opiates found in *somniferum* (morphine, codeine, thebaine, noscapine and papaverine) [23,121–127]. If the levels of bioactive alkaloids are so low in *setigerum*, it is difficult to explain how it got selected for cultivation and eventual domestication, especially for medicinal or mind-altering purposes. Unless, of course, *setigerum* was mainly used for food purposes or if the low alkaloid contents of *setigerum* were, nonetheless, mildly effective painkillers or narcotics. If, however, we assume that *somniferum* was already growing naturally in Western or Central Europe, the first farmers there may have discovered its properties, or learned them from local Mesolithic groups and cultivated it (no poppy seed remains have been recovered in Mesolithic contexts yet). A similar scenario has recently been proposed for another dual-purpose crop, *Cannabis sativa* (psychoactive drug and hemp fibre source), with no wild progenitor and with present-day

hemp and drug varieties descending from an ancestral gene pool in East Asia that exists today as a mixture of feral plants and landraces [128].

The fact that both *setigerum* and *somniferum* possess the morphine synthesis pathway could be explained by *setigerum* being the wild progenitor (previous hypothesis) or, indeed, by the two taxa sharing a recent common ancestor. In this scenario, those outlier *setigerum* accessions would be feral forms: once cultivated *somniferum* that acquired a wild morphotype (e.g. PAP371 would be a feral form of PAP760). Feralization of opium poppy has been described not just in the Mediterranean area but also in locations where *setigerum* is unlikely to occur spontaneously, such as Korea or New Zealand [126,129–131]. In this case, those *setigerum* outlier accessions would be recent feralized forms of their genetically closer *somniferum* accessions (e.g. PAP371 would be a feral form of PAP760). The other bona fide *setigerum* would be a separate species that diverged recently from a common ancestor to *somniferum* (maybe through a polyploidization event, which would explain the tetraploid forms of *setigerum*). This is supported by the distinctiveness of the Pop3 (dark red) gene pool in STRUCTURE (figure 2). Moreover, the high number of differences that have been reported at the chloroplast genome level between *setigerum* and *somniferum* strengthens the case for them being two separate species [48].

Another possibility is that the outlier accessions are hybrids between *setigerum* and *somniferum*. Although considered a self-pollinator, *somniferum* is known to have varying degrees of outcrossing, including with *setigerum* [132]. This would explain why these eight accessions are separated from both the *setigerum* and *somniferum* clusters in the PCA. STRUCTURE suggests the accessions classified as *setigerum* are in fact *somniferum* introgressed by *setigerum* (they get most of their alleles from the *somniferum* blue and red populations), and that the Romanian accession PAP 770, on the contrary, is probably a *setigerum* introgressed by a *somniferum* (it gets most of its alleles from the *setigerum* 'dark red' population). This is a more likely scenario for those eight accessions, although it would fit both the '*setigerum* as wild progenitor' and the '*setigerum* and *somniferum* as independent species' hypotheses.

Future work will be needed to disentangle these scenarios. This will include (i) whole genome sequencing of *somniferum* and *setigerum* accessions; (ii) use of computational methods that disentangle hybrid swarms from feralization events; (iii) RNA-seq experiments to determine which genes in which metabolic pathways are exclusive to *somniferum*; (iv) field collections of *setigerum* plants for karyotyping alongside *setigerum* accessions held in germplasm banks. The latter should sample locations where *setigerum* has been described in disparate literature, but no records found (e.g. Canary Islands and North Africa) as well as putatively feral accessions from Korea and New Zealand. Reflecting the archaeobotanical data, we consider the hypothesis that *setigerum* is indeed the wild progenitor of *somniferum,* and that the outlier accessions are hybrids between *setigerum* and *somniferum,* to be the most parsimonious one.

(c) Domestication syndrome in the opium poppy

The domestication syndrome is the suite of developmental, metabolic and phenotypic traits that differentiate a crop from its wild ancestor(s) [133]. In the case of opium poppy, this has rarely been studied, except for the difference in BIA levels and leaf shape [134]. In addition to the detection of genomic differences between *setigerum* and *somniferum,* we used GMM to investigate differences in the seed shape of the two taxa. Results showed that the morphometric differences between seeds of the *setigerum* and *somniferum* accessions shown here affected mostly the hilum area (figure 5), suggesting that the selective pressure associated with domestication of the opium poppy affected seeds (if we consider that *setigerum* is indeed the wild ancestor). It is unclear if this was a secondary effect of selection for some other traits or if the use of seeds (in culinary or oil production) led to selection for larger seeds.

Interestingly, GWAS also detected SNPs affected by domestication that are associated with seed weight and colour, supporting the suggestion that seeds were one of the organs most affected by domestication (assuming *setigerum* is indeed the wild progenitor). Other SNPs detected by GWAS are associated with cell wall synthesis and plant defence against pathogens, traits that have been recently discovered to be involved in the domestication process of other crops, such as barley and tomato [135,136]. Intriguingly, no gene involved in the synthesis of alkaloids was detected, suggesting that differences between *setigerum* and *somniferum* might be due not so much to the selection of genetic variants but to changes in the expression levels of particular genes. This has been shown for the domestication of cotton and maize [137–140]. One SNP was associated with a gene (deoxyloganetic acid glucosyltransferase-like) involved in alkaloid accumulation, reinforcing the idea that domestication involved changes in the level of specific alkaloids rather than the ability of them being synthesized. Further GWAS with alkaloid level phenotyping as well as an RNA-seq study comparing gene expression between low and high alkaloid content varieties (as well as wild *Papaver* accessions) could elucidate these issues.

