

RESEARCH ARTICLE

Wild Carrot Differentiation in Europe and Selection at *DcAOX1* Gene?

Tânia Nobre^{1*}, Manuela Oliveira², Birgit Arnholdt-Schmitt¹

1 EU Marie Curie Chair, ICAAM - Instituto de Ciências Agrárias e Ambientais Mediterrânicas, Universidade de Évora, Évora, Portugal, **2** Centro de Investigação em Matemática e Aplicações, Instituto de Investigação e Formação Avançada, Universidade de Évora, Évora, Portugal

* tnobre@uevora.pt



 OPEN ACCESS

Citation: Nobre T, Oliveira M, Arnholdt-Schmitt B (2016) Wild Carrot Differentiation in Europe and Selection at *DcAOX1* Gene? PLoS ONE 11(10): e0164872. doi:10.1371/journal.pone.0164872

Editor: Kenneth M Olsen, Washington University, UNITED STATES

Received: June 1, 2016

Accepted: October 3, 2016

Published: October 21, 2016

Copyright: © 2016 Nobre et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All sequences are available from the NCBI database (for *DcAOX1*, KX359844 - KX359963; and for D28n KX359964 - KX360018).

Funding: TN was supported by a Marie Curie Career Integration grant (FP7-PEOPLE-2012-CIG ref. 321725) and a Foundation for Science and Technology (FCT, Portugal) fellowship (SFRH/BCC/52187/2013). MO thanks FCT for the support given under the program UID/MAT/04674/2013 (CIMA). This work is funded by National Funds through FCT under the Project UID/AGR/00115/2013. The funders had no role in study design, data collection

Abstract

By definition, the domestication process leads to an overall reduction of crop genetic diversity. This led to the current search of genomic regions in wild crop relatives (CWR), an important task for modern carrot breeding. Nowadays massive sequencing possibilities can allow for discovery of novel genetic resources in wild populations, but this quest could be aided by the use of a surrogate gene (to first identify and prioritize novel wild populations for increased sequencing effort). Alternative oxidase (AOX) gene family seems to be linked to all kinds of abiotic and biotic stress reactions in various organisms and thus have the potential to be used in the identification of CWR hotspots of environment-adapted diversity. High variability of *DcAOX1* was found in populations of wild carrot sampled across a West-European environmental gradient. Even though no direct relation was found with the analyzed climatic conditions or with physical distance, population differentiation exists and results mainly from the polymorphisms associated with *DcAOX1* exon 1 and intron 1. The relatively high number of amino acid changes and the identification of several unusually variable positions (through a likelihood ratio test), suggests that *DcAOX1* gene might be under positive selection. However, if positive selection is considered, it only acts on some specific populations (i.e. is in the form of adaptive differences in different population locations) given the observed high genetic diversity. We were able to identify two populations with higher levels of differentiation which are promising as hot spots of specific functional diversity.

Introduction

Crop plants typically include only a portion of the genetic diversity of their wild relatives. Since genetic variation is the raw material of evolution, low genetic diversity has as direct consequence a reduction on the ability of the species to evolve in response to changes in its environment. If until recently we have perceived plant breeding as overall reducing crop genetic diversity, recent assessments have shown spatial and temporal patterns of genetic diversity losses (e.g. [1,2]). Particularly, breeding objectives have changed across time, from just yield

and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

improvement to also accommodate adaptation capacity and stress resistance under different agricultural systems and climatic conditions. This, in turn, has as consequence different selective pressures within breeding populations, and thus variable genetic diversity in released cultivars of a given crop [2]. Also, the extent of this loss of diversity depends on the population size during the domestication period, the mating system and the duration of that period, and it is not experienced equally by all genes in the genome [3]. Genetic diversity is lowered by intraspecific hybridization as well as by the selection process, which enhances genetic differentiation [4].

For many crops, like maize and cauliflower, this diversity loss due to domestication has made the plant totally dependent on humans in such a way that plant crop is no longer capable of propagating itself in nature [5]. In others however, such as in carrot (*Daucus carota* L.), the domestication process rendered more modest changes when compared to their progenitors—and they can even revert to the wild or become self-propagating weeds [5]. Even though there is clear evidence for diversification between wild and cultivated *D. carota* (e.g. [6,7]) and for the separation of the cultivated germplasm into two distinct groups (the Eastern–Asian and Western–European and American- gene pools [8]), widespread hybridization and introgression events have been reported (e.g. [9,10]). The outcrossing nature of carrot without any clonal propagation, associated with the fact that open-pollinated seed production was (likely) used to propagate carrot during domestication, lead to a small reduction in the genetic diversity of cultivated vs. wild carrot [6]. Nonetheless, Grzebelus and co-workers [11] encountered an overall higher gene diversity of wild accessions. They found private markers not present in any gene pool of cultivated carrot, which encourages the search of genomic regions potentially important for modern carrot breeding.

Close relatives of domesticated plants -crop wild relatives (CWRs)- represent a practical gene pool that can be exploited by plant breeders in the quest to address modern agricultural needs: higher productivity, climate resilience and nutritional security (reviewed in [12]). In the case of carrot, and because primary CWR are inter-fertile with the crop species, this genetic variability is easily accessible for plant improvement. To access this genetic pool, there is a need to screen wild populations. This screen can be done randomly, but can potentially be made more efficient by making use of climatic information, to help in selecting the most appropriate populations for further testing. This assumes that CWRs are adapted to their environment and thus present a set of traits involved in adaptation. Usually this means that a few genes of large effect should account for a relatively large proportion of the genetic differentiation between adapted populations together with many loci of smaller effect [13,14]. Nowadays massive sequencing possibilities can allow for discovery of novel genetic resources in wild populations, and comparison of genome variation in contrasting environments may bring new options for use of genetic variability in plant breeding for climate resilience [15]. An alternative option to massive sequencing would be the use of a surrogate gene(s) to first identify and prioritize novel wild populations for increased sequencing effort. In this case, a surrogate gene would be a proxy for the broader genome (as this is too big to allow to be considered individually for a greater number of populations) given a certain trait, for example, specificities of the environment of the adapted population. Such a surrogate gene would have to show a set of polymorphisms whose presence would link to particular environmental conditions, and indicate the presence of wider genetic variability associated with evolution of adaptation.

The adaptation to different environments involves adaptation to biotic and abiotic stresses, and the concomitant variation in CWR genes could be explored as a way to increase tolerance and hence improve long term crop productivity. Alternative oxidase (AOX) gene family seems to be linked to all kinds of abiotic and biotic stress reactions in various organisms (e.g. [16–18]) and have the potential to be used in a model towards identification of CWR hotspots of

environment-adapted diversity. The choice of this gene relates to the fact that AOX is not only part of the stress response in plants, but it also plays a central role in defining the stress response [19]. By this reason, AOX was previously proposed as a source for functional markers for breeding [20]. The rendering proteins are active in mitochondria, organelles of crucial importance for environmental stress perception and stress signal transduction. This gene family is present in the respiratory chains of all plants, as well as in certain fungi, protists, animals and bacteria, and it is crucially involved in the adaptive regulation of metabolism. Within the same species, individual genotypes and/or groups of genotypes can be distinguished by polymorphic AOX gene family sequences (e.g. [20–23]) and this is also true for carrot (e.g. [24–27]). In carrot, this multigene family is encoded by three genes distributed in two discrete gene subfamilies—*DcAOX1* and *DcAOX2* (a and b). Generally, and while AOX1 is induced by stress stimuli, AOX2 is referred as constitutive or developmental. Even though this paradigm begins to be challenged [28–30] the current view is still that the subfamilies have different physiological roles. They are thus expected to have evolved under different selection pressures, being AOX1 gene subfamily likely most responsive to environmental stimuli. Campos and co-workers [31] showed that *DcAOX1* gene expression respond to different growth temperature conditions, and that this response was genotype dependent. In carrot, the complete gene shows different lengths and the less typical AOX structure of three exons interrupted by two introns, with the highest sequence variability found on intron 1 (including a hyper- variable region) followed far by exon 1 [27].

In this study we look into the variability of *DcAOX1* in populations of carrot CWR in Europe, subjected to different climatic stress and thus putatively adapted to different environments. By scanning wild crop relatives with a stress related gene, we expect to highlight hot spots of specific functional diversity.

Material and Methods

Plant material

Sampling was performed following an environmental gradient across Europe. This gradient accommodated sampling points with deviating climatic conditions, such as in Sierra de Guadarrama (considerable temperature changes between summer and winter and a very dry summer; wild carrots could not be found above 1100 m) or in central Pyrenees and the French Massif Central (with a cold continental climate at equivalent height). In general, the sampling was made on easily accessible non-cultivated fields (thus close to a road), with altitudes ranging from the referred 1100 m to sea level. In total, 13 sites were sampled (Table 1) by collecting wild carrot roots at each location. The samples were dried in silica gel and stored at -80°C. Due to the hard and woody characteristics of the wild carrot roots, an adaptation of the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) standard extraction protocol was used to extract the DNA: 1) extra initial grinding step (with liquid nitrogen and using a tissue grinder) to further pulverize the hard root tissues; 2) addition of polyvinylpyrrolidone (PVP, 10 000 mol wt at 3%) to the extraction buffer to remove phenolic and other compounds that can inhibit PCR and 3) lysis was performed overnight at 60 rpm. DNA concentration of all samples was determined with the NanoDrop-2000C spectrophotometer.

