Drivers of genetic differentiation and recent evolutionary history of an Eurasian wild pea

Timo Hellwig^{1,2}, Shahal Abbo¹, Ron Ophir²

¹The Levi Eshkol School of Agriculture, Hebrew University of Jerusalem, Rehovot 7610001, IL

²Volcani Center, Agricultural Research Station, Rishon LeZion, IL

Abstract

Genetic diversity a major determinant for the capacity of species to persist and adapt to their environments. Unraveling the factors affecting genetic differentiation is crucial to understand how genetic diversity is shaped and species may react to changing environments. We employed genotyping by sequencing to test the influence of climate, space, latitude, altitude and land cover on genetic differentiation in a collection of 81 wild pea samples (*Pisum sativum* ssp. *elatius*) from across its distribution range from western Europe to central Asia. We also attempted to elucidate the species recent evolutionary history and its effect on the current distribution of genetic diversity. Association of single SNPs with climate variables were analyses to test for signatures of local adaptation. Genetic variation was geographically structured into six distinct genetic cluster. Two of which were associated with a taxonomic group (*Pisum sativum* ssp. *humile*) that according to some researchers does not qualify for a sub-species rank due to its alleged lack of genetic distinctness from other conspecific groups. The effect of the tested factors influencing genetic differentiation were rather variable among genetic clusters. The climate predictors were most important in all clusters. Land use was more important in clusters from areas strongly influenced by human land use, especially by agriculture. We found a statistically significant association of 3,623 SNPs (2.4 % of all SNPs) with one of the environmental predictors. Most of them were correlated with latitude followed by temperature, precipitation and altitude. Estimation of SNP effects of the candidates resulted in a missense to silent ratio of 0.45, suggesting many of the observed candidates SNPs may alter the encoded amino acid sequence. Wild peas went through a genetic bottleneck during the last glacial period followed by population recovery. Probably associated with this population recovery, we detected a range expansion, which may have led to an eastward range expansion of the European cluster to Turkey and thereof southwards and eastwards. Overall, the interplay of several environmental factors and the recent evolutionary history affected the distribution of genetic diversity in wild peas where each subpopulations were differently affected by those factors and processes.

Keywords: genetic diversity, genetic differentiation, crop wild relative, range expansion, demographic history

Introduction

Genetic diversity is the raw material of evolution and, therefore, a major determinant of the ability of species to persist and adapt in changing environments (Crozier, 1997; Barret & Schluter, 2008; Prentis et al., 2008; Matuszewski et al., 2015; Lai et al., 2019). It plays a fundamental role in the evolutionary trajectory of species and thus affects all levels of biodiversity (Crutsinger et al., 2006, 2007; Hughes et al., 2008; Stange et al., 2021). This has direct implications for species utilization in agriculture (Kannenberg & Falk, 1995; Hoisington et al., 1999), pharmacology (Elisabetsky & Costa-Campos, 1996; Barata et al., 2016) or bioenergy (Jaradat, 2010; Kiru & Nasenko, 2010) as well as on ecosystems' stability and scope for recovery (Reusch et al., 2005; Reynolds et al., 2012). Therefore, understanding how species genetic diversity is distributed is crucial. A major determinant of genetic differentiation is the balance between the degree of geneflow between populations and random genetic drift. Geneflow can be restricted due to spatial distribution, if populations are separated by large geographic distances (Slatkin, 1993). A process known as isolation by distance (IBD) that leads to spatial autocorrelation of genetic variation (Wright, 1943). IBD was extensively studied and is widespread in nature (Meirmans, 2012; Sexton et al., 2014). Genetic variation may also be shaped by (expressed as correlation with) environmental variables, which may lead to a pattern of isolation by environment (IBE; Shafer & Wolf, 2013; Wang & Bradburd, 2014). Similarly to the case of IBD, geneflow between populations that inhabit contrasting environments is limited. However, the causes of restriction to geneflow between populations in contrasting environments are more difficult to determine. Several processes like reduced hybrid fitness, genotype-dependent dispersal and natural or sexual selection against immigrants may cause IBE (Servedio, 2004; Nosil et al., 2005; Edelaar et al., 2008; McBride & Singer, 2010; Bolnick & Otto, 2013). IBE has been documented in a variety of different species including barley (*Hordeum vulgare*), soft snow-grass (*Poa hiemata*) or Japanese larch (*Larix kaempferi*) among others (Tanto Hadado et al., 2010; Byars et al., 2009; Nishimura & Setoguchi, 2011; for review see Shafer & Wolf, 2013 and Sexton et al., 2014). A third process, termed isolation by resistance (IBR;

McRae, 2006) takes a third dimension of restriction to dispersal and gene flow into account, the heterogeneity of landscapes. The IBR model describes genetic variation in relation to resistance, where the resistance depends on the habitability of landscape features, which connects populations (McRae & Beier, 2007). Cities, cropland or large water bodies present higher resistances to many species compared with forests or grassland and, by that pose larger restrictions on the connectivity of populations. IBR takes this effect into account in addition to the geographic distance and can therefore be considered an extension of the simple IBD model. IBR is particularly important for plants because they cannot easily move through inhabitable areas like cities, cropland or large water bodies and was found to be a better predictor of plants dispersal capabilities than IBD (Cruzan & Hendrikson, 2020, Grasty et al., 2020). Besides those processes that directly affect gene flow patterns among populations, the demographic history of a species can indirectly influence the genetic differentiation between populations. Increased random genetic drift due to a genetic bottleneck may promote the divergence of populations (Wright, 1931). This effect is amplified in peripheral populations, but can also occur in central populations, if the reduction of population size leads to fragmentation, which causes altered gene flow patterns (Lowe et al., 2005; Bouzat et al., 2008; Stevens et al., 2018). An increase in population size can also lead to increased genetic differentiation if accompanied by range expansion. Population expansions may occur in a continuous-gradual manner or via piecemeal punctuations as in a simple stepping-stone model (Excoffier et al., 2009). During a range expansion under the stepping-stone model, the low effective population size of a founding population may lead to increased genetic drift in the population which can drive rapid differentiation between the derived and the source population (Baker & Moeed, 1987). In a series of colonizations, this effect is likely to persist and thereby create an IBD pattern (Ramachandran et al., 2005; Orsini et al., 2013). Naturally, additional parameters will complicate the effects an expansion may have on the involved populations, e.g. the expansion rate or the degree of ongoing gene flow with respective parent populations (Excoffier et al., 2009). Another cause of genetic differentiation following a founder event is described by the monopolization hypothesis (De Meester et al., 2002). After the

colonization of a new habitat, rapid adaptation may occur (van Boheemen et al., 2018; Szűcs et al., 2017; Yin et al., 2021) and can provide an adaptive advantage to early founders over late comers leading to evolution-mediated priority effects (De Meester et al., 2016). The resulting adaptive divergence can cause reduced geneflow with the parent population due to the same processes like in the case of IBE (above). This monopolization scenario is equivalent to niche pre-emption (Fukami, 2015), but its priority effects are the result genetic adaptation (De Meester et al., 2016).

The Pea (*Pisum* L.) genus is commonly divided into two species, *Pisum fulvum* and *Pisum sativum*. *Pisum fulvum* is a rather homogenous entity and a typical east Mediterranean element restricted to Syria, Lebanon, Israel and Jordan where it forms small scattered patches in a variety of habitats (Ben-Ze'ev & Zohary, 1973; Ladizinsky & Abbo, 2015). *Pisum sativum* is divided into at least two subspecies, the domesticated *P. s. ssp. sativum* and an aggregate of wild forms collectively named *P. s. ssp. elatius* (e.g., Palmer et al., 1985; Hoy et al., 1996; Vershinin et al., 2003; Jing et al., 2007, 2010; Kosterin et al., 2010; Zaytseva et al. 2012; Bogdanova et al., 2018). Older classifications recognized another species, *P. humile* (Boissier, 1867) which was later included within *P. sativum* as a wild subspecies and divided into two varieties, *P. s. ssp. humile* var. *syriacum* (often termed northern humile) and *P. s. ssp. humile* var. *humile* (known as southern humile, *sensu* Ben-Ze'ev & Zohary, 1973; and see Ladizinsky & Abbo, 2015). Others, however, consider *P. s. ssp. humile* a synonym of *P. s. ssp. elatius* (Maxted & Ambrose, 2001; Vershinin et al., 2003; Ellis, 2011). Wild *P. sativum* forms (*P. s. ssp. elatius - sensu lato*) occur across a large geographic range from the Atlantic districts of southern Europe all around the Mediterranean basin, the Black Sea and the Caspian Sea, at least as far east as Turkmenistan (Karakala region) and the Iranian Zagros Mountains (Ladizinsky & Abbo, 2015). Isolated scattered populations also occur as far north as Hungary and Bulgaria (Smykal et al. 2017). Like many Mediterranean elements, *P. s. ssp. elatius* exhibits a winter-annual life-cycle, where seeds germinate in autumn after the first precipitation events and overwinter as short statured plants until temperatures rise in the spring. Wild peas are similar to other wild legumes in their strong seed dormancy causing

them to build large soil seed banks (Russi et al. 2008; Smýkal et al., 2014; Baskin & Baskin, 2014; Ladizinsky & Abbo, 2015). Genetic diversity in the *P. sativum* complex has been extensively investigated, yet the majority of studies were based on germplasm accessions with a focus on domesticated *P. s. ssp. sativum* or rather small sample sizes (Palmer et al., 1985; Hoy et al., 1996; Vershinin et al., 2003; Jing et al., 2007, 2010; Kosterin et al., 2010; Zaytseva et al. 2012; Bogdanova et al., 2018). Exceptions are the studies of Smýkal et al., (2011) and Trněný et al., (2018) which employed a large number of wild samples covering the distribution range of *P. s. ssp. elatius*. Yet, neither of them put much emphasis on the factors that may influence genetic variation in wild pea. Smýkal et al., (2018) did investigate the geography and environmental effects on genetic differentiation of wild pea, but only across south-east Turkey.

We attempted to fill the gap left by previous studies and investigated the drivers of genetic differentiation in *P. s. ssp. elatius* from a sample set covering its entire distribution range. More specifically, we addressed the following questions: (1) How is genetic diversity in *P. s. ssp. elatius* distributed across its native range? (2) What may be the main drivers of genetic differentiation in this taxon? (3) What impact may have the demographic history of this species on its genetic structure? and (4) Can we detect regions of the genome associated with climatic adaptation? Due to its biological characteristics common to many Mediterranean wild legumes and large distribution range covering a great variety of habitats, *P. s. ssp. elatius* is a suitable model to study factors driving genetic differentiation that influence the distribution of genetic diversity. Therefore, we consider the present results relevant not only to this species but rather promoting a more general understanding of the factors under which genetic diversity of wild plants evolves.

Methods

Plant Material and Genotyping

Our collection contained 81 *P. sativum* ssp. *elatius* individuals which originated from

throughout its distribution range (Fig. 1). Some of these accessions were sampled by ourselves, some of them were obtained from genebanks or other sources (Tab. S1). Location details of some genebank accessions were imprecise and appeared in the sea. These were corrected by moving the sample location in a straight line to the adjacent shore (less than ~2 km in 11 cases). Following the classification of Ladizinsky & Abbo (2015), we further subdivided our samples into *P. s. ssp. elatius* (48 samples) and *P. s. ssp. humile* including its two varieties, *P. s. ssp. humile* var. *syriacum* (northern humile; 21 samples) and *P. s. ssp. humile* var. *humile* (southern humile; 12 samples) based on morphological characteristics and the respective ecology of the sampling sites. DNA was extracted from the accessions and restriction site associated DNA (RAD) sequencing was conducted by Elshire Group Ltd. (<https://www.elshiregroup.co.nz/>). Restriction enzyme ApeKI was used for complexity reduction and the protocol of (Elshire et al., 2011) was followed for genotyping by sequencing (GBS). However, 150 bp paired end reads with combinatorial bar codes were produced by the X Ten sequencing platform instead of single end reads with single barcodes by HiSeq 2500 as described in the protocol. Axem demux (Murray & Borevitz, 2017) was used for demultiplexing. Trimming of adapter and reverse-barcode was done with GBS-PreProcess (<https://github.com/Lanilen/GBS-PreProcess>; accessed: Feb. 2020). After mapping the processed reads against the *Pisum sativum* ssp. *sativum* reference genome (Kreplak et al., 2019). Stacks (ver. 2.5; Catchen et al., 2011; Catchen et al., 2013) was used to call variants.

