



HAL
open science

Molecular footprints of local adaptation in two Mediterranean conifers

Delphine Grivet, Federico Sebastiani, Ricardo Alia, Thomas Bataillon, Sara Torre, Mario Zabal, Giovanni G Vendramin, Santiago González-Martínez

► **To cite this version:**

Delphine Grivet, Federico Sebastiani, Ricardo Alia, Thomas Bataillon, Sara Torre, et al.. Molecular footprints of local adaptation in two Mediterranean conifers. *Molecular Biology and Evolution*, 2010, 28 (1), pp.101-116. 10.1093/molbev/msq190 . hal-00610661

HAL Id: hal-00610661

<https://hal.science/hal-00610661>

Submitted on 23 Jul 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Molecular footprints of local adaptation in two Mediterranean conifers

2
3 Research Article

4
5 Delphine Grivet ¹; Federico Sebastiani ^{2,3}; Ricardo Alía ^{1,5}; Thomas Bataillon ⁴; Sara Torre ³;
6 Mario Zabal-Aguirre ¹; Giovanni G. Vendramin ³; Santiago C. González-Martínez ^{1,5}

7
8 ¹Department of Forest Ecology & Genetics, Center of Forest Research, CIFOR-INIA, Carretera
9 de la Coruña km 7.5, E28040 Madrid, Spain;

10 ²Department of Agricultural Biotechnology, Genexpress, University of Florence, Via della
11 Lastruccia 14, I50019 Sesto Fiorentino (FI), Italy;

12 ³Plant Genetics Institute, Division of Florence, National Research Council, via Madonna del
13 Piano 10, I50019 Sesto Fiorentino (FI), Italy;

14 ⁴BiRC – Bioinformatics Research Center, Aarhus University, C.F. Møllers Allé 8, DK8000
15 Aarhus C, Denmark

16 ⁵Sustainable Forest Management Research Institute, University of Valladolid-INIA.

17
18 Corresponding author:

19 Santiago C. González-Martínez

20 Department of Ecology & Genetics, Center of Forest Research, CIFOR-INIA, Carretera de la
21 Coruña km 7.5, E28040 Madrid, Spain

22 Phone: (++34) 913471499

23 Fax: (++34) 913572293

24 e-mail: santiago@inia.es

25

26 Key words (3 to 6): natural selection, neutrality tests, environmental associations, candidate

27 genes, drought stress, *4cl*

28

29 Running head (max of 50 characters and spaces): Contrasting patterns of polymorphism in two

30 pines

31

32

33 **ABSTRACT**

34 This study combines neutrality tests and environmental correlations to identify non-neutral
35 patterns of evolution in candidate genes related to drought stress in two closely-related
36 Mediterranean conifers, *Pinus pinaster* Ait. and *Pinus halepensis* Mill. Based on previous
37 studies, we selected 12 amplicons covering six candidate genes that were sequenced in a large
38 sample spanning the full range of these two species. Neutrality tests relatively robust to
39 demography (*DHEW* compound test and *ML-HKA* test) were used to detect selection events at
40 different temporal scales. Environmental associations between variation at candidate genes and
41 climatic variables were also examined. These combined approaches detected distinct genes that
42 may be targeted by selection, most of them specific to only one of the two conifers, despite their
43 recent divergence (< 10 Ma). An exception was *4-coumarate: CoA ligase (4cl)*, a gene involved
44 in the production of various important secondary products that appeared to play a role in local
45 adaptation processes of both pines. Another remarkable result was that all significant
46 environmental correlations involved temperature indices, highlighting the importance of this
47 climatic factor as a selective driver on Mediterranean pines. The ability to detect natural
48 selection at the DNA sequence level depends on the nature and the strength of the selection
49 events, on the timescale at which they occurred and on the sensitivity of the methods to other
50 evolutionary forces that can mimic selection (e.g. demography, population structure). Using
51 complementary approaches can help to capture different aspects of the evolutionary processes
52 that govern molecular variation at both intra- and interspecific levels.

53

54

INTRODUCTION

55

56

57 Identifying candidate genes underlying genetic differences for adaptive traits can help to
58 understand how species have adapted to their environment and to predict how they will respond
59 to future climatic changes, which is a special concern in regions such as the Mediterranean
60 Basin, where a substantial decrease in precipitation and a pronounced warming is expected in the
61 near future (Petit, Hampe, and Cheddadi 2005; Giorgi and Lionello 2008). The ability to detect
62 the footprints of natural selection in population DNA sequence samples depends on the nature
63 and the strength of the selection events (Nielsen 2005), on the evolutionary scale at which they
64 occur (Zhai, Nielsen, and Slatkin 2009) and on the sensitivity of the methods to other
65 evolutionary forces that can mimic selection (e.g. demography, population structure; Biswas and
66 Akey 2006).

67 Positive selection, which drives the increase in frequency of advantageous mutations, is
68 of particular interest because it underlies local adaptation. Recent or on-going positive selection
69 events can be detected using methods based on polymorphism within species (Biswas and Akey
70 2006; Zhai, Nielsen, and Slatkin 2009). Recently, powerful methods based on site- and
71 haplotype- frequency spectra that are relatively insensitive to demography or population
72 structure have been developed (Zeng, Shi, and Wu 2007). Detecting selection events over a
73 wider evolutionary scale (e.g. recurrent positive selection or ancient selective sweeps), can be
74 efficiently done using methods contrasting within species polymorphism with among species
75 divergence, as these methods are less sensitive to assumptions regarding recombination or
76 demography. The *HKA* (Hudson-Kreitman-Aguadé) neutrality test (Hudson, Kreitman, and
77 Aguade 1987) and its extensions (e.g. Wright and Charlesworth 2004) are among the most

78 powerful tests to detect positive selection (Zhai, Nielsen, and Slatkin 2009). In addition, the
79 action of natural selection can be reflected in statistical associations between genetic and
80 environmental data (Manel et al. 2003; Hancock et al. 2008; Coop et al. 2009). Several studies
81 have shown that environmental heterogeneity influences the distribution of genetic diversity
82 across plant populations (see Savolainen, Pyhäjärvi, and Knürr 2007, for forest trees; Nakazato,
83 Bogonovich, and Moyle 2008; Montesinos et al. 2009). Detecting clinal genetic variation along
84 environmental gradients can thus provide evidence for the action of selection (Gram and Sork
85 2001; Vasemagi and Primmer 2005; Parisod and Christin 2008). Approaches that detect adaptive
86 clinal variation are very attractive because they can directly link environmental gradients with
87 genotypes and phenotypes. However, detecting environmental associations depends on the
88 spatial scale examined and the associations have to be controlled for selectively neutral processes
89 that can also generate clines (e.g. Storz 2002; Hancock et al. 2008; Coop et al. 2009; Keller et al.
90 2009). Using environmental association approaches in conjunction with neutrality tests helps
91 capturing different aspects of the evolutionary processes that govern molecular variation.

92 In the present study, we use the combined strategy outlined above to investigate patterns
93 of polymorphism for a set of candidate genes related to drought stress in maritime pine (also
94 known as cluster pine, *Pinus pinaster* Ait.), and in Aleppo pine (*Pinus halepensis* Mill.). *P.*
95 *pinaster* is an economically important pine native to the occidental Mediterranean Basin and the
96 European Atlantic front (Figure 1). This species is ecologically versatile, growing in a variety of
97 substrates, in a wide range of elevations and under a range of Mediterranean and Atlantic climate
98 regimes (semi-arid to very humid; Table 1). *P. halepensis*, a pine closely related to maritime
99 pine, is one of the most abundant and widespread pine species of the Mediterranean Basin
100 (Figure 1). This species, which also has a wide ecological breadth, is well adapted to dry

101 conditions as well as high-intensity fire regimes (Tapias et al. 2004), which makes it a species of
102 choice for afforestation in xeric Mediterranean areas. These two conifers, because they have
103 extensive and partially overlapping distributions across the Mediterranean Basin, may possibly
104 display adaptation to similar sets of contrasted environmental conditions, *i.e.* regional climate
105 types and spatial variability. Among the key environmental factors to *Pinus* distribution in
106 Mediterranean climates, temperature and rainfall have been shown to constitute good biological
107 descriptors (Richardson and Bond 1991; Soto et al. 2010). Additionally, several common garden
108 studies have documented differences in drought tolerance and/or gene expression among
109 Mediterranean pine provenances (for *P. pinaster* see Nguyen and Lamant 1989; Costa et al.
110 1998; Chambel, Climent, and Alía 2007; Aranda et al. 2009; for *P. halepensis* see Atzmon,
111 Moshe, and Schiller 2004; Sathyan, Newton, and Loopstra 2005; Voltas et al. 2008). Overall,
112 these studies point to Mediterranean pines as an interesting system to study local adaptation
113 mediated by abiotic stress response. In that context, studying allelic (SNP-allele and haplotype)
114 frequency in drought-related candidate genes along temperature or precipitation clines can help
115 to improve our understanding of pine adaptation to environmental heterogeneity.

116 Although *P. pinaster* and *P. halepensis* share various ecological requirements and may
117 have responded to their environment in a similar manner, they present quite contrasted
118 biogeographic and demographic histories that can also have a strong effect on the pattern of
119 variation of the candidate genes examined. *Pinus pinaster* has a long-presence in the western
120 Mediterranean Basin surviving the last glaciations in multiple refugia located in southeastern
121 Spain, northern Africa and the Atlantic coast of Portugal (Bucci et al. 2007). In contrast, *P.*
122 *halepensis* would have undergone long-range colonization of the western Mediterranean from
123 ancient populations located in Greece and Turkey (Bucci et al. 1998; Morgante, Felice, and

124 Vendramin 1998; Grivet et al. 2009). In addition, the dynamics of *P. halepensis* populations
125 throughout the Mediterranean Basin highly depend on forest fires. This species-specific
126 information on population dynamics has to be taken into account when looking at the molecular
127 footprint of selection.

128 Based on previous population genomic studies identifying drought tolerance genes
129 potentially under selection (see Eveno et al. 2008; Grivet et al. 2009) as well as on gene
130 expression studies identifying genes affected by drought stress (Watkinson et al. 2003; Rani et al.
131 2009), we examine here patterns of nucleotide variation in an ortholog set of targeted candidate
132 genes in *P. pinaster* and *P. halepensis*. We sample a broad geographic range of populations to: i)
133 provide new evidence about positive selection acting on these genes, ii) identify the time scale at
134 which selection events may have occurred, and iii) examine which environmental factors
135 underlie molecular signatures of selection in both species. Our pluralistic approach provides
136 insights on the adaptive strategy of two conifers that live under the same Mediterranean climate,
137 but present distinct demographic, (re)colonization and life histories.

138

139

MATERIALS AND METHODS

140

141 **Study species.** *Pinus pinaster* Ait.: Maritime pine populations have been assessed using various
142 molecular markers: chloroplast and mitochondrial (e.g. Burban and Petit 2003; Bucci et al.
143 2007), as well as nuclear (e.g. Salvador et al. 2000; Eveno et al. 2008). In particular, chloroplast
144 markers were able to identify different gene clusters related to the history of the species (Bucci et
145 al. 2007). Because of its economic importance, various programs of genetic improvement have
146 been developed in *P. pinaster*, in particular for the Atlantic provenances. Within this framework,

147 several adaptive traits, such as growth, tolerance to drought and cold, and resistance to pests and
148 diseases, have been the subject of genetic variability studies. Numerous approaches have been
149 used to unravel the basis of quantitative traits in maritime pine: genetic (genetic cartography),
150 physiologic (mechanisms implicated in traits), functional and structural genomic (candidate
151 genes and proteins) and population genetic and genomic (genes under selection).

152 *Pinus halepensis* Mill.: Aleppo pine genetic variability has been studied with both
153 chloroplast and nuclear markers (Bucci et al. 1998; Morgante, Felice, and Vendramin 1998;
154 Grivet et al. 2009). A recent study revealed that the pattern of polymorphism observed in some
155 candidate genes related to drought-tolerance in this species reflected long-range colonization and
156 possibly natural selection during range expansion (Grivet et al. 2009). Various common garden
157 studies have assessed Aleppo pine intraspecific variability in order to study the role of ecological
158 factors in shaping adaptive strategies. Some experiments revealed adaptive variation to climate
159 (total precipitation and dry season duration; Voltas et al. 2008) or reproductive features (Climent
160 et al. 2008), highlighting thus the selective role of climate variables in determining population
161 and family fitness in this species.

