SUPPORTING INFORMATION for

Castellanos M.C., S. C. González-Martínez & J. G. Pausas. Field heritability of a plant adaptation to fire in heterogeneous landscapes. ***Molecular Ecology***, doi:10.1111/mec.13421

Appendix s1.

Estimation of genomic relatedness with *synbreed* (Wimmer *et al.* 2012):

Previous to relatedness estimation, missing genotypic data were imputed using the *codeGeno* command in *synbreed*; 804 out of 82942 genotypes were imputed for *P. halepensis*, and 570 out of 48694 in *P. pinaster*.

Genomic relatedness was estimated with the function *kin*, using the "realizedAB" method. The resulting G matrix is, by definition, semi-positive definite. For the animal models, G matrices were forced to be positive-definite with the command *make.positive.definite* in package *corpcor* (Schäfer *et al.* 2014). This command computes the nearest positive definite symmetric matrix, using the algorithm by Higham (1988). This transformation implies adding very small values to the original relatedness estimates, e.g. the addition of a values around $3*10^{-13}$, and has a negligible effect on the matrix structure.

Appendix s2.

Estimation of heritability (h^2) in *MCMCglmm* (Hadfield 2010); see Villemereuil (2012) and Wilson *et al.* (2010) for tutorials.

The unweighted, inverted positive-definite matrix was then used to provide the covariance structure for the random 'animal' predictor in animal models for serotiny (input as *ginverse* in *MCMCglmm*). By not weighting our G matrix, we are assuming that all sampled SNPs are equally linked to the genetic causal variants of the studied traits. We modelled serotiny as a binary trait (each cone per tree is coded as open or closed) using the "ordinal" family distribution. Modelling serotiny at the cone level prevents complications related to using a proportion as the response variable. For the prior distribution of variance components we used a parameter expanded prior with the χ^2 distribution with 1 degree of freedom, with the residual variance component fixed to 1, as shown here for two random effects:

```
prior <- list(R = list(V = 1, fix = 1),
G = list(G1 = list(V = 1, nu = 1000, alpha.mu = 0, alpha.V = 1),
G2 = list(V = 1, nu = 1000, alpha.mu = 0, alpha.V = 1)))
```

For the estimation of h^2 in binary traits, the usual priors with long tails such as the inverse-Gamma distribution often bias the results towards $h^2 = 1$, while χ2 expanded priors have the closest cumulative distribution to a uniform distribution on heritability (Villemereuil *et al.* 2013).

MCMC chains for binary models often result in high levels of autocorrelation for variance components, so we let MCMCglmm run for 4 million iterations with a high thinning interval of 2000 after a burn-in of 10 000. This resulted in an effective sample size of 2 000. Posterior modes and means of h^2 estimations were very similar to each other, as expected, and modes are used throughout the paper.

Appendix s4.

Comparison of relatedness estimates calculated with a large SNP dataset

For the test explained below, we used the genotypes from 1745 SNPs in Jaramillo-Correa *et al.* (2015) for *Pinus pinaster*, chosen from a larger Illumina Infinium assay of SNPs by Chancerel *et al.* (2013). The dataset includes randomly-chosen SNPs detected from transcriptome sequence data (from various tissues without prior experimental treatment). The chosen SNPs are expected to be neutral variants. A total of 150 trees were included in this dataset, all from Eslida, one of the study stands in our study.

To test for the robustness of relatedness estimates based on 250 SNP markers (as in our study), we calculated the correlation between the G matrix calculated from 250 SNPs randomly chosen from the total 1745 SNPs and the G matrix calculated with the remaining 1481 SNPs. This was iterated 100 times. G matrices were estimated with *synbreed* as explained above, after removing loci with minimum allele frequency < 1% and randomly imputing missing genotypes. The correlations between pairs of G matrices averaged $r = 0.79 \pm 0.006$ SD.

Appendix s3.

