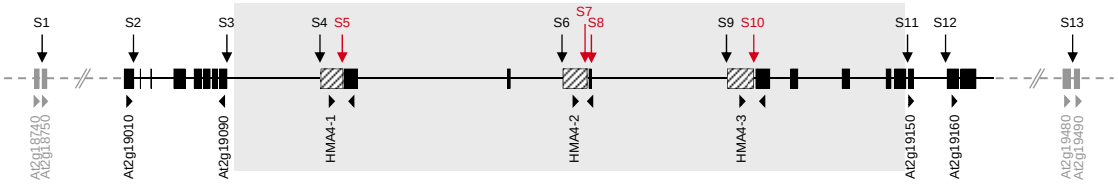
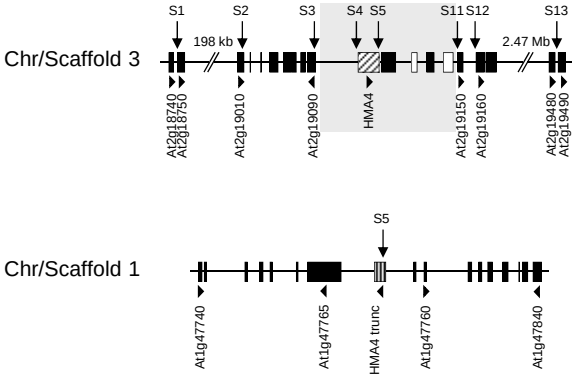
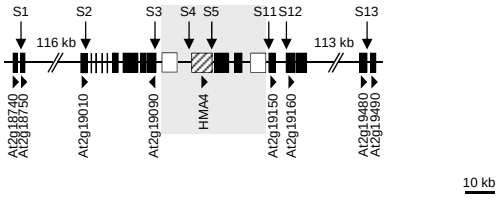
A *Arabidopsis halleri* (Chr/LG 3)



B *Arabidopsis lyrata*

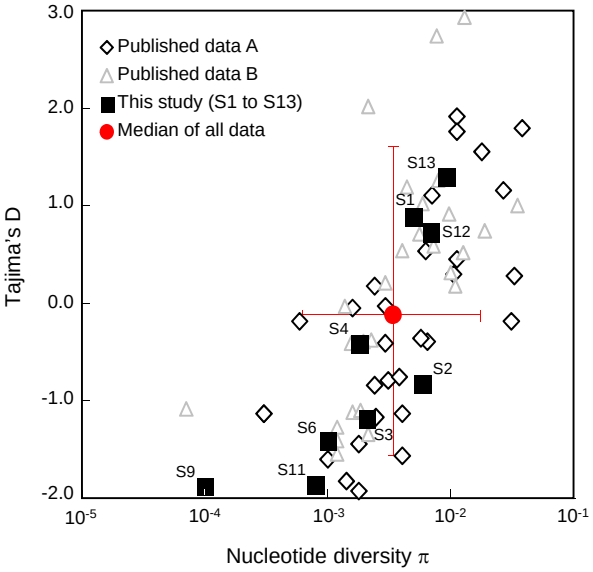


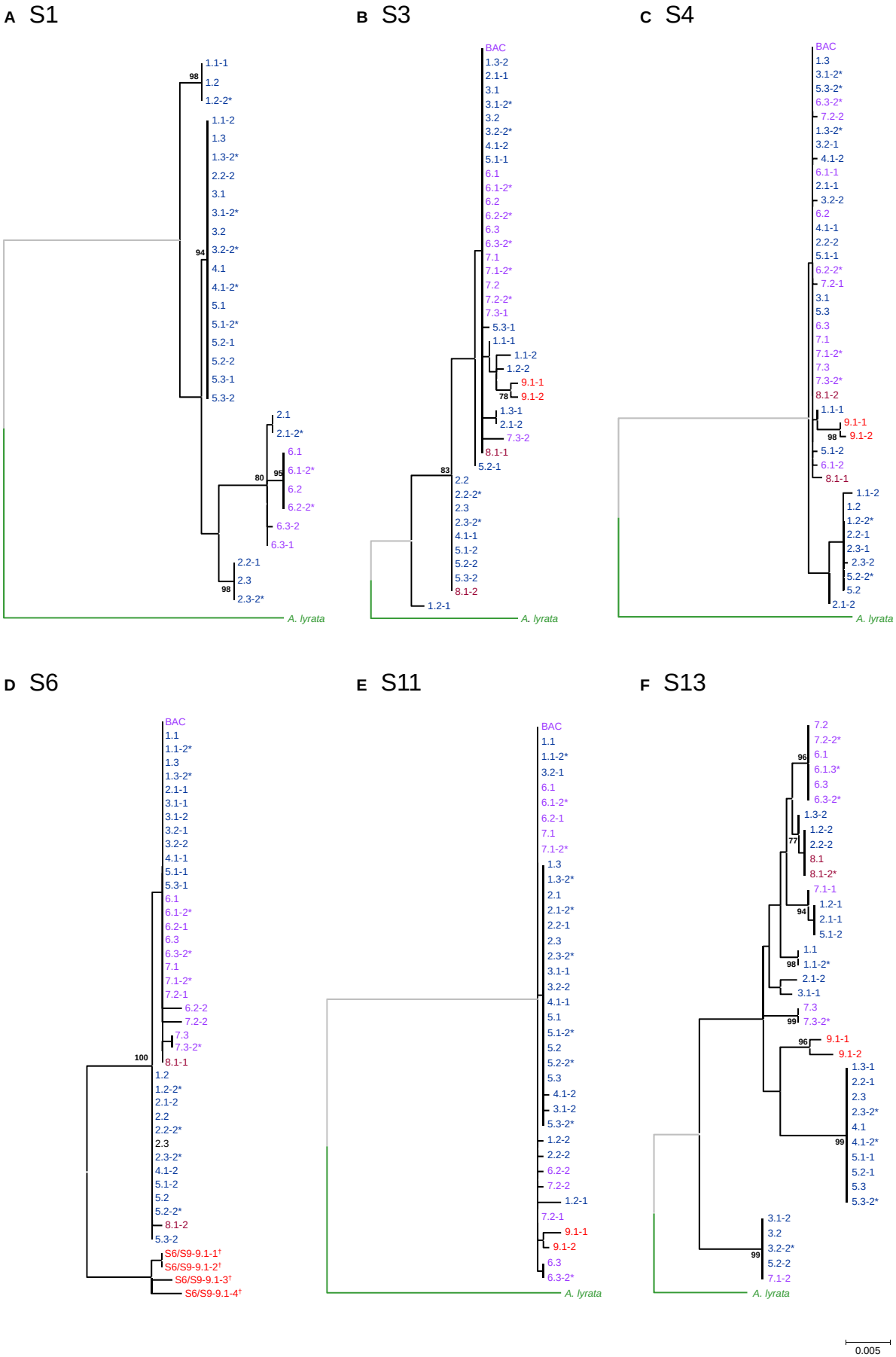
C *Arabidopsis thaliana* (Chr 2)

