Supplementary figures, tables and note for Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding

Yusheng Zhao[a], Zuo Li[a], Guozheng Liu[a], Yong Jiang[a], Hans P. Maurer[b], Hans-Peter Mock[c], Andrea Matros[c], Erhard Ebmeyer[d], Ralf Schachschneider[e], Ebrahim Kazman[f], Johannes Schacht[g], Manje Gowda[h], C. Friedrich H. Longin[b], Jochen C. Reif[a,1]

[a] Department of Breeding Research, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Corrensstraße 3, 06466 Stadt Seeland, Germany

[b] State Plant Breeding Institute, University of Hohenheim, 70593 Stuttgart, Germany

[c] Department of Physiology and Cell Biology, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Corrensstraße 3, 06466 Stadt Seeland, Germany

[d] KWS LOCHOW GMBH, 29296 Bergen, Germany

[e] Nordsaat Saatzuchtgesellschaft mbH, 38895 Langenstein, Germany

[f] Lantmännen SW Seed Hadmersleben GmbH, 39398 Hadmersleben, Germany

[g] Limagrain GmbH, 31226 Peine-Rosenthal, Germany

[h] International Maize and Wheat Improvement Center (CIMMYT), P. O. Box 1041-00621, Nairobi, Kenya

[1] To whom correspondence may be addressed. Department of Breeding Research, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstraße 3, 06466 Gatersleben, Germany, Telephone: +49 (0)39482 5-840, Fax: +49 (0)39482 5-137, E-mail: reif@ipk-gatersleben.de

Supplementary Tables

Table S1 Detailed description of the 135 wheat parental lines for grain yield (GY, Mg ha^{-1}), 1000-kernel weight (TKW, g), gluten content (GC, %), kernel hardness (KH), protein content (PC, %), sedimentation volume (SV), starch content (SC, %), test weight (TW, g), brown rust severity (BR, scale 1-9), Fusarium head blight severity (FH, scale 1-9), powdery mildew severity (PM, scale 1-9), Septoria tritici blotch severity (ST, scale 1-9), and yellow rust severity (YR, %), frost tolerance (FT, scale 1-9).

Genotype	Code	LP[1]	GCA[2]	TKW	GC	KH	PC	SV	SC	TW	BR	FH	PM	ST	YR	FT
WS-001	F001	10.0	0.20	44.9	26.3	46.8	12.0	44.0	67.3	76.1	1.4	3.7	2.7	4.2	9.7	7.2
KWS-002	F002	9.5	-0.11	44.4	28.0	51.1	12.6	52.8	68.6	76.2	2.0	3.2	3.0	3.9	7.0	7.2
KWS-003	F003	10.1	0.19	45.9	25.6	53.5	11.9	46.2	68.7	76.9	3.0	4.0	2.1	2.7	6.4	5.3
KWS-004	F004	9.8	-0.22	41.8	25.0	39.7	11.8	36.0	69.4	77.5	0.9	4.3	3.0	5.3	31.1	6.7
KWS-005	F005	9.3	-0.18	48.8	28.0	55.1	12.7	55.9	68.9	79.2	4.1	4.6	2.7	4.1	12.9	3.8
KWS-006	F006	10.5	0.29	45.0	24.6	45.6	11.6	37.1	68.8	73.6	0.5	5.6	1.8	5.4	15.7	7.1
KWS-007	F007	9.8	0.07	46.5	27.1	53.9	12.2	34.2	68.2	75.9	4.4	5.0	3.2	5.9	6.5	6.8
KWS-008	F008	9.6	0.02	42.3	25.9	61.5	12.0	30.9	68.7	72.8	3.2	4.7	2.8	5.4	9.3	7.1
KWS-009	F009	9.7	-0.09	42.7	26.8	40.2	12.2	32.7	67.0	74.5	1.6	4.9	2.4	3.3	6.8	6.6
KWS-010	F010	9.5	-0.01	40.2	28.6	49.3	12.5	51.2	66.8	75.6	1.7	3.9	1.5	4.2	10.2	7.0
KWS-011	F011	9.7	-0.05	44.3	28.6	37.1	12.6	52.3	66.4	76.7	1.3	4.6	2.4	4.6	10.1	7.4
KWS-012	F012	9.8	0.34	43.8	26.0	47.5	11.8	28.2	69.1	71.9	2.0	4.1	2.0	4.5	7.7	6.9
KWS-013	F013	9.4	0.00	42.0	27.5	38.2	12.7	36.6	69.3	75.5	0.9	5.1	2.0	5.3	28.0	6.9
KWS-014	F014	9.7	0.11	39.5	25.6	42.5	12.0	33.6	68.8	73.7	1.3	4.4	2.1	3.9	24.0	6.1
KWS-015	F015	9.3	-0.10	46.0	27.4	43.5	12.4	40.5	66.4	74.8	3.3	3.6	1.5	4.8	6.5	4.7
KWS-016	F016	9.9	0.07	41.3	27.5	42.1	12.2	38.1	68.1	76.0	4.5	5.3	1.5	4.8	8.2	6.3
KWS-017	F017	9.5	-0.08	43.6	27.1	23.7	12.2	43.8	69.4	76.8	2.3	4.3	2.1	4.1	6.5	6.2
KWS-018	F018	9.1	-0.18	44.7	28.8	40.9	12.7	40.2	67.3	72.4	4.1	5.0	1.4	4.8	7.0	6.9
KWS-019	F019	8.4	-0.51	49.8	33.2	44.2	14.0	58.7	67.1	78.0	2.7	3.4	3.3	5.1	8.4	7.7
KWS-020	F020	10.4	0.25	45.6	29.2	39.9	12.5	48.2	67.6	74.6	2.4	4.9	3.8	3.1	9.8	3.9
KWS-021	F021	9.6	-0.16	46.8	27.7	29.5	13.0	45.8	67.7	77.5	1.5	4.3	3.9	5.6	8.2	6.1
KWS-022	F022	9.7	0.17	42.3	26.9	37.9	12.2	41.9	68.6	75.5	1.1	4.3	1.8	3.9	20.3	5.7
KWS-023	F023	9.3	-0.04	43.3	29.6	50.6	13.0	56.7	68.3	78.0	1.8	5.8	3.5	3.5	6.5	5.7
KWS-024	F024	9.9	0.00	49.0	27.5	57.4	12.2	34.9	68.2	77.5	1.7	3.9	3.9	3.8	7.8	8.2
KWS-025	F025	9.6	0.04	45.4	26.3	20.8	12.3	35.4	68.0	73.9	1.9	3.9	3.1	2.9	6.7	7.8
KWS-026	F026	9.7	-0.03	42.1	28.5	34.3	13.1	46.6	68.5	76.2	0.9	5.3	1.9	4.5	9.9	7.6
NOS-001	F027	9.0	-0.38	44.1	26.3	44.7	12.6	38.4	68.3	72.9	4.4	5.1	1.6	5.1	5.0	6.8
NOS-002	F028	9.7	-0.10	41.9	25.7	37.3	11.9	31.0	68.4	73.0	1.0	5.5	1.9	4.0	6.9	3.5
NOS-003	F029	9.4	-0.07	42.3	31.9	51.7	13.6	54.1	66.4	73.2	1.0	5.9	2.0	2.7	20.0	6.5
NOS-004	F030	9.4	-0.02	45.3	27.8	51.9	12.6	47.6	68.1	74.8	3.3	4.3	2.9	5.1	17.6	4.6
NOS-005	F031	9.3	-0.25	41.8	28.4	45.6	13.0	52.1	66.8	74.4	2.1	4.2	1.9	3.3	47.8	7.3
NOS-006	F032	10.0	0.07	43.2	27.7	57.9	12.4	53.1	68.0	71.9	5.2	5.3	1.9	5.5	10.0	6.7
NOS-007	F033	9.4	-0.11	45.8	28.4	32.8	12.8	43.7	68.5	74.2	2.3	4.1	1.9	5.1	10.4	7.2
NOS-008	F034	9.4	0.00	41.8	29.9	38.2	13.0	37.6	66.5	76.6	4.8	4.2	4.3	4.2	9.8	5.0
NOS-009	F035	9.4	-0.11	50.1	31.3	45.3	13.4	58.1	66.5	77.0	2.5	4.9	2.9	4.4	6.4	5.4
NOS-010	F036	9.5	-0.17	39.8	27.1	38.3	12.6	49.4	67.5	75.0	1.3	4.3	2.1	5.2	21.7	5.4
NOS-011	F037	9.7	0.09	46.6	27.4	56.0	12.2	46.0	68.7	75.1	2.1	3.8	3.0	4.1	6.6	6.8

