

Supporting Information

Causes and consequences of large clonal assemblies in a
poplar hybrid zone

David Macaya-Sanz, Myriam Heuertz, Dorothea Lindtke, Giovanni G.
Vendramin, Christian Lexer and Santiago C. González-Martínez

Table S1 Nuclear microsatellites, including locus code, chromosome and genetic position, and the marker subsets used in the different data analyses.

Loci (primer code)	Chromosome	Genetic position (cM)	Initial marker subset	STRUCTURE marker subset	MSVAR marker subset	FRANz marker subset	Genet age estimation marker subset
ORPM 137	1	-1.00	X	X	X	X	X
GCPM 124	1	0.22		X	X		X
ORPM 030_1	1	17.34	X	X		X	X
ASP376	1	3.53		X	X	X	X
PMGC 2852	1	6.21	X	X	X	X	X
ASP302	1	79.48		X	X	X	X
GCPM 1719	1	25.22		X	X	X	X
GCPM 1274	1	51.21		X			
GCPM 1158	2	-1.00		X		X	X
GCPM 1376	2	102.18		X			
GCPM 1133	3	-1.00		X	X	X	X
ORPM 30_2	3	30.11	X	X	X	X	X
GCPM 1887	3	1.79		X	X		X
GCPM 1629	3	1.07		X	X	X	X
ORPM 203	3	3.50		X	X	X	X
GCPM 1869	3	23.98		X			
GCPM 1688	3	3.84		X			X
ORPM 127	4	-1.00	X	X		X	X
ORPM 220	4	6.66	X	X		X	X
GCPM 1809	4	6.66		X		X	X
GCPM 1255	5	-1.00		X		X	X
GCPM 1192	5	11.78		X	X	X	X
GCPM 1838	5	23.59		X	X	X	X
GCPM 20	5	71.85		X	X	X	X
WPMS 15	5	11.27	X	X	X	X	X
GCPM 139	6	-1.00		X	X	X	X
GCPM 1831	6	6.95		X	X	X	X
GCPM 1074	6	1.59		X			X
ORPM 26	6	8.99		X			
ORPM 167	6	0.17	X	X		X	X
GCPM 1485	6	6.73		X	X	X	X
ASP933	6	29.26		X		X	X
ORPM 190	6	3.49		X	X	X	X
WPMS 12	6	28.77		X	X	X	X
GCPM 2034	6	13.74		X	X	X	X
ORPM 369	6	2.89		X		X	X
ORPM 60	6	4.29	X	X			X
GCPM 1065	6	2.32		X		X	X
ASP322	6	5.33		X	X	X	X
GCPM 1260	7	-1.00		X	X	X	X
WPMS 17	7	26.29		X	X	X	X
GCPM 1416	7	2.76		X	X	X	X
ORPM 312_1	7	11.89	X	X	X	X	X
GCPM 2062	8	-1.00		X	X	X	X
ORPM 374	8	7.62		X			
ORPM 202	8	32.88	X	X	X	X	X
ORPM 268	8	1.37		X	X		X
GCPM 1949	9	-1.00		X	X	X	X
ORPM 23	9	13.56		X		X	X
ORPM 21	9	5.11	X	X			X
GCPM 2020	10	-1.00		X	X	X	X
ORPM 344	10	29.98	X	X	X	X	X
GCPM 1574	10	8.95		X	X		X
ORPM 149	10	0.27	X	X	X	X	X
GCPM 114	10	20.28		X	X	X	X
GCPM 1037	11	-1.00		X			
ORPM 29	11	33.77		X		X	X
GCPM 154	12	-1.00		X			
WPMS 05	12	2.06	X	X		X	X
GCPM 1186	12	23.32		X	X	X	X
GCPM 1353	13	-1.00		X	X	X	X
GCPM 1812	14	-1.00		X	X	X	X
GCPM 1306	14	39.86		X		X	X
GCPM 1894	15	-1.00		X	X	X	X
GCPM 1454	15	0.52		X	X		X
GCPM 1608	15	20.33		X	X	X	X
ORPM 14	16	-1.00	X	X	X	X	X
ORPM 214	18	-1.00	X	X		X	X
GCPM 1577	18	7.45		X	X	X	X
ORPM 28	18	29.60	X	X	X	X	X
GCPM 162	18	14.44		X		X	X
ORPM 206	19	-1.00	X	X		X	X
ORPM 312_2	U	-1.00	X			X	X
TOTAL			20	72	44	56	66

Primer sources: loci coded GCPM and PMGC, from http://www.ornl.gov/sci/ipgc/ssr_resource.htm; loci coded ORPM, from Tuskan et al. (2004; Can J Forest Res); loci coded WPMS, from van der Schoot et al. (2000; Theor Appl Genet) and Smulders et al. (2001; Mol Ecol Notes); and loci coded ASP, from de Carvalho et al. (2010, Mol Ecol).

Table S2 Basic statistics per sample for GBS data: number of reads, coverage and number of stacks.