(d) Spread of opium poppy cultivation

We detected two major populations within *somniferum* each subdivided into two sub-populations (figure 2B). This agrees with a previous study on population structure of cultivated *somniferum* [42]. The geographical distribution of these populations likely reflects the spread of opium poppy cultivation. If so, there seems to have been two major stages in this process, which can also explain the separation of the two populations in the PCA plot (electronic supplementary material, S5). Pop1 is the most cosmopolitan and encompasses most accessions, including those that occur in the putative initial domestication area (Western Mediterranean). This can be interpreted as corresponding to the earliest cultivated stock that spread during the Neolithic period throughout Europe. This is supported by the fact that ABBA–BABA detects significant introgression between this population and *setigerum*, which does not happen with Pop2, despite the fact both populations have comparable levels of genetic diversity

(table 1). Pop1B predominates in Eastern and Northern Europe, whereas Pop1A is more (broadly speaking) present in Western and Southern Europe. The fact that this geographical distribution is not clearly demarcated suggests that movement of cultivars occurred later. For example, accessions in China belong to Pop1B, and historical sources indicate it was introduced there during recent historical times [1]. The distribution of Pop2 in North Africa, Turkey, Afghanistan and India suggests this might correspond to accessions cultivated recently during the Arabic, Ottoman and Mogul periods, most likely for the purpose of obtaining opium [1]. All accessions from Afghanistan and India belong to Pop2B, as expected from a recent introduction of *somniferum* in this region. However, in the phylogeny shown in electronic supplementary material, S6, some Pop2 accessions seem to be ancestral to all the other *somniferum* and Pop1 is monophyletic. This might suggest that the separation between these two populations started early on during the domestication process, and it might just be that Pop2 accessions existed in the core domestication region in the past but are no longer found there. Two independent domestication events are not supported by the phylogeny or PCA analysis.

This geographical distribution cannot be explained by adaptation to different environmental regimes as the associations between allele frequencies and environmental variables were weak, both for the whole *somniferum* panel and for the two populations. This contrasts with other crops for which environmental adaptation played a role in shaping the geographical distribution of genetic diversity [141–144]. The absence of an environmental effect in *somniferum* genetic diversity might reflect a small-scale cultivation, practised mostly in gardens rather than in extensive rainfed fields, like cereals or some legumes. Nevertheless, results suggest that geographical distances and amount of rainfall have an effect—albeit a small one—on the distribution of genetic diversity in *somniferum*. The weak association observed between genetic diversity and environmental variables may be due to insufficient sampling or phenotyping, and this would have to be validated by common garden experiments [113].

Interestingly, there is also a correspondence between seed shapes defined by GMM and the clusters determined by STRUCTURE in the $K = 5$ model (figure 6 and electronic supplementary material, S15). GWAS also revealed SNPs associated with seed colour in chromosome 6. Could the (mostly abandoned) taxonomical categories of *somniferum*, *album* and *nigrum*, based on seed colour reflect the predominant use for opium production or culinary purposes [145]? Unfortunately, passport data for *somniferum* accessions do not indicate the end-use, context of its cultivation or opiate content, so this is difficult to infer. The study by Wong *et al.* [146] does not corroborate this hypothesis; though, from the six accessions with the highest codeine content, some had yellow others grey or black seeds. In addition, most *somniferum* accessions (but not *setigerum*) had high morphine contents regardless of seed colour.

The genomic and seed GMM data here help answer the questions proposed. (i) None of the most common *Papaver* species was the wild progenitor of *somniferum* or had any genetic contribution to its gene pool; *setigerum* was the wild progenitor or, also compatible with the data, *somniferum* does not have a wild progenitor, being the cultivated form of a naturally occurring species of which *setigerum* is a feral form. (ii) Considering the archaeobotanical data and the geographical distribution of genetic diversity in *somniferum*, the Western Mediterranean is the place where this species was originally domesticated (or cultivated); the hypothesis of an origin in the Fertile Crescent area followed by an introduction in Europe as a weed can be discarded. (iii) Phylogenetic analysis corroborates a single domestication wave, in the aforementioned region. (iv) After or during the domestication process, there were two major gene pool formation events in *somniferum*, the second of which is likely associated with the cultivation of varieties meant for opium production. (v) In addition to the higher content in BIAs, seed shape, cell walls and biotic defence mechanisms (and genes involved in these traits) were affected by the domestication process.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. The raw sequence data datasets were deposited in the BioProject database of NCBI [147]. The raw vcf file for Datasets 1 and 2 and raw data for the GMM analysis are available in [148].

Supplementary material is available online [149].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. R.S.R.M.: data curation, formal analysis, investigation, methodology, resources, software, visualization, writing—review and editing; V.B.: formal analysis, investigation, methodology, software, visualization, writing—review and editing; R.S.: investigation, resources; A.J.: formal analysis, investigation, methodology, resources, visualization, writing—review and editing; L.B.: methodology, writing—review and editing; A.E.: formal analysis, investigation, methodology, writing—review and editing; M.J.F.M.: formal analysis, investigation, methodology, resources, visualization, writing—review and editing; S.G.: resources, supervision, writing—review and editing; F.A.: conceptualization, funding acquisition, investigation, methodology, project administration, writing—review and editing; A.S.: conceptualization, funding acquisition, investigation, methodology, project administration, resources, writing—review and editing; H.R.O.: conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, visualization, writing—original draft.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. This work and RSRM grant have been funded by the Portuguese Fundação para a Ciência e Tecnologia (FCT) through a Projects ICDT grant (PTDC/HAR-ARQ/1709/2021); by the French ANR-Agence Nationale de la Recherche (ANR-21-CE27-0006); and by the ICArEHB Seed Fund (ISF). H.R.O. and M.J.F.M. were supported by a Concurso Estímulo ao Emprego Científico contract (CEECIND/00848/2017; 2021.01680.CEECIND; OWLDER-Old World Legume Domestication, Evolution and Resilience). A.E. was funded by a European Research Council grant (ERC-2019-STG-852573-DEMETER).