Data collection

The available climatic data of the last 15 years was collected for the closest by station of the sampling points (<http://globalweather.tamu.edu/>). Month averages were obtained to have a workable characterization of the long-term climatic conditions at the sampling locations. A Principal Component Analysis (PCA) was used to define patterns of climatic conditions of the

Table 1. Sampling locations and geographic coordinates of collection sites.

Population code	Name of closest location	Country code	Coordinates		
			Longitude	Latitude	Altitude
1	Guadalupe	PT	38.5825	-7.9980	310
3	Sant Carles de la Rapita	SP	40.6267	0.6529	0
4	Rostock	DE	54.0779	12.1126	24
5	Romangordo	SP	39.7795	-5.6966	327
6	Rascafria	SP	40.8840	-3.8867	1190
7	Torla-Huerdesa	SP	42.6511	-0.1380	1279
8	Gruissan	FR	43.1010	3.1201	84
9	Saint-Privat-d'Allier	FR	45.0012	3.6850	1009
10	Agencourt	FR	47.1324	4.9836	281
11	Kapellen	LU	49.6401	6.0184	362
12	Nijkerk	NL	52.2154	5.4779	2
13	Capbreton	FR	43.6307	-1.4322	64
14	Torquemada	SP	42.0364	-4.2824	785

doi:10.1371/journal.pone.0164872.t001

sampling locations. Data was grouped into clusters for data summarization, through an Agglomerative Hierarchical Clustering (Ward agglomeration method, on Euclidian distances for five classes). These analyses were performed using XLSTAT-ecology, an add-on to MS Excel[®].

For the isolation of AOX1 amplicon in wild carrots, specific primers were designed based on the already available cDNA sequence at the NCBI GenBank (EU286573.2). A nested PCR approach was selected. For the first reaction, the primers used were located at the beginning of exon 1 (*DcAOX1*_24Fw or *DcAOX1*_94Fw) as forward and at the end of exon 3 as reverse (*DcAOX1*_1032Rev; more details in [27]). A standard 25 ul reaction, at 2.5 mM of magnesium and BSA at 0.4 ug/ul, was run with an annealing temperature of 55°C. The PCR product was then diluted in 1:50 and used as template for the second reaction. The second reaction was performed with the same forward primer and the degenerated primer P2 [32], in a 25 ul reaction at 1.5 mM of magnesium. Annealing temperature was 60°C. Amplicons were purified from the agarose gel with GFX PCR DNA and Gel Band Purification Kit, directly cloned into pGEM[®]-T Easy vector (Promega, Madison, WI, USA), transformed into bacterial strain JM109 (Promega, USA) and bacterial colonies were tested using T7 and SP6 primers. Sequence was done from the PCR product, in sense and antisense strands. Because carrot is an outcrossing species, a minimum of four plants per population and two clones per plant were sequenced (S1 Table). For the amplification of D28n marker (for taxonomic certainty), we followed the procedure described by Spooner and co-workers [33], in a standard 25 ul reaction at 1.5 mM of magnesium and an annealing temperature of 55°C. Sequence was done directly from the PCR product, in sense and antisense strands.

Phylogenetic analyses

Sequence visualization was performed in CLC Main Workbench vs 6.8.1 software and, for AOX1 amplicon, exons and intron regions were identified on the sequences and aligned separately. The alignment of segments of the exon regions was relatively straightforward, with few insertions/deletions needing to be inferred. For the intron region an iterative refinement method, which accounts for larger gaps in the sequences (E-INS-i) as implemented in the program MAFFT [34], was used to align the segment. Alignment of D28n amplicon was made under the option L-INS-I (iterative refinement method incorporating local pairwise

alignments; gap opening penalty: 1.5 and gap extension penalty 0.14; 1PAM/k $\frac{1}{4}$ 2 scoring matrix for nucleotide sequences). The alignment of D28n fragment was straightforward, as no insertions/deletions had to be inferred. The optimal models of evolution were tested independently of the sequence region in MrModeltest 2.2 [35] and selection took place on the basis of the BIC scores (Bayesian Information Criterion; the lowest the value the better the substitution pattern).

Bayesian inference was conducted using MrBayes version 3.0 [36,37]. The default settings of MrBayes were used and with MCMC (considering 100 000 generations) runs being repeated three times as a safeguard against spurious results. The first 1 000 trees were discarded as burn-in, and the remaining trees were used to calculate a majority rule consensus tree. Stationarity was confirmed by analysis of the log-likelihoods and the consistency between runs. For AOX1, the same analysis was also done considering the full fragment, the exons alone and also considering the full fragment excluding the intron insertions suspected to be the result of introgression events (S1, S2 and S3 Figs).

AOX1 variability

The AOX1 sequences alignment was used to locate polymorphic positions and ambiguities encountered were, whenever possible, resolved by resequencing. Summary statistics and tests of neutrality were calculated with DnaSPv.4.0 [38] on the basis of the number of segregating sites (since we observed three different bases per site in some populations, we also performed the analyses on the basis of the total number of mutations (h) and obtained qualitatively similar results; data not shown). Tajima's D statistics compares the average number of pairwise differences with the number of segregating sites [39]. Over the all sequenced AOX1 fragment, linkage disequilibrium was measured using the ZnS statistic (the squared allele frequency correlation r^2 , [40]) on the basis of the parsimony informative sites. Statistical significance for ZnS and Tajima's D was assessed by coalescent simulations with 10 000 replicates as implemented in DnaSP v. 4.0, conducted considering all segregating sites and an intermediate level of recombination [38]. In addition, for the coding regions of *DcAOX1*, tests for positive selection were performed using the maximum likelihood methods implemented in the CODEML program of PAML [41]. The dN/dS ratio (ω) was calculated using models M0 (one-ratio) and M3 (discrete), and M1a (nearly neutral) and M2a (positive selection). The relevant likelihood ratio tests were performed to assess significance: M0-M3 tests for variable ω among sites and M1a-M2a tests positive selection.

We determined pairwise F_{ST} values among wild *D. carota* populations using Arlequin3.5.22 [42]; levels of significance were assessed on the basis of 10 000 permutations. The significance of isolation-by-distance was tested with a Mantel test.

Results

Based on the D28n phylogeny, all plants were confirmed belonging to *Daucus carota*. Within *D. carota* the phylogeny is highly unresolved (Fig 1), with the exception of a cluster comprising the four most western Iberian populations (Fig 1, red box). The reconstructed AOX1 phylogeny (Fig 2) however, does not show any obvious population clade or relates to a clear geographical origin. Through the use of a PCA (Fig 3a and 3b), the set of climatic observations of possible correlated variables were converted into a set of values of uncorrelated variables (the principal components). The percentage of the variance explained by the two first components is of 86.8% (70.4% for PC1 and 16.4% for PC2). The PCA representation permits distinguishing temperatures (max and min) and also solar radiation, from the variables precipitation, humidity and wind. In the components space, population 7 sampling site is differentiated from

