MOLECULAR ECOLOGY

Supplemental Information for:

Geographic isolation versus dispersal: Relictual alpine grasshoppers support a model of interglacial diversification with limited hybridization

Joaquín Ortego and L. Lacey Knowles

Contents:

Supplementary methods

METHODS S1 Genomic data filtering and assembling

Raw sequences were demultiplexed and preprocessed using STACKS v. 1.35 (Catchen, Amores, Hohenlohe, Cresko, & Postlethwait, 2011; Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013; Hohenlohe et al., 2010) and assembled using PYRAD v. 3.0.66 (Eaton, 2014). Libraries were demultiplexed and filtered for overall quality using *process_radtags* (Catchen et al., 2011, 2013), retaining reads with a Phred score > 10 (using a sliding window of 15%), no adaptor contamination, and that had an unambiguous barcode and restriction cut site. Raw sequence data quality was checked in FASTQC v. 0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and sequences were trimmed to 129 bp using SEQTK (Heng Li, https://github.com/lh3/seqtk) in order to remove barcodes and low-quality reads near the 3´ ends. We assembled our sequences into *de novo* loci using PYRAD v. 3.0.66 (Eaton, 2014). Briefly, reads retained after *process_radtags* were further quality-filtered with PYRAD to convert base calls with a Phred score <20 into *N*s and discard reads with >2 *N*s. Retained reads were clustered within- and across samples considering two different thresholds of sequence similarity (W_{clust} = 85 and 90%) and clusters with a coverage depth <5 were discarded. Loci containing one or more heterozygous sites across more than 5 samples (~10% of individuals) were excluded, as we expect that this represents a fixed difference among clustered paralogs rather than a true polymorphism (Eaton, 2014; e.g., Eaton, Hipp, González-Rodríguez, & Cavender-Bares, 2015). In a final filtering step, we excluded loci that were not recovered in at least 10, 20 or 30 samples (*minCov* = 10, 20 and 30). The choice of different filtering thresholds had little impact on the obtained inferences (see Results; e.g., Eaton et al., 2015; Ortego, Gugger, & Sork, 2018). For this reason, unless otherwise indicated, all downstream analyses were performed using a SNP dataset obtained with PYRAD considering a clustering threshold of sequence similarity of 0.85 (W_{clust} = 0.85) and discarding loci that were not present in at least 20 individuals (*minCov* = 20).

Supplementary tables

TABLE S1 Geographical location and number of analyzed individuals (*n*) for the studied populations of *Podisma pedestris*, *P. carpetana* and *P. cantabricae*. *Cophopodisma pyrenaea* was used as an outgroup in phylogenomic analyses.

Species	Locality	Code	*n*	Latitude	Longitude
P. pedestris	Aigüestortes	AIG	7	42.599790	0.980960
P. pedestris	Lac d'Aule	AUL	7	42.878730	-0.462730
P. carpetana ignatii	Sierra del Moncayo	MON	7	41.783860	-1.836900
P. carpetana ignatii	Sierra de Urbión	URB	4	42.006440	-2.867267
P. carpetana ignatii	Sierra de la Demanda	DEM	3	42.261250	-2.994350
P. carpetana carpetana	Sierra de Guadarrama	GUA	7	40.807840	-3.875100
P. carpetana ignatii	Picos de Europa	EUR	6	43.212280	-4.713650
P. cantabricae	Huertos del Diablo	DIA	7	43.064320	-5.979070
C. pyrenaea	Setcases	SET	7	42.427560	2.266300

TABLE S2 Distribution of topologies (%) contained in 95% highest posterior density (HPD) tree sets reconstructed with SNAPP. Population codes are described in Table S1.

Tree no.	%	Tree
1	60.74	(((AIG,AUL),((((MON,(URB,DEM)),GUA),EUR),DIA)),SET)
2	32.76	(((AIG,AUL),(((((MON,URB),DEM),GUA),EUR),DIA)),SET)
3	3.96	(((AIG,AUL),(((((MON,DEM),URB),GUA),EUR),DIA)),SET)

TABLE S3 Analyses of introgression using four-taxon *D*-statistic (ABBA/BABA) tests. Analyses were performed for each of the six species-population combinations separately and using six genomic datasets generated in PYRAD by setting different clustering thresholds (W_{clust} = 0.85 and 0.90) and values for minimum taxon coverage (*minCov* = 10, 20 and 30). *Cophopodisma pyrenaea* was used as an outgroup. All tests were highly significant (q < 0.001) after a false discovery rate (FDR) adjustment (5%) to control for multiple tests. Population codes are described in Table S1.

P1 (*P. cantabricae)*	P2 (*P. carpetana*)	P3 (*P. pedestris)*	*n*	BABA	ABBA	*D* (± S.D.)	*z*	*q*
W_{clust} = 0.85; *minCov* = 10								
DIA	EUR	AUL	5,210	935	1,353	0.183 ± 0.027	6.68	<0.001
DIA	MON	AUL	4,713	774	1,319	0.260 ± 0.027	9.61	<0.001
DIA	GUA	AUL	4,634	809	1,314	0.238 ± 0.029	8.12	<0.001
DIA	EUR	AIG	5,307	939	1,391	0.194 ± 0.027	7.11	<0.001
DIA	MON	AIG	4,806	785	1,367	0.271 ± 0.026	10.54	<0.001
DIA	GUA	AIG	4,708	822	1,325	0.234 ± 0.028	8.52	<0.001
W_{clust} = 0.85; *minCov* = 20								
DIA	EUR	AUL	3,772	595	932	0.221 ± 0.031	7.08	<0.001
DIA	MON	AUL	3,601	546	990	0.289 ± 0.031	9.29	<0.001
DIA	GUA	AUL	3,553	577	948	0.244 ± 0.033	7.45	<0.001
DIA	EUR	AIG	3,808	592	961	0.238 ± 0.031	7.70	<0.001
DIA	MON	AIG	3,640	551	1,026	0.301 ± 0.031	9.86	<0.001
DIA	GUA	AIG	3,587	566	962	0.259 ± 0.031	8.46	<0.001
W_{clust} = 0.85; *minCov* = 30								
DIA	EUR	AUL	1,767	210	360	0.264 ± 0.045	5.87	<0.001
DIA	MON	AUL	1,774	213	413	0.319 ± 0.044	7.24	<0.001
DIA	GUA	AUL	1,760	213	409	0.314 ± 0.043	7.37	<0.001
DIA	EUR	AIG	1,765	218	370	0.260 ± 0.043	6.00	<0.001
DIA	MON	AIG	1,772	219	418	0.313 ± 0.040	7.74	<0.001
DIA	GUA	AIG	1,758	218	416	0.312 ± 0.041	7.51	<0.001
W_{clust} = 0.90; *minCov* = 10								
DIA	EUR	AUL	5,771	1,004	1,453	0.183 ± 0.026	6.94	<0.001
DIA	MON	AUL	5,233	810	1,472	0.290 ± 0.024	11.90	<0.001
DIA	GUA	AUL	5,162	846	1,437	0.259 ± 0.026	10.02	<0.001
DIA	EUR	AIG	5,888	1,013	1,495	0.192 ± 0.025	7.64	<0.001
DIA	MON	AIG	5,344	829	1,514	0.293 ± 0.025	11.87	<0.001
DIA	GUA	AIG	5,253	856	1,468	0.263 ± 0.026	10.04	<0.001
W_{clust} = 0.90; *minCov* = 20								
DIA	EUR	AUL	4,101	627	994	0.227 ± 0.030	7.57	<0.001
DIA	MON	AUL	3,941	570	1,071	0.305 ± 0.030	10.18	<0.001
DIA	GUA	AUL	3,905	586	1,023	0.272 ± 0.030	8.93	<0.001
DIA	EUR	AIG	4,155	632	1,022	0.236 ± 0.031	7.61	<0.001
DIA	MON	AIG	4,001	576	1,098	0.312 ± 0.028	11.16	<0.001
DIA	GUA	AIG	3,958	584	1,029	0.276 ± 0.030	9.34	<0.001
W_{clust} = 0.90; *minCov* = 30								