Acknowledgements. We would like to thank the germplasm banks that provided seeds for this study; Uyen Vu Thuy Hong and Mathew G. Lewsey (La Trobe University, Australia), for sharing with us the .vcf file of their study on opium poppy; Christian Parisod (University of Bern), for providing DNA extractions of P. occidentale; Roza Parol-Kryger, Katarzyna Wolinska-Griese, Berthold Fartman and Silke Winkler at LGC Genomics for assistance with bioinformatics and project planning.

References

1. Inglis L. 2018 *Milk of paradise: a history of opium*. London, UK: Macmillan.
2. Melo D, Álvarez-Ortí M, Nunes MA, Espírito Santo L, Machado S, Pardo JE, Oliveira MBPP. 2022 Nutritional and chemical characterization of poppy seeds, cold-pressed oil, and cake: poppy cake as a high-fibre and high-protein ingredient for novel food production. *Foods* **11**, 3027. (doi:10.3390/foods11193027)
3. Jesus A, Antolín F. 2022 Evidence of the presence and use of neolithic oil plants in the NW of the mediterranean and the Swiss plateau. In *The missing woodland resources*, pp. 23–40. Kooiweg, The Netherlands: Barkhuis Publishing. (doi:10.2307/j.ctv23wf366.5)
4. Food and Agriculture Organization of the United Nations. 2024 *Crops and livestock products*. See https://www.fao.org/faostat/en/#data/QCL (accessed 3 December 2024).
5. Statista. 2024 *Worldwide area for the illegal cultivation of opium poppy from 1994 to 2022*. See https://www.statista.com/statistics/268765/area-for-the-illegal-cultivation-of-opium-poppy/ (accessed 3 December 2024).
6. Kadereit JW. 1988 Sectional affinities and geographical distribution in the genus *Papaver* L. (Papaveraceae). *Beitrage Zur Biologie Der Pflanzen* **63**, 139–156.
7. Zohary D, Hopf M, Weiss E. 2012 *Domestication of plants in the old world*. Oxford, UK: Oxford University Press.
8. Carlin MG, Dean JR, Ames JM. 2020 Opium alkaloids in harvested and thermally processed poppy seeds. *Front. Chem.* **8**, 737. (doi:10.3389/fchem.2020.00737)
9. Jesus A, Bonhomme V, Evin A, Ivorra S, Soteras R, Salavert A, Antolín F, Bouby L. 2021 A morphometric approach to track opium poppy domestication. *Sci. Rep.* **11**, 1. (doi:10.1038/s41598-021-88964-4)
10. Salavert A *et al*. 2020 Direct dating reveals the early history of opium poppy in western Europe. *Sci. Rep.* **10**, 1. (doi:10.1038/s41598-020-76924-3)
11. Salavert A, Martin L, Antolín F, Zazzo A. 2018 The opium poppy in Europe: exploring its origin and dispersal during the Neolithic. *Antiquity* **92**, e1. (doi:10.15184/aqy.2018.154)
12. Bellwood P. 2022 *First farmers: the origins of agricultural societies.* Hoboken, NJ: John Wiley & Sons.
13. Fuller DQ, Denham T, Allaby R. 2023 Plant domestication and agricultural ecologies. *Curr. Biol.* **33**, R636–R649. (doi:10.1016/j.cub.2023.04.038)
14. Wu D, Lao S, Fan L. 2021 De-domestication: an extension of crop evolution. *Trends Plant Sci.* **26**, 560–574. (doi:10.1016/j.tplants.2021.02.003)
15. Scossa F, Fernie AR. 2021 When a crop goes back to the wild: feralization. *Trends Plant Sci* **26**, 543–545. (doi:10.1016/j.tplants.2021.02.002)
16. Guo W *et al*. 2020 Origin and adaptation to high altitude of tibetan semi-wild wheat. *Nat. Commun.* **11**, 5085. (doi:10.1038/s41467-020-18738-5)
17. Mabry ME *et al*. 2021 The evolutionary history of wild, domesticated, and feral *Brassica oleracea* (Brassicaceae). *Mol. Biol. Evol.* **38**, 4419–4434. (doi:10.1093/molbev/msab183)
18. Charbonneau A *et al*. 2018 Weed evolution: genetic differentiation among wild, weedy, and crop radish. *Evol. Appl.* **11**, 1964–1974. (doi:10.1111/eva.12699)
19. Gao G *et al*. 2024 Most Tibetan weedy barleys originated via recombination between Btr1 and Btr2 in domesticated barley. *Plant Commun.* **5**, 100828. (doi:10.1016/j.xplc.2024.100828)
20. Page A, Gibson J, Meyer RS, Chapman MA. 2019 Eggplant domestication: pervasive gene flow, feralization, and transcriptomic divergence. *Mol. Biol. Evol.* **36**, 1359–1372. (doi:10.1093/molbev/msz062)