Genet		First trimmed 60-bp reads								Second trimmed 60-bp reads						
MLL009	Ramet	UTMX	UTMY	Illumina reads (a)	Number trimmed reads	Stacks	Coverage (mean)	Coverage (sd)	Utilized reads (b)	Utilized reads (c)	Number trimmed reads	Stacks	Coverage (mean)	Coverage (sd)	Utilized reads (b)	Utilized reads (c)
	A087	433478	4600385	669457	587583	18322	17.0814	139.961	312966	344771	239824	8738	13.4973	146.329	117939	129084
	A195	406620	4603949	733333	677245	22958	15.5252	128.485	356427	402043	287163	10603	13.1857	161.46	139808	155480
	A240	412985	4596131	758800	698266	26734	14.0928	93.0963	376756	428426	268683	11544	10.6861	112.326	123360	139507
	A268	442660	4613326	614751	540147	14179	20.1961	177.619	286360	311809	207005	5893	17.4682	184.019	102940	110999
	A272	461216	4607852	597606	532461	15763	15.0238	104.339	236820	264905	227972	8209	12.0118	116.427	98605	110109
	A272 (rep)	NA	NA	687666	609579	19882	14.8621	106.763	295488	329195	266373	10646	11.7255	118.152	124830	138897
	A337	419077	4611995	701978	616141	22676	13.2436	179.879	300312	342594	280591	12150	10.9984	235.568	133630	150729
	A358	453532	4608936	560326	499853	16528	13.1199	96.573	216846	245992	215469	8803	10.4905	117.28	92348	104886
	C033	467898	4586658	644721	588032	11466	25.7127	238.642	294822	317697	281708	6504	22.6479	262.544	147302	157332
	C049	492681	4589532	501720	464978	9438	20.5804	151.946	194238	211363	210040	5530	16.883	160.161	93363	100863
MLL006																
	A052	423681	4614886	634721	566619	8659	27.6631	249.357	239535	259403	277887	5891	22.1526	255.518	130501	139947
	A052 (rep)	NA	NA	792795	702529	16944	20.8941	214.152	354030	390385	305265	8623	17.4752	236.969	150689	163486
	A082	434462	4598318	804971	718576	20062	16.4746	131.364	330514	375835	300753	10833	12.0582	123.69	130627	147187
	A244	442907	4612671	886440	774376	25272	15.2595	114.386	385639	440611	295436	11493	11.7389	130.301	134915	151545
	C004	381399	4660402	810028	721412	11629	31.709	313.693	368744	396324	315887	6096	26.7726	311.921	163206	173333
	C038	475166	4576950	594138	532310	6780	34.1819	256.824	231753	250714	231986	4118	25.5986	238.024	105415	113007
	C055	512215	4614202	666809	591159	6733	41.6124	317.724	280176	300440	263315	3688	35.1643	312.899	129686	137808
	C059	455835	4655330	587148	532071	5146	50.9246	450.253	262058	276778	251748	3053	42.357	463.393	129316	135147
	C066	491584	4616554	590371	529466	3801	66.6911	445.361	253493	267424	249420	2570	50.163	422.948	128919	135364
	C081	497545	4584454	587812	529105	8694	23.9829	183.998	208507	229071	243569	5633	19.2413	192.965	108386	117497

- (a) After preliminary filtering (using process_radtags fromStack software)
- (b) Before matching remainder reads
- (c) After matching remainder reads

Table S3 General information on MLLs sampled, including the code of an exemplary ramet, the species (Pa for *P. alba*, Pc for *P. × canescens*, and Pt for *P. tremula*), the number of ramets, the spatial extension of the MLL, the number of offspring assigned by COLONY, the identity of both parents assigned by COLONY, and the number of offspring assigned by FRANz.

