Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/5/5/eaav0536/DC1

Supplementary Materials for

Worldwide phylogeography and history of wheat genetic diversity

François Balfourier*, Sophie Bouchet, Sandra Robert, Romain De Oliveira, Hélène Rimbert, Jonathan Kitt, Frédéric Choulet;
International Wheat Genome Sequencing Consortium, BreedWheat Consortium, Etienne Paux*

*Corresponding author. Email: francois.balfourier@inra.fr (F.B.); etienne.paux@inra.fr (E.P.)

Published 29 May 2019, *Sci. Adv.* **5**, eaav0536 (2019)
DOI: 10.1126/sciadv.aav0536

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/5/5/eaav0536/DC1)

Data file S1 (Microsoft Excel format). List of 4506 wheat accessions and related information.
Data file S2 (.zip format). Genotyping data of 4506 wheat accession with 113,457 genome-wide SNPs.
Data file S3 (.zip format). Haplotyping data of 4403 wheat accessions with 8741 haplotypic blocks.
Members of the IWGSC Consortium (Microsoft Word format)
Members of the BreedWheat Consortium (Microsoft Word format)

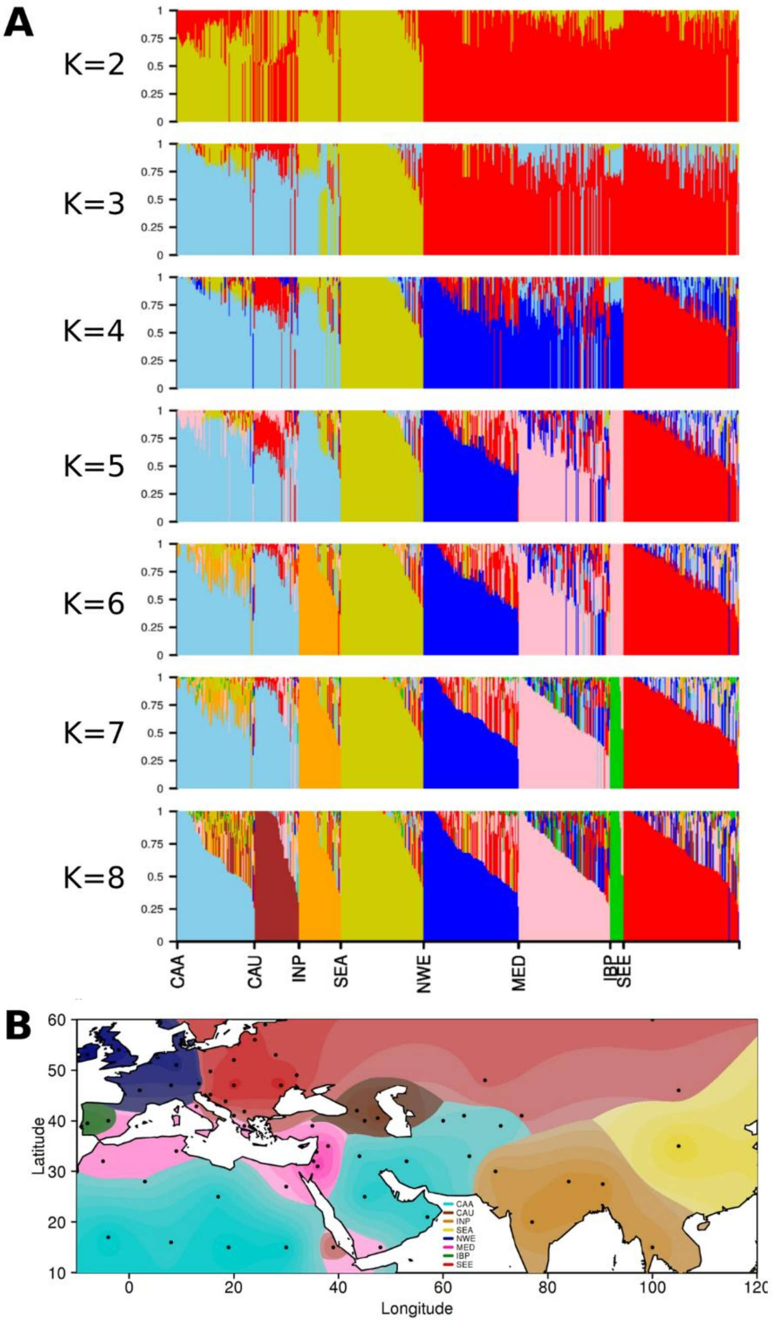

Fig. S1. Assignment of 632 landraces to genetic groups. (A) STRUCTURE outputs for K=2 to K=8. (B) Interpolated map representing the structure of wheat landraces (K=8). Cyan: CAA, Brown: CAU, Orange: INP, Yellow: SEA, Blue: NEW, Pink: MED, Green: IBP, Red: SEE.

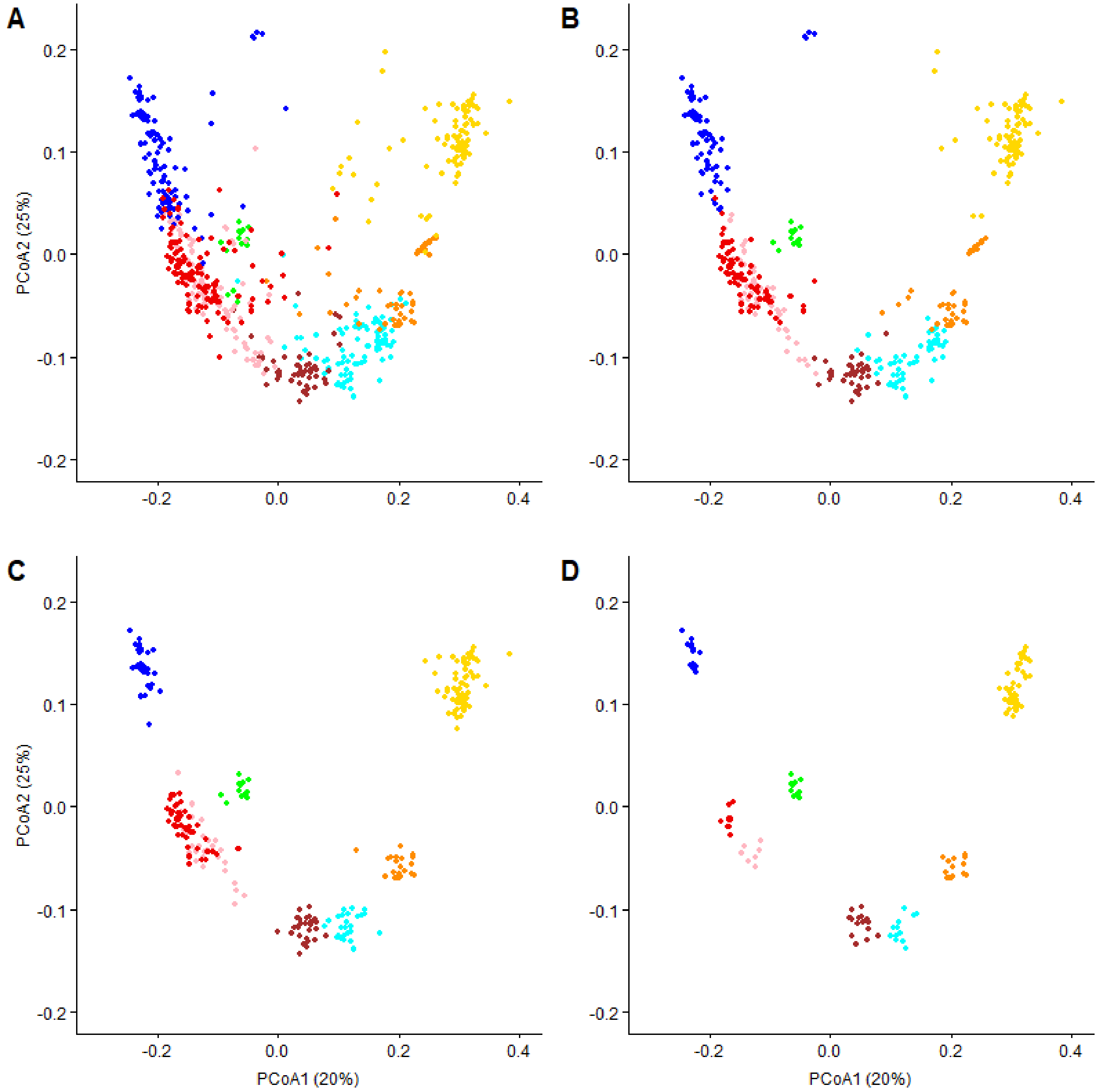

Fig. S2. PCoA calculated with 8741 haplotypes on landraces. (A) All landraces. (B) Landraces with assignment > 0.6. (C) Landraces with assignment > 0.8. (D) Landraces with assignment > 0.99. Simple Matching distance was calculated using Darwin software. Cyan: Central Asia & Africa (CAA); Brown: Caucasus (CAU); Orange: Indian Peninsula (INP); Yellow: South-East Asia (SEA); Blue: North West Europe (NWE); Pink: Mediterranean basin (MED); Green: Iberian Peninsula (IBP); Red: South East Europe (SEE).