Only biallelic SNPs were retained. As a minor allele frequency (MAF) threshold we used 0.01. SNPs with less than 40 % genotype calls were removed. We also removed SNPs with minimum depth > 3, minimum mean depth across samples > 5 and minimum genotype quality > 25. The filtering was done with VCFTools (ver. 0.1.17; Danecek et al., 2011). The resulting SNP set was pruned for linkage disequilibrium (LD) with PLINK (v1.90b5.4; Purcell et al., 2007). We shifted 50 kb windows by 5 variants and SNPs were removed when they had $r^2 \geq 0.5$ within one window.

Population Structure

We used the LD pruned variant set to estimate sample's subpopulation fractions with sparse non-negative matrix factorization (sNMF; Frichot & François, 2015) in the R statistics environment (R Core Team, 2020) using the package *LEA* (Frichot et al., 2014). We performed sNMF 20 times with numbers of subpopulations (K) ranging from 1 through 20. Since the regularization parameters (α) can lead to biased sNMF results, especially in small data sets, we tested different values to determine an optimum α , which was used in the final analysis. The parameters applied were: 5000 iterations, tolerance = 0.00001, $\alpha = 1$. For the computation of the cross-entropy criterion, we used 10 % masked genotypes.

We calculated pairwise GENPOFAD distances (Joly et al., 2015) from the LD pruned SNP set to be used in principal coordinate analysis (PCoA). The analysis was done in R with the package *poppr* (Kamvar et al., 2015; R Core Team, 2020).

SplitsTree4 (Huson & Bryant, 2005) was used to create a neighbor-net phylogenetic network based on the entire SNP set. We used the Hasegawa-Kishino-Yano model of nucleotide substitution (Hasegawa et al., 1985). The employed parameters were: transition/transversion ratio = 1.5, empiric base frequencies and a proportion of invariable sites = 0. Since we used a SNP set for the analysis we expected to violate the assumption of equal substitution rates among sites. Therefore, the shape factor of the gamma distribution was set to 0.5 to take wide range of unequal rates among sites into consideration.

We defined a genetic cluster based on subpopulation fractions estimated by sNMF with K=6. An individual was assigned to a genetic cluster when at least 50 % of its subpopulation fraction came from the respective subpopulation (Tab. S1). Two accessions, PeAb15 from Greece and PeAb37 from Turkey, were strongly admixed. They were not assigned to any genetic cluster and excluded from the all subpopulation specific analyses. We calculated pairwise F_{st} between genetic clusters according to (Weir & Cockerham, 1984). Test of significance was done with 9,999 bootstrap values (H_0 : $F_{st} =$

0). The R package *stamp* (Pembleton et al., 2013; R Core Team, 2020) was used for the computations.

Landscape genetic analysis

We employed gradient forests analysis, an extension of random forests, to explore non-linear association between climate, land cover, environmental, spatial and allelic variation (Ellis, et al., 2012). The entire analysis was done with the R packages *SoDA* (Chambers, 2013), *vegan* (Jari, 2018), *adespatial* (Dray et al., 2020) and *gradientForest* (Ellis, et al., 2012). The described procedure was done for the entire populations and for each genetic cluster separately.

As environmental variables we used Latitude, Altitude and bioclimatic variables from WorldClim2.1 (Fick & Hijmans, 2017) at 30-seconds spatial resolution. Several of those variables were strongly correlated. There is controversy regarding the question if correlated predictors should be excluded in random forests analyses and if so, where to determine the threshold of correlation (e.g. Cuttler et al., 2012; Murray & Conner, 2009; Knudby et al., 2010). We opted to use an intermediate approach and removed predictors with $r^2 > 0.9$ before the analysis. The environmental variables employed in the analysis after removing strongly correlated variables were Latitude, Altitude, BIO11 (mean temperature of coldest quarter), BIO18 (precipitation of warmest quarter), BIO19 (precipitation of coldest quarter), BIO15 (precipitation seasonality), BIO17 (precipitation of driest quarter), BIO10 (mean temperature of warmest quarter), BIO14 (precipitation of driest month) and BIO13 (precipitation of wettest month). In the gradient forest analysis, we used conditional permutation to calculate variables' importance for predictors with $r^2 > 0.7$ to further account for non-independence among predictors (Strobl et al., 2008).

A detailed land cover classification map was obtained from the Copernicus Global Land Service (<https://www.copernicus.eu/en>; Buchhorn et al., 2019). Due to the high computational demand of the analysis we scaled the resolution down to 1 km. A single resistance value was assigned to each of the 22 land cover classifications (e.g. shrubland: 10, open evergreen needle-leaved forest: 20, cropland: 80, build-up area: 90, sea; 200;

Tab. S3). Circuitscape (Circuitscape.jl ver. 5.7.1; Anantharaman et al., 2020) was used to calculate pairwise resistance distances between all samples from the land cover map. In order to use the resistance values in the gradient forest analysis, we transformed the resistance matrix to vectors by performing classical multidimensional scaling (cMDS). We accounted for overfitting by calculating the goodness of fit with varying numbers of retained cMDS coordinates to determine the minimum number of dimensions that retained a sufficient amount of the variation of the original resistance matrix. The cMDS coordinates that still explained 99 % of the variation in the resistance matrix was used as predictors representing land cover in the gradient forests analysis.

Moran Eigenvector Maps (MEMs; Dray et al., 2006) were constructed to model the spatial pattern. For the construction of MEMs, we tested two different algorithms to create the connectivity matrix: Gabriel's graph and relative neighborhood graph. Each algorithm was tested with three different weights to calculate the spatial weighting matrix (SWM): no weights, linear ($1 - D/d_{\max}$, where D was the distance of the node to be weighted and $d_{\max}$ the maximum distance between two nodes in the connectivity matrix) and concave down ($1 - (D/d_{\max})^y$, where y ranged from 2 through 5). The optimum combination of graph and SWM was selected based on the procedure presented by Bauman et al., (2018). The global test of significance was performed with 9,999 permutations. We created a subset of MEMs for each graph - SWM combination using forward selection with double stopping criterion (999 permutations; Blanchet et al., 2008) to avoid overfitting. The graph - SWM combination that showed the highest adjusted R^2 with the selected MEMs was used for the analysis. Its MEMs represented the spatial pattern in the data.

Gradient forest analysis was implemented in the R package *gradientForest* (Ellis, et al., 2012) where the selected environmental variables, cMDS coordinates and MEMs were used to test their association with the SNP allele frequencies.

Environmental genome-wide association study

We used redundancy analysis (RDA; Forester et al., 2018) and a latent factor mixed

model (LFMM; Caye et al., 2019) to identify SNP associated with environmental predictors. We grouped all temperature and precipitation related variables into the groups temperature and precipitation. Principle component analysis (PCA) was performed on these two groups and the first PCs were used as input for the analysis. We used Beagle (v5.0; Browning and Browning, 2007; Browning and Browning, 2009; Browning et al., 2018) with 30 iterations, 10 burning steps and effective population size of 50,000 to impute missing genotypes and removed SNPs on scaffolds which left 150,456 SNPs for the analysis. Since we found six well-distinguished populations, we used six latent factors for LFMM analysis. Other input parameters were 100,000 iteration, 10,000 burn-in rounds and ten repetitions per environmental variable. In the RDA, SNPs were considered as candidates if their loadings in the ordination space deviated more than three standard deviations from the mean (approximately equivalent to a two-tailed p-value of 0.0027). Note, that this threshold is rather low, but after the analyses, we only considered SNPs that appeared to be outliers in both analyses as candidates. This approach has been shown to produce rather conservative results and to be a good procedure for a first exploratory analyses of loci under selection (Forester et al., 2018). The analysis were done using ‘LEA’ and ‘vegan’ R packages (Frichot and François, 2015; Jari, 2018). We used the program SnpEff (v. 5.0) to predict effects of candidate SNPs (Cingolani, 2012).

Statistics of diversity and neutrality

Several diversity statistics were calculated for each of the genetic cluster. Nei’s gene diversity (H_{exp} ; Nei, 1978), observed heterozygosity (H_{obs}) and nucleotide diversity (π) were calculated with *adegenet* (Jombart & Ahmed, 2011) and *VCFTools* (ver. 0.1.17; Danecek et al., 2011). Haplotypes were generated by HaploView (Barrett et al., 2005) using the four gamete rule (Wang et al., 2002). Markers which heavily violated HWE were removed from the analysis to ensure that SNPs under selection are not included in the analysis. We applied Bonferroni correction to account for multiple testing and removed markers with a p-value below 0.05/(number of markers). We calculated number of haplotypes (H_{tn}) and haplotype diversity (H_{td}) from the results of HaploView using R core functions (R Core Team, 2020).

Tests of neutrality were conducted by calculating Tajima's D (Tajima, 1989) and R_2 (Ramos-Onsins & Rozas, 2002) for each genetic cluster using the R packages *r2vcftools* and *pegas* (Pope, 2020; Paradis, 2010). Significance of both statistics were assessed with simulations of 100,000 neutral sequences.

We estimated selfing rates of the entire population using the method of David et al. (2007). P-values of the identity disequilibrium g_2 were calculated with 9,999 permutations.

Inference of demographic history

Inference of demographic history was done using Stairway Plot 2 (Liu & Fu, 2020). The analysis was run with the LD pruned variant set and 1,000 replicates to estimate confidence intervals. Although wild peas are annual plants, their generation time is longer than one year due to their strong seed coat imposed physical dormancy. Based on observations in our experimental farm where volunteers emerged up to six years after growing wild peas, we opted to use an average generation time of three years. Mutation rates in peas are unknown. In line with the generation time hypothesis, which states that species tend to have accelerated rates of molecular evolution if they have shorter generation times, we used a mutation rate of $2e-9$, since this is at the lower end of the estimated mutation rates of *Arabidopsis thaliana* (Exposito-Alonso et al., 2018; Lynch, 2010). This reflects the higher generation time that we used compared to the 1-year generation time of *A. thaliana*. We used the unfolded sites frequency spectrum (SFS) as input for Stairway Plot 2 and ignored singletons as suggested by the authors (Liu and Fu, 2015).

Detection of range expansion

We tested for a past range expansion using the directionality index (ψ) (Peter and Slatkin, 2013; Peter and Slatkin, 2015). Since this approach assumes neutrality, we excluded SNPs that violate HWE using the same approach as for the calculation of neutrality statistics (see above). We additionally removed sites with more than 10 % missing

genotypes to ensure that ψ , which detects clines in allele frequencies, can be calculated reliably. The null distribution of the ψ was generated by permuting the allele frequencies of each SNP and assigning them randomly to a population. P-values were calculated based on this null distribution ($H_0: \psi = 0$). Since permutations tend to underestimate the variance in the data if SNPs are in linkage disequilibrium (Peter & Slatkin, 2013), we used the LD pruned variant set for the analysis.