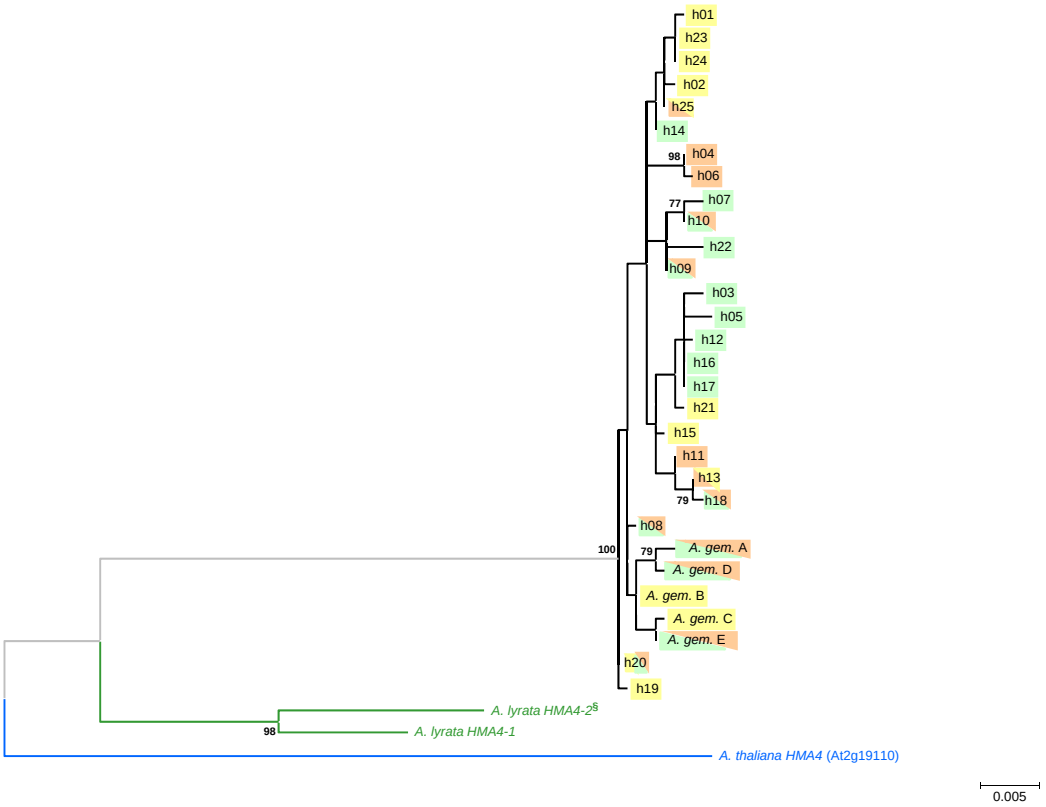


D

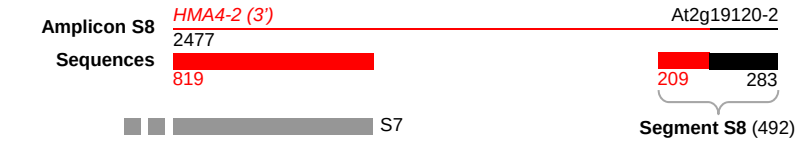
Segments			S4								S6		S9		S6/S9	
		Size (bp)	*At*		*Al*		*Ah* ssp. *h*		*Ah* ssp. *g*		*Ah* ssp. *h*		*Ah* ssp. *h*		*Ah* ssp. *g*	
			id	cov	id	cov	id	cov	id	cov	id	cov	id	cov	id	cov
S4	*At*	1549	100	100	80.6	99	80.1	99	79.4	78	74.7	29	71.3	56	71.3-71.9	37
	Al	1885			100	100	88.3	100	88.7	100	74.8	51	74.5	60	74.4	58
	Ah ssp. *h*	2000					100	100	99.3	100	77.0	39	74.1	62	73.9-74	60
	Ah ssp. *g*	1534							100	100	76.9	47	73.4	75	73.9	73.9-74
S6	*Ah* ssp. *h*	2245									100	100	92.2	85	91.2-91.6	100
S9	*Ah* ssp. *h*	3372											100	100	97.1-98.2	96-97
S6/S9	*Ah* ssp. *g*	2818-3372													98.3-99.8	100


Published data A
Published data B
This study (S1 to S13)
Median of all data
S13
S1
S12
S4
S2
S6
S3
S9
S11
3.0
2.0
1.0
0.0
-1.0
-2.0
Tajima's D
10-5
10-4
10-3
10-2
10-1
Nucleotide diversity π