MLL code	Code of exemplary ramet	Species	Number of ramets	Extension (km)	# offspring (COLONY)	Parent A identity	Parent B identity	# offspring (FRANz)
9	A012	Pa	189	99.53	29	*4	#3	21
25	A051	Pa	26	74.62	23	*6	#7	23
11	A022	Pa	6	4.14	13	A051	#1	8
120	C064	Pa	1	NA	8	*2	#18	4
6	A008	Pc	124	158.56	7	*13	#17	2
126	C084	Pa	3	4.89	5	*3	#18	2
86	A321	Pa	2	1.55	5	A051	A012	4
58	A217	Pa	1	NA	3	A321	#1	2
59	A220	Pa	2	0.12	2	A321	#1	0
135	T014	Pt	1	NA	2	C056	#19	1
73	A286	Pa	17	5.60	1	C084	A012	2
74	A287	Pa	7	10.68	1	*2	#5	1
42	A141	Pa	2	0.04	1	A022	A012	1
47	A147	Pa	2	0.06	1	A051	#5	1
101	A373	Pa	2	1.16	1	*2	A012	1
129	T003	Pt	2	2.17	1	*11	T014	0
117	C056	Pt	1	NA	1	*11	#16	0
133	T010	Pt	1	NA	1	*11	#19	1
2	AD111	Pa	8	22.60	0	*10	#5	0
49	A150	Pa	5	17.46	0	*6	#11	0
57	A209	Pc	5	0.62	0	A008	A012	2
83	A314	Pa	5	2.59	0	A321	#1	2
128	T001	Pt	4	6.07	0	*11	#19	0
134	T011	Pt	4	28.07	0	*11	#19	0
34	A132	Pa	3	2.38	0	A051	#9	1
108	C002	Pa	3	134.92	0	*5	#6	1
140	T021	Pt	3	1.49	0	*11	#19	0
14	A001	Pa	2	0.05	0	A217	#4	1
12	A003	Pa	2	0.31	0	A022	C064	0
21	A025	Pa	2	0.12	0	A051	#5	0
23	A043	Pa	2	0.06	0	A051	#1	0
24	A049	Pa	2	10.45	0	*5	#6	0
33	A130	Pa	2	3.91	0	A051	#5	1
53	A200	Pa	2	5.24	0	A220	#1	4
55	A203	Pc	2	0.08	0	A008	A012	1
56	A207	Pa	2	0.09	0	A321	#1	1
61	A225	Pa	2	6.80	0	*10	#4	0
82	A312	Pa	2	2.84	0	C084	A012	1
103	A383	Pa	2	0.33	0	*3	A012	0
105	A386	Pa	2	13.62	0	*3	A012	1
4	AD201	Pa	2	0.56	0	A022	#4	0
107	C003	Pa	2	7.71	0	*12	#15	0
111	C045	Pa	2	17.12	0	C084	C064	2
118	C057	Pc	2	9.55	0	*1	#3	1
136	T016	Pt	2	22.95	0	*11	#16	0
138	T018	Pt	2	3.56	0	*11	#16	0
15	A004	Pa	1	NA	0	A022	A012	0
16	A005	Pa	1	NA	0	A022	C064	1
19	A011	Pc	1	NA	0	A008	A012	1
22	A038	Pa	1	NA	0	A051	A012	0
28	A062	Pa	1	NA	0	*7	#2	0
30	A088	Pc	1	NA	0	*4	#3	1
32	A129	Pa	1	NA	0	A051	#8	0

Table S3 (continued)

MLL code	Code of exemplary ramet	Species	Number of ramets	Extension (km)	# offspring (COLONY)	Parent A identity	Parent B identity	# offspring (FRANz)
35	A134	Pa	1	NA	0	A051	C064	0
36	A135	Pa	1	NA	0	*8	#7	0
37	A136	Pa	1	NA	0	A022	A012	1
38	A137	Pa	1	NA	0	A051	#8	0
39	A138	Pa	1	NA	0	A051	#5	0
40	A139	Pa	1	NA	0	A051	#8	0
41	A140	Pa	1	NA	0	A022	#10	0
43	A143	Pa	1	NA	0	A022	#10	0
44	A144	Pa	1	NA	0	*9	A141	0
45	A145	Pa	1	NA	0	*4	#5	0
46	A146	Pa	1	NA	0	A051	#9	0
48	A149	Pa	1	NA	0	*6	A147	0
50	A151	Pa	1	NA	0	A051	A012	0
51	A183	Pc	1	NA	0	A008	A012	0
54	A202	Pa	1	NA	0	A051	#1	0
60	A222	Pa	1	NA	0	A217	A012	1
64	A253	Pa	1	NA	0	*6	#1	1
66	A263	Pa	1	NA	0	A051	C064	0
67	A266	Pa	1	NA	0	*9	#11	0
69	A277	Pc	1	NA	0	*6	#12	0
70	A281	Pa	1	NA	0	*6	#9	0
71	A282	Pa	1	NA	0	A051	#13	0
72	A283	Pa	1	NA	0	*8	#2	0
75	A289	Pa	1	NA	0	C084	A012	0
76	A296	Pa	1	NA	0	*6	#12	1
77	A302	Pa	1	NA	0	*7	#12	0
78	A303	Pa	1	NA	0	*6	#14	0
79	A307	Pa	1	NA	0	A220	#15	0
81	A311	Pc	1	NA	0	A008	A012	1
85	A318	Pa	1	NA	0	A217	A012	0
89	A329	Pa	1	NA	0	A321	#1	0
91	A343	Pa	1	NA	0	A051	A012	0
94	A361	Pa	1	NA	0	A051	A373	0
96	A363	Pa	1	NA	0	A051	#2	0
97	A367	Pa	1	NA	0	A051	A012	0
98	A368	Pa	1	NA	0	A051	#1	0
99	A369	Pc	1	NA	0	*1	#3	0
100	A370	Pa	1	NA	0	A287	A286	0
104	A384	Pa	1	NA	0	*3	A012	0
3	AD171	Pa	1	NA	0	C084	A012	0
5	AD203	Pa	1	NA	0	A022	A012	0
7	AD221	Pa	1	NA	0	A022	A012	1
8	AD222	Pa	1	NA	0	A022	A012	0
10	AD231	Pa	1	NA	0	A022	C064	0
13	AD251	Pa	1	NA	0	A022	#14	0
1	AD91	Pt	1	NA	0	*11	#16	0
109	C010	Pc	1	NA	0	A008	A012	0
114	C050	Pc	1	NA	0	A008	C064	1
115	C052	Pa	1	NA	0	*12	#9	0
119	C060	Pc	1	NA	0	*1	A012	0
122	C067	Pt	1	NA	0	*11	#16	0
123	C068	Pt	1	NA	0	*11	#19	0
124	C078	Pa	1	NA	0	*2	A012	0
127	C101	Pa	1	NA	0	*6	C064	0
130	T006	Pt	1	NA	0	*11	T014	0
131	T007	Pt	1	NA	0	*11	#16	0
132	T008	Pt	1	NA	0	*11	#16	0
137	T017	Pt	1	NA	0	*11	#19	0
141	T024	Pt	1	NA	0	T010	T003	0

NA: Not applicable, as only one ramet was found.