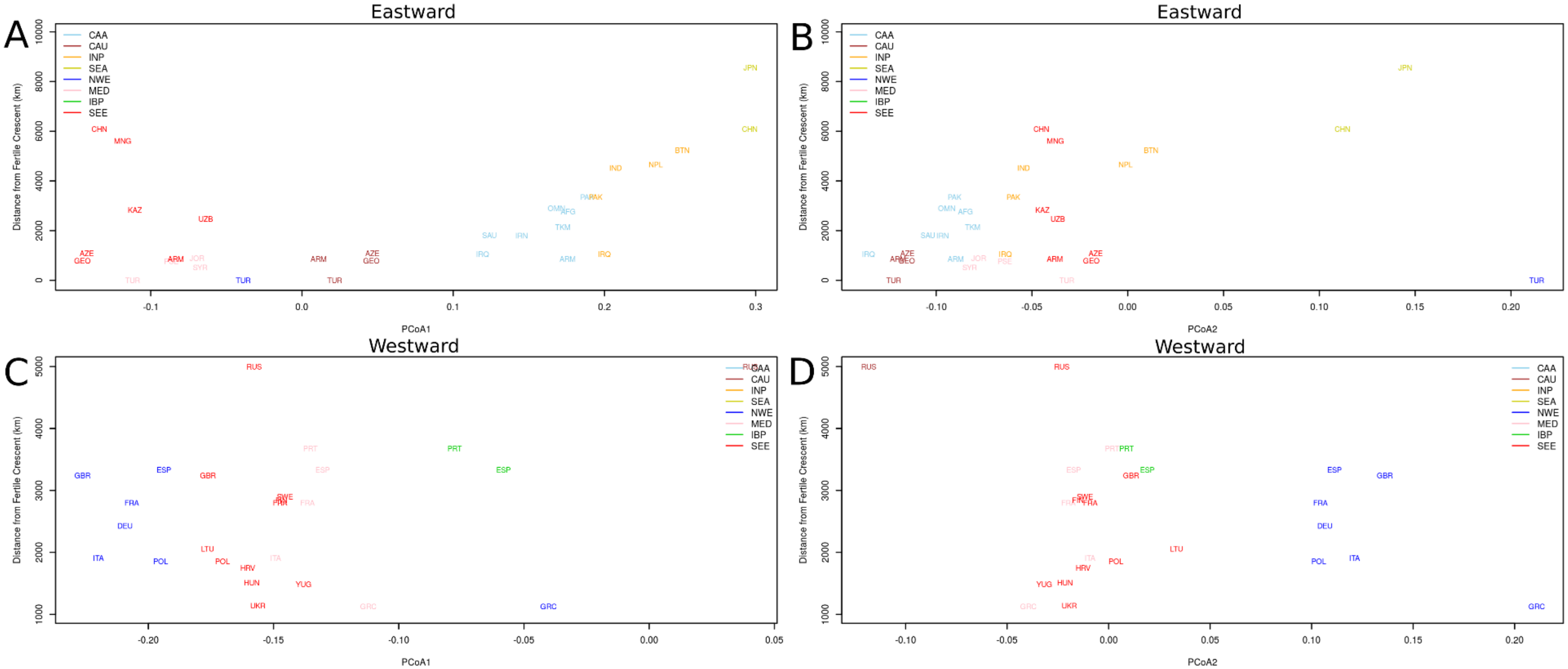

Fig. S3. Isolation by distance. (A) Relations between geographical distance from the Fertile Crescent eastward and PCoA1; (B) Relations between geographical distance from the Fertile Crescent eastward and PCoA2; (C) Relations between geographical distance from the Fertile Crescent westward and PCoA1 (PCoA); (D) Relations between geographical distance from the Fertile Crescent eastward and PCoA2.

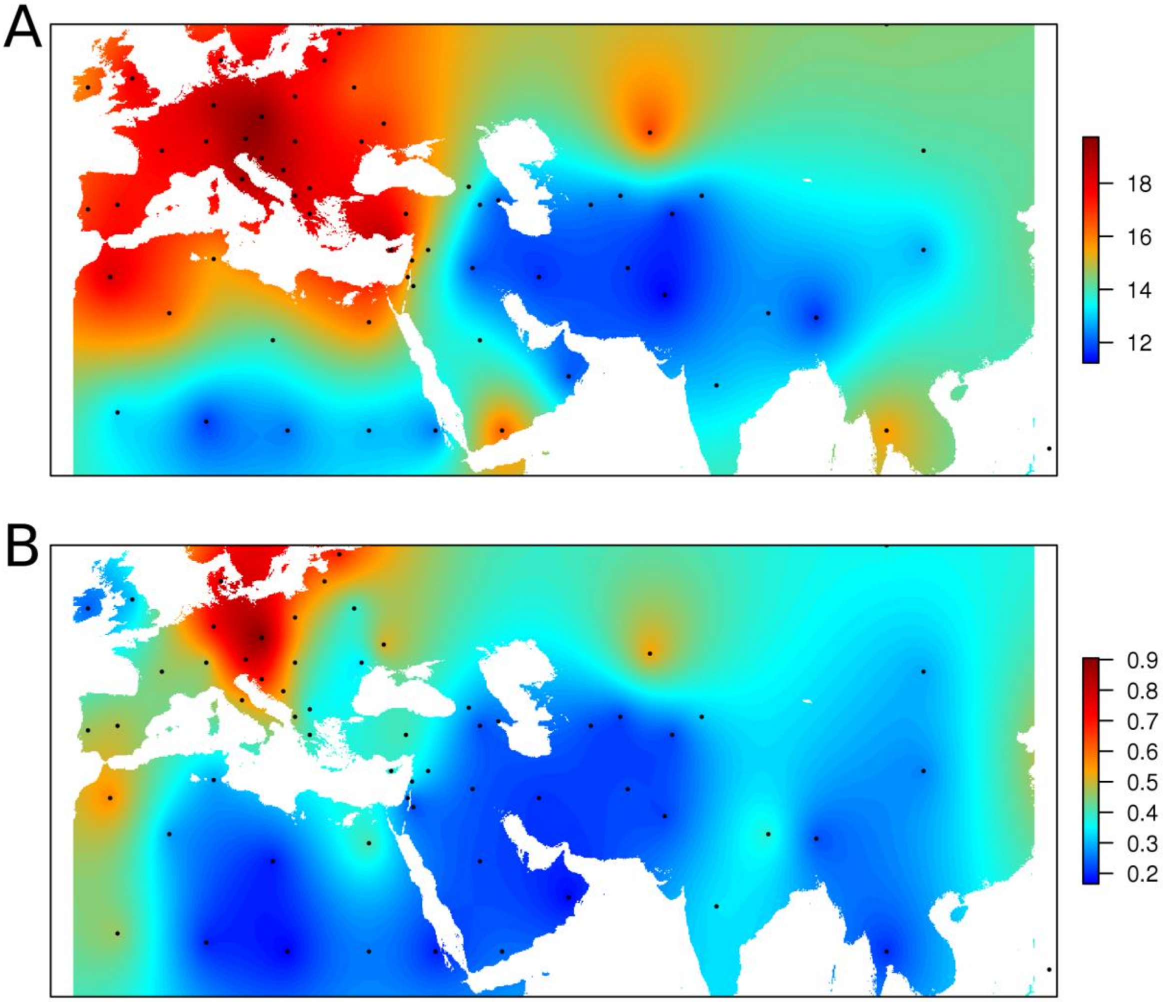

Fig. S4. Geographical projection of the frequency of rare alleles and singletons in landraces. (A) Rare alleles; (B) singletons.