162

163 **Sampling.** Twelve populations of *P. pinaster* (77-122 individuals, depending on the gene; [Table](#)
164 [S1](#)) and nine populations of *P. halepensis* (72-93 individuals; [Table S1](#)) were collected spanning
165 the full range of each species ([Figure 1](#), [Table 1](#)). Populations were selected considering not only
166 spatial distribution but also environmental heterogeneity in both species, prioritizing populations
167 that represent contrasted environments ([Table 1](#)). Our sampling covers also different soil
168 (siliceous, calcareous) and Mediterranean forest types. Finally, we included representations of
169 the five traditional varieties or landraces described in maritime pine (Resch 1974): *mesogeensis*

170 (e.g. Olba), *atlantica* (e.g. Mimizan), *corteensis* (e.g. Pinia), *maghrebiana* (e.g. Tamrabta) and
171 *renoui* (e.g. Tabarka), as well as all genecological groups normally considered in Aleppo pine:
172 Western Europe, Eastern Europe and Northern Africa.

173 From each population, cones were collected from mother trees separated by at least 50 m
174 without any phenotypic selection. Seeds from each mother tree were kept in individual paper
175 bags and stored at 4 °C in a dry environment till DNA extraction (see below).

176

177 ***Spatial, environmental and genetic data.***

178 *Spatial and environmental data.* Three spatial variables were recorded for each population:
179 altitude, latitude and longitude (Table 1). Eight climatic variables were considered: annual mean
180 temperature (AMT), temperature seasonality (expressed as the standard deviation across months
181 multiplied by 100; TS), maximum temperature of the warmest month (MTWM), minimum
182 temperature of the coldest month (MTCM), annual precipitation (AP), precipitation of wettest
183 month (PWM), precipitation of driest month (PDM), and precipitation seasonality (coefficient of
184 variation; PS). Climatic data for *P. pinaster*'s Iberian populations were obtained from a
185 functional phytoclimatic model based on raw data from meteorological stations (Gonzalo 2007).
186 Climatic data for *P. pinaster*'s non-Iberian populations as well as for all *P. halepensis*'
187 populations were obtained from the WorldClim – Global Climate Data at 5 minute resolution
188 (Hijmans et al. 2005) (Table 1). Graphical pairwise correlations between these 11 spatial and
189 environmental variables are presented in Figure S1.

190 *DNA extraction and candidate gene sequencing.* Genomic DNA from *P. pinaster* (haploid)
191 megagametophytes was extracted with a modified Dellaporta et al. (1983) protocol. DNA
192 extraction for *P. halepensis* was carried out as reported in Grivet et al. (2009).

193 Candidate genes related to drought tolerance were originally identified on the basis of
194 functional studies performed in *Pinus taeda* and other conifers, or derived from model species
195 such as *Arabidopsis thaliana*, as described elsewhere (González-Martínez et al. 2006; Eveno et
196 al. 2008; Grivet et al. 2009; Wachowiak, Balk, and Savolainen 2009). Altogether the candidate
197 genes selected for this study belong to three well-known and relatively small multigene families:
198 the ASR (*lp31-Pt* and *lp33-Pt*), dehydrin (*dhn2-Pp*, *dhn2-Ps* and *dhn5-Ps*) and 4-coumarate:
199 *CoA ligase* (*4cl-Pt*) families. The ASR family is named after the Abscisic acid (ABA), stress and
200 ripening response (Frankel et al. 2006). These proteins are also induced by water deficit stress
201 (WDS). The dehydrin gene, *dhn2-Pp*, previously described in maritime pine (Eveno et al. 2008),
202 is not orthologous to the one described in Scots pine, i.e. *dhn2-Ps* (Wachowiak et al. 2009).

203 Previously published primer pairs of these putative candidate genes were transferred from
204 *P. taeda* (*Pt*), *P. sylvestris* (*Ps*) or *P. pinaster* (*Pp*) to either *P. pinaster* or *P. halepensis* or both,
205 and three new primer pairs, which overlap with pre-described amplicons, were also designed in
206 order to extend the coverage of these target genes (Table S2). PCR conditions for the two pines
207 studied here are given in Table S3. Outgroup sequences for each amplicon were obtained from
208 GenBank: *P. taeda* for amplicons *lp31-Pt*, *lp33-Pt*, *dhn2-Pp* and *4cl-Pt*, and *P. sylvestris* for
209 *dhn5-Ps* and one *dhn2-Ps* amplicon; or produced with our newly designed primers (*P. nigra* for
210 the second *dhn2-Ps* amplicon) (Table S2). Sequence data from this article have been deposited
211 with the EMBL/GenBank Data Libraries under accession nos. HM481479-HM483364.

212 PCR products were purified and sequenced from both ends following standard protocols
213 (for *P. pinaster*, see Eveno et al. 2008; for *P. halepensis*, see Grivet et al. 2009). Multiple
214 sequence alignments and manual adjustments were done using SeqMan v7 (DNASTAR
215 Lasergene software) and BIOEDIT (<http://www.mbio.ncsu.edu/BioEdit/page2.html>).

216 *Chloroplast microsatellites*. To control for associations between candidate genes and
217 environmental data that could be due to neutral processes, we included variation of chloroplast
218 microsatellites (the only neutral genetic markers available for the two species at the wide range
219 scale) as a covariate in the multivariate logistic regressions (see *Environmental associations*
220 section below). We performed population-based Principal Component Analysis (PCA) on
221 chloroplast markers, and kept the Principal Components (PCs) that explained the majority of the
222 overall inertia of the data. For *P. pinaster*, the three first PCs of the 16 most common haplotypes
223 (determined on the basis of five chloroplast microsatellites) – accounting for 74% of the overall
224 variation – were extracted from the lattice dataset (430 records × 16 variables) of Bucci et al.
225 (2007). For populations that did not match any of the 430 locations in this lattice, the mean of the
226 closest 3-4 locations was computed (Table S4). For *P. halepensis*, PCs were computed in R (R
227 Development Core Team) for each of the nine populations based on the 22 haplotypes
228 determined from three chloroplast microsatellites (our unpublished data) and the three first PCs,
229 which explain 82% of the overall variation, were used for subsequent analysis (Table S4).

230 Chloroplast microsatellites were also used to determine groups with similar evolutionary
231 history to conduct neutrality tests (Figure 1). A previous study based on chloroplast
232 microsatellites pointed at eight gene pools in *P. pinaster* (Bucci et al. 2007). Within the sampling
233 used in the present study, Mimizan (Continental French lineage) and Pinia (Northern Italy and
234 Corsican lineages) represent single lineages. Tabarka (Tunisia), Tamrabta (Moroccan Middle
235 Atlas) and Sidi Meskour (Moroccan High Atlas) are all part of a North African lineage that
236 combines western and eastern African origins. The rest of *P. pinaster* populations sampled are
237 part of a wide Western Mediterranean lineage, except for Quatretonda and Oria that are
238 considered marginal populations and may thus present a distinct evolutionary history (González-

239 Martínez et al. 2007a; Eveno et al. 2008). In *P. halepensis*, eastern Mediterranean populations
240 (Elea in Greece and Shaharia in Israel) present a higher level of cpSSR diversity and substantial
241 genetic differentiation from the rest and are considered single-population lineages (Grivet et al.
242 2009); Zaouia Ifrane (Morocco) and Aures Beni Melloul (Algeria) form part of the North
243 African group, while all four Spanish populations and the Italian population of Imperia belong to
244 the Western Mediterranean group. These regions (North Africa and Western Mediterranean)
245 define relatively homogeneous zones of the Mediterranean Basin in terms of soil and climate
246 (Barbéro et al. 1998).

247

248 ***Statistical analyses.***

249 *Gene diversity and divergence.* Number of segregating sites (S), nucleotide diversity statistics
250 (θ_π , Tajima 1989; θ_w , Watterson 1975), number of haplotypes (K), and haplotypic diversity
251 (H_e) were computed for both North African and Western Mediterranean groups using scripts
252 kindly provided by S.E. Ramos-Onsins (Department of Genetics, Faculty of Biology, University
253 of Barcelona, Spain) and the program DnaSP v5 (Librado and Rozas 2009). Average divergence
254 per site (K_{all}) for each geographical group was computed between each of the pines and the
255 outgroup *Pinus taeda*, as well as between *P. pinaster* and *P. halepensis* themselves, using scripts
256 also provided by S.E. Ramos-Onsins.

257 *Neutrality tests.* All neutrality tests were performed at the regional level (see group definition
258 above) considering all sequenced gametes in each amplicon, using one sequence from another
259 pine species as outgroup when appropriate, and considering all substitutions as well as only
260 silent sites (except for the *MK* test where all sites must be used, see below). First, we conducted
261 neutrality tests based on within-species population genetic data with (to determine ancestral

262 states) or without outgroup, including Tajima's D test (Tajima 1989), Fay and Wu's H normalized
263 test (Zeng et al. 2006), both based on the site-frequency spectrum (SFS), Ewens-Watterson EW
264 homozygosity test (Watterson 1978), and the $DHEW$ neutrality test (Zeng, Shi, and Wu 2007) – a
265 compound test that combines the properties of D , H normalized and EW tests – using scripts
266 provided by K. Zeng (State Key Laboratory of Biocontrol and Key Laboratory of Gene
267 Engineering of the Ministry of Education, Sun Yat-sen University, Guangzhou, China). Second,
268 we used methods that compare patterns of polymorphism within species and divergence between
269 species. The McDonald-Kreitman (MK) test, which compares the ratio of nonsynonymous and
270 synonymous mutations between and within species (McDonald and Kreitman 1991), was
271 computed using DnaSP v5 (Librado and Rozas 2009). $ML-HKA$ (Maximum-Likelihood
272 multilocus Hudson-Kreitman-Aguadé) neutrality tests were computed using the $mlhka$ program
273 (Wright and Charlesworth 2004). This test is an extension of the HKA test that allows to test
274 specific locus or groups of loci against a group of neutrally evolving genes for their levels of
275 polymorphism and divergence. First, each amplicon was compared to all the rest and the
276 selection parameter K was computed. The K parameter measures the degree to which diversity is
277 increased or decreased with respect to divergence. Second, for significant tests, the analysis was
278 repeated considering only amplicons from different genes in order to avoid potential linkage. All
279 runs consisted in 100,000 cycles of the Markov chain; long runs of 1,000,000 cycles were also
280 computed for significant tests as well as different starting values for the divergence time
281 parameter (T). However, no differences were found in the results and only basic runs are
282 presented here. Standard likelihood ratio tests (LRTs) were used to compare among models
283 considering one amplicon under selection at a time and the null model of no selection (degrees of
284 freedom equal one). The $ML-HKA$ test was performed using different outgroups to gain insights

285 on evolutionary times at which positive selective wave(s) may have taken place: i) with *P. taeda*
286 as outgroup for both *P. pinaster* and *P. halepensis* amplicons. In this case, the time frame would
287 correspond to the split between the *Trifoliae* section (to which *P. taeda* belongs) and the *Pinus*
288 section (to which the Mediterranean pines belong) in the Oligocene (~25 Ma; Gernandt et al.
289 2005; Gernandt et al. 2008). Three amplicons were discarded from these analyses because they
290 had *P. sylvestris* or *P. nigra* as outgroups (two from *dhn2-Ps*, and one from *dhn5-Ps*); ii) with *P.*
291 *halepensis* as outgroup for *P. pinaster* amplicons and *P. pinaster* as outgroup for *P. halepensis*
292 amplicons. The time frame, in this case, would correspond to the diversification of
293 Mediterranean piness (in the Miocene, ~10 Ma; Gernandt et al. 2008). One amplicon was
294 discarded from this series of tests (*dhn2-Ps_b*; see Table S2 for amplicon nomenclature) due to
295 bad alignment between species and, thus, the possible amplification of a paralogous fragment in
296 *P. halepensis* (*dhn2-Ps_b* from *P. halepensis* had more similarity with the outgroup, *P. nigra*,
297 than with its close relative *P. pinaster*).