Prefixed star (*) in the column 'Parent A identity' indicates that the potential parent in not is the sample

Prefixed hash (#) in the column 'Parent B identity' indicates that the potential parent is not in the sample

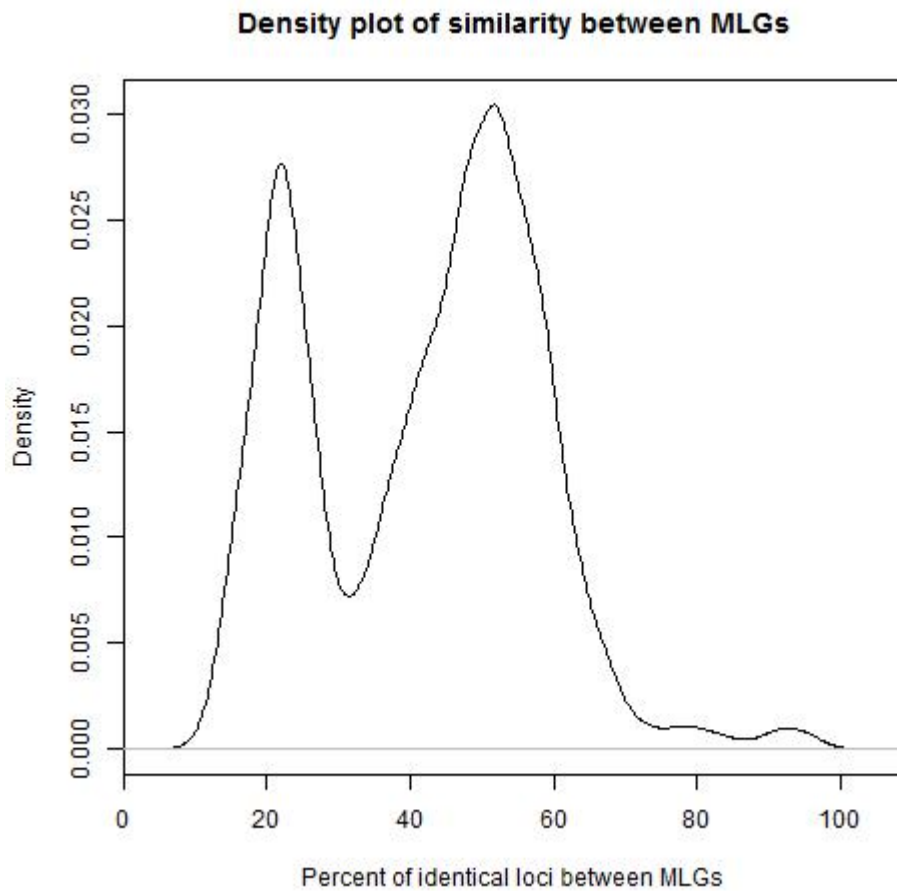


Figure S1 Distribution of pairwise genetic similarity between MLGs, measured as percent of identical alleles. Note that the distribution has more than two modes due to the population including two different species. The mode with less percentage of similarity corresponds to the pairwise comparisons between species, the mode in the middle to the pairwise comparisons within species, and the small mode with higher genetic similarity represents the comparison between MLGs within the same MLL.