DIA	EUR	AUL	1,923	232	398	0.263 ± 0.044	5.91	<0.001
DIA	MON	AUL	1,928	237	456	0.317 ± 0.043	7.28	<0.001
DIA	GUA	AUL	1,914	232	450	0.320 ± 0.043	7.38	<0.001
DIA	EUR	AIG	1,923	246	404	0.244 ± 0.043	5.67	<0.001
DIA	MON	AIG	1,928	248	463	0.303 ± 0.043	7.08	<0.001
DIA	GUA	AIG	1,914	241	455	0.307 ± 0.043	7.09	<0.001

n, number of retained SNPs; *D* (± S.D.), *D*-statistic and corresponding standard deviation; *z*, *z*-statistic; *q*, *p*-values adjusted at a FDR of 5%

TABLE S4 Comparison of alternative species divergence models (detailed in Figure S1) tested using FASTSIMCOAL2, with best supported models highlighted in bold. Analyses were performed for each of the six species-population combinations separately. Population codes are described in Table S1. The number of loci retained for the calculation of the SFS is indicated in parentheses.

	lnL	*k*	AIC	ΔAIC	ω_i	lnL	*k*	AIC	ΔAIC	ω_i
Model	Populations: AUL-EUR-DIA (2,353 loci)					Populations: AIG-EUR-DIA (1,991 loci)				
A	-3,358.70	10	6,737.39	2.27	0.24	-2,862.65	10	5,745.30	1.63	0.30
B	**-3,358.56**	**9**	**6,735.12**	**0.00**	**0.76**	**-2,862.84**	**9**	**5,743.67**	**0.00**	**0.67**
C	-3,368.18	9	6,754.36	19.24	0.00	-2,867.56	9	5,753.12	9.45	0.01
D	-3,372.46	10	6,764.92	29.80	0.00	-2,869.95	10	5,759.91	16.24	0.00
E	-3,376.99	9	6,771.97	36.85	0.00	-2,872.41	9	5,762.81	19.14	0.00
F	-3,373.20	9	6,764.40	29.28	0.00	-2,871.71	9	5,761.43	17.75	0.00
G	-3,364.65	10	6,749.29	14.17	0.00	-2,866.50	10	5,752.99	9.32	0.01
H	-3,367.84	9	6,753.68	18.56	0.00	-2,870.45	9	5,758.90	15.23	0.00
I	-3,370.07	9	6,758.14	23.02	0.00	-2,870.49	9	5,758.98	15.30	0.00
J	-3,369.33	6	6,750.67	15.55	0.00	-2,869.59	6	5,751.17	7.50	0.02
	Populations: AUL-MON-DIA (819 loci)					Populations: AIG-MON-DIA (710 loci)				
A	-1,177.62	10	2,375.25	2.83	0.18	-1,014.50	10	2,048.99	3.57	0.08
B	**-1,177.21**	**9**	**2,372.41**	**0.00**	**0.76**	**-1,013.71**	**9**	**2,045.42**	**0.00**	**0.48**
C	-1,185.11	9	2,388.21	15.80	0.00	-1,016.05	9	2,050.10	4.69	0.05
D	-1,185.39	10	2,390.78	18.37	0.00	-1,015.74	10	2,051.48	6.06	0.02
E	-1,186.94	9	2,391.88	19.47	0.00	-1,016.83	9	2,051.66	6.24	0.02
F	-1,184.59	9	2,387.18	14.76	0.00	-1,016.25	9	2,050.50	5.08	0.04
G	-1,180.87	10	2,381.74	9.33	0.01	-1,015.69	10	2,051.38	5.97	0.02
H	-1,183.91	9	2,385.81	13.40	0.00	-1,018.24	9	2,054.48	9.07	0.01
I	-1,183.76	9	2,385.52	13.10	0.00	-1,019.00	9	2,056.00	10.58	0.00
J	-1,183.03	6	2,378.06	5.65	0.05	-1,017.27	6	2,046.53	1.12	0.28
	Populations: AUL-GUA-DIA (1,393 loci)					Populations: AIG-GUA-DIA (1,247 loci)				
A	-1,991.36	10	4,002.73	1.37	0.33	-1,803.52	10	3,627.03	2.27	0.23
B	**-1,991.68**	**9**	**4,001.36**	**0.00**	**0.66**	**-1,803.38**	**9**	**3,624.76**	**0.00**	**0.72**
C	-2,000.01	9	4,018.02	16.66	0.00	-1,807.44	9	3,632.88	8.12	0.01
D	-1,999.89	10	4,019.77	18.42	0.00	-1,808.04	10	3,636.08	11.32	0.00
E	-2,004.84	9	4,027.68	26.32	0.00	-1,809.83	9	3,637.66	12.90	0.00
F	-2,003.01	9	4,024.03	22.67	0.00	-1,811.49	9	3,640.98	16.22	0.00
G	-1,995.64	10	4,011.27	9.91	0.00	-1,805.76	10	3,631.52	6.76	0.02
H	-2,003.56	9	4,025.11	23.75	0.00	-1,811.93	9	3,641.85	17.09	0.00
I	-2,002.81	9	4,023.61	22.25	0.00	-1,812.26	9	3,642.52	17.76	0.00
J	-2,002.79	6	4,017.58	16.23	0.00	-1,811.31	6	3,634.62	9.86	0.01

lnL, maximum likelihood estimate of the model; *k*, number of parameters in the model; AIC, Akaike's information criterion value; ΔAIC, difference in AIC value from that of the strongest model; ω_i, AIC weight.