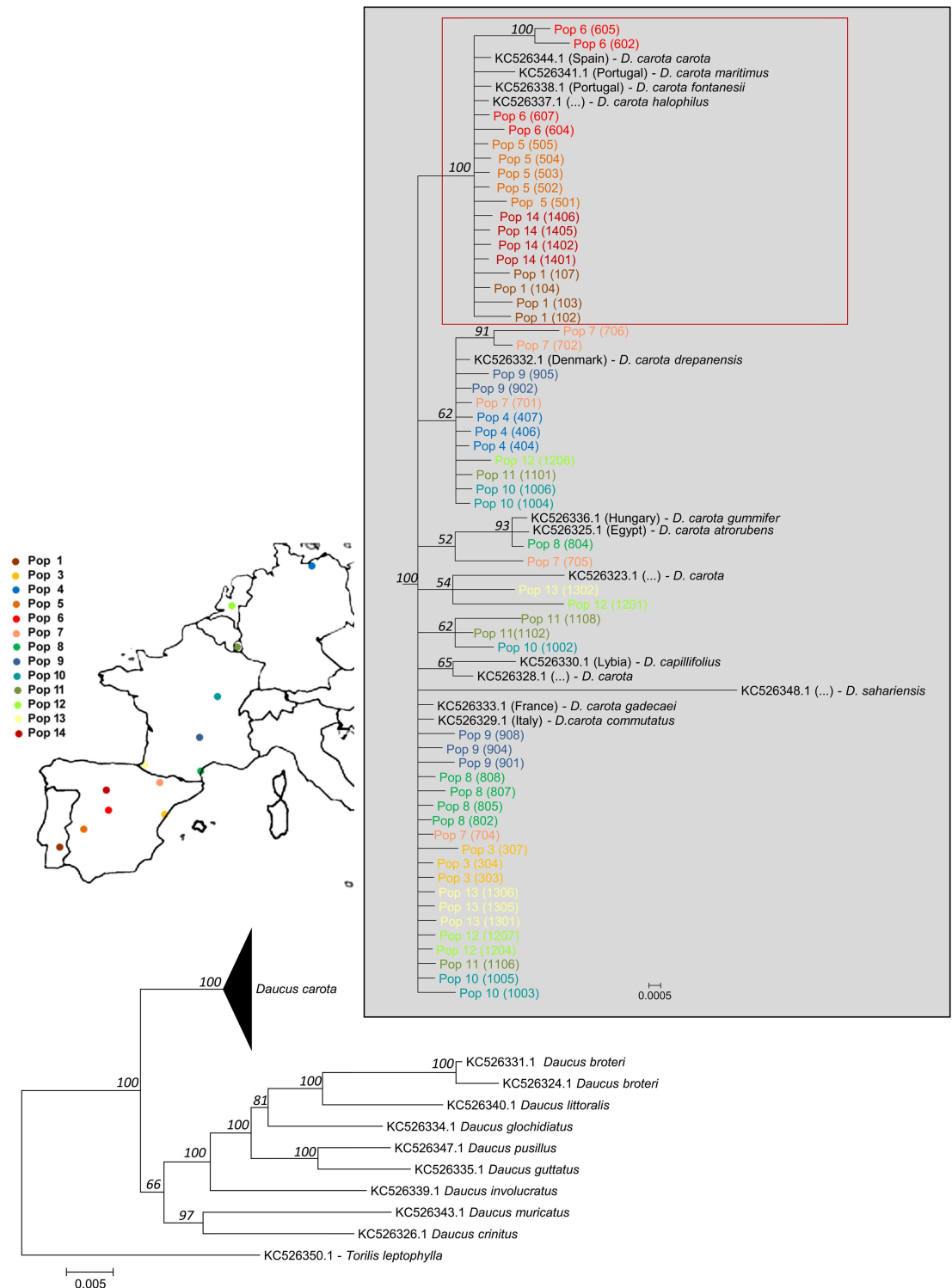


Fig 1. Reconstructed phylogeny based on de conserved orthologue marker D28n [33]. The phylogeny corresponds to the majority rule consensus tree of trees sampled in a Bayesian analysis (K2P+G as substitution model). *Torilis leptophylla* was used as outgroup. The numbers above the branches refer to the Bayesian posterior probability of the nodes (more than 50%) derived from 19500 Markov chain Monte Carlo-sampled trees. The red box encompasses a clade of western Iberian populations.

doi:10.1371/journal.pone.0164872.g001

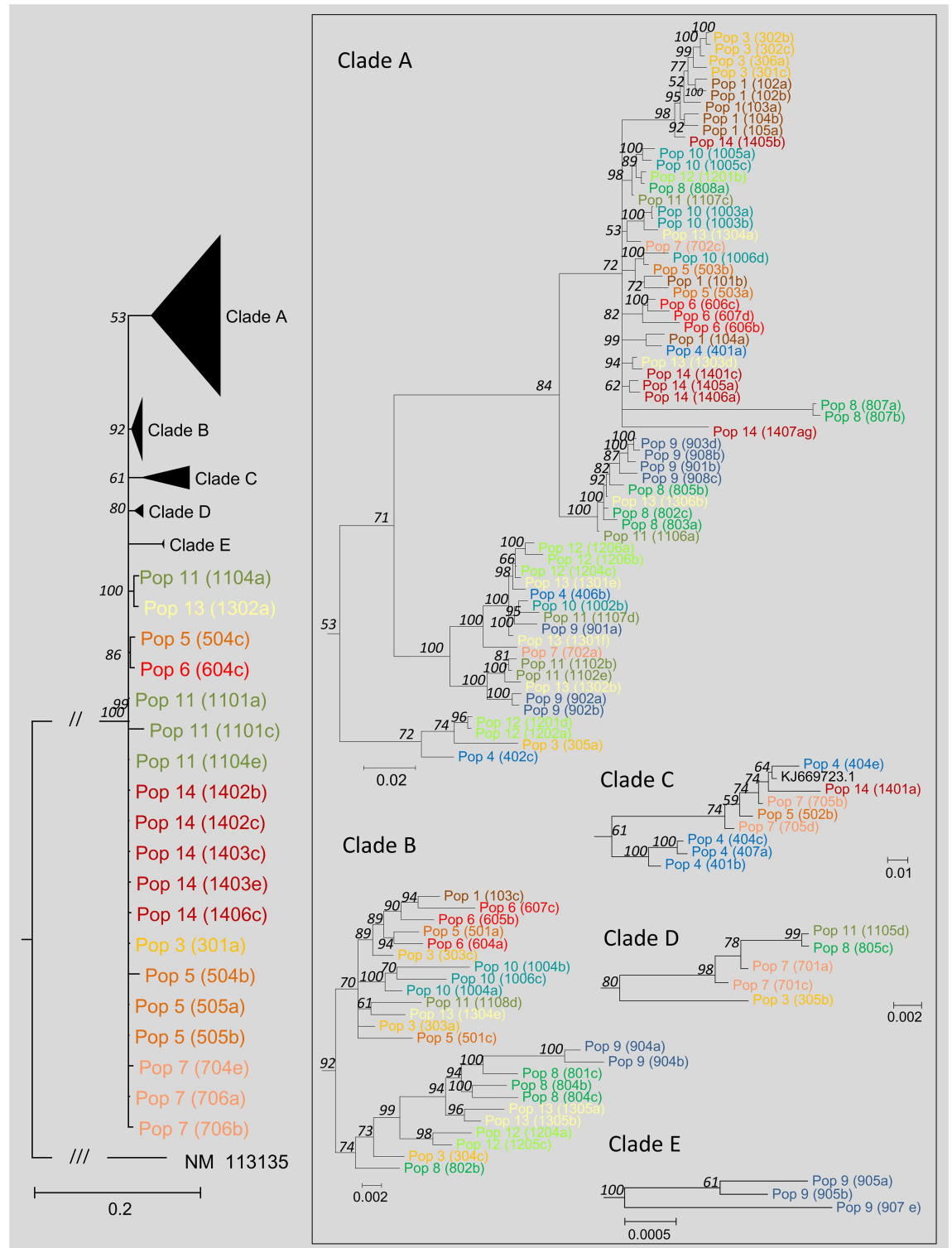


Fig 2. Reconstructed phylogeny based on the AOX1 fragment (2353 bp in the alignment). DNA models of evolution were tested independently of the sequence region, and selection took place on the basis of the BIC scores (Bayesian Information Criterion; the lowest the value the better the substitution pattern): Exon 1—K2+G+I; Intron 1—GTR+G and Exon 2—GTR+G. The phylogeny corresponds to the majority rule consensus tree of trees sampled in a Bayesian analysis. *Arabidopsis thaliana* was used as outgroup. The numbers above the branches refer to the Bayesian posterior probability of the nodes (more than 50%) derived from 19500 Markov chain Monte Carlo-sampled trees.