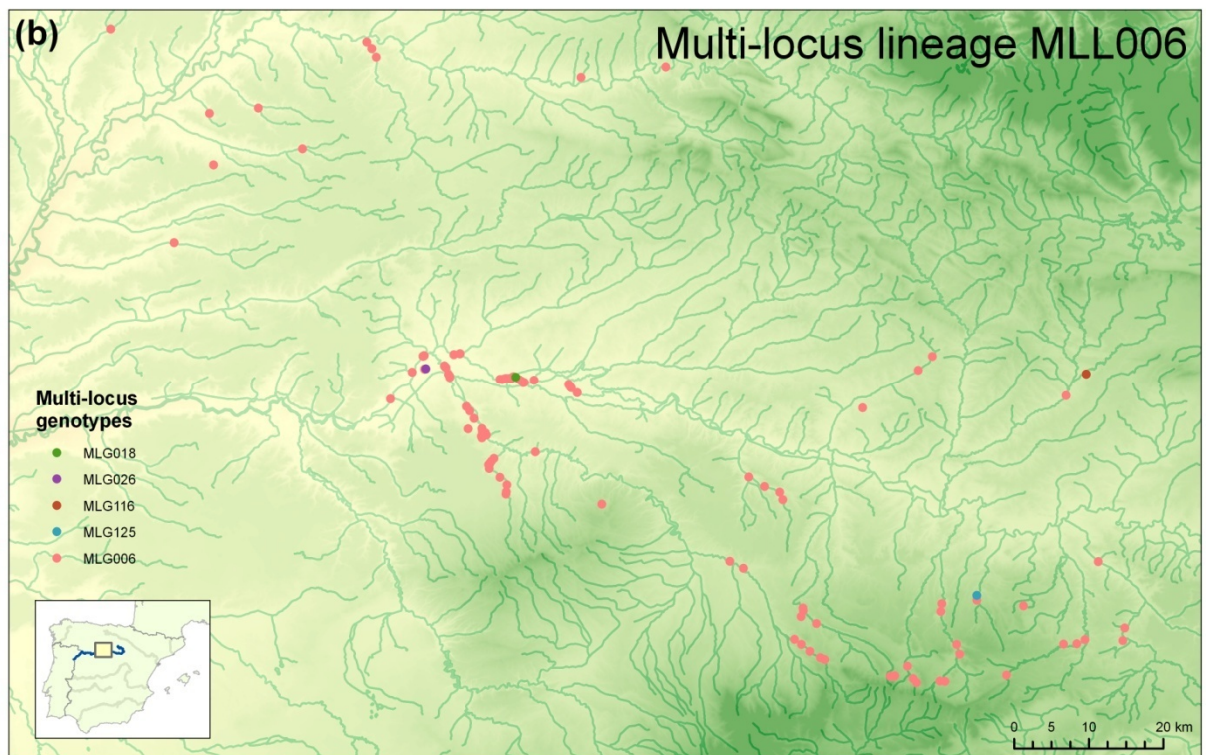
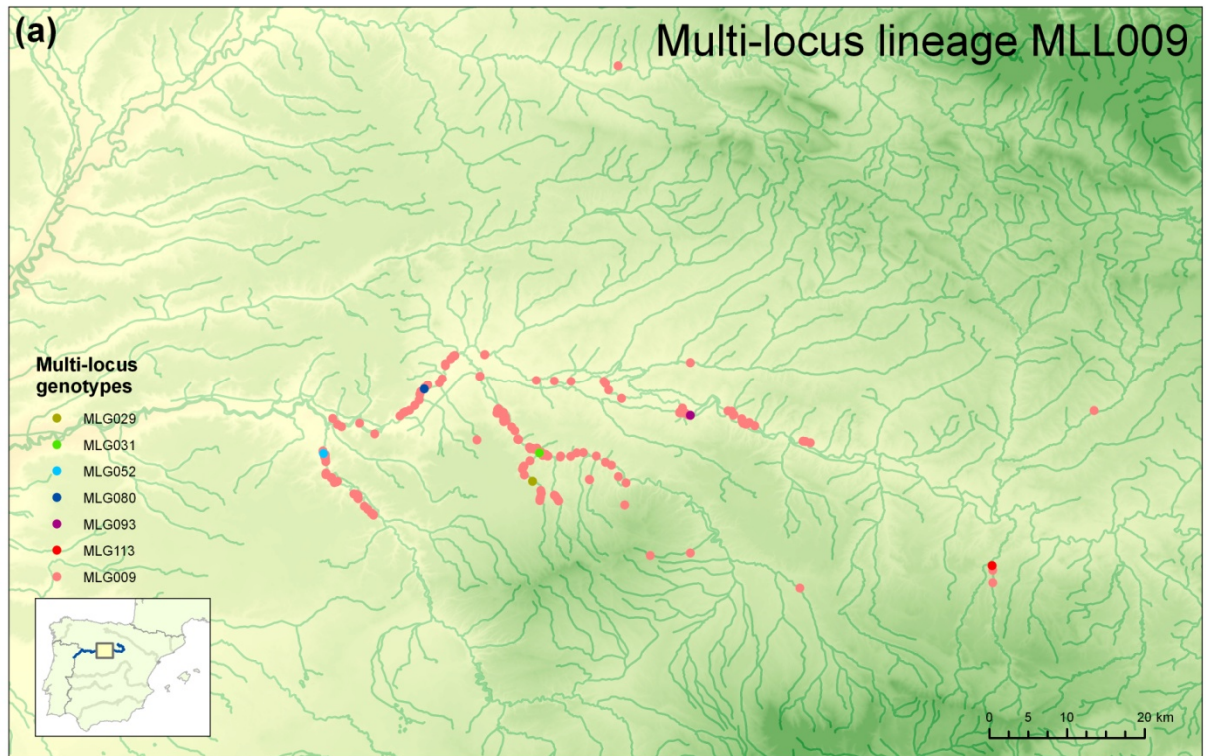


Figure S2 Geographic distribution of ramets for clones MLL009 (a) and MLL006 (b). Color key indicates the genotype (i.e. MLG) of each ramet.