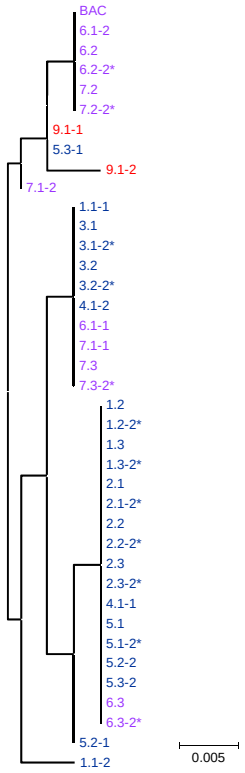

A S1
1.1-1
1.2
1.2-2*
1.1-2
1.3
1.3-2*
2.2-2
3.1
3.1-2*
3.2
3.2-2*
4.1
4.1-2*
5.1
5.1-2*
5.2-1
5.2-2
5.3-1
5.3-2
2.1
2.1-2*
6.1
6.1-2*
6.2
6.2-2*
6.3-2
6.3-1
2.2-1
2.3
2.3-2*
A. lyrata
B S3
BAC
1.3-2
2.1-1
3.1
3.1-2*
3.2
3.2-2*
4.1-2
5.1-1
6.1
6.1-2*
6.2
6.2-2*
6.3
6.3-2*
7.1
7.1-2*
7.2
7.2-2*
7.3-1
5.3-1
1.1-1
1.1-2
1.2-2
9.1-1
9.1-2
1.3-1
2.1-2
7.3-2
8.1-1
5.2-1
2.2
2.2-2*
2.3
2.3-2*
4.1-1
5.1-2
5.2-2
5.3-2
8.1-2
1.2-1
A. lyrata
C S4
BAC
1.3
3.1-2*
5.3-2*
6.3-2*
7.2-2
1.3-2*
3.2-1
4.1-2
6.1-1
2.1-1
3.2-2
6.2
4.1-1
2.2-2
5.1-1
6.2-2*
7.2-1
3.1
5.3
6.3
7.1
7.1-2*
7.3
7.3-2*
8.1-2
1.1-1
9.1-1
9.1-2
5.1-2
6.1-2
8.1-1
1.1-2
1.2
1.2-2*
2.2-1
2.3-1
2.3-2
5.2-2*
5.2
2.1-2
A. lyrata
D S6
BAC
1.1
1.1-2*
1.3
1.3-2*
2.1-1
3.1-1
3.1-2
3.2-1
3.2-2
4.1-1
5.1-1
5.3-1
6.1
6.1-2*
6.2-1
6.3
6.3-2*
7.1
7.1-2*
7.2-1
6.2-2
7.2-2
7.3
7.3-2*
8.1-1
1.2
1.2-2*
2.1-2
2.2
2.2-2*
2.3
2.3-2*
4.1-2
5.1-2
5.2
5.2-2*
8.1-2
5.3-2
S6/S9-9.1-1†
S6/S9-9.1-2†
S6/S9-9.1-3†
S6/S9-9.1-4†
E S11
BAC
1.1
1.1-2*
3.2-1
6.1
6.1-2*
6.2-1
7.1
7.1-2*
1.3
1.3-2*
2.1
2.1-2*
2.2-1
2.3
2.3-2*
3.1-1
3.2-2
4.1-1
5.1
5.1-2*
5.2
5.2-2*
5.3
4.1-2
3.1-2
5.3-2*
1.2-2
2.2-2
6.2-2
7.2-2
1.2-1
7.2-1
9.1-1
9.1-2
6.3
6.3-2*
A. lyrata
F S13
7.2
7.2-2*
6.1
6.1.3*
6.3
6.3-2*
1.3-2
1.2-2
2.2-2
8.1
8.1-2*
7.1-1
1.2-1
2.1-1
5.1-2
1.1
1.1-2*
2.1-2
3.1-1
7.3
7.3-2*
9.1-1
9.1-2
1.3-1
2.2-1
2.3
2.3-2*
4.1
4.1-2*
5.1-1
5.2-1
5.3
5.3-2*
3.1-2
3.2
3.2-2*
5.2-2
7.1-2
A. lyrata
0.005


h01
h23
h24
h02
h25
h14
98
h04
h06
77
h07
h10
h22
h09
h03
h05
h12
h16
h17
h21
h15
h11
h13
79
h18
h08
100
79
A. gem. A
A. gem. D
A. gem. B
A. gem. C
A. gem. E
h20
h19
A. lyrata HMA4-2§
98
A. lyrata HMA4-1
A. thaliana HMA4 (At2g19110)
0.005

