Supplemental Information for:

Population genomics of the invasive Argentine ant (*Linepithema humile*) – adaptive evolution in the introduced supercolonies despite low genetic diversity

Ida Holmberg, Jenni Paviala, Jes Søe Pedersen, Heikki Helanterä, Lumi Viljakainen

Table of Contents:

S1. Sampling locations of the Argentine ant supercolonies studied. Sampling was done from four geographic areas to represent the native and introduced ranges. In the native range, ten individuals were collected from the ten different nests (a total of 100 individuals) belonging to ten supercolonies from four different localities. In the introduced range, ten individuals were collected from the ten different nests belonging to the same supercolony (a total of 100 individuals). Finally, in the case of invasive supercolonies, these 100 individuals were pooled to represent final supercolony-specific samples. The native range sample captured total genetic diversity in the native range of species instead of representing one native supercolony. Sampled nests were a subset of those used in Vogel et al. (2010)*, but worker specimens were different.

Supercolony name	Population genomic sample name	Geographic range [native (N) or introduced (I)]	Locality	Supercolony ID	Nest sample ID [10 workers from each nest]
The native supercolony	Native	N	Argentina, Corrientes, Arroyo Cuay Grande	Corrientes S1	14
The native supercolony	Native	N	Argentina, Corrientes, Arroyo Cuay Grande	Corrientes S2	21
The native supercolony	Native	N	Argentina, Buenos Aires, Otamendi	Otamendi S2	106
The native supercolony	Native	N	Argentina, Buenos Aires, Otamendi	Otamendi S3	70
The native supercolony	Native	N	Argentina, Buenos Aires, Otamendi	Otamendi S4	65
The native supercolony	Native	N	Argentina, Buenos Aires, Otamendi	Otamendi S6	30
The native supercolony	Native	N	Argentina, Buenos Aires, Otamendi	Otamendi S9	01
The native supercolony	Native	N	Argentina, Buenos Aires, Buenos Aires City, Boca	Boca S1	02
The native supercolony	Native	N	Argentina, Buenos Aires, Buenos Aires City, Boca	Boca S2	07
The native supercolony	Native	N	Argentina, Buenos Aires, Santa Coloma	Santa Coloma	04
The European Main supercolony	European Main	I	France, Port-Leucate	European Main	58, 60, 61, 63, 64, 65, 68, 71, 76, 80
The Catalonian supercolony	Catalonia	I	Spain, Pals	Catalonia	03, 14, 15, 17, 18, 20, 22, 24, 25, 33
The Chilean supercolony	Chile	I	Chile, Santiago	Chile	02, 04, 05, 06, 07, 08, 09, 10, 11, 13

* Vogel, V., Pedersen, J. S., Giraud, T., Krieger, M. J. B., & Keller, L. (2010). The worldwide expansion of the Argentine ant. Diversity and Distributions, 16(1), 170–186. https://doi.org/10.1111/j.1472-4642.2009.00630.x

S2. Analyses including Catalonia

Materials and methods – including Catalonia

Genomic diversity (π and Θ_w) and test of neutrality (Tajima's D) were conducted for the shared positions of all four supercolonies, i.e., including Catalonia, in the same way as described in the main text. The only exception was the requirement for sequencing depth criteria, which dropped to 25% (from 60% used for analyses excluding Catalonia). This quite loose window requirement stems from Catalonia's uneven read alignment. In addition, the genomic differentiation (F_{ST}) was estimated in the same way as described in the main text. However, when looking for targets of selection from the extreme ends of the Tajima's D distributions, we left Catalonia out and considered only the three other supercolonies (see main text). In this way, we could minimise the bias in the results of selection analyses, as Catalonia had an uneven alignment covering a small fraction of the reference genome. Similarly, while looking for the most differentiated genomic regions (F_{ST}) from the extreme ends of the F_{ST} distributions, we left Catalonia out to avoid possible bias in the results.

Results and discussion – including Catalonia

Properties of the data and genomic diversity

After filtering, the sample-specific pileup files with shared positions of all four samples with a sequencing depth of at least 20 covered only 20% of the reference genome. Catalonia's read alignment to the reference genome was uneven, with a considerable variation in sequencing depth along the genome (see below Tables S5). After filtering, the multi-sample joint mpileup file (used for genomic F_{ST}) with shared positions of all four samples with a sequencing depth of at least 4 covered 56% of the reference genome (see Tables S5).

We found 256 545 SNPs from Native, which was about two-fold higher than the number of SNPs found in any of the invasive supercolony samples, European Main, Catalonia or Chile (Table S2.1. Statistics of genomic diversity (π and Θ_W) including Catalonia). Thus, the magnitude of differences remains similar to those in the analyses with higher reference genome coverage, excluding Catalonia (see main text). Both genomic diversity estimators, π and Θ_w, indicated that the native supercolonies are genetically more diverse than the invasive supercolonies (Mann-Whitney U-test: $p<2.2e^{-16}$ in all pairwise comparisons of π and Θ_w estimated in 20 kbp windows between the native and invasive supercolonies) giving us similar conclusions as results in the main text (excluding Catalonia).

Table S2.1. Genomic diversity (π and Θ_W) including Catalonia. The genome-wide means and standard deviations (sd) of π and Θ_W and the number of SNPs used for estimations.

Sample	**Mean π**	**Sd π**	**Mean Θ_W**	**Sd Θ_W**	**SNPs**
Including Catalonia					
Native	0.0016	0.0009	0.0017	0.0010	256 545
European Main	0.0008	0.0007	0.0007	0.0007	128 124
Chile	0.0010	0.0010	0.0009	0.0009	141 101
Catalonia	0.0008	0.0008	0.0009	0.0008	115 054

Test of neutrality

The density curves for both 20 kbp genomic windows and annotated genes of the native and Catalonian supercolonies were located on the negative side. In contrast, the density curves of the European Main and Chilean supercolonies were located on the positive side, but the density curves of 20 kbp windows were skewed toward the negative side. In addition, the density curves of the native and Catalonian supercolonies were narrow with high kurtosis, while the density curves of the European Main and Chilean supercolonies were broad, with low kurtosis and more extreme values than the other two supercolonies (Figure S2.2. Density curves of Tajima's D including Catalonia). Thus, the patterns detected in the density plots for the native, European Main, and Chilean supercolonies in the main text remain similar when analysing shared positions for all four supercolony samples covering less from the reference genome. While the lowest extremes of Tajima's D values indicated genomic regions and genes under positive selection, the genome-wide negative Tajima's D might signal population size expansion instead. The negatively located Tajima's D distributions of the Catalonian supercolony might indicate that this introduced supercolony is still expanding.