doi:10.1371/journal.pone.0164872.g002

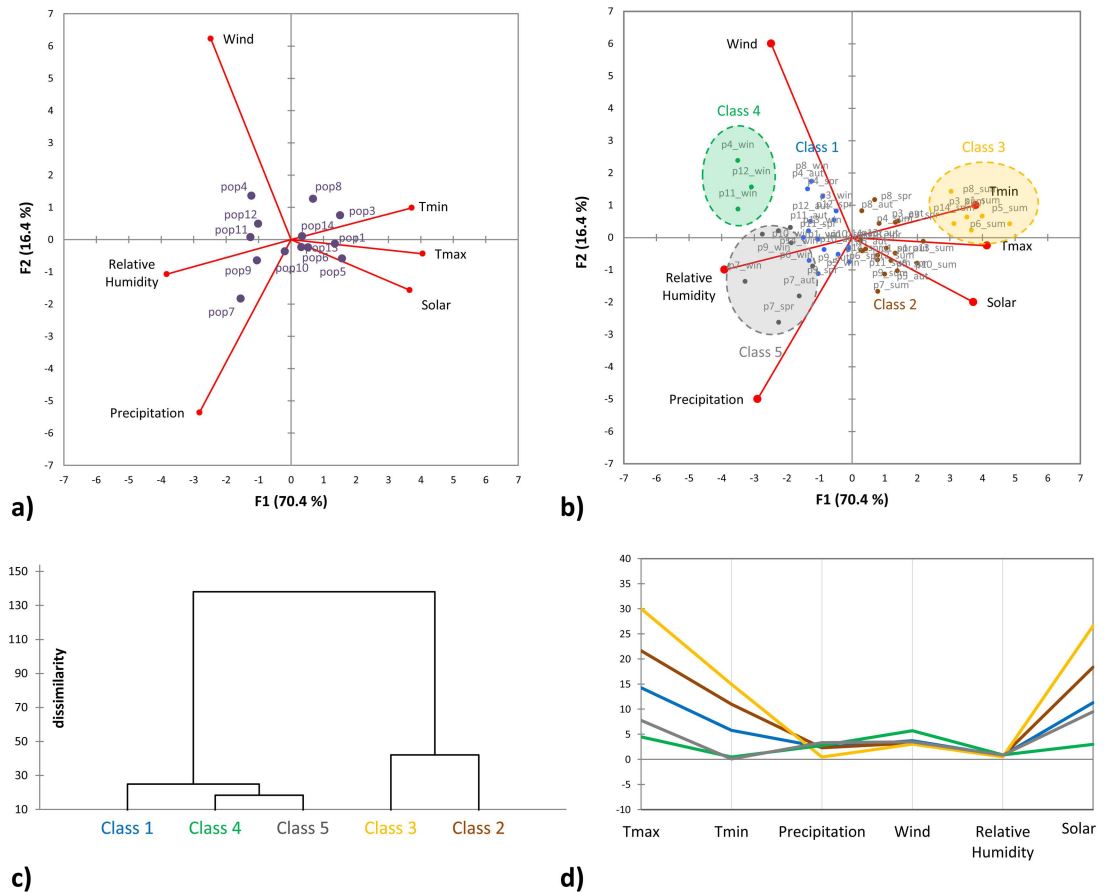


Fig 3. Sampling locations analysis based on average monthly weather data. A plot on the two main components of a Principal Component Analysis (86.9%): a) variables yearly averaged per population as additional category data; b) variables grouped by meteorological season. Agglomerative Hierarchical Clustering analysis: c) euclidian distance dendrogram for 5 classes; d) profile of the classes. tmin, minimum temperature ($^{\circ}\text{C}$); tmax, maximum temperature ($^{\circ}\text{C}$); precipitation (mm); relative humidity (%); wind (m/s); solar (MJ/m^2).

doi:10.1371/journal.pone.0164872.g003

the rest and the sampling sites of populations 4, 12 and 11 are similar for the analyzed variables (grouped based mainly on winter conditions, Fig 3b). The cluster analysis into five groups (Fig 3c and 3d) conforms with the PCA, dividing the locations into two highly differentiated clusters: a) high temperatures and solar radiation -Class 2 and Class 3 (grouping populations 1, 3, 5, 6, 8 and 14 summer characteristics); b) stronger winds and higher moisture -Class 1 and Class 4 (comprising populations 4, 11 and 12 winter conditions). Even though it is clear that the sampling locations show diversified climatic conditions of temperature, precipitation, humidity, solar radiation and even wind, no relation is apparent of any of these variables to the clades formed in the phylogenetic analysis of AOX1 gene fragment.

DcAOX1 variability in the CWR

The obtained 122 sequences were annotated for exonic regions and those were analyzed separately from the intron. Overall, the data show much higher gene diversity than initially expected: the amplicon size varies considerably due to intron 1 size, which ranged from 325 bp to 951 bp. In the current dataset, comprising a large West-European sampling of wild carrots, we found 9 haplotypes with an insertion ranging 200 bp to 252 bp in the beginning of the

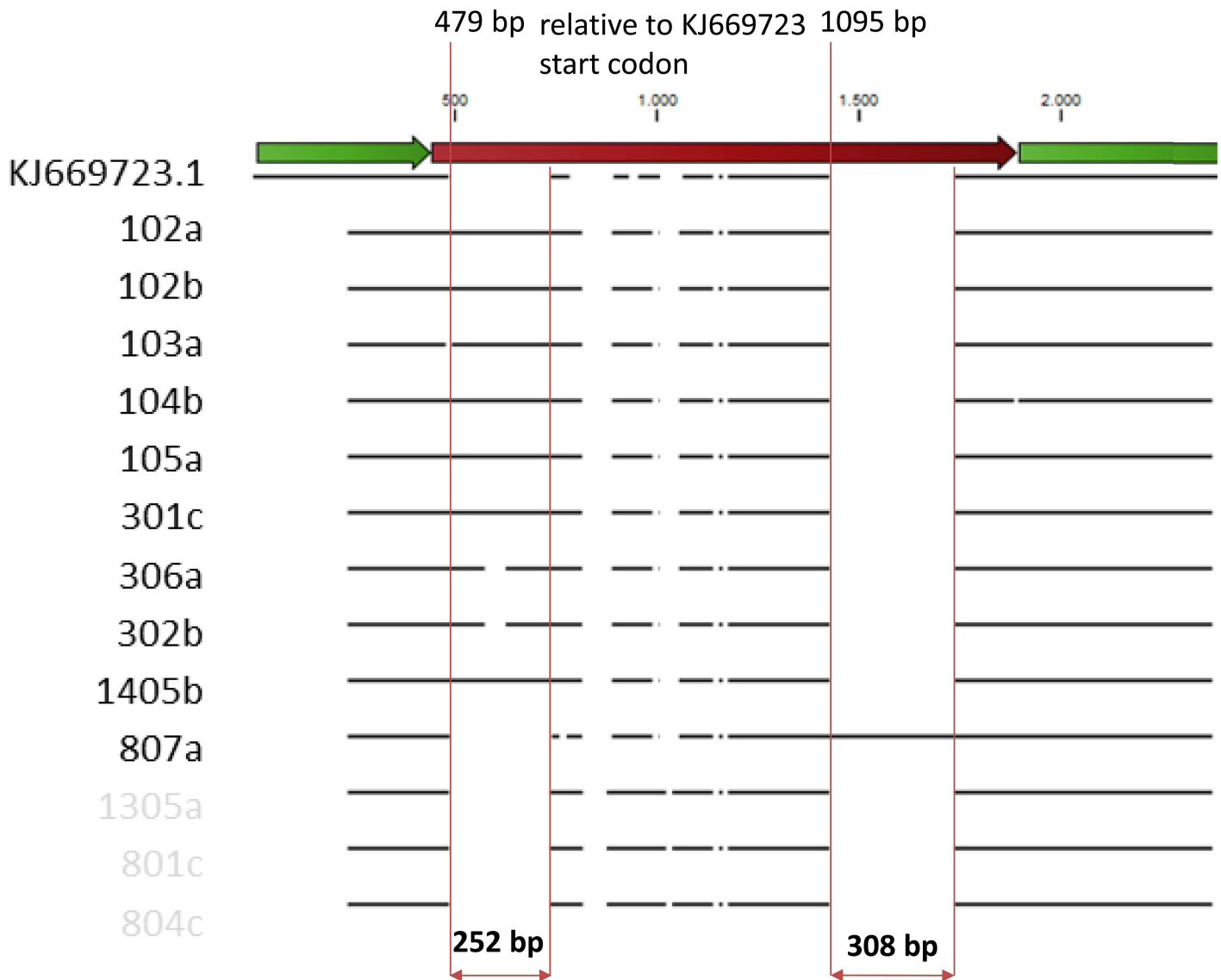


Fig 4. Diagram showing the long intron insertions relative to a known full DcAOX1 gene sequence (KJ669723.1).

doi:10.1371/journal.pone.0164872.g004

intron 1 (479 bp relative to KJ669723.1 start codon; Fig 4). This insertion was only present in individuals from three Iberian populations:

1. population 1–5 of the 8 haplotypes show here an insertion (I1a) which shows levels of similarity between themselves higher than 90%;
2. population 3–3 of the 8 haplotypes have here an insertion, being one of them equivalent to I1a (average similarity higher than 90%) whereas the two other haplotypes present a different insertion (I1b) but similar between themselves (99%);
3. population 14—only 1 of the 8 haplotypes present an insertion with 96% similarity with I1a.

I1a blasts primarily with *Rhizophagus intraradices* clone JGIBTPH-93C11 (AC237375, S2 Table). The deposited sequence is originated from a genome sequencing project of this

Table 2. Sequence variability analyses of a *DcAOX1* fragment covering part of gene exon 1 and exon 2 in 122 wild carrot haplotypes.

	Exon 1	Exon 2
Number of sequences	119	122
Number of sites (bp)	207	479
Total number of sites (excl. missing data)	205	474
Number of haplotypes	71	94
Polymorphic sites (S)	46	55
Parsimony informative	23	34
Total number of mutations	49	56
Synonymous changes	25	31
Non-Synonymous	24	25
Haplotype diversity (Hd)	0.979	0.995
Nucleotide diversity (Pi)	0.029	0.015

doi:10.1371/journal.pone.0164872.t002

arbuscular mycorrhiza fungus, where the fungus culture was root culture of carrot [43]. This means that the database hits to *R. intraradices* are most likely an artifact of contamination of the fungus DNA with co-cultured carrot cell DNA. All subsequent homologies found were with carrot samples (S2 Table) albeit not annotated as belonging to the AOX gene family. In only one individual, from population 8, a second insertion was found towards the end of the intron (1095 bp relative to KJ669723.1 start codon; Fig 4). This insertion of 308 bp shows high homology with genomic sequence of either *Daucus sahariensis* (KJ519787.1, KJ519789.1 and KJ519806.1) or *Daucus syrticus* (KJ519808.1 and KJ519807.1).