Regional genetic structure- STRUCTURE analysis

Before performing the heritability estimates, we tested for population genetic structure and differentiation among stands within each study species using microsatellite genotypes. Because serotiny varies across the study area, it is essential to ensure that the study of heritability is performed within a genetically homogenous group of stands that are interconnected, i.e., that associations between phenotypes and genotypes are not determined by population structure. The analysis was based on nine nuclear microsatellite markers for *Pinus halepensis* (data in Budde 2014a, b) and eleven for *P. pinaster* (Budde et al. 2014), using the Bayesian clustering approach implemented in the software STRUCTURE v. 2.3.4 (Pritchard *et al.* 2000). This method attempts to assign individuals to the optimal number of *K* genetic clusters based on allele frequencies at each locus. It is not designed to deal with populations with high isolation-by-distance, but this is likely not the case for stands of wind-pollinated *Pinus* species, where gene flow is expected to reach long distances and maintain large natural populations (De-Lucas *et al.* 2009; Steinitz *et al.* 2011). For *P. halepensis*, we ran simulations including stand identifiers and with the LOCPRIOR option, to make sure that even weak genetic structuring could be detected. Simulation runs calculated the likelihood of clustering in $K = 1$ to 9 subpopulations (one more than the actual sampled stands), and were run for 2×10^5 iterations after a 1×10^4 burn-in period. Ten such runs were carried out for each value of *K*. The results were visualized with STRUCTURE HARVESTER (Earl & vonHoldt 2012). As shown in Fig. S1 below, the analysis showed no evidence for $K > 1$, and individuals appeared as strongly admixed. Analogous STRUCTURE simulations with the same results for *P. pinaster* stands are detailed in Budde *et al.* (2014).