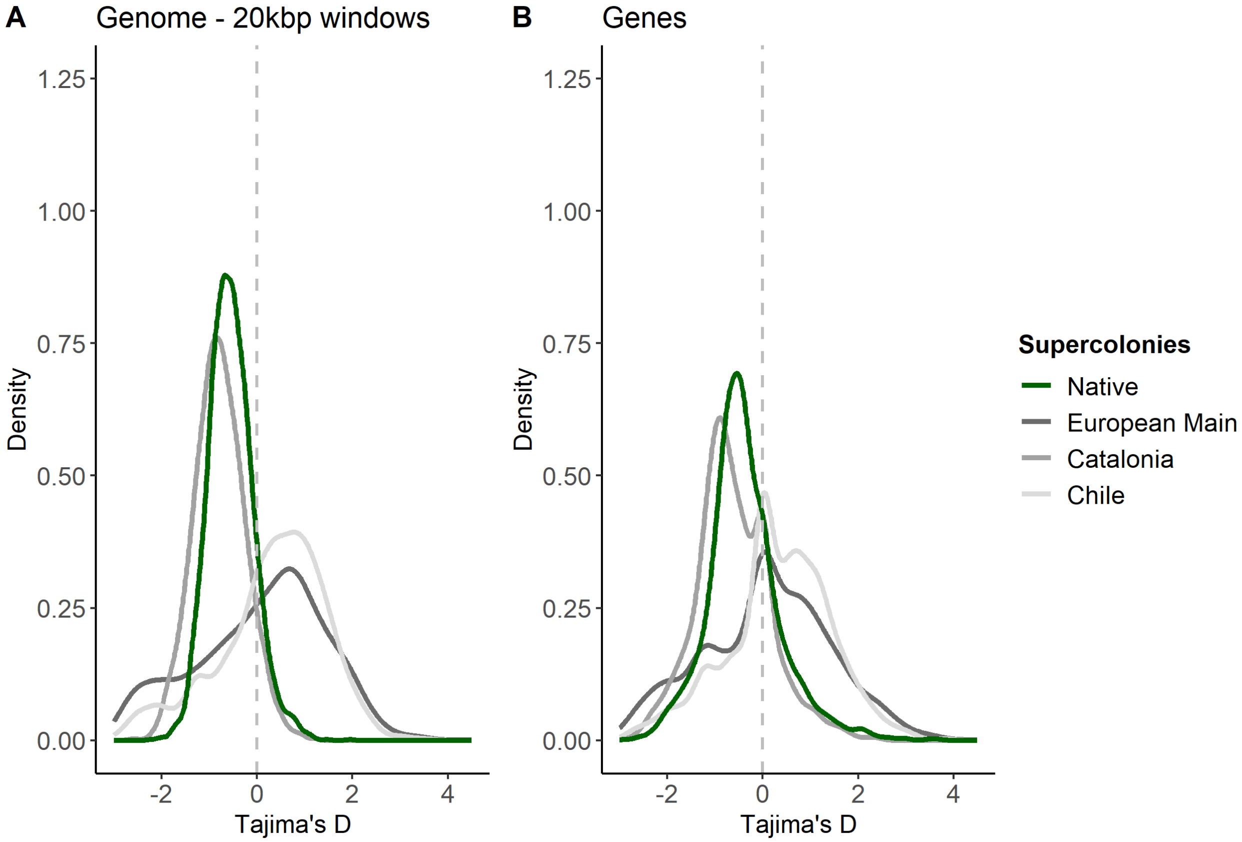


Figure S2.2. *The density curves of Tajima's D values were estimated from (A) whole genomes with 20 kbp windows and (B) genes including Catalonia.*

Genetic differentiation

We used pairwise F_{ST} values to evaluate genomic differentiation between supercolonies. These analyses revealed that the invasive supercolonies seemed to be more differentiated from each other than from the native supercolonies (Mann-Whitney U-test: $p<2.2e^{-16}$ in all pairwise comparisons; Figure S2.3. The boxplots of genomic pairwise F_{ST} values including Catalonia, Table S14). Thus, also F_{ST} gave us similar conclusions as in the main text with the more comprehensive data.

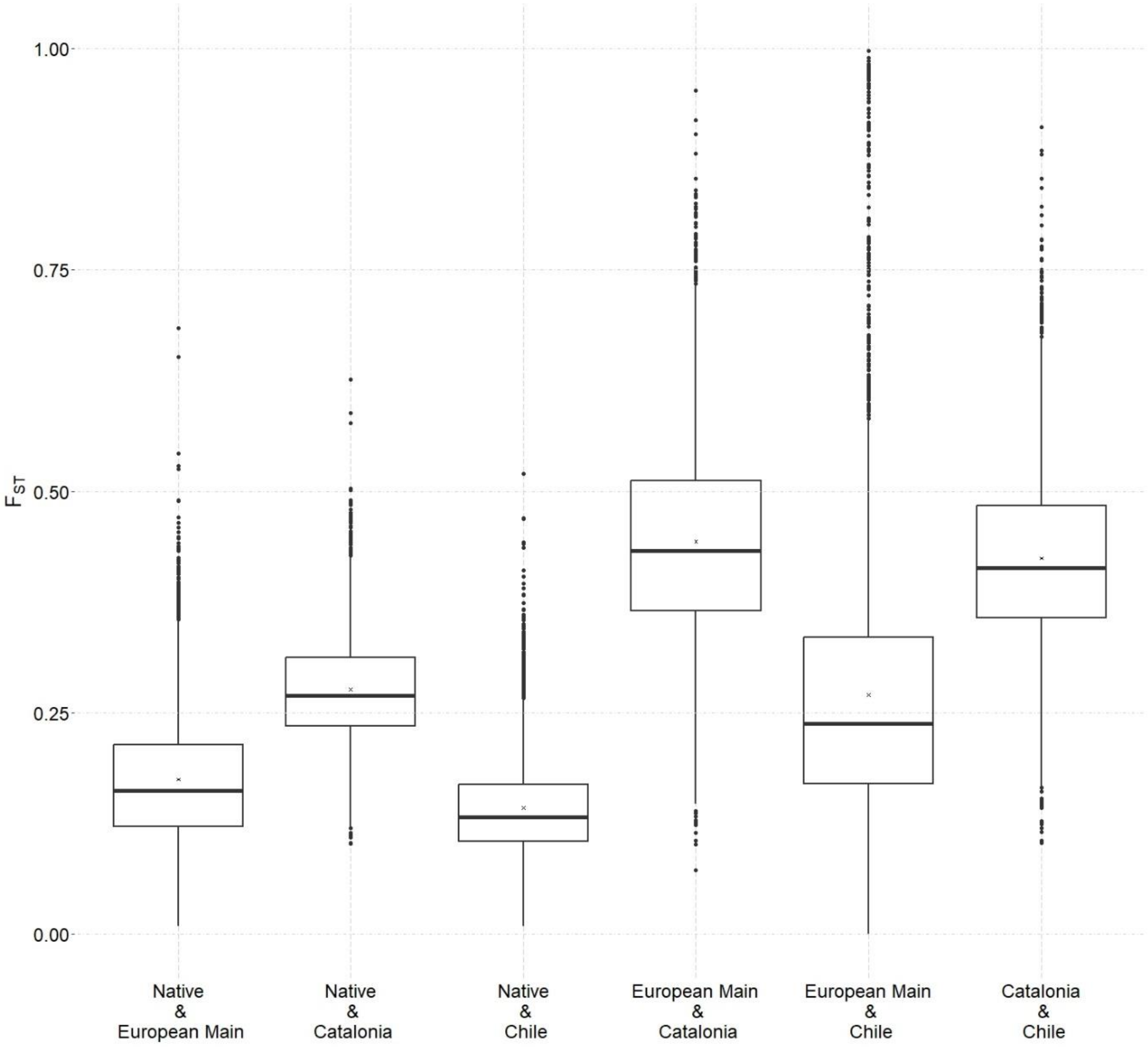


Figure S2.3. *The boxplots of genomic pairwise F_{ST} values including Catalonia.. The mean values are indicated with "x".*

Using a Cochran-Mantel-Haenszel test, we detected similar allele frequency changes between the native and each of the invasive supercolonies. When the Catalonian supercolony was included, 65% of the SNPs found showed significant and similar allele frequency changes (BH-adjusted p-values<0.05; Table S4). As an alternative method to detect similar allele frequency changes between the native and each of the invasive supercolonies, we used qbGLM (See Table S4).