21. Wang H, Vieira FG, Crawford JE, Chu C, Nielsen R. 2017 Asian wild rice is a hybrid swarm with extensive gene flow and feralization from domesticated rice. *Genome Res.* **27**, 1029–1038. (doi:10.1101/gr.204800.116)
22. Burger JC, Holt JM, Ellstrand NC. 2007 Rapid phenotypic divergence of feral rye from domesticated cereal rye. *Weed Sci.* **55**, 204–211. (doi:10.1614/ws-06-177.1)
23. Dittbrenner A, Mock HP, Börner A, Lohwasser U. 2012 Variability of alkaloid content in Papaver somniferum L. *J. Appl. Bot. Food Qual* **82**, 103–107. https://core.ac.uk/download/pdf/235697529.pdf
24. Lavania UC, Srivastava S. 1999 Quantitative delineation of karyotype variation in *Papaver* as a measure of phylogenetic differentiation and origin. *Curr. Sci.* **77**, 429–435.
25. Mallet J. 2007 Hybrid speciation. *Nature* **446**, 279–283. (doi:10.1038/nature05706)
26. Petit C, Bretagnolle F, Felber F. 1999 Evolutionary consequences of diploid–polyploid hybrid zones in wild species. *Trends Ecol. Evol.* **14**, 306–311. (doi:10.1016/s0169-5347(99)01608-0)
27. Price TD. 2000 *Europe's first farmers*. Cambridge, UK: Cambridge University Press.
28. van Willigen S. 2018 Between Cardial and Linearbandkeramik: from no-man's-land to communication sphere. *Quat. Int.* **470**, 333–352. (doi:10.1016/j.quaint.2017.08.031)
29. Huang X, Huang S, Han B, Li J. 2022 The integrated genomics of crop domestication and breeding. *Cell* **185**, 2828–2839. (doi:10.1016/j.cell.2022.04.036)
30. Kantar MB, Nashoba AR, Anderson JE, Blackman BK, Rieseberg LH. 2017 The genetics and genomics of plant domestication. *BioScience* **67**, 971–982. (doi:10.1093/biosci/bix114)
31. Olsen KM, Wendel JF. 2013 A bountiful harvest: genomic insights into crop domestication phenotypes. *Annu. Rev. Plant Biol.* **64**, 47–70. (doi:10.1146/annurev-arplant-050312-120048)
32. Liber M, Duarte I, Maia AT, Oliveira HR. 2021 The history of lentil (*Lens culinaris* subsp. *culinaris*) domestication and spread as revealed by genotyping-by-sequencing of wild and landrace accessions. *Front. Plant Sci.* **12**, 628439. (doi:10.3389/fpls.2021.628439)
33. Naeem M, Demirel U, Yousaf MF, Caliskan S, Caliskan ME. 2021 Overview on domestication, breeding, genetic gain and improvement of tuber quality traits of potato using fast forwarding technique (GWAS): a review. *Plant Breed.* **140**, 519–542. (doi:10.1111/pbr.12927)
34. Kim H *et al*. 2016 Population dynamics among six major groups of the *Oryza rufipogon* species complex, wild relative of cultivated Asian rice. *Rice* **9**, 56. (doi:10.1186/s12284-016-0119-0)
35. Swarts K *et al*. 2017 Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. *Science* **357**, 512–515. (doi:10.1126/science.aam9425)
36. Taranto F, D'Agostino N, Greco B, Cardi T, Tripodi P. 2016 Genome-wide SNP discovery and population structure analysis in pepper (*Capsicum annuum*) using genotyping by sequencing. *BMC Genom.* **17**, 943. (doi:10.1186/s12864-016-3297-7)
37. Mittell EA, Cobbold CA, Ijaz UZ, Kilbride EA, Moore KA, Mable BK. 2020 Feral populations of *Brassica oleracea* along Atlantic coasts in western Europe. *Ecol. Evol.* **10**, 11810–11825. (doi:10.1002/ece3.6821)
38. Acquadro A, Barchi L, Gramazio P, Portis E, Vilanova S, Comino C, Plazas M, Prohens J, Lanteri S. 2017 Coding SNPs analysis highlights genetic relationships and evolution pattern in eggplant complexes. *PLoS One* **12**, e0180774. (doi:10.1371/journal.pone.0180774)
39. Muli JK, Neondo JO, Kamau PK, Michuki GN, Odari E, Budambula NLM. 2022 Genetic diversity and population structure of wild and cultivated *Crotalaria* species based on genotyping-by-sequencing. *PLoS One* **17**, e0272955. (doi:10.1371/journal.pone.0272955)
40. Guo L *et al*. 2018 The opium poppy genome and morphinan production. *Science* **362**, 343–347. (doi:10.1126/science.aat4096)
41. Yang X *et al*. 2021 Three chromosome-scale *Papaver* genomes reveal punctuated patchwork evolution of the morphinan and noscapine biosynthesis pathway. *Nat. Commun.* **12**, 6030. (doi:10.1038/s41467-021-26330-8)