298 *Environmental associations.* We tested if SNP-allele or haplotype frequencies at candidate loci
299 correlated with climatic variables using logistic regression. We first carried out series of
300 univariate logistic regressions to test for association between SNP/haplotype frequencies and
301 environmental variables using the program SAM (Joost, Kalbermatten, and Bonin 2008). We
302 considered a correlation as significant only when two likelihood ratio tests (G and Wald tests)
303 rejected the null hypothesis of no association between the genetic and the environmental
304 variables (at the 5% level). A strict Bonferroni correction was applied to correct for multiple
305 testing of univariate models.

306 In a second step, whenever a SNP/haplotype was found significantly correlated with an
307 environmental variable, we performed a multivariate logistic regression using neutral marker

308 PCs (see above) as covariates (together with the SNP and the environmental variables) to control
309 for neutral processes (e.g. postglacial migrations, geographical isolation) that may have also
310 generated clines in the absence of any local adaptation. Multiple logistic regressions were
311 performed in *R* (R Development Core Team) using the *glm* function (assuming that allelic counts
312 at SNPs were binomially distributed) for each SNP/haplotype separately. Model selection was
313 carried out using the *drop1* function and likelihood ratio tests (LRTs). Whenever the models
314 compared were not nested (and LRTs were thus not appropriate), we used Akaike's Information
315 Criterion (AIC). Tests for environmental associations were performed at the amplicon level
316 (counting the SNPs located within overlapping regions only once) and including only
317 populations with more than three gametes sampled (average number of gametes sampled per
318 population: 7.98 in *P. pinaster* and 9.27 in *P. halepensis*, see Table S1).

319

320

RESULTS

321

322 Gene diversity and divergence

323 Overall primer transfer rates across the two pine species were high (91.6% for *P. pinaster* and
324 83.3% for *P. halepensis*) and we obtained a total of 533 kbp and 380 kbp of sequence in *P.*
325 *pinaster* and *P. halepensis*, respectively. All sequences represented putative orthologs
326 (determined on the basis of similarities with available sequences in GenBank and construction of
327 phylogenies including multiple gene members of the target families in different conifers) except
328 probably for *dhn2-Ps_b* in *P. halepensis* (see Methods). Maritime pine showed a higher level of
329 gene diversity than Aleppo pine (Table 2), and this result held when considering only the nine
330 orthologous amplicons in the two species (data not shown). In total, 131 SNPs (24 non-

331 synonymous) were detected in *P. pinaster* (4 to 27 per amplicon) and 65 SNPs (12 non-
332 synonymous) in *P. halepensis* (1 to 21 per amplicon) (Table 2). There were only four shared
333 polymorphisms (i.e. 2 % of polymorphic sites were shared) across species, with two of them
334 (those within *4cl-Pt*) appearing only in one *P. pinaster* individual. Number of
335 insertions/deletions was similar between the two species for the same locus, but these indels
336 were not shared across species. None of them caused frame shifts. At the haplotype level, not
337 only the number of haplotypes substantially differed between the two species (86 for *P. pinaster*
338 vs. only 59 for *P. halepensis*) but also their frequency spectrum: *P. pinaster* had multiple
339 common haplotypes along with some less frequent ones, while *P. halepensis* generally displayed
340 one major haplotype together with some low-frequency ones (Figure S2). Finally, there was also
341 a contrasted pattern of gene diversity between the regional groups defined by chloroplast
342 microsatellites in *P. pinaster*: the Western Mediterranean group displayed higher nucleotide
343 diversity than the North African group (average θ_w per population of 0.0039 vs. 0.0022). The
344 SNP and haplotypes detected were used to conduct neutrality tests, as well as to detect
345 associations with environmental variables.

346 Divergence at the DNA sequence level was examined using two outgroups: *P. taeda*, a
347 New World pine, and each of the Mediterranean pines as reference for the other (i.e. *P.*
348 *halepensis* was used as outgroup for *P. pinaster* and *vice versa*). For both species, nucleotide
349 divergence per site with *P. taeda* ($K_{all} = 0.0397$ for *P. pinaster*; $K_{all} = 0.0498$ for *P. halepensis*)
350 was higher than with the other Mediterranean species ($K_{all} = 0.0235$ for *P. pinaster* vs. *P.*
351 *halepensis*) except for *dhn2-Pp_b* in *P. pinaster* that displayed a slightly higher divergence
352 between the Mediterranean pines than between each of them with *P. taeda*.

353

354 **Neutrality tests**

355 Neutrality tests rejected the null neutral model for distinct genes depending on the species and
356 the geographic group (Table 3). The D and H_{norm} tests identified some genes departing
357 significantly from expectation under the neutral model that are not detected by the more robust
358 $DHEW$ test: $lp31-Pt_a$ and $lp31-Pt_b$ in the North African group for *P. pinaster*; $lp33-Pp$ in the
359 Western Mediterranean group for *P. halepensis*. As shown below, these two tests are probably
360 reflecting demographical events and not selective processes. More relevantly, two fragments of
361 the $4cl-Pt$ gene were identified as potential targets of selection by $DHEW$ tests: $4cl-Pt_b$ in the
362 Western Mediterranean group for *P. pinaster* and $4cl-Pt_d$ in the North African group for *P.*
363 *halepensis*. Considering all polymorphic sites or only silent sites did not change the qualitative
364 results of the test (Table 3 and Table S5).

365 None of the MK tests showed any departure from the null hypothesis of an equal ratio of
366 nonsynonymous to synonymous variation within and between species (data not shown). The $ML-$
367 HKA tests based on silent sites (with *P. taeda* as outgroup) revealed uncoupled patterns of
368 polymorphism and divergence in only one amplicon in *P. pinaster* in the Western Mediterranean
369 group ($dhn2-Pp_b$), while in *P. halepensis* one amplicon in this same group ($lp33-Pp$) and one in
370 the North African group ($4cl-Pt_b$) were significant (Table 3). These results still stood when
371 analyses were restricted to only one amplicon per gene (in order to avoid potential linkage
372 between fragments from the same gene) and when fragments potentially under selection were not
373 used as control loci (data not shown). When all substitution sites were included in the analyses,
374 three more genes were significant for *P. halepensis* in the Western Mediterranean group ($dhn2-$
375 Ps_a , $4cl-Pt_a$, $4cl-Pt_b$) (Table S5). In contrast, when Mediterranean pines were used as
376 outgroup for each other, patterns of polymorphism and divergence either considering all sites

377 (Table S5) or silent sites (Table 3) seemed to evolve neutrally and a similar trend of the
378 maximum likelihood estimate of the selection parameter (*ML-HKA*'s *K*) was observed in the two
379 series of analyses.

380

381 **Environmental associations**

382 Environmental associations were examined at both the SNP and the haplotype level, and, overall,
383 similar associations were detected. Nonetheless, a few correlations differed, with some found
384 only at the haplotype level, which highlights how these two levels complement each other as
385 haplotypes may reflect interactions among linked mutations. Within the two species, all
386 significant associations detected between SNPs or haplotypes and climatic variables involved
387 temperature indices as environmental variables.

388 As many as 23 significant associations were initially found in *P. pinaster* (non-corrected
389 model as provided by SAM, see Methods). Only three associations remained after integrating
390 neutral marker PCs as covariates, two of them with spatial variables (one with altitude, one with
391 latitude) and one (*4cl-Pt_c_229*) with the minimum temperature of the coldest month (MTCM)
392 (Figure 2; Table S6). At the haplotype level, initially 32 associations were found in maritime
393 pine but only six remained in the corrected model: two (*4cl-Pt_c_A* and *4cl-Pt_c_B*) with
394 altitude and one (*dhn5-Ps_C*) with annual mean temperature (AMT), one (*dhn5-Ps_E*) with
395 maximum temperature of the warmest month (MTWM), and two (*4cl-Pt_c_A* and *4cl-Pt_c_B*)
396 with minimum temperature of the coldest month (MTCM) (Figure 2; Table S7). In the case of *P.*
397 *halepensis*, only one association was found by the corrected models, both at the SNP (between
398 *4cl-Pt_c_131* and temperature seasonality, TS) and at the haplotype (*4cl-Pt_c_B* and TS) level
399 (Figure 2; Tables S6 and S7).

400 In summary, association analysis identified two loci exhibiting significant correlations
401 with temperature indices: *4cl-Pt_c* that is associated with MTCM in *P. pinaster* and with TS in
402 *P. halepensis* at both the SNP and haplotype levels, and *dhn5-Ps* that is associated with AMT
403 and MTWM at the haplotype level in *P. pinaster* only.

404

405 DISCUSSION

406

407 In this study, we assessed the impact of natural selection on the same set of candidate genes
408 related to drought tolerance in two widespread Mediterranean pine species. Our results revealed
409 distinct selection patterns according to species, geographic regions and loci. Below, we discuss
410 these findings in the light of the history of each species and the specificities of each of the
411 methods used to reveal footprints of selection.

412

413 **Neutrality tests**

414 Neutrality tests examining selection events based on site- and haplotype-frequency spectrum
415 identified distinct genes potentially targeted by selection. The *D* and/or *H_{norm}* tests detected three
416 loci departing from neutrality that were not detected by the *DHEW* compound tests. These results
417 may be explained by the sensitivity of the *D* and *H_{norm}* tests to demographic factors and different
418 degrees of background selection (Zeng et al. 2006). Especially, Tajima's *D* is sensitive to
419 background selection and population growth while the *H_{norm}* is more sensitive to population
420 shrinkage and subdivision. In a previous study in *P. halepensis* based on ten candidate genes, we
421 have shown that this species has undergone historical bottlenecks and that the western
422 populations of the species harbored some signatures of this demographic event (Grivet et al.

423 2009). In addition, based on the present data set and extensive coalescent simulation, we have
424 found that observed values of D and H_{norm} for *P. pinaster* both in the Western Mediterranean and
425 North African groups reject the standard neutral model and suggest past bottlenecks in this
426 species too (see [Figure S3](#)). To test the robustness of the *DHEW* compound test, neutral
427 coalescent simulations under realistic bottlenecks scenarios were further simulated in *P. pinaster*
428 (1,000 coalescent simulations per scenario) and the significant values for this test were recorded.
429 In all cases, the number of significant tests obtained was equal or lower than the number
430 expected by chance (see [Figure S3](#)), highlighting the robustness of the *DHEW* test in the
431 presence of bottleneck events and geographic population structure of magnitudes similar to the
432 ones found in maritime and Aleppo pines. In addition to our simulations, the *DHEW* test has
433 been shown to be robust to recombination and to have high sensitivity to detect positive selection
434 (Zeng et al. 2007), as it combines powerful (and insensitive to recombination) haplotype-
435 frequency spectrum tests (i.e the *EW* test) with SFS statistics, such as the D test that maintains
436 power to detect positive selection across a wide period of time (Zeng et al. 2006; Zhai, Nielsen,
437 and Slatkin 2009). Only two on-going events of selection were detected by the *DHEW* test, one
438 in *P. pinaster* and one in *P. halepensis*. Both cases involved the *4cl-Pt* gene (albeit in different
439 geographical regions and distinct amplicons), suggesting a potential role of *4cl* in local adaption
440 in both pine species (see below).

441 Among the tests examining polymorphism within species and divergence among
442 species, the *MK* test did not detect any loci under selection. While this test is somewhat robust to
443 demography, it seems less powerful in detecting positive selection than *HKA* type tests (Zhai,
444 Nielsen, and Slatkin 2009). The *ML-HKA* tests for both silent and all sites, using *P. taeda* as
445 outgroup, identified three amplicons with uncoupled levels of polymorphism and divergence in

446 the two pines. Two genes (*dhn2-Pp_b* in *P. pinaster* in the Western Mediterranean group and
447 *4cl-Pt_b* in *P. halepensis* in the North African group) showed high levels of diversity compared
448 to divergence (selection parameter higher than one), a pattern compatible with balancing
449 selection. One other gene (*lp33-Pp* in the Western Mediterranean group in *P. halepensis*) showed
450 low diversity compared to divergence (selection parameter lower than one), which could reflect
451 the transient reduction in variability occurring during a selective sweep. Three other genes
452 showed uncoupled levels of polymorphism and divergence in *P. halepensis* when examining all
453 sites, but not when examining silent sites alone. Two of these genes (*4cl-Pt_a* and *4cl-Pt_b* in the
454 Western Mediterranean group) had an excess of nucleotide diversity, which could reflect recent
455 balancing selection acting on non-synonymous sites, while for one gene (*dhn2-Pp_a* in the
456 Western Mediterranean group) purifying selection reducing diversity in non-synonymous sites
457 could explain the observed pattern. Evidence for extensive purifying selection in conifers comes
458 from an approximately four-fold nucleotide diversity at synonymous compared to non-
459 synonymous sites in most species (see Table 1 in González-Martínez et al. 2010), including
460 maritime ($dN/dS = 0.169$) and Aleppo ($dN/dS = 0.344$) pines (this study). Although
461 demographical events can result in a significant *ML-HKA* test, its multilocus nature combined
462 with the *HKA* framework should produce a more robust test than those based on comparing
463 different aspects of polymorphism at a single locus (Wright and Charlesworth 2004).