TABLE S5 Demographic parameters inferred with FASTSIMCOAL2 for the most likely species divergence model (Model B, illustrated in Figure S1). Table shows point estimates and lower and upper 95% confidence intervals for each parameter: the ancestral (θ_{ANC}, $\theta_{CAN\text{-}CAR}$) and contemporary (θ_{PED}, θ_{CAR}, θ_{CAN}) effective population sizes, migration rates (*m*), timing of species split (T_{DIV1}, T_{DIV2}), and timing (beginning and end) of interspecific gene flow ($T_{INTROG1}$, $T_{INTROG2}$), with time given in units of generations (or years, with 1 generation per year). Analyses were performed for each of the six species-population combinations separately. Population codes are described in Table S1.

Parameter	Point estimate	Lower bound	Upper bound	Point estimate	Lower bound	Upper bound
	Populations: AUL-EUR-DIA			Populations: AIG-EUR-DIA		
θ_{ANC}	155,653	29,151	418,990	365,207	16,508	469,512
$\theta_{CAR\text{-}CAN}$	127,882	47,710	194,084	119,061	56,916	399,925
θ_{PED}	46,363	39,453	55,707	59,827	49,237	73,318
θ_{CAR}	64,394	60,041	79,249	67,878	60,268	83,764
T_{DIV1}	992,317	178,271	1,086,881	637,958	181,899	921,786
T_{DIV2}	135,537	132,699	169,645	148,149	138,496	177,519
$T_{INTROG1}$	108,047	106,212	158,920	140,573	98,998	167,235
$T_{INTROG2}$	104,545	74,192	132,094	119,116	86,690	151,569
$m_{CAR\leftarrow PED}$	7.75×10^{-04}	1.00×10^{-05}	8.44×10^{-04}	1.34×10^{-04}	1.13×10^{-05}	2.65×10^{-03}
	Populations: AUL-MON-DIA			Populations: AIG-MON-DIA		
θ_{ANC}	319,411	26,417	475,277	264,395	26,319	442,362
$\theta_{CAR\text{-}CAN}$	139,221	68,231	328,598	117,867	43,202	355,943
θ_{PED}	50,083	41,548	75,281	62,154	46,577	77,687
θ_{CAR}	57,362	49,036	83,244	61,875	50,281	82,257
T_{DIV1}	765,354	181,493	1,027,446	993,133	176,183	856,788
T_{DIV2}	131,434	123,170	185,097	141,661	132,731	192,455
$T_{INTROG1}$	119,742	86,114	167,863	114,038	77,799	176,863
$T_{INTROG2}$	87,035	59,249	131,918	105,837	77,048	134,862
$m_{CAR\leftarrow PED}$	8.37×10^{-05}	1.14×10^{-05}	8.50×10^{-04}	3.97×10^{-04}	1.84×10^{-05}	4.25×10^{-03}
	Populations: AUL-GUA-DIA			Populations: AIG-GUA-DIA		
θ_{ANC}	254,765	33,257	365,025	253,670	27,857	408,976
$\theta_{CAR\text{-}CAN}$	153,194	44,365	325,856	149,143	67,931	365,232
θ_{PED}	53,842	44,056	70,451	70,936	59,326	89,204
θ_{CAR}	57,736	51,028	78,886	60,759	54,166	79,820
T_{DIV1}	829,739	185,338	897,286	813,984	189,836	956,447
T_{DIV2}	143,207	134,614	188,693	155,922	145,880	203,562
$T_{INTROG1}$	133,485	105,901	177,337	147,312	105,473	191,253
$T_{INTROG2}$	106,278	65,836	140,233	120,584	68,674	161,032
$m_{CAR\leftarrow PED}$	9.05×10^{-05}	5.02×10^{-06}	6.96×10^{-04}	1.09×10^{-04}	3.25×10^{-06}	1.03×10^{-03}

TABLE S6 Environmental niche modeling (ENM) for *Podisma pedestris*, *P. carpetana* and *P. cantabricae*. Table shows the parameters of the best species-specific model and the variables retained sorted from higher to lower values of permutation importance. Variables in bold are those that cumulatively contributed > 50% to the model based on the permutation importance statistic.

Species	*n*	*FC*	*RM*	AUC_{TEST}	OR_{MTP}	MTSS	Environmental variables
P. pedestris	34	T	1	0.988	0.00	0.016	**BIO5**,BIO17,BIO11,BIO9,BIO4
P. carpetana	36	T	1	0.997	0.00	0.366	**BIO1**,**BIO15**,BIO18,BIO6,BIO16,BIO3, BIO4,BIO5,BIO9,BIO8
P. cantabricae	5	LPQ	0	1.000	0.00	0.813	**BIO4**,**BIO16**,**BIO8**,**BIO18**,BIO15,BIO3, BIO2,BIO11,BIO5,BIO9

n, number of occurrence records used for ENM; *FC*, feature class for variable transformation; *RM*, regularization multiplier; AUC_{TEST}, area under the curve of the receiver-operating characteristic plot on the testing data; OR_{MTP}, omission rate based on the minimum training presence threshold; MTSS, maximum training sensitivity plus specificity logistic threshold for species presence/absence

Supplementary figures

FIGURE S1 Alternative species divergence models tested using FASTSIMCOAL2. Parameters include ancestral (θ_{ANC}, $\theta_{CAN\text{-}CAR}$) and contemporary (θ_{PED}, θ_{CAR}, θ_{CAN}) effective population sizes, migration rates (*m*, arrows), timing of species split (T_{DIV1}, T_{DIV2}), and timing (beginning and end) of interspecific gene flow ($T_{INTROG1}$, $T_{INTROG2}$).