Results

Genotyping by sequencing

We detected 1,973,460 variants in total while 167,766 SNPs were retained after quality filtering and 106,770 after LD pruning. In the entire SNP set as well as in the LD pruned set, SNPs were distributed throughout the entire genome without notable large gaps (Fig. S1). In both SNP sets the majority of sites had low missing genotype calls with a median and mean of 11 % and 14 %, respectively (Tab. S2). Rare alleles were abundant in both SNP sets. The third quantile of MAF was 0.13 and 0.11 in the filtered and the LD pruned set, respectively (Tab. S2). Observed heterozygosity was extremely low with only 2 % at the third quantile and a mean of 0.01 (Tab. S2). In the entire variant set the mean and median of the nucleotide diversity (π) were 0.09 and 0.16, respectively and ranged from 0.04 in the first quantile to 0.233 in the third quantile. After LD pruning similar values were observed (Tab. S2).

Population Structure

Analysis of population structure exposed six well distinct clusters (Cl1 - Cl6; Fig. 2A) with F_{st} values ranging from 0.162 (Cl1 - Cl4) to 0.438 between Cl3 and Cl6 (Tab. 1). The majority of individuals only showed small degrees of admixture in the sNMF analysis and only two accessions (PeAb15 and PeAb37) did not have a subpopulation fraction of > 0.5 of any genetic cluster and were not assigned to any cluster (Fig. 2A). Individuals from Cl1 were sampled from the northern and eastern part of the Fertile Crescent. Cluster 4 and Cl6 showed a wide distribution range from Turkey to the

Caucasus and around the Black Sea while Cl2 and Cl5 were located in geographically restricted areas in southern and northern Israel, respectively. The sample locations of Cl3 were spread all over southern Europe (Fig. 1). Cluster 1 and Cl4 split from a common node in the phylogenetic network and their individuals cluster closely together in PCoA (Fig. 1 B,C). They also cluster together until $K = 4$ in the sNMF analysis (Fig. 2A). Cl2 and Cl5 also shared the same origin in the phylogenetic network. Interestingly, Cl3 emerged from the same node, although it was located at the opposite end of the distribution range of the germplasm collection relative to Cl2 and Cl5. PCoA results were in line with these observations (Fig. 1 B, C). Cluster 6 was placed between the two main branches. Despite the apparent geographic pattern, some outliers were observed. One sample from Cl5 was sampled from eastern and one from western Turkey. The sample location of two individuals assigned to Cl1 were located in Jordan and on the Golan Heights close to the Israeli - Syrian border.

Landscape genetic analysis

In the gradient forest analysis 21.00 % of the genetic variation could be explained by the employed variables, where the spatial predictors were most influential (10.60 %) followed by the climate (7.28 %; Fig. 3). Within the climate predictors precipitation related variables had higher importance relative to temperature related variables (Fig. S2). Altitude, land cover and latitude accounted for minute fractions of the genetic variation with 0.47 %, 1.20 % and 1.45 %, respectively. In the individual genetic clusters the effect of the different predictors were overall similar with few exceptions. Notably, small association was estimated for Cl6 with only 1.06 % in total. The very small sample size of this cluster has to be kept in mind and its estimate should be considered with caution. Climate showed the strongest association with genetic variation in all clusters, altitude the weakest. Within the climate predictors, precipitation related variables had an overall stronger effect than temperature related variables (Fig. S2). The effect of land cover was particularly strong in Cl1 (9.55 %) and Cl2 (6.17 %), where it was the second strongest effect after climate and followed by the geography. The spatial predictors showed strong association in Cl1, Cl3 and Cl4, which were also the clusters with the widest distribution

range.

Environmental genome-wide association study

We detected 3,623 SNPs (2.4 % of all SNPs) that were significantly correlated with one of the environmental predictors. Most SNPs were associated with latitude (1,140) followed by temperature (963), precipitation (929) and altitude (591). Candidate SNPs were distributed rather evenly across the genome with gaps in centromeric regions. Notably, only very few of the SNPs with the highest correlation were located on chromosomes six and seven (Fig. S3). The ratio of missense to silent SNPs was 0.45. Intron variants were the most abundant predicted effect (29.9 %) followed by synonymous SNPs (27.4 %). Thirty seven SNPs possibly caused a premature start codon and one SNP created a stop codon within a gene. All candidate SNPs, their predicted effects and the corresponding genes are depicted in Tab. S4.

Statistics of diversity and neutrality

Observed heterozygosity was very small and barely varied across genetic clusters (Tab. 2). Expected heterozygosity was also small and only little variation was observed among clusters ranging from 0.07 in Cl3 to 0.1 in Cl1 (Tab. 2). Nucleotide diversity was highest in Cl5 and Cl6 with 0.38 and 0.46, respectively. These two clusters also had the highest H_t (both 0.92) and the fewest number of haplotypes with 3,811 and 2,893, respectively (Tab. 2). The lowest π (0.24) and H_t (0.74) was found in Cl4 which also had the lowest H_n (4045) after Cl4 and Cl5. Cluster 1, Cl2 and Cl3 showed intermediate values of π , H_t , H_n (Tab. 2). Cluster 4 had a low negative Tajima's D value (-0.12; $p < 1e-05$). Clusters 1 to 3 had very low positive Tajima's D values, yet all of them were significantly different from 0. Tajima's D was high for Cl5 and Cl6 with 0.58 ($p < 1e-05$) and 0.89 ($p < 1e-05$), respectively. Contrary to the Tajima's D values none of the R_2 values were significant even though they ranged from 0.12 in Cl1 to 0.2 in Cl6 (Tab. 2). We estimated a selfing rate of 0.700 for the total population

Inference of demographic history

Estimations of past effective population sizes (N_e) depicted an increase of N_e during the early phase of the last glacial period from around 150 k to 90 k year ago. Afterwards, a genetic bottleneck slowly started and became increasingly rapid until N_e dropped from ~180 k to 60k roughly 3 k years ago. The population quickly recovered after the bottleneck and depicted constant N_e on a similar level as before the bottleneck for the last 1.5 k year. Note that we refrained from depicting results deeper into the past beyond around 100 - 200 thousand years, because estimations tend to become less reliable for such periods. This is especially true for results of Stairway Plot 2 (Liu & Fu, 2015).

Detection of range expansion

Significant directionality indices (ψ) were detected for CI3 indicating an expansion towards CI6 and CI1 which were also the highest ψ values with 0.32 and 0.30, respectively. Both, CI4 and CI5 had significant ψ leading to CI1 and CI2, but ψ values originating from CI4 were higher than those of CI5 (Tab. S5). Overall, we detected expansion from the eastern part of the distribution area to the central part and from there southwards and further to the east (Fig. 5).

Discussion

Structure of genetic diversity

We observed six well distinct genetic clusters showing a hierarchical structure with a notable geographic pattern (Tab. 1, Fig. 1, 2). Overall our genetic structure is congruent with the structure reported by Smýkal et al., (2017). Large overlap could be observed between CI3, CI4, and CI1 with the subpopulations Q3, Q4, and Q5 reported in Smýkal et al. (2017), respectively. These are also largely equivalent to the clusters 9, 10 and 1 of Trněný et al. (2018) who used the same wild germplasm array of Smýkal et al., (2017) but additionally included the two domesticated taxa. Yet, if viewed in greater detail, there are some differences, e.g. Smýkal et al., (2017) and Trněný et al. (2018) could not detect distinct subpopulations in the southern Levant, most likely due to the smaller number of

samples from this area. Therefore, they could not have observe that CI2, CI3 and CI5, which were sampled from Israel (CI2, CI5) and southern Europe (CI3), showed close genetic proximity even though they are geographically remote. This pattern was also observed by Kosterin et al. (2010) who investigated the phylogeny of wild peas based on three dimorphic markers (nuclear, plastid and mitochondrial) and divided the genus into four lineages (A through D). They suggested that lineage A represents the ancestral state since its marker combination is present also in *P. fulvum*. Combination A is also dominant in *P. s. ssp. elatius* samples from Israel and the Mediterranean islands Sardinia and Menorca while lineage C, which is distinct from A by only a single mutation, could be found in continental southern Europe (France, Greece, Italy). These authors proposed a westward expansion during the Pleistocene climate cooling when oscillating sea levels temporally connected certain Mediterranean islands. Analysis of spatial diversity of niche patterns in wild peas suggested that the center of diversity is currently located in the Near East (Smýkal et al., 2017) which is also supported by our findings as evident from the diverse population structure and the higher diversity statistics of the Near Eastern clusters compared to the European CI3 (Tab. 2). The same analysis predicted a center of high diversity in northern Africa and the authors proposed that the route through north Africa might be a so far overlooked route during the westward expansion of wild pea. Both proposed scenarios (Kosterin et al. 2010; Smýkal et al., 2017) are consistent with our findings of close genetic proximity of the southern Levantine clusters (CI2, CI5) with the southern European CI3. Unfortunately, our sample collection did not contain any samples from north Africa and the number of available genebank accessions from north Africa is very limited and often with questionable passport data. Sample expeditions to this area may enable researchers to shed more light on the role of this area in past expansion of wild pea. Indeed, the germplasm collection of Smýkal et al., (2017) held a small number of genebank accessions from north Africa, yet the study was criticized for data confusion, including a misclassification of at least one domesticated sample as wild which suggests that one should view parts of this collection with some caution (Kosterin & Bogdanova, 2018).

For decades botanists followed the classification of Boissier (1867) who recognized three wild pea species, *P. fulvum*, *P. elatius* and *P. humile*. Ben-Ze'ev & Zohary (1973) showed by crossing experiments that although chromosomal interchanges distinguish between *P. s. ssp. elatius* and the so called northern *P. humile*, *P. s. ssp. elatius* and the southern *P. humile* share the same cytotype. Therefore, Ben-Ze'ev & Zohary (1973) considered *P. elatius* and *P. humile* as interfertile (and likewise with the domesticated *P. s. ssp. sativum*) and consequently, adopted the classification proposed by Davis (1970), who recognized only two *Pisum* species, *P. fulvum* and the *P. sativum* complex (with its several subspecies), which is mostly accepted today. More detailed classifications within the *P. sativum* complex were proposed in a number of studies, which led to a confusing number of proposed epithets and taxonomic entities that often were considered synonymous with each other (e.g., Maxted & Ambrose, 2001; Ellis, 2011). A unified taxonomic classification of the *P. sativum* complex is not yet agreed upon, but most researchers today follow the classification of Maxted & Ambrose (2001), who divided the *P. sativum* complex into a domesticated subspecies, *P. s. ssp. sativum*, and a single wild subspecies, *P. s. ssp. elatius*. Consequently, *Pisum s. ssp. humile* is nowadays widely considered a synonym of *P. s. ssp. elatius*, because it was argued that it is not genetically distinct from studied *P. s. ssp. elatius* accessions (Vershinin et al., 2003; Ellis, 2011). However, these studies included only a single *P. s. ssp. humile* accession. This may be due to lack of genuine wild samples or due to the fact that the different wild *P. sativum* subgroups (in our view, ecotypes) are often difficult to distinguish morphologically. This may be problematic when accessions are grown in greenhouses, outside of their natural ecological zone, by researchers who never observed them in their native habitats or in common garden experiments when these stocks maintain their distinct phenotypes. For an example of the profound effect of growth conditions on the morphology see comments and photos in Kosterin et al., (2020). Besides *P. s. ssp. elatius*, we included *P. s. ssp. humile* and its proposed separation into two varieties, *P. s. ssp. humile* var. *syriacum* (northern humile; 21 samples) and *P. s. ssp. humile* var. *humile* (southern humile; 12 samples) following Ladizinsky & Abbo (2015). Out of 21 northern *humile* samples 17 clustered in Cl1, which also included the *P. s. ssp. elatius* accession PeAb03 (Fig. 2; Tab.