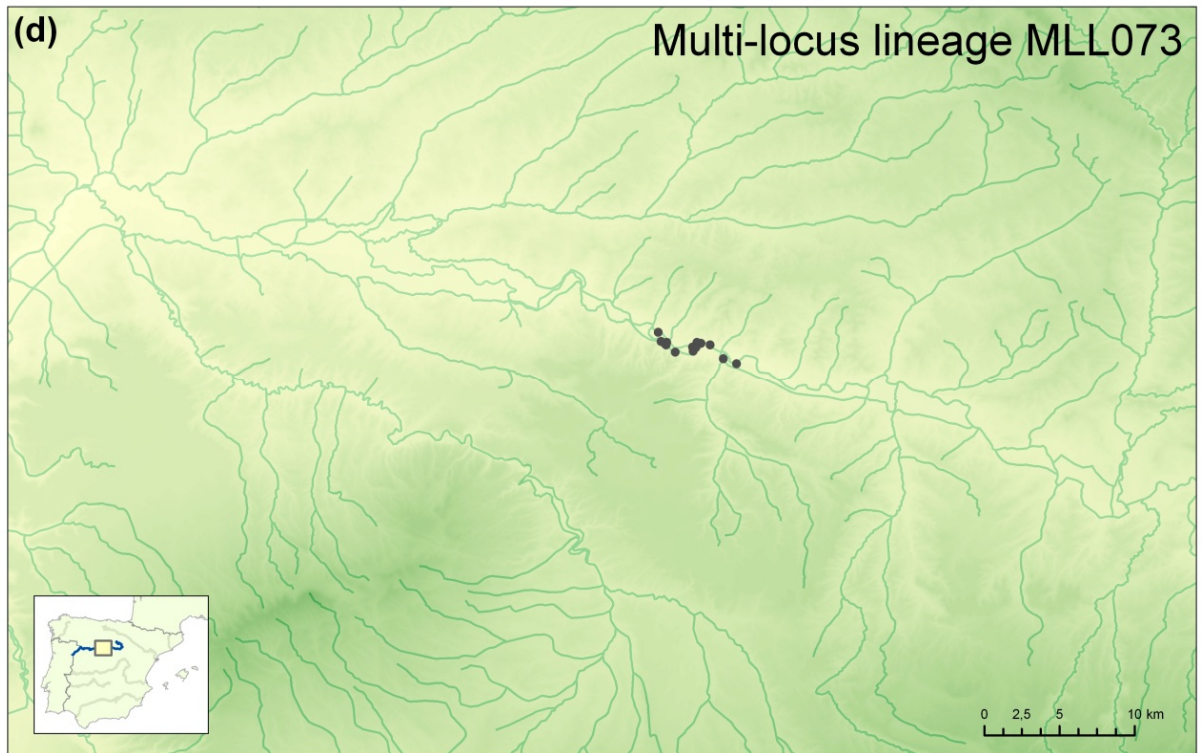
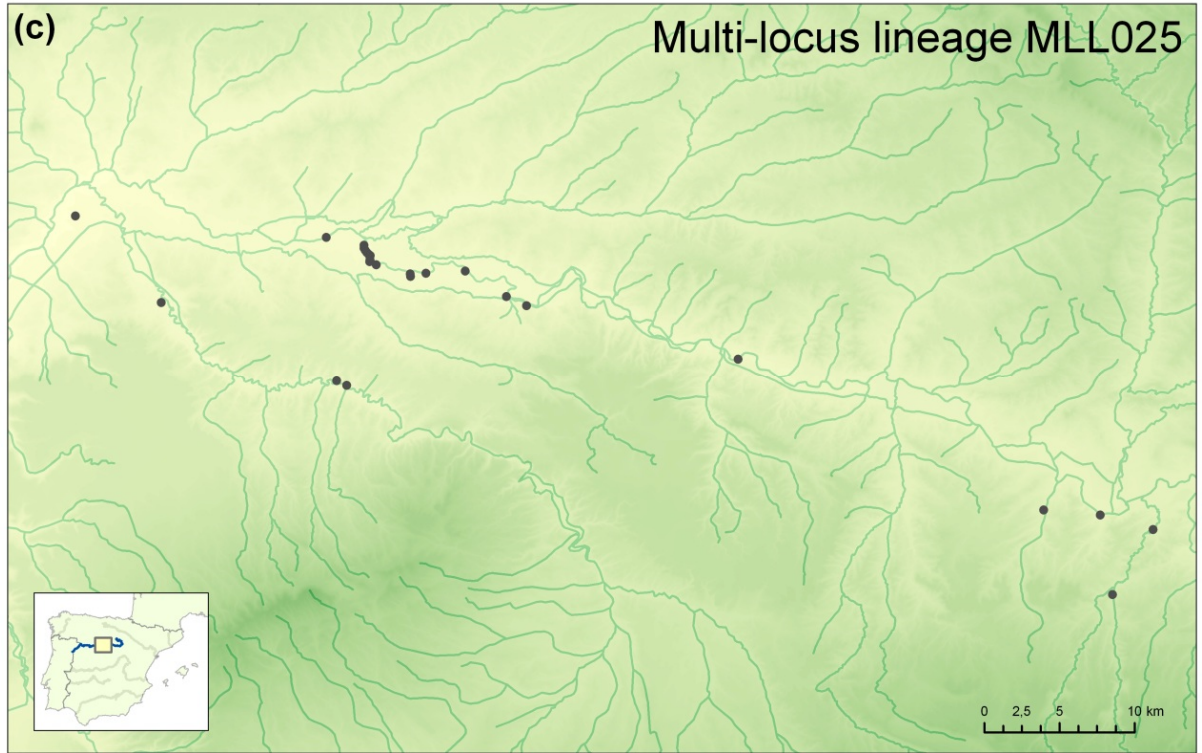


Figure S2 (continued) Clones MLL025 (c) and MLL073 (d).