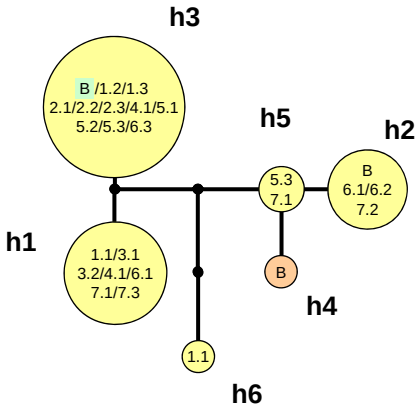
A



B

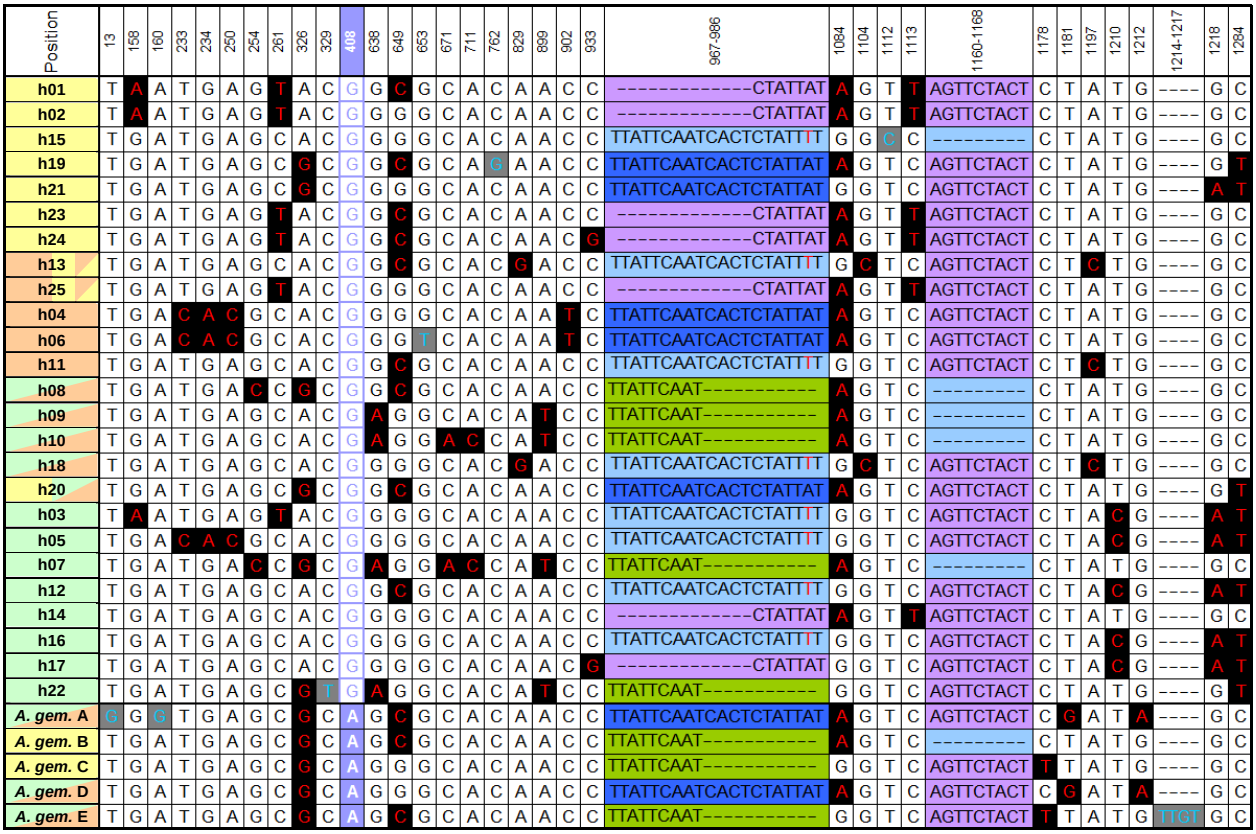


C

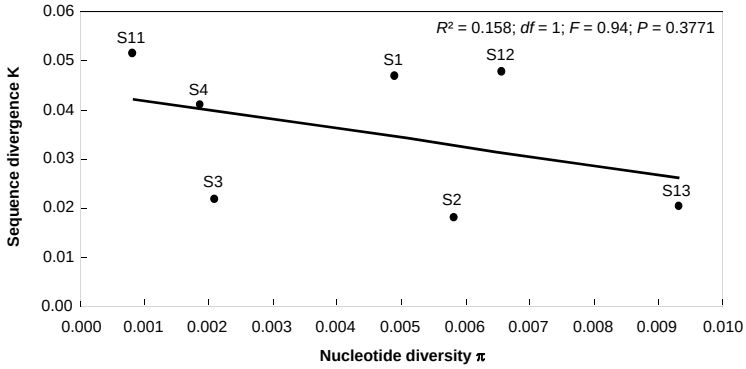


D

Position	4	7	15	62	88	127	133-163	173	189
h1	G	C	G	G	G	G	----------	A	C
h2	T	C	C	G	A	G	----------	G	C
h3	T	C	G	A	G	G	----------	A	C
h4	T	C	G	G	A	A	----------	G	C
h5	T	C	G	G	A	G	----------	G	C
h6	T	C	G	G	G	G	GCTCACCAAAAAAAAAAACATTCTCGTAAGA	G	T
***A. gem.* A**	T	C	G	G	A	G	----------	G	C
***A. gem.* B**	G	A	G	G	A	G	----------	G	C



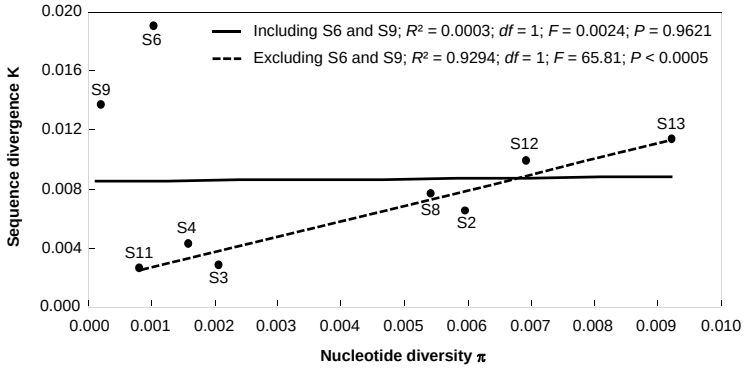
A *A. halleri* ssp. *halleri* compared to *A. lyrata* ssp. *lyrata*



Locus	# sites (excl. gaps)	π	K
S1	1546	0.00490	0.04686
S2	745	0.00582	0.01794
S3	1227	0.00209	0.02180
S4	1755	0.00188	0.04105
S6[a]	-	-	-
S8[a]	-	-	-
S9[a]	-	-	-
S11	1439	0.00083	0.05163
S12	1779	0.00657	0.04771
S13	1593	0.00932	0.02049

[a]There is only a single full *HMA4* gene copy in the *A. lyrata* genome, which corresponds to *AhHMA4-1* (S4); S6, S8 and S9 are missing in *A. lyrata*.

B *A. halleri* ssp. *halleri* compared to *A. halleri* ssp. *gemmifera*



Locus	# sites (excl. gaps)	π	K
S1[b]	-	-	-
S2	752	0.00596	0.00654
S3	1234	0.00208	0.00280
S4	1528	0.00160	0.00430
S6[c]	778	0.00104	0.01900
S8	470	0.00542	0.00769
S9[c]	2159	0.00011	0.01373
S11	1480	0.00081	0.00261
S12	1980	0.00692	0.00987
S13	1608	0.00923	0.01133

[b]Data are missing for S1 of *A. halleri* ssp. *gemmifera.*
[c]Both S6 and S9 primer pairs amplified the same four alleles in *A. halleri* ssp. *gemmifera*, which were used for estimation of K.

Table S1. Information on DNA amplicons and on the sequencing of cloned amplicons.

Amplicon		Composition		Target gene (position)	Annotation[2]	Missing individuals[3]	Sequences of PCR primers used for amplification[4]	
Name	Size (bp)[1]	Coding (bp)[1]	Non-coding (bp)[1]				Forward	Reverse
S1	1593	224	1369	At2g18750 (5')	Calmodulin-binding protein	7.1-7.3, 8.1, 9.1	CAACATAACTCTGATGATGAACACG	CTCGGCGAAGAATCGGTTCC
S2	1451	398	1053	At2g19010 (3')	GDSL-like Lipase/Acylhydrolase	-	GCACACCAAAGATAATGAAGTCACAC	GACTTGCTTAGCGGCTTCCTCG
S3	1305	540	765	At2g19090 (5')	Protein of unknown function	-	CGAAAATAAACATGAAAAGCGAGG	CCAAAACGAAGATGAAGTAGACGC
S4	1885	30	1855	*HMA4-1*$_P$[5] (5')	Heavy Metal ATPase 4	-	GACATTCCACTTTTGGGGGTTTCC	TTTCTCTTCTTCTTTGTTTTGTGACGCC
S5	1419	1301	118	*HMA4-1*[5] (3')	Heavy Metal ATPase 4	-	GCGATGATGATGATGCTGTGGAC	TCAAGCACTCACATGGTGATGGTG
S6	917	30	887	*HMA4-2*$_P$[5] (5')	Heavy Metal ATPase 4	[6]	GCTAAAAACACGCCGATTAAGAAG	TTTCTCTTCTTCTTTGTTTTGTGACGCC
S7	1406	1301	105	*HMA4-2*[5] (3')	Heavy Metal ATPase 4	[6]	GCGATGATGATGATGCTGTGGAC	TCAAGCACTCACATGGTGATGGTG
S8	2477[7]	1346	1131	*HMA4-2*[5] (3')/ At2g19120-2	Heavy Metal ATPase 4	8.1[6]	GACACATCTTACCTGGAGAAG	GCCAGAGTTATATGGTGATATG
S9	2245	30	2215	*HMA4-3*$_P$[5] (5')	Heavy Metal ATPase 4	3.2, 7.3[6]	TTAAAAGGGGTATTGAAAAAGGAGC	TTTCTCTTCTTCTTTGTTTTGTGACGCC
S10	1425	1307	118	*HMA4-3*[5] (3')	Heavy Metal ATPase 4	[6]	GCGATGATGATGATGCTGTGGAC	TCAAGCACTCACATGGTGATGGTG
S11	1326	458	868	At2g19150 (5')	Pectin lyase-like protein	7.3, 8.1	TGAAAATTTCAACAAATAGTGTACCG	GCAGCCACCGCTGGTTTGG
S12	2132	311	1821	At2g19160 (5')	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase	-	TAACATGCTTTCCTTTGACGGG	GAACTCTTTTTTATAACACGAGGGG
S13	1694	302	1392	At2g19490 (5')	recA DNA recombination protein	6.2	GGTCAAGCAGGTGAGAGGCCA	CCTCCTTGTTTTTGTGCTTCTGC

[1]Sizes based on *A. halleri* BAC sequences (Genbank accession numbers: EU382073.1, EU382072.1) [8], and on *A. thaliana* genome sequence (Genbank accession numbers: AC005724 and AC005917) for amplicons S1 and S13. Note that primers designed for amplicons S5, S7 and S10 non-specifically cross-amplified sequences from all three of these target gene copies in *A. halleri*; assignment to copies was then done based on S8 and according to Materials and Methods.