From exon1 the dataset is not complete and from the expected fragment of 432 sites, only 205 sites have no alignment gaps or missing data (analyzed only for 119 sequences due to absence of data on 302c, 402c and 807b). Data on the exons (number of haplotypes, polymorphic sites and diversity estimates) is summarized in Table 2. Both exon 1 and 2 exhibit high level of nucleotide diversity, showing also almost equivalent numbers of non-synonymous and synonymous sites. For haplotype and nucleotide diversity estimates (Hd and Pi), we chose to analyze synonymous and nonsynonymous sites jointly because the amount of coding sequence was limited and treating them separately would have led to estimates based on even a smaller number of sites [44].

Population genetics inferences are primarily based on two sources of information: the site frequency spectrum of mutations (SFS) -being Tajima's *D* one of the most popular summary statistics- and the statistical association among those, that is, linkage disequilibrium (LD). Considering an intermediate level of recombination (gene recombination parameter *R* determined to be 18.60) the observed Tajima's *D* value (-1.32) was significant ($p = 0.01$; [-1.19, 1.01] 95% confidence interval). The observed *ZnS* value (0.08) was not significant ($p = 0.87$; [0.04, 0.10] 95% confidence interval).

Likelihood ratio tests (LRTs) revealed that PAML models that allowed for adaptive positive selection fitted the exon 1 sequence data better than those which did not; this was, however, not true for exon 2 (Table 3) although models M3 fit the data significantly better than the null models M0. A total of 12 sites were identified as being under positive selection ($\omega > 1$ for $\alpha = 0.05$).

Population differentiation based on AOX1

Fig 5 shows population relationships as described by F_{ST} computed between pairs of populations (measure of population differentiation due to genetic structure, ranging from 0 to 1; with zero

Table 3. Parameter estimates and likelihood scores under models of variable ω ratios among sites for exon1 and exon2 of the *DcAOX1* gene.

	Nested model pairs	Parameter estimates	-lnL	LRT	PSS
Exon1	M0: one ratio	$\omega = 0.125$	3432.757	438.092 ($p = 0.000$)	na
	M3: discrete	$\omega_0 = 0.037; \omega_1 = 1.014; \omega_2 = 5.925; P_0 = 0.885; P_1 = 0.100; P_2 = 0.015$	3213.711		15, 16, 30, 51, 54, 60, 63, 64, 66, 67, 68, 74
	M1a: nearly neutral	$\omega_0 = 0.034; \omega_1 = 1.000 P_0 = 0.886; P_1 = 0.114$	3236.036	44.644 ($p = 0.000$)	na
	M2a: positive selection	$\omega_0 = 0.0346; \omega_1 = 1.000; \omega_2 = 5.907 P_0 = 0.884; P_1 = 0.101; P_2 = 0.015;$	3213.714		54, 66
Exon2	M0: one ratio	$\omega = 0.053$	2444.668	11.115 ($p = 0.029$)	na
	M3: discrete	$\omega_0 = 0.020; \omega_1 = 0.020; \omega_2 = 0.306; P_0 = 0.239; P_1 = 0.632; P_2 = 0.128$	2439.110		
	M1a: nearly neutral	$\omega_0 = 0.036; \omega_1 = 1.000 P_0 = 0.968; P_1 = 0.032$	2441.611	0.000 ($p = 1.000$)	na
	M2a: positive selection	$\omega_0 = 0.036; \omega_1 = 1.000; \omega_2 = 1.000 P_0 = 0.968; P_1 = 0.018; P_2 = 0.014;$	2441.611		na

doi:10.1371/journal.pone.0164872.t003

value indicating no population structuring or subdivision and one indicating that all genetic variation at the analyzed markers can be explained by population structure). It shows that variability at exon 1 (Fig 5c) is the main responsible for differentiation, except for the case of population 1 where the indel at intron 1 (presented above; Fig 5b) characterizes the population. No isolation

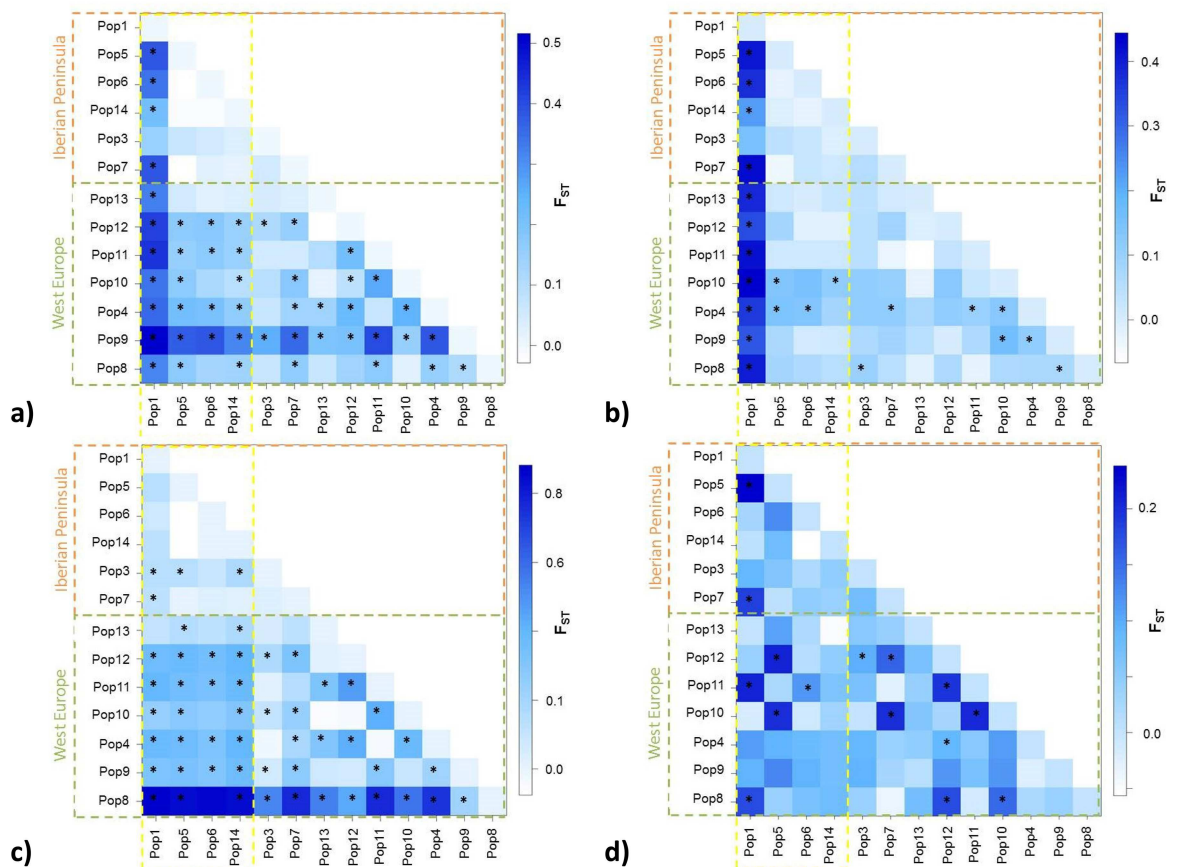


Fig 5. Representation of pairwise F_{ST} values among wild *D. carota* populations AOX1 fragment, with significance at 0.05 being highlighted with a *; a) complete fragment; b) intron1 only; c) exon1; and d) exon 2. Graphs present different y scale to highlight differences within data.

doi:10.1371/journal.pone.0164872.g005

by distance was observed (Mantel test considering the entire sequenced fragment, $p = 0.31$; only intron 1, $p = 0.12$; only exon 1, $p = 0.65$; only exon 2, $p = 0.77$).

Discussion

Daucus carota is a highly diverse group, which certainly contributes to the poorly developed (or even lack of) barriers for interbreeding among either CWR or domesticated forms. The reconstructed phylogeny of carrot CWR (based on the conserved orthologue marker D28n) suggests a Eurosiberian (Boreal)–Mediterranean division, with the exception of the Miranda del Ebro population (population 3). This population, even though still in a Mediterranean climate (Csa, according to Köppen-Geiger climate classification system), grows in an environment close to an oceanic climate zone (Cfb) much more similar, in overall weather conditions, to other more central European locations. The reconstructed phylogeny based on the analyzed *DcAOX1* gene fragment does not seem to identify any clear clade that could directly be linked to the population origin. The observed structure in the reconstructed tree is mainly due to the intron region, as the one obtained when using the exons only is generally unresolved (S2 Fig).

There is a lack of structure among the *DcAOX1* genetic pool within the Iberian Peninsula (Fig 4), and these populations seem to be differentiated from the ones of the rest of the Western Europe, consistent with the information obtained based on the D28n marker. The overall level of population differentiation is relatively low, but population 9 (Saint-Privat-d'Allier, France) and population 1 (Guadalupe, Portugal) are highlighted as potentially interesting in terms of differentiation, which results mainly from the polymorphisms associated with exon 1 and intron 1. On the contrary to what was observed with CWR of lettuce in Europe for example [45], no correlation between geographic and genetic distances was found for wild carrot based on the studied *DcAOX* gene sequences. The outcrossing nature of carrot (contrasting with the selfing habits of lettuce) increases the probability of long-distance gene flow, and hence the absence of a distance effect. The level of genetic differentiation depends on gene flow and genetic drift, so the lack of genetic differentiation among Iberian populations likely results from high rates of historical gene flow between populations and/or large effective population sizes. The intron 1 of *DcAOX1* is a lengthy variable region, including even a hyper-variable region of simple sequence repeats (SSRs) [27], and shows in the present study two large insertions (> 200 bp).