464 In order to gain power by increasing sample size while taking the specific history of
465 each species into account, we performed the neutrality tests in groups of populations that present
466 similar evolutionary history (i.e. North African and Western Mediterranean groups). However,
467 grouping populations may bias the output of the SFS-based neutrality tests as it can increase the
468 proportion of singletons, leading to an excess of low frequency variants and thus to more

469 negative D (Städler et al. 2009). In our study, populations were grouped according to
470 homogeneous gene pools (see Bucci et al. 2007 for *P. pinaster* and Grivet et al. 2009 for *P.*
471 *halepensis*), the consequence of which should be that no substantial differences in the number of
472 rare variants at the population and group levels are found. The extensive gene flow normally
473 found in conifers with large and continuous distribution would also support this approach.
474 Nevertheless, to check that our results were not biased because of population grouping, we
475 performed neutrality tests at the population level too and found that values of D test at the group
476 level were not more negative than at the single population level (data not shown). Moreover,
477 results of the H_{nom} and $DHEW$ tests should not be affected by the accumulation of singletons as
478 the first is based on intermediary and high frequency variants while the second integrates all
479 variants of the site frequency spectrum. Thus, our analyses appear reasonably robust to
480 population grouping.

481 Selective events appear to have affected distinct genes in *P. pinaster* and in *P.*
482 *halepensis* despite their overlapping environment and close phylogenetic relationship, a result
483 that may be connected to the different histories of the two pines that have likely resulted in
484 different selective pressures. In *P. pinaster*, selective events were detected only within the
485 Western Mediterranean group, while in *P. halepensis* footprints of selection were identified both
486 in the Western Mediterranean and North African groups. These results point to different
487 geographical selection pressures that may have led to the process of regional adaptation of the
488 pines (see Barbéro et al. 1998 for a description of the different Mediterranean environmental
489 zones; and Gómez and Zamora 2000; Nakazato, Bogonovich, and Moyle 2008; Montesinos et al.
490 2009 for some examples on adaptive variation across heterogeneous environments). It is also
491 noticeable that *P. pinaster* had lower levels of nucleotide variation within the North African

492 range. This fact can be attributed to population history and/or interpreted in the light of more
493 extreme environmental conditions constraining population sizes and/or resulting in stronger
494 selection in this range.

495

496 **Environmental associations**

497 Environmental associations identified two loci that were correlated with temperature, suggesting
498 the importance of this climatic variable as selective agent (Saxe et al. 2001; Jump et al. 2006).
499 One of these genes was common for both pines (*4cl-Pt* correlated with MTCM in *P. pinaster* and
500 with TS in *P. halepensis*), while the other gene association was only significant in *P. pinaster*
501 (*dhn5-Ps* correlated with AMT and MTWM). Some of the correlations between genetic and
502 climatic data in *P. pinaster* were due basically to the extreme values of a few North African
503 populations (Sidi Meskour and Tamrabta for *4cl-Pt_c*; Tabarka for *dhn5-Ps_C*; see Figures 2 and
504 3) and may not represent true adaptive responses to environmental gradients but local adaptation
505 to particular environments or genetic drift due to population isolation (see references in
506 Alleaume-Benharira, Pen, and Ronce 2006; Rosenblum, Hickerson, and Moritz 2007). In
507 contrast, the other associations (*4cl-Pt_c* in *P. halepensis* and *dhn5-Ps_E* in *P. pinaster*) showed
508 more robust patterns. For instance, *dhn5-Ps_E* tended to be absent from populations
509 characterized by the highest MTWM (Arenas de San Pedro in central Spain, Tamrabta in
510 Moroccan Middle Atlas and Tabarka in coastal Tunisia) regardless of their geographical location
511 (Table 1 and Figure 3).

512 Environmental correlations with allelic variation can be spuriously inflated by neutral
513 processes that may also generate genetic clines, such as population history or population genetic
514 structure. This is particularly true for forest trees that may have followed a postglacial

515 colonization pathway overlapping temperature and rainfall clines. Here, to control for
516 confounding associations between candidate genes and environmental data, we included
517 variation of chloroplast microsatellites as a covariate in the multivariate logistic regressions. In
518 *P. pinaster*, the use of cpSSRs has allowed a more accurate description of genetic structure
519 (Bucci et al. 2007) than previous studies based on biochemical markers, such as terpenes or
520 allozymes (e.g. Baradat and Marpeau-Bezard 1988; Salvador et al. 2000). Recently, nuclear
521 molecular data for a geographically limited set of populations (nuSSRs and SNPs) (Eveno et al.
522 2008; our unpublished results) have identified similar gene pools as Bucci et al. (2007). More
523 extensive sampling with highly-polymorphic biparentally-inherited nuclear markers could, in
524 principle, reveal a more complex spatial genetic pattern affecting the outcome of some of the
525 multivariate logistic regressions reported in this study. However, because the available
526 chloroplast data set we used strongly reflects the species' history, they are expected to 'correct'
527 for the presence of overall neutral genetic gradients. Accordingly, including these markers as
528 covariates identified a substantial fraction of the correlations initially retained as significant as
529 false positives (83.6 % in *P. pinaster* and 85.7% in *P. halepensis*) (Tables S6 and S7).

530

531 ***The dehydrin gene family***

532 Neutrality tests and environmental associations both point to dehydrins as potential targets of
533 natural selection in *P. pinaster*. Some dehydrins were also suggested to be under selection in a
534 study based on detection of outlier loci in this species (Eveno et al. 2008). Dehydrins displayed
535 also non-neutral patterns of nucleotide diversity in *Pinus sylvestris* populations showing
536 divergence for cold tolerance in Europe (Wachowiak et al. 2009) and were associated with
537 carbon isotope discrimination (and, thus, potentially with drought tolerance) in *P. taeda*

538 (González-Martínez et al. 2008). Altogether these studies suggest the involvement of dehydrins
539 in the adaptive response of pines to abiotic stress.

540 Dehydrins are part of a relatively small multigene family of intracellular stabilizers that
541 plays a major role in cell protection against desiccation. These proteins are produced in response
542 to any type of stress that causes dehydration at the cellular level, such as cold, drought or salinity
543 (Close 1997). Changes in dehydrin gene expression have been reported in response to drought
544 and/or cold stress in many plants such as cowpea (Ismail, Hall, and Close 1999), barley
545 (Suprunova et al. 2004), wheat (Lopez et al. 2001), apple (Wisniewski et al. 2008), tomato
546 (Weiss and Egea-Cortines 2009) and blueberry (Panta, Rieger, and Rowland 2001); while
547 transgenic experiments confirmed the role of dehydrin genes in enhancing tolerance to drought
548 or freezing stress in plants (e.g. in *Arabidopsis*, Puhakainen et al. 2004). In conifers, at least eight
549 dehydrin genes have been identified in *Pinus sylvestris* (Joosen et al. 2006) and *Picea abies*
550 (Yakovlev et al. 2008), and dehydrin expression has been shown to increase under wounding,
551 cold and drought stress (Richard et al. 2000; Watkinson et al. 2003). Thus, it is not surprising to
552 find in our study significant associations between dehydrins and temperatures, as critical low
553 temperatures can cause tissue injury while high temperatures accompany dehydration, both
554 stimulating the accumulation of dehydrins (Lewitt 1980; Ingram and Bartels 1996; Rizhsky,
555 Liang, and Mittler 2002). Although our results, and others in *P. pinaster*, suggest that selection
556 may have acted on some of the dehydrins tested, further genetic association and functional
557 studies are necessary to confirm the role of dehydrins in local adaptation of this Mediterranean
558 pine.

559

560 **The 4-coumarate: CoA ligase (*4cl*) gene**

561 There was converging evidence both from neutrality tests (*DHEW* in *P. pinaster* and both
562 *DHEW* and *ML-HKA* in *P. halepensis*) and environmental associations that the *4-coumarate:*
563 *CoA ligase (4cl)* gene may be under selection in the two Mediterranean pines studied.
564 Preliminary results on phenotypical association corrected by neutral gradients in *P. pinaster* also
565 point at *4cl-Pt*. Indeed, multiple regressions reveal significant associations between
566 polymorphism in *4cl-Pt_c* at the SNP and haplotype levels and total height in populations
567 growing in sites characterized by dry and intermediate humidity (Table S8). The *4cl* family has
568 been extensively studied in plants, where it is encoded by four to five genes in the fully
569 sequenced genomes of *Arabidopsis*, rice and poplar (reviewed in Souza et al. 2008). The *4cl*
570 gene is involved, among other processes, in the production of basic enzymes of the
571 phenylpropanoid metabolism that are important metabolites acting as protectants against biotic
572 and abiotic stresses (Rani et al. 2009). The *4cl* gene also encodes key enzymes in the
573 biosynthesis of lignin and several studies have demonstrated its involvement in plant growth
574 (e.g. Yun et al. 2005; Wagner et al. 2009; Yun et al. 2009). Thus, we expect that changes in *4cl*
575 function would have significant repercussions on tree physiology and morphology. Implication
576 of *4cl* in pine morphology and physiology has been shown in gene association studies (for *P.*
577 *taeda* see González-Martínez et al. 2007b) as well as in gene suppression studies (for *P. radiata*
578 see Wagner et al. 2009).

579

580 **Neutrality tests versus Environmental correlations**

581 The different approaches used in this study suggested different loci under selection within each
582 of the two Mediterranean pine species. This fact has to be connected to the specificities of the
583 statistical methods used, whose performance has been recently studied under various

584 demographic and recombination scenarios (Zeng et al. 2007; Zeng, Shi, and Wu 2007; Ramírez-
585 Soriano et al. 2008; Zhai, Nielsen, and Slatkin 2009). Within-species site and haplotype
586 frequency spectrum-based methods are suitable to detect ongoing or recently fixed selective
587 sweeps - *EW* tending to be more powerful around the time when a selected mutation reaches
588 fixation (Zeng, Shi, and Wu 2007). In contrast, tests comparing levels of (within-species)
589 polymorphism and (among-species) divergence are able to detect the cumulative effects of
590 positive selection events over a wider evolutionary scale. Within this category, the *ML-HKA* test
591 is suitable to quantify the amount of selection in the genome and, due to its multilocus nature, it
592 is expected to be more robust to changes in population size such as population bottlenecks and
593 expansion than traditional tests (Wright and Charlesworth 2004). Simulations have shown that
594 the *HKA* test is relatively insensitive to change of divergence time as most of its power comes
595 from the transient reduction in variability occurring during selective sweeps (Zhai, Nielsen, and
596 Slatkin 2009). However, we only obtained significant tests when using the less closely-related
597 outgroup (i.e. a New World pine, *P. taeda*). This points to either lower power when divergence
598 time from the outgroup is low or, alternatively, to selection events that took place before the split
599 of the two Mediterranean pines considered. In our study, none of the *EW* neutrality tests were
600 significant (data not shown), while both the *DHEW* and the *ML-HKA* tests revealed distinct
601 genes potentially under selection. As a consequence, potential selective events detected by the
602 neutrality tests assessed in this study would correspond to either on-going or relatively old
603 selective events (i.e. before the diversification of the Mediterranean pines in the Miocene, ~10
604 Ma; Gernandt et al. 2008).