[2]TAIR 10 (http://www.arabidopsis.org).

[3]No PCR products were obtained for the given individuals despite attempts using a range of alternative primer pairs and PCR conditions.

[4]Amplicons were sequenced after cloning, using vector-specific primers M13f (CGCCAGGGTTTTCCCAGTCACGAC) and M13r (TCACACAGGAAACAGCTATGAC). When necessary, sequencing was completed using additional gene-specific primers.

[5]All *HMA4* gene copies correspond to At2g19110. $_P$: promoter region.

[6]There is only a single complete *HMA4* gene in the *A. lyrata* genome, which corresponds to *AhHMA4-1* (S4, S5). See also Figure S1.

[7]For S8, only the 3'-end of the amplicon encompassing 492 bp (209 bp coding and 283 bp non-coding[1]) was used as segment S8 in all analyses (see Figure S5A). The 5'-end of amplicon S8 (819 bp[1]) was designed to overlap with S7 and used to identify S7 sequences (see Materials and Methods; Figure S5A).

Red fonts: segments comprising repeated sequence stretches present in several, almost identical copies in the *HMA4* genomic region.

Table S2. Summary of statistics of nucleotide sequence diversity as shown in Figure 1.

		Segment												
Alleles	Statistics	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13
Sequenced and inferred	π	0.0049	0.0059	0.0021	0.0018	0.0042	0.0010	0.0032	0.0052	0.0001	0.0035	0.0008	0.0069	0.0091
(as shown in Fig. 1)	Tajima's *D*	0.877	-0.835	-1.182	-0.417	n.a.	-1.418	n.a.	n.a.	**-1.886**[§]	n.a.	**-1.863**[§]	0.720	1.294
	Fu and Li's *D**	**1.410**[§]	0.875	-2.283	-1.677	n.a.	-2.442	n.a.	n.a.	**-3.124**[§]	n.a.	**-3.437**[§]	1.082	**1.693**[§]
	Fu and Li's *F**	1.458	0.353	-2.266	-1.490	n.a.	-2.486	n.a.	n.a.	**-3.206**[§]	n.a.	**-3.450**[§]	1.136	**1.845**[§]
	# sites (excluding gaps)	1554	760	1234	1827	1320	841	1327	470	2189	1336	1480	1985	1622
	# haplotypes	7	6	10	13	n.a.	6	n.a.	8	5	n.a.	10	12	11
	N	30	39	39	39	35	39	34	37	35	34	35	39	36
Sequenced only	π	0.0048	0.0077	0.0024	0.0019	/	0.0010	/	0.0055	0.0002	/	0.0010	0.0071	0.0093
	Tajima's *D*	0.392	-0.466	-1.110	-0.491	/	-1.546	/	n.a.	**-1.978**[§]	/	**-1.901**[§]	0.676	1.076
	Fu and Li's *D**	1.164	0.943	-1.981	-1.430	/	**-2.870**[§]	/	n.a.	**-2.716**[§]	/	**-3.404**[§]	1.090	0.905
	Fu and Li's *F**	1.088	0.596	-2.003	-1.331	/	**-2.882**[§]	/	n.a.	**-2.864**[§]	/	**-3.443**[§]	1.126	1.128
	# sites (excluding gaps)	1554	760	1234	1827	/	841	/	470	2189	/	1480	1985	1622
	# haplotypes	7	6	10	13	/	6	/	8	5	/	10	12	11
	N	20	24	30	30	/	29	/	25	22	/	25	33	26
No. of alleles	BAC	-	1	1	1	1	1	1	1	1	1	1	1	-
observed per genotype	1.1	2	2	2	2	1	1	2	2	2	1	1	2	1
	1.2	1	2	2	1	1	1	2	1	1	1	2	2	2
	1.3	1	1	2	1	1	1	2	1	2	2	1	2	2
	2.1	1	1	2	2	1[+]	2	2	1	1	1[+]	1	2	2
	2.2	2	2	1	2	1	1	1	1	1	1	2	2	2
	2.3	1	1	1	2	1	1	1	1	1	1	1	1	1
	3.1	1	1	1	1	2	2	2	1	1	2	2	2	2
	3.2	1	1	1	2	2	2	2	1	-	2	2	2	1
	4.1	1	1	2	2	2	2	2	2	1	2	2	1	1
	5.1	1	1	2	2	2	2	2	1	2	2	1	2	2
	5.2	2	1	2	1	2	1	2	2	1	2	1	1	2
	5.3	2	1	2	1	2	2	2	2	1	2	1	1	1
	6.1	1	1	1	2	2	1	2	2	1	1	1	2	1
	6.2	1	1	1	1	1	2	1	1	1	1	2	1	-
	6.3	2	1	1	1	2	1	2	1	1	2	1	2	1
	7.1	-	1	1	1	2	1	2	2	1	1	1	2	2
	7.2	-	1	1	2	2	2	2	1	2	2	2	1	1
	7.3	-	2	2	1	2	1	2	1	-	2	-	2	1
	8.1	-	1	2	2	2	2	1	-	1	1	-	2	1
	A. halleri ssp. *gemmifera* (9.1)[1]	-	1	2	2	1[+]	4*	2	2	4*	1[+]	2	2	2
	A. lyrata ssp. *lyrata*[1]	1	1	1	1	1	x	x	x	x	x	1	1	1

Values for BAC and *A. lyrata* are based on public data [8,39]. Genotypes for which only a single consensus sequence was obtained were concluded to be homozygous, and the second allele was inferred. Assignment to *HMA4* gene copy of S5, S7 and S10 was done by inference upon the assignment to S7 through anchoring *via* overlapping S8 sequences according to Materials and Methods (marked "/"; [+]: ambiguous assignment to S5 or S10). Grey shade: amplicons in triplicated region (*HMA4* and two downstream genes, see Figure 1). Red fonts: segments comprising repeated sequence stretches present in several, almost identical copies in the *HMA4* genomic region. n.a.: not applicable; sequences did not evolve independently because of ectopic gene conversion, thus precluding tests of molecular population genetics. [§]: $P < 0.05$. -: no PCR product obtained. x: not present in the genome. *: Primer pairs designed to obtain S6 and S9 (see Table S1) both yielded the same set of four highly similar sequences (see Figure S1). [1]Parameters for the *A. halleri* ssp. *gemmifera* and *A. lyrata* ssp. *lyrata* individuals are only shown for comparison; their sequences were not included in the analyses presented above and in Figure 1.