S1). All samples classified as southern *humile* clustered in Cl2, which contained only another sample that was classified as northern *humile* (PeAb28). This contradicts the results of Vershinin et al. (2003) and Ellis (2011) and suggests that northern and southern *humile* are distinct groups within the *P. sativum* complex. Yet, they are just as distinct from each other as they are from other *P. s. ssp. elatius* groups. It should also be noted that according to our results *P. s. ssp. humile* does not take up a special place in the *P. sativum* complex. From a genetic standpoint, the *P. s. ssp. humile* groups are just two out of six well distinct entities within this complex.

Drivers of genetic diversity

We observed notable variation among genetic differentiation explained by the employed variables between clusters. Yet, some common patterns could be detected. In total, all variables accounted for a comparable fraction of genetic variation (Fig. 3). Cluster 6 was an outlier with a very small fraction of genetic variation associated with the employed predictors. As mentioned above, the small sample size of this cluster (5) could have biased the analysis and the results of this cluster should be viewed with caution.

The bioclimate appeared to be the strongest driver of genetic differentiation in *P. s. ssp. elatius* since the bioclimate variables were most strongly associated with genetic structure in all genetic clusters and accounted on average for 57 % of the explained variation. Overall, precipitation related variables had a stronger effect than temperature related variables (Fig. S2). These findings differ from the results of Smýkal et al. (2018) who did not find any signal of IBE in *P. s. ssp. elatius* populations from south-east Turkey. The limited geographic scale and lower environmental heterogeneity may be the reason why Smýkal et al. (2018) did not detect IBE. Limited IBE was also reported for *P. fulvum* from the southern Levant in a rather small study area, yet with great environmental heterogeneity. The authors attributed the absence of a strong IBE to strong genetic bottleneck in *P. fulvum*, which may have caused random genetic drift to obscure the signal of genome-wide differentiation following the environmental cline (Hellwig et al., 2020, 2021). Isolation by environment is very common in nature, for example, in a meta-

study 74 % (52 out of 70) of phylogeographic studies reported patterns of IBE (Sexton et al., 2013). The average effect size of IBE was estimated to be 0.34 when controlled for spatial autocorrelation (Shafer & Wolf, 2013). Our estimated effect was significantly higher which emphasizes the importance of the climate on genetic differentiation in *P. s. ssp. elatius*. These findings are particularly important for the utilization of *P. s. ssp. elatius* for breeding efforts, since it suggests a large fraction of adaptive variation is related to climatic factors. We found 963 (0.64 % of total) and 926 (0.62 % of total) SNPs significantly correlated with temperature and precipitation, respectively (Tab. S4; Fig. S3). A large fraction of these were non-synonymous polymorphisms within predicted genes (Tab. S4). If these candidates can be validated, they may be useful for crop improvement, because *P. s. ssp. elatius* is interfertile with the domesticated *P. s. ssp. sativum*.

Although most of the detected candidate SNPs putatively under selection were associated with latitude, the genome-wide effect of latitude was comparably small. These were rather unexpected results, especially for the entire population as well as for clusters with a large north-south distribution like Cl4 and Cl3. It is expected that a number of developmental traits change with latitude as an adaptation to varying day length. This has been observed in a variety of plant species like *Arabidopsis* (Riikimäli et al., 2005; Sami et al., 2008; Fabian et al., 2012; Debieu et al., 2013), soybean (Lu et al., 2015), wild lupin (Mousavi-Derazmahalleh et al., 2018), wild wheat (Matsuoka et al., 2008) and a number of different forest herbs (De Frenne et al., 2009). Local adaptation without genome-wide differentiation following the respective environmental cline has also been observed in *P. fulvum* (Hellwig et al., 2021) and likewise with lava flow lizards (Krohn et al., 2019). Krohn et al. (2019) argued that ongoing migration between populations may counteract genome-wide divergence. It has also been pointed out that genome-wide differentiation is less likely to occur if adaptive loci are not linked to traits involved in reproductive isolation (Slatkin, 1987; Presgraves, 2013). Yet, extensive gene flow over thousands of kilometers across the entire north-south distribution range of *P. s. ssp. elatius* is highly unlikely and loci involved in adaptation to latitude (i.e. growth temperature, day length)

are likely to be linked to reproductive traits like flowering time. Hellwig et al. (2021) explained their observed pattern with the genetic bottleneck underwent by the studied populations. They hypothesize that neutral areas of the genome are strongly influenced by random genetic drift caused by the reduced N_e while adaptive areas are shaped by selection. Even though we also observed a genetic bottleneck during the last glacial period, the scenario hypothesized for *P. fulvum* cannot explain the pattern observed for latitude, because we would expect to see a similar pattern for the other predictors. The age of the adaptation to latitude could be a possible explanation. With time, the linkage between adaptive and the surrounding neutral loci may break apart, leaving adaptive SNPs without associated genome-wide differentiation. However, we do not have any evidence to support this scenario, which therefore remains speculative.

The spatial predictors were most important in clusters with large geographic range. Euclidean distances between sample locations are not considered as meaningful predictor for genetic differentiation since it cannot account for landscape heterogeneity (Slatkin, 1985; McRae, 2005). Indeed, a straight line between sample locations is not a good approximation of the connectivity of plant populations. Despite this criticism, it is still commonly included in landscape genetics analyses. Therefore, we included absolute distance as a predictor in our analysis to make our results comparable with other studies, even though we agree with the criticism and advocate to use more meaningful variables to model the spatial connectivity between populations (e.g. resistance maps like used herein or least cost paths; for review see Zeller et al., 2012).

Land cover had a particularly large impact on genetic differentiation in central Israel (C12) and south-east Turkey (C11). Central Israel is highly populated and the areas suitable for wild peas is heavily fragmented by roads, industrial estates, urban areas and cropland. Similarly, there are large blocks of arable land in southern Turkey. Land cover was also an important predictor for the distribution of genetic variation in south Europe (C13), yet, to a smaller degree relative to the Levant. Southern Europe is also strongly impacted by agriculture, but contrary to the Levant, there are large areas covered by forests, herbaceous vegetation and shrubland that are overall rather well connected. These areas

may serve as routes that connect *P. s. ssp. elatius* populations and can reduce the effect of habitat fragmentation. Habitat fragmentation is certainly an important process affecting genetic diversity. Habitat fragmentation reduces N_e which negatively affects genetic diversity due to increased genetic drift (Ellstrand & Elam, 1993; Young et al., 1996). Honnay & Jacquemyn (2007) supported this theoretical prediction in a meta-study and emphasized that such a population genetic response to habitat fragmentation can be observed in rare as well as common species. Another meta-study demonstrated that regardless of the ecological or life-history traits of the investigated species, fragmentation is likely to negatively impact plant sexual reproduction (Aguilar et al., 2006). However, not all species react in the same way to habitat fragmentation. Encinas-Viso et al. (2020) showed that reduced forested area had a significant impact on the genetic differentiation of the Australian legume *Acacia salicina* but not on *A. stenophylla*, which inhabits largely the same area. Foré et al. (1992) even found decreased genetic differentiation in post-fragmentation populations compared to pre-fragmentation populations in *Acer saccharum* from the USA. The results of Prober & Brown (1994), who studied the effect of habitat fragmentation, due to cropland expansion, on *Eucalyptus albens*, indicate that there may be a fragmentation threshold and it is likely that the duration and intensity of habitat fragmentation plays an important role in the plant population genetic response. Our results suggest that one possible explanation is that habitat fragmentation due to intensive human land use, especially agriculture, causes increased genetic differentiation in wild pea. Yet, we cannot say which of the possible responses to habitat fragmentation have caused this increased genetic differentiation.

Demographic history

Our results suggested that *P. s. ssp. elatius* went through a genetic bottleneck during the last glacial period (Fig. 4). A similar population contraction was reported for the wild pea species *P. fulvum* (Hellwig et al., 2020, 2021). Yet, contrary to the situation in *P. fulvum*, N_e of *P. s. ssp. elatius* increased shortly after the bottleneck and post-recovery estimates of N_e after the recovery were even slightly higher than before the contraction. Hellwig et al., (2020, 2021) mentioned several possible reasons why pea populations may have

decreased during the last ice age like sea level fluctuations or the increased area covered by forests which is a less suitable habitat for peas. The bottleneck continues into the early phase of the Holocene when agricultural practices were already widespread within our study area and therefore, human made environmental disturbances like deforestation and increased grazing pressure by livestock likely affected pea populations as well. However, this cannot explain why *P. s. ssp. elatius* recovered from the bottleneck while *P. fulvum* did not. The fairly similar ecological affinities (including spatial distribution patterns) of the two wild pea species adds to this riddle. Our estimate of the overall selfing rate of *P. s. ssp. elatius* (0.700) is higher than the reported estimate of *P. fulvum* (0.495; Hellwig et al., 2021). Populations can benefit from higher selfing rates after a bottleneck, because it ensures reproduction in absence of mating partners (Cheptou, 2019). Therefore, the difference in reproduction pattern between *P. sativum* and *P. fulvum* may be considered as one of the factors that contributed to the difference in the recovery from the genetic bottleneck.

It is expected that a range expansion will create a gradient of genetic diversity along the expansion route due to serial founder effects where variation decreases with increasing distance from the origin (Austerlitz et al., 1997; Edmonds et al., 2004; Ramachandran et al., 2005; Slatkin & Excoffier, 2012). We did not observe such a gradient of genetic variation in our data and statistics of neutrality also did not support a recent population expansion (Tab. 2), although the estimation of historic N_e resulted in a population recovery after a strong genetic bottleneck during the last glacial period (Fig. 4). Cluster 4 had a significantly negative Tajima's D value indicating a possible population expansion after a bottleneck. However, this value was rather low and, hence, this expansion was likely not severe enough to explain the detected increase of N_e . In the other clusters, values of Tajima's D were significantly positive, which rather speak for a recent population contraction or alternatively, for balancing selection. Yet, Tajima's D in those clusters was very close to 0 except for C15 and C16, where the low sample size may have biased the results. Estimates of R_2 were not significantly different from 0 in any cluster contradicting the calculations of Tajima's D. The results of summary statistics are

ambiguous and do not allow any clear conclusions. Another population genetic signal of recent range expansion are clines in allele frequencies along the expansion trajectory (Slatkin & Excoffier, 2012). The results of the approach of Peter & Slatkin (2013, 2015) suggest a range expansion rejecting the null model of isolation-by-distance at equilibrium.

Note, that while the center of diversity (hence, the likely origin) of the genus as a whole is in the east Mediterranean basin, the detected clines of allele frequencies in our work is most likely the result of a recent expansion associated with the population recovery after the genetic bottleneck during the last glacial period. Genetic signals from the mid or early Pleistocene, when peas first expanded from the Near East westwards, were probably lost from recent populations due to the effect of other evolutionary forces. We detected significant clines in allele frequencies suggesting expansion routes from the West to the East (Cl3 → Cl6 & Cl1) and from Turkey to the South (Cl4 → Cl1; Cl5 → Cl2) and further to the East (Cl4 → Cl1; Cl5 → Cl2; Cl1 → Cl6; Fig. 5). It seems like population Cl4, which is located around the Black Sea expanded towards the South and East. At the same time the European cluster expanded eastwards. It is possible that the southward expansion of glaciers during the last ice age reduced the size of European populations of *P. s. ssp. elatius* and forced those into refugia as was observed in numerous Mediterranean plant species (Médail & Diadema, 2009). Following the retreat of glaciers, pea populations may have expanded from those refugia northward as well as eastwards. A process that could still be ongoing. Such an eastward expansion from refugia was also observed in other pan-Mediterranean legumes like narrow-leafed lupin (*Lupinus angustifolius*; Mousavi-Derazmahalleh et al., 2018) and pitch trefoil (*Bituminaria bituminosa*; García-Verdugo et al., 2021). The main difference between the *Lupinus* and *Bitumenaria* cases and wild peas is that the center of diversity of narrow-leafed lupin and pitch trefoil is located in the western rather than the eastern Mediterranean. Those species likely originated from the western Mediterranean, while the origin of *P. s. ssp. elatius* is located in eastern Mediterranean. The range expansion we detected may be the result of an eastward expansion from glacial refugia of the European populations which mixed with the eastern populations during the expansion and a simultaneous expansion of

Turkish populations to the South and further to the east.