605 Last generation tests, such as *DHEW* or *ML-HKA*, are powerful for detecting positive
606 selection and are relatively insensitive to other evolutionary forces, but still do not integrate

607 recombination rate, a factor that can produce substantial biases (Nielsen et al. 2007; Ramírez-
608 Soriano et al. 2008), nor provide insights on selection drivers. Methods based on correlation
609 between genetic and environmental data are appealing because they aim at understanding
610 species-specific adaptations and processes that connect an organism to its environment by
611 looking at allelic frequencies or the genetic structure of populations (Foll and Gaggiotti 2006;
612 Joost, Kalbermatten, and Bonin 2008). These correlation methods are also more appealing than
613 methods based on detection of outliers (e.g. Eveno et al. 2008) as they target particular selection
614 drivers and provide a hypothesis testing framework. In addition, they are well suited for high-
615 throughput genotyping even for non-model species (Namroud et al. 2008; Eckert et al.
616 forthcoming). However, correlation methods are not without limitations, as it may be challenging
617 to find the environmental factors that are relevant for each species adaptation and associations
618 have to be controlled for historical and demographical processes, in particular for allelic clines
619 produced by postglacial migrations in Europe and the Americas.

620 Together, neutrality tests and environmental association approaches complement each
621 other by looking at different evolutionary scales and types of selection. In our study, they
622 detected a relatively-high number of genes showing non-neutral patterns of evolution, a result
623 that can be attributed to a selection of candidate genes based on earlier studies (González-
624 Martínez et al. 2006; Eveno et al. 2008; Grivet et al. 2009). This fact supports a candidate gene
625 approach with targeted genes, at least for organisms that have large genomes (e.g. conifers),
626 which, so far, preclude dense genome-wide sampling.

627

628 **Conclusion**

629 Our pluralistic approach revealed a dynamic action of natural selection in space and time within
630 maritime and Aleppo pines (Felsenstein 1976; see Vasemagi and Primmer 2005 for other
631 examples). Selection events along with environmental associations were detected; some of these
632 events differ between the two species reflecting individual histories (recolonization,
633 demography, adaptation) while others are shared, which translates partly as a common history of
634 these closely-related and partially sympatric Mediterranean pines. The simultaneous search for
635 patterns of selection in two closely-related species can help understanding the evolutionary
636 forces responsible for adaptive responses, and thus provides an effective way of assessing the
637 degree of local adaptation, a key factor to integrate in future management and conservation
638 strategies (Wright and Gaut 2005).

639

SUPPLEMENTARY MATERIAL

640

641

642 **Table S1.** Sampling size per amplicon (six putative candidate genes) and per population for *P.*
643 *pinaster* and *P. halepensis*.

644 **Table S2.** Characteristics of the primers used to amplify the candidate genes in *P. pinaster* and
645 *P. halepensis*.

646 **Table S3.** PCR conditions for amplicon amplification (six putative candidate genes) for *P.*
647 *pinaster* and *P. halepensis*.

648 **Table S4.** First three Principal Components (PCs) of the chloroplast haplotype frequency for *P.*
649 *pinaster* and *P. halepensis*.

650 **Table S5.** Polymorphism, divergence and neutrality tests (all sites) in *P. pinaster* and *P.*
651 *halepensis* for the phylogeographical groups defined with chloroplast microsatellites.

652 **Table S6.** SAM output (uncorrected models) showing the significant associations (at 5%)
653 between a given allele (SNP) at a given locus and a given environmental variable for *P. pinaster*
654 and *P. halepensis*.

655 **Table S7.** SAM output (uncorrected models) showing the significant associations (at 5%)
656 between a given haplotype and a given environmental variable for *P. pinaster* and *P. halepensis*.

657 **Table S8.** Phenotypic data for *P. pinaster* populations (unpublished data kindly provided by R.
658 Alía and P.H. Garnier-Géré).

659 **Figure S1.** Graphical pairwise relationships between the 11 spatial and environmental variables
660 recorded.

661 **Figure S2.** Haplotype distribution in *P. pinaster* and *P. halepensis* for the nine common
662 amplicons.

663 **Figure S3.** Demographic inferences in *P. pinaster* using multilocus coalescent simulation of
664 genetic bottleneck scenarios.
665

666

ACKNOWLEDGEMENTS

667

668 We are grateful to C. García-Barriga for her assistance in producing sequences of *P. pinaster*
669 candidate genes, to C. Collada, M.T. Cervera, M.A. Guevara, G. Gill, G.R. Brown and D.B.
670 Neale for providing some of the primers, as well as to A.I. de-Lucas and C. Ordoñez for
671 supplying some of the *P. pinaster* seeds. We thank J. Gonzalo for kindly providing the climatic
672 data from his functional phytoclimatic model, as well as S.E. Ramos-Onsins for his scripts to
673 compute the diversity parameters and K. Zeng for his scripts to compute the *DHEW* neutrality
674 test. Thanks are extended to M. Robson, M.A. Zavala and J. Climent for helpful discussion on
675 the study. We are also thankful to A.J. Eckert for thoughtful comments on the manuscript, N.
676 Takebayashi and two anonymous reviewers for constructive comments and suggestions, and P.C.
677 Grant for revising the English. The EUFORGEN program (Bioversity International) provided the
678 distribution maps used in Fig. 1.

679

680

FUNDING

681

682 This work was supported by the EU EVOLTREE Network of Excellence, the Collaborative
683 Project on ‘Conservation of Forest Genetic Resources’ between the Spanish Ministry of
684 Environment and INIA (AEG06-054) and Projects CGL2008-05289-C02-02 /BOS (VaMPiro)
685 and EUI2008-03713 (BiodivERsA-LinkTree) from the Spanish Ministry of Science and
686 Innovation, the later through the FP6 BiodivERsA Eranet. TB acknowledges support from the
687 Danish Research Council (FNU) through a Steno fellowship.

688

LITERATURE CITED

- 689
- 690
- 691 Alleaume-Benharira M, Pen IR, Ronce O. 2006. Geographical patterns of adaptation within a
692 species' range: interactions between drift and gene flow. *J Evol Biol.* 19:203-215.
- 693 Aranda I, Alía R, Ortega U, Dantas AK, Majada J. 2009. Intra-specific variability in biomass
694 partitioning and carbon isotopic discrimination under moderate drought stress in
695 seedlings from *Pinus pinaster* populations. *Tree Genet Genomes* 6:169-178.
- 696 Atzmon N, Moshe Y, Schiller G. 2004. Ecophysiological response to severe drought in *Pinus*
697 *halepensis* Mill. trees of two provenances. *Plant Ecol.* 171:15-22.
- 698 Baradat PH, Marpeau-Bezard A. 1988. Le pin maritime, *Pinus pinaster* Ait. Biologie et
699 génétique des terpènes pour la connaissance et l'amélioration de l'espèce. PhD Thesis,
700 University of Bordeaux I, France.
- 701 Barbéro M, Loisel R, Quézel P, Richardson DM, Romane F. 1998. Pines of the Mediterranean
702 Basin. In Richardson DM, editor. *Ecology and Biogeography of Pinus*. Cambridge
703 University Press, Cambridge. Pp. 153-170.
- 704 Biswas S, Akey JM. 2006. Genomic insights into positive selection. *Trends Genet.* 22:437-446.
- 705 Bucci G, Anzidei M, Madaghiele A, Vendramin GG. 1998. Detection of haplotypic variation and
706 natural hybridization in halepensis-complex pine species using chloroplast simple
707 sequence repeat (SSR) markers. *Mol Ecol.* 7:1633-1643.
- 708 Bucci G, González-Martínez SC, Le Provost G, Plomion C, Ribeiro MM, Sebastiani F, Alía R,
709 Vendramin GG. 2007. Range-wide phylogeography and gene zones in *Pinus pinaster* Ait.
710 revealed by chloroplast microsatellite markers. *Mol Ecol.* 16:2137-2153.

711 Burban C, Petit RJ. 2003. Phylogeography of maritime pine inferred with organelle markers
712 having contrasted inheritance. *Mol Ecol.* 12:1487-1495.

713 Chambel MR, Climent J, Alía R. 2007. Divergence among species and populations of
714 Mediterranean pines in biomass allocation of seedlings grown under two watering
715 regimes. *Ann For Sci.* 64:87-97.

716 Climent J, Prada MA, Calama R, Chambel MR, De Ron DS, Alía R. 2008. To grow or to seed:
717 Ecotypic variation in reproductive allocation and cone production by young female
718 Aleppo pine (*Pinus halepensis*, Pinaceae). *Am J Bot.* 95:833-842.

719 Close T. 1997. Dehydrins: a commonality in the response of plants to dehydration and low
720 temperature. *Physiol Plant.* 100:291-296.

721 Coop G, Pickrell JK, Novembre J, Kudaravalli S, Li J, Absher D, Myers RM, Cavalli-Sforza LL,
722 Feldman MW, Pritchard JK. 2009. The role of geography in human adaptation. *PLoS*
723 *Genetics* 5(6): e1000500.

724 Costa P, Bahrman N, Frigerio JM, Kremer A, Plomion C. 1998. Water-deficit-responsive
725 proteins in maritime pine. *Plant Mol Biol.* 38:587-596.

726 Dellaporta SL, Wood J, Hicks JB. 1983. A plant DNA miniprep: version II. *Plant Mol*
727 *Biol Rep.* 1:19-21.

728 Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, González-Martínez SC,
729 Neale DB. Forthcoming. Patterns of population structure and environmental associations
730 to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics*.

731 Eveno E, Collada C, Guevara MA, et al. (11 co-authors). 2008. Contrasting patterns of selection
732 at *Pinus pinaster* Ait. drought stress candidate genes as revealed by genetic
733 differentiation analyses. *Mol Biol Evol.* 25:417-437.

- 734 Felsenstein J. 1976. Theoretical population-genetics of variable selection and migration. *Annu*
735 *Rev Genet.*10:253-280.
- 736 Foll M, Gaggiotti O. 2006. Identifying the environmental factors that determine the genetic
737 structure of populations. *Genetics* 174:875-891.
- 738 Frankel N, Carrari F, Hasson E, Iusem ND. 2006. Evolutionary history of the *Asr* gene family.
739 *Gene* 378:74-83.
- 740 Gernandt DS, Lopez GG, Garcia SO, Liston A. 2005. Phylogeny and classification of *Pinus*.
741 *Taxon* 54:29-42.
- 742 Gernandt DS, Magallon S, Lopez GG, Flores OZ, Willyard A, Liston A. 2008. Use of
743 simultaneous analyses to guide fossil-based calibrations of *Pinaceae* phylogeny. *Int J*
744 *Plant Sci.*169:1086-1099.
- 745 Giorgi F, Lionello P. 2008. Climate change projections for the Mediterranean region. *Glob*
746 *Planet Change* 63:90-104.
- 747 Gómez JM, Zamora R. 2000. Spatial variation in the selective scenarios of *Hormathophylla*
748 *spinosa* (Cruciferae). *Am Nat.* 155:657-668.
- 749 González-Martínez SC, Dillon S, Garnier-Géré P, et al. (16 co-authors). Forthcoming. Patterns of
750 nucleotide diversity and association mapping In Kole C, editor. *Genomics of Conifers*.
751 Science Publishers, Inc., New Hampshire.
- 752 González-Martínez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB. 2006. DNA sequence
753 variation and selection of tag single-nucleotide polymorphisms at candidate genes for
754 drought-stress response in *Pinus taeda* L. *Genetics* 172:1915-1926.
- 755 González-Martínez SC, Gómez A, Carrión JS, Agúndez D, Alía R, Gil L. 2007a. Spatial genetic
756 structure of an explicit glacial refugium of maritime pine (*Pinus pinaster* Aiton) in

757 southeastern Spain. In Weiss S, Ferrand N, editors. Phylogeography in southern
758 European refugia: evolutionary perspectives on the origins of European biodiversity,
759 London. Pp. 257-269.

760 González-Martínez SC, Huber D, Ersoz E, Davis JM, Neale DB. 2008. Association genetics in
761 *Pinus taeda* L. II. Carbon isotope discrimination. *Heredity* 101:19-26.

762 González-Martínez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB. 2007b. Association
763 genetics in *Pinus taeda* L. I. Wood property traits. *Genetics* 175:399-409.

764 Gonzalo J. 2007. Phytoclimatic analysis of the Spanish Peninsula. Update and geostatistical
765 analysis. PhD Thesis, University of Valladolid, Spain.

766 Gram WK, Sork VL. 2001. Association between environmental and genetic heterogeneity in
767 forest tree populations. *Ecology* 82:2012-2021.

768 Grivet D, Sebastiani F, González-Martínez SC, Vendramin GG. 2009. Patterns of polymorphism
769 resulting from long-range colonization in the Mediterranean conifer Aleppo pine. *New*
770 *Phytol.* 184:1016-1028.