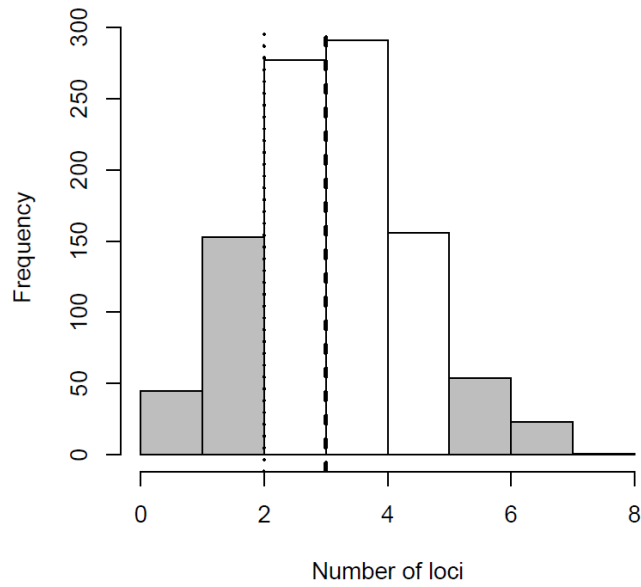


Figure S3 Distribution of expected number of loci with extreme LSA scores in a subset of 10 clones, but without extreme values in the remaining clones, as identified by 1,000 permutations (without replacement). Gray bars indicate lower 5% and upper 95% quantiles of the distribution. The dotted and dashed lines show the observed number of loci with extreme LSA scores in the 10 largest clones in terms of spatial extension and number of ramets, respectively.

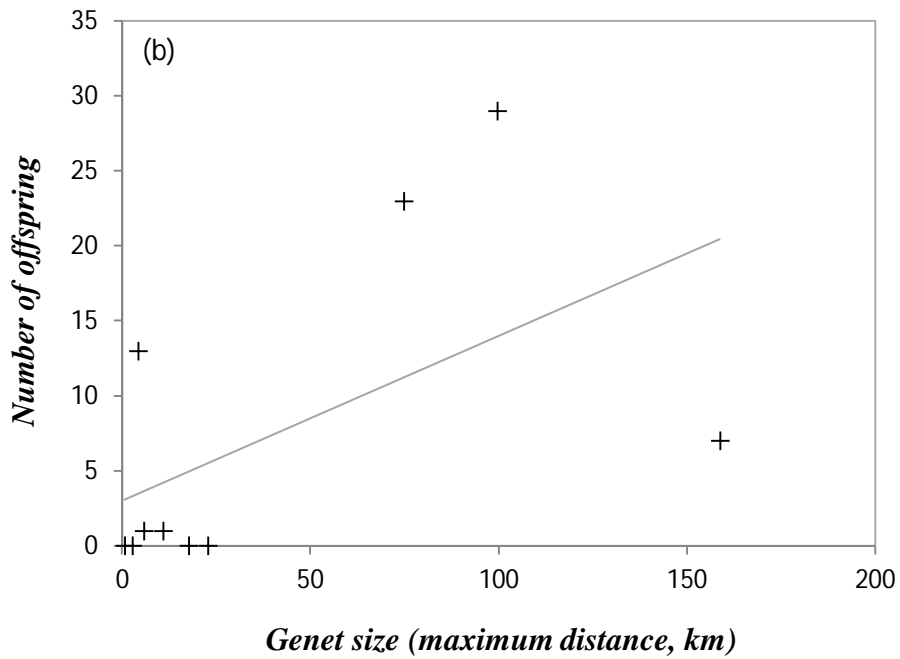
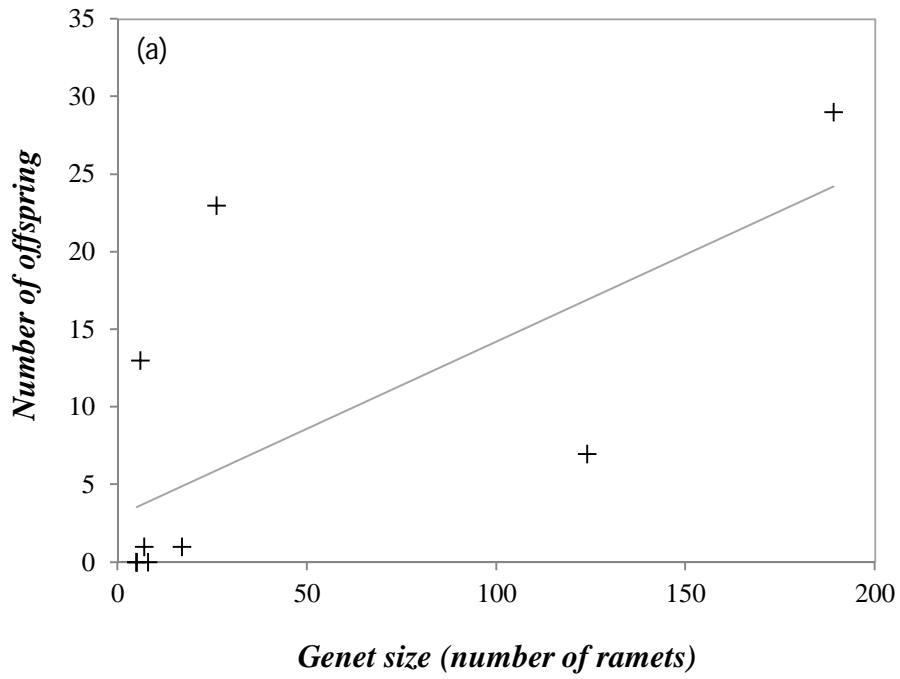


Figure S4 Correlation between genet size and number of offspring, either estimated as number of ramets (a) or as maximum distance between two ramets (b). Only the ten largest clones are represented. A linear trend is also provided.