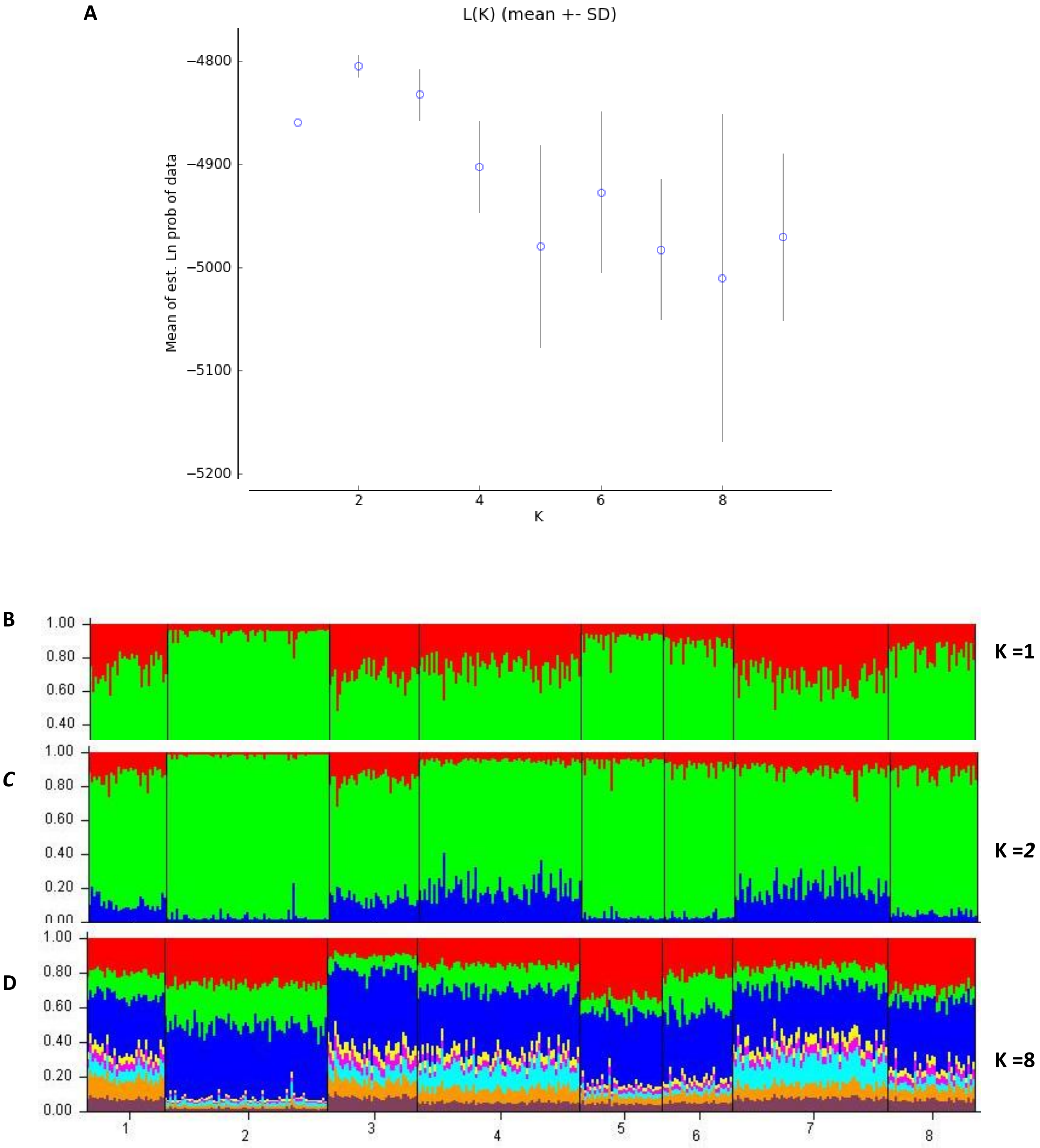

Figure (appendix s4). Results of STRUCTURE analysis for *Pinus halepensis* stands, based on nine nuclear microsatellites. A. The probabilities (LnP) for K = 2 and 3 are slightly higher than for K=1, but example figures in B and C show no differentiation into two or three genetic groups. D shows an example of the data with K=8, the actual number of stands, showing strong admixture.

References

Budde KB (2014a) Data from: Effects of fire regime on the population genetics of natural pine stands, in Genetic structure of forest trees in biodiversity hotspots at different spatial scales (Ph.D. thesis). *Dryad Digital Repository* **http://dx.doi.org/10.5061/dryad.n193d**.

Budde KB (2014b) *Genetic structure of forest trees in biodiversity hotspots at different spatial scales*, PhD Thesis, Complutense University of Madrid. www.worldcat.org/oclc/896778059

Budde KB, Heuertz M, Hernández-Serrano A, *et al.* (2014) In situ genetic association for serotiny, a fire-related trait, in Mediterranean maritime pine (*Pinus pinaster*). *New Phytologist* **201**, 230-241.

Chancerel E, Lamy J-B, Lesur I, *et al.* (2013) High-density linkage mapping in a pine tree reveals a genomic region associated with inbreeding depression and provides clues to the extent and distribution of meiotic recombination. *BMC Biology* **11**, 50.

De-Lucas AI, González-Martínez SC, Vendramin GG, Hidalgo E, Heuertz M (2009) Spatial genetic structure in continuous and fragmented populations of *Pinus pinaster* Aiton. *Molecular Ecology* **18**, 4564-4576.

Earl D, vonHoldt B (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359-361.

Hadfield JD (2010) MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software* **33**, 1-22.

Higham NJ (1988) Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and its Applications* **103**, 103-118.

Jaramillo-Correa JP, Grivet D, Lepoittevin C, *et al.* (2015) Molecular proxies of climate maladaptation in a long-lived tree (*Pinus pinaster* Aiton, Pinaceae). *Genetics* **199**, 793-807.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155**, 945-959.

Schäfer J, Opgen-Rhein R, Zuber V, *et al.* (2014) corpcor: Efficient Estimation of Covariance and (Partial) Correlation. R package version 1.6.7., http://CRAN.R-project.org/package=corpcor.

Steinitz O, Troupin D, Vendramin GG, Nathan R (2011) Genetic evidence for a Janzen–Connell recruitment pattern in reproductive offspring of Pinus halepensis trees. *Molecular Ecology* **20**, 4152-4164.

Villemereuil P (2012) *Estimation of a biological trait heritability using the animal model. How to use the MCMCglmm R package.*, http://devillemereuil.legtux.org/downloads/

Villemereuil P, Gimenez O, Doligez B (2013) Comparing parent–offspring regression with frequentist and Bayesian animal models to estimate heritability in wild populations: a simulation study for Gaussian and binary traits. *Methods in Ecology and Evolution* **4**, 260-275.

Wilson AJ, Réale D, Clements MN, *et al.* (2010) An ecologist's guide to the animal model. *Journal of Animal Ecology* **79**, 13-26.

Wimmer V, Albrecht T, Auinger H, Schoen C (2012) *synbreed*: a framework for the analysis of genomic prediction data using R. *Bioinformatics* **28**, 2086-2087.