To summarise, the results of the data set, including Catalonia, led to conclusions similar to those in the main text, with the data set excluding Catalonia. However, due to its poor reference genome coverage and following limitations, we decided not to keep Catalonia in all analyses and thus dropped it from the main text.

S3. Comparing the results of genomic data and microsatellite marker data

Materials and methods

Genetic marker data on five variable microsatellite loci (*Lhum*-3, *Lhum*-11, *Lhum*-19, *Lhum*-28, and Lhum-35) were obtained from Vogel et al. (2010) applying a similar sample size and design as the genomic samples, i.e. native with the same 10 supercolonies (10 worker individuals each) as subsamples and each of the three introduced supercolonies with 10 nests as subsamples. Estimations of genetic diversity (expected heterozygosity, H_{exp}, and allelic richness, *k'*) and genetic differentiation (pairwise F_{ST}) were performed as in Vogel et al. (2010) with dispersal measures (SE) calculated by jackknifing over subsamples.

Vogel, V., Pedersen, J. S., Giraud, T., Krieger, M. J. B., & Keller, L. (2010). The worldwide expansion of the Argentine ant. Diversity and Distributions, 16(1), 170–186. https://doi.org/10.1111/j.1472-4642.2009.00630.x

Results

Genetic diversity

Table S3.1. Genetic marker diversity expressed as expected heterozygosity (H_{exp}) and allelic richness (*k'*), both on average per five loci. *k'* is adjusted to a minimum sample size of 10 individuals.

Sample	**H_{exp}**	**SE_{Hexp}**	**k'**	**$SE_{k'}$**	**N**
Native	0.9036	0.0040	9.980	0.185	10
European Main	0.4644	0.0078	3.254	0.043	10
Chile	0.5426	0.0053	3.440	0.056	10
Catalonia	0.5710	0.0073	4.040	0.057	10

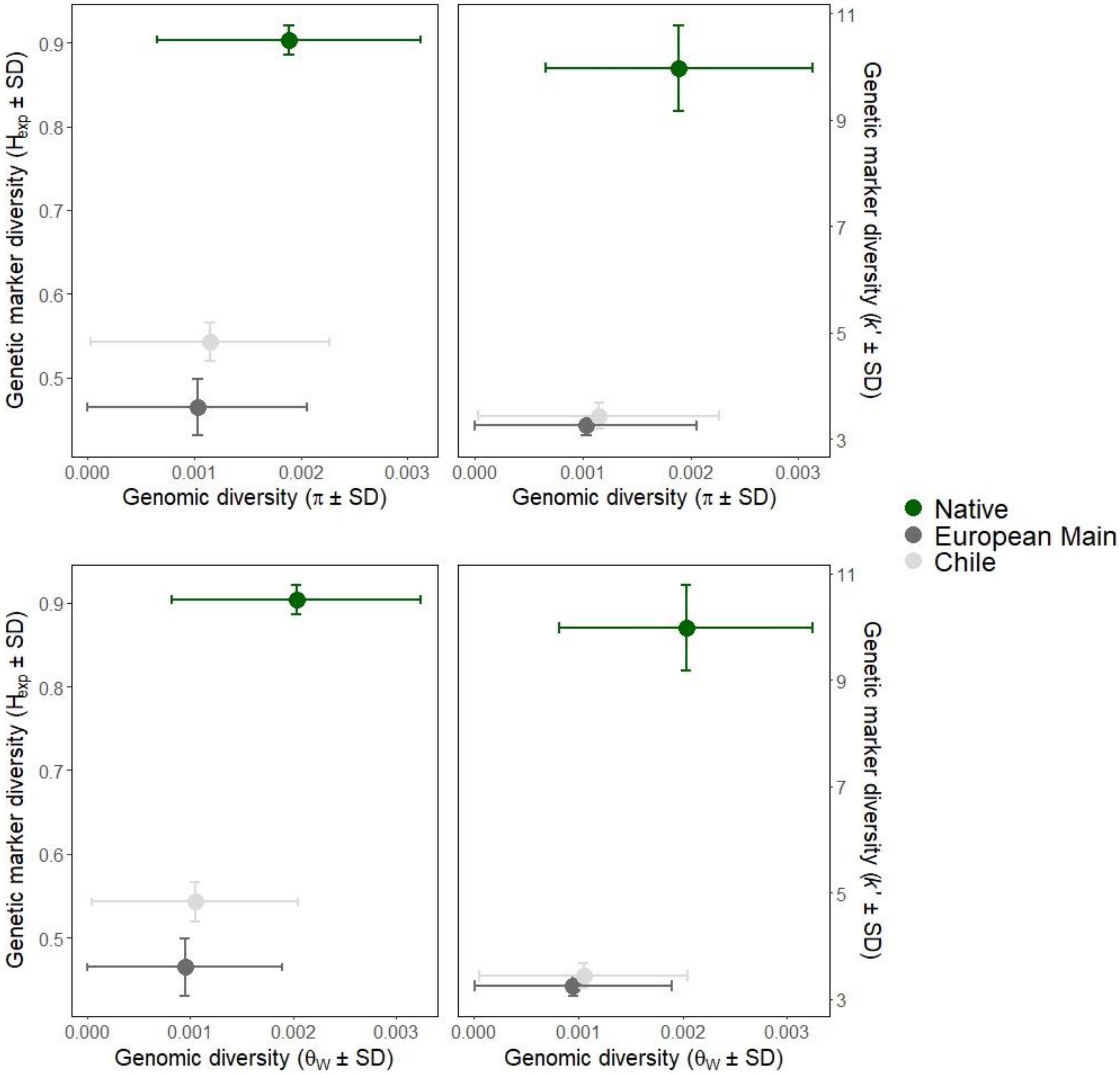


Figure S3.2. *Genomic diversity (mean π and Θ$_w$ estimated in 20 kbp windows for the whole genome) comparison with genetic marker diversity (average* H$_{exp}$ *and allelic richness, k'* per five loci)*, excluding Catalonia.*