Genotype	Code	LP[1]	GCA[2]	TKW	GC	KH	PC	SV	SC	TW	BR	FH	PM	ST	YR	FT
NOS-012	F038	8.3	-0.15	45.7	29.1	49.6	13.2	49.0	68.3	75.6	3.7	4.2	2.0	6.4	5.3	6.3
NOS-013	F039	10.1	0.15	42.7	27.9	54.2	12.3	45.0	68.1	76.9	1.4	3.2	1.2	4.0	5.1	2.8
NOS-014	F040	9.3	-0.23	41.3	26.5	35.2	11.9	37.0	68.0	73.0	1.0	4.7	3.0	6.2	6.6	4.9
NOS-015	F041	10.0	0.15	41.1	27.0	55.1	11.8	40.5	69.4	75.8	3.2	3.7	2.4	3.4	6.7	3.9
NOS-016	F042	9.7	0.04	43.3	26.9	47.4	12.1	30.1	67.4	76.5	1.0	3.1	1.3	3.8	16.1	5.2
NOS-017	F043	9.6	-0.11	51.8	26.2	40.3	11.9	36.6	68.9	72.5	1.4	5.1	2.5	6.0	4.9	4.7
NOS-018	F044	10.0	0.20	44.3	29.0	49.2	12.7	50.6	68.2	76.5	1.3	3.7	2.6	5.4	6.8	4.4
NOS-019	F045	9.5	-0.06	43.0	25.9	45.1	11.9	44.7	68.4	77.2	2.6	3.2	3.4	4.0	6.8	5.6
NOS-020	F046	10.0	0.08	45.2	26.7	36.8	12.1	30.7	68.1	75.4	1.5	4.7	1.3	2.4	6.7	2.8
NOS-021	F047	9.9	0.04	45.6	25.6	56.1	11.7	41.3	68.9	75.2	1.3	3.4	2.3	2.5	5.0	5.0
NOS-022	F048	9.6	-0.34	46.3	28.5	32.7	12.8	39.0	67.8	75.1	1.5	4.4	2.2	3.9	6.6	3.0
NOS-023	F049	9.7	-0.02	41.9	26.2	35.9	11.9	45.4	68.9	76.6	1.7	3.5	2.2	4.5	4.7	2.4
NOS-024	F050	9.6	0.00	45.8	28.1	46.5	12.3	43.2	67.4	75.2	3.1	3.6	3.1	3.8	10.3	3.4
NOS-025	F051	9.4	-0.24	47.4	26.4	34.9	12.3	38.0	69.0	74.7	1.3	4.9	2.6	4.6	6.8	4.6
NOS-026	F052	9.7	-0.12	39.4	30.2	44.5	12.9	38.5	67.1	78.1	1.4	3.8	1.8	2.7	6.7	3.6
NOS-027	F053	9.2	-0.24	42.4	28.6	44.2	13.0	49.5	67.4	74.0	1.9	4.2	1.7	4.7	8.0	4.2
NOS-028	F054	8.9	-0.30	41.1	27.8	45.4	13.0	44.7	67.4	76.8	1.7	3.0	2.3	3.5	6.5	7.3
NOS-029	F055	8.8	-0.40	40.0	30.3	39.6	13.5	46.1	67.5	76.1	2.4	3.7	1.5	4.2	5.1	4.0
NOS-030	F056	9.1	-0.10	45.1	25.0	52.3	11.6	44.7	69.3	78.3	2.2	5.2	2.0	4.5	6.7	5.7
NOS-031	F057	9.6	-0.07	46.3	25.7	49.8	12.0	49.0	69.2	75.9	2.2	4.5	2.3	5.7	24.1	7.6
NOS-032	F058	9.9	-0.03	43.3	27.7	43.7	12.4	42.5	67.7	74.9	3.8	4.0	2.2	4.8	10.2	5.9
NOS-033	F059	9.9	0.02	45.6	29.1	48.3	12.7	49.2	68.2	76.4	1.1	4.6	1.7	2.9	34.4	6.2
NOS-034	F060	10.3	0.27	45.7	27.3	47.6	12.2	51.6	69.6	77.1	2.6	3.9	2.1	2.5	8.3	5.6
NOS-035	F061	10.1	0.14	38.9	26.3	60.1	12.0	45.0	69.2	79.7	3.5	3.0	2.7	3.3	8.6	3.6
NOS-036	F062	9.7	0.13	43.7	29.1	47.2	12.9	49.7	67.2	79.2	4.0	3.4	1.7	3.8	12.2	1.7
NOS-037	F063	9.3	-0.22	45.7	28.3	49.5	12.9	57.0	68.7	77.7	2.2	3.9	2.5	5.0	8.6	3.2
NOS-038	F064	9.9	0.06	44.6	26.2	51.0	12.1	41.9	68.2	77.8	3.2	3.3	2.4	4.2	6.4	3.0
NOS-039	F065	10.4	0.19	48.3	27.0	48.9	12.3	39.0	68.7	76.9	2.4	3.8	3.0	2.7	10.7	5.5
NOS-040	F066	9.7	0.05	41.4	28.5	59.4	12.5	51.6	68.5	74.7	2.9	4.1	2.1	3.3	4.9	6.6
NOS-041	F067	10.3	0.16	48.8	27.7	45.2	12.2	33.6	67.0	73.8	5.0	3.9	1.5	4.2	6.5	3.5
NOS-042	F068	9.3	0.00	42.8	26.4	44.4	11.9	35.8	66.8	66.6	2.2	6.5	1.3	2.7	4.3	4.5
NOS-043	F069	9.3	-0.14	40.8	27.9	54.0	12.6	46.7	68.3	77.5	3.8	4.6	4.6	3.4	7.7	3.5
NOS-044	F070	9.9	0.01	42.7	28.3	50.3	12.5	38.4	66.5	70.5	1.2	5.7	1.9	3.4	17.0	5.2
NOS-045	F071	10.0	0.16	43.1	27.5	39.1	12.4	37.2	68.6	76.7	1.6	5.7	1.9	3.7	30.3	3.6
NOS-046	F072	9.3	-0.07	48.8	29.0	56.2	13.1	49.2	68.0	77.7	3.7	4.8	1.9	4.4	6.2	7.2
NOS-047	F073	9.8	-0.07	39.4	27.1	53.5	12.1	51.0	68.2	77.2	3.1	4.4	3.8	5.3	9.3	7.6
NOS-048	F074	9.6	0.02	43.0	26.9	49.1	12.0	46.9	68.6	76.0	1.7	5.1	2.5	3.9	10.0	4.9
NOS-049	F075	9.8	0.10	44.0	29.1	38.5	12.6	37.3	67.8	76.3	1.8	3.8	3.3	3.8	6.9	5.2
NOS-050	F076	9.7	-0.02	44.9	27.3	45.3	12.2	38.2	68.1	73.7	2.5	4.9	2.2	3.1	12.8	5.4
NOS-051	F077	9.3	0.07	46.8	26.7	47.0	12.5	46.1	69.2	76.3	2.4	3.4	2.4	4.6	6.9	3.9
NOS-052	F078	8.0	-0.33	50.5	28.2	55.4	12.9	46.9	68.8	75.2	2.7	4.8	2.0	4.3	6.5	7.4
NOS-053	F079	9.3	-0.03	46.0	30.1	43.7	12.9	33.6	66.5	71.8	2.6	4.8	1.3	4.0	7.0	7.4
NOS-054	F080	9.6	0.01	43.4	28.5	40.5	12.7	51.4	67.5	76.6	2.4	4.5	1.9	3.1	17.9	7.3
NOS-055	F081	9.1	-0.43	43.0	28.1	42.8	12.9	50.7	67.7	76.6	3.8	3.4	2.6	2.8	7.9	4.2
NOS-056	F082	9.7	0.03	47.8	29.7	58.3	12.9	47.3	67.7	75.7	4.1	3.7	3.6	3.7	13.4	5.2

Genotype	Code	LP[1]	GCA[2]	TKW	GC	KH	PC	SV	SC	TW	BR	FH	PM	ST	YR	FT
NOS-057	F083	10.0	0.21	42.2	27.4	42.7	12.5	54.8	69.2	76.4	2.4	4.3	2.0	3.4	11.8	4.7
NOS-058	F084	9.9	0.12	47.5	28.9	54.9	12.7	48.1	67.1	75.5	3.5	3.4	3.6	4.1	11.8	6.8
LG-001	F085	10.0	0.03	49.9	28.4	57.4	12.7	51.5	68.2	78.1	2.6	5.5	3.6	4.5	13.0	4.7
LG-002	F086	10.2	0.15	43.9	26.2	40.0	11.9	33.9	68.2	72.7	1.1	5.6	2.5	2.3	6.9	5.9
LG-003	F087	9.5	-0.25	48.1	28.1	56.8	12.9	47.9	68.3	76.4	3.4	3.9	3.0	4.3	10.8	4.9
LG-004	F088	9.8	-0.08	52.9	27.6	54.9	12.4	44.5	69.1	77.6	2.7	4.6	3.1	4.7	13.7	6.6
LG-005	F089	9.6	0.07	43.5	26.8	53.9	12.5	45.7	67.8	76.5	2.5	4.4	2.7	5.1	10.1	7.9
LG-006	F090	9.9	0.19	47.3	28.2	42.8	12.8	49.2	66.9	77.8	3.3	4.4	2.5	3.4	18.8	5.1
LG-007	F091	10.1	0.07	46.6	26.6	46.8	12.6	50.7	68.4	77.1	3.7	3.1	3.1	2.7	6.1	5.9
LG-008	F092	8.6	-0.15	45.6	27.0	47.5	12.2	44.4	69.0	75.5	3.0	3.6	2.5	5.6	8.2	8.6
LG-009	F093	9.8	-0.11	44.1	27.5	43.2	12.3	42.6	68.4	73.6	2.6	4.1	1.6	4.4	5.3	8.0
LG-010	F094	9.2	-0.02	48.5	27.1	45.0	12.3	50.0	68.1	75.1	3.2	6.1	2.3	6.0	10.8	5.8
LG-011	F095	9.6	0.05	44.7	26.0	34.7	12.0	28.4	67.9	70.2	3.0	5.0	3.1	4.4	5.3	7.0
LG-012	F096	9.4	0.15	44.1	26.5	43.2	12.0	39.8	68.7	73.5	4.2	4.6	1.7	4.7	8.1	6.8
LG-013	F097	10.3	0.47	50.0	25.9	36.0	12.0	31.5	68.5	73.8	1.2	5.0	1.5	3.1	21.4	7.3
LG-014	F098	10.1	0.16	49.4	27.6	52.5	12.5	33.0	66.7	75.1	1.3	4.8	1.7	1.9	6.4	4.2
LG-015	F099	10.1	0.27	42.9	26.1	47.3	11.7	40.6	68.5	72.4	1.2	5.4	1.5	4.4	15.4	3.9
LG-016	F100	10.1	0.17	49.8	25.6	51.8	11.8	41.5	68.4	76.3	1.4	4.0	2.9	3.5	10.4	5.1
LG-017	F101	10.4	0.29	49.3	25.1	45.4	11.7	41.1	69.2	75.9	1.0	4.8	3.4	2.8	12.3	7.1
LG-018	F102	10.4	0.41	52.4	27.7	46.6	12.1	43.5	69.0	78.6	2.1	3.9	1.5	3.4	6.5	3.0
LG-019	F103	9.8	0.05	48.1	26.9	53.1	12.5	46.8	69.5	76.8	2.4	4.1	1.8	3.8	9.8	6.4
LG-020	F104	10.0	0.13	47.2	26.2	53.1	12.1	48.6	68.8	79.2	2.0	4.3	2.2	5.2	8.6	5.5
LG-021	F105	10.0	0.24	45.3	26.2	48.1	11.9	44.1	68.2	77.7	3.8	3.7	2.3	4.3	19.7	5.9
LG-022	F106	10.0	0.07	48.5	26.0	45.7	11.9	42.5	69.5	76.7	2.2	4.7	1.6	3.0	61.6	5.1
LG-023	F107	9.8	0.07	46.3	28.3	42.3	13.0	58.9	68.5	76.3	2.4	4.0	3.5	3.3	6.5	4.2
SW-001	F108	8.9	-0.15	54.1	29.1	40.8	13.2	48.7	67.8	75.7	2.9	3.2	1.8	4.5	5.0	5.1
SW-002	F109	9.5	-0.02	42.6	27.9	50.3	12.8	42.2	67.7	78.1	4.4	3.5	2.5	4.3	15.7	3.2
SW-003	F110	9.0	-0.16	44.4	28.1	46.4	13.1	51.7	67.8	78.0	2.6	3.8	2.8	4.5	10.1	3.4
SW-004	F111	9.7	-0.12	45.2	28.9	50.5	12.8	52.1	68.5	76.3	3.6	3.6	2.7	2.9	11.6	3.5
SW-005	F112	9.8	-0.07	45.6	27.3	56.9	12.0	40.3	68.6	73.7	1.6	5.4	4.2	3.4	14.1	6.2
SW-006	F113	9.6	-0.01	46.1	29.2	58.0	12.8	54.8	66.4	75.6	1.9	4.9	2.6	4.9	6.6	7.3
SW-007	F114	9.8	0.07	48.7	29.3	44.0	13.0	54.2	67.4	76.1	3.9	4.7	2.9	3.5	7.7	2.2
SW-008	F115	10.4	0.30	43.9	28.3	48.3	12.6	48.0	67.3	76.0	6.4	3.5	1.4	3.2	6.7	3.6
SW-009	F116	9.1	0.00	50.0	27.6	42.7	12.8	51.3	67.2	78.2	2.3	4.8	1.8	2.9	11.4	3.3
SW-010	F117	9.2	-0.13	47.5	29.4	43.5	13.7	51.0	66.8	77.4	3.0	3.5	1.8	3.7	6.8	6.8
SW-011	F118	9.9	-0.01	47.8	28.4	41.8	12.3	40.1	67.2	75.9	1.6	4.5	2.8	4.9	12.2	7.0
SW-012	F119	9.8	0.12	44.5	28.8	39.9	12.9	33.5	67.6	73.3	1.6	5.2	1.6	2.6	11.1	5.4
SW-013	F120	10.2	0.13	49.1	26.2	54.3	11.9	45.4	69.0	76.2	4.6	4.4	3.0	3.3	7.1	3.6
NOS-059	M001	9.9	-0.01	47.5	27.1	61.4	11.9	47.4	68.8	77.5	3.9	3.7	2.5	3.4	5.2	3.4
NOS-060	M002	9.8	-0.04	48.6	28.0	53.3	12.4	43.4	67.5	77.3	3.5	3.4	3.1	4.6	10.0	4.2
NOS-061	M003	9.3	-0.20	43.2	27.4	47.1	12.2	48.6	67.8	75.3	2.9	4.3	3.4	3.6	6.8	4.3
NOS-062	M004	9.6	-0.25	45.8	29.2	46.2	12.5	50.6	67.2	76.0	2.7	3.4	2.8	3.8	6.9	3.3
NOS-063	M005	9.8	0.12	48.2	27.7	47.3	12.2	49.3	67.8	77.5	3.3	4.0	3.6	3.7	8.5	3.7
NOS-064	M006	10.4	0.20	47.3	27.8	48.4	12.0	43.2	68.0	76.6	3.8	3.9	2.8	3.7	13.5	3.1
NOS-065	M007	9.8	0.01	48.6	27.1	59.4	12.1	43.6	68.5	77.9	2.9	4.7	1.9	4.3	10.2	4.2