Overall, the interplay between environmental factors and recent evolutionary history affected the distribution of genetic diversity of *P. s. ssp elatius* where each subpopulations were differently affected by those factors and processes.

Authors' credit statement

Timo Hellwig: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Writing; **Shahal Abbo:** Review & Editing, Supervision, Funding acquisition; **Ron Ophir:** Review & Editing, Data curation, Supervision, Funding acquisition.

Acknowledgements

We thank Dr. Naama Teboul for their critical comments on DNA extraction and Prof. Zvi Peleg for hosting T.H. in his lab during the preparation of DNA samples. The work was supported by The Israel Science Foundation [Grant/Award Number 307/17 to RO and SA]. Shahal Abbo is the incumbent of The Jacob & Rachel Liss Chair in Agronomy at the Hebrew University of Jerusalem. Timo Hellwig is a recipient of The Robert H. Smith stipend.

Competing Interests

All authors declare that they have no competing interests.

Tables and Figures

Table 1: Pairwise F_{st} values between genetic clusters (upper triangle) with corresponding p-values based on 999 bootstrap values (lower triangle).

	Cl1	Cl2	Cl3	Cl4	Cl5	Cl6
Cl1	---	0.327	0.347	0.162	0.249	0.258
Cl2	<0.001	---	0.409	0.366	0.342	0.395
Cl3	<0.001	<0.001	---	0.372	0.350	0.438
Cl4	<0.001	<0.001	<0.001	---	0.283	0.297
Cl5	<0.001	<0.001	<0.001	<0.001	---	0.340
Cl6	<0.001	<0.001	<0.001	<0.001	<0.001	---

Table 2: Statistics of diversity (H_o : observed heterozygosity, H_e : expected heterozygosity, π : nucleotide diversity, H_{tn} : number of haplotypes, H_{td} : haplotype diversity) and neutrality of each genetic cluster (Tajima's D and R_2).

	H_o	H_e	π	H_{tn}	H_{td}	Tajima's D	R_2
Cl1	0.007	0.103	0.26	6538	0.88	0.02 (0.05)	0.12 (0.5)
Cl2	0.016	0.082	0.29	5384	0.86	0.09 (< 1e-05)	0.13 (0.56)
Cl3	0.009	0.065	0.28	4333	0.83	0.16 (< 1e-05)	0.13 (0.55)
Cl4	0.006	0.092	0.24	4045	0.74	-0.12 (< 1e-05)	0.09 (0.41)
Cl5	0.009	0.078	0.38	3811	0.92	0.58 (< 1e-05)	0.17 (0.55)
Cl6	0.006	0.080	0.46	2893	0.92	0.89 (< 1e-05)	0.20 (0.51)

Figure 1: Map of sample locations. Pie charts are located at the sample location of the analyzed accessions and represent sNMF subpopulation fractions with K=6 (for a more detailed map of the Levantinian accessions see Fig. S4).

Figure 2: Results of population structure analysis. A: subpopulation fraction estimated by sNMF with K ranging from 2 through 6. B: PcoA; percentages in axes titles represent the fraction of explained variation; accessions that did not have a subpopulation fraction of > 0.5 of any genetic cluster were defined as admixed. C: Neighbor-net phylogenetic network.

Figure 4: Inferences of past effective population size. The solid line represents the median N_e . Red and pink shaded areas represent the 97.5 and 87.5 percentiles, respectively. Both axes are on $\log_{10}$ scale.

Figure 5: Distribution centers of genetic clusters with expansion routes estimated by pairwise directionality indices ψ . Arrows indicate the putative directions of expansions. Arrows' thickness are proportional to the respective values of ψ . Only significant ψ are depicted (for exact values of ψ and their respective p-values see Tab. S5).

Supplementary Figures

Figure S1: Distribution of SNP density over linkage groups. Outer circle (green): filtered variant set. Inner circle (red): LD pruned variant set.

Figure S2: R²-weighted importance estimated in gradient forest analysis of each individual predictor for each genetic cluster and the whole population.

Figure S3: Genome-wide distribution of the SNPs associated with environmental factors. A: entire set of candidates. B: 10 % of candidates of each environmental predictor with the highest r values in RDA.

Figure S4: Detailed maps of Levantine sample locations. Pie charts are located at the sample location of the analyzed accessions and represent sNMF subpopulation fractions with K=6 (for a map depicting the entire sample collection see Fig. 1).

References

- Aguilar, R., Ashworth, L., Galetto, L., & Aizen, M. A. (2006). Plant reproductive susceptibility to habitat fragmentation: Review and synthesis through a meta-analysis. *Ecology Letters*, 9(8), 968–980. doi: <https://doi.org/10.1111/j.1461-0248.2006.00927.x>
- Anantharaman, R., Hall, K., Shah, V. B., & Edelman, A. (2020). Circuitscape in Julia: High Performance Connectivity Modelling to Support Conservation Decisions. *Proceedings of the JuliaCon Conferences*, 1(1), 58. doi: [10.21105/jcon.00058](https://doi.org/10.21105/jcon.00058)
- Austerlitz, F., Jung-Muller, B., Godelle, B., & Gouyon, P.-H. (1997). Evolution of Coalescence Times, Genetic Diversity and Structure during Colonization. *Theoretical Population Biology*, 51(2), 148–164. doi: [10.1006/tpbi.1997.1302](https://doi.org/10.1006/tpbi.1997.1302)
- Baker, A. J., & Moeed, A. (1987). Rapid Genetic Differentiation and Founder Effect in Colonizing Populations of Common Mynas (acridotheres Tristis). *Evolution*, 41(3), 525–538. doi: <https://doi.org/10.1111/j.1558-5646.1987.tb05823.x>
- Barata, A. M., Rocha, F., Lopes, V., & Carvalho, A. M. (2016). Conservation and sustainable uses of medicinal and aromatic plants genetic resources on the worldwide for human welfare. *Industrial Crops and Products*, 88, 8–11. doi: [10.1016/j.indcrop.2016.02.035](https://doi.org/10.1016/j.indcrop.2016.02.035)
- Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics (Oxford, England)*, 21(2), 263–265. doi: [10.1093/bioinformatics/bth457](https://doi.org/10.1093/bioinformatics/bth457)
- Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, 23(1), 38–44. doi: [10.1016/j.tree.2007.09.008](https://doi.org/10.1016/j.tree.2007.09.008)
- Baskin, C. C., & Baskin, J. M. (1998). *Seeds: Ecology, Biogeography, and, Evolution of Dormancy and Germination*. Elsevier.
- Baucom, R. S., Estill, J. C., & Cruzan, M. B. (2005). The effect of deforestation on the genetic diversity and structure in *Acer saccharum* (Marsh): Evidence for the loss and restructuring of genetic variation in a natural system. *Conservation Genetics*, 6(1), 39–50. doi: [10.1007/s10592-004-7718-9](https://doi.org/10.1007/s10592-004-7718-9)
- Bauman, D., Drouet, T., Fortin, M.-J., & Dray, S. (2018). Optimizing the choice of a spatial weighting matrix in eigenvector-based methods. *Ecology*, 99(10), 2159–2166. doi: <https://doi.org/10.1002/ecy.2469>
- Ben-Ze'ev, N., & Zohary, D. (1973). Species relation in the genus *Pisum* L. *Israel Journal of Botany*, 22, 73–91.
- Blanchet, F. G., Legendre, P., & Borcard, D. (2008). Forward Selection of Explanatory Variables. *Ecology*, 89(9), 2623–2632. doi: <https://doi.org/10.1890/07-0986.1>
- Bogdanova, V. S., Mglinets, A. V., Shatskaya, N. V., Kosterin, O. E., Solovyev, V. I., & Vasiliev, G. V. (2018). Cryptic divergences in the genus *Pisum* L. (peas) as revealed by phylogenetic analysis of plastid genomes. *Molecular Phylogenetics and Evolution*, 129, 280–290. doi: [10.1016/j.ympev.2018.09.002](https://doi.org/10.1016/j.ympev.2018.09.002)

- Boheemen, L. A. van, Atwater, D. Z., & Hodgins, K. A. (2019). Rapid and repeated local adaptation to climate in an invasive plant. *New Phytologist*, 222(1), 614–627. doi: <https://doi.org/10.1111/nph.15564>
- Boissier, E. (1867). *Flora Orientalis: Sive, Enumeratio plantarum in Oriente a Graecia et Aegypto ad Indiae fines hucusque observatarum. Thalamiflorae. Volumen primum* (Vol. 1). G. Bibliopolam.
- Bolnick, D. I., & Otto, S. P. (2013). The magnitude of local adaptation under genotype-dependent dispersal. *Ecology and Evolution*, 3(14), 4722–4735. doi: [10.1002/ece3.850](https://doi.org/10.1002/ece3.850)
- Bouzat, J. L., Cheng, H. H., Lewin, H. A., Westemeier, R. L., Brawn, J. D., & Paige, K. N. (1998). Genetic Evaluation of a Demographic Bottleneck in the Greater Prairie Chicken. *Conservation Biology*, 12(4), 836–843. doi: <https://doi.org/10.1111/j.1523-1739.1998.97164.x>
- Browning, B. L., & Browning, S. R. (2009). A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. *The American Journal of Human Genetics*, 84(2), 210–223. doi: [10.1016/j.ajhg.2009.01.005](https://doi.org/10.1016/j.ajhg.2009.01.005)
- Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *The American Journal of Human Genetics*, 103(3), 338–348. doi: [10.1016/j.ajhg.2018.07.015](https://doi.org/10.1016/j.ajhg.2018.07.015)
- Browning, S. R., & Browning, B. L. (2007). Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *The American Journal of Human Genetics*, 81(5), 1084–1097. doi: [10.1086/521987](https://doi.org/10.1086/521987)
- Byars, S. G., Parsons, Y., & Hoffmann, A. A. (2009). Effect of altitude on the genetic structure of an Alpine grass, *Poa hiemata*. *Annals of Botany*, 103(6), 885–899. doi: [10.1093/aob/mcp018](https://doi.org/10.1093/aob/mcp018)
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124–3140. doi: [10.1111/mec.12354](https://doi.org/10.1111/mec.12354)
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes, Genomes, Genetics*, 1(3), 171–182. doi: [10.1534/g3.111.000240](https://doi.org/10.1534/g3.111.000240)
- Caye, K., Jumentier, B., Lepeule, J., & François, O. (2019). LFMM 2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies. *Molecular Biology and Evolution*, 36(4), 852–860. doi: [10.1093/molbev/msz008](https://doi.org/10.1093/molbev/msz008)
- Chambers, J. M. (2013). SoDA: Functions and Examples for “Software for Data Analysis.” <https://CRAN.R-Project.Org/Package=SoDA>.
- Cheptou, P.-O. (2019). Does the evolution of self-fertilization rescue populations or increase the risk of extinction? *Annals of Botany*, 123(2), 337–345. doi: [10.1093/aob/mcy144](https://doi.org/10.1093/aob/mcy144)
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w¹¹¹⁸; iso-2; iso-3. *Fly*, 6(2), 80–92. doi: [10.4161/fly.19695](https://doi.org/10.4161/fly.19695)
- Crozier, R. H. (1997). Preserving the Information Content of Species: Genetic Diversity, Phylogeny, and Conservation Worth. *Annual Review of Ecology and Systematics*, 28(1), 243–268. doi: [10.1146/annurev.ecolsys.28.1.243](https://doi.org/10.1146/annurev.ecolsys.28.1.243)