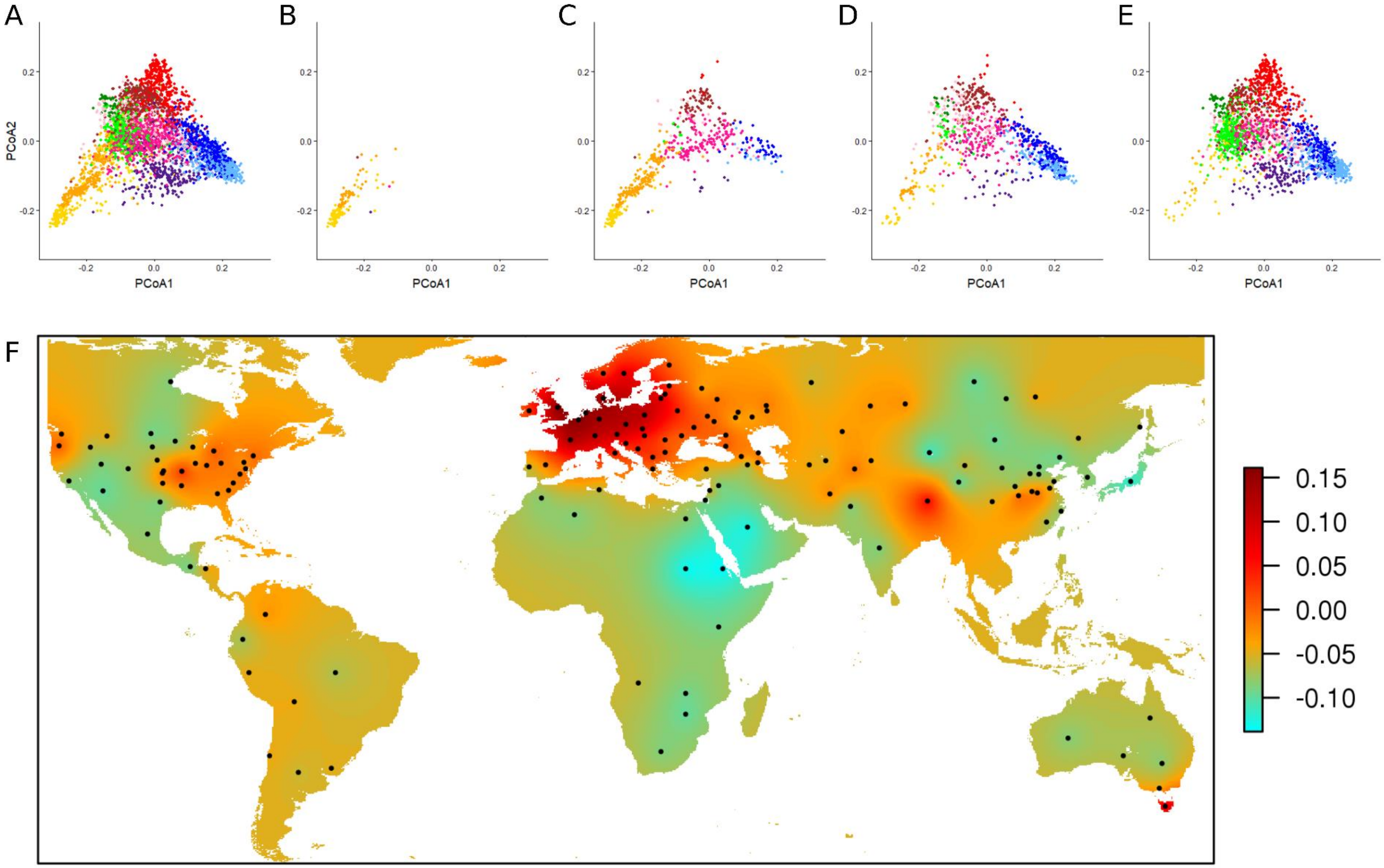

Fig. S5. Temporal evolution of worldwide genetic diversity. (**A**) PCoA calculated with 8741 haplotypes on 4403 accessions. (**B**) PCoA calculated with 8741 haplotypes on 139 SEA and INP landraces. (**C**) PCoA calculated with 8741 haplotypes on 632 landraces. (**D**) PCoA calculated with 8741 haplotypes on 947 traditional cultivars. (**E**) PCoA calculated with 8741 haplotypes on 2210 modern varieties. (**F**) Geographical projection of the first axis of the PCoA for modern 2210 varieties. The different colors correspond to the eleven groups.

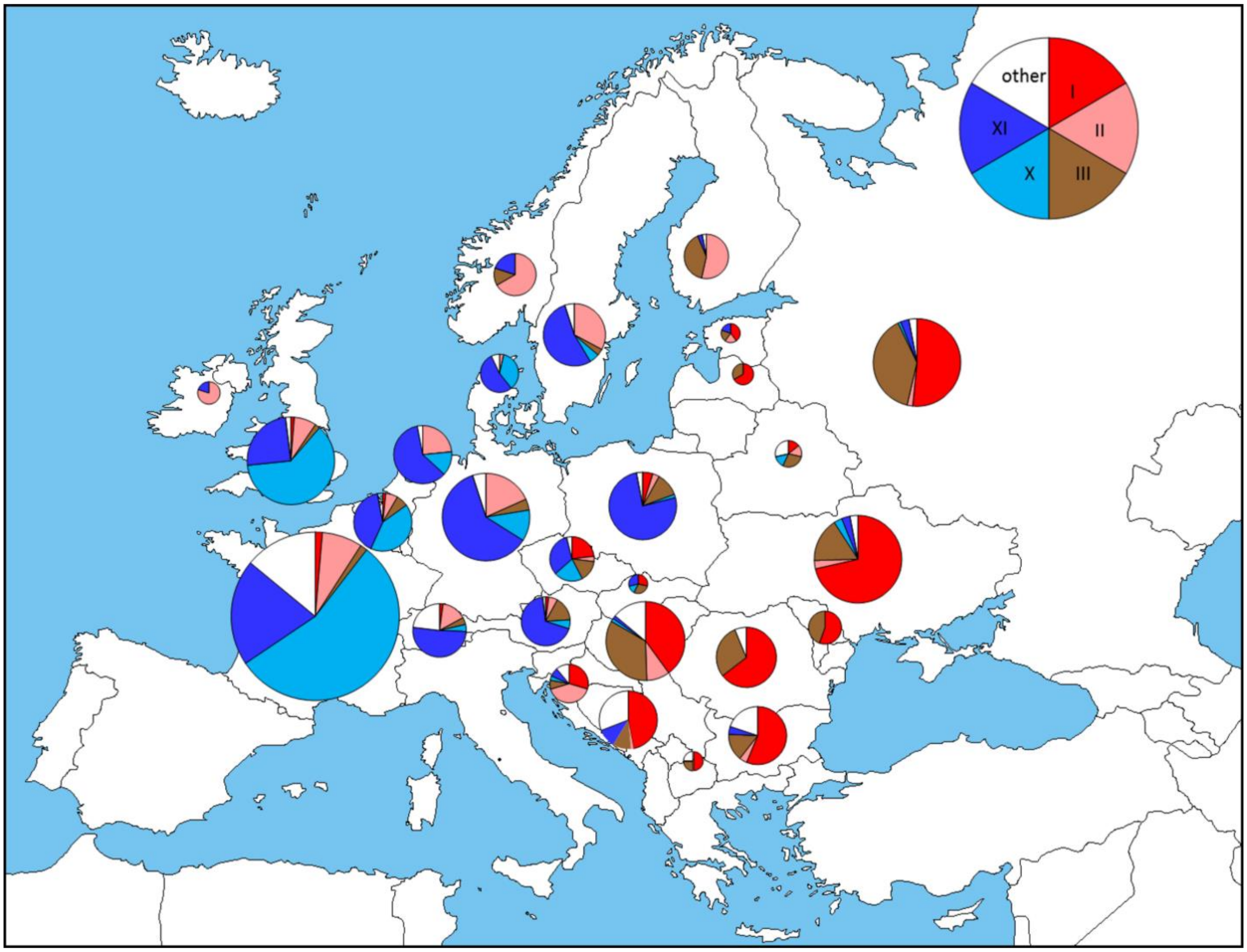

Fig. S6. Frequency of major haplotype groups in Europe. The size of a pie is proportional to the number of accessions in each country.