Genotype	Code	LP[1]	GCA[2]	TKW	GC	KH	PC	SV	SC	TW	BR	FH	PM	ST	YR	FT
NOS-066	M008	10.0	0.04	46.4	26.9	49.4	12.0	44.2	68.5	78.7	4.6	4.4	2.8	2.8	6.5	2.4
NOS-067	M009	10.0	0.20	45.8	25.5	45.1	11.7	48.3	69.1	77.3	2.3	4.0	3.2	3.6	19.4	3.3
NOS-068	M010	9.4	-0.06	43.2	26.6	59.5	11.9	37.8	68.7	78.2	1.3	3.3	3.0	4.3	60.0	4.9
NOS-069	M011	10.1	-0.05	47.9	27.9	51.8	12.5	53.8	68.5	77.6	3.4	3.2	2.2	4.2	10.2	4.0
NOS-070	M012	9.1	-0.13	42.5	27.6	49.4	12.4	44.5	69.1	76.1	3.5	3.5	2.6	4.6	12.7	3.3
LG-024	M013	9.9	-0.01	45.7	27.4	50.6	12.3	48.5	68.1	76.7	2.7	4.0	2.5	3.2	8.9	6.3
LG-025	M014	10.1	0.13	51.3	26.2	55.1	12.0	34.0	69.6	77.0	2.8	3.3	1.9	4.1	13.4	5.2
LG-026	M015	10.1	0.06	46.0	26.6	38.9	11.8	29.8	68.3	74.0	2.5	3.4	2.4	3.7	16.9	3.6

[1] LP refers to the line per se performance

[2] GCA refers to the general combining ability effects of the lines

Table S2. Variance of genotypes (σ^2_G), genotypes times environment interactions (σ^2_{GxE}), and of the residuals (σ^2_e) as well as broad-sense heritability estimates (h^2) for grain yield (Mg ha^{-1}) evaluated in 11 environments. The hybrid performance was decomposed into general (GCA) and specific combining ability (GCA) effects

Source	Grain yield
Lines	
σ^2_G	0.14**
$\sigma^2_{G\times E}$	0.22**
h^2	0.79
Hybrids	
σ^2_G	0.08**
σ^2_{GCA}	0.03**
σ^2_{SCA}	0.02**
$\sigma^2_{G\times E}$	0.12**
$\sigma^2_{GCA\times Et}$	0.05**
$\sigma^2_{SCA\times E}$	0.03**
σ^2_e	0.24
h^2	0.73

** Significantly different from zero at 0.01 level of probability

Table S3 Summary of whole-genome prediction accuracies with standard deviations for six methods evaluated with five-fold cross validation. T2 test sets included hybrids sharing both parental lines, T1 test sets comprised hybrids sharing one parental line, and T0 test sets contained hybrids having no parental line in common with the hybrids in the related training sets

Model	T0	T1	T2
GBLUP (A^1)	0.31±0.08	0.59±0.05	0.88±0.05
GBLUP ($A+D^2$)	0.32±0.08	0.65±0.05	0.89±0.05
GBLUP ($A+D+AA^3$)	0.32±0.09	0.65±0.05	0.90±0.05
GBLUP ($A+D+AA+AD^4$)	0.33±0.09	0.66±0.05	0.90±0.05
GBLUP ($A+D+AA+AD+DD^5$)	0.31±0.09	0.66±0.05	0.91±0.04
Bayes-Cπ (A+D)	0.32±0.08	0.66±0.05	0.90±0.05

[1, 2, 3, 4, 5] additive, dominance, additive × additive, additive × dominance, and dominance × dominance effects, respectively.

Table S4 Correlation among kinship matrices of additive (A), dominance (D) and respective digenic epistatic (AA, AD, DD) variance components (VC)

VC	A	D	AA	AD	DD
A	1.00	0.06	0.70	0.12	0.09
D		1.00	0.08	0.35	0.68
AA			1.00	0.06	0.11
AD				1.00	0.39
DD					1.00

Table S5 Variance components of genotypes (σ^2_G), genotypes times environment interactions (σ^2_{GxE}), and of the residuals (σ^2_{error}) as well as broad-sense heritability estimates (h^2) for the metabolites used to examine the 135 parental wheat lines

Metabolite	σ^2_G	σ^2_{GxE}	σ^2_{error}	h^2
Oxalic-acid	0.11***	0.02	0.78	0.58
L-Leucine	0.11***	0.17***	0.88	0.44
L-Valine	0.08***	0.11***	0.36	0.54
Ethanolamine	0.04*	0.18***	0.46	0.28
Norleucine	0.06***	0.13***	0.37	0.44
Isoleucine	0.07***	0.12***	0.30	0.49
L-Proline	0.16***	0.27***	0.45	0.55
Glycine	0.09***	0.06***	0.41	0.61
Glyceric_acid	0.08***	0.11***	0.42	0.52
Unknown_1	0.10*	0.30***	1.41	0.31
Threonic_acid-1_4-lactone	0.05*	0.13***	0.67	0.32
L-Threonine	0.62*	4.72***	1.48	0.26
Pyroglutamic_acid	0.15***	0.26***	1.20	0.43
Aspartic_acid	0.08***	0.11***	0.82	0.40
Butanoic_acid_2-amino	0.08***	0.13***	0.84	0.41
L-Glutamic_acid	0.11**	0.30***	1.28	0.33
L-Phenylalanine	0.11***	0.12***	0.80	0.49
Pentose_I	0.12***	0.23***	0.91	0.42
Pentose_alcohol_I	0.13***	0.11***	0.39	0.65
Putrescine	0.16***	0.19***	0.57	0.59
Pentose_alcohol_II	0.11***	0.09***	0.64	0.55
cis-Aconitic_acid	0.34***	0.96***	1.79	0.42
Lyxonic_acid	0.07***	0.16***	0.54	0.41
Ornithin	0.07*	0.25***	1.37	0.27
D-Fructose	0.14*	0.59***	1.82	0.28
Citric_acid	0.17***	0.24***	0.83	0.52
L-Arginine	0.49***	0.25***	1.02	0.74
D-Galactose_I	0.14***	0.14***	0.59	0.58
Quinic_acid	0.25***	0.20***	1.08	0.61
Unknown_3	0.05*	0.23***	0.45	0.28
Hexose_II_b	0.42***	0.38***	1.18	0.64
Pentose_alcohol_III	0.18***	0.13***	0.45	0.69
L-Tyrosine	0.11**	0.32***	1.01	0.36
Oligo_II	0.08**	0.24***	0.84	0.33

Supplementary Figures

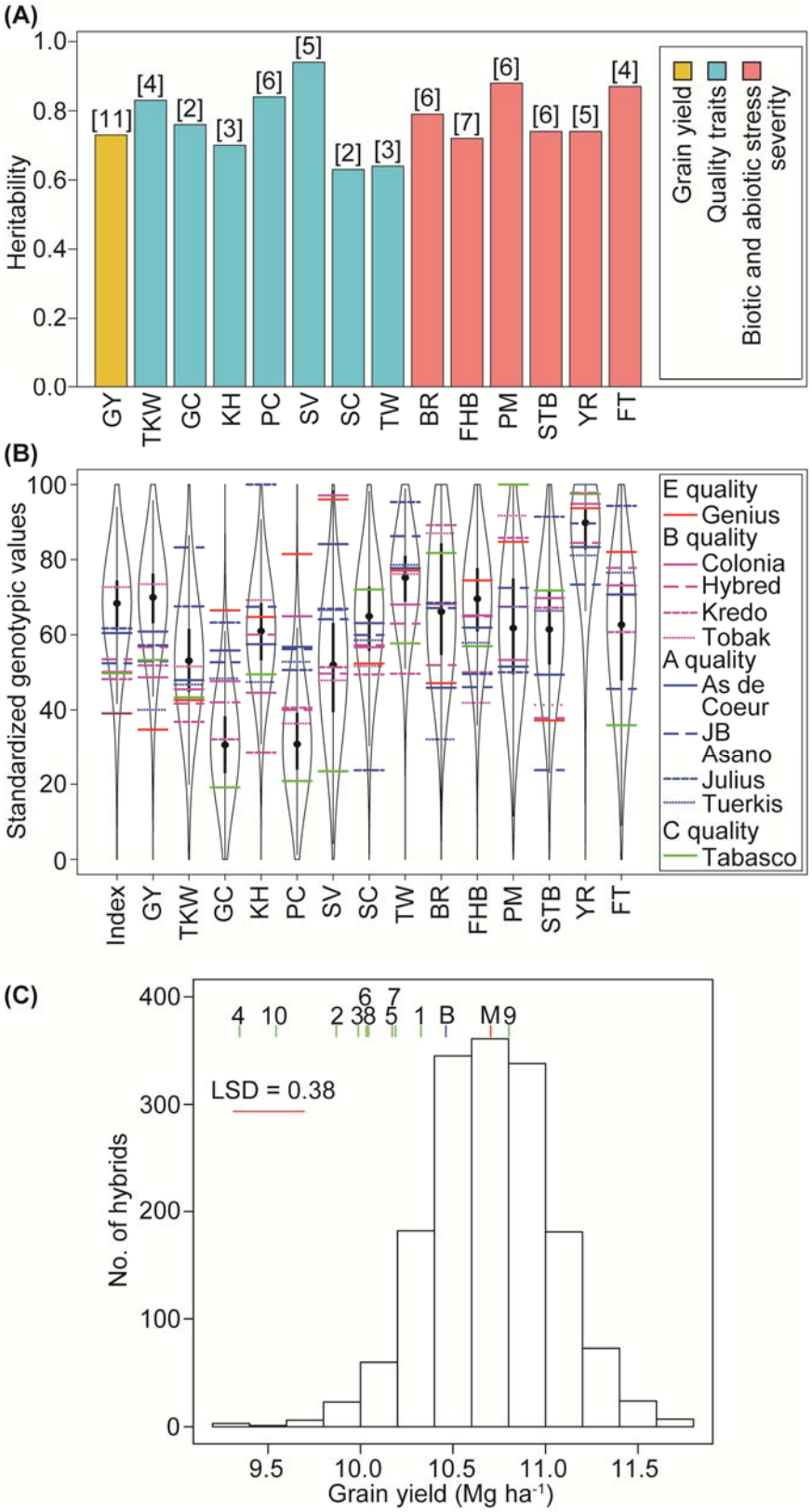

Fig. S1. (A) Broad-sense heritability estimates and **(B)** Violin plots of standardized best linear unbiased estimates for grain yield (GY, Mg ha^{-1}), 1000-kernel weight (TKW, g), gluten content (GC, %), kernel hardness (KH), protein content (PC, %), sedimentation volume (SV), starch content (SC, %), test weight (TW, g), brown rust severity (BR, scale 1-9), Fusarium

head blight severity (FHB, scale 1-9), powdery mildew severity (PM, scale 1-9), Septoria tritici blotch severity (STB, scale 1-9), yellow rust severity (YR, %), frost tolerance (FT, scale 1-9) and a weighted index (60% grain yield, 20% biotic and biotic stress severity, 20% quality) for the 1,604 hybrids and ten released varieties classified into four quality groups according to official variety testing in Germany. Number of environments in which the genotypes have been evaluated is given in brackets. **(C)** Distribution of hybrid performance for grain yield (Mg ha^{-1}) evaluated in 11 environments in Germany. B refers to the performance of the best parental line and M to the mean performance of the 1,604 hybrids. The numbers 1 to 10 indicate the performance of different commercial checks: 1 = *As de Coeur*, 2 = *Colonia*, 3 = *Kredo*, 4 = *Genius*, 5 = *Hybred*, 6 = *JB Asano*, 7 = *Julius*, 8 = *Tabasco*, 9 = *Tobak*, and 10 = *Tuerkis*. LSD refers to the least significant difference at a significant threshold of $P < 0.05$.