771 Hancock AM, Witonsky DB, Gordon AS, Eshel G, Pritchard JK, Coop G, Di Rienzo A. 2008.
772 Adaptations to climate in candidate genes for common metabolic disorders. *PLoS*
773 *Genetics* 4(2): e32.

774 Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated
775 climate surfaces for global land areas. *Int J Climatol.* 25:1965-1978.

776 Hudson RR, Kreitman M, Aguade M. 1987. A test of neutral molecular evolution based on
777 nucleotide data. *Genetics* 116:153-159.

778 Ingram J, Bartels D. 1996. The molecular basis of dehydration tolerance in plants. *Annu Rev*
779 *Plant Physiol Plant Mol Biol.* 47:377-403.

780 Ismail AM, Hall AE, Close TJ. 1999. Allelic variation of a dehydrin gene cosegregates with
781 chilling tolerance during seedling emergence. Proc Natl Acad Sci USA. 96:13566-13570.

782 Joosen RVL, Lammers M, Balk PA, Bronnum P, Konings MCJM, Perks M, Stattin E, Van
783 Wordragen MF, van der Geest AHM. 2006. Correlating gene expression to physiological
784 parameters and environmental conditions during cold acclimation of *Pinus sylvestris*,
785 identification of molecular markers using cDNA microarrays. Tree Physiol. 26:1297-
786 1313.

787 Joost S, Kalbermatten M, Bonin A. 2008. Spatial analysis method (SAM): a software tool
788 combining molecular and environmental data to identify candidate loci for selection. Mol
789 Ecol Resources 8:957-960.

790 Jump AS, Hunt JM, Martínez-Izquierdo JA, Peñuelas J. 2006. Natural selection and climate
791 change: temperature-linked spatial and temporal trends in gene frequency in *Fagus*
792 *sylvatica*. Mol Ecol. 15:3469-3480.

793 Keller SR, Sowell DR, Neiman M, Wolfe LM, Taylor DR. 2009. Adaptation and colonization
794 history affect the evolution of clines in two introduced species. New Phytol. 183:678-
795 690.

796 Kimura M, Ohta T. 1974. Some principles governing molecular evolution. Proc Natl Acad Sci
797 USA. 71:2848-2852.

798 Lewitt J. 1980. Responses of plants to environmental stresses, Vol I. Chilling, freezing and high
799 temperature stresses. Academic Press, New York, USA.

800 Librado P, Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA
801 polymorphism data. Bioinformatics 25:1451-1452.

802 Lopez CG, Banowetz G, Peterson CJ, Kronstad WE. 2001. Differential accumulation of a 24-kd
803 dehydrin protein in wheat seedlings correlates with drought stress tolerance at grain
804 filling. *Hereditas* 135:175-181.

805 Manel S, Schwartz MK, Luikart G, Taberlet P. 2003. Landscape genetics: combining landscape
806 ecology and population genetics. *Trends Ecol Evol.* 18:189-197.

807 McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the ADH locus in *Drosophila*
808 *Nature* 351:652-654.

809 Montesinos A, Tonsor SJ, Alonso-Blanco C, Picó FX. 2009. Demographic and genetic patterns
810 of variation among populations of *Arabidopsis thaliana* from contrasting native
811 environments. *PLoS One* 4:e7213.

812 Morgante M, Felice N, Vendramin GG. 1998. Analysis of hypervariable chloroplast
813 microsatellites in *Pinus halepensis* reveals a dramatic genetic bottleneck. In Karp A,
814 Isaac PG, Ingram DS, editors. *Molecular tools for screening biodiversity*. Chapman &
815 Hall, London, UK. Pp. 407-412.

816 Nakazato T, Bogonovich M, Moyle LC. 2008. Environmental factors predict adaptive
817 phenotypic differentiation within and between two wild Andean tomatoes. *Evolution*
818 62:774-792.

819 Namroud MC, Beaulieu J, Juge N, Laroche J, Bousquet J. 2008. Scanning the genome for gene
820 single nucleotide polymorphisms involved in adaptive population differentiation in white
821 spruce. *Mol Ecol.* 17:3599-3613.

822 Nei M. 1987. *Molecular evolutionary genetics*. Columbia Univ. Press, New York.

823 Nguyen A, Lamant A. 1989. Variation in growth and osmotic regulation of roots of water-
824 stressed maritime pine (*Pinus pinaster* Ait) provenances. *Tree Physiol.* 5:123-133.

- 825 Nielsen R. 2005. Molecular signatures of natural selection. *Annu Rev Genet.* 39:197-218.
- 826 Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG. 2007. Recent and ongoing selection
827 in the human genome. *Nat Rev Genet.* 8:857-868.
- 828 Panta GR, Rieger MW, Rowland LJ. 2001. Effect of cold and drought stress on blueberry
829 dehydrin accumulation. *J Horticult Sci Biotechnol.* 76:549-556.
- 830 Parisod C, Christin PA. 2008. Genome-wide association to fine-scale ecological heterogeneity
831 within a continuous population of *Biscutella laevigata* (Brassicaceae). *New Phytol.*
832 178:436-447.
- 833 Petit RJ, Hampe A, Cheddadi R. 2005. Climate changes and tree phylogeography in the
834 Mediterranean. *Taxon* 54:877-885.
- 835 Puhakainen T, Hess MW, Mäkelä P, Svensson J, Heino P, Palva ET. 2004. Overexpression of
836 multiple dehydrin genes enhances tolerance to freezing stress in *Arabidopsis*. *Plant Mol*
837 *Biol.* 54:743-753.
- 838 Quézel P. 2000. Taxonomy and biogeography of Mediterranean pines. In Ne'eman G, Trabaud L,
839 editors. Ecology, biogeography and management of *Pinus halepensis* and *P. brutia* forest
840 ecosystems in the Mediterranean Basin. Backhuys Publishers, Leiden, The Netherlands.
841 Pp. 1-12.
- 842 Ramírez-Soriano A, Ramos-Onsins SE, Rozas J, Calafell F, Navarro A. 2008. Statistical power
843 analysis of neutrality tests under demographic expansions, contractions and bottlenecks
844 with recombination. *Genetics* 179:555-567.
- 845 Rani A, Singh K, Sood P, Kumar S, Ahuja PS. 2009. p-Coumarate:CoA ligase as a key gene in
846 the yield of catechins in tea [*Camellia sinensis* (L.) O. Kuntze]. *Funct Integr Genomics* 9:
847 271-275.

848 Resch T. 1974. Essai de distinction des races majeurs de *Pinus pinaster*. Annales de la
849 Recherche Forestière au Maroc 14:91-102.

850 Richard S, Morency MJ, Drevet C, Jouanin L, Séguin A. 2000. Isolation and characterization of
851 a dehydrin gene from white spruce induced upon wounding, drought and cold stresses.
852 Plant Mol Biol. 43:1-10.

853 Richardson DM, Bond WJ. 1991. Determinants of plant-distribution - Evidence from pine
854 invasions. Am Nat. 137:639-668.

855 Rizhsky L, Liang HJ, Mittler R. 2002. The combined effect of drought stress and heat shock on
856 gene expression in tobacco. Plant Physiol. 130:1143-1151.

857 Rosenblum EB, Hickerson MJ, Moritz C. 2007. A multilocus perspective on colonization
858 accompanied by selection and gene flow. Evolution 61:2971-2985.

859 Salvador L, Alía R, Agúndez D, Gil L. 2000. Genetic variation and migration pathways of
860 maritime pine (*Pinus pinaster* Ait) in the Iberian peninsula. Theor Appl Genet. 100:89-
861 95.

862 Sathyan P, Newton RJ, Loopstra CA. 2005. Genes induced by WDS are differentially expressed
863 in two populations of Aleppo pine (*Pinus halepensis*). Tree Genet Genomes 1:166-173.

864 Savolainen O, Pyhäjärvi T, Knürr T. 2007. Gene flow and local adaptation in trees. Annu Rev
865 Ecol Evol Syst. 38:595-619.

866 Saxe H, Cannell MGR, Johnsen B, Ryan MG, Vourlitis G. 2001. Tree and forest functioning in
867 response to global warming. New Phytol. 149:369-399.

868 Soto A, Robledo-Arnuncio JJ, González-Martínez SC, Smouse PE, Alía R. 2010. Climatic niche
869 and neutral genetic diversity of the six Iberian pine species: a retrospective and
870 prospective view. Mol Ecol. 19:1396-1409.

871 Souza CD, Barbazuk B, Ralph SG, Bohlmann J, Hamberger B, Douglas CJ. 2008. Genome-wide
872 analysis of a land plant-specific acyl: coenzymeA synthetase (ACS) gene family in
873 *Arabidopsis*, poplar, rice and *Physcomitrella*. *New Phytol.* 179:987-1003.

874 Städler T, Haubold B, Merino C, Stephan W, Pfaffelhuber P. 2009. The impact of sampling
875 schemes on the site frequency spectrum in nonequilibrium subdivided populations.
876 *Genetics* 182:205-216.

877 Storz JF. 2002. Contrasting patterns of divergence in quantitative traits and neutral DNA
878 markers: analysis of clinal variation. *Mol Ecol.* 11:2537-2551.

879 Suprunova T, Krugman T, Fahima T, Chen G, Shams I, Korol A, Nevo E. 2004. Differential
880 expression of dehydrin genes in wild barley, *Hordeum spontaneum*, associated with
881 resistance to water deficit. *Plant Cell Environ.* 27:1297-1308.

882 Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA
883 polymorphism. *Genetics* 123:585-595.

884 Tapias R, Climent J, Pardos JA, Gil L. 2004. Life histories of Mediterranean pines. *Plant Ecol.*
885 171:53-68.

886 Vasemagi A, Primmer CR. 2005. Challenges for identifying functionally important genetic
887 variation: the promise of combining complementary research strategies. *Mol Ecol.*
888 14:3623-3642.

889 Voltas J, Chambel M, Prada M, Ferrio J. 2008. Climate-related variability in carbon and oxygen
890 stable isotopes among populations of Aleppo pine grown in common-garden tests. *Trees-*
891 *Struct Funct.* 22:759-769.

892 Wachowiak W, Balk PA, Savolainen O. 2009. Search for nucleotide diversity patterns of local
893 adaptation in dehydrins and other cold-related candidate genes in Scots pine (*Pinus*
894 *sylvestris* L.). *Tree Genet Genomes* 5:117-132.

895 Wagner A, Donaldson L, Kim H, Phillips L, Flint H, Steward D, Torr K, Koch G, Schmitt U,
896 Ralph J. 2009. Suppression of 4-Coumarate-CoA ligase in the coniferous gymnosperm
897 *Pinus radiata*. *Plant Physiol.* 149:370-383.

898 Watkinson JI, Sioson AA, Vasquez-Robinet C, et al. (14 co-authors). 2003. Photosynthetic
899 acclimation is reflected in specific patterns of gene expression in drought-stressed
900 loblolly pine. *Plant Physiol.* 133: 1702-1716.

901 Watterson GA. 1975. On the number of segregating sites in genetical models without
902 recombination. *Theor Pop Biol.* 7: 256-276.

903 Watterson GA. 1978. The homozygosity test of neutrality. *Genetics* 88:405-417.

904 Weiss J, Egea-Cortines M. 2009. Transcriptomic analysis of cold response in tomato fruits
905 identifies dehydrin as a marker of cold stress. *J Appl Genetics* 50:311-319.

906 Wisniewski M, Bassett C, Norelli J, Macarisin D, Artlip T, Gasic K, Korban S. 2008. Expressed
907 sequence tag analysis of the response of apple (*Malus x domestica* 'Royal Gala') to low
908 temperature and water deficit. *Physiol Plant.* 133:298-317.

909 Wright SI, Charlesworth B. 2004. The HKA test revisited: a Maximum-Likelihood-Ratio test of
910 the standard neutral model. *Genetics* 168:1071-1076.

911 Wright SI, Gaut BS. 2005. Molecular population genetics and the search for adaptive evolution
912 in plants. *Mol Biol Evol.* 22:506-519.

913 Yakovlev IA, Asante DKA, Fossdal CG, Partanen J, Junttila O, Johnsen O. 2008. Dehydrins
914 expression related to timing of bud burst in Norway spruce. *Planta* 228:459-472.