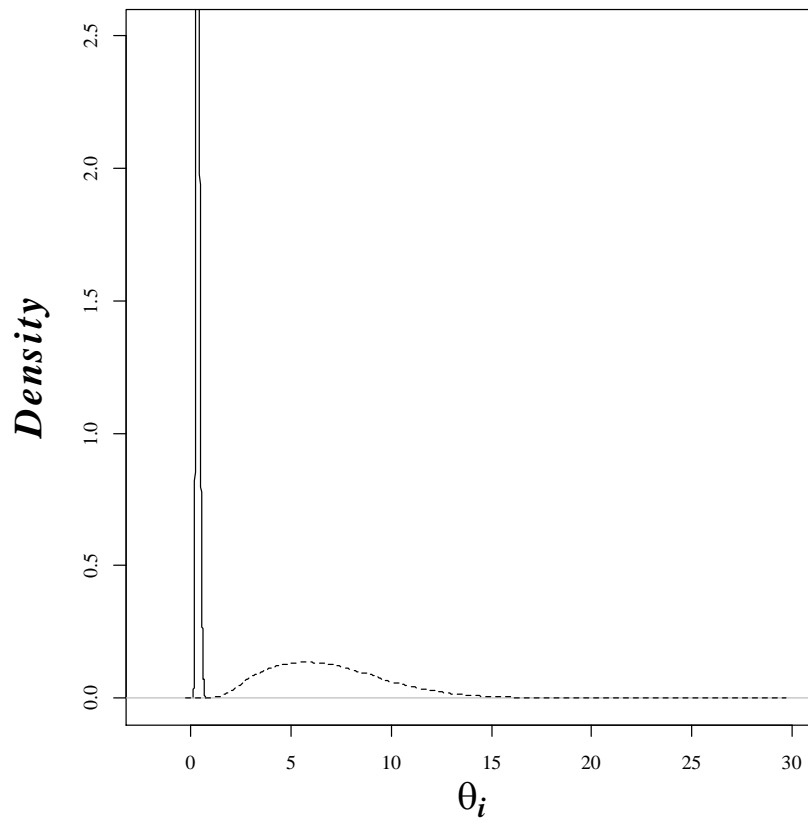
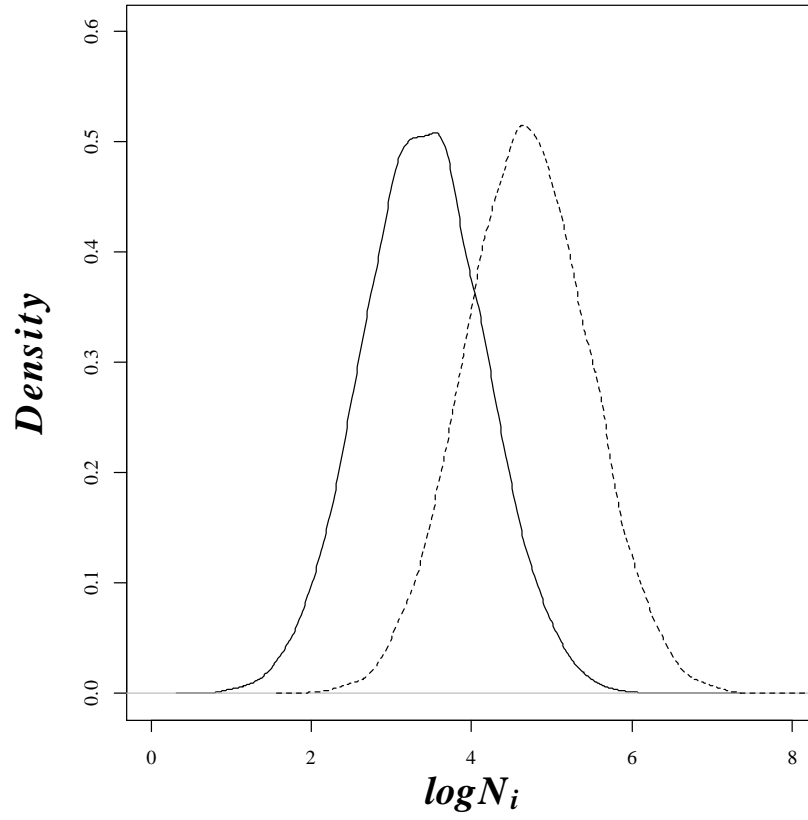


Figure S5 Posterior distributions of main population parameters obtained by MSVAR in an exponential size-change scenario. Solid line corresponds to current values, and dashed line to ancestral values.