42. Hong UVT, Tamiru-Oli M, Hurgobin B, Okey CR, Abreu AR, Lewsey MG. 2022 Insights into opium poppy (*Papaver* spp.) genetic diversity from genotyping-by-sequencing analysis. *Sci. Rep.* **12**, 111. (doi:10.1038/s41598-021-04056-3)
43. Li Y, Winzer T, He Z, Graham IA. 2020 Over 100 million years of enzyme evolution underpinning the production of morphine in the papaveraceae family of flowering plants. *Plant Commun.* **1**, 100029. (doi:10.1016/j.xplc.2020.100029)
44. Xu T *et al*. 2022 A global survey of the transcriptome of the opium poppy (*Papaver somniferum*) based on single-molecule long-read isoform sequencing. *Plant J.* **110**, 607–620. (doi:10.1111/tpj.15689)
45. Gurkok T, Turktas M, Parmaksiz I, Unver T. 2015 Transcriptome profiling of alkaloid biosynthesis in elicitor induced opium poppy. *Plant Mol. Biol. Report.* **33**, 673–688. (doi:10.1007/s11105-014-0772-7)
46. Ziegler J *et al*. 2009 Evolution of morphine biosynthesis in opium poppy. *Phytochemistry* **70**, 1696–1707. (doi:10.1016/j.phytochem.2009.07.006)
47. Hu Y, Zhao R, Xu P, Jiao Y. 2018 The genome of opium poppy reveals evolutionary history of morphinan pathway. *Genom. Proteom. Bioinform.* **16**, 460–462. (doi:10.1016/j.gpb.2018.09.002)
48. Liu L, Du Y, Shen C, Li R, Lee J, Li P. 2020 The complete chloroplast genome of *Papaver setigerum* and comparative analyses in Papaveraceae. *Genet. Mol. Biol* **43**, 1–10. (doi:10.1590/1678-4685-gmb-2019-0272)
49. Kadereit JW, Schwarzbach AE, Jork KB. 1997 The phylogeny of *Papaver* s. l. (Papaveraceae): polyphyly or monophyly? *Plant Syst. Evol.* **204**, 75–98. (doi:10.1007/bf00982533)
50. Lack HW. 2019 The discovery and naming of *Papaver orientale* s.l. (Papaveraceae) with notes on its nomenclature and early cultivation. *Candollea* **74**, 47. (doi:10.15553/c2019v741a7)
51. Carolan JC, Hook ILI, Chase MW, Kadereit JW, Hodkinson TR. 2006 Phylogenetics of papaver and related genera based on DNA sequences from its nuclear ribosomal DNA and plastid trnL Intron and trnL–F intergenic spacers. *Ann. Bot.* **98**, 141–155. (doi:10.1093/aob/mcl079)
52. Cheng L, Lin Z, Xin C, Sun H, Li X. 2022 Molecular identification and phylogenetic analysis of *Papaver* based on ITS2 barcoding. *J. Forensic Sci.* **67**, 712–719. (doi:10.1111/1556-4029.14925)
53. Rohlf FJ. 1990 Morphometrics. *Annu. Rev. Ecol. Syst.* **21**, 299–316.
54. Rohlf FJ, Marcus LF. 1993 A revolution morphometrics. *Trends Ecol. Evol.* **8**, 129–132. (doi:10.1016/0169-5347(93)90024-j)
55. Terral J *et al*. 2004 Historical biogeography of olive domestication (*Olea europaea* L.) as revealed by geometrical morphometry applied to biological and archaeological material. *J. Biogeogr.* **31**, 63–77. (doi:10.1046/j.0305-0270.2003.01019.x)
56. Terral JF *et al*. 2010 Evolution and history of grapevine (*Vitis vinifera*) under domestication: new morphometric perspectives to understand seed domestication syndrome and reveal origins of ancient European cultivars. *Ann. Bot.* **105**, 443–455. (doi:10.1093/aob/mcp298)
57. Bonhomme V, Forster E, Wallace M, Stillman E, Charles M, Jones G. 2017 Identification of inter- and intra-species variation in cereal grains through geometric morphometric analysis, and its resilience under experimental charring. *J. Archaeol. Sci.* **86**, 60–67. (doi:10.1016/j.jas.2017.09.010)
58. Banfi E, Bartolucci F, Tison JM, Galasso G. 2022 A new genus for *Papaver* sect. *Meconella* and new combinations in *Roemeria* (Papaveraceae) in Europe and the Mediterranean area. *Nat. Hist. Sci* **9**, 67–72. (doi:10.4081/nhs.2022.556)
59. Martin M. 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10. (doi:10.14806/ej.17.1.200)
60. Li Q *et al*. 2020 Gene clustering and copy number variation in alkaloid metabolic pathways of opium poppy. *Nat. Commun.* **11**, 1190. (doi:10.1038/s41467-020-15040-2)
61. Li H. 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv.* (doi:10.48550/arXiv.1303.3997)
62. Garrison E, Marth G. 2012 Haplotype-based variant detection from short-read sequencing. *arXiv.* (doi:10.48550/arXiv.1207.3907)
63. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES. 2014 TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS One* **9**, e90346. (doi:10.1371/journal.pone.0090346)
64. Danecek P *et al*. 2011 The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158. (doi:10.1093/bioinformatics/btr330)
65. Knaus BJ, Grünwald NJ. 2017 vcfr: A package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* **17**, 44–53. (doi:10.1111/1755-0998.12549)
66. Jombart T, Kamvar ZN, Collins C, Lustrik R, Beugin MP, Knaus BJ. 2018 Package 'adegenet'.Github repository. https://github com/thibautjombart/adegenet
67. Wickham H, Chang W, Wickham MH. 2016 Package 'ggplot2.' Create elegant data visualisations using the grammar of graphics Version **2**, 1–189.
68. Allaire JJ. 2012 RStudio: integrated development environment for R. Boston, MA: RStudio Inc
69. R Core Team. 2013 R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. See http://www.R-project.org/.
70. Pritchard JK, Stephens M, Donnelly P. 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959. (doi:10.1093/genetics/155.2.945)
71. Evanno G, Regnaut S, Goudet J. 2005 Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* **14**, 2611–2620. (doi:10.1111/j.1365-294x.2005.02553.x)
72. Earl DA, vonHoldt BM. 2012 STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361. (doi:10.1007/s12686-011-9548-7)
73. Bivand R, Keitt T, Rowlingson B, Pebesma E, Sumner M, Hijmans R. 2015 *Package 'rgdal*. Bindings for the Geospatial Data Abstraction Library. See https://cran r-project org/web/packages/rgdal/index html (accessed 15 October 2017).
74. Gräler B, Pebesma E, Heuvelink G. 2016 Spatio-Temporal Interpolation using gstat. *R J.* **8**, 204. (doi:10.32614/rj-2016-014)
75. Nei M. 1987 *Molecular evolutionary genetics*. New York, NY: Columbia university press.
76. Tajima F. 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595. (doi:10.1093/genetics/123.3.585)
77. Jombart T. 2008 *adegenet* : a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405. (doi:10.1093/bioinformatics/btn129)
78. Stamatakis A. 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313. (doi:10.1093/bioinformatics/btu033)
79. Ortiz EM. 1052 vcf2phylip v2.0: Convert a VCF matrix into several matrix formats for phylogenetic analysis. *Zenodo.* (doi:10.5281/zenodo.2540861)
80. Darriba D, Taboada GL, Doallo R, Posada D. 2012 jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**, 772–772. (doi:10.1038/nmeth.2109)
81. Ronquist F *et al*. 2012 MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542. (doi:10.1093/sysbio/sys029)
82. Owens GL, Baute GJ, Rieseberg LH. 2016 Revisiting a classic case of introgression: hybridization and gene flow in Californian sunflowers. *Mol. Ecol.* **25**, 2630–2643. (doi:10.1111/mec.13569)
83. Malinsky M, Matschiner M, Svardal H. 2021 Dsuite—fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* **21**, 584–595. (doi:10.1111/1755-0998.13265)
84. Ellis N, Smith SJ, Pitcher CR. 2012 Gradient forests: calculating importance gradients on physical predictors. *Ecology* **93**, 156–168. (doi:10.1890/11-0252.1)

85. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005 Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978. (doi:10.1002/joc.1276)
86. Hijmans RJ *et al*. 2015 Package 'raster'. See https://CRAN.R-project.org/package=raster
87. Alvarado AH, Bossu CM, Harrigan RJ, Bay RA, Nelson ARP, Smith TB, Ruegg KC. 2022 Genotype–environment associations across spatial scales reveal the importance of putative adaptive genetic variation in divergence. *Evol. Appl.* **15**, 1390–1407. (doi:10.1111/eva.13444)
88. Hijmans RJ. 2019 Introduction to the 'geosphere' package (Version 1.5-10). https://cran.r-project.org/web/packages/geosphere/vignettes/geosphere.pdf.
89. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara RB. 2013 Package 'vegan. *Community Ecol. Package* **2**, 1–295. (doi:10.32614/CRAN.package.vegan)
90. Schneider CA, Rasband WS, Eliceiri KW. 2012 NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675. (doi:10.1038/nmeth.2089)
91. Bonhomme V, Picq S, Gaucherel C, Claude J. 2014 Momocs: outline analysis using R. *J. Stat. Softw* **56**, 1–24. (doi:10.18637/jss.v056.i13)
92. Evin A, Cucchi T, Cardini A, Strand Vidarsdottir U, Larson G, Dobney K. 2013 The long and winding road: identifying pig domestication through molar size and shape. *J. Archaeol. Sci.* **40**, 735–743. (doi:10.1016/j.jas.2012.08.005)
93. Peng H *et al*. 2024 A revised tribal classification of Papaveraceae (poppy family) based on molecular and morphological data. *Taxon* **73**, 762–783. (doi:10.1002/tax.13175)
94. Martin SH, Davey JW, Jiggins CD. 2015 Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244–257. (doi:10.1093/molbev/msu269)
95. Asada K, Salim V, Masada-Atsumi S, Edmunds E, Nagatoshi M, Terasaka K, Mizukami H, De Luca V. 2013 A 7-deoxyloganetic acid glucosyltransferase contributes a key step in secologanin biosynthesis in Madagascar periwinkle. *Plant Cell* **25**, 4123–4134. (doi:10.1105/tpc.113.115154)
96. Jin T, An J, Xu H, Chen J, Pan L, Zhao R, Wang N, Gai J, Li Y. 2022 A soybean sodium/hydrogen exchanger GmNHX6 confers plant alkaline salt tolerance by regulating Na+/K+ homeostasis. *Front. Plant Sci.* **13**, 938635. (doi:10.3389/fpls.2022.938635)
97. Sun TJ, Fan L, Yang J, Cao RZ, Yang CY, Zhang J, Wang DM. 2019 A glycine max sodium/hydrogen exchanger enhances salt tolerance through maintaining higher Na+ efflux rate and K+/Na+ ratio in *Arabidopsis*. *BMC Plant Biol.* **19**, 10. (doi:10.1186/s12870-019-2084-4)
98. Zhang WM, Cheng XZ, Fang D, Cao J. 2022 AT-HOOK MOTIF NUCLEAR LOCALIZED (AHL) (AHL) proteins of ancient origin radiate new functions. *Int. J. Biol. Macromol.* **214**, 290–300. (doi:10.1016/j.ijbiomac.2022.06.100)
99. Zhao J, Favero DS, Qiu J, Roalson EH, Neff MM. 2014 Insights into the evolution and diversification of the *AT-hook Motif Nuclear Localized* gene family in land plants. *BMC Plant Biol.* **14**, 266. (doi:10.1186/s12870-014-0266-7)
100. Wu Z, Ren H, Xiong W, Roje S, Liu Y, Su K, Fu C. 2018 Methylenetetrahydrofolate reductase modulates methyl metabolism and lignin monomer methylation in maize. *J. Exp. Bot.* **69**, 3963–3973. (doi:10.1093/jxb/ery208)
101. Condori-Apfata JA, Batista-Silva W, Medeiros DB, Vargas JR, Valente LML, Pérez-Díaz JL, Fernie AR, Araújo WL, Nunes-Nesi A. 2021 Downregulation of the E2 Subunit of 2-oxoglutarate dehydrogenase modulates plant growth by impacting carbon–nitrogen metabolism in *Arabidopsis thaliana*. *Plant Cell Physiol.* **62**, 798–814. (doi:10.1093/pcp/pcab036)
102. Condori-Apfata JA *et al*. 2019 The *Arabidopsis* E1 subunit of the 2-oxoglutarate dehydrogenase complex modulates plant growth and seed production. *Plant Mol. Biol.* **101**, 183–202. (doi:10.1007/s11103-019-00900-3)
103. Sharma V, Fedoseyenko D, Joshi S, Abdelwahed S, Begley TP. 2024 Phosphomethylpyrimidine synthase (ThiC): trapping of five intermediates provides mechanistic insights on a complex radical cascade reaction in thiamin biosynthesis. *ACS Cent. Sci.* **10**, 988–1000. (doi:10.1021/acscentsci.4c00125)