A sole wild carrot specimen from population 8 (Port-la-Nouvelle, France) showed an insertion of 308 bp at the end of intron 1 with high identity with clones of *Daucus sahariensis* and *Daucus syrticus*. Arbizu and co-workers [10] found that, even with combined molecular and morphological studies, there are particular problems in distinguishing these two species. *Daucus syrticus*, and *D. sahariensis* together with *D. gracilis*, are probably the most closely related plants to *D. carota* [46]. This insertion can either be vertically inherited from a common ancestor, or the result of a hybridization event between wild *D. carota* and *D. sahariensis* or *D. syrticus* (that even though native in North Africa could have been locally introduced in France). Remains to be tested whether the high variation at the intron, has an effect on gene expression and on the functionality of the encoded alternative oxidase protein. Introns can have large influence in the control of gene expression in plants (e.g. [47,48]). Particularly, introns at the proximity of the 5' end of a gene are of relevance, as they can affect the binding of transcription factors [49], the process of alternative splicing [50], the coding of intronic regulatory elements [51] and also nonsense-mediated mRNA decay [52].

Nucleotide diversity at the *DcAOX1* fragment of carrot CWR was high (higher at exon 1 than exon 2), no linkage disequilibrium was found and the significant negative value of Tajima's D suggests either a recent selective sweep (or linkage to it) or a recent population expansion

following a bottleneck (the current data, based on one gene only, does not allow to distinguish between both events).

The likelihood ratio test identified several positions at exon 1 which are unusually variable (indication of positive selection), and a relatively high number of amino acid changes is observed, together indicating that *DcAOX1* gene might be under positive selection. However, the high genetic diversity, including several indels, suggests that, if there is positive selection, it only acts on some specific populations (i.e. is in the form of adaptive differences in different population locations). The hypothesis that balanced selection is driving the high genetic diversity at *DcAOX1* gene cannot be excluded.

Additionally, the observed high diversity suggests that the *DcAOX1* gene is not present in a single copy. The adaptive role of copy number variation (CNV) is suspected to be of high relevance, and for specific genes has been linked to important traits such as flowering time, plant height and resistance to biotic and abiotic stress (reviewed in [53]). Knowledge on the extent of CNV of the *DcAOX1* gene in natural populations will elucidate on its adaptive role.

Exon 1 in the *DcAOX1* gene seems to result from a fusion of two exons (via loss of an intron; [31]) and this intron loss might have been adaptive. It has been suggested the existence of high selection force against introns in rapidly regulated genes, with the rationality that an intron-less allele will produce its protein product more rapidly than a corresponding intron-containing allele because splicing is relatively slow compared with transcription [54]. Nonetheless, intron loss-gain seems to be a dynamic process difficult to generalize, as the importance of several introns for the functions of the respective gene might imply that, in some cases, intron insertion is favored by natural selection so that evolutionary conserved genes may accumulate introns [55].

It has been proposed that positive selection promotes the functional divergence of gene family members encoding enzymes involved in secondary metabolism as its products are thought to be a response to challenges imposed by the environment ([e.g. [55,56,57]). In plants, the AOX is present as a small multigene family in individual species with the overall proposed role in regulation of growth rate homeostasis under various environmental conditions that a plant undergoes. Umbach and co-workers [58] showed that alterations in the AOX pathway provoked changes that were largely chloroplast and carbohydrate metabolism related, and not only moderating ROS, thus contributing to the accumulation of secondary metabolites [59]. Therefore, the signatures suggesting local positive selection, the indications of high CNV and the *a priori* knowledge of the AOX1 gene involvement in homeostasis under stress conditions, calls for further characterization.

If, with the present data, AOX does not seem to work as a surrogate for diversity directly linked to the climatic conditions analyzed, we were able to clearly identify two populations with higher levels of differentiation which are promising as hot spots of specific functional diversity. These two populations are thus good targets for a wider approach, either considering several candidate genes or a genome-wide approach, towards the identification of novel genetic resources relevant for modern carrot breeding. Enlarging this study to a wider geo-climatic region, including the center of domestication of carrot, has the potential of identifying further genetic diversity hotspots, particularly if combined with a more fine scale environmental analysis.

Supporting Information

S1 Fig. Reconstructed phylogeny based on a AOX1 fragment. The phylogeny corresponds to the majority rule consensus tree of trees sampled in a Bayesian analysis. Two insertions at the intron were removed. *Arabidopsis thaliana* was used as outgroup. The numbers above the branches refer to the Bayesian posterior probability of the nodes (more than 50%) derived

from 19500 Markov chain Monte Carlo-sampled trees.
(TIF)

S2 Fig. Reconstructed phylogeny based on a AOX1 fragment. The phylogeny corresponds to the majority rule consensus tree of trees sampled in a Bayesian analysis. Only fragments in exons were considered. *Arabidopsis thaliana* was used as outgroup. The numbers above the branches refer to the Bayesian posterior probability of the nodes (more than 50%) derived from 19500 Markov chain Monte Carlo-sampled trees.
(TIF)

S3 Fig. Reconstructed phylogeny based on a AOX1 fragment. The phylogeny corresponds to the majority rule consensus tree of trees sampled in a Bayesian analysis. Only intron 1 was considered. *Arabidopsis thaliana* was used as outgroup. The numbers above the branches refer to the Bayesian posterior probability of the nodes (more than 50%) derived from 19500 Markov chain Monte Carlo-sampled trees.
(TIF)

S1 Table. Sample locations, geographic coordinates, populations and individual plants codes.
(PDF)

S2 Table. Intron 1 insertions and homologies according to NCBI.
(PDF)

Acknowledgments

The authors would like to thank the reviewers for the valuable comments, to Vera Valadas for the help with the laboratory work and to Amaia Nogales for constructive discussions during the preparation of this manuscript.

Author Contributions

Conceptualization: TN BAS.

Data curation: TN.

Formal analysis: TN MO.

Funding acquisition: TN BAS.

Investigation: TN.

Methodology: TN.

Resources: TN BAS.

Visualization: TN.

Writing – original draft: TN.

Writing – review & editing: TN.

References

1. van de Wouw M, van Hintum T, Kik C, van Treuren R, Visser B. Genetic diversity trends in twentieth century crop cultivars: A meta analysis. *Theor Appl Genet.* 2010; 120: 1241–1252. doi: [10.1007/s00122-009-1252-6](https://doi.org/10.1007/s00122-009-1252-6) PMID: 20054521