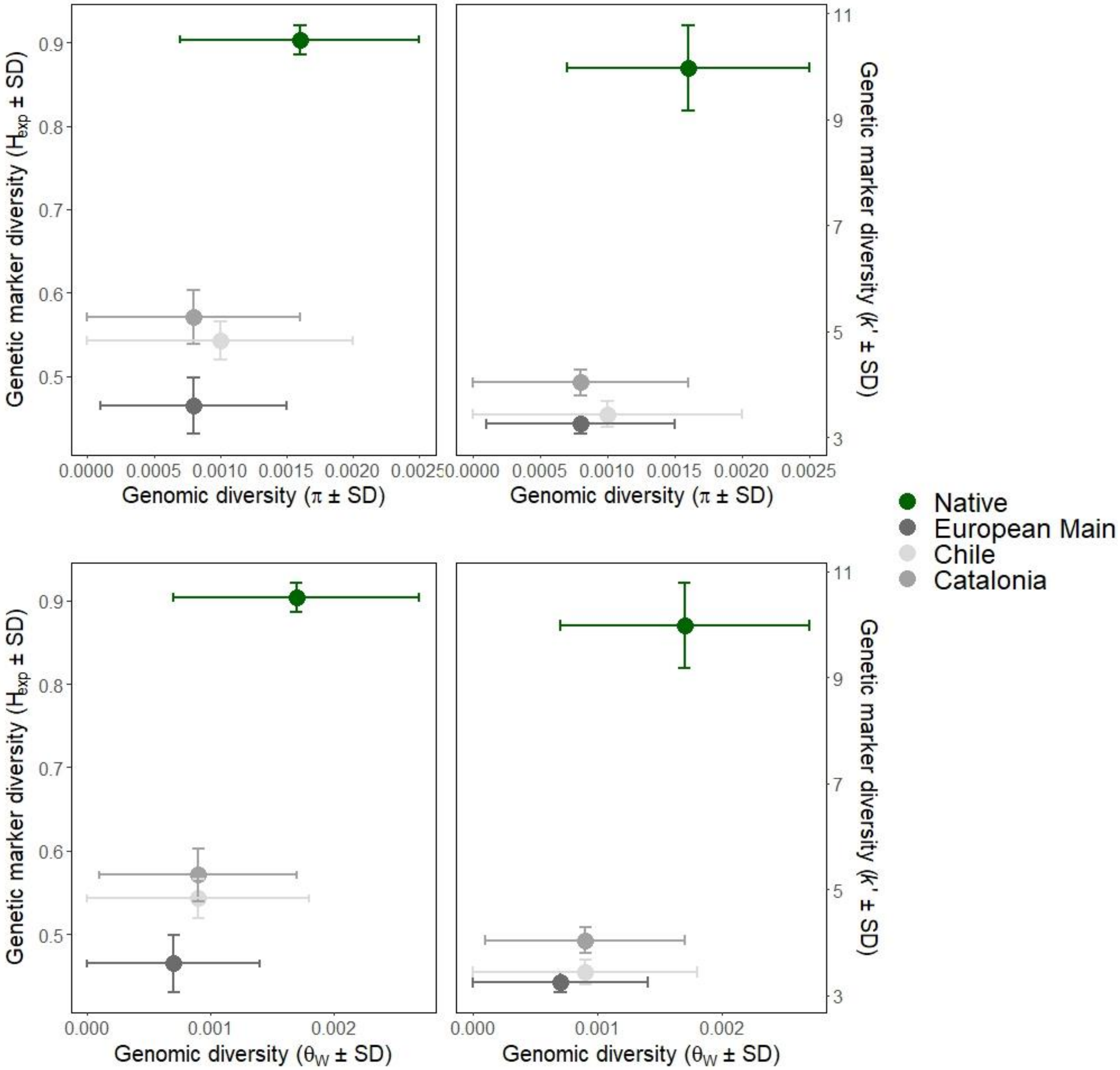


Figure S3.3. *Genomic diversity (mean π and* Θ_w *estimated in 20 kbp windows for the whole genome) comparison with genetic marker diversity (average* H_{exp} *and allelic richness, k'* per five loci)*, including Catalonia.*

Genetic differentiation

Table S3.4. Genetic marker differentiation as pairwise F_{ST}, average per five loci. Please note that for the marker data, unlike the genomic data, F_{ST} is not affected by whether or not Catalonia is included in the analysis.

Sample pair	**F_{ST}**	**SE**	**N**
Native – European Main	0.2419	0.0066	10
Native – Catalonia	0.1945	0.0061	10
Native – Chile	0.1902	0.0083	10
European Main – Catalonia	0.4565	0.0066	10
European Main – Chile	0.3706	0.0055	10
Chile – Catalonia	0.3942	0.0041	10