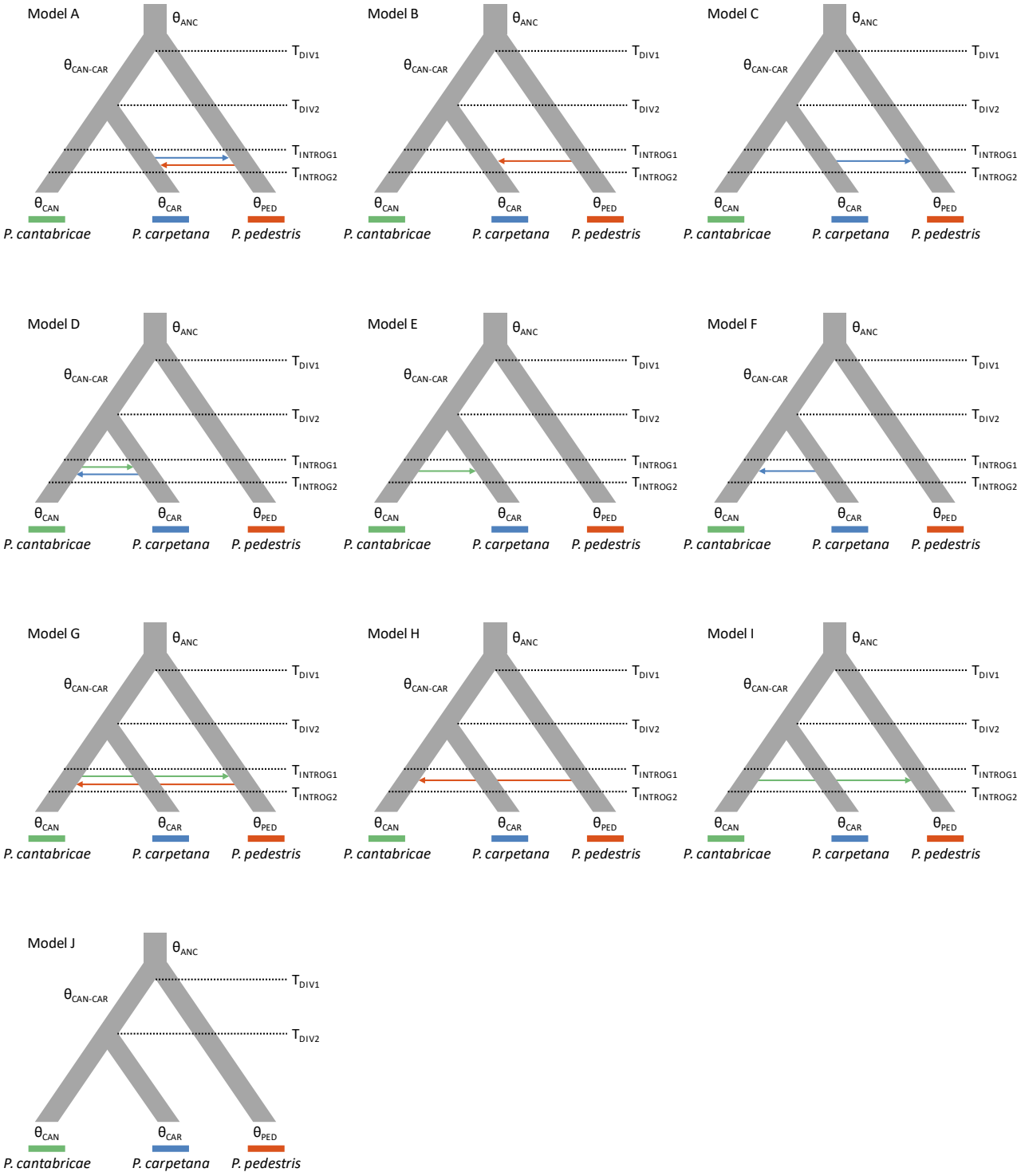


FIGURE S2 Number of reads per individual before and after different quality filtering steps by PYRAD. The cumulative stacked bars represent the total number of raw reads for each individual. Dark red color represents the reads that were discarded by *process_radtags* in STACKS due to low quality, adapter contamination or ambiguous barcode. Light red color represents the reads that were discarded during *step 2* in PYRAD after filtering out reads that did not comply with the quality criteria (reads with >2 sites with a Phred quality score < 20 were discarded). Green color represents the total number of retained reads used to identify homologous loci. Individuals are sorted by species and populations following the same order and codes presented in Table S1.