2. Rauf S, da Silva JDAT, Khan A. Consequences of plant breeding on genetic diversity. *Int J Plant Breed.* 2010; 4: 1–21.
3. Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, et al. The effects of artificial selection on the maize genome. *Sci.* 2005; 308: 1310–1314. doi: [10.1126/science.1107891](https://doi.org/10.1126/science.1107891) PMID: [15919994](https://pubmed.ncbi.nlm.nih.gov/15919994/)
4. Fu Y-B. Understanding crop genetic diversity under modern plant breeding. *Theor Appl Genet.* 2015; 128: 2131–2142. doi: [10.1007/s00122-015-2585-y](https://doi.org/10.1007/s00122-015-2585-y) PMID: [26246331](https://pubmed.ncbi.nlm.nih.gov/26246331/)
5. Doebley JF, Gaut BS, Smith BD. The molecular genetics of crop domestication. *Cell.* 2006; 1309–1321. doi: [10.1016/j.cell.2006.12.006](https://doi.org/10.1016/j.cell.2006.12.006) PMID: [17190597](https://pubmed.ncbi.nlm.nih.gov/17190597/)
6. Iorizzo M, Senalik D a, Ellison SL, Grzebelus D, Cavagnaro PF, Allender C, et al. Genetic structure and domestication of carrot (*Daucus carota* subsp. *sativus*) (Apiaceae). *Am J Bot.* 2013; 100: 930–8. doi: [10.3732/ajb.1300055](https://doi.org/10.3732/ajb.1300055) PMID: [23594914](https://pubmed.ncbi.nlm.nih.gov/23594914/)
7. Bradeen JM, Bach IC, Briard M, Clerc V, Senalik DA, Simon PW, et al. Molecular diversity analysis of cultivated carrot (*Daucus carota* L.) and wild *Daucus* populations reveals a genetically nonstructured composition. *J Am Soc Horticultural Sci.* 2002; 127: 383–391.
8. Baranski R, Maksylewicz-Kaul A, Nothnagel T, Cavagnaro PF, Simon PW, Grzebelus D. Genetic diversity of carrot (*Daucus carota* L.) cultivars revealed by analysis of SSR loci. *Genet Resour Crop Evol.* 2012; 59: 163–170. doi: [10.1007/s10722-011-9777-3](https://doi.org/10.1007/s10722-011-9777-3)
9. Magnussen LS, Hauser TP. Hybrids between cultivated and wild carrots in natural populations in Denmark. *Heredity.* 2007; 185–192. doi: [10.1038/sj.hdy.6800982](https://doi.org/10.1038/sj.hdy.6800982) PMID: [17473862](https://pubmed.ncbi.nlm.nih.gov/17473862/)
10. Arbizu C, Ruess H, Senalik D, Simon PW, Spooner DM. Phylogenomics of the carrot genus (*Daucus*, Apiaceae). *Am J Bot.* 2014; 101: 1–20. doi: [10.3732/ajb.1400106](https://doi.org/10.3732/ajb.1400106) PMID: [25077508](https://pubmed.ncbi.nlm.nih.gov/25077508/)
11. Grzebelus D, Iorizzo M, Senalik D, Ellison S, Cavagnaro P, Macko-Podgorni A, et al. Diversity, genetic mapping, and signatures of domestication in the carrot (*Daucus carota* L.) genome, as revealed by Diversity Arrays Technology (DARt) markers. *Mol Breed.* 2014; 33: 625–637. doi: [10.1007/s11032-013-9979-9](https://doi.org/10.1007/s11032-013-9979-9) PMID: [24532979](https://pubmed.ncbi.nlm.nih.gov/24532979/)
12. Brozynska M, Furtado A, Henry RJ. Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotechnol J.* 2015; 1–16. doi: [10.1111/pbi.12454](https://doi.org/10.1111/pbi.12454) PMID: [26311018](https://pubmed.ncbi.nlm.nih.gov/26311018/)
13. Orr HA. The genetic theory of adaptation: a brief history. *Nat Rev Genet.* 2005; 6: 119–127. doi: [10.1038/nrg1523](https://doi.org/10.1038/nrg1523) PMID: [15716908](https://pubmed.ncbi.nlm.nih.gov/15716908/)
14. Stapley J, Reger J, Feulner PGD, Smadja C, Galindo J, Ekblom R, et al. Adaptation genomics: The next generation. *Trends Ecol Evol.* 2010; 25: 705–712. doi: [10.1016/j.tree.2010.09.002](https://doi.org/10.1016/j.tree.2010.09.002) PMID: [20952088](https://pubmed.ncbi.nlm.nih.gov/20952088/)
15. Henry RJ, Nevo E. Exploring natural selection to guide breeding for agriculture. *Plant Biotechnol J.* 2014; 12: 655–662. doi: [10.1111/pbi.12215](https://doi.org/10.1111/pbi.12215) PMID: [24975385](https://pubmed.ncbi.nlm.nih.gov/24975385/)
16. Vanlerberghe GC. Alternative Oxidase: A mitochondrial respiratory pathway to maintain metabolic and signaling homeostasis during abiotic and biotic stress in plants. *Int J Mol Sci.* 2013; 14: 6805–47. doi: [10.3390/ijms14046805](https://doi.org/10.3390/ijms14046805) PMID: [23531539](https://pubmed.ncbi.nlm.nih.gov/23531539/)
17. Cvetkovska M, Vanlerberghe GC. Alternative oxidase impacts the plant response to biotic stress by influencing the mitochondrial generation of reactive oxygen species. *Plant, Cell Environ.* 2013; 36: 721–732. doi: [10.1111/pce.12009](https://doi.org/10.1111/pce.12009) PMID: [22978428](https://pubmed.ncbi.nlm.nih.gov/22978428/)
18. McDonald AE. Alternative oxidase: an inter-kingdom perspective on the function and regulation of this broadly distributed “cyanide-resistant” terminal oxidase. *Funct Plant Biol.* 2008; 35: 535–552.
19. Van Aken O, Giraud E, Clifton R, Whelan J. Alternative oxidase: a target and regulator of stress responses. *Physiol Plant.* 2009; 137: 354–361. doi: [10.1111/j.1399-3054.2009.01240.x](https://doi.org/10.1111/j.1399-3054.2009.01240.x) PMID: [19470093](https://pubmed.ncbi.nlm.nih.gov/19470093/)
20. Arnholdt-Schmitt B, Costa JH, de Melo DF. AOX- a functional marker for efficient cell reprogramming under stress? *Trends Plant Sci.* 2006; 11: 281–287. doi: [10.1016/j.tplants.2006.05.001](https://doi.org/10.1016/j.tplants.2006.05.001) PMID: [16713324](https://pubmed.ncbi.nlm.nih.gov/16713324/)
21. Abe F, Saito K, Miura K, Toriyama K. A single nucleotide polymorphism in the alternative oxidase gene among rice varieties differing in low temperature tolerance. 2002; 527: 181–185.
22. Santos Macedo E, Cardoso HG, Hernández A, Peixe AA, Polidoros A, Ferreira A, et al. Physiologic responses and gene diversity indicate olive alternative oxidase as a potential source for markers involved in efficient adventitious root induction. *Physiol Plant.* 2009; 137: 532–552. doi: [10.1111/j.1399-3054.2009.01302.x](https://doi.org/10.1111/j.1399-3054.2009.01302.x) PMID: [19941624](https://pubmed.ncbi.nlm.nih.gov/19941624/)
23. Costa JH, De Melo DF, Gouveia Z, Cardoso HG, Peixe A, Arnholdt-Schmitt B. The alternative oxidase family of *Vitis vinifera* reveals an attractive model to study the importance of genomic design. *Physiol Plant.* 2009; 137: 553–565. doi: [10.1111/j.1399-3054.2009.01267.x](https://doi.org/10.1111/j.1399-3054.2009.01267.x) PMID: [19682279](https://pubmed.ncbi.nlm.nih.gov/19682279/)