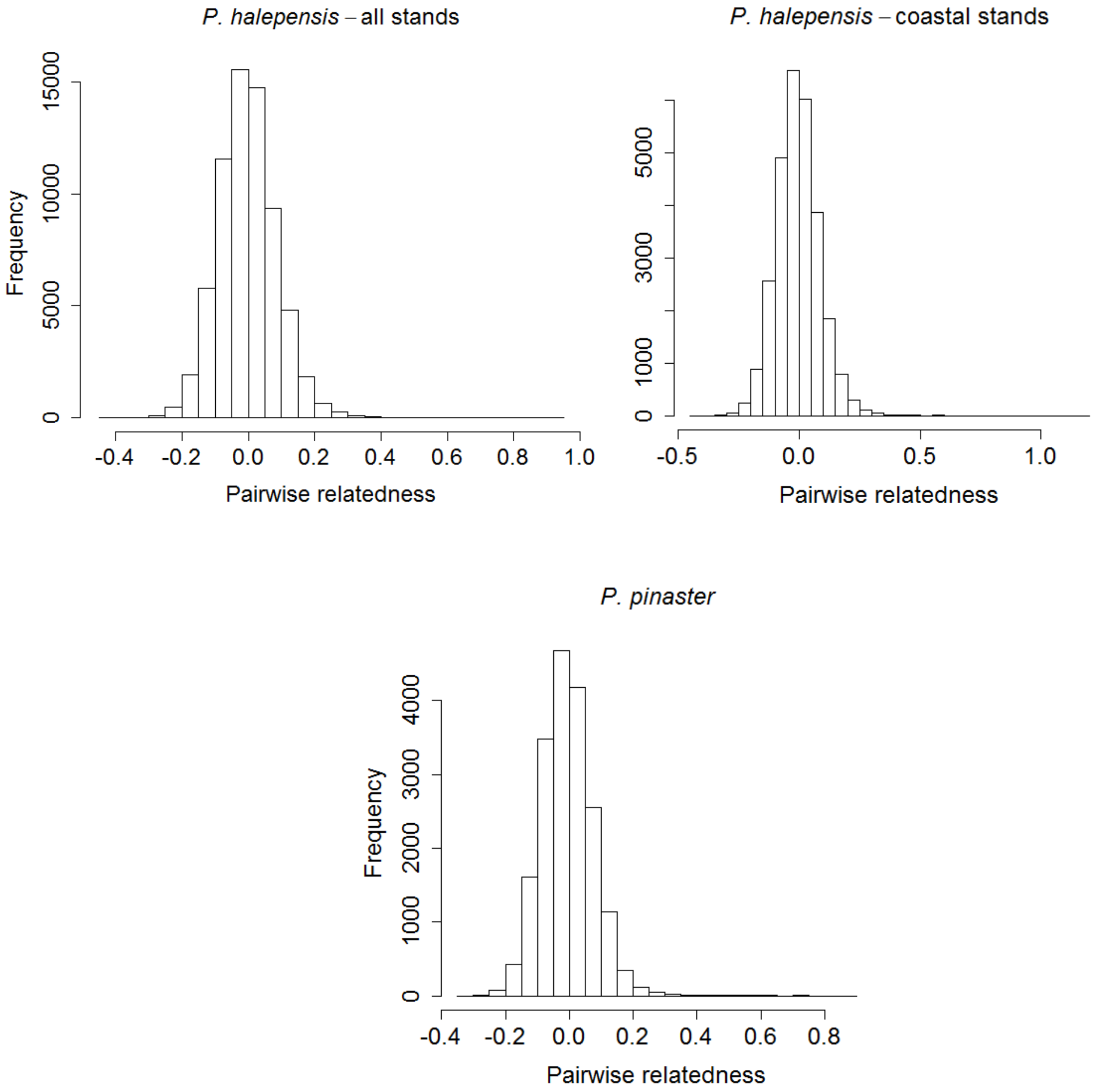

Fig. S1. Histograms with pairwise relatedness estimates (e.g. the values in G matrices) for the three data sets analyzed.

Table S1. Location and characteristics of the study stands of *Pinus halepensis and P. pinaster.* Habitat, crown-fire recurrence, number of sampled trees, geographic coordinates, mean (± SD) elevation, mean annual temperature (T), total annual precipitation (P), mean (± SD) diameter at breast height (DBH), mean (± SD) percentage of serotiny, and mean pairwise genomic relatedness for each study stand.