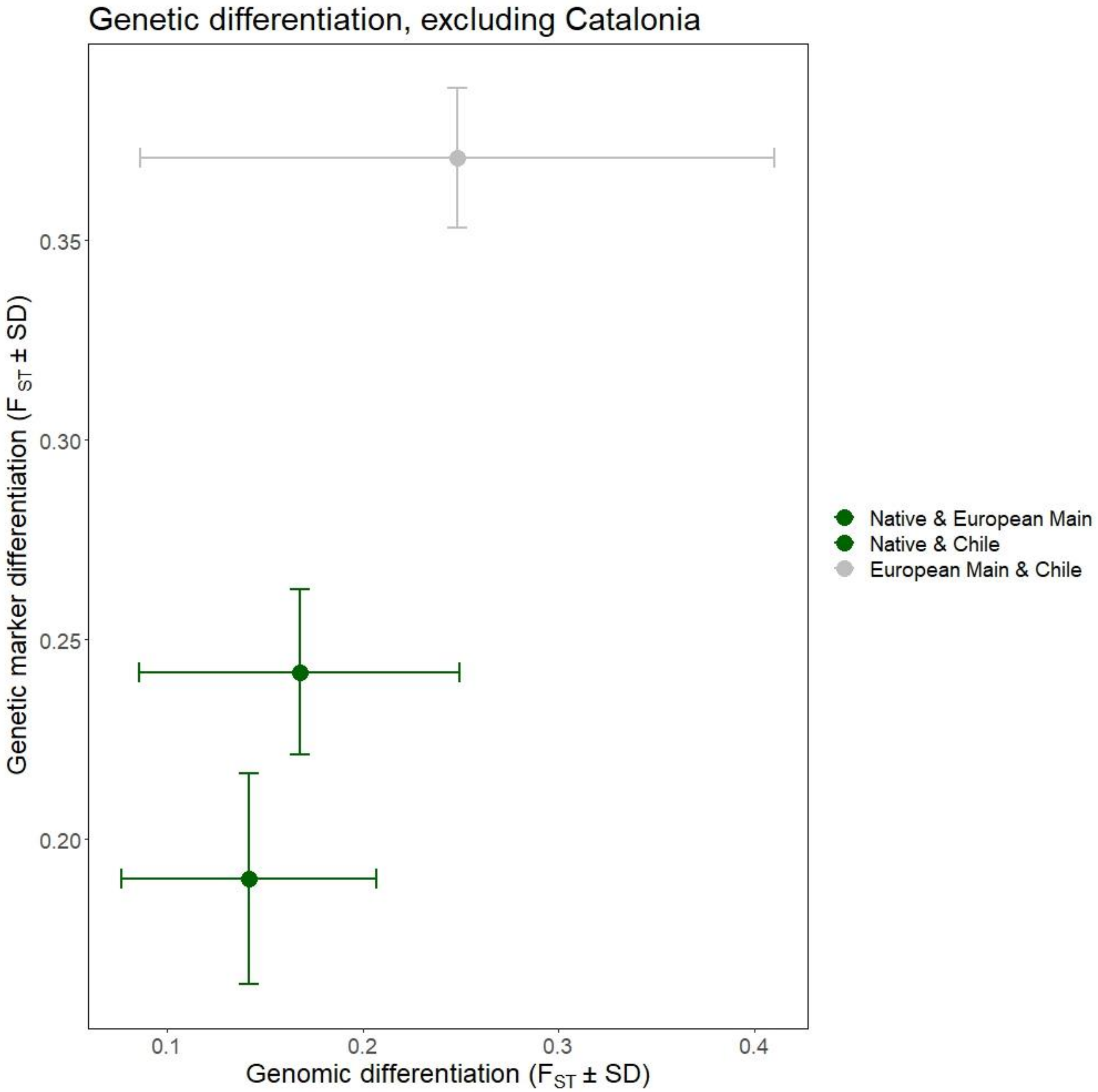


Figure S3.5. *Genomic differentiation (mean F_{ST} estimated in 20 kbp windows for the whole genome) comparison with genetic marker diversity (average F_{ST} per five loci), including Catalonia.*

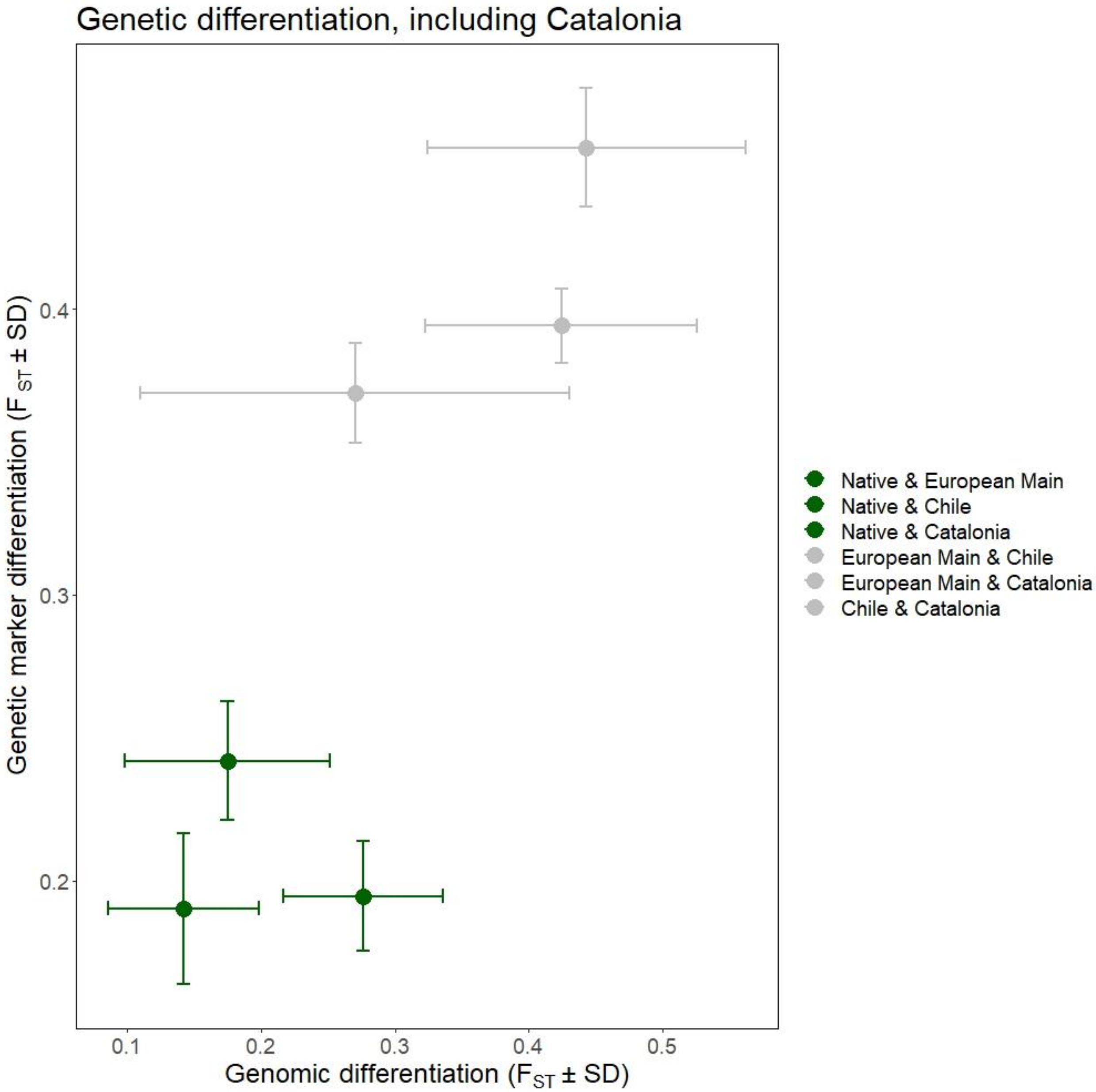


Figure S3.6. *Genomic differentiation (mean F_{ST} estimated in 20 kbp windows for the whole genome) comparison with genetic marker differentiation (average* F_{ST} per five loci)*, excluding Catalonia.*

S4. The number of detected SNPs and associated p-values while looking for similar allele frequency changes between the native and invasive supercolonies using a Cochran-Mantel-Haenszel test (cmh) and a quasibinomial GLM (qbGLM).

In addition to a cmh-test, a quasibinomial generalised linear model (qbGLM) for each segregating site was used to detect significant allele frequency differences between the native and invasive supercolonies. qbGLM was done using the modified multi-sample joint mpileup files with poolFreqDiffTest_QBGLM.py script from the poolFreqDiff program (Wiberg et al., 2017). The minimum and maximum sequencing depth requirements were 4 and 250, respectively. Furthermore, the minimum count required to consider a SNP was 4. This method detected significant and similar allele frequency changes for 31% and 64% of the SNPs, including and excluding Catalonia, respectively (BH-corrected p-values<0.05).

Samples	**Program**	**All p-values**	**P-values < 0.05**	**Adjusted p-values (BH) <= 0.05**	**Proportion of statistically significant SNPs (adj. p-val < 0.05)**
All	cmh	956 798	649 416	625 808	0.6541
All excluding Catalonia	cmh	1 475 443	973 984	932 655	0.6321
All	qbGLM	1 082 636	466 243	334 875	0.3093
All excluding Catalonia	qbGLM	1 692 572	1 118 242	1 076 207	0.6358

S5. Properties of the data

A. The number of raw and filtered reads after specific filtering steps.

Sample	**Raw reads**	**Base quality 20 & min read length 70 bp**	**Mapping quality 20 & flag “properly paired”**	**Deduplication**
Native	351 474 642	338 876 714	249 429 687	224 678 771
European Main	346 653 204	334 410 182	274 324 386	256 626 900
Chile	346 869 216	333 868 165	247 763 921	204 902 571
Catalonia	322 926 608	309 976 495	213 423 509	146 643 559

B. The number of positions with different sequencing depths and the proportions of these positions from the reference genome (coverage) were calculated from the sample-specific indel-filtered pileup files before filtering for shared positions with a sequencing depth of at least 20.