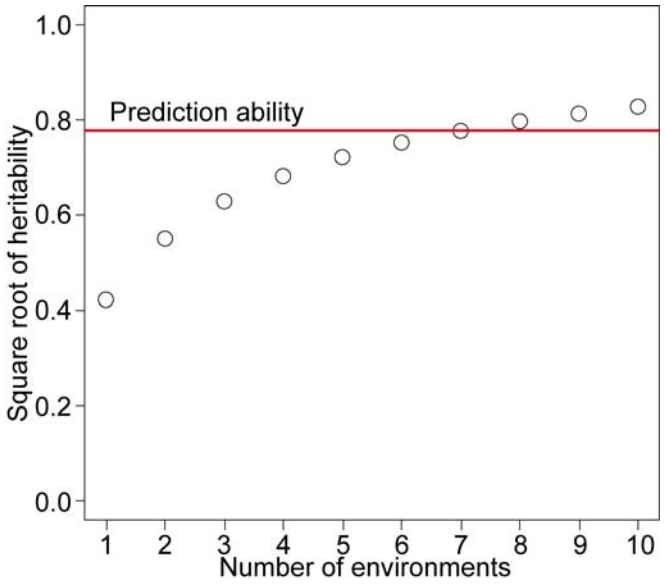

Fig. S2. Relationship between the square root of the broad-sense heritability (h) for grain yield and the number of environments evaluated. The prediction ability is given as red horizontal line.

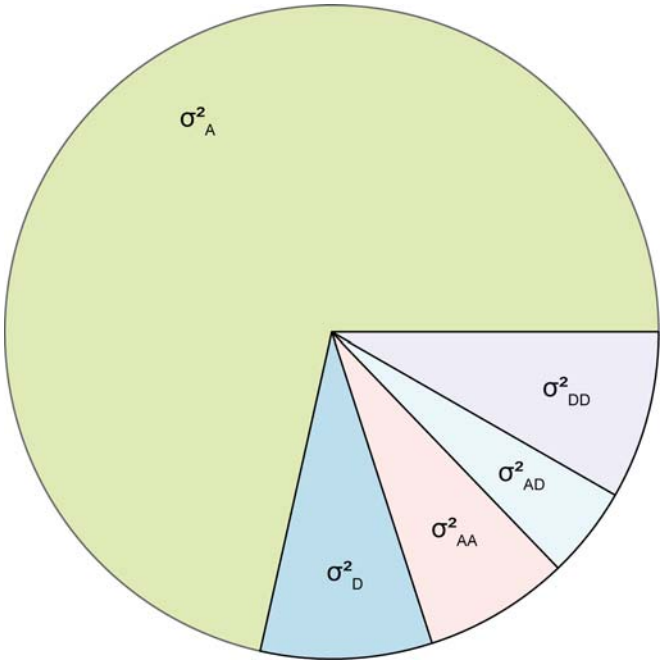

Fig. S3. Pie chart of genetic components of variance (additive variance σ^2_A, dominance variance σ^2_D and respective epistatic variance components) estimated with Bayesian generalized linear regression.

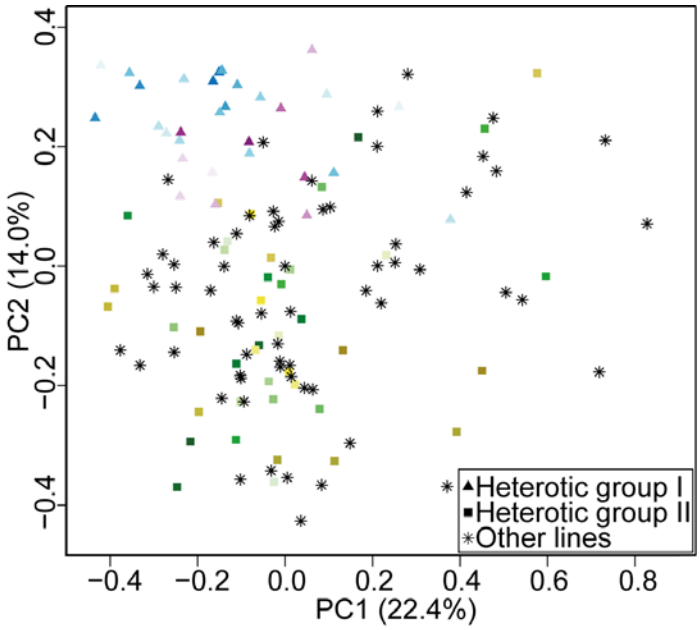

Fig. S4. Principal component analyses of the 135 parental wheat lines fingerprinted with a 90k SNP array. The color code reflects the membership to the heterotic groups depicted in Fig. 4.

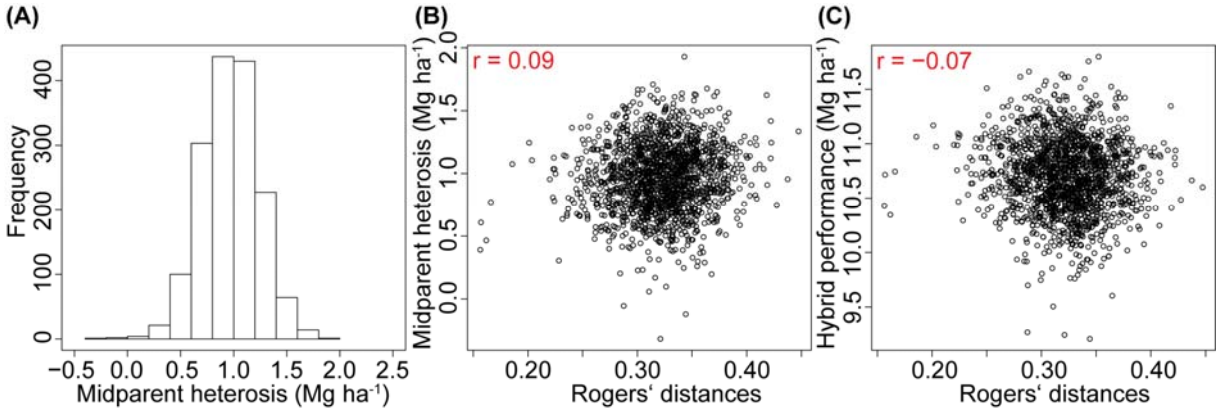

Fig. S5. (A) Distribution of midparent heterosis of the 1,604 hybrids, **(B)** association between Rogers' distances estimated based on 17,372 SNPs and midparents heterosis for grain yield, and **(C)** association between Rogers' distances and hybrid performance.

Supplementary Note

a. Absence of genetically distinct subpopulations

The presence of genetically distinct subpopulations impacts the implementation of cross validation scenarios for hybrid prediction and association mapping. We inspected the presence of subpopulations by using a complete linkage clustering method and observed absence of genetically distinct subpopulations with only minor influence of the breeding program from which the lines have been derived (Fig. Supplementary Note a-1). These findings are in accordance with previous observations for wheat lines adapted to Central Europe (17) and can be explained by the constant exchange of germplasm between breeding programs.

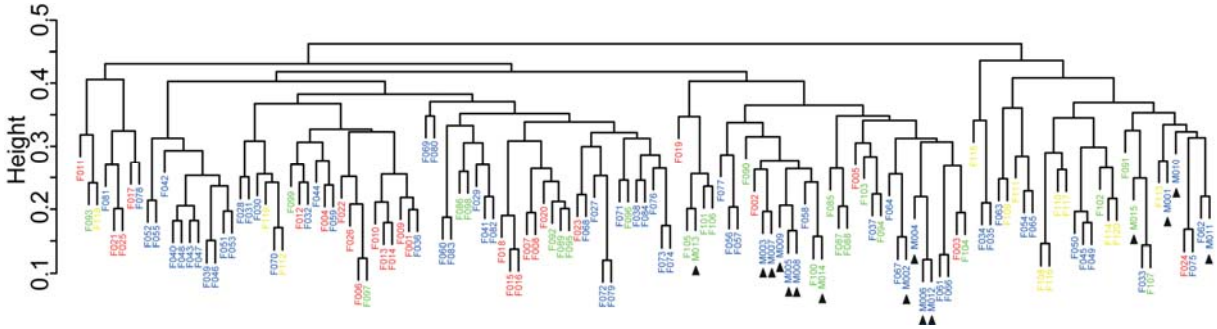

Fig. Supplementary Note a-1. Associations among the 135 wheat inbred lines revealed by complete linkage clustering method based on Rogers' distances among pairs of lines. Different colors refer to the origin of the inbred lines into breeding programs (red = KWS, blue = NOS, green = LG, yellow = SW). Labels marked with black triangles refer to males.

b. Chess-board-like cross validation and reliability criterion revealed high-quality of the predicted hybrid performances for all 9,045 pairwise single crosses

As in factorial mating designs relatedness between training and test set influences prediction accuracy, we followed previous suggestions (14) and sampled training sets consisting of 10 out of 15 male and 80 out of 120 female parental lines as well as 610 hybrids derived from them. From the remaining hybrids, test sets with three successively decreasing degrees of relatedness to the training set were formed (Fig. Supplementary Note b-1). Test set T2 most closely related to the training set included only hybrids derived from the same parents as the hybrids that had been evaluated, while the less related test set T1 included hybrids sharing one parent with the hybrids in the training set and the least related test set T0 included only hybrids having no parents in common with the training set. For each test set, we used 100 cross validations and estimated marker effects. The obtained marker effects were then used to predict the performance of the hybrids in the T2, T1, and T0 test sets. The prediction accuracy for each test set was estimated as the Pearson correlation coefficient between the predicted and the observed hybrid performances standardized with the square root of the heritability.

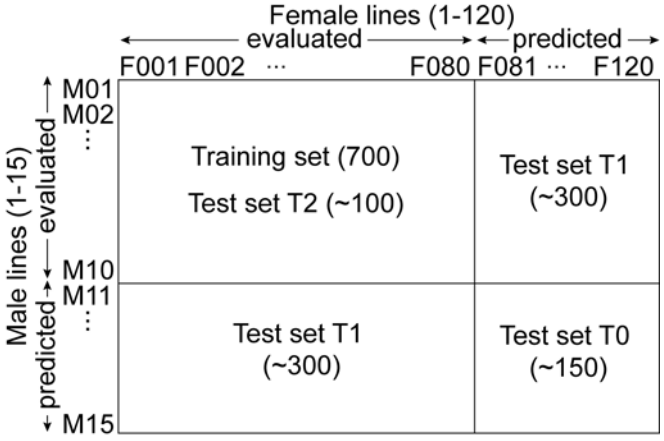

Fig. Supplementary Note b-1. Cross validation scenarios applied in our study. T2 test sets included hybrids sharing both parental lines, T1 test sets comprised hybrids sharing one parental line and T0 test sets contained hybrids having no parental line in common with the hybrids in the related training sets

Cross validations are expected to deliver similar estimates of prediction accuracies compared to independent validations as highlighted in the context of QTL mapping (43). One important requirement, however, is that the cross-validation scenarios mimic the relevant relatedness patterns. Our study is based on 1,604 single crosses. Thus, the applied cross validation scenarios yield robust estimates for crosses between the female and male lines. Nevertheless, the question arises whether the findings of the T2 scenario can be safely expanded to the expected prediction accuracies for the intra-female and intra-male crosses. We observed absence of a population structure among the male and female lines (Fig. Supplementary Note a-1), which clearly suggests that the prediction accuracies observed between female and male lines are comparable to those observed for intra-female and intra-male crosses.