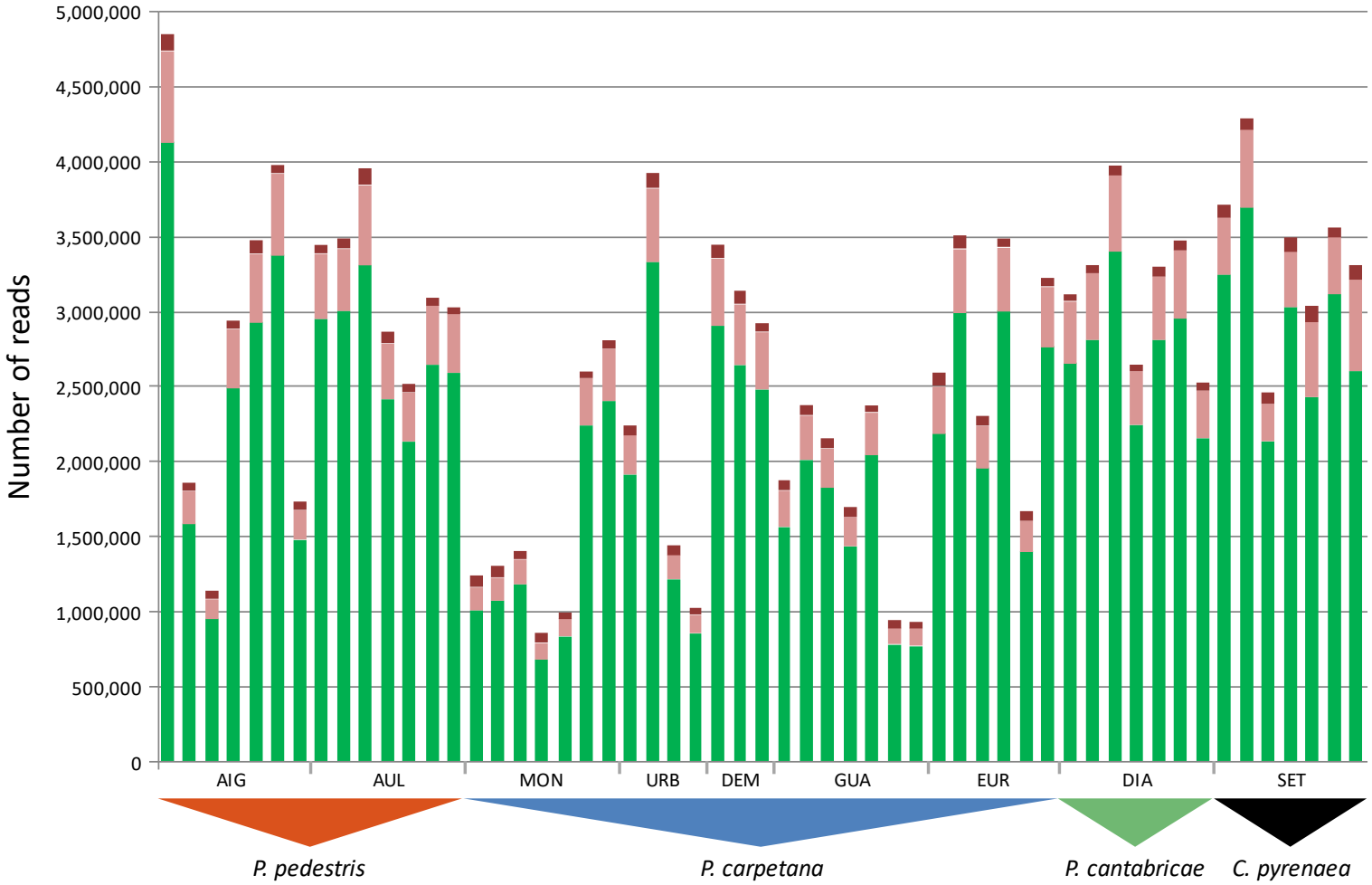


FIGURE S3 Mean (±SD) log probability of the data (LnPr(X|*K*)) over 10 runs of STRUCTURE (left axes, black dots and error bars) for each value of *K* and the magnitude of Δ*K* (right axes, blue dots). Hierarchical STRUCTURE analyses were run for (a) all species and independently for populations of (b) *Podisma pedestris* and (c) *P. carpetana*. Analyses are based on a random subset of 10,000 SNPs.

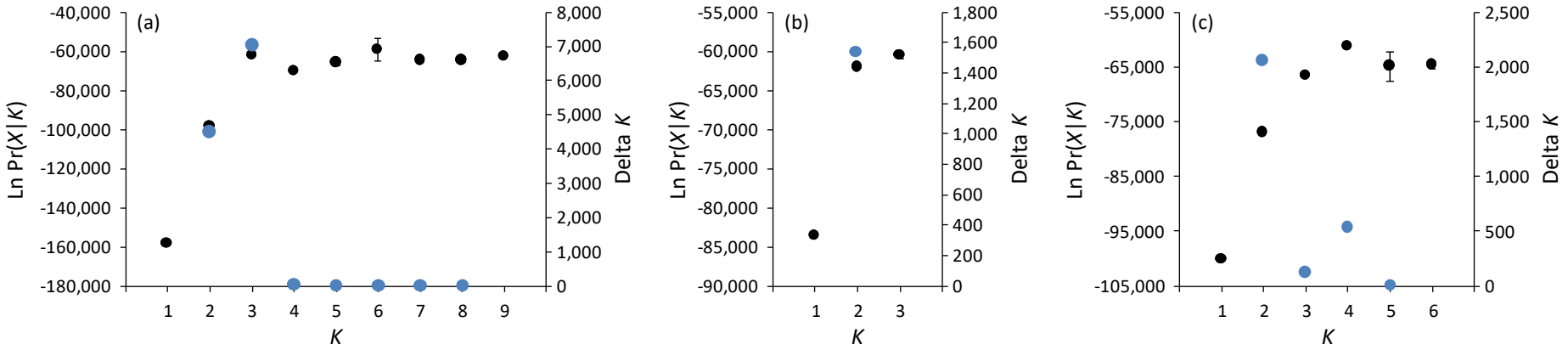


FIGURE S4 Principal component analyses (PCAs) of genetic variation for *Podisma pedestris*, *P. carpetana*, and *P. cantabricae*. Analyses were run for all (a) species (23,333 SNPs) and independently for populations of (b) *P. carpetana* (13,003 SNPs) and (c) *P. pedestris* (11,999 SNPs). Population codes are described in Table S1.