Depth	**0**		**1**		**2**		**>0**		**<20**		**>=20**	
Sample	**Positions**	**Coverage**	**Positions**	**Coverage**	**Positions**	**Coverage**	**Positions**	**Coverage**	**Positions**	**Coverage**	**Positions**	**Coverage**
Native	13 096 157	0.0597	388 593	0.0018	306 435	0.0014	206 404 593	0.9403	10 316 199	0.0470	196 090 509	0.8933
European Main	11 288 881	0.0514	361 368	0.0017	283 907	0.0013	208 211 869	0.9486	10 184 652	0.0464	198 029 317	0.9022
Chile	11 881 078	0.0541	424 274	0.0019	326 747	0.0015	207 619 672	0.9459	13 629 922	0.0621	19 399 2171	0.8838
Catalonia	28 101 445	0.1280	24 501 135	0.1116	23 164 305	0.1055	191 399 305	0.8720	145 208 860	0.6615	46 269 712	0.2108

C. After specific filtering steps, the coverage of the reference genome and the mean sequencing depth of sample-specific alignments. Finally, the filtered sample-specific files containing shared positions with a minimum sequencing depth of 20 covered only 20% of the reference genome. When Catalonia was excluded, the filtered files of Native, European Main and Chile, containing shared positions with a minimum sequencing depth of 20 covered 85% of the reference genome.

	Before deduplication		**After deduplication**		**Indel filtering**		**Shared positions and depth filtering‡**	
Sample	**Coverage†**	**Mean depth**	**Coverage†**	**Mean depth**	**Coverage†**	**Mean depth**	**Coverage†**	**Mean depth**
Native	96	91	96	87	94	87	20/85	137/94
European Main	96	100	96	97	95	97	20/85	163/104
Chile	96	85	96	81	95	81	20/85	131/87
Catalonia	88	63	88	56	87	54	20/-	209/-

† The percentage covered of the reference genome

‡ Coverage and sequencing depth of shared positions, including/excluding Catalonia

D. After specific filtering steps, the reference genome coverage of the multi-sample joint mpileup files. After filtering for shared positions with a minimum sequencing depth of 4, the coverage of the mpileup file containing all four supercolonies covered 56% of the reference genome, and the mpileup file containing Native, European Main, and Chile covered 93% of the reference genome.

	Before indel filtering	**Indel filtering**	**Depth filtering**
	Coverage†	Coverage†	Coverage†
All	96	94	56
All except Catalonia	96	94	93

† The percentage covered of the reference genome

S6. Statistical tests

Comparison groups	**Tested parameter**	**Test**	**Statistic**	**Sample size**	**p-value**
Native vs. European Main	π	Mann-Whitney U-test	W=78016761	$n_{Native}=9402$, $n_{EuropeanMain}=9382$	$<2.2e^{-16}$
Native vs. Chile	π	Mann-Whitney U-test	W=73978979	$n_{Native}=9402$, $n_{Chile}=9467$	$<2.2e^{-16}$
Native vs. European Main	Θ_w	Mann-Whitney U-test	W=83527951	$n_{Native}=9402$, $n_{EuropeanMain}=9382$	$<2.2e^{-16}$
Native vs. Chile	Θ_w	Mann-Whitney U-test	W=80667746	$n_{Native}=9402$, $n_{Chile}=9467$	$<2.2e^{-16}$
Native's genome vs. genes	Tajima's D	Mann-Whitney U-test	W=5965426	$n_{Genome}=9404$ $n_{Genes}=10935$	$p<2.2e^{-16}$
European Main's genome vs. genes	Tajima's D	Mann-Whitney U-test	W=64310360	$n_{Genomes}=9832$ $n_{Genes}=10884$	$p<2.2e^{-16}$
Chile's genome vs. genes	Tajima's D	Mann-Whitney U-test	W=59169477	$n_{Genomes}=9467$ $n_{Genes}=10956$	$p<2.2e^{-16}$
Native and European Main vs. European Main and Chile	Overlap of genes (highest 1% of Tajima's D)	Two-Sample Test for Equality Proportions	$X^2=9.627$ df=1	$n_{Native-EurMain}=109$ $n_{EurMain-Chile}=109$	p-value=0.0019
Native and Chile vs. European Main and Chile	Overlap of genes (highest 1% of Tajima's D)	Two-Sample Test for Equality Proportions	$X^2=17.107$ df=1	$n_{Native-Chile}=110$ $n_{EurMain-Chile}=109$	p-value=$3.5e^{-5}$
Native and European Main vs. European Main and Chile	Overlap of genes (highest 5% of Tajima's D)	Two-Sample Test for Equality Proportions	$X^2=41.995$ df=1	$n_{Native-EurMain}=545$ $n_{EurMain-Chile}=545$	p-value=$9.2e^{-11}$
Native and Chile vs. European Main and Chile	Overlap of genes (highest 5% of Tajima's D)	Two-Sample Test for Equality Proportions	$X^2=25.027$ df=1	$n_{Native-Chile}=547$ $n_{EurMain-Chile}=545$	p-value=$5.7e^{-7}$
Native and European Main vs.	Overlap of genomic windows	Two-Sample Test for	$X^2=17.299$ df=1	$n_{Native-EurMain}=109$ $n_{EurMain-Chile}=109$	p-value=$3.2e^{-5}$

European Main and Chile	(highest 1% of Tajima's D)	Equality Proportions			
Native and Chile vs. European Main and Chile	Overlap of genomic windows (highest 1% of Tajima's D)	Two-Sample Test for Equality Proportions	X^2=15.372 df=1	$n_{Native-Chile}$=110 $n_{EurMain-Chile}$=109	p-value=8.8e^{-5}
Native and European Main vs. European Main and Chile	Overlap of genomic windows (highest 5% of Tajima's D)	Two-Sample Test for Equality Proportions	X^2=92.808 df=1	$n_{Native-EurMain}$=545 $n_{EurMain-Chile}$=545	p-value<2.2e^{-16}
Native and Chile vs. European Main and Chile	Overlap of genomic windows (highest 5% of Tajima's D)	Two-Sample Test for Equality Proportions	X^2=20.492 df=1	$n_{Native-Chile}$=547 $n_{EurMain-Chile}$=545	p-value=6.0e^{-6}
Native vs. European Main	F_{ST}	Mann-Whitney U-test	W=25382	n=10153	p<2.2e-^{16}
Native vs. Chile	F_{ST}	Mann-Whitney U-test	W=25382	n=10153	p<2.2e-^{16}
European Main vs. Chile	F_{ST}	Mann-Whitney U-test	W=25382	n=10153	p<2.2e-^{16}

S7. The effect of different sliding window lengths and the inclusion of deletion sizes for the results of π and Θ_w were tested with European Main and Catalonia.

A. The genome-wide average π values using shared positions for all four supercolonies with different window sizes and the “no discard deletions” parameter.