To further confirm that the prediction accuracies observed between female and male lines are comparable to those observed for intra-female and intra-male crosses, we assessed prediction accuracy of particular individuals merely based on genotypic data using the reliability criterion (19). The reliability can be calculated via the GBLUP model, which is of the form $y = 1_n\mu + Zg + e$, where g is the vector of genotypic values, Z is the corresponding design matrix and e is the vector of residuals. We assume that $g \sim N(0, G\sigma_g^2)$, where G is the $n \times n$ genomic relationship matrix (41), and $e \sim N(0, \sigma_e^2)$. The reliability of the estimated genotypic value of the i^{th} genotype was defined as the correlation between the true and estimated genotypic value: $r_i = cor(g_i, \hat{g}_i)$. To calculate it, we need to extract the coefficient matrix of the mixed model equations (44): $C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} n & 1_n{}'Z \\ Z'1_n & Z'Z + G^{-1}\sigma_e^2/\sigma_g^2 \end{bmatrix}$. Let $\begin{bmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{bmatrix}$ be a generalized inverse matrix of C. Then, the reliability can be calculated as $r_i = \sqrt{1 - \frac{d_i\sigma_e^2}{\sigma_g^2}}$, where d_i is the diagonal element in C^{22} corresponding to the i^{th} genotype. Note, that $d_i\sigma_e^2 = SE(\hat{g}_i)^2 = var(g_i - \hat{g}_i)$ is the squared standard error or the prediction error variance of $\hat{g}_i$ (44). We observed a similar distribution of the reliability criterion for the factorial crosses as compared to the intra-

female and intra-male crosses (Fig. Supplementary Note b-2). Consequently, the prediction accuracies estimated based on the experimental data of the factorial crosses well approximate the prediction accuracy for all 9,045 pairwise single crosses.

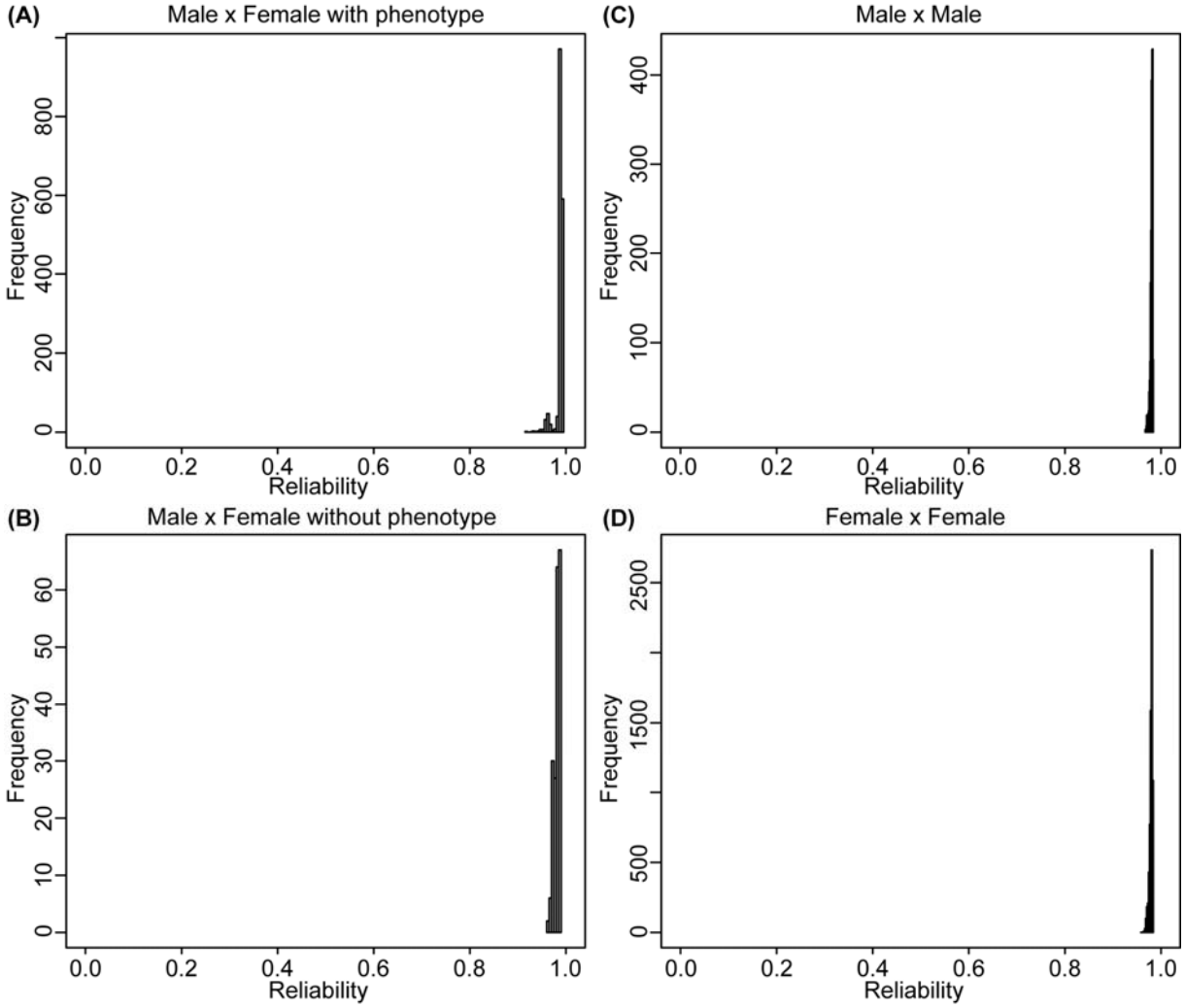

Fig. Supplementary Note b-2. Histograms of reliabilities estimated for the intra-female and intra-male hybrids as well as for the phenotyped and non-phenotyped single crosses between males and female lines

c. Association mapping reveals absence of large effect QTL for grain yield

We performed association mapping using the software ASReml-R 3.0 (45). and corrected for population stratification by fitting a polygenic effect for the breeding values of the parents, where the covariance was modelled using the kinship matrix estimated from the marker data. We also tested a model correcting for population structure with the kinship matrix and the first ten principal coordinates (Q+K model). We contrasted these approaches with a model not correcting for population stratification except considering a group effect (lines versus hybrids) using quantile-quantile plots. We compared the different models by plotting the observed versus the expected P-values (Fig. Supplementary Note c-1). Assuming that none of the SNPs is associated with the trait, the P-values are expected to follow the diagonal. The distribution is expected to deviate from the diagonal, however, if true marker-trait associations are present causing a bulge at the left side of the plot. As true marker-trait associations are not known, the true distribution of P-values cannot be predicted but strong bulges clearly indicate inflated

false-positive rates. Inspecting the quantile-quantile plots for our data revealed that population stratification had to be considered through modelling the kinship matrix for the lines and hybrids but correcting for population structure with principal coordinates was not required (Fig. Supplementary Note c-1).

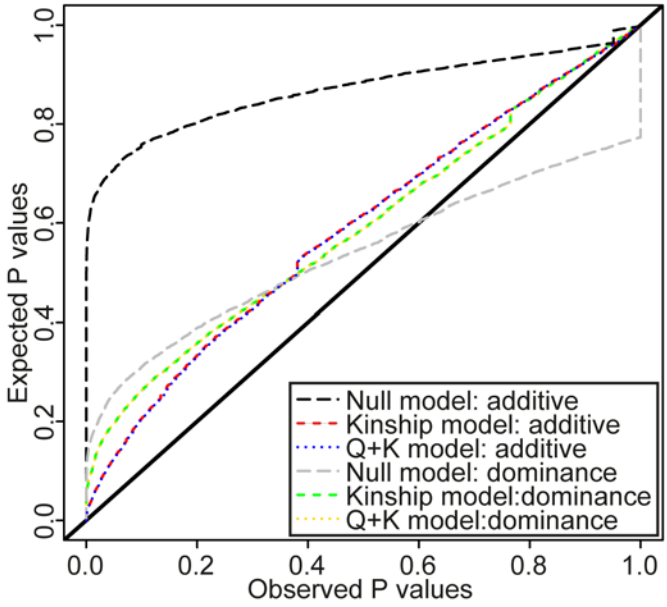

Fig. Supplementary Note c-1. Quantile-quantile plots for association mapping for additive and dominance effects without correction for population structure (Null model) and correcting for population structure with a kinship matrix (Kinship model) and the first ten principal coordinates (Q+K model).

Presence of large effect QTL was examined by studying the prediction accuracy of the association mapping applying cross validations. In the cross validations, we split the total data set into training and test sets as outlined in the material and methods part (Fig. Supplementary Note b-1). We used 100 cross validations and estimated marker effects of the QTL identified in the genome-wide association mapping scan in each cross validation run within the training sets. We applied the association mapping model outlined above correcting for population stratification with a kinship matrix. The effects for the detected QTL were estimated with a linear model and then used to predict the performance of the hybrids in the T0 test sets. The prediction accuracy for each test set was estimated as the Pearson correlation coefficient (r) between the predicted and the observed hybrid performance standardized with the square root of the heritability. We observed for a wide range of significance values prediction accuracies close to zero despite detecting a substantial number of QTL in every scenario (Fig. Supplementary Note c-2). These findings clearly suggested absence of large effect QTL for grain yield in our study.

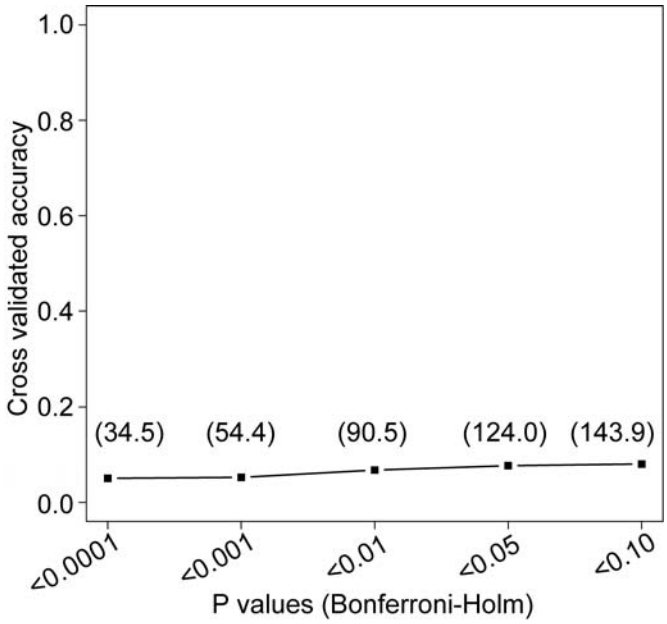

Fig. Supplementary Note c-2. Cross validated accuracies of prediction of association mapping for grain yield in the T0 scenario with hybrids in the test sets having no parents in common with the hybrids in the training set. Number in parenthesis refers to the average number of marker-trait associations detected in the 100 training populations.

d. Efficient designs of training populations

We evaluated the possibility to optimize the design of the training population to predict grain yield performance of related T2 hybrids using the G-BLUP(A+D) approach in combination with 100 cross validations. We compared four different designs of the training population: (1) A nested factorial mating design with 120 hybrids (single crosses of each male line with eight female lines; Fig. Supplementary Note d-1A). (2) An incomplete factorial design comprising 360 hybrids (single crosses among each female line with 3 male lines, there are 2 overlapping male parents in each neighboring single cross; Fig. Supplementary Note d-1B) with balanced missing data structure. (3) Top-cross design comprising 396 hybrids resulting out of crosses among all 120 females with three randomly selected male tester lines and among all male lines crossed with three randomly selected female tester lines (Fig. Supplementary Note d-1C). (4) As standard scenario we implemented an incomplete factorial design with random missing hybrid combinations using two population sizes of 120 and 360 hybrids (Fig. Supplementary Note d-1D). For all four scenarios, we studied the prediction accuracy including and excluding the 135 parental lines in the training populations. Our findings clearly showed that a drastic reduction in training population size from 1,604 to 360 hybrids is possible without substantially reducing the prediction accuracy (Table Supplementary Note d-1). Interestingly, a random sampling of 360 hybrids yielded equal or higher prediction accuracies than targeted designs such as the nested factorial, balanced incomplete factorial, or top-cross designs (Table Supplementary Note d-1).