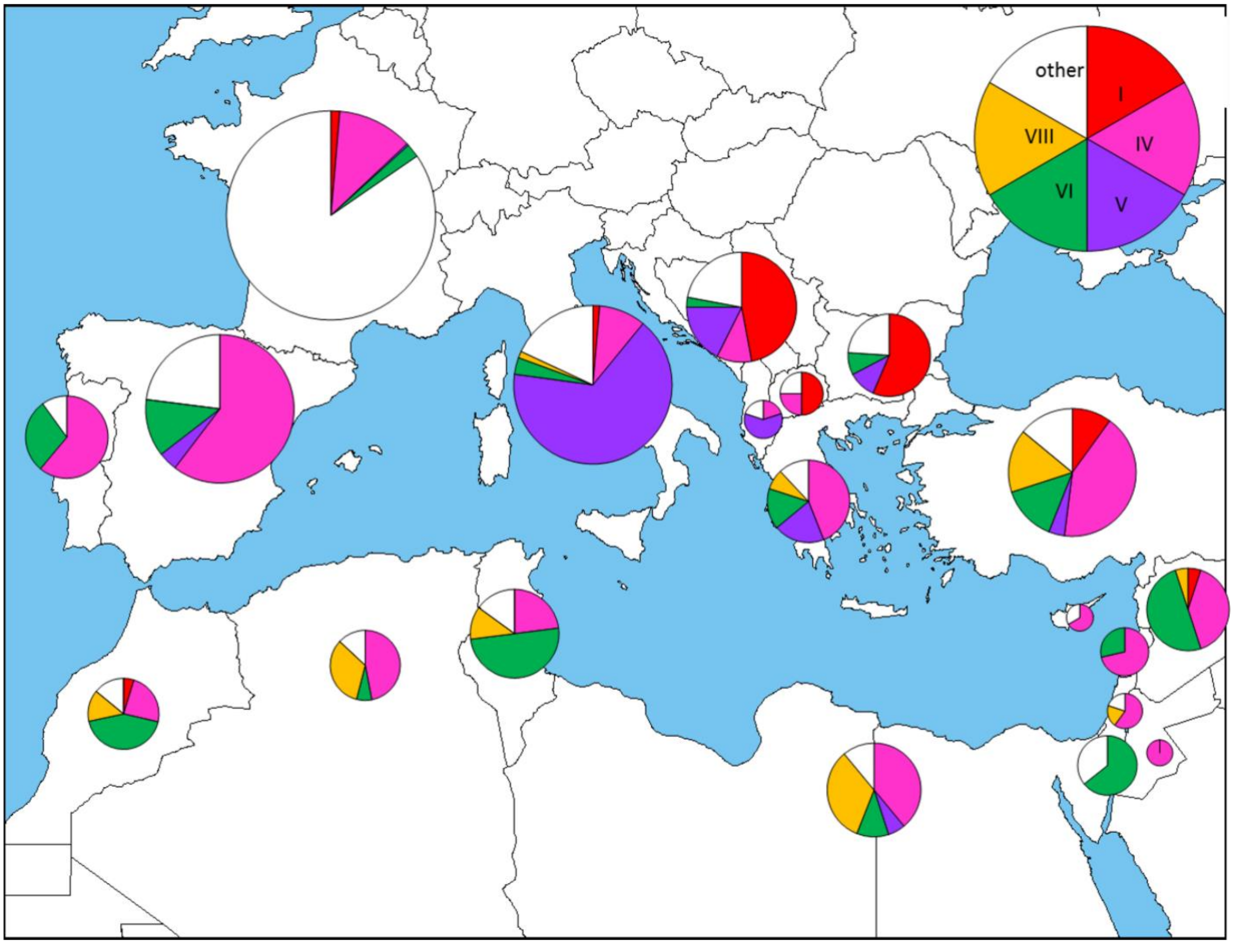

Fig. S7. Frequency of major haplotype groups in the MED. The size of a pie is proportional to the number of accessions in each country.

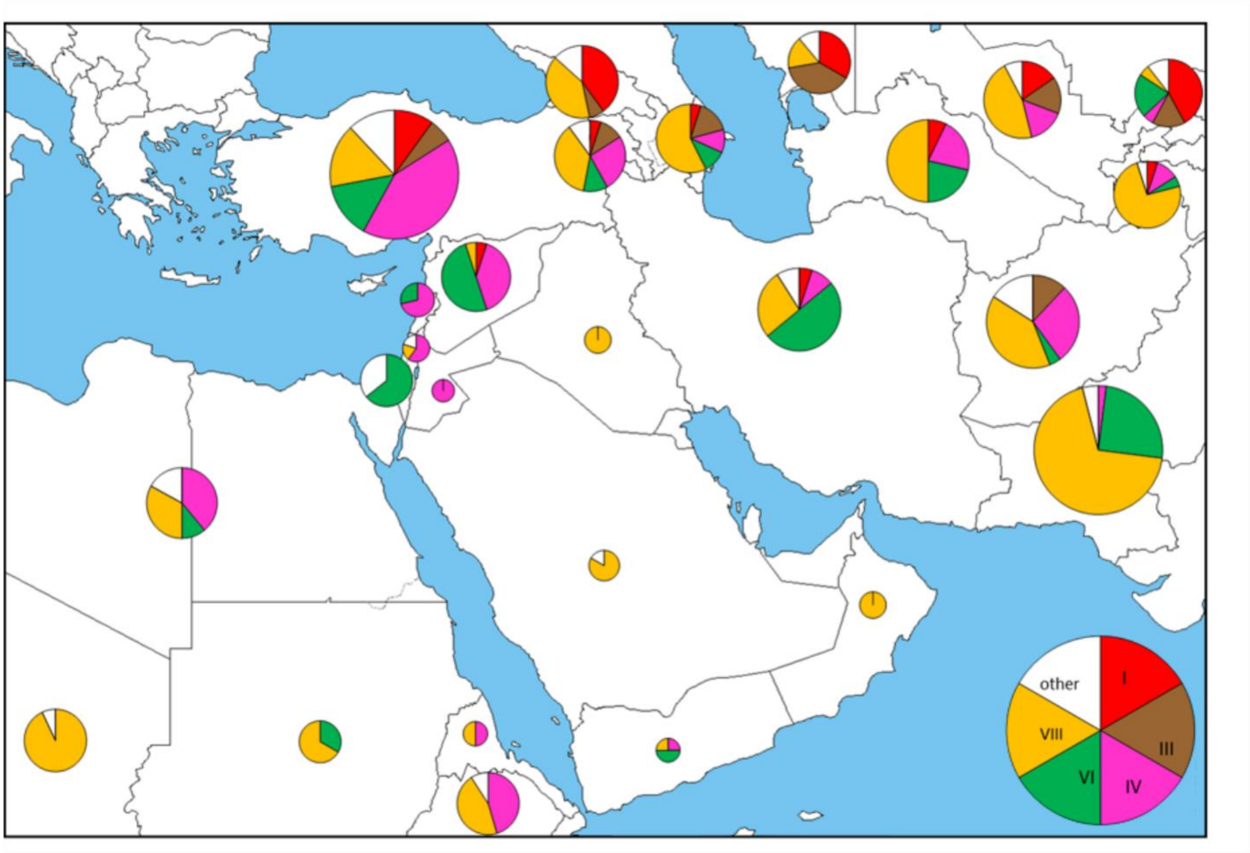

Fig. S8. Frequency of major haplotype groups in Middle East and Central Asia. The size of a pie is proportional to the number of accessions in each country.

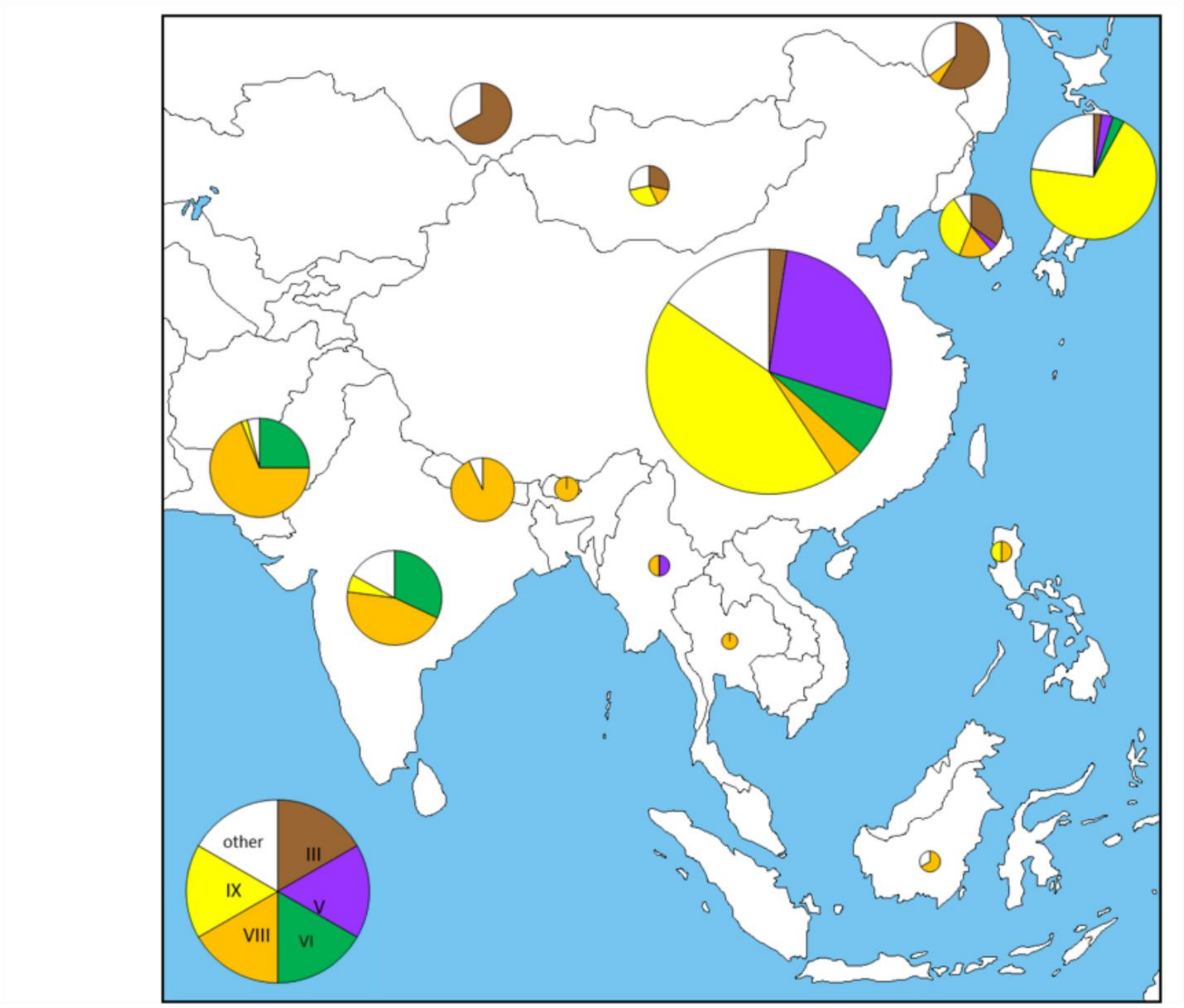

Fig. S9. Frequency of major haplotype groups in South Eastern Asia. The size of a pie is proportional to the number of accessions in each country.