915 Yun MS, Chen WJ, Deng F, Yogo Y. 2005. Differential properties of 4-coumarate: CoA ligase
916 related to growth suppression by chalcone in maize and rice. *Plant Growth Regul.*
917 46:169-176.

918 Yun MS, Chen WJ, Deng F, Yogo Y. 2009. Selective growth suppression of five annual plant
919 species by chalcone and naringenin correlates with the total amount of 4-coumarate:
920 coenzyme A ligase. *Weed Biol Manag.* 9:27-37.

921 Zeng K, Fu Y-X, Shi S, Wu C-I. 2006. Statistical tests for detecting positive selection by
922 utilizing high-frequency variants. *Genetics* 174:1431-1439.

923 Zeng K, Mano SH, Shi SH, Wu CI. 2007. Comparisons of site- and haplotype-frequency
924 methods for detecting positive selection. *Mol Biol Evol.* 24:1562-1574.

925 Zeng K, Shi S, Wu CI. 2007. Compound tests for the detection of hitchhiking under positive
926 selection. *Mol Biol Evol.* 24:1898-1908.

927 Zhai WW, Nielsen R, Slatkin M. 2009. An investigation of the statistical power of neutrality
928 tests based on comparative and population genetic data. *Mol Biol Evol.* 26:273-283.

929

930

Table 1. Spatial coordinates and climatic variables for *P. pinaster* and *P. halepensis* populations.

Population	Country	Spatial variables			Climatic variables							
		Altitude (m)	Latitude (degrees)	Longitude (degrees)	AMT (°C)	TS	MTWM (°C)	MTCM (°C)	AP (mm)	PWM (mm)	PDM (mm)	PS
<i>P. pinaster</i>												
Arenas de San Pedro	Spain	733	40.194822	-5.116213	14.2	668.9	33.4	1.4	1318	199	12	60.8
Cazorla	Spain	1100	37.919675	-2.925765	11.5	650.2	30.6	-1.4	1257	179	11	58.3
Coca	Spain	800	41.254705	-4.497827	12.3	655.9	31.2	-0.6	454	55	15	30.0
Cómpeta	Spain	903	36.834265	-3.953989	14.0	512.1	27.5	4.7	899	132	5	64.1
Olba	Spain	1002	40.173309	-0.622966	12.4	600.7	28.3	0.6	509	63	24	31.5
Oria	Spain	1223	37.531165	-2.351138	13.1	633.1	30.6	0.4	357	46	5	46.2
Quatretonda	Spain	435	38.971645	-0.358844	15.3	553.7	30.3	3.8	777	120	9	52.1
Mimizan	France	19	44.134167	-1.303056	13.2	490.9	24.9	3.1	1235	149	62	24.3
Pinia	France	15	42.021083	9.464861	15.6	515.5	27.0	6.3	585	80	8	49.6
Tabarka	Tunisia	121	36.958397	8.703792	17.7	573.4	31.8	6.7	916	162	4	72.3
Tamrabta	Morocco	1758	33.600000	-5.016667	11.8	665.3	31.3	-3.2	721	96	8	54.4
Sidi Meskour	Morocco	1931	31.439375	-6.903864	11.4	711.3	32.5	-5.2	514	71	5	56.4
<i>P. halepensis</i>												
Cabanellas	Spain	210	42.235556	2.790000	14.8	575.5	27.4	3.9	713	97	35	26.0
Carratraca	Spain	650	36.841111	-4.834444	15.7	568.6	30.6	4.4	693	106	3	69.4
S'avall	Spain	10	39.287222	3.047778	16.8	525.9	28.7	6.8	566	92	5	54.7
Tarrasa	Spain	117	41.466667	2.100000	15.9	551.1	27.7	5.8	619	83	29	29.5
Imperia	Italy	109	43.900000	8.050000	15.1	550.5	27.5	4.6	804	107	18	38.5
Zaouia Ifrane	Morocco	1512	33.570000	-5.140000	11.5	668.7	31.2	-3.8	849	123	8	57.4

Aures Beni Melloul	Algeria	936	35.166667	6.833333	13.4	657.2	31.2	-0.1	375	53	10	36.1
Elea	Greece	155	37.766667	21.533333	16.7	592.0	30.8	5.4	808	150	6	77.1
Shaharia	Israel	236	31.600000	34.833333	20.0	496.4	31.8	8.0	397	100	0	116

AMT: annual mean temperature; TS: temperature seasonality (STD *100); MTWM: maximum temperature of the warmest month; MTCM: minimum temperature of the coldest month; AP: annual precipitation; PWM: precipitation of the wettest month; PDM: precipitation of the driest month; PS: precipitation seasonality (CV *100).

Table 2. Gene diversity (all sites) for 12 amplicons from six putative candidate genes across the twelve *P. pinaster* populations and the nine *P. halepensis* populations.

Amplicon	<i>P. pinaster</i>							<i>P. halepensis</i>						
	<i>N</i>	<i>L</i>	<i>S</i>	θ_π	θ_w	<i>K</i>	<i>He</i>	<i>N</i>	<i>L</i>	<i>S</i>	θ_π	θ_w	<i>K</i>	<i>He</i>
<i>lp31-Pt</i>														
<i>a</i>	122	456	12	6.26	4.89	6	0.717	79	353	3	2.48	1.72	4	0.460
<i>b</i>	111	560	7	3.91	2.37	3	0.446	74	488	7	3.93	2.94	4	0.464
<i>lp33-Pp</i>	97	449	9	4.29	3.89	9	0.798	90	375	1	0.17	0.53	2	0.065
<i>dhn2-Pp</i>														
<i>a</i>	120	472	12	6.06	4.74	11	0.807	89	448	2	0.20	0.88	3	0.088
<i>b</i>	85	596	21	8.42	7.03	11	0.811	93	346	4	0.55	2.26	5	0.105
<i>dhn2-Ps</i>														
<i>a</i>	92	743	27	8.10	7.13	14	0.899	na	na	na	na	na	na	na
<i>b</i>	80	513	15	5.92	5.90	14	0.855	92	457	11	1.93	4.73	6	0.204
<i>dhn5-Ps</i>	77	449	4	4.00	1.81	5	0.735	na	na	na	na	na	na	na
<i>4cl-Pt</i>														
<i>a</i>	91	551	13	5.67	4.64	4	0.678	88	461	10	5.88	4.30	13	0.805
<i>b</i>	92	256	7	3.56	5.37	5	0.291	79	688	21	10.67	6.18	13	0.793
<i>c</i>	87	543	4	0.69	1.46	4	0.341	78	528	3	2.47	1.15	5	0.697
<i>d</i>	na	na	na	na	na	na	na	72	452	3	2.80	1.37	4	0.484
<i>Total</i>	1054	5588	131	-	-	86	-	834	4596	65	-	-	59	
<i>Mean</i>	-	-	-	5.17	4.48	-	0.671	-	-	-	3.11	2.61	-	0.416

N : number of sequences; L : total analyzed length in base pair; S : polymorphic sites; θ_{π} : Tajima's nucleotide diversity $\times 10^{-3}$ (Tajima 1989) and θ_w : Watterson's nucleotide diversity (Watterson 1975) per site ($\times 10^{-3}$); K : number of haplotypes; He : haplotype diversity; na: amplicons that failed to transfer across species.

Table 3a. Polymorphism, divergence and neutrality tests (silent sites) in *P. pinaster* for the phylogeographical groups defined with chloroplast microsatellites (see text for details).

Amplicon	<i>S</i>	<i>D</i>	<i>H_{norm}</i>	<i>DHEW</i> <i>P</i> -value	<i>ML-HKA's K</i> (<i>P. taeda</i>)	<i>K_s</i> (<i>P. taeda</i>)	<i>ML-HKA's K</i> (<i>P. halepensis</i>)	<i>K_s</i> (<i>P. halepensis</i>)
Western Mediterranean ^a								
<i>lp31-Pt</i>								
<i>a</i>	8	2.172	-0.635	0.958	0.961	38.350	1.292	20.310
<i>b</i>	6	1.834	-1.380	0.913	0.698	44.250	0.720	24.360
<i>lp33-Pp</i>								
<i>dhn2-Pp</i>	6	1.133	0.911	0.984	0.696	60.060	1.839	15.990
<i>a</i>	7	0.269	-0.853	0.406	1.605	38.160	1.307	31.580
<i>b</i>	19	0.186	0.825	0.727	3.256 *	31.940	1.975	39.720
<i>dhn2-Ps</i>								
<i>a</i>	15	0.631	0.703	0.562	ptna	ptna	phna	phna
<i>b</i>	8	-0.126	0.232	0.209	ptna	ptna	d	d
<i>dhn5-Ps</i>								
<i>a</i>	2	1.960	0.136	0.907	ptna	ptna	phna	phna
<i>4cl-Pt</i>								
<i>a</i>	2	0.422	0.067	0.610	0.307	31.480	0.250	26.870
<i>b</i>	2	-1.041	-3.173 *	0.014 *	0.741	43.930	0.446	41.220
<i>c</i>	1	0.214	0.565	0.663	0.542	23.080	0.000	1.190
<i>d</i>	na	na	na	na	na	na	na	na
North Africa ^b								
<i>lp31-Pt</i>								
<i>a</i>	7	-0.276	-3.065 *	0.336	0.878	39.300	0.952	22.280

<i>b</i>	6	-2.046	-3.813	0.073	0.820	44.250	0.880	26.150
<i>lp33-Pp</i>	5	0.014	0.926	0.899	0.614	57.860	1.946	13.040
<i>dhn2-Pp</i>								
<i>a</i>	4	0.799	-0.554	0.615	0.868	36.610	0.792	28.620
<i>b</i>	13	1.095	0.361	0.718	2.036	32.780	1.362	33.600
<i>dhn2-Ps</i>								
<i>a</i>	6	0.190	-0.232	0.322	ptna	ptna	phna	phna
<i>b</i>	6	-1.116	-1.223	0.085	ptna	ptna	d	d
<i>dhn5-Ps</i>	2	0.688	-0.809	0.540	ptna	ptna	phna	phna
<i>4cl-Pt</i>								
<i>a</i>	8	2.797	0.411	0.995	1.279	39.160	0.442	28.880
<i>b</i>	3	2.266	0.090	0.982	1.303	48.190	1.017	34.730
<i>c</i>	2	0.201	0.387	0.494	1.039	25.240	0.000	3.410
<i>d</i>	na	na	na	na	na	na	na	na

D: Tajima's *D* (Tajima 1989); H_{norm} : Fay and Wu's normalized *H* (Zeng et al. 2006); *DHEW* test: compound test (Zeng et al. 2007); *ML-HKA's K*: Selection parameter of the Maximum-Likelihood multilocus Hudson-Kreitman-Aguadé neutrality tests (Wright and Charlesworth 2004); the species used as outgroup is given in parenthesis; *S*: number of segregating sites; K_s : average proportion of nucleotide differences between species per silent site $\times 10^{-3}$.

na: amplicons that failed to transfer across species; ptna: *P. taeda* outgroup sequence not available; phna: *P. halepensis* outgroup sequence not available; d: discarded because possible paralog.

P-values [P neutrality test (neutral) \leq P neutrality test (observed)]: 0 < *** < 0.001 < ** < 0.01 < * < 0.05

^a Western Mediterranean: Arenas de San Pedro, Cazorla, Coca, Cómpea and Olba.

^b North Africa: Tabarka, Tamrabta and Sidi Meskour.

Table 3b. Polymorphism, divergence and neutrality tests (silent sites) in *P. halepensis* for the phylogeographical groups defined with chloroplast microsatellites (see text for details).