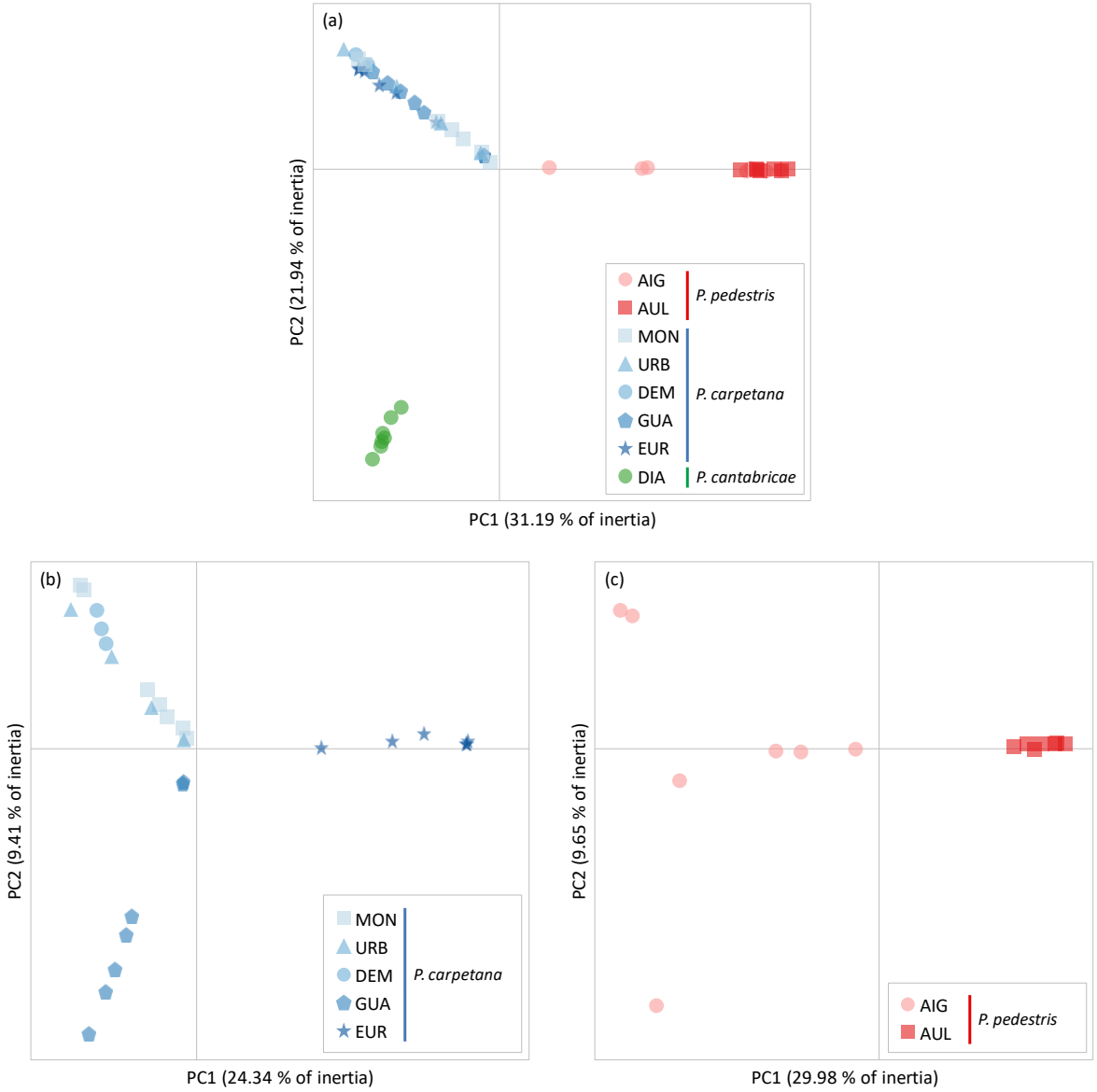


FIGURE S5 Phylogenetic trees inferred with SVDQUARTETS using six genomic datasets generated in PYRAD by setting different clustering thresholds (W_{clust} = 0.85 and 0.90) and values for minimum taxon coverage (*minCov* = 10, 20 and 30). *Cophopodisma pyrenaea* was used as an outgroup. Node colors indicate bootstrapping support (BS) values based on 100 replicates (green: BS > 95 %; orange: 95 % > BS > 90 %; red: BS < 90 %). The number of SNPs retained for each analysis is presented in parentheses. Population codes are described in Table S1.

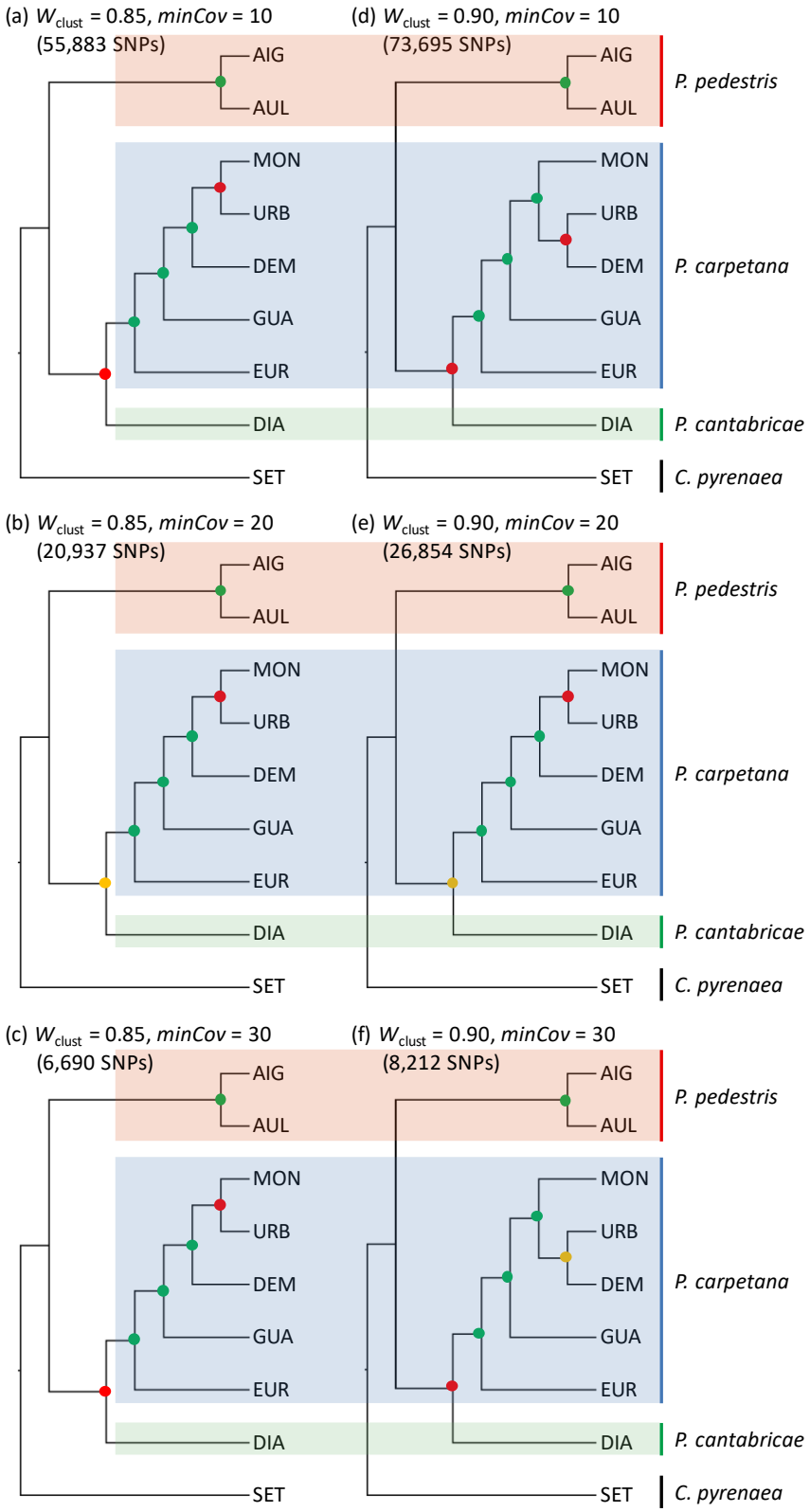


FIGURE S6 Summary of model fit with PHYLONETWORKS. The figure shows the negative log pseudo-likelihood for models with different number of introgression events (*h* from 0 to 5).