104. Fitzpatrick TB, Chapman LM. 2020 The importance of thiamine (vitamin B1) in plant health: from crop yield to biofortification. *J. Biol. Chem.* **295**, 12002–12013. (doi:10.1074/jbc.rev120.010918)
105. Yonekura-Sakakibara K, Tohge T, Matsuda F, Nakabayashi R, Takayama H, Niida R, Watanabe-Takahashi A, Inoue E, Saito K. 2008 Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene–metabolite correlations in *Arabidopsis*. *Plant Cell* **20**, 2160–2176. (doi:10.1105/tpc.108.058040)
106. Oka T, Nemoto T, Jigami Y. 2007 Functional analysis of *Arabidopsis thaliana* RHM2/MUM4, a multidomain protein involved in UDP-D-glucose to UDP-L-rhamnose conversion. *J. Biol. Chem.* **282**, 5389–5403. (doi:10.1074/jbc.m610196200)
107. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011 *Stacks*: building and genotyping loci *de novo* from short-read sequences. G3 (Bethesda) **1**, 171–182. (doi:10.1534/g3.111.000240)
108. Wickland DP, Battu G, Hudson KA, Diers BW, Hudson ME. 2017 A comparison of genotyping-by-sequencing analysis methods on low-coverage crop datasets shows advantages of a new workflow, GB-eaSy. *BMC Bioinform.* **18**, 586. (doi:10.1186/s12859-017-2000-6)
109. Crossman CA, Hamilton PK, Brown MW, Conger LA, George RC, Jackson KA, Radvan SN, Frasier TR. 2024 Effects of inbreeding on reproductive success in endangered North Atlantic right whales. *R. Soc. Open Sci.* **11**, 240490. (doi:10.1098/rsos.240490)
110. Voglmayr H, Montes-Borrego M, Landa BB. 2014 Disentangling *Peronospora* on *Papaver*: phylogenetics, taxonomy, nomenclature and host range of downy mildew of opium poppy (*Papaver somniferum*) and related species. *PLoS One* **9**, e96838. (doi:10.1371/journal.pone.0096838)
111. Liu YC, Liu YN, Yang FS, Wang XQ. 2014 Molecular phylogeny of Asian *Meconopsis* based on nuclear ribosomal and chloroplast DNA sequence data. *PLoS One* **9**, e104823. (doi:10.1371/journal.pone.0104823)
112. Tavakkoli Z, Assadi M. 2016 Evaluation of seed and leaf epidermis characters in the taxonomy of some annual species of the genus *Papaver* (Papaveraceae). *Nord. J. Bot.* **34**, 302–321. (doi:10.1111/njb.00833)
113. de Villemereuil P, Gaggiotti OE, Mouterde M, Till-Bottraud I. 2016 Common garden experiments in the genomic era: new perspectives and opportunities. *Heredity* **116**, 249–254. (doi:10.1038/hdy.2015.93)
114. Iob A, Botigué LR. Wild emmer contribution in wheat domestication and adaptation to new environments. *bioRxiv* 2023.10.25.563983. (doi:10.1101/2023.10.25.563983)
115. Zhou T *et al*. 2022 Nucleotide evolution, domestication selection, and genetic relationships of chloroplast genomes in the economically important crop genus *Gossypium*. *Front. Plant Sci.* **13**, 873788. (doi:10.3389/fpls.2022.873788)
116. Tittarelli R, Gismondi A, Di Marco G, Mineo F, Vernich F, Russo C, Marsella LT, Canini A. 2022 Forensic application of genetic and toxicological analyses for the identification and characterization of the opium poppy (*Papaver somniferum* L.). *Biology* **11**, 672. (doi:10.3390/biology11050672)
117. Stranska I, Skalicky M, Novak J, Matyasova E, Hejnak V. 2013 Analysis of selected poppy (*Papaver somniferum* L.) cultivars: pharmaceutically important alkaloids. *Ind. Crop. Prod.* **41**, 120–126. (doi:10.1016/j.indcrop.2012.04.018)
118. Laughlin JC, Munro D. 1982 The effect of fungal colonization on the morphine production of poppy (*Papaver somniferum* L.) capsules. *J. Agric. Sci.* **98**, 679–687. (doi:10.1017/s0021859600054484)
119. Mahdavi-Damghani A, Kamkar B, Al-Ahmadi MJ, Testi L, Muñoz-Ledesma FJ, Villalobos FJ. 2010 Water stress effects on growth, development and yield of opium poppy (*Papaver somniferum* L.). *Agric. Water Manag.* **97**, 1582–1590. (doi:10.1016/j.agwat.2010.05.011)