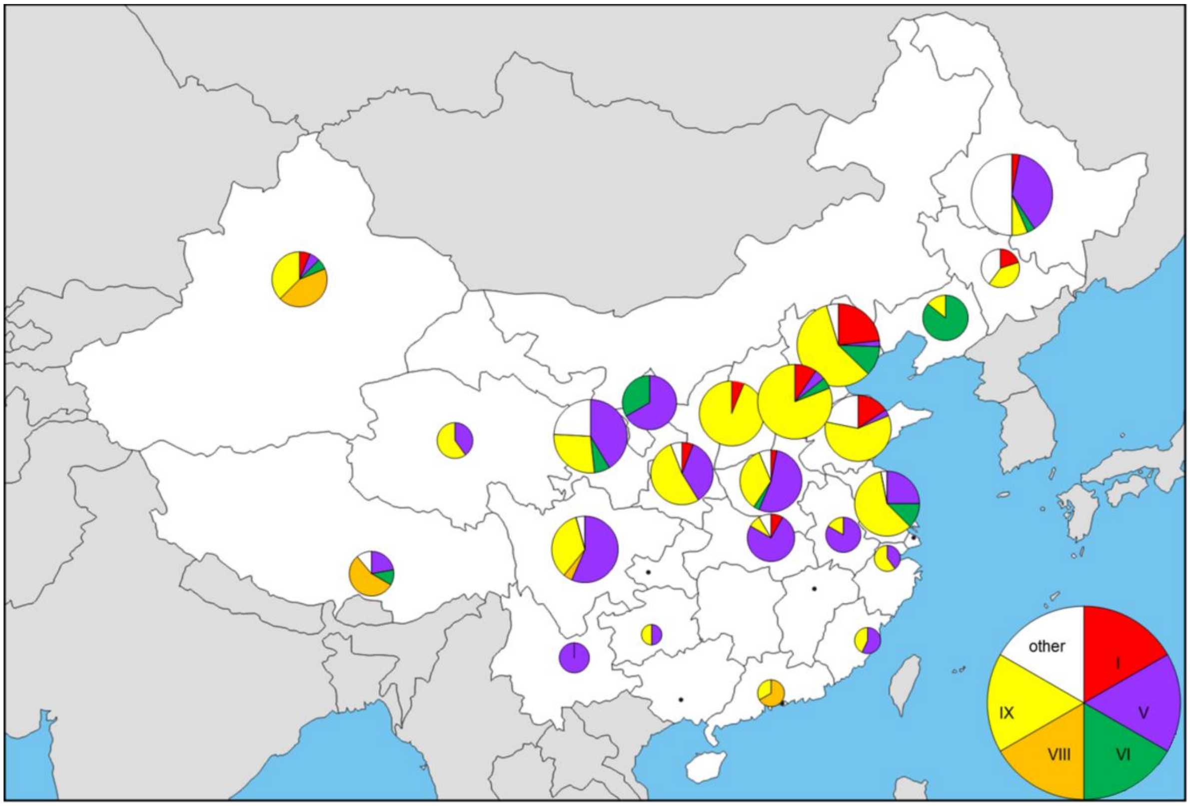

Fig. S10. Frequency of major haplotype groups in China. The size of a pie is proportional to the number of accessions in each country.

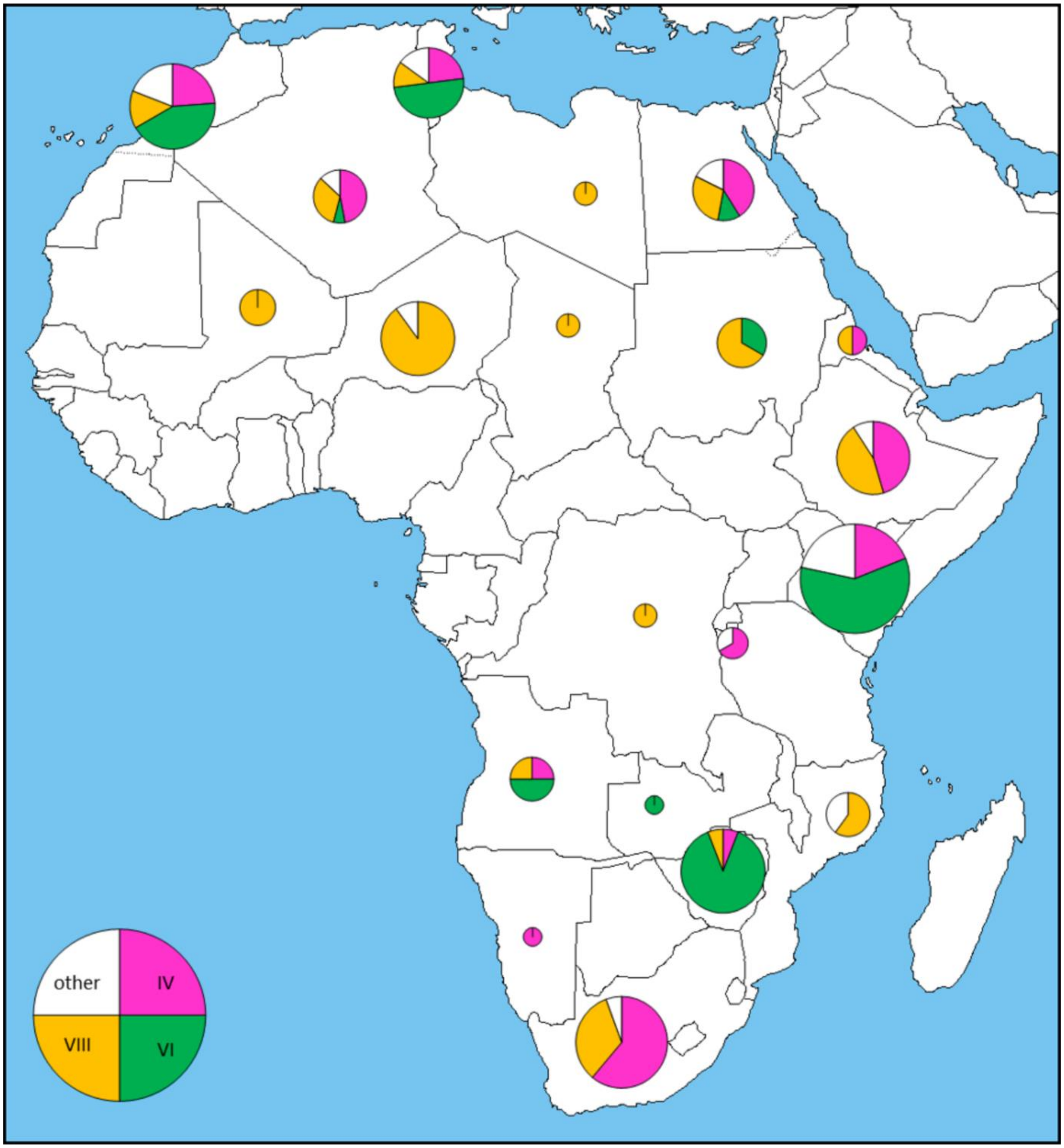

Fig. S11. Frequency of major haplotype groups in Africa. The size of a pie is proportional to the number of accessions in each country.