Window size	**Sample**	**European Main**	**Catalonia**
20 kbp		0.0008	0.0008
15 kbp		0.0008	0.0008
10 kbp		0.0009	0.0008
5 kbp		0.0009	0.0009
“no discard deletions”		0.0008	0.0008

B. The genome-wide average Θ_w values using shared positions for all four supercolonies with different window sizes and the “no discard deletions” parameter.

Window size	**Sample**	**European Main**	**Catalonia**
20 kbp		0.0007	0.0009
15 kbp		0.0007	0.0009
10 kbp		0.0008	0.0009
5 kbp		0.0008	0.0010
“no discard deletions”		0.0007	0.0009

S8. The genome-wide minimum, maximum, median, mean and standard deviation (sd) values of Tajima's D from the 20 kbp windows and the number of SNPs used for calculations, using shared positions including and excluding Catalonia.

Sample	Min	Max	Median	Mean	Sd	SNPs
Including Catalonia						
Native	-2.2225	1.9528	-0.5745	-0.5498	0.4698	350 745
European Main	-3.0459	4.2802	0.2661	0.0785	1.3514	169 276
Chile	-2.9635	3.6980	0.4598	0.3174	1.1302	176 517
Catalonia	-2.6092	1.2345	-0.8033	-0.7851	0.5374	185 461
Excluding Catalonia						
Native	-2.4585	2.0845	-0.3799	-0.3764	0.3978	1 537 157
European Main	-3.0226	3.3766	0.3550	0.1242	1.2234	722 735
Chile	-3.0038	2.9987	0.5215	0.3258	1.0182	732 980

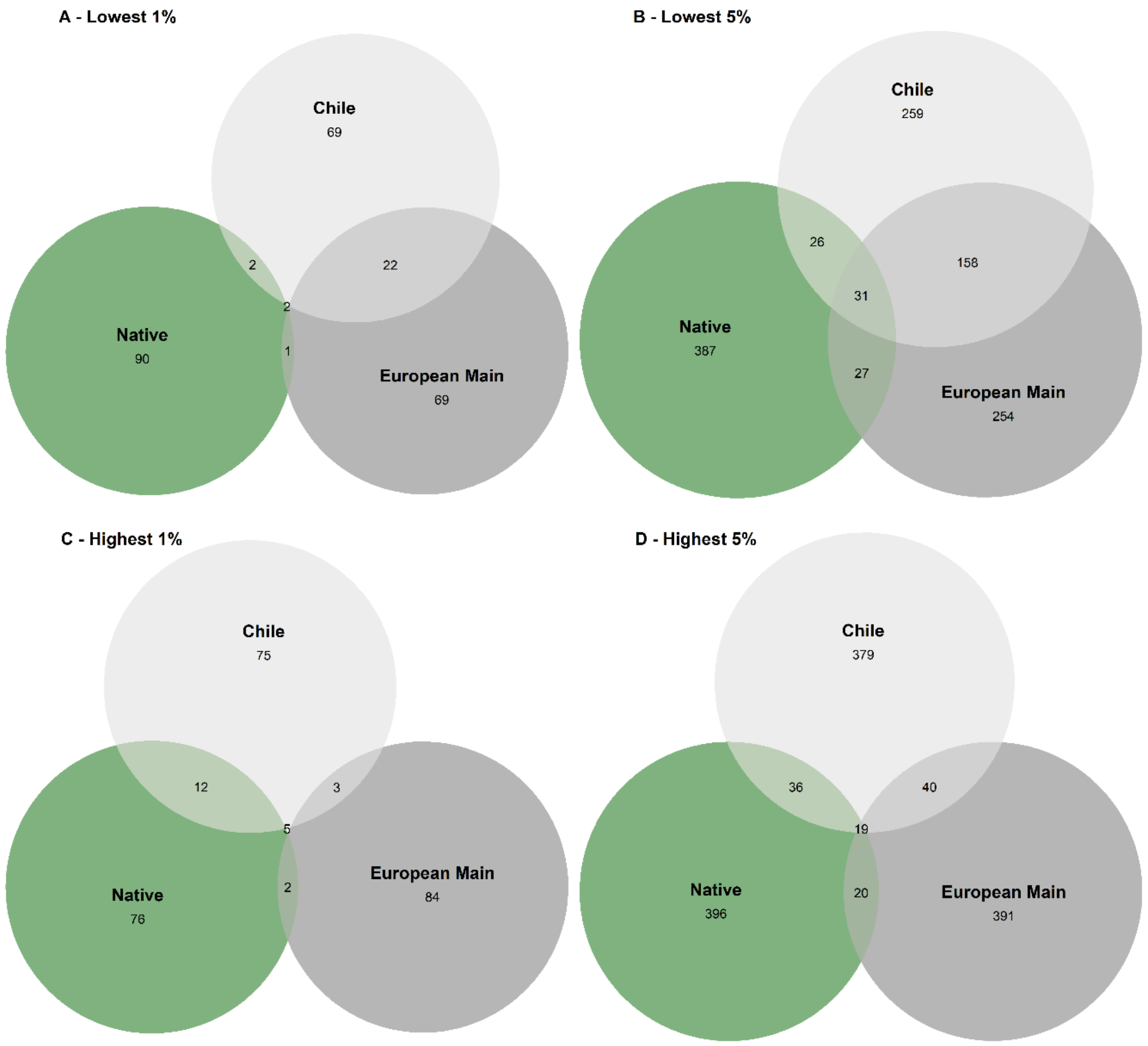
S9. The minimum, maximum, median, mean and standard deviation (sd) values of Tajima's D over all genes in the genome and the number of SNPs used for calculations, using shared positions including and excluding Catalonia.

Sample	**Min**	**Max**	**Median**	**Mean**	**Sd**	**SNPs**
Including Catalonia						
Native	-2.9099	3.6887	-0.4702	-0.3864	0.7811	263 790
European Main	-2.9700	4.5208	0.0205	0.0959	1.3370	111 103
Chile	-3.0177	3.8728	0.3572	0.3232	1.1051	115 092
Catalonia	-2.8815	2.4116	-0.6984	-0.6152	0.7802	126 149
Excluding Catalonia						
Native	-2.9099	3.6856	-0.2990	-0.2569	0.6839	1 188 303
European Main	-3.0002	4.2852	0.1538	0.1866	1.2462	499 276
Chile	-2.9465	4.1395	0.3963	0.3699	1.0276	516 415