Table Supplementary Note d-1. Prediction accuracy (r_G) and its standard deviation (SD) for four different designs of the training population including and excluding the 135 parental lines

Design	Size of training population	r_G	SD(r_G)
Including parents			
Nested design	120 hybrids	0.78	0.02
Balanced incomplete factorial design	360 hybrids	0.85	0.01
Topcross design	396 hybrids	0.79	0.03
Random design	120 hybrids	0.79	0.02
Random design	360 hybrids	0.86	0.01
Excluding parents			
Nested design	120 hybrids	0.62	0.05
Balanced incomplete factorial design	360 hybrids	0.82	0.02
Topcross design	396 hybrids	0.74	0.05
Random design	120 hybrids	0.62	0.05
Random design	360 hybrids	0.82	0.02

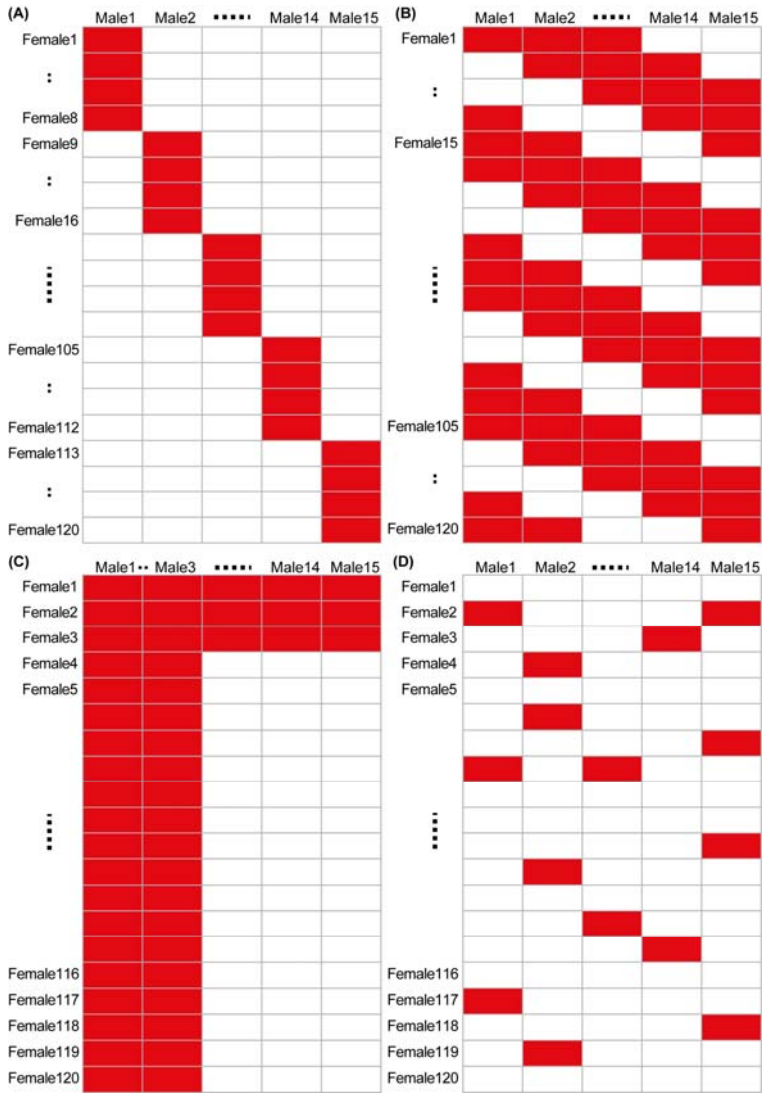

Fig. Supplementary Note d-1. Four different designs of training populations: (A) Nested design, (B) Balanced incomplete factorial design, (C) Topcross design, and (D) Random design.

e. Implementation, performance and robustness of the developed simulated annealing algorithm as well as long-term success of the identified heterotic pattern

We implemented a simulated annealing algorithm (46) to maximize the hybrid performance to search for heterotic groups. The details of the simulated annealing process are described as follows. Suppose we have a population of n parents $L = \{l_1, l_2, \dots, l_n\}$, and we need to find the heterotic groups of size m $\left(m < {}^{n}/_{2}, m \in N \text{ with } N \text{ referring to the natural numbers}\right)$. For the simulated annealing method, we could define the state space as $\Omega = \left\{(G_1, G_2) \middle| G_1 = \left\{l_{i_1}, l_{i_2}, \dots, l_{i_m},\ i_1, i_2, \dots, i_m \in N_L\right\}, G_2 = \left\{l_{j_1}, l_{j_2}, \dots, l_{j_m},\ j_1, j_2, \dots, j_m \in N_L\right\},\ G_1 \cap G_2 = \emptyset\right\}$, where $N_L = \{1,2, \dots, n\}$, and the goal function is: $F(G_1, G_2) = \frac{1}{m^2}\sum_{u=1}^{m}\sum_{v=1}^{m} f_{i_u,j_v}$, where $f_{i,j}$ is the estimated performance of the hybrid generated by crossing l_i with l_j *($i, j \in N_L$ and $i \neq j$).* The annealing schedule is $T = T_0 \frac{k_{max}-k}{k_{max}-1}$, where T_0 is the initial temperature, k is the cycle of iterations, and k_{max} is the maximum cycle of iterations. We employed $P(F(G_1, G_2), F(G_1', G_2'), T) = min\left\{1, e^{\frac{F(G_1', G_2') - F(G_1, G_2)}{T}}\right\}$ as the acceptance probability function from state (G_1, G_2) to state (G_1', G_2') at temperature T, which was proposed elsewhere (46). The central steps of the simulated annealing algorithm are summarized as follows:

Step1: Initializing all parameters. Set the initial values for temperature T_0, the maximal iteration cycles k_{max}, and the restart threshold r. Start from $k = 0$ and randomly sample *2m* parents from L as the initial heterotic groups (G_1, G_2). Let $F_{current} = F(G_1, G_2)$. The parents not chosen into the initial heterotic groups will form the "left set" $LEFT = L - (G_1 \cup G_2)$.

Step 2: Let $\left(G_{1,best}, G_{2,best}\right) = (G_1, G_2)$, $LEFT_{best} = LEFT$, and $F_{best} = F_{current}$.

Step 3: Let $k = k + 1$, calculate T, randomly choose one parent l_{left} from LEFT and one parent $l_{current}$ from $G_1 \cup G_2$, respectively, replace $l_{current}$ with l_{left}, and form the new state (G_1', G_2') and the new "left set" *LEFT´*.

Step 4: Calculate $F(G_1', G_2')$, and accept the new state in the following ways: (a) If $F(G_1', G_2') > F(G_1, G_2)$, accept the new state (G_1', G_2'), i.e., set $(G_1, G_2) = (G_1', G_2')$, $LEFT = LEFT'$, and $F_{current} = F(G_1', G_2')$. Further, if $F_{current} > F_{best}$, set $\left(G_{1,best}, G_{2,best}\right) = (G_1, G_2)$, $LEFT_{best} = LEFT$, and $F_{best} = F_{current}$. (b) If $F(G_1', G_2') \leq F(G_1, G_2)$, accept the new state (G_1', G_2') with the probability of $p = e^{\frac{F(G_1', G_2') - F(G_1, G_2)}{T}}$. If the new state is accepted, set $(G_1, G_2) = (G_1', G_2')$, $LEFT = LEFT'$, and $F_{current} = F(G_1', G_2')$.

Step 5: If $F_{best} - F_{current} > r$, set $(G_1, G_2) = \left(G_{1,best}, G_{2,best}\right)$, $LEFT = LEFT_{best}$, and $F_{current} = F_{best}$.

Step 6: Repeat Step 3 to Step 5 until k reaches k_{max}, then output (G_1, G_2), $F_{current}$, $\left(G_{1,best}, G_{2,best}\right)$, and F_{best}.

To try to avoid being trapped in the local optimum, we started the iterations at a relatively high temperature and repeated the above simulated annealing algorithm at least 10 times, each starting with a randomly chosen state. As final result we used the best performing replication.

We evaluated the performance of the developed simulated annealing algorithm to identify heterotic groups by maximizing the hybrid performance. First, we studied the computational time for different heterotic group sizes setting $T_0 = 0.75$ and $r = 7.5$ and repeating the algorithm for 10 times with an appropriate k_{max} (Table Supplementary Note e-1).

Table Supplementary Note e-1. Computational time of the developed simulated annealing algorithm under different parameter settings for identifying heterotic groups among the 135 parental lines maximizing the hybrid performance

Size of heterotic Groups (s)	T_0	r	k_{max}	Time for one run[1]	Time for all 10 runs
$s \leq 10$	0.75	7.5	100000	≤ 5.5 sec	≤ 55 sec
$10 < s \leq 20$	0.75	7.5	1000000	≤ 54 sec	≤ 9 min
$20 < s \leq 26$	0.75	7.5	10000000	< 10 min	< 100 min
$26 < s \leq 36$	0.75	7.5	60000000	< 1 h	< 10 h

[1] Computation time was assessed using following setting: processor - Inter(R) Core(TM) i5-3470 CPU @ 3.20GHz 3.20GHz, RAM: 8GB, 64-bit; Operating System, Software: R x64 3.1.1

For small heterotic group sizes, we tested the robustness of the developed simulated annealing algorithm against an exhausted enumeration algorithm (15). Both algorithms were used to identify heterotic groups of size 8 from 20 parents. The detected heterotic groups maximizing the hybrid performance were the same for both algorithms. The computational efficiency, however, was up to 13 times faster for the simulated annealing versus the exhausted enumeration algorithm. We further studied the repeatability of the simulated annealing algorithm for larger heterotic group sizes and observed reasonable repeatability with the difference of the group constitution reflecting only small differences in the maximum hybrid performance below 0.02%. Thus, the performance of the developed simulated annealing algorithm facilitates a robust identification of heterotic groups maximizing the hybrid performance.

The simulated annealing algorithm is based on the predicted hybrid performances of the 9,045 single crosses. The use of predicted and not observed values can, however, introduce a bias. Therefore, we devised a cross validation scenario exclusively for the phenotyped hybrids to examine the stability of the groups identified with the simulated annealing algorithm (Table Supplementary Note e-2). As no intra-female and intra-male hybrids were phenotyped, the cross validation scenario selects promising lines for heterotic groups only for the pattern male times female lines.