120. Lachman J, Hejtmánková A, Miholová D, Kolihová D, Tluka P. 2006 Relations among alkaloids, cadmium and zinc contents in opium poppy (*Papaver somniferum* L.). *Plant Soil Environ.* **52**, 282–288. (doi:10.17221/3442-pse)
121. Shaghaghi A, Alirezalu A, Nazarianpour E, Sonboli A, Nejad-Ebrahimi S. 2019 Opioid alkaloids profiling and antioxidant capacity of *Papaver* species from Iran. *Ind. Crop. Prod.* **142**, 111870. (doi:10.1016/j.indcrop.2019.111870)
122. Panicker S, Wojnoa HL, Ziska LH. 2007 Quantitation of the major alkaloids in opium from *Papaver setigerum* DC. *Microgram* **5**, 13–19.
123. Ziska LH, Panicker S, Wojno HL. 2008 Recent and projected increases in atmospheric carbon dioxide and the potential impacts on growth and alkaloid production in wild poppy (*Papaver setigerum* DC.). *Clim. Chang.* **91**, 395–403. (doi:10.1007/s10584-008-9418-9)
124. Choe S, Kim S, Lee C, Yang W, Park Y, Choi H, Chung H, Lee D, Hwang BY. 2011 Species identification of *Papaver* by metabolite profiling. *Forensic Sci. Int.* **211**, 51–60. (doi:10.1016/j.forsciint.2011.04.015)
125. Bernáth J, Dános B, Veres T, Szantó J, Tétényi P. 1988 Variation in alkaloid production in poppy ecotypes: responses to different environments. *Biochem. Syst. Ecol.* **16**, 171–178. (doi:10.1016/0305-1978(88)90092-0)
126. Garnock-Jones PJ, Scholes P. 1990 Alkaloid content of *Papaver somniferum* subsp. *setigerum* from New Zealand. *NZ J. Bot.* **28**, 367–369. (doi:10.1080/0028825x.1990.10412320)
127. Valva VL, Sabato S, Gigliano GS. 1985 Morphology and alkaloid chemistry of *Papaver setigerum* DC. (Papaveraceae). *Taxon* **34**, 191–196. (doi:10.2307/1221777)
128. Ren G *et al*. 2021 Large-scale whole-genome resequencing unravels the domestication history of *Cannabis sativa*. *Sci. Adv.* **7**, g2286. (doi:10.1126/sciadv.abg2286)
129. Kim CS, Song GP, Moon MO, Lee EJ, Kim CH. 2006 Unrecorded naturalized species: *Papaver setigerum* DC. (Papaveraceae) and *Veronica serpyllifolia* L. (Scrophulariaceae). *Korean J. Plant Taxon.* **36**, 145–151. (doi:10.11110/kjpt.2006.36.2.145)
130. Choe S, Lee E, Jin G, Lee YH, Kim SY, Choi H, Chung H, Hwang BY, Kim S. 2012 Genetic and chemical components analysis of *Papaver setigerum* naturalized in Korea. *Forensic Sci. Int.* **222**, 387–393. (doi:10.1016/j.forsciint.2012.08.002)
131. Campbell LG, Blanchette CM, Small E. 2019 Risk analysis of gene flow from cultivated, addictive, social-drug plants to wild relatives. *Bot. Rev.* **85**, 149–184. (doi:10.1007/s12229-019-09206-x)
132. Kumar B, Patra NK. 2010 Gene frequency-based estimation of natural outcrossing in opium poppy (*Papaver somniferum* L.). *Mol. Breed.* **26**, 619–626. (doi:10.1007/s11032-010-9395-3)
133. Denham T *et al*. 2020 The domestication syndrome in vegetatively propagated field crops. *Ann. Bot.* **125**, 581–597. (doi:10.1093/aob/mcz212)
134. Gleissberg S. 1998 Comparative analysis of leaf shape development in Papaveraceae–Papaveroideae. *Flora* **193**, 269–301. (doi:10.1016/s0367-2530(17)30849-6)
135. Gutierrez A, Grillo MA. 2022 Effects of domestication on plant–microbiome interactions. *Plant Cell Physiol.* **63**, 1654–1666. (doi:10.1093/pcp/pcac108)
136. Shoji T, Umemoto N, Saito K. 2022 Genetic divergence in transcriptional regulators of defense metabolism: insight into plant domestication and improvement. *Plant Mol. Biol.* **109**, 4. (doi:10.1007/s11103-021-01159-3)
137. Rapp RA, Haigler CH, Flagel L, Hovav RH, Udall JA, Wendel JF. 2010 Gene expression in developing fibres of Upland cotton (*Gossypium hirsutum* L.) was massively altered by domestication. *BMC Biol.* **8**, 139. (doi:10.1186/1741-7007-8-139)
138. Swanson-Wagner R, Briskine R, Schaefer R, Hufford MB, Ross-Ibarra J, Myers CL, Tiffin P, Springer NM. 2012 Reshaping of the maize transcriptome by domestication. *Proc. Natl Acad. Sci. USA* **109**, 11878–11883. (doi:10.1073/pnas.1201961109)
139. Sauvage C *et al*. 2017 Domestication rewired gene expression and nucleotide diversity patterns in tomato. *Plant J.* **91**, 631–645. (doi:10.1111/tpj.13592)
140. Zhou Y, Lu G, Sun G, Sun D, Ren X. 2022 Transcriptome and metabolite insights into domestication process of cultivated barley in China. *Plants* **11**, 209. (doi:10.3390/plants11020209)
141. Menamo T, Kassahun B, Borrell AK, Jordan DR, Tao Y, Hunt C, Mace E. 2021 Genetic diversity of Ethiopian sorghum reveals signatures of climatic adaptation. *Theor. Appl. Genet.* **134**, 731–742. (doi:10.1007/s00122-020-03727-5)
142. Zhang Q, Chen X, Guo H, Trindade LM, Salentijn EMJ, Guo R, Guo M, Xu Y, Yang M. 2018 Latitudinal adaptation and genetic insights into the origins of *Cannabis sativa* L. *Front. Plant Sci.* **9**, 1876. (doi:10.3389/fpls.2018.01876)
143. Wu ZY *et al*. 2024 Genomic variation, environmental adaptation, and feralization in ramie, an ancient fiber crop. *Plant Commun.* **5**, 100942. (doi:10.1016/j.xplc.2024.100942)
144. Khazaei H, Caron CT, Fedoruk M, Diapari M, Vandenberg A, Coyne CJ, McGee R, Bett KE. 2016 Genetic diversity of cultivated lentil (*Lens culinaris* Medik.) and its relation to the world's agro-ecological zones. *Front. Plant Sci.* **7**, 1093. (doi:10.3389/fpls.2016.01093)
145. Muhammad A, Akhtar A, Aslam S, Khan RS, Ahmed Z, Khalid N. 2021 Review on physicochemical, medicinal and nutraceutical properties of poppy seeds: a potential functional food ingredient. *Funct. Foods Health Dis.* **11**, 522–547. (doi:10.31989/ffhd.v11i10.836)
146. Wong AK, Somogyi AA, Rubio J, Philip J. 2022 The role of pharmacogenomics in opioid prescribing. *Curr. Treat. Options Oncol.* **23**, 1353–1369. (doi:10.1007/s11864-022-01010-x)
147. National Library of Medicine. 2024 DOPE Domestication of the Opium poppy in Prehistoric Europe. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1160476
148. Machado R *et al*. 2025 The origins and spread of the opium poppy (*Papaver somniferum* L.) revealed by genomics and seed morphometrics. *Zenodo*. (doi:10.5281/zenodo.15019934)
149. Machado R *et al*. 2025 Supplementary Material from: The Origins and Spread of the Opium Poppy (*Papaver Somniferum* L.) Revealed by Genomics and Seed Morphometrics. Figshare (doi:10.6084/m9.figshare.c.7742946)