Stand	**Habitat**	**Crown-fire recurrence**	**n trees**	**Latitude (º)**	**Longitude (º)**	**Elevation (m)**	**T (°C)**	**P (mm)**	**DBH (cm)**	**% Serotiny** (± SD)	**Pairwise genomic relatedness**	
											mean	**range**
P. halepensis											-0,003	-0.409 to 0.905
Alzira	Coastal	high	39	39.12	-0.39	155 ± 13.9	16.9	512	26.5 ± 4.47	54.8 ± 17.96	0.019	-0.296 to 0.344
Cabanes	Coastal	high	31	40.10	0.04	449 ± 27.6	15.2	647	29.5 ± 9.44	42.6 ± 21.98	0.089	-0.183 to 0.520
Serra Calderona	Coastal	high	62	39.74	-0.48	711 ± 33.3	13.9	574	28.2 ± 7.16	52.3 ± 16.07	0.013	-0.282 to 0.905
Eslida	Coastal	high	67	39.87	-0.29	508 ± 57.5	14.9	589	27.6 ± 5.47	52.7 ± 16.92	0.057	-0.340 to 0.563
Serra d'Irta	Coastal	high	39	40.35	0.32	361 ± 33.1	15.7	694	25.5 ± 3.98	61.4 ± 12.09	0.048	-0.218 to 0.349
Montan	Inland	low	31	40.05	-0.59	838 ± 29.2	13.3	576	27.6 ± 4.49	30.4 ± 14.88	0.027	-0.214 to 0.626
Sinarcas	Inland	low	65	39.8	-1.20	912 ± 10.2	13.0	470	32.4 ± 6.33	26.9 ± 14.76	0.019	-0.276 to 0.414
Titaguas	Inland	low	33	39.89	-1.13	904 ± 20.4	13.0	485	26.9 ± 3.85	25.1 ± 11.63	0.012	-0.243 to 0.383
P. pinaster											-0.005	-0.304 to 0.884
Serra Calderona	Costal	high	64	39.75	-0.50	807 ± 15.9	13.5	582	21.6 ± 3.28	57.8 ± 19.61	0.023	-0.234 to 0.720
Eslida	Costal	high	66	39.88	-0.30	437 ± 18.8	15.2	580	27.1 ± 3.58	29.0 ± 14.16	0.023	-0.226 to 0.570
Sinarcas	Inland	low	64	39.79	-1.20	885 ± 8.3	13.1	467	33.6 ± 4.19	21.7 ± 18.39	0.019	-0.267 to 0.884

Table S2. Heritability estimates (h^2) for cone serotiny in *Pinus halepensis* and *P. pinaster*, along with their credible intervals (CI) and deviance information criterion values (DIC) for complete final models, "naïve" models with no spatial or environmental predictors (as in Table 2 in main text), and models where the relatedness among individuals ("id") was removed. In this later case, h^2 values cannot be estimated but DIC values are reported.

Dataset	model		h^2	CI	DIC
***P. halepensis*, all stands**	full	serot ~ ta + diam + rec, random= id + stand	0.10	0.073- 0.142	14211.97
	naïve	serot ~ 1, random= id	0.25	0.193 - 0.296	14396.24
	no relatedness	serot ~ ta + diam + rec, random= stand	-	-	14689.42
***P. halepensis*, coastal stands**	full	serot ~ ta + diam, random= id + stand	0.14	0.087 - 0.201	10577.92
	naïve	serot ~ 1, random= id	0.19	0.146 - 0.258	10597.39
	no relatedness	serot ~ ta + diam, random= stand	-	-	11097.04
***P. pinaster*, all stands**	full	serot ~ ta + diam + rec, random= id + stand	0.12	0.054- 0.234	6564.94
	naïve	serot ~ 1, random= id	0.33	0.272 - 0.415	6565.15
	no relatedness	serot ~ ta + diam + rec, random= stand	-	-	7063.56

serot= serotiny, coded at the cone level as a binary variable; ta= mean annual temperature, diam= tree diameter at breast heigh, rec=fire recurrence, id= individual identifier (='animal') with covariance structure provided by their pairwise relatedness matrix.