Table S3. Identification within amplicons of repeated sequence stretches present in several copies in the *HMA4* genomic region.

A. Sequence similarities between amplified segments and sequences within BAC contig of the *A. halleri* genomic *HMA4* region.

Amplicon	Sequence identity (%)[1]	Hit length (bp)[1]	Coordinates in amplicon	Coordinates in *HMA4* BAC contig	Annotation
S2	67	336	1094-1413 r.c.	22145-22456	intergenic region At2g19050/At2g19060[2]
S3	-	-	-	-	-
S4	84	310	4-298	197971-198272	*AhHMA4-3*$_P$
	77	364	1111-1471	146010-146336	*AhHMA4-2*$_P$
S5	99	1425	1-1419	208610-210034	*AhHMA4-3* 3'-end
	99	1419	1-1419	153021-154436	*AhHMA4-2* 3'-end
S6	88	753	224-917	200743-201470	*AhHMA4-3*$_P$
	77	364	317-643	64760-65120	*AhHMA4-1*$_P$
S7	99	1419	1-1406	71255-72673	*AhHMA4-1* 3'-end
	99	1407	1-1394	208610-210016	*AhHMA4-3* 3'-end
S8	95	2263	1-2194	209197-211416	intergenic region *AhHMA4-3*/At2g19120-3
	90	2208	1-2194	71842-73865	intergenic region *AhHMA4-1*/At2g19120-1
S9	80	2589	43-2245	144154-146630	*AhHMA4-2*$_P$
	77	313	1714-1989	64809-65120	*AhHMA4-1*$_P$
S10	99	1407	1-1407	71255-72661	*AhHMA4-1* 3'-end
	99	1407	1-1407	153021-154414	*AhHMA4-2* 3'-end
S11	76	103	133-234	61738-61639	*AhHMA4-1*$_P$
S12	77	71	410-480	167443-167509	intergenic region At2g19120-2(c)/*AhHMA4-3*

[1]Sequence identities and hit lengths are given for the top-scoring non-self hits running amplicon sequences (predicted based on BAC contig) as queries against the *HMA4* BAC contig in BLASTN with default settings at NCBI (Genbank accession numbers EU382073.1 and EU382072.1; contig length 289,768 bp [8]).

[2]All three genes (At2g19010/S2, At2g19050 and At2g19060) encode putative GDSL-motif lipase/hydrolase family proteins.

Note that sequences corresponding to S1 and S13 are not present in the *HMA4* BAC contig. Red fonts: segments comprising repeated sequence stretches present in several, almost identical copies in the genomic *HMA4* region. r.c.: reverse complement. -: no significant hit.

B. Sequence similarities between amplified segments of *A. halleri* and sequences in the *A. thaliana* and *A. lyrata* ssp. *lyrata* genomes.

Top hits	*A. thaliana*			*A. lyrata*		
Loci	Sequence identity (%)[1]	Hit length (nt)[1]	Annotation	Sequence identity (%)[1]	Hit length (nt)[1]	Annotation
S1	92	328	At2g19750	96	149	GENE ID: 9322332 ARALYDRAFT_900373
S2	96	217	At2g19010	99	217	GENE ID: 9305466 ARALYDRAFT_491739
S3	92	546	At2g19090	94	540	GENE ID: 9322308 ARALYDRAFT_480823
S4	91	64	5'-UTR of *AtHMA4*	97	33	GENE ID: 9322307 ARALYDRAFT_480822[2]
S5	92	690	*AtHMA4*	95	692	GENE ID: 9322307 ARALYDRAFT_480822[2]
				95	692	GENE ID: 9330132 ARALYDRAFT_473917[3]
S6	79	186	BAC t20k24 (in *AtHMA4* promoter)	85	197	*Arabidopsis lyrata* clone JGIFAFI-61E14
S7	92	690	*AtHMA4*	95	692	GENE ID: 9322307 ARALYDRAFT_480822[2]
				95	692	GENE ID: 9330132 ARALYDRAFT_473917[3]
S8	96	632	At2g19120	-	-	-
S9	71	295	BAC t20k24 (in *AtHMA4* promoter)	82	95	GENE ID: 9307927 ARALYDRAFT_909928
S10	92	690	*AtHMA4*	95	692	GENE ID: 9322307 ARALYDRAFT_480822[2]
				95	692	GENE ID: 9330132 ARALYDRAFT_473917[3]
S11	92	235	At2g19150	98	236	GENE ID: 9322304 ARALYDRAFT_480819
S12	92	450	At2g19160	90	454	GENE ID: 9320242 ARALYDRAFT_480818
S13	96	222	At2g19490	83	289	GENE ID: 9320055 ARALYDRAFT_480524
Second hits	*A. thaliana*			*A. lyrata*		
Loci	Sequence identity (%)[1]	Hit length (nt)[1]	Annotation	Sequence identity (%)[1]	Hit length (nt)[1]	Annotation
S1	89	162	At2g19740	75	226	Flowering locus T promoter (FN813300.1)
S2	72	186	At2g19060	71	211	GDSL-motif lipase/hydrolase family protein
	< 80	variable	multiple hits[4]			
S3	84	224	At4g30130 (segmental duplication)	73	102	GENE ID: 9305000 ARALYDRAFT_656131
S4	-	-	-	-	-	-
S5	70	207	At4g30110 - *AtHMA2* (segmental duplication)	70	207	GENE ID: 9303439 ARALYDRAFT_657042
S6	-	-	-	-	-	-
S7		207	At4g30110 - *AtHMA2* (segmental duplication)	70	207	GENE ID: 9303439 ARALYDRAFT_657042
S8	86	556	At4g30100 (segmental duplication)	-	-	-
S9	-	-	-	-	-	-
S10		207	At4g30110 - *AtHMA2* (segmental duplication)	70	207	GENE ID: 9303439 ARALYDRAFT_657042
S11	76	158	At2g47280	-	-	-
S12	77	339	At4g30060 (segmental duplication)	76	366	GENE ID: 9305472 ARALYDRAFT_491750
S13	86	327	At2g19480	88	41	GENE ID: 9328930 ARALYDRAFT_471744

[1]Sequence identities and hit lengths are given for the two top-scoring hits from BLASTN with default settings using amplicon sequences as queries against all *A. thaliana* and *A. lyrata* ssp. *lyrata* sequences at NCBI.