As reference base for the cross validations, we identified heterotic groups with two population sizes (male lines = {5, 7}; female lines = {10, 25}) based on the observed hybrid performances of all 1,604 single crosses. In every cycle of the 100 cross validation runs, we randomly selected 360 out of the 1,604 hybrids. The 360 hybrids and 135 parents served as training population for genomic predictions of the performances of all 1,604 single crosses. This scenario mimics the procedure for the prediction of the 9,045 hybrids based on the

phenotyped 1,604 single crosses and the 135 parents. We used the predicted hybrid performances of the 1,604 hybrids based on the training population of the 360 hybrids and 135 parents and identified heterotic groups with the above specified population sizes (male lines = {5, 7}; female lines = {10, 25}). We then compared the heterotic groups based on the predicted versus the observed hybrid performances and confirmed the stability of the groups identified with the simulated annealing algorithm: Around 80% of the individuals overlapped between the groupings based on predicted and observed hybrid performances (Table Supplementary Note e-3).

Table Supplementary Note e-2. Composition of the heterotic groups identified with the simulated annealing algorithm as well as average hybrid performance within (Intra) and between both heterotic groups (Inter)

Group	Group 1		Group 2		Inter
size	Selected lines	Intra	Selected lines	Intra	
2	M006 M009	11.12	F102 F097	11.29	11.43
4	M006 M005 M009 M014	11.08	F097 F006 F102 F012	11.05	11.33
6	M014 F061 M009 M005 M006 F115	11.09	F012 F101 F099 F006 F097 F102	11.06	11.26
8	F061 F115 M006 M015 M008 M014 M005 M009	11.02	F102 F012 F006 F020 F099 F101 F097 F060	11.08	11.20
10	F100 M005 F003 M006 F061 F090 M009 M015 M014 M008	10.99	F006 F099 F101 F012 F097 F060 F102 F020 F115 F001	11.08	11.16
12	F090 F105 F003 M006 F061 M015 F100 M014 F067 M005 M009 M008	10.99	F001 F006 F097 F012 F120 M001 F102 F101 F020 F099 F060 F115	11.05	11.12
14	F105 M014 M009 F067 F100 M015 F090 F104 M005 F061 M007 F003 M008 M006	10.96	F022 F001 F083 F097 M001 F006 F101 F020 F120 F060 F012 F115 F102 F099	11.01	11.10
16	M009 M007 F061 F105 F100 M008 M005 F003 F104 M014 F090 F065 M006 M015 F084 F067	10.95	F101 F006 F099 F086 F102 F012 F044 F022 F115 F097 F060 F001 F120 F020 M001 F083	10.99	11.07
18	F061 F104 M006 M008 F003 F084 F105 F100 F065 F090 M014 F067 M005 M009 M015 M007 F064 M013	10.93	F101 F115 F099 F060 F120 F102 F044 F071 F020 F097 F012 F022 F086 F083 F001 F032 M001 F006	10.95	11.05
20	M006 M009 F003 F104 M002 F100 F062 F090 F084 M007 M015 F105 M014 F064 F061 M013 F067 M008 M005 F065	10.91	F012 F097 F071 F041 F044 F020 F119 F001 F120 F060 F032 F101 F099 F102 F022 F006 F086 F083 M001 F115	10.92	11.03
22	M013 M008 F065 M006 F105 M009 M007 F067 F084 M010 F061 M005 M014 F106 F104 M002 F003 F100 F090 F062 F064 M015	10.90	F119 F120 F020 F097 F098 F086 F071 F099 F041 F083 F006 M001 F022 F044 F101 F115 F102 F001 F032 F060 F014 F012	10.91	11.01
24	F104 F037 F090 M008 F067 M013 M005 M009 M014 M002 F105 M015 F065 M007 M010 F106 F084 F066 F061 M006 F064 F003 F100 F062	10.88	F060 F115 F041 F119 F120 F032 F091 F098 F006 F083 F020 F001 F101 F096 F086 F102 F097 F022 F012 M001 F014 F099 F071 F044	10.88	10.98
26	M002 M010 F037 F090 M006 F065 M014 F003 F066 M005 F105 F062 M008 F067 F084 F104 M011 M007 F106 M015 F061 F064 M009 M013 F100 F103	10.86	F014 F060 F096 M001 F101 F012 F032 F086 F083 F001 F115 F071 F102 F098 F097 F099 F006 F022 F091 F120 F039 F020 F041 F016 F119 F044	10.86	10.96
28	F100 M009 M007 F062 M005 F106 F067 F104 F085 F061 F084	10.87	F016 F102 F022 F071 F032 F097 F060 F041 F096 F046	10.82	10.94

	M006 F037 F105 F101 M008 M011 F065 F064 M014 M013 F090 F066 M002 M010 F103 M015 F003		F039 F120 F012 F091 M001 F119 F075 F099 F007 F083 F115 F086 F020 F006 F044 F014 F098 F001		
30	M002 F064 M013 F090 F003 M009 F046 F104 F084 F065 F106 F062 F101 F103 F105 M014 F100 M015 M006 M008 F067 F085 F037 M005 M011 M012 F098 F061 F066 M007	10.85	F016 F107 F091 F006 F114 F119 F022 F115 F102 F071 M001 M010 F097 F020 F086 F001 F075 F014 F039 F012 F096 F032 F007 F099 F060 F044 F120 F041 F083 F077	10.80	10.91
32	M013 F085 F104 F084 M010 F062 F106 M005 F037 F064 M006 F105 F065 M014 M015 F066 F090 M009 F003 F061 M002 M007 M012 F100 F103 F024 F098 M011 F067 F046 F101 M008	10.85	F096 F032 F001 F119 F022 F059 F006 F041 F089 F120 F016 F044 F095 F086 F115 F107 F091 F075 F014 F020 F071 F114 F077 F039 F007 F083 F102 F099 F012 M001 F097 F060	10.77	10.89
34	F067 F062 M008 M013 F085 F106 F102 M012 M006 F024 F103 F064 F104 F120 F058 F101 M007 F065 F066 M014 M010 F100 F037 F105 M015 M005 F061 F090 F003 M002 F084 M011 M009 F114	10.85	F075 F089 F044 F012 F107 F007 F096 F083 F014 F032 F001 F119 F071 M001 F026 F115 F039 F006 F041 F047 F042 F098 F095 F016 F097 F059 F022 F086 F099 F060 F091 F077 F046 F020	10.72	10.87
36	M007 F065 F084 F114 F067 F064 F024 M012 M002 M005 F085 F058 M008 F003 M015 F002 M010 F061 M009 F106 F105 F102 F104 F062 F066 F037 M013 F103 F115 M006 F090 F120 M014 F100 M011 F101	10.85	F086 F007 F039 F008 M001 F071 F080 F016 F042 F091 F096 F060 F118 F099 F046 F077 F107 F014 F098 F001 F083 F059 F047 F012 F044 F022 F006 F097 F095 F020 F041 F075 F119 F089 F032 F026	10.69	10.85

Table Supplementary Note e-3. Overlap (%) of identified genotypes for the male (Male) and female heterotic group (Female) based on observed versus predicted hybrid performances of the 1,604 single crosses

Group size	Male	Female	Total
5+10	81.00	74.60	76.73
7+25	76.43	75.72	75.88

We contrasted the average hybrid performances of the heterotic pattern identified with the simulated annealing algorithm with those of potential heterotic patterns defined using standard clustering approaches. As clustering approaches we used the Ward's clustering based on the predicted hybrid performances and Rogers' distances (Fig. Supplementary Note e-1, 2). The average predicted hybrid performances of all potential pairs of groups identified by the clustering methods were smaller than the comparable scenarios of the heterotic patterns identified by the simulated annealing algorithm (Table Supplementary Note e-2). The highest yielding heterotic pattern identified based on the Ward's clustering method has a performance of 10.69 Mg ha^{-1} (Fig. Supplementary Note e-1) or 10.81 Mg ha^{-1} (Fig. Supplementary Note e-2), i.e., resulting in an increase of 0.17 Mg ha^{-1} or 0.29 Mg ha^{-1} compared to the average hybrid performance of all 9,045 single crosses. The performance of the corresponding heterotic pattern identified with the simulated annealing algorithm amounted to 10.97 Mg ha^{-1} for the corresponding group sizes of the Ward's clustering results based on Rogers' distances or 11.03 Mg ha^{-1} for the corresponding group sizes of the Ward's clustering results based on

predicted hybrid performances. This results in increases of 0.45 Mg ha^{-1} or 0.51 Mg ha^{-1} compared to the average hybrid performance of all 9,045 single crosses. Thus, using the simulated annealing algorithm doubled the gain in performance contributed by the Ward's clustering method underlining the value of the simulated annealing algorithm in order to identify high-yielding heterotic patterns.

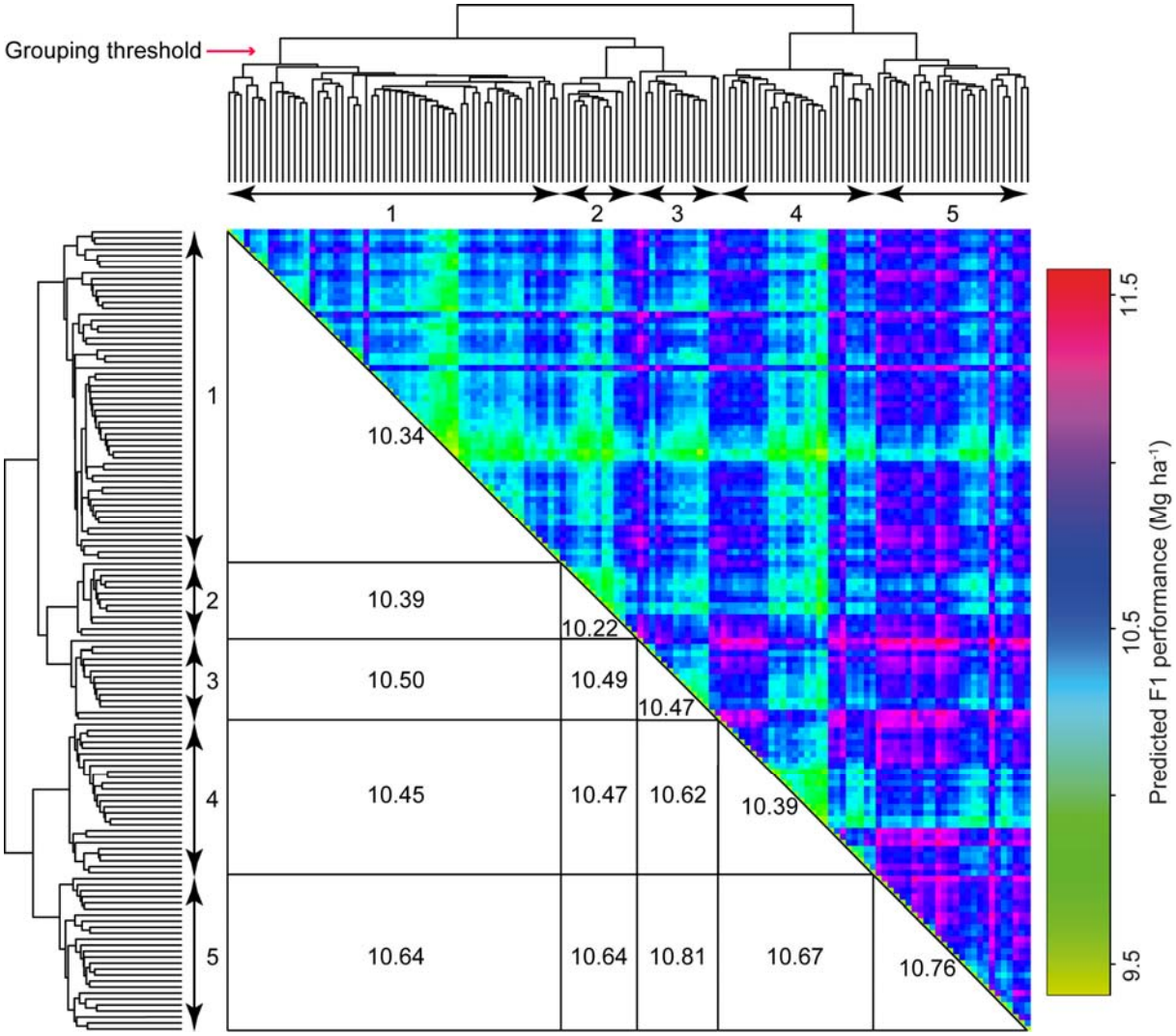

Fig. Supplementary Note e-1. Average hybrid performances (Mg ha^{-1}) of heterotic patterns (below diagonal) identified using the Ward's clustering approach based on the predicted hybrid performances of 9,045 single crosses (above diagonal).