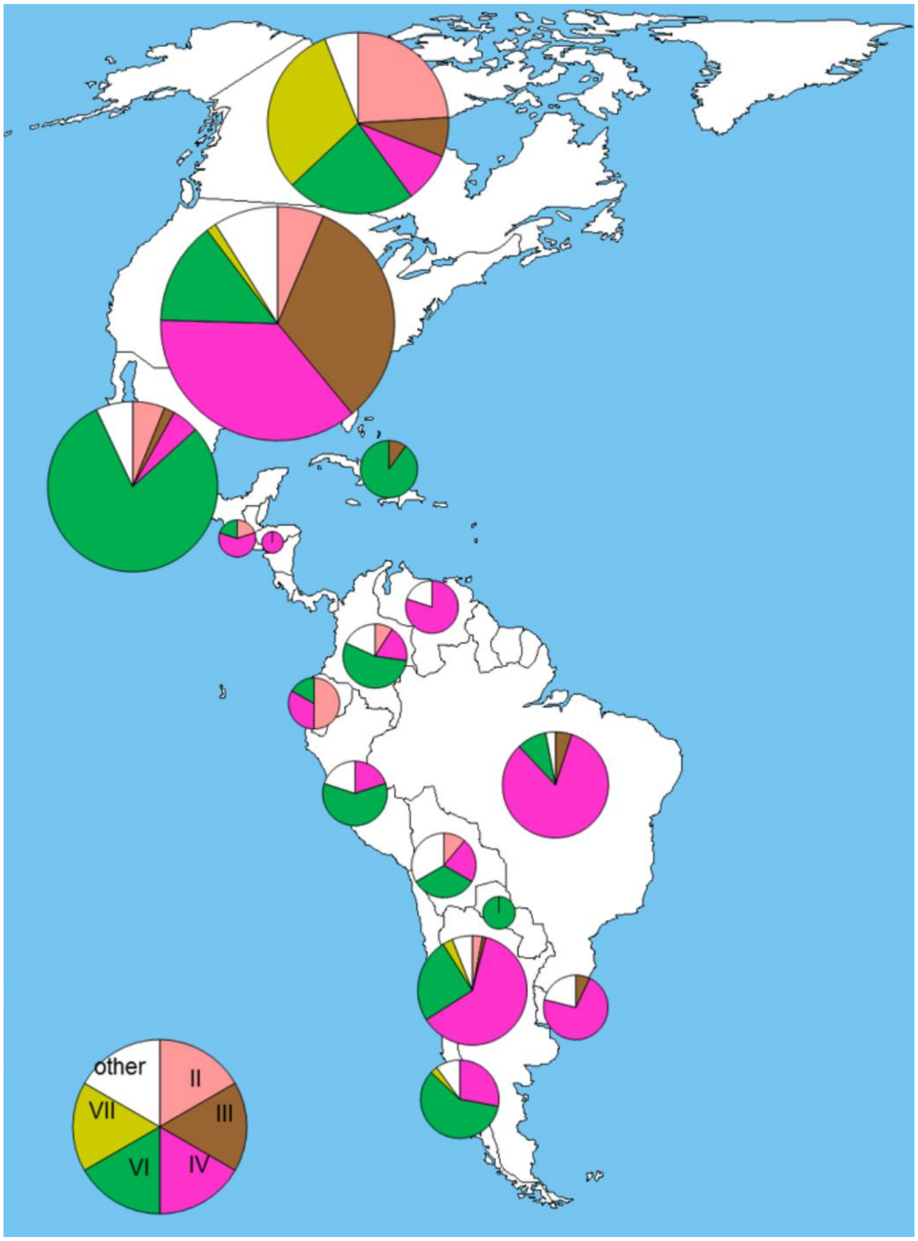

Fig. S12. Frequency of major haplotype groups in America. The size of a pie is proportional to the number of accessions in each country.

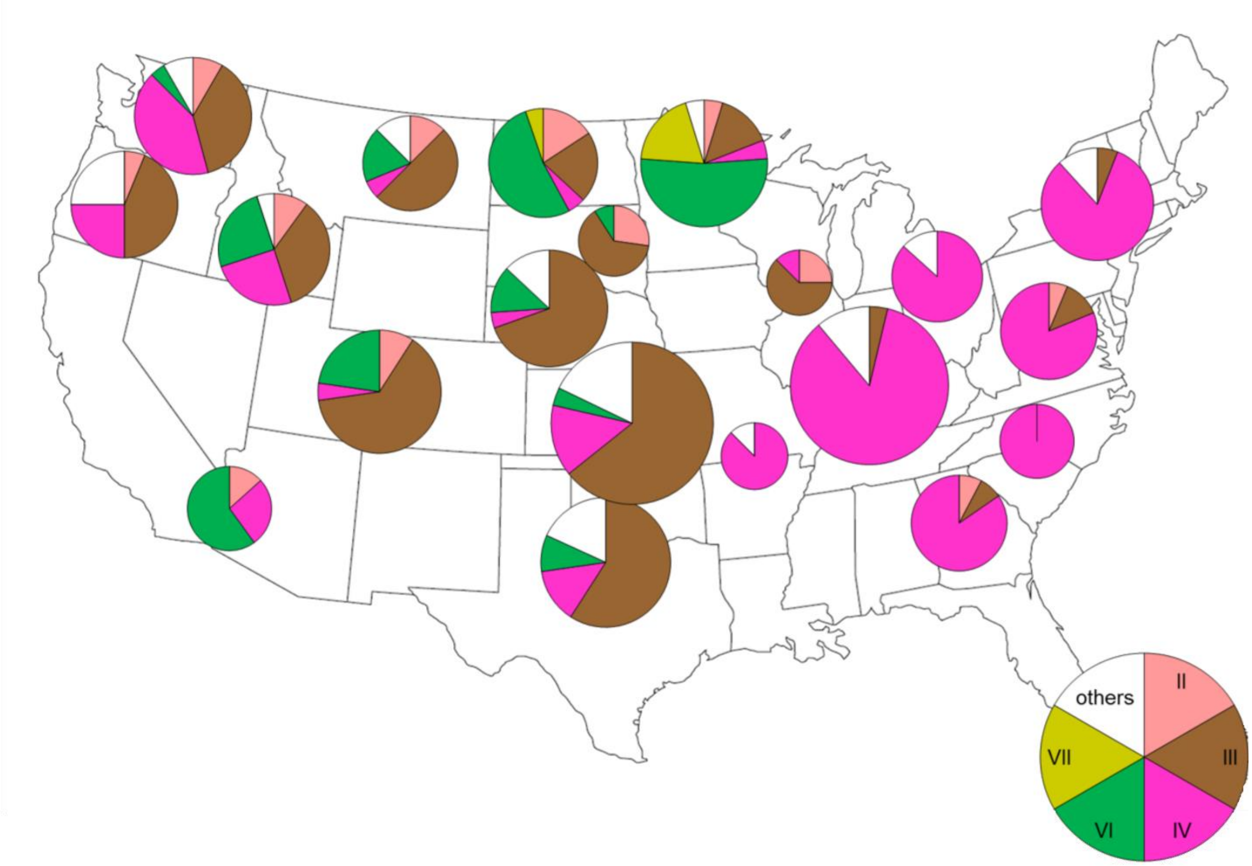

Fig. S13. Frequency of major haplotype groups in the United States. The size of a pie is proportional to the number of accessions in each country.

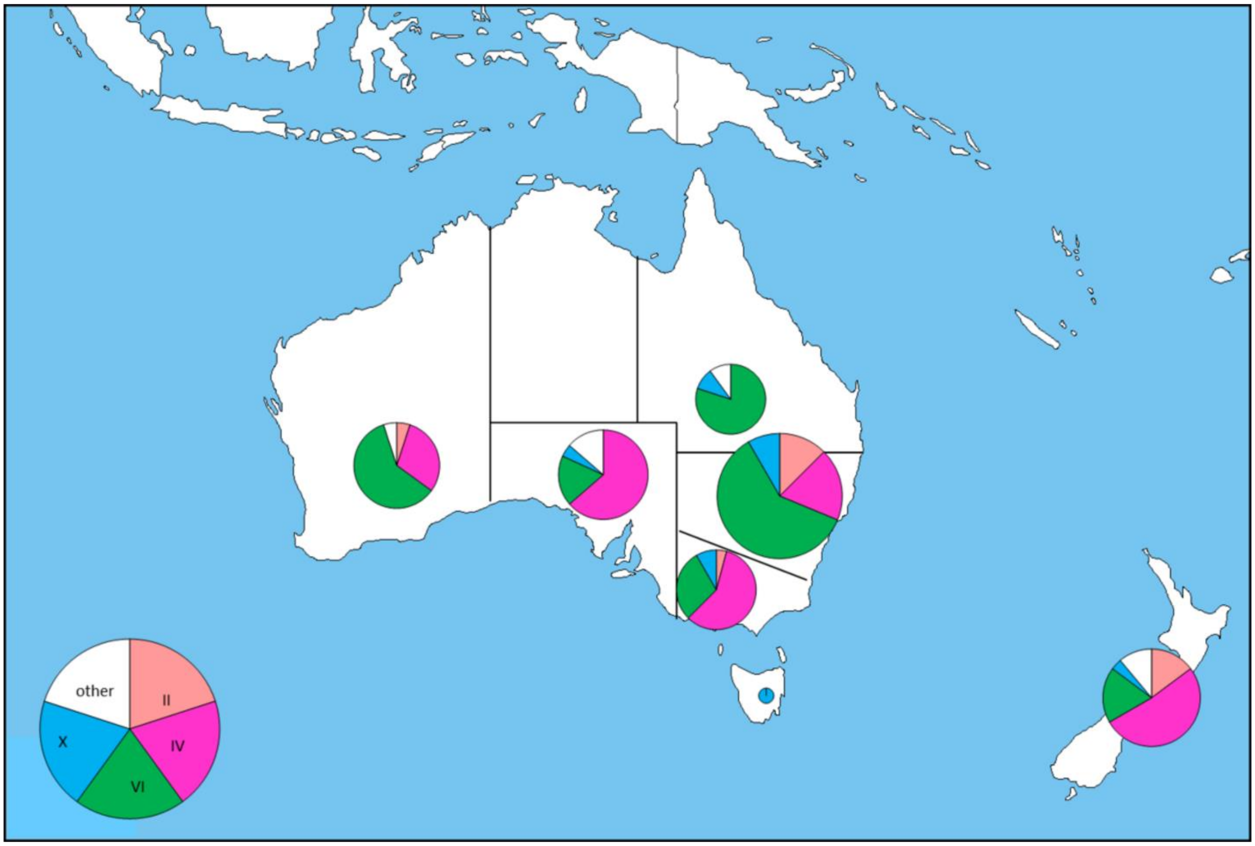

Fig. S14. Frequency of major haplotype groups in Oceania. The size of a pie is proportional to the number of accessions in each country.