[2]*AlHMA4*.

[3]Truncated pseudogene copy of *HMA4* present only in *A. lyrata* and corresponding to the 3'-end of *HMA4* (see Figure S1).

[4]Part of the At2g19010/At19020 intergenic region of the *A. halleri* S2 sequences contains repeats found at multiple sites on multiple chromosomes of *A. thaliana*.

Red fonts: segments comprising repeated sequence stretches present in several, almost identical copies in the *HMA4* genomic region of *A. halleri*. r.c.: reverse complement.-: no significant hit.

Table S4. Robustness of π estimation.

A. Estimation of error in π resulting from uncertainties in the assignment of sequences to *HMA4* gene copy (for S5, S7 and S10).

	Genotypes								
	All *A. halleri* ssp. *halleri*			Harz Mountains			Thuringian Forest		
Locus	*N*	No. of sites	π	*N*	No. of sites	π	*N*	No. of sites	π
S5 only (*HMA4-1*)	35	1320	0.0042	22	1320	0.0042	11	1336	0.0025
S5 + ambiguous S5/S10	40	1320	0.0043	25	1320	0.0042	13	1336	0.0033
S7 anchored based on S8	34	1327	0.0032	22	1327	0.0032	10	1327	0.0020
S7 only (*HMA4-2*)	37	1327	0.0031	23	1327	0.0033	12	1327	0.0019
S7 + ambiguous S7/S10	39	1327	0.0032	25	1327	0.0033	12	1327	0.0019
S10 only (*HMA4-3*)	34	1336	0.0035	21	1336	0.0039	11	1349	0.0035
S10 + ambiguous S5/S10	41	1320	0.0036	26	1320	0.0039	13	1349	0.0036

No. of sites are numbers of informative sites, excluding those with gaps. Sequences that could not be assigned to a single *HMA4* gene copy are named ambiguous (see Figure 4, Figure S4).

B. Dependence of π on variation in sampling.

	Segment												
Genotypes	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13
All *A. halleri* ssp. *halleri*	0.0049	0.0059	0.0021	0.0018	0.0042	0.0010	0.0032	0.0052	0.0001	0.0035	0.0008	0.0069	0.0091
Harz Mountains	0.0033	0.0063	0.0026	0.0023	0.0042	0.0006	0.0032	0.0034	0.0001	0.0039	0.0007	0.0068	0.0098
Thuringian Forest	0.0012	0.0066	0.0004	0.0003	0.0025	0.0011	0.0020	0.0062	0.0001	0.0035	0.0005	0.0057	0.0061

Grey shade: amplicons in triplicated region (*HMA4* and two downstream genes, see Figure 1). Red fonts: segments comprising repeated sequence stretches present in several, almost identical copies in the *HMA4* genomic region.

Table S5. Analyses of Molecular Variance (AMOVA) in *A. halleri*.

A. Collection sites were assigned to three geographic regions (referred to as groups): Harz Mountains (*A. halleri* ssp. *halleri*: 1 to 5; see Table 1), Thuringian Forest (*A. halleri* ssp. *halleri*: 6, 7), and Japan (*A. halleri* ssp. *gemmifera*: 9).

Segment	Source of variation	*df*	Sum of squares	Variance components	Percentage of variation	Fixation Indices	*P* value
S1	Between groups	1	44.8	3.76	57.1	F_{CT}: 0.57	0.1515 (± 0.0137)
	Between sites within groups	4	29.1	1.21	18.4	F_{SC}: 0.43	0.0049 (± 0.0020)
	Within sites	24	38.8	1.62	24.5	F_{ST}: 0.75	$< 5{\cdot}10^{-6}$ ($\pm < 5{\cdot}10^{-6}$)
	Total	29	112.8	6.59			
S2	Between groups	2	26.0	0.77	22.0	F_{CT}: 0.22	0.1652 (± 0.0126)
	Between sites within groups	5	33.2	0.99	28.3	F_{SC}: 0.36	0.0059 (± 0.0022)
	Within sites	30	52.2	1.74	49.7	F_{ST}: 0.50	$< 5{\cdot}10^{-6}$ ($\pm < 5{\cdot}10^{-6}$)
	Total	37	111.3	3.50			
S3	Between groups	2	10.3	0.35	21.4	F_{CT}: 0.21	0.0254 (± 0.0040)
	Between sites within groups	5	10.2	0.19	11.7	F_{SC}: 0.15	0.1241 (± 0.0128)
	Within sites	30	32.7	1.09	66.8	F_{ST}: 0.33	$< 5{\cdot}10^{-6}$ ($\pm < 5{\cdot}10^{-6}$)
	Total	37	53.2	1.63			
S4	Between groups	2	17.4	0.69	30.6	F_{CT}: 0.31	0.0078 (± 0.0031)
	Between sites within groups	5	11.9	0.21	9.4	F_{SC}: 0.14	0.2033 (± 0.0153)
	Within sites	30	40.2	1.34	59.9	F_{ST}: 0.40	0.0010 (± 0.0010)
	Total	37	69.6	2.24			
S6[a]	Between groups	1	2.4	0.12	25.2	F_{CT}: 0.25	0.1134 (± 0.0098)
	Between sites within groups	5	2. 7	0.05	10.3	F_{SC}: 0.14	0.0635 (± 0.0065)
	Within sites	29	8. 7	0.30	64.5	F_{ST}: 0.35	0.0020 (± 0.0014)
	Total	35	13. 8	0.46			
S9[a]	Between groups	1	0.1	0.00	0.8	F_{CT}: 0.01	0.6188 (± 0.0151)
	Between sites within groups	5	0.5	-0.07	-5.1	F_{SC}: -0.05	0.6901 (± 0.0084)
	Within sites	25	3.2	0.13	104.2	F_{ST}: -0.04	0.6843 (± 0.0154)
	Total	31	3.9	0.12			
S11	Between groups	2	8.8	0.44	44.6	F_{CT}: 0.45	0.0049 (± 0.0020)
	Between sites within groups	5	3.5	0.04	4.2	F_{SC}: 0.08	0.0841 (± 0.0082)
	Within sites	28	14.1	0.50	51.2	F_{ST}: 0.49	$< 5{\cdot}10^{-6}$ ($\pm < 5{\cdot}10^{-6}$)
	Total	35	26.4	0.98			
S12	Between groups	2	45.0	1.23	15.3	F_{CT}: 0.14	0.0860 (± 0.0077)
	Between sites within groups	5	60.2	1.33	16.5	F_{SC}: 0.20	0.0293 (± 0.0048)
	Within sites	30	164.7	5.49	68.2	F_{ST}: 0.31	$< 5{\cdot}10^{-6}$ ($\pm < 5{\cdot}10^{-6}$)
	Total	37	270.0	8.06			
S13	Between groups	2	40.8	0.74	8.6	F_{CT}: 0.09	0.2248 (± 0.0104)
	Between sites within groups	5	81.9	2.32	27.2	F_{SC}: 0.30	0.0078 (± 0.0034)
	Within sites	28	153.2	5.47	64.1	F_{ST}: 0.36	$< 5{\cdot}10^{-6}$ ($\pm < 5{\cdot}10^{-6}$)
	Total	35	275.9	8.53			

B. Collection sites were assigned to four geographic regions (referred to as groups): Groups were as in (**a**) above, and additionally including France as a fourth region (group) represented by Auby (*A. halleri* ssp. *halleri*: 8; see Table 1).