24. Campos C, Cardoso H, Nogales A, Svensson J, Lopez-Ráez JA, Pozo MJ, et al. Intra and Inter-spore variability in *Rhizophagus irregularis* AOX Gene. *PLoS One*. 2015; 10: e0142339. doi: [10.1371/journal.pone.0142339](https://doi.org/10.1371/journal.pone.0142339) PMID: [26540237](https://pubmed.ncbi.nlm.nih.gov/26540237/)
25. Cardoso HG, Campos MD, Costa A, Campos C, Nothnagel T, Arnholdt-Schmitt B. Carrot alternative oxidase gene AOX2a demonstrates allelic and genotypic polymorphisms in intron 3. *Physiol Plant*. 2009; 137: 592–608. doi: [10.1111/j.1399-3054.2009.01299.x](https://doi.org/10.1111/j.1399-3054.2009.01299.x) PMID: [19941625](https://pubmed.ncbi.nlm.nih.gov/19941625/)
26. Cardoso H, Doroteia Campos M, Nothnagel T, Arnholdt-Schmitt B. Polymorphisms in intron 1 of carrot AOX2b—a useful tool to develop a functional marker? *Plant Genet Resour*. 2011; 9: 177–180. doi: [10.1017/S1479262111000591](https://doi.org/10.1017/S1479262111000591)
27. Nogales A, Nobre T, Cardoso HG, Muñoz-Sanhueza L, Valadas V, Campos MD, et al. Allelic variation on DcAOX1 gene in carrot (*Daucus carota* L.): An interesting simple sequence repeat in a highly variable intron. *Plant Gene*. 2016; 5: 49–55. doi: [10.1016/j.plgene.2015.11.001](https://doi.org/10.1016/j.plgene.2015.11.001)
28. Costa JH, Mota EF, Cambursano MV, Lauxmann MA, de Oliveira LMN, Silva Lima MDG, et al. Stress-induced co-expression of two alternative oxidase (VuAox1 and 2b) genes in *Vigna unguiculata*. *J Plant Physiol*. 2010; 167: 561–70. doi: [10.1016/j.jplph.2009.11.001](https://doi.org/10.1016/j.jplph.2009.11.001) PMID: [20005596](https://pubmed.ncbi.nlm.nih.gov/20005596/)
29. Clifton R, Millar AH, Whelan J. Alternative oxidases in Arabidopsis: A comparative analysis of differential expression in the gene family provides new insights into function of non-phosphorylating bypasses. *Biochim Biophys Acta—Bioenerg*. 2006; 1757: 730–741. doi: [10.1016/j.bbabi.2006.03.009](https://doi.org/10.1016/j.bbabi.2006.03.009) PMID: [16859634](https://pubmed.ncbi.nlm.nih.gov/16859634/)
30. Cavalcanti JHF, Oliveira GM, Saraiva KDDC, Torquato JPP, Maia IG, de Melo DF, et al. Identification of duplicated and stress-inducible Aox2b gene co-expressed with Aox1 in species of the *Medicago* genus reveals a regulation linked to gene rearrangement in leguminous genomes. *J Plant Physiol*. 2013; 170: 1609–19. doi: [10.1016/j.jplph.2013.06.012](https://doi.org/10.1016/j.jplph.2013.06.012) PMID: [23891563](https://pubmed.ncbi.nlm.nih.gov/23891563/)
31. Campos MD, Nogales A, Cardoso HG, Kumar SR, Nobre T, Sathishkumar R, et al. Stress-induced accumulation of DcAOX1 and DcAOX2a transcripts coincides with critical time point for structural biomass prediction in carrot primary cultures (*Daucus carota* L.). *Front Genet*. 2016; 7: 1–17. doi: [10.3389/fgene.2016.00001](https://doi.org/10.3389/fgene.2016.00001) PMID: [26858746](https://pubmed.ncbi.nlm.nih.gov/26858746/)
32. Saisho D, Nambara E, Naito S, Tsutsumi N, Hirai A, Nakazono M. Characterization of the gene family for alternative oxidase from *Arabidopsis thaliana*. *Plant Mol Biol*. 1997; 89875: 585–596.
33. Spooner D, Rojas P, Bonierbale M, Mueller L a., Srivastav M, Senalik D, et al. Molecular phylogeny of *Daucus* (Apiaceae). *Syst Bot*. 2013; 38: 850–857. doi: [10.1600/036364413X670449](https://doi.org/10.1600/036364413X670449)
34. Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform*. 2008; 9: 286–298. doi: [10.1093/bib/bbn013](https://doi.org/10.1093/bib/bbn013) PMID: [18372315](https://pubmed.ncbi.nlm.nih.gov/18372315/)
35. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics*. 1998; 14: 817–818. PMID: [9918953](https://pubmed.ncbi.nlm.nih.gov/9918953/)
36. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. 2001; 17: 754–755. PMID: [11524383](https://pubmed.ncbi.nlm.nih.gov/11524383/)
37. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003; 19: 1572–1574. doi: [10.1093/bioinformatics/btg180](https://doi.org/10.1093/bioinformatics/btg180) PMID: [12912839](https://pubmed.ncbi.nlm.nih.gov/12912839/)
38. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*. 2003; 19: 2496–2497. doi: [10.1093/bioinformatics/btg359](https://doi.org/10.1093/bioinformatics/btg359) PMID: [14668244](https://pubmed.ncbi.nlm.nih.gov/14668244/)
39. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989; 123: 585–595. PMID: [2513255](https://pubmed.ncbi.nlm.nih.gov/2513255/)
40. Kelly JK. A test of neutrality based on interlocus associations. *Genetics*. 1997; 146: 1197–1206. PMID: [9215920](https://pubmed.ncbi.nlm.nih.gov/9215920/)
41. Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007; 24: 1586–1591. doi: [10.1093/molbev/msm088](https://doi.org/10.1093/molbev/msm088) PMID: [17483113](https://pubmed.ncbi.nlm.nih.gov/17483113/)
42. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online*. 2005; 1: 47–50.
43. Tisserant E, Malbreil M, Kuo A, Kohler A, Symeonidi A, Balestrini R, et al. Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proc Natl Acad Sci*. 2013; 110: 20117–20122. doi: [10.1073/pnas.1313452110](https://doi.org/10.1073/pnas.1313452110) PMID: [24277808](https://pubmed.ncbi.nlm.nih.gov/24277808/)
44. Larsson H, Kallman T, Gyllenstrand N, Lascoux M. Distribution of long-range linkage disequilibrium and Tajima's D values in scandinavian populations of Norway spruce (*Picea abies*). *Genes|Genomes|Genetics*. 2013; 3: 795–806. doi: [10.1534/g3.112.005462](https://doi.org/10.1534/g3.112.005462) PMID: [23550126](https://pubmed.ncbi.nlm.nih.gov/23550126/)
45. van de Wiel CCM, Sretenović Rajčić T, van Treuren R, Dehmer KJ, van der Linden CG, van Hintum T.J.L. Distribution of genetic diversity in wild European populations of prickly lettuce (*Lactuca serriola*):

- implications for plant genetic resources management. *Plant Genet Resour.* 2010; 8: 171–181. doi: [10.1017/S1479262110000134](https://doi.org/10.1017/S1479262110000134)
46. Lee BY, Park C-W. Molecular phylogeny of *Daucus* (Apiaceae): Evidence from nuclear ribosomal DNA ITS sequences. *J Species Res.* 2014; 3: 39–52. doi: [10.12651/JSR.2014.3.1.039](https://doi.org/10.12651/JSR.2014.3.1.039)
 47. Emami S, Arumainayagam D, Korf I, Rose AB. The effects of a stimulating intron on the expression of heterologous genes in *Arabidopsis thaliana*. *Plant Biotechnol J.* 2013; 11: 555–563. doi: [10.1111/pbi.12043](https://doi.org/10.1111/pbi.12043) PMID: [23347383](https://pubmed.ncbi.nlm.nih.gov/23347383/)
 48. Giani S, Morello L, Bardini M, Breviaro D. Tubulin intron sequences: multi-functional tools. *Cell Biol Int.* 2003; 27: 203–205. doi: [10.1016/S1065-6995\(02\)00301-3](https://doi.org/10.1016/S1065-6995(02)00301-3) PMID: [12681308](https://pubmed.ncbi.nlm.nih.gov/12681308/)
 49. Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-K, et al. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature.* 2005; 434: 338–345. doi: [10.1038/nature03441](https://doi.org/10.1038/nature03441) PMID: [15735639](https://pubmed.ncbi.nlm.nih.gov/15735639/)
 50. Baek J-M, Han P, Iandolino A, Cook DR. Characterization and comparison of intron structure and alternative splicing between *Medicago truncatula*, *Populus trichocarpa*, *Arabidopsis* and rice. *Plant Mol Biol.* 2008; 67: 499–510. doi: [10.1007/s11103-008-9334-4](https://doi.org/10.1007/s11103-008-9334-4) PMID: [18438730](https://pubmed.ncbi.nlm.nih.gov/18438730/)
 51. Li SC, Shiau CK, Lin WC. Vir-Mir db: Prediction of viral microRNA candidate hairpins. *Nucleic Acids Res.* 2008; 36: 184–189. doi: [10.1093/nar/gkm610](https://doi.org/10.1093/nar/gkm610) PMID: [17702763](https://pubmed.ncbi.nlm.nih.gov/17702763/)
 52. Jaillon O, Bouhouche K, Gout J-F, Aury J-M, Noel B, Sautemont B, et al. Translational control of intron splicing in eukaryotes. *Nature.* 2008; 451: 359–362. doi: [10.1038/nature06495](https://doi.org/10.1038/nature06495) PMID: [18202663](https://pubmed.ncbi.nlm.nih.gov/18202663/)
 53. Żmieńko A, Samelak A, Kozłowski P, Figlerowicz M. Copy number polymorphism in plant genomes. *Theor Appl Genet.* 2014; 127: 1–18. doi: [10.1007/s00122-013-2177-7](https://doi.org/10.1007/s00122-013-2177-7) PMID: [23989647](https://pubmed.ncbi.nlm.nih.gov/23989647/)
 54. Jeffares DC, Penkett CJ, Bähler J. Rapidly regulated genes are intron poor. *Trends Genet.* 2008; 24: 375–378. doi: [10.1016/j.tig.2008.05.006](https://doi.org/10.1016/j.tig.2008.05.006) PMID: [18586348](https://pubmed.ncbi.nlm.nih.gov/18586348/)
 55. Carmel L, Rogozin IB, Wolf YI, Koonin EV. Evolutionarily conserved genes preferentially accumulate introns. *Genome Res.* 2007; 17: 1045–1050. doi: [10.1101/gr.5978207](https://doi.org/10.1101/gr.5978207) PMID: [17495009](https://pubmed.ncbi.nlm.nih.gov/17495009/)
 56. Kroymann J. Natural diversity and adaptation in plant secondary metabolism. *Curr Opin Plant Biol.* 2011; 14: 246–251. doi: [10.1016/j.pbi.2011.03.021](https://doi.org/10.1016/j.pbi.2011.03.021) PMID: [21514879](https://pubmed.ncbi.nlm.nih.gov/21514879/)
 57. Benderoth M, Textor S, Windsor AJ, Mitchell-Olds T, Gershenzon J, Kroymann J. Positive selection driving diversification in plant secondary metabolism. *Proc Natl Acad Sci U S A.* 2006; 103: 9118–9123. doi: [10.1073/pnas.0601738103](https://doi.org/10.1073/pnas.0601738103) PMID: [16754868](https://pubmed.ncbi.nlm.nih.gov/16754868/)
 58. Umbach AL, Fiorani F, Siedow JN. Characterization of transformed *Arabidopsis* with altered alternative oxidase levels and analysis of effects on reactive oxygen species in tissue. *Plant Physiol.* 2005; 139: 1806–1820. doi: [10.1104/pp.105.070763](https://doi.org/10.1104/pp.105.070763) PMID: [16299171](https://pubmed.ncbi.nlm.nih.gov/16299171/)
 59. Sitaramam V, Pachapurkar S, Gokhale T. The alternative oxidase mediated respiration contributes to growth, resistance to hyperosmotic media and accumulation of secondary metabolites in three species. *Physiol Mol Biol Plants.* 2008; 14: 235–251. doi: [10.1007/s12298-008-0023-1](https://doi.org/10.1007/s12298-008-0023-1) PMID: [23572891](https://pubmed.ncbi.nlm.nih.gov/23572891/)