- Crutsinger, G. M., Collins, M. D., Fordyce, J. A., Gompert, Z., Nice, C. C., & Sanders, N. J. (2006). Plant Genotypic Diversity Predicts Community Structure and Governs an Ecosystem Process. *Science*, 313(5789), 966–968. doi: [10.1126/science.1128326](https://doi.org/10.1126/science.1128326)
- Crutsinger, G. M., Souza, L., & Sanders, N. J. (2008). Intraspecific diversity and dominant genotypes resist plant invasions. *Ecology Letters*, 11(1), 16–23. doi: <https://doi.org/10.1111/j.1461-0248.2007.01118.x>
- Cruzan, M. B., & Hendrickson, E. C. (2020). Landscape Genetics of Plants: Challenges and Opportunities. *Plant Communications*, 1(6). doi: [10.1016/j.xplc.2020.100100](https://doi.org/10.1016/j.xplc.2020.100100)
- Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random Forests for Classification in Ecology. *Ecology*, 88(11), 2783–2792. doi: <https://doi.org/10.1890/07-0539.1>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... and, R. D. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. doi: [10.1093/bioinformatics/btr330](https://doi.org/10.1093/bioinformatics/btr330)
- David, P., Pujol, B., Viard, F., Castella, V., & Goudet, J. (2007). Reliable selfing rate estimates from imperfect population genetic data. *Molecular Ecology*, 16(12), 2474–2487. doi: [10.1111/j.1365-294x.2007.03330.x](https://doi.org/10.1111/j.1365-294x.2007.03330.x)
- Davis, P. H. (1970). *Pisum L. In Flora of Turkey and the East Aegean Islands* (Vol. 3, pp. 370–373). Edinburgh: Edinburgh University Press.
- De Meester, L., Gómez, A., Okamura, B., & Schwenk, K. (2002). The Monopolization Hypothesis and the dispersal–gene flow paradox in aquatic organisms. *Acta Oecologica*, 23(3), 121–135. doi: [10.1016/S1146-609X\(02\)01145-1](https://doi.org/10.1016/S1146-609X(02)01145-1)
- De Meester, L., Vanoverbeke, J., Kilsdonk, L. J., & Urban, M. C. (2016). Evolving Perspectives on Monopolization and Priority Effects. *Trends in Ecology & Evolution*, 31(2), 136–146. doi: [10.1016/j.tree.2015.12.009](https://doi.org/10.1016/j.tree.2015.12.009)
- Debieu, M., Tang, C., Stich, B., Sikosek, T., Effgen, S., Josephs, E., ... Meaux, J. de. (2013). Co-Variation between Seed Dormancy, Growth Rate and Flowering Time Changes with Latitude in *Arabidopsis thaliana*. *PLOS ONE*, 8(5), e61075. doi: [10.1371/journal.pone.0061075](https://doi.org/10.1371/journal.pone.0061075)
- Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., ... Wagner, H. H. (2020). *adespatial: Multivariate Multiscale Spatial Analysis*. <https://CRAN.R-Project.Org/Package=adespatial>.
- Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, 196(3), 483–493. doi: [10.1016/j.ecolmodel.2006.02.015](https://doi.org/10.1016/j.ecolmodel.2006.02.015)
- Edelaar, P., Siepielski, A. M., & Clobert, J. (2008). Matching Habitat Choice Causes Directed Gene Flow: A Neglected Dimension in Evolution and Ecology. *Evolution*, 62(10), 2462–2472. doi: [10.1111/j.1558-5646.2008.00459.x](https://doi.org/10.1111/j.1558-5646.2008.00459.x)
- Edmonds, C. A., Lillie, A. S., & Cavalli-Sforza, L. L. (2004). Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences*, 101(4), 975–979. doi: [10.1073/pnas.0308064100](https://doi.org/10.1073/pnas.0308064100)
- Elisabetsky, E., & Costa-Campos, L. (1996). Medicinal plant genetic resources and international cooperation: The Brazilian perspective. *Journal of Ethnopharmacology*, 51(1), 111–120. doi: [10.1016/0378-8741\(95\)01353-9](https://doi.org/10.1016/0378-8741(95)01353-9)

- Ellis, N., Smith, S. J., & Pitcher, C. R. (2012). Gradient forests: Calculating importance gradients on physical predictors. *Ecology*, 93(1), 156–168. doi: <https://doi.org/10.1890/11-0252.1>
- Ellis, T. H. N. (2011). Pisum. In *Wild Crop Relatives: Genomic and Breeding Resources: Legume Crops and Forages* (Kole, Chittaranjan, Vol. 10, pp. 237–248). Heidelberg: Springer-Verlag. Retrieved from <https://www.springer.com/us/book/9783642143861>
- Ellstrand, N. C., & Elam, D. R. (1993). Population Genetic Consequences of Small Population Size: Implications for Plant Conservation. *Annual Review of Ecology and Systematics*, 24(1), 217–242. doi: [10.1146/annurev.es.24.110193.001245](https://doi.org/10.1146/annurev.es.24.110193.001245)
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLOS ONE*, 6(5), e19379. doi: [10.1371/journal.pone.0019379](https://doi.org/10.1371/journal.pone.0019379)
- Encinas-Viso, F., McDonald-Spicer, C., Knerr, N., Thrall, P. H., & Broadhurst, L. (2020). Different landscape effects on the genetic structure of two broadly distributed woody legumes, *Acacia salicina* and *A. stenophylla* (Fabaceae). *Ecology and Evolution*, 10(23), 13476–13487. doi: <https://doi.org/10.1002/ece3.6952>
- Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), 481–501. doi: [10.1146/annurev.ecolsys.39.110707.173414](https://doi.org/10.1146/annurev.ecolsys.39.110707.173414)
- Exposito-Alonso, M., Becker, C., Schuenemann, V. J., Reiter, E., Setzer, C., Slovak, R., ... Weigel, D. (2018). The rate and potential relevance of new mutations in a colonizing plant lineage. *PLOS Genetics*, 14:2. doi: [10.1371/journal.pgen.1007155](https://doi.org/10.1371/journal.pgen.1007155)
- Fabian, D. K., Kapun, M., Nolte, V., Kofler, R., Schmidt, P. S., Schlötterer, C., & Flatt, T. (2012). Genome-wide patterns of latitudinal differentiation among populations of *Drosophila melanogaster* from North America. *Molecular Ecology*, 21(19), 4748–4769. doi: <https://doi.org/10.1111/j.1365-294X.2012.05731.x>
- Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Molecular Ecology*, 27(9), 2215–2233. doi: [10.1111/mec.14584](https://doi.org/10.1111/mec.14584)
- Frenne, P. D., Kolb, A., Verheyen, K., Brunet, J., Chabrierie, O., Decocq, G., ... Graae, B. J. (2009). Unravelling the effects of temperature, latitude and local environment on the reproduction of forest herbs. *Global Ecology and Biogeography*, 18(6), 641–651. doi: <https://doi.org/10.1111/j.1466-8238.2009.00487.x>
- Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6(8), 925–929. doi: [10.1111/2041-210x.12382](https://doi.org/10.1111/2041-210x.12382)
- Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and Efficient Estimation of Individual Ancestry Coefficients. *Genetics*, 196(4), 973–983. doi: [10.1534/genetics.113.160572](https://doi.org/10.1534/genetics.113.160572)
- Fukami, T. (2015). Historical Contingency in Community Assembly: Integrating Niches, Species Pools, and Priority Effects. *Annual Review of Ecology, Evolution, and Systematics*, 46(1), 1–23. doi: [10.1146/annurev-ecolsys-110411-160340](https://doi.org/10.1146/annurev-ecolsys-110411-160340)
- García-Verdugo, C., Mairal, M., Tamaki, I., & Msanda, F. (2021). Phylogeography at the crossroad: Pleistocene range expansion throughout the Mediterranean and back-colonization from the Canary

Islands in the legume *Bituminaria bituminosa*. *Journal of Biogeography*. doi: <https://doi.org/10.1111/jbi.14100>

- Grasty, M. R., Thompson, P. G., Hendrickson, E. C., Pheil, A. E., & Cruzan, M. B. (2020). Fine-scale habitat heterogeneity and vole runways influence seed dispersal in *Plagiobothrys nothofulvus*. *American Journal of Botany*, 107(3), 413–422. doi: [10.1002/ajb2.1433](https://doi.org/10.1002/ajb2.1433)
- Hasegawa, M., Kishino, H., & Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, 22(2), 160–174. doi: [10.1007/bf02101694](https://doi.org/10.1007/bf02101694)
- Hellwig, T., Abbo, S., Sherman, A., Coyne, C. J., Saranga, Y., Lev-Yadun, S., ... Ophir, R. (2020). Limited divergent adaptation despite a substantial environmental cline in wild pea. *Molecular Ecology*, 29(22). doi: [10.1111/mec.15633](https://doi.org/10.1111/mec.15633)
- Hellwig, T., Abbo, S., Sherman, A., & Ophir, R. (2021). Prospects for genetic diversity of crop wild relatives under climate change: The case of the wild pea *Pisum fulvum*. *Plant Science*, 110957. doi: [10.1016/j.plantsci.2021.110957](https://doi.org/10.1016/j.plantsci.2021.110957)
- Hoey, B. K., Crowe, K. R., Jones, V. M., & Polans, N. O. (1996). A phylogenetic analysis of *Pisum* based on morphological characters and allozyme and RAPD markers. *Theoretical and Applied Genetics*, 92(1), 92–100. doi: [10.1007/bf00222957](https://doi.org/10.1007/bf00222957)
- Hoisington, D., Khairallah, M., Reeves, T., Ribaut, J.-M., Skovmand, B., Taba, S., & Warburton, M. (1999). Plant genetic resources: What can they contribute toward increased crop productivity? *Proceedings of the National Academy of Sciences*, 96(11), 5937–5943. doi: [10.1073/pnas.96.11.5937](https://doi.org/10.1073/pnas.96.11.5937)
- Honnay, O., & Jacquemyn, H. (2007). Susceptibility of Common and Rare Plant Species to the Genetic Consequences of Habitat Fragmentation. *Conservation Biology*, 21(3), 823–831. doi: <https://doi.org/10.1111/j.1523-1739.2006.00646.x>
- Hughes, A. R., Inouye, B. D., Johnson, M. T. J., Underwood, N., & Vellend, M. (2008). Ecological consequences of genetic diversity. *Ecology Letters*, 11(6), 609–623. doi: [10.1111/j.1461-0248.2008.01179.x](https://doi.org/10.1111/j.1461-0248.2008.01179.x)
- Huson, D. H., & Bryant, D. (2005). Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution*, 23(2), 254–267. doi: [10.1093/molbev/msj030](https://doi.org/10.1093/molbev/msj030)
- Jaradat, A. A. (2010). *Genetic resources of energy crops: Biological systems to combat climate change*. Retrieved from <https://pubag.nal.usda.gov/catalog/49241>
- Jari, O. (2018). vegan: Community Ecology Package. <https://CRAN.R-Project.Org/Package=vegan>.
- Jing, R., Johnson, R., Seres, A., Kiss, G., Ambrose, M. J., Knox, M. R., ... Flavell, A. J. (2007). Gene-Based Sequence Diversity Analysis of Field Pea (*Pisum*). *Genetics*, 177(4), 2263–2275. doi: [10.1534/genetics.107.081323](https://doi.org/10.1534/genetics.107.081323)
- Jing, R., Vershinin, A., Grzebyta, J., Shaw, P., Smýkal, P., Marshall, D., ... Flavell, A. J. (2010). The genetic diversity and evolution of field pea (*Pisum*) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. *BMC Evolutionary Biology*, 10(1), 44. doi: [10.1186/1471-2148-10-44](https://doi.org/10.1186/1471-2148-10-44)
- Joly, S., Bryant, D., & Lockhart, P. J. (2015). Flexible methods for estimating genetic distances from single nucleotide polymorphisms. *Methods in Ecology and Evolution*, 6(8), 938–948. doi: [10.1111/2041-210x.12343](https://doi.org/10.1111/2041-210x.12343)

- Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070–3071. doi: [10.1093/bioinformatics/btr521](https://doi.org/10.1093/bioinformatics/btr521)
- Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, 6. doi: [10.3389/fgene.2015.00208](https://doi.org/10.3389/fgene.2015.00208)
- Kannenberg, L. W., & Falk, D. E. (2011). Models for activation of plant genetic resources for crop breeding programs. *Canadian Journal of Plant Science*. (Ottawa, Canada). doi: [10.4141/cjps95-008](https://doi.org/10.4141/cjps95-008)
- Kiru, S., & Nasenko, I. (2010). Use of genetic resources from Jerusalem artichoke collection of N. Vavilov Institute in breeding for bioenergy and health security. *Agronomy Research*, 8, 625–632.
- Knudby, A., Brenning, A., & LeDrew, E. (2010). New approaches to modelling fish–habitat relationships. *Ecological Modelling*, 221(3), 503–511. doi: [10.1016/j.ecolmodel.2009.11.008](https://doi.org/10.1016/j.ecolmodel.2009.11.008)
- Kosterin, O E, & Bogdanova, V. S. (2018). Obscuring the routes: Confused data cannot reveal phylogeography of pea crop wild relatives (refutation to ‘Genomic diversity and macroecology of the crop wild relatives of domesticated pea’ by Smýkal et al. 2017). *Acta Biologica Sibirica*, 4(3), 66–68.
- Kosterin, O. E., Bogdanova, V. S., & Mglinets, A. V. (2020). Wild pea (*Pisum sativum* L. subsp. *elatius* (Bieb.) Aschers. et Graebn. s.l.) at the periphery of its range: Zagros Mountains. *Vavilovskii Zhurnal Genetiki I Seleksii*, 24(1), 60–68. doi: [10.18699/VJ20.596](https://doi.org/10.18699/VJ20.596)
- Kosterin, Oleg E., Zaytseva, O. O., Bogdanova, V. S., & Ambrose, M. J. (2010). New data on three molecular markers from different cellular genomes in Mediterranean accessions reveal new insights into phylogeography of *Pisum sativum* L. subsp. *Elatius* (Bieb.) Schmalh. *Genetic Resources and Crop Evolution*, 57(5), 733–739. doi: [10.1007/s10722-009-9511-6](https://doi.org/10.1007/s10722-009-9511-6)
- Kreplak, J., Madoui, M., Cápál, P., Novák, P., Labadie, K., Aubert, G., ... Burstin, J. (2019). A reference genome for pea provides insight into legume genome evolution. *Nat Genet*, 51, 1411–1422.
- Ladizinsky, G., & Abbo, S. (2015). *The Search for Wild Relatives of Cool Season Legumes*. Cham: Springer International Publishing.
- Lai, Y.-T., Yeung, C. K. L., Omland, K. E., Pang, E.-L., Hao, Y., Liao, B.-Y., ... Li, S.-H. (2019). Standing genetic variation as the predominant source for adaptation of a songbird. *Proceedings of the National Academy of Sciences*, 116(6), 2152–2157. doi: [10.1073/pnas.1813597116](https://doi.org/10.1073/pnas.1813597116)
- Laird, C. D., McConaughy, B. L., & McCarthy, B. J. (1969). Rate of Fixation of Nucleotide Substitutions in Evolution. *Nature*, 224(5215), 149–154. doi: [10.1038/224149a0](https://doi.org/10.1038/224149a0)
- Liu, X., & Fu, Y.-X. (2020). Stairway Plot 2: Demographic history inference with folded SNP frequency spectra. *Genome Biology*, 21(1), 280. doi: [10.1186/s13059-020-02196-9](https://doi.org/10.1186/s13059-020-02196-9)
- Lowe, A. J., Boshier, D., Ward, M., Bacles, C. F. E., & Navarro, C. (2005). Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. *Heredity*, 95(4), 255–273. doi: [10.1038/sj.hdy.6800725](https://doi.org/10.1038/sj.hdy.6800725)
- Lu, S., Li, Y., Wang, J., Srinives, P., Nan, H., Cao, D., ... Kong, F. (2015). QTL mapping for flowering time in different latitude in soybean. *Euphytica*, 206(3), 725–736. doi: [10.1007/s10681-015-1501-5](https://doi.org/10.1007/s10681-015-1501-5)
- Lynch, M. (2010). Evolution of the mutation rate. *Trends in Genetics*, 26:8, 345–352. doi: [10.1016/j.tig.2010.05.003](https://doi.org/10.1016/j.tig.2010.05.003)

- Marcel Buchhorn, Bruno Smets, Luc Bertels, Bert De Roo, Myroslava Lesiv, Nandin-Erdene Tsendbazar, ... Steffen Fritz. (2020). *Copernicus Global Land Service: Land Cover 100m: collection 3: epoch 2019: Globe* [Data set]. Zenodo. doi: [10.5281/zenodo.3939050](https://doi.org/10.5281/zenodo.3939050)
- Matsuoka, Y., Takumi, S., & Kawahara, T. (2008). Flowering Time Diversification and Dispersal in Central Eurasian Wild Wheat *Aegilops tauschii* Coss.: Genealogical and Ecological Framework. *PLOS ONE*, 3(9), e3138. doi: [10.1371/journal.pone.0003138](https://doi.org/10.1371/journal.pone.0003138)
- Matuszewski, S., Hermisson, J., & Kopp, M. (2015). Catch Me if You Can: Adaptation from Standing Genetic Variation to a Moving Phenotypic Optimum. *Genetics*, 200(4), 1255–1274. doi: [10.1534/genetics.115.178574](https://doi.org/10.1534/genetics.115.178574)
- Maxted, N., & Ambrose, M. (2001). Peas (*Pisum L.*). In *Current Plant Science and Biotechnology in Agriculture. Plant Genetic Resources of Legumes in the Mediterranean* (pp. 181–190). Springer, Dordrecht. doi: [10.1007/978-94-015-9823-1_10](https://doi.org/10.1007/978-94-015-9823-1_10)
- McBride, C. S., & Singer, M. C. (2010). Field Studies Reveal Strong Postmating Isolation between Ecologically Divergent Butterfly Populations. *PLOS Biology*, 8(10), e1000529. doi: [10.1371/journal.pbio.1000529](https://doi.org/10.1371/journal.pbio.1000529)
- McRae, B. H. (2006). ISOLATION BY RESISTANCE. *Evolution*, 60(8), 1551–1561. doi: [10.1554/05-321.1](https://doi.org/10.1554/05-321.1)
- McRae, B. H., & Beier, P. (2007). Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences*, 104(50), 19885–19890. doi: [10.1073/pnas.0706568104](https://doi.org/10.1073/pnas.0706568104)
- Médail, F., & Diadema, K. (2009). Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography*, 36(7), 1333–1345. doi: <https://doi.org/10.1111/j.1365-2699.2008.02051.x>
- Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, 21(12), 2839–2846. doi: [10.1111/j.1365-294X.2012.05578.x](https://doi.org/10.1111/j.1365-294X.2012.05578.x)
- Mousavi-Derazmahalleh, M., Bayer, P. E., Nevado, B., Hurgobin, B., Filatov, D., Kilian, A., ... Nelson, M. N. (2018). Exploring the genetic and adaptive diversity of a pan-Mediterranean crop wild relative: Narrow-leaved lupin. *Theoretical and Applied Genetics*, 131(4), 887–901. doi: [10.1007/s00122-017-3045-7](https://doi.org/10.1007/s00122-017-3045-7)
- Murray, K., & Conner, M. M. (2009). Methods to quantify variable importance: Implications for the analysis of noisy ecological data. *Ecology*, 90(2), 348–355. doi: <https://doi.org/10.1890/07-1929.1>
- Murray, K. D., & Borevitz, J. O. (2017). Axe: Rapid competitive sequence read demultiplexing using a trie. *Bioinformatics*, 34(22), 3924–3925. doi: [10.1101/160606](https://doi.org/10.1101/160606)
- Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89, 583–590.
- Nishimura, M., & Setoguchi, H. (2011). Homogeneous genetic structure and variation in tree architecture of *Larix kaempferi* along altitudinal gradients on Mt. Fuji. *Journal of Plant Research*, 124(2), 253–263. doi: [10.1007/s10265-010-0370-1](https://doi.org/10.1007/s10265-010-0370-1)
- Nosil, P., Vines, T. H., & Funk, D. J. (2005). Reproductive Isolation Caused by Natural Selection Against Immigrants from Divergent Habitats. *Evolution*, 59(4), 705–719. doi: [10.1111/j.0014-3820.2005.tb01747.x](https://doi.org/10.1111/j.0014-3820.2005.tb01747.x)

- Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., & Meester, L. D. (2013). Drivers of population genetic differentiation in the wild: Isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology*, 22(24), 5983–5999. doi: <https://doi.org/10.1111/mec.12561>
- Palmer, J. D., Jorgensen, R. A., & Thompson, W. F. (1985). Chloroplast DNA variation and evolution in *Pisum*: Patterns of change and phylogenetic analysis. *Genetics*, 109.
- Paradis, E. (2010). pegas: An R package for population genetics with an integrated–modular approach. *Bioinformatics*, 26(3), 419–420. doi: [10.1093/bioinformatics/btp696](https://doi.org/10.1093/bioinformatics/btp696)
- Pembleton, L. W., Cogan, N. O. I., & Forster, J. W. (2013). StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources*, 13(5), 946–952. doi: [10.1111/1755-0998.12129](https://doi.org/10.1111/1755-0998.12129)
- Peter, B. M., & Slatkin, M. (2013). Detecting Range Expansions from Genetic Data. *Evolution*, 67(11), 3274–3289. doi: <https://doi.org/10.1111/evo.12202>
- Peter, B. M., & Slatkin, M. (2015). The effective founder effect in a spatially expanding population. *Evolution; International Journal of Organic Evolution*, 69(3), 721–734. doi: [10.1111/evo.12609](https://doi.org/10.1111/evo.12609)
- Pope, N. (2020). r2vcftools: An R interface for vcftools. <https://github.com/Nspope/R2vcftools>.
- Prentis, P. J., Wilson, J. R. U., Dormontt, E. E., Richardson, D. M., & Lowe, A. J. (2008). Adaptive evolution in invasive species. *Trends in Plant Science*, 13(6), 288–294. doi: [10.1016/j.tplants.2008.03.004](https://doi.org/10.1016/j.tplants.2008.03.004)
- Presgraves, D. C. (2013). Hitchhiking to Speciation. *PLOS Biology*, 11(2), e1001498. doi: [10.1371/journal.pbio.1001498](https://doi.org/10.1371/journal.pbio.1001498)
- Prober, S. M., & Brown, A. H. D. (1994). Conservation of the Grassy White Box Woodlands: Population Genetics and Fragmentation of *Eucalyptus albens*. *Conservation Biology*, 8(4), 1003–1013.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, 81(3), 559–575. doi: [10.1086/519795](https://doi.org/10.1086/519795)
- Ramachandran, S., Deshpande, O., Roseman, C. C., Rosenberg, N. A., Feldman, M. W., & Cavalli-Sforza, L. L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences*, 102(44), 15942–15947. doi: [10.1073/pnas.0507611102](https://doi.org/10.1073/pnas.0507611102)
- Ramos-Onsins, S. E., & Rozas, J. (2002). Statistical Properties of New Neutrality Tests Against Population Growth. *Molecular Biology and Evolution*, 19(12), 2092–2100. doi: [10.1093/oxfordjournals.molbev.a004034](https://doi.org/10.1093/oxfordjournals.molbev.a004034)
- Rannala, B., & Yang, Z. (2003). Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, 164(4), 1645–1656.
- Reusch, T. B. H., Ehlers, A., Hämmerli, A., & Worm, B. (2005). Ecosystem recovery after climatic extremes enhanced by genotypic diversity. *Proceedings of the National Academy of Sciences*, 102(8), 2826–2831. doi: [10.1073/pnas.0500008102](https://doi.org/10.1073/pnas.0500008102)
- Reynolds, L. K., McGlathery, K. J., & Waycott, M. (2012). Genetic Diversity Enhances Restoration Success by Augmenting Ecosystem Services. *PLOS ONE*, 7(6), e38397. doi: [10.1371/journal.pone.0038397](https://doi.org/10.1371/journal.pone.0038397)

- Riihimäki, M., Podolsky, R., Kuittinen, H., Koelewijn, H., & Savolainen, O. (2005). Studying genetics of adaptive variation in model organisms: Flowering time variation in *Arabidopsis lyrata*. In R. Mauricio (Ed.), *Genetics of Adaptation* (pp. 63–74). Dordrecht: Springer Netherlands. doi: [10.1007/1-4020-3836-4_7](https://doi.org/10.1007/1-4020-3836-4_7)
- Russi, L., Cocks, P. S., & Roberts, E. H. (1992). Hard-seededness and seed bank dynamics of six pasture legumes. *Seed Science Research*, 2(4), 231–241. doi: [10.1017/S0960258500001422](https://doi.org/10.1017/S0960258500001422)
- Samis, K. E., Heath, K. D., & Stinchcombe, J. R. (2008). Discordant Longitudinal Clines in Flowering Time and Phytochrome C in *Arabidopsis Thaliana*. *Evolution*, 62(12), 2971–2983. doi: <https://doi.org/10.1111/j.1558-5646.2008.00484.x>
- Servedio, M. R. (2004). The Evolution of Premating Isolation: Local Adaptation and Natural and Sexual Selection Against Hybrids. *Evolution*, 58(5), 913–924. doi: [10.1111/j.0014-3820.2004.tb00425.x](https://doi.org/10.1111/j.0014-3820.2004.tb00425.x)
- Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2013). Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution*, 68:1, 1–15. doi: [10.1111/evo.12258](https://doi.org/10.1111/evo.12258)
- Shafer, A. B. A., & Wolf, J. B. W. (2013). Widespread evidence for incipient ecological speciation: A meta-analysis of isolation-by-ecology. *Ecology Letters*, 16(7), 940–950. doi: [10.1111/ele.12120](https://doi.org/10.1111/ele.12120)
- Slatkin, M. (1985). Gene Flow in Natural Populations. *Annual Review of Ecology and Systematics*, 16(1), 393–430. doi: [10.1146/annurev.es.16.110185.002141](https://doi.org/10.1146/annurev.es.16.110185.002141)
- Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science*, 236(4803), 787–792. doi: [10.1126/science.3576198](https://doi.org/10.1126/science.3576198)
- Slatkin, Montgomery. (1993). Isolation by Distance in Equilibrium and Non-Equilibrium Populations. *Evolution*, 47(1), 264–279. doi: [10.1111/j.1558-5646.1993.tb01215.x](https://doi.org/10.1111/j.1558-5646.1993.tb01215.x)
- Slatkin, Montgomery, & Excoffier, L. (2012). Serial Founder Effects During Range Expansion: A Spatial Analog of Genetic Drift. *Genetics*, 191(1), 171–181. doi: [10.1534/genetics.112.139022](https://doi.org/10.1534/genetics.112.139022)
- Smýkal, P., Hradilová, I., Trněný, O., Brus, J., Rathore, A., Bariotakis, M., ... Pirintsos, S. (2017). Genomic diversity and macroecology of the crop wild relatives of domesticated pea. *Scientific Reports*, 7(1), 17384. doi: [10.1038/s41598-017-17623-4](https://doi.org/10.1038/s41598-017-17623-4)
- Smýkal, P., Kenicer, G., Flavell, A. J., Corander, J., Kosterin, O., Redden, R. J., ... Ellis, N. T. H. (2011). Phylogeny, phylogeography and genetic diversity of the *Pisum* genus. *Plant Genetic Resources*, 9(1), 4–18. doi: [10.1017/S147926211000033X](https://doi.org/10.1017/S147926211000033X)
- Smýkal, P., Trněný, O., Brus, J., Hanáček, P., Rathore, A., Roma, R. D., ... Toker, C. (2018). Genetic structure of wild pea (*Pisum sativum* subsp. *Elatius*) populations in the northern part of the Fertile Crescent reflects moderate cross-pollination and strong effect of geographic but not environmental distance. *PLOS ONE*, 13(3), e0194056. doi: [10.1371/journal.pone.0194056](https://doi.org/10.1371/journal.pone.0194056)
- Smýkal, P., Vernoud, V., Blair, M. W., Soukup, A., & Thompson, R. D. (2014). The role of the testa during development and in establishment of dormancy of the legume seed. *Frontiers in Plant Science*, 5, 351–351. doi: [10.3389/fpls.2014.00351](https://doi.org/10.3389/fpls.2014.00351)
- Soria-Hernanz, D. F., Fiz-Palacios, O., Braverman, J. M., & Hamilton, M. B. (2008). Reconsidering the generation time hypothesis based on nuclear ribosomal ITS sequence comparisons in annual and perennial angiosperms. *BMC Evolutionary Biology*, 8(1), 344. doi: [10.1186/1471-2148-8-344](https://doi.org/10.1186/1471-2148-8-344)
- Speidel, L., Forest, M., Shi, S., & Myers, S. R. (2019). A method for genome-wide genealogy estimation for thousands of samples. *Nature Genetics*, 51(9), 1321–1329. doi: [10.1038/s41588-019-0484-x](https://doi.org/10.1038/s41588-019-0484-x)

- Stange, M., Barrett, R. D. H., & Hendry, A. P. (2021). The importance of genomic variation for biodiversity, ecosystems and people. *Nature Reviews Genetics*, 22(2), 89–105. doi: [10.1038/s41576-020-00288-7](https://doi.org/10.1038/s41576-020-00288-7)
- Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9(1), 307. doi: [10.1186/1471-2105-9-307](https://doi.org/10.1186/1471-2105-9-307)
- Szűcs, M., Vahsen, M. L., Melbourne, B. A., Hoover, C., Weiss-Lehman, C., & Hufbauer, R. A. (2017). Rapid adaptive evolution in novel environments acts as an architect of population range expansion. *Proceedings of the National Academy of Sciences*, 114(51), 13501–13506. doi: [10.1073/pnas.1712934114](https://doi.org/10.1073/pnas.1712934114)
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585–595.
- Tanto Hadado, T., Rau, D., Bitocchi, E., & Papa, R. (2010). Adaptation and diversity along an altitudinal gradient in Ethiopian barley (*Hordeum vulgare* L.) landraces revealed by molecular analysis. *BMC Plant Biology*, 10(1), 121. doi: [10.1186/1471-2229-10-121](https://doi.org/10.1186/1471-2229-10-121)
- Team, R. C. (2020). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria*, <https://www.R-project.org/>.
- Thomas, J. A., Welch, J. J., Lanfear, R., & Bromham, L. (2010). A Generation Time Effect on the Rate of Molecular Evolution in Invertebrates. *Molecular Biology and Evolution*, 27:5, 1173–1180. doi: [10.1093/molbev/msq009](https://doi.org/10.1093/molbev/msq009)
- Trněný, O., Brus, J., Hradilová, I., Rathore, A., Das, R., Kopecký, P., ... Smýkal, P. (2018). Molecular Evidence for Two Domestication Events in the Pea Crop. *Genes*, 9(11), 535. doi: [10.3390/genes9110535](https://doi.org/10.3390/genes9110535)
- Vershinin, A. V. (2003). Transposable Elements Reveal the Impact of Introgression Rather than Transposition, in Pisum Diversity, Evolution, and Domestication. *Molecular Biology and Evolution*, 20(12), 2067–2075. doi: [10.1093/molbev/msg220](https://doi.org/10.1093/molbev/msg220)
- Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, 23(23), 5649–5662. doi: [10.1111/mec.12938](https://doi.org/10.1111/mec.12938)
- Wang, N., Akey, J. M., Zhang, K., Chakraborty, R., & Jin, L. (2002). Distribution of Recombination Crossovers and the Origin of Haplotype Blocks: The Interplay of Population History, Recombination, and Mutation. *The American Journal of Human Genetics*, 71(5), 1227–1234. doi: [10.1086/344398](https://doi.org/10.1086/344398)
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, 38(6), 1358. doi: [10.2307/2408641](https://doi.org/10.2307/2408641)
- Wright, S. (1931). Evolution in Mendelian Populations. *Genetics*, 16(2), 97–159.
- Wright, S. (1943). Isolation by Distance. *Genetics*, 28(2), 114–138.
- Yin, X., Martinez, A. S., Sepúlveda, M. S., & Christie, M. R. (2021). Rapid genetic adaptation to recently colonized environments is driven by genes underlying life history traits. *BMC Genomics*, 22(1), 269. doi: [10.1186/s12864-021-07553-x](https://doi.org/10.1186/s12864-021-07553-x)
- Young, A., Boyle, T., & Brown, T. (1996). The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology & Evolution*, 11(10), 413–418. doi: [10.1016/0169-5347\(96\)10045-8](https://doi.org/10.1016/0169-5347(96)10045-8)

- Zaytseva, O. O., Bogdanova, V. S., & Kosterin, O. E. (2012). Phylogenetic reconstruction at the species and intraspecies levels in the genus *Pisum* (L.) (peas) using a histone H1 gene. *Gene*, 504(2), 192–202. doi: [10.1016/j.gene.2012.05.026](https://doi.org/10.1016/j.gene.2012.05.026)
- Zaytseva, O. O., Gunbin, K. V., Mglinets, A. V., & Kosterin, O. E. (2015). Divergence and population traits in evolution of the genus *Pisum* L. as reconstructed using genes of two histone H1 subtypes showing different phylogenetic resolution. *Gene*, 556(2), 235–244. doi: [10.1016/j.gene.2014.11.062](https://doi.org/10.1016/j.gene.2014.11.062)
- Zeller, K. A., McGarigal, K., & Whiteley, A. R. (2012). Estimating landscape resistance to movement: A review. *Landscape Ecology*, 27(6), 777–797. doi: [10.1007/s10980-012-9737-0](https://doi.org/10.1007/s10980-012-9737-0)