Segment	Source of variation	*df*	Sum of squares	Variance components	Percentage of variation	Fixation Indices	*P* value
S1	Between groups	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	Between sites within groups	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	Within sites	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	Total	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
S2	Between groups	3	26.9	0.53	16.6	F_{CT}: 0.17	0.2493 (± 0.0132)
	Between sites within groups	5	33.2	1.00	31.5	F_{SC}: 0.37	0.0059 (± 0.0022)
	Within sites	31	52.2	1.68	52.3	F_{ST}: 0.48	0.0010 (± 0.0010)
	Total	39	112.2	3.22			
S3	Between groups	3	10.8	0.25	16.5	F_{CT}: 0.16	0.0616 (± 0.0072)
	Between sites within groups	5	10.2	0.19	11.9	F_{SC}: 0.14	0.1241 (± 0.0117)
	Within sites	31	34. 7	1.12	71.5	F_{ST}: 0.28	0.0039 (± 0.0018)
	Total	39	55.6	1.56			
S4	Between groups	3	18.4	0.56	26.3	F_{CT}: 0.27	0.0371 (± 0.0057)
	Between sites within groups	5	11.9	0.21	10.3	F_{SC}: 0.14	0.2014 (± 0.0125)
	Within sites	31	41.2	1.33	63.4	F_{ST}: 0.37	0.0078 (± 0.0034)
	Total	39	71.6	2.11			
S6[a]	Between groups	3	50.0	1.99	81.3	F_{CT}: 0.81	0.0371 (± 0.0048)
	Between sites within groups	5	2.7	0.02	0.8	F_{SC}: 0.04	0.0743 (± 0.0069)
	Within sites	33	14. 7	0.44	17.9	F_{ST}: 0.82	$< 5{\cdot}10^{-6}$ ($\pm < 5{\cdot}10^{-6}$)
	Total	41	67.1	2.44			
S9[a]	Between groups	3	84.0	3.77	82.9	F_{CT}: 0.83	0.0987 (± 0.0110)
	Between sites within groups	5	0.5	-0.20	-4.5	F_{SC}: -0.26	0.7077 (± 0.0157)
	Within sites	29	28.5	0.98	21.6	F_{ST}: 0.78	$< 5{\cdot}10^{-6}$ ($\pm < 5{\cdot}10^{-6}$)
	Total	37	113.0	4.55			
S11	Between groups	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	Between sites within groups	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	Within sites	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	Total	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
S12	Between groups	3	59.8	1.37	16.8	F_{CT}: 0.17	0.0753 (± 0.0081)
	Between sites within groups	5	60.2	1.34	16.5	F_{SC}: 0.20	0.0283 (± 0.0059)
	Within sites	31	168.2	5.42	66.7	F_{ST}: 0.33	$< 5{\cdot}10^{-6}$ ($\pm < 5{\cdot}10^{-6}$)
	Total	39	288.3	8.14			
S13	Between groups	3	50.6	0.60	7.3	F_{CT}: 0.07	0.2698 (± 0.0169)
	Between sites within groups	5	81.9	2.36	28.6	F_{SC}: 0.31	0.0039 (± 0.0018)
	Within sites	29	153.2	5.28	64.0	F_{ST}: 0.36	$< 5{\cdot}10^{-6}$ ($\pm < 5{\cdot}10^{-6}$)
	Total	37	285.7	8.25			

[a]excluding *A. halleri* ssp. *gemmifera*. n.a.: not applicable (data for S1 and S11 are missing for the individual from Auby; see Table S1)

Table S6. Sequences and reaction efficiencies of primer pairs used for quantitative PCR determination of genomic copy number.

Amplicon		Species			Primer efficiency[†]
		A. halleri	*A. lyrata* ssp. *lyrata*	*A. thaliana* (Col-0)	
HMA4	Forward	GCTGCAGCGATGAAAAACAAAC	GCTGCAACGATGAAAAGCAAAC	GCTGCAGCGATGAAAAACAAAC	1.91 ± 0.01
	Reverse	TCCATACAACATCCCGAGGAAC	TCCATACAACATCCCGAGGAAC	TTCACACAACATCCCGAGGAGC	*n* = 290
FRD3 5'	Forward	TGGCAGAGGAAGACACGATG	TGGCAGAGGAAGACACGATG	TGGCAGAGGAAGACACGATG	1.91 ± 0.01
	Reverse	TGGCTTTGTTCGCTTCTTCTTT	TGGCTTTGTTCGCTTCTTCTTT	TGGCTTTGTTTGCTTCTTCTTT	*n* = 276
FRD3 3'	Forward	TTTATAGCAGCCACGCAGCC	TTTATAGCAGCCACGCAACC	TTTATAGCAGCAACGCAGCC	1.89 ± 0.01
	Reverse	TCCATCCAATACAAAGGCGAGT	TCCATCCAATACAAAGGCGAGT	TCCATCCAATACAAAGGCGAGA	*n* = 294
S13	Forward	ACTTCTGGGTAGTTGGAATTTTCC	ACTTCTGGGTAGTTGGAATTTTCC	ACTTCTGGGTAGTTGGAATTTTCC	1.89 ± 0.01
	Reverse	CGTTTCGGAGAATCCTCGC	CGTTTCGGAGAATCCTCGC	CGTTTCGGAGAATCCTCGC	*n* = 276

All primer sequences are given in 5'- to 3'-direction. The *HMA4* and S13 primers for *A. halleri* were designed based on an alignment of all S5/S7/S10 and S13 consensus sequences (see Figure 1, Table S1), respectively, in order to ensure the absence of polymorphisms within primer binding sites. The *FRD3* primers for *A. halleri* were designed based on available sequences from the Lan3.1 individual of the Langelsheim population [12]. The *A. thaliana* and *A. lyrata* ssp. *lyrata* primers were designed based on available genome sequences [39,73].

[†]: arithmetic mean ± s. e. m., *n* equals the number of replicate PCR reactions per amplicon.