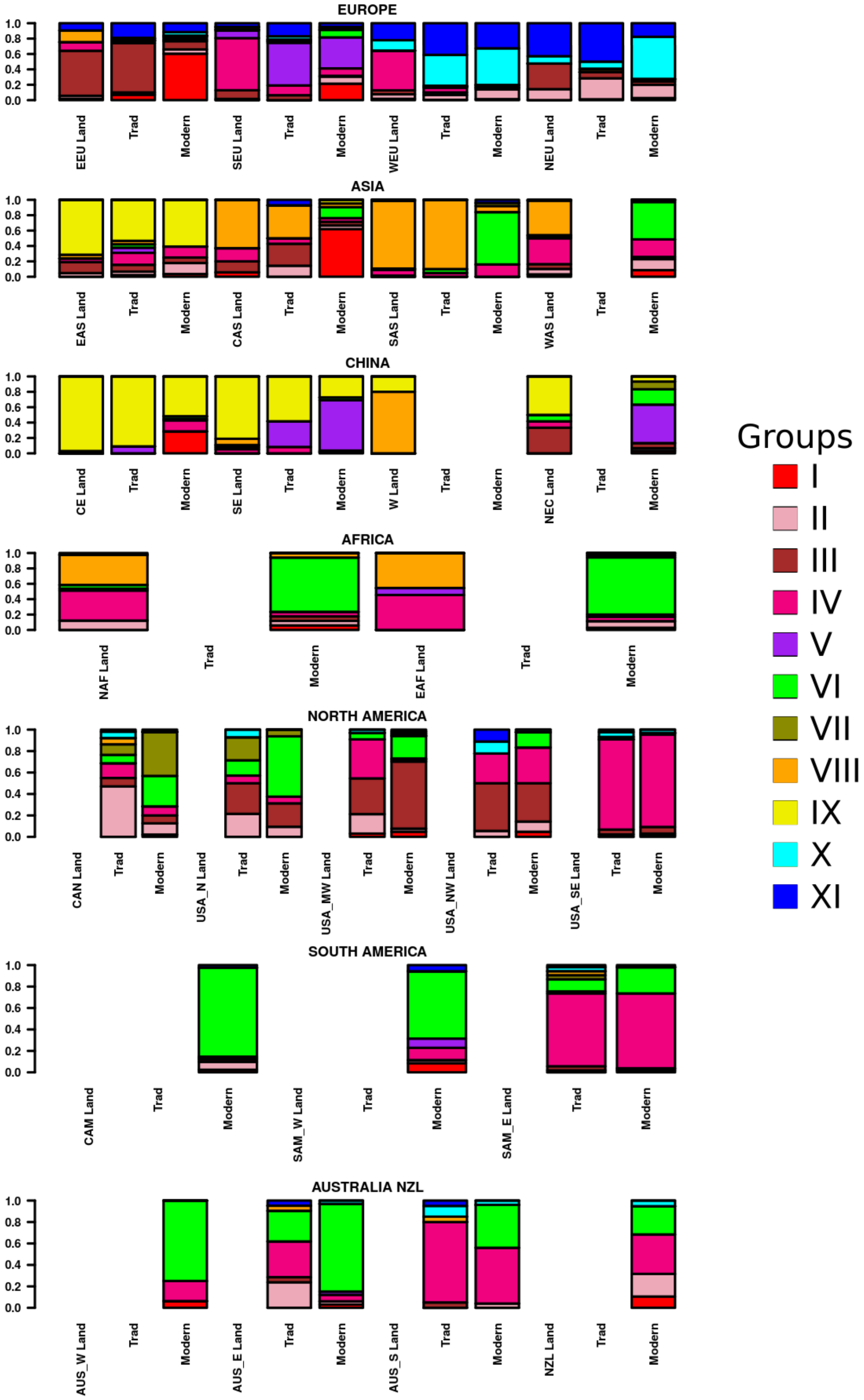
EUROPE
1.0
0.8
0.6
0.4
0.2
0.0
EEU Land
Trad
Modern
SEU Land
Trad
Modern
WEU Land
Trad
Modern
NEU Land
Trad
Modern
ASIA
EAS Land
Trad
Modern
CAS Land
Trad
Modern
SAS Land
Trad
Modern
WAS Land
Trad
Modern
CHINA
CE Land
Trad
Modern
SE Land
Trad
Modern
W Land
Trad
Modern
NEC Land
Trad
Modern
AFRICA
NAF Land
Trad
Modern
EAF Land
Trad
Modern
NORTH AMERICA
CAN Land
Trad
Modern
USA_N Land
Trad
Modern
USA_MW Land
Trad
Modern
USA_NW Land
Trad
Modern
USA_SE Land
Trad
Modern
SOUTH AMERICA
CAM Land
Trad
Modern
SAM_W Land
Trad
Modern
SAM_E Land
Trad
Modern
AUSTRALIA NZL
AUS_W Land
Trad
Modern
AUS_E Land
Trad
Modern
AUS_S Land
Trad
Modern
NZL Land
Trad
Modern
Groups
I
II
III
IV
V
VI
VII
VIII
IX
X
XI

Fig. S15. Proportion of the 11 groups among the landraces, traditional cultivars, and modern varieties from different regions of the world.

EUROPE - EEU: Eastern Europe; NEU: Northern Europe; SEU: Southern Europe; WEU: Western Europe
ASIA - CAS: Central Asia; EAS: Eastern Asia; SAS: Southern Asia; WAS: Western Asia
CHINA - SE: South-Eastern; W: Western; NEC: North-Eastern and Central
AFRICA - NAF: Northern Africa; EAF: Eastern Africa
NORTH AMERICA - CAN: Canada; USA_N: Northern USA; USA_MW: Mid-Western USA; USA_SE: Southern and Eastern USA, USA_NW: North-Western USA
SOUTH AMERICA - CAM: Central America, SAM_E: Eastern Coast of Southern America; SAM_W: Western Coast of Southern America
ASUTRALIA_NZL - AUS_W: Western Coast Australia; AUS_E: Eastern Coast Australia; AUS_S: Southern Coast Australia; NZL: New-Zealand

Table S1. Description of the eight subpopulations of landraces.

Group[a]	N[b]	Assign[c]	Countries[d]	Regions[e]	Winter[f]	He[g]	Poly Rate[h]	N rare[i]	N sing[j]
CAA	87 (54)	68 (65)	OMN (9),SDN (7),SAU (7),PAK (7),IRN (7)	WAS (30),SAS (22),NAF (20),WAF (11),CAS (9)	14	0.41	0.80	9.33	0.2
CAU	50 (38)	80 (95)	AZE (24),RUS (19),TUR (16),GEO (14),ARM (11)	WAS (68),EEU (22),SEU (3),SEA (3),EAS (3)	35	0.40	0.75	8.48	0.3
INP	47 (36)	77 (73)	IND (28),PAK (25),NPL (22),IRQ (6),EGY (6)	SAS (83),WAS (6),SEA (6),NAF (6)	0	0.35	0.71	8.51	0.32
SEA	93 (80)	88 (99)	CHN (85),JPN (10)	EAS (98)	53	0.34	0.77	6.75	0.17
NWE	107 (68)	68 (66)	FRA (46),DEU (13),GBR (12),ESP (4),TUR (3)	WEU (60),NEU (15),SEU (12),EEU (6),WAS (4)	84	0.45	0.88	20.41	0.5
MED	103 (59)	66 (67)	SYR (12),FRA (10),ESP (10),ITA (7),TUR (5)	WAS (29),SEU (27),NAF (17),WEU (10),SAS (3)	15	0.47	0.90	23.09	1.9
IBP	15 (12)	89 (100)	ESP (75),PRT (25)	SEU (100)	25	0.30	0.51	19.6	0.53
SEE	130 (84)	71 (71)	FRA (14),RUS (12),HUN (11),UKR (5),CHN (5)	EEU (35),WEU (14),WAS (13),NEU (11),SEU (7)	51	0.44	0.92	21.31	0.77

a Groups of landraces as defined by STRUCTURE algorithm. CAA: Central Asia – Africa; CAU: Caucasus; IBP: Iberian Peninsula; INP: Indian Peninsula; NWE: North Western Europe; MED: Mediterranean region; SEA: South Eastern Asia; SEE: South Eastern Europe

b Number of accessions with highest genome proportion (with at least 60% of its genome) assigned to this group.

c Mean (median) genome proportion assigned to this group.

d Percentage of accessions originating from each country (International Organization for Standardization code). ARM: Armenia; AZE: Azerbaijan; CHN: China; DEU: Germany; EGY: Egypt; ESP: Spain; FRA: France; GBR: United Kingdom of Great Britain and Northern Ireland; GEO: Georgia; HUN: Hungary; IND: India; IRN: Islamic Republic of Iran; IRQ: Iraq; ITA: Italy; JPN: Japan; NPL: Nepal; OMN:Oman; PAK:Pakistan; PRT: Portugal; RUS: Russian Federation; SAU: Saudi Arabia; SDN: Sudan; SYR: Syrian Arab Republic; TUR: Turkey; UKR: Ukraine

e Percentage of accessions originating from each region (ISO code). CAS: Central Asia; EAS: Eastern Asia; EEU: Eastern Europe; NAF: Northern Africa; NEU: Northern Europe; SAS: Southern Asia; SEA: South-Eastern Asia; SEU: Southern Europe; WAF: Western Africa; WAS: Western Asia; WEU: Western Europe

f Percentage of winter type accessions. Most of other accessions are spring type, a few are alternative type.

g	Nei He diversity index
h	Percentage of polymorphic alleles
i	Average number of rare allele (frequency < 0.01) per line
j	Average number of singleton per line

Table S2. Pairwise comparisons of five subpopulations of landraces westward of the Fertile Crescent.