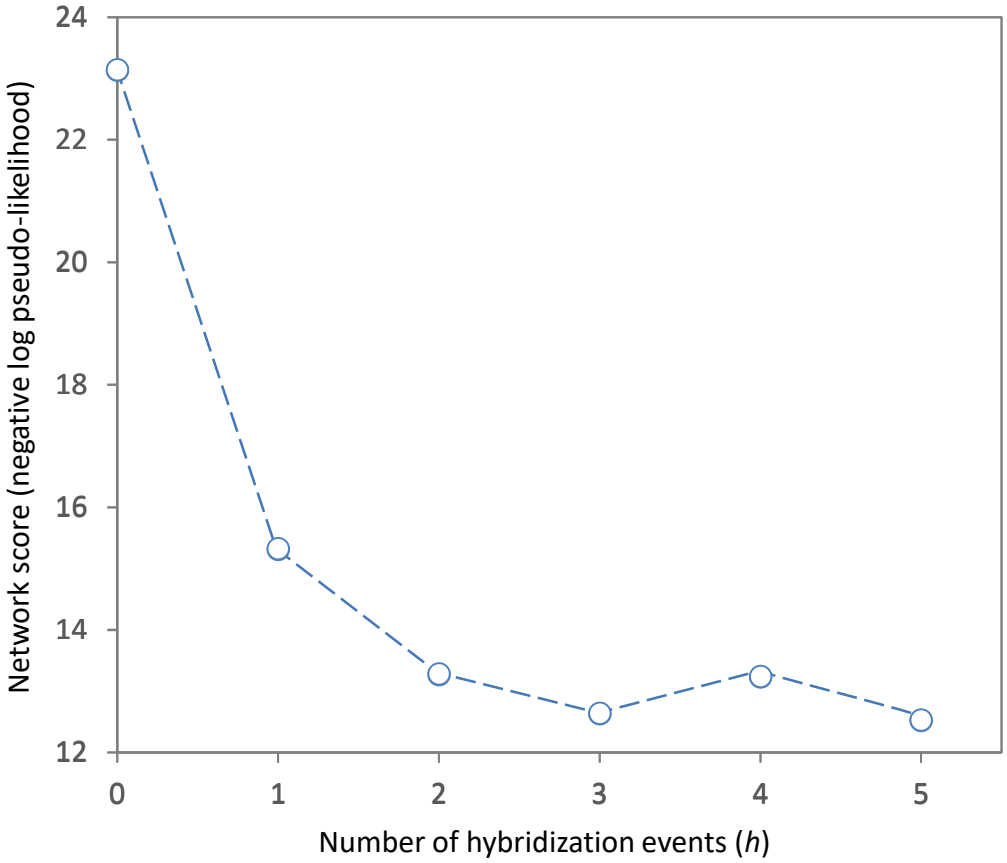


FIGURE S7 Summary of model fit with TREEMIX. The figure shows the Ln(likelihood) for models with different number of migration events (*m* from 0 to 4) over three independent runs (open circles). Analyses were run including one terminal per species and considering six species-population combinations. Codes of the specific populations of each species included in the different analyses are indicated in each panel and described in Table S1.