Amplicon	<i>S</i>	<i>D</i>	<i>H_{norm}</i>	<i>DHEW</i> <i>P</i> -value	<i>ML-HKA's K</i> (<i>P. taeda</i>)	<i>K_s</i> (<i>P. taeda</i>)	<i>ML-HKA's K</i> (<i>P. pinaster</i>)	<i>K_s</i> (<i>P. pinaster</i>)
Western Mediterranean ^a								
<i>lp31-Pt</i>								
<i>a</i>	2	1.689	-1.714	0.860	0.597	34.590	0.452	20.960
<i>b</i>	4	0.535	-0.948	0.501	0.889	48.280	1.380	23.090
<i>lp33-Pp</i>								
<i>dhn2-Pp</i>	1	-0.629	-4.343 *	0.270	0.145 *	69.100	0.361	14.020
<i>a</i>	0	nps	nps	nps	nps	55.770	nps	44.020
<i>b</i>	0	nps	nps	nps	nps	57.180	nps	38.220
<i>dhn2-Ps</i>								
<i>a</i>	na	na	na	na	na	na	na	na
<i>b</i>	2	-0.595	0.317	0.382	ptna	ptna	d	d
<i>dhn5-Ps</i>								
<i>a</i>	na	na	na	na	na	na	na	na
<i>4cl-Pt</i>								
<i>a</i>	6	0.752	-1.222	0.599	2.236	43.260	1.618	22.500
<i>b</i>	16	2.386	-0.320	0.985	2.279	52.000	1.242	30.660
<i>c</i>	2	1.701	-1.305	0.867	1.350	31.970	2.976	9.440
<i>d</i>	3	2.324	-0.981	0.965	0.674	54.780	ppna	ppna
North Africa ^b								
<i>lp31-Pt</i>								
<i>a</i>	2	1.818	-1.053	0.898	0.819	34.140	1.123	12.710

<i>b</i>	2	1.882	0.602	0.909	0.530	49.000	0.924	22.700
<i>lp33-Pp</i>	0	nps	nps	nps	nps	69.410	nps	9.550
<i>dhn2-Pp</i>								
<i>a</i>	1	-0.774	0.289	0.313	0.329	56.280	0.525	19.310
<i>b</i>	1	-1.147	0.160	0.189	0.215	57.320	0.205	33.600
<i>dhn2-Ps</i>								
<i>a</i>	na	na	na	na	na	na	na	na
<i>b</i>	0	nps	nps	nps	ptna	ptna	d	d
<i>dhn5-Ps</i>	na	na	na	na	na	na	na	na
<i>4cl-Pt</i>								
<i>a</i>	4	1.306	-0.201	0.791	2.072	44.030	1.815	22.450
<i>b</i>	14	2.316	0.044	0.981	3.251 *	50.760	1.951	31.290
<i>c</i>	2	1.077	0.221	0.834	1.665	27.720	3.309	5.200
<i>d</i>	3	-1.494	-6.584	0.009 **	0.860	60.020	ppna	ppna

D: Tajima's *D* (Tajima 1989); H_{norm} : Fay and Wu's normalized *H* (Zeng et al. 2006); *DHEW* test: compound test (Zeng et al. 2007); *ML-HKA*'s *K*: Selection parameter of the Maximum-Likelihood multilocus Hudson-Kreitman-Aguadé neutrality tests (Wright and Charlesworth 2004); the species used as outgroup is given in parenthesis; *S*: number of segregating sites; K_s : average proportion of nucleotide differences between species per silent site $\times 10^{-3}$.

na: amplicons that failed to transfer across species; ptna: *P. taeda* outgroup sequence not available; ppna: *P. pinaster* outgroup sequence not available; d: discarded because possible paralog; nps: monomorphic fragment.

P-values [P neutrality test (neutral) \leq P neutrality test (observed)]: 0 < *** < 0.001 < ** < 0.01 < * < 0.05

^a Western Mediterranean: Cabanellas, Carratraca, Imperia, S'avall and Tarrasa.

^b North Africa: Aures Beni Melloul and Zaouia Ifrane.

Figure 1. Location of the 12 *P. pinaster* and the 9 *P. halepensis* populations and the distribution of the two species. Grey areas represent the two groups (Western Mediterranean and North African) defined by chloroplast SSRs (see text for details).

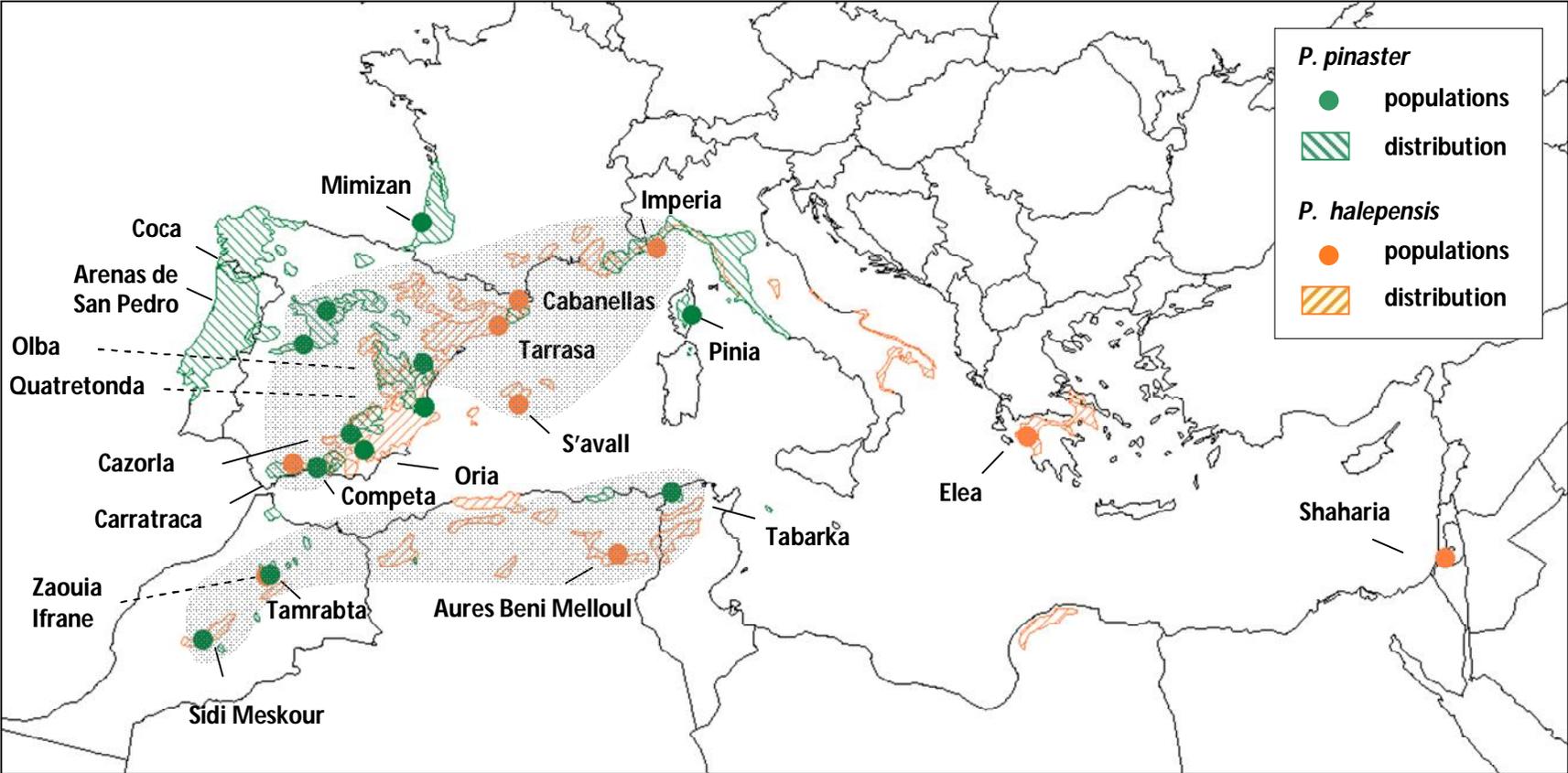


Figure 2. Predicted and observed patterns of clinal variation at *4cl-Pt_c*. Sigmoid curves denote the clines in either allelic (SNP position given) or haplotypic (haplotypes represented by capital letters) frequencies under a logistic regression model. The logistic regression model uses a climatic variable (MTCM or TS) after controlling for neutral processes (i.e. using the PCs of neutral molecular markers as covariates) for *P. pinaster* and *P. halepensis*. MTCM: minimum temperature of the coldest month; TS: temperature seasonality.

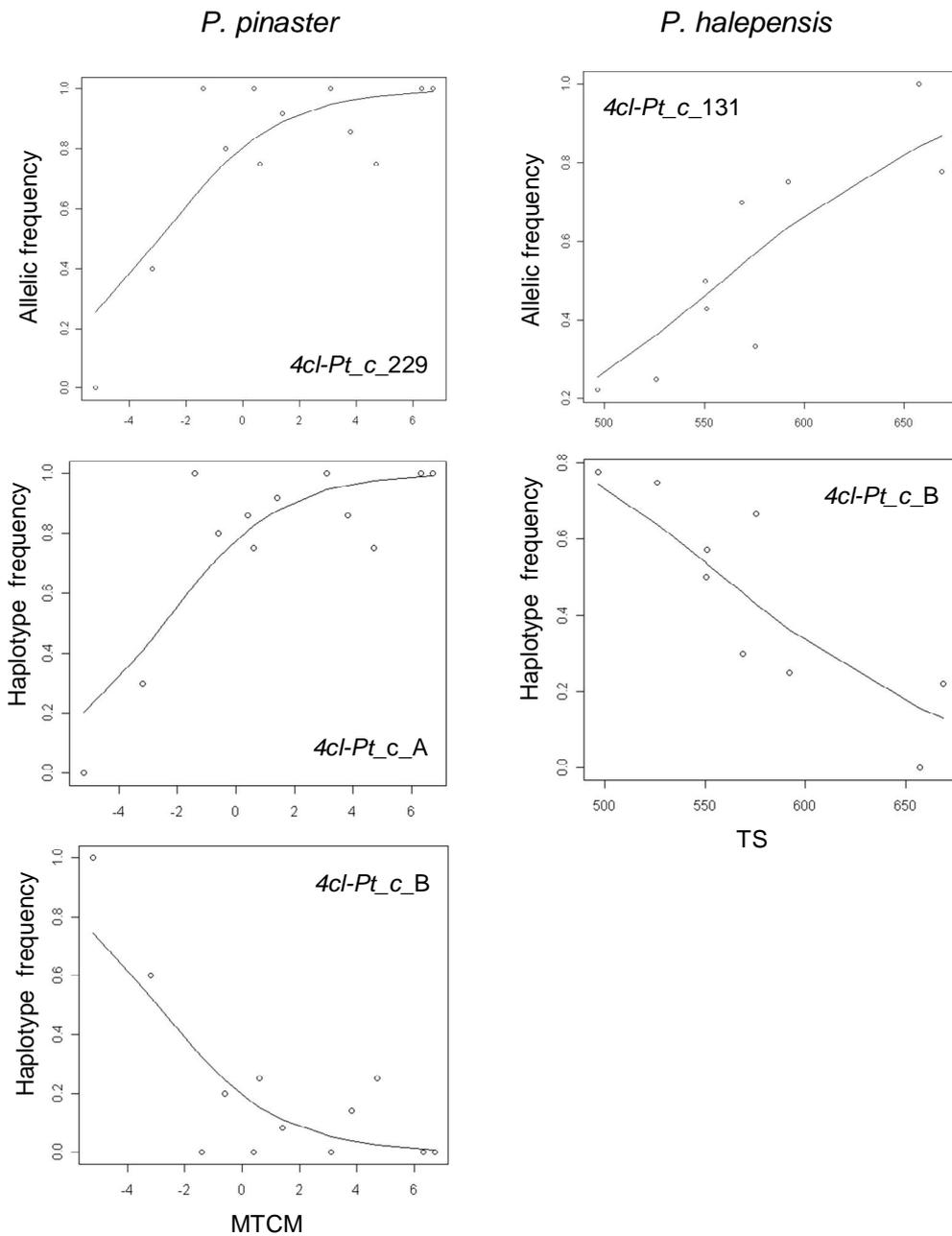


Figure 3. Upper section: *dhn5-Ps* haplotype distribution showing the proportion of haplotypes C (in grey), E (in black) and all others (in white) in *P. pinaster*. The size of the pie is proportional to the sample size at each location. **Lower section:** expected (sigmoid curves) and observed covariation in haplotype frequencies of *dhn5-Ps_C* and *dhn5-Ps_E* and climatic variables. Expected curves were obtained under a logistic regression model including a single environmental variable (AMT or MTWM) after controlling for neutral processes (i.e. using the PCs of neutral molecular markers as covariates) for *P. pinaster*. AMT: annual mean temperature; MTWM: maximum temperature of the warmest month.