	MED	SEE	NWE	CAU	CAA
MED	615	271	166	161	242
SEE	271	553	197	209	112
NWE	166	197	322	43	53
CAU	161	209	43	306	235
CAA	242	112	53	235	430

Number of rare alleles (frequency<0.05) specific to one genetic group on the diagonal (in blue)
Number of rare alleles shared between populations below and above the diagonal.

Table S3. Pairwise comparisons of five subpopulations of landraces eastward of the Fertile Crescent.

	MED	SEA	INP	CAU	CAA
MED	927	173	53	215	244
SEA	173	505	179	163	187
INP	53	179	190	51	80
CAU	215	163	51	341	135
CAA	244	187	80	135	220

Number of rare alleles (frequency<0.05) specific to one genetic group on the diagonal (in blue)
Number of rare alleles shared between populations below and above the diagonal.

Table S4. Description of the 11 groups of accessions.

Gp[a]	Name	N[b]	Countries[c]	Regions[d]	Landraces[e]	Registration[f]	Modern[g]	Winter[h]	He[i]	Poly[j]
I	SEE_modern	413	UKR (17),HUN (14),RUS (13),YUG (8),ROU (8)	EEU (64),SEU (11),EAS (7),CAS (4),WAS (3)	SEE (7)	1984 (1984)	95	93	0.39	0.95
II	NWE_NAM	353	FRA (14),CAN (10),DEU (8),USA (7),SWE (5)	WEU (30),NEU (17),NAM (16),EEU (9),SEU (7)	SEE (19)	1962 (1970)	66	20	0.47	0.97
III	NAM_SEE	419	USA (28),RUS (14),HUN (12),UKR (4),ROU (3)	EEU (40),NAM (30),WEU (7),NEU (6),EAS (5)	SEE (69)	1960 (1965)	57	86	0.43	0.97
IV	NW_MED	674	USA (20),FRA (11),ARG (6),AUS (6),ESP (4)	NAM (22),SAM (15),WEU (14),SEU (12),AUS (8)	IBP (100),MED (100),NWE (9),CAU (8),SEE (5)	1956 (1962)	54	39	0.51	0.99
V	CHN_IT_modern	265	CHN (40),ITA (37),YUG (5),HUN (4),GRC (2)	SEU (45),EAS (42),EEU (6)	NWE (10)	1967 (1968)	81	39	0.45	0.94
VI	CIMMYT_NW_CHN_modern	528	MEX (15),AUS (11),USA (10),CAN (6),CHN (5)	NAM (16),CAM (15),AUS (12),SAM (11),SAS (8)	-	1976 (1977)	90	7	0.47	0.97
VII	CAN_CHN_USA_modern	63	CAN (70),CHN (11),USA (8),ARG (3)	NAM (78),EAS (11),SAM (5)	-	1971 (1974)	77	0	0.24	0.72
VIII	SEA_INP	292	PAK (11),NPL (9),IND (8),CHN (5),TJK (5)	SAS (35),WAS (17),CAS (11),NAF (7),EAS (7)	INP (100),CAA (100),CAU (92)	1938 (1934)	16	23	0.44	0.91
IX	CHN_JPN	255	CHN (66),JPN (23),KOR (3)	EAS (93)	SEA (98)	1961 (1965)	63	63	0.45	0.93
X	FRA_GBR_modern	581	FRA (62),GBR (17),BEL (4),DEU (3)	WEU (72),NEU (19),EEU (4)	NWE (24)	1969 (1974)	73	90	0.40	0.94
XI	FRA_DEU	560	FRA (24),DEU (16),POL (8),GBR (7),NLD (6)	WEU (61),NEU (16),EEU (14),SEU (4)	NWE (56)	1957 (1964)	57	96	0.42	0.94

a Haplotypic Groups, from I to XI

b Number of accessions

c Percentage of accessions originating from each country (ISO code). ARG: Argentina; AUS: Australia; BEL: Belgium; CAN: Canada; CHN: China; DEU: Germany; ESP: Spain; FRA: France; GBR: United Kingdom of Great Britain and Northern Ireland; GRC: Greece; HUN: Hungary; IND: India; ITA: Italy; JPN: Japan; KOR: Republic of Korea; MEX: Mexico; NLD: Netherlands; NPL: Nepal; PAK: Pakistan; POL: Poland; ROU: Romania; RUS: Russian Federation; SWE: Sweden; TJK: Tajikistan; UKR: Ukraine; USA: United States of America; YUG: Yugoslavia d Percentage of accessions originating from each region (ISO code). AUS: Australia and NewZealand; CAM: Central America; CAS: Central Asia; EAS: Eastern Asia; EEU: Eastern Europe; NAF: Northern Africa; NAM: Northern America; NEU: Northern Europe; SAM: South America; SAS: Southern Asia; SEU: Southern Europe; WAS: Western Asia; WEU: Western Europe

e Percentage of accessions from the eight landrace groups defined by STRUCTURE assigned to this group

f Mean (Median) registration date

g Percentage of accessions registered after 1959

h Percentage of winter type accessions. Most of other accessions are spring type, a few are alternative type

i Nei He diversity index

j Percentage of polymorphic alleles

Table S5. Main SVs (>5 Mb) detected in the wheat genome.

Chromosome	Start (Mb)	Stop (Mb)	Size (Mb)	Chromosome length (Mb)	Fraction of the chromosome
1A	0	214	214	593	36%
1A	0	28	28	593	5%
1A	1	22	21	593	4%
1A	565	593	28	593	5%
1B	0	689	689	689	100%
1B	0	236	236	689	34%
1B	0	218	218	689	32%
1B	625	689	64	689	9%
1D	0	12	12	495	2%
1D	413	418	5	495	1%
1D	413	424	11	495	2%
2A	0	10	10	780	1%
2A	13	19	6	780	1%
2A	611	780	169	780	22%
2A	712	780	68	780	9%
2A	734	780	46	780	6%
2B	90	757	667	800	83%
2B	54	733	679	800	85%
2B	90	609	519	800	65%
2B	90	154	64	800	8%
2B	455	733	278	800	35%
2B	591	768	177	800	22%
2B	670	743	73	800	9%
2B	0	30	30	800	4%
2D	0	18	18	651	3%
2D	0	46	46	651	7%
2D	0	457	457	651	70%
2D	636	651	15	651	2%
3A	0	41	41	751	5%
3A	0	19	19	751	3%
3A	8	14	6	751	1%
3A	732	751	19	751	3%
3B	7	14	7	829	1%
3B	1	33	32	829	4%
3B	1	6	5	829	1%
3B	0	123	123	829	15%
3B	819	829	10	829	1%

Table S5. Main SVs (>5 Mb) detected in the wheat genome (Cont.)

Chromosome	Start (Mb)	Stop (Mb)	Size (Mb)	Chromosome length (Mb)	Fraction of the chromosome
3D	504	615	111	615	18%
3D	600	615	15	615	2%
4A	640	744	104	744	14%
4A	725	744	19	744	3%
4A	642	655	13	744	2%
4B	640	672	32	672	5%
4B	656	672	16	672	2%
4D	0	8	8	509	2%
5A	0	31	31	710	4%
5A	0	19	19	710	3%
5A	464	475	11	710	2%
5A	534	539	5	710	1%
5B	533	545	12	713	2%
5B	486	513	27	713	4%
5B	507	525	18	713	3%
5B	588	713	125	713	18%
5D	501	564	63	565	11%
6A	306	617	311	617	50%
6A	0	50	50	617	8%
6A	1	20	19	617	3%
6A	585	617	32	617	5%
6B	0	39	39	721	5%
6B	0	9	9	721	1%
6B	704	721	17	721	2%
6D	0	7	7	473	1%
6D	0	24	24	473	5%
6D	456	473	17	473	4%
6D	462	470	8	473	2%
7A	0	184	184	736	25%
7A	0	30	30	736	4%
7A	150	162	12	736	2%
7A	592	736	144	736	20%
7A	672	736	64	736	9%
7A	699	736	37	736	5%

Table S5. Main SVs (>5 Mb) detected in the wheat genome (Cont.)

Chromosome	Start (Mb)	Stop (Mb)	Size (Mb)	Chromosome length (Mb)	Fraction of the chromosome
7B	0	73	73	751	10%
7B	34	40	6	751	1%
7B	633	751	118	751	16%
7B	668	751	83	751	11%
7B	727	732	5	751	1%
7B	693	698	5	751	1%
7B	739	749	10	751	1%
7D	0	10	10	639	2%
7D	49	59	10	639	2%
7D	0	457	457	639	72%
7D	375	639	264	639	41%
7D	563	639	76	639	12%
7D	612	639	27	639	4%

Data S1. List of 4506 wheat accessions and related information.

Data S2. Genotyping data of 4506 wheat accession with 113,457 genome-wide SNPs. (https://urgi.versailles.inra.fr/download/wheat/genotyping/Balfourier_et_al_Wheat_Phylogeography_DataS2.zip)

Data S3. Haplotyping data of 4403 wheat accessions with 8741 haplotypic blocks. (https://urgi.versailles.inra.fr/download/wheat/genotyping/Balfourier_et_al_Wheat_Phylogeography_DataS3.zip)