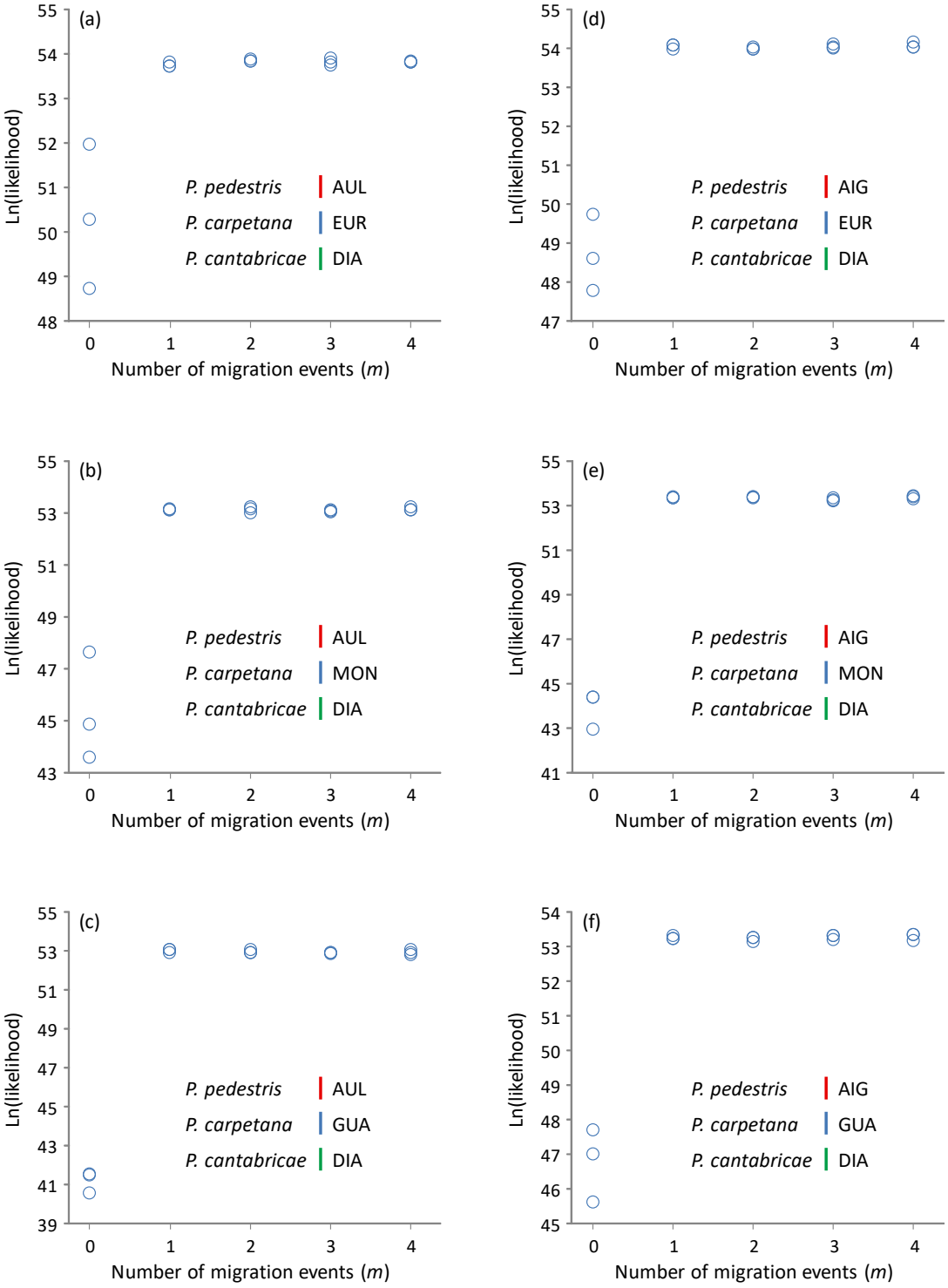
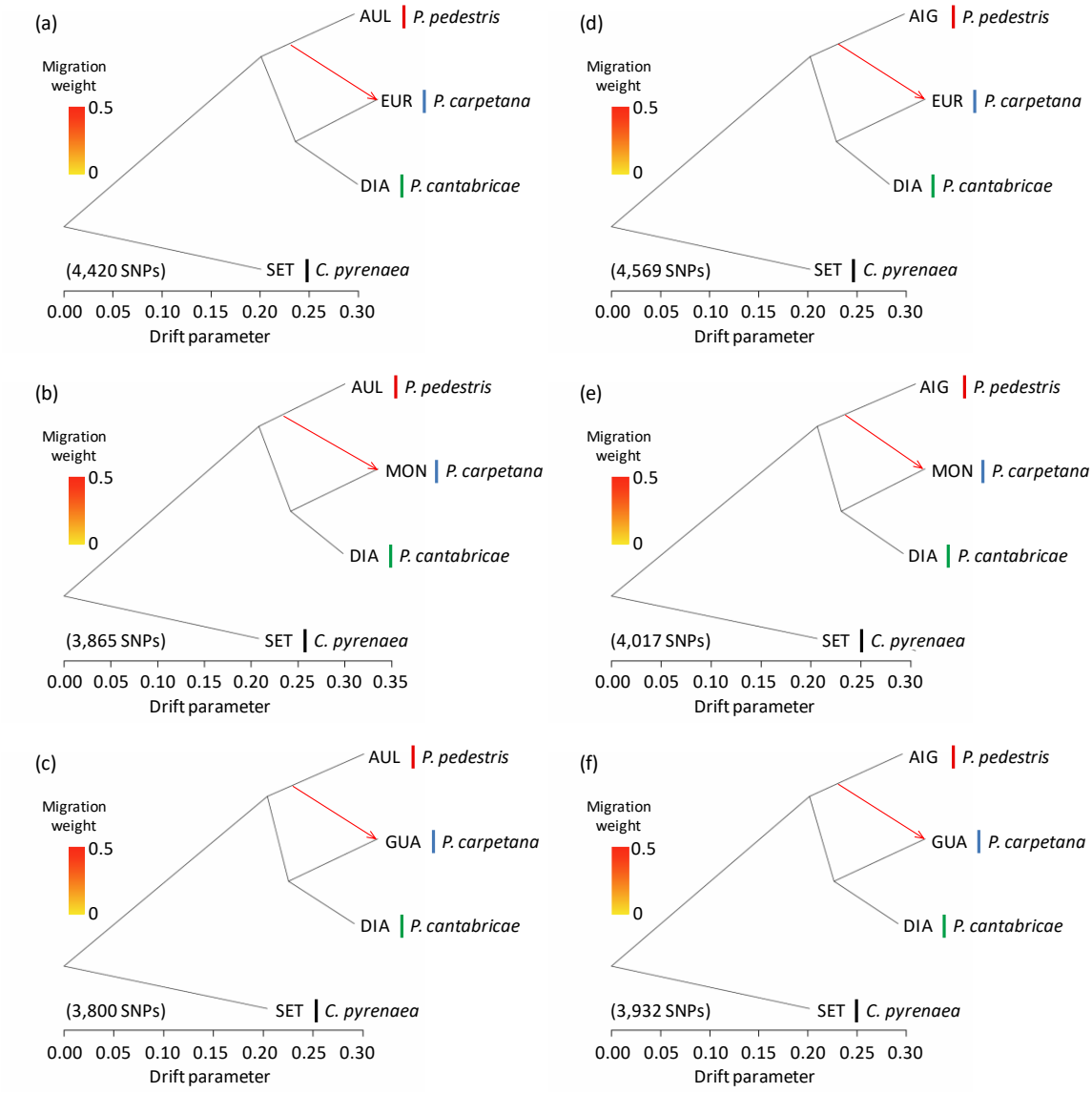


FIGURE S8 Maximum-likelihood trees inferred with TREEMIX showing the most likely migration event (*m* = 1). The direction of gene flow is represented with an arrow colored according to the percentage of alleles (weight) originating from the source. Analyses were run including one terminal per species and considering six species-population combinations. The number of SNPs retained for each analysis is presented in parentheses. Codes of the specific populations of each species included in the different analyses are described in Table S1.



References

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). STACKS: Building and genotyping loci *de novo* from short-read sequences. *G3-Genes Genomes Genetics, 1*(3), 171-182. doi:10.1534/g3.111.000240

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). STACKS: an analysis tool set for population genomics. *Molecular Ecology, 22*(11), 3124-3140. doi:10.1111/mec.12354

Eaton, D. A. R. (2014). PYRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics, 30*(13), 1844-1849. doi:10.1093/bioinformatics/btu121

Eaton, D. A. R., Hipp, A. L., González-Rodríguez, A., & Cavender-Bares, J. (2015). Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution, 69*(10), 2587-2601. doi:10.1111/evo.12758

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics, 6*(2), e1000862. doi:10.1371/journal.pgen.1000862

Ortego, J., Gugger, P. F., & Sork, V. L. (2018). Genomic data reveal cryptic lineage diversification and introgression in Californian golden cup oaks (section *Protobalanus*). *New Phytologist, 218*(2), 804-818. doi:10.1111/nph.14951