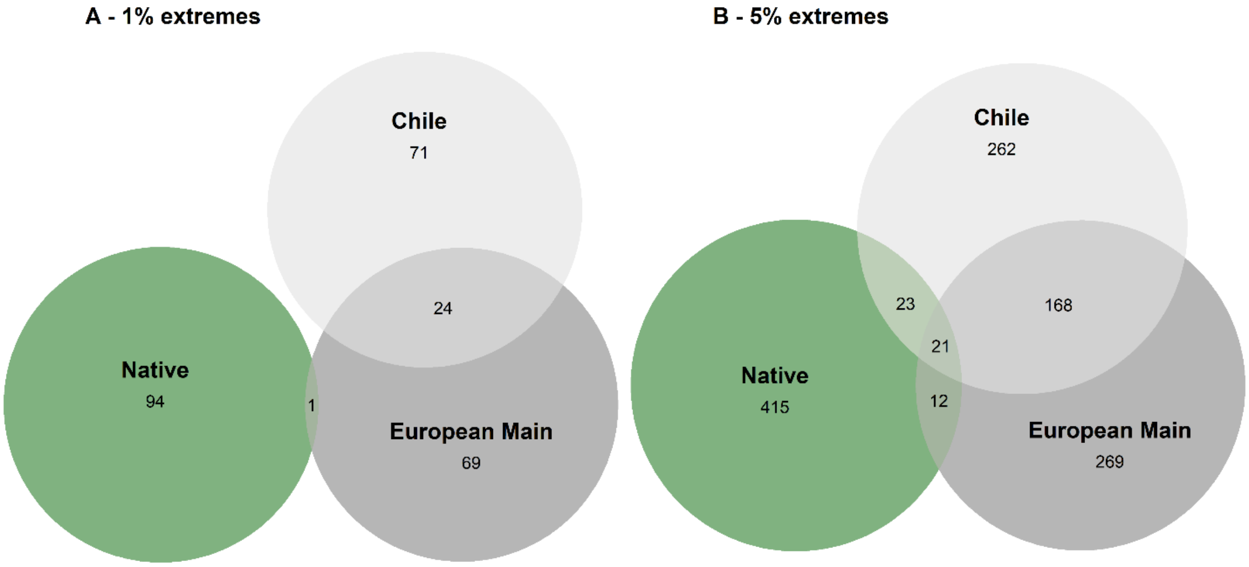
S10. See the separate Excel file (file S10) – the complete lists of all 20 kbp genomic windows and annotated genes of each supercolony sample with the lowest and highest Tajima's D values and the overlaps of these lists.

S11. Overlaps of the targets of selection – 20 kbp windows

A. Venn diagrams showed the overlaps of the 20 kbp windows in the (A) lowest 1%, (B) lowest 5%, (C) highest 1%, and (D) highest 5% extremes of Tajima's D values, excluding Catalonia.



B. Venn diagrams showed the overlaps of the 20 kbp windows with the highest Tajima's D of Native and the lowest Tajima's D values of the invasive supercolonies in (A) 1% extremes and (B) 5% extremes, excluding Catalonia.



S12. Pairwise Jaccard Indexes for 20 kbp genomic windows in the lowest 1% and 5% and highest 1% and 5% extremes of Tajima's D values.

The extreme end of Tajima's D distribution	**Sample pair**	**Jaccard Index**
Low 1%	Native & European Main	0.0161
Low 1%	Native & Chile	0.0215
Low 1%	European Main & Chile	0.1455
Low 5%	Native & European Main	0.0657
Low 5%	Native & Chile	0.0642
Low 5%	European Main & Chile	0.2503
High 1%	Native & European Main	0.0385
High 1%	Native & Chile	0.0983
High 1%	European Main & Chile	0.0442
High 5%	Native & European Main	0.0432
High 5%	Native & Chile	0.0618
High 5%	European Main & Chile	0.0667

S13. Enrichment Analysis of GO terms

GO ID	GO Name	GO Category	FDR (BH-adj. pval)	P-value	Nr tests	Test set
GO:0015074	DNA integration	BIOLOGICAL_PROCESS	0.0073	<0.0001	7	Native – the highest 1% of Tajima's D
GO:0015074	DNA integration	BIOLOGICAL_PROCESS	0.0039	<0.0001	14	Native – the highest 5% of Tajima's D
GO:0004984	olfactory receptor activity	MOLECULAR_FUNCTION	0.0012	<0.0001	26	European Main – the highest 5% of Tajima's D
GO:0007608	sensory perception of smell	BIOLOGICAL_PROCESS	0.0024	<0.0001	26	European Main – the highest 5% of Tajima's D
GO:0005549	odorant binding	MOLECULAR_FUNCTION	0.0033	<0.0001	26	European Main – the highest 5% of Tajima's D
GO:0006259	DNA metabolic process	BIOLOGICAL_PROCESS	0.0058	<0.0001	30	Chile – the highest 5% of Tajima's D
GO:0015074	DNA integration	BIOLOGICAL_PROCESS	0.0128	<0.0001	4	Native & European Main – the highest 1% of Tajima's D
GO:0015074	DNA integration	BIOLOGICAL_PROCESS	0.0064	<0.0001	6	Native & European Main – the highest 5% of Tajima's D
GO:0015074	DNA integration	BIOLOGICAL_PROCESS	0.0133	<0.0001	5	Native & Chile – the highest 5% of Tajima's D
GO:0015074	DNA integration	BIOLOGICAL_PROCESS	0.0001	<0.0001	8	Native, European Main & Chile – the highest 5% of Tajima's D

S14. The genome-wide minimum, maximum, median, mean, and standard deviation (sd) values of pairwise F_{ST}, including and excluding Catalonia.

Sample pair	**Min**	**Max**	**Median**	**Mean**	**Sd**
Including Catalonia					
Native – European Main	0.0089	0.6848	0.1617	0.1747	0.076498
Native – Catalonia	0.1020	0.6260	0.2691	0.2761	0.059928
Native – Chile	0.0086	0.5201	0.1317	0.1422	0.0564287
European Main – Catalonia	0.0721	0.9529	0.4332	0.4432	0.1190575
European Main – Chile	0.0000	0.9979	0.2379	0.2701	0.1605818
Catalonia – Chile	0.1027	0.9115	0.4133	0.4245	0.1017088
Excluding Catalonia					
Native – European Main	0.0110	0.7059	0.1569	0.1677	0.0817383
Native – Chile	0.0107	0.6605	0.1323	0.1419	0.0652877
European Main – Chile	0.0000	1.0000	0.2207	0.2482	0.1620753

S15. See separate Excel file (file S15) – the complete lists of all 20 kbp genomic windows and annotated genes with the highest F_{ST} values between sample pairs.

S16. Enrichment Analysis of GO terms (F_{ST})

GO ID	**GO Name**	**GO Category**	**FDR (BH-adj. pval)**	**P-value**	**Nr tests**	**Test set**
GO:0005515	protein binding	MOLECULAR_FUNCTION	0.0025	<0.0001	85	European Main & Chile– the most differentiated genes (5%)
GO:0005215	transporter activity	MOLECULAR_FUNCTION	0.0131	<0.0001	28	European Main & Chile– the most differentiated genes (5%)
GO:0006816	calcium ion transport	BIOLOGICAL_PROCESS	0.0220	<0.0001	5	European Main & Chile– the most differentiated genes (5%)
GO:0098662	inorganic cation transmembrane transport	BIOLOGICAL_PROCESS	0.0427	<0.0001	10	European Main & Chile– the most differentiated genes (5%)
GO:0034703	cation channel complex	CELLULAR_COMPONENT	0.0427	<0.0001	5	European Main & Chile– the most differentiated genes (5%)
GO:0005262	calcium channel activity	MOLECULAR_FUNCTION	0.0474	0.0001	4	European Main & Chile– the most differentiated genes (5%)