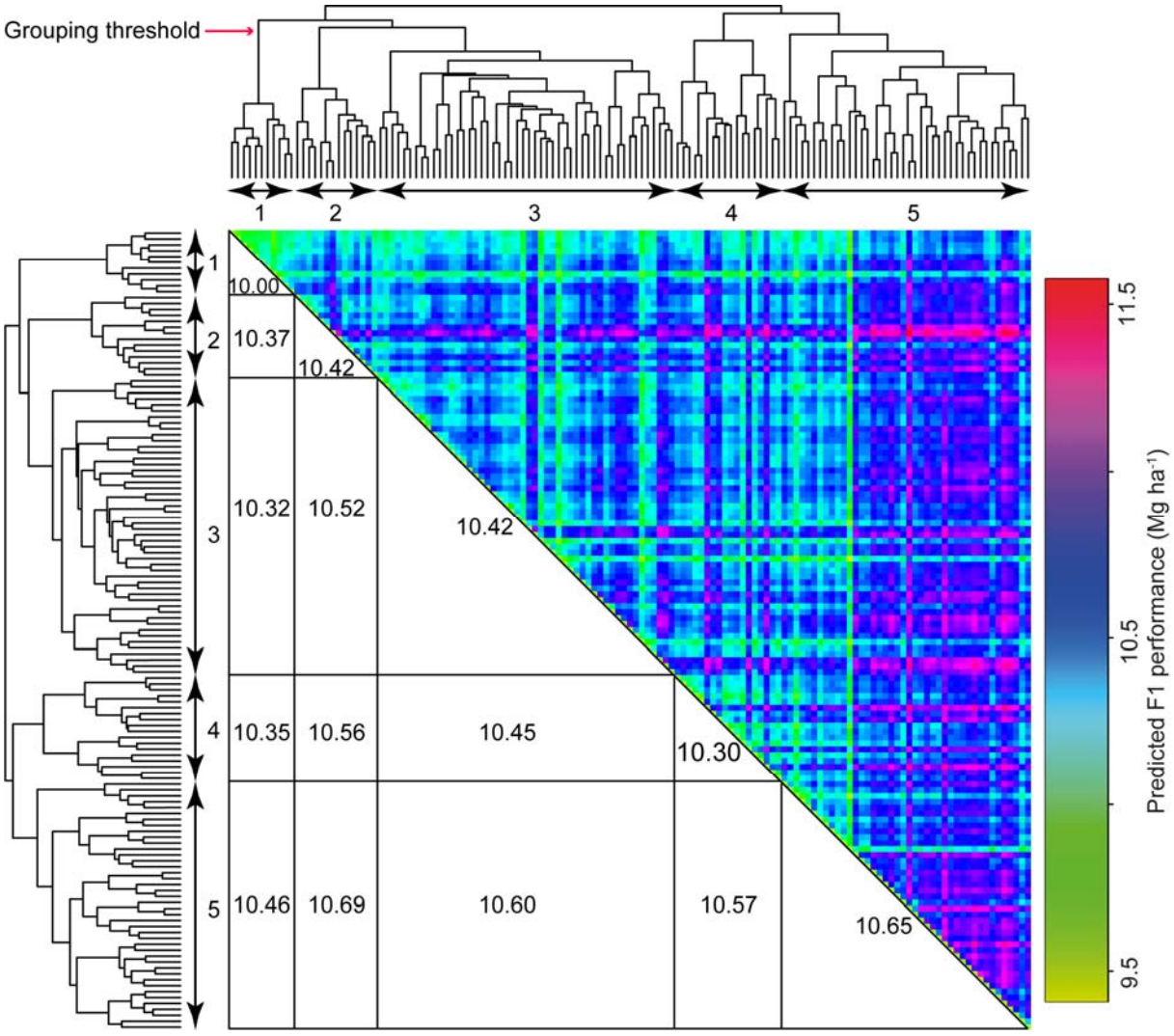

Fig. Supplementary Note e-2. Average hybrid performances (Mg ha^{-1}) of heterotic patterns (below diagonal) identified using the Ward's clustering approach based on the Rogers' distance among 135 parental lines. The predicted hybrid performances of all 9,045 single crosses were depicted above the diagonal.

In order to highlight the advantages of the developed strategy to search for an optimal heterotic pattern, we contrasted it with several alternative approaches: a) Parents were selected based on their *per se* performances followed by a random clustering of superior lines into heterotic groups. b) Parents were selected based on their general combining ability effects followed by a random clustering of superior lines into heterotic groups. c) Midparent heterosis values were used as input matrix of the simulated annealing algorithm. d) Better-parent heterosis values were used as input matrix of the simulated annealing algorithm. The findings of the selection of parents based on their *per se* performances were discussed in detail in the main part of the manuscript (Table 1). Therefore, we focused here on alternative scenarios relying on general combining ability effects, mid- and better-parent heterosis (Table Supplementary Note e-4). The developed strategy to search for an optimal heterotic pattern based on the hybrid performances outperformed all alternatives with respect to the most relevant criterion grain yield. Using the developed simulated annealing algorithm with heterosis values as input matrix resulted in a drastic reduction in grain yield performance of the hybrid population corresponding to up to 20.9 years of selection gain. Therefore, using the

simulated annealing algorithm based on heterosis values is despite the benefits in the ratio of additive versus the dominance genetic variance, no promising alternative compared to the developed approach using the simulated annealing algorithm based on a matrix of the hybrid performances.

Choosing parents for heterotic groups based on their general combining ability effects showed the smallest difference in grain yield performance still corresponding to up to 7.5 years of selection gain. Nevertheless, we observed a higher ratio (up to 5.3 times higher) of the additive versus the dominance genetic variance for the heterotic groups identified applying the simulated annealing algorithm based on estimated hybrid performances compared to the scenario of random grouping of lines selected based on their general combining ability effects. The enhanced relevance of additive genetic variance contributes to an increased recurrent selection gain (16). Moreover, predictions based on additive effects are more accurate than those based only on dominance effects (14). Summarizing, these findings clearly underline the value of the simulated annealing algorithm based on the hybrid performances in order to identify high-yielding heterotic patterns.

The long-term success of the identified heterotic pattern is assessed by estimating usefulness, selection limit, and representativeness of the heterotic pattern with respect to a defined base population. The concept of usefulness was introduced to quantify the performance of a selected fraction of a population after one cycle of selection of line breeding (22). Here, we implemented a modified usefulness criterion U_{G_1,G_2} to quantify the performance of a hybrid population resulting from crossing two heterotic groups (G_1, G_2) after one, five, and ten cycles of selection. For such moderate lengths of selection cycles and low selection intensities, we assumed that the genetic variance of the hybrid population is not impacted by selection. This assumption is based on previous simulation studies highlighting that for complex traits influenced by a couple hundred QTL, the genetic variance will not decrease within 10 cycles of selection (23). The modified usefulness of a heterotic pattern (G_1, G_2) is then defined as $U_{G_1,G_2} = \mu_G + n_c \sigma_G i_s h_s$, where μ_G is the mean genotypic value of the hybrid population generated by randomly crossing parents in G_1 with those in G_2, n_c denotes the number of the selection cycles, σ_G^2 is the genetic variance in the hybrid population, i_s is the selection intensity and h_s the square root of the heritability, whose product was defined as 1.75 ($h^2 = 0.73$; $i_{s=0.05} = 2.05$). The parameters μ_G and σ_G^2 were estimated for each pair of heterotic groups by combining computer simulations using the software package Plabsoft (47) with the developed genomic selection models. First, we generated SNP profiles of 1,000 gametes extracted from a gene orthogonal population, which was established by crossing the detected parental lines of each heterotic group. Second, the two simulated populations were crossed to generate a hybrid population. The genomic profiles of the hybrids were used in combination with the developed genomic selection models and the mean (μ_G) and variance of the hybrid populations σ_G^2 were estimated. Following the defining of the maximal long-term selection response (23), we estimated the theoretical selection limit of a hybrid population resulting from crossing two heterotic groups. The theoretical selection limit was estimated using the marker effects determined through genomic prediction in the population of the 1,604 hybrids and assuming absence of migration, mutations, and epistasis. Assume two alleles per QTL and n_M QTL underlying the trait under consideration. Suppose that the two alleles of the xth QTL are denoted as $A_{x,1}$ and $A_{x,2}$ $(x = 1,2, \dots, n_M)$. For the

heterotic groups G_1 and G_2 as defined previously, the theoretical selection limit is $TSL(G_1, G_2) = \mu + \sum_{x=1}^{n_M} max_{p_x \in P_x, q_x \in Q_x}\{ g_{A_{x,p_x}A_{x,q_x}}\}$, where μ is the overall mean of the hybrid population, $g_{A_{x,s}A_{x,t}}$ is the genotypic value of the genotype $A_{x,s}A_{x,t}$ $(s, t = 1,2)$ which was estimated through genomic prediction, and $G_{1,x} = \left\{A_{x,p_x},\ p_x \in P_x \subset \{1,2\}\right\}$, $G_{2,x} = \left\{A_{x,q_x},\ q_x \in Q_x \subset \{1,2\}\right\}$ are the sets of alleles at the xth QTL segregating in the individuals of G_1 and G_2, respectively.

The proportion of the unique genomes in the population represented by a subset of individuals was measured by the genetic representativeness (48). The genetic representativeness of two specific heterotic groups (G_1, G_2) is defined as follows. For the L, G_1, G_2, and N_L as defined previously, and for individual l_w $(w \in N_L)$, the proportion (Pr_w) of l_w's genome present in $G_1 \cup G_2$ is:

$$Pr_w = \left(\begin{bmatrix} a_{i_1i_1} & \cdots & a_{i_1i_m} & a_{i_1j_1} & \cdots & a_{i_1j_m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{i_mi_1} & \cdots & a_{i_mi_m} & a_{i_mj_1} & \cdots & a_{i_mj_m} \\ a_{j_1i_1} & \cdots & a_{j_1i_m} & a_{j_1j_1} & \cdots & a_{j_1j_m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{j_mi_1} & \cdots & a_{j_mi_m} & a_{j_mj_1} & \cdots & a_{j_mj_m} \end{bmatrix}^{-1} \begin{bmatrix} a_{i_1w} \\ \vdots \\ a_{i_mw} \\ a_{j_1w} \\ \vdots \\ a_{j_mw} \end{bmatrix}\right)^T \begin{bmatrix} Pr_{i_1} \\ \vdots \\ Pr_{i_m} \\ Pr_{j_1} \\ \vdots \\ Pr_{j_m} \end{bmatrix} = (A_{G_1 \cup G_2}{}^{-1} \vec{a}_{G_1 \cup G_2, w})^T \overrightarrow{Pr}_{G_1 \cup G_2},$$

where a_{ij} is the additive relationship between individual l_i and l_j, $i, j \in N_L$; $A_{G_1 \cup G_2}$ is the additive relationship matrix among all individuals in $G_1 \cup G_2$; $\vec{a}_{G_1 \cup G_2, w}$ is the vector of additive relationships between all individuals in $G_1 \cup G_2$ and individual l_w; $\overrightarrow{Pr}_{G_1 \cup G_2}$ is the vector of the genome proportion present in $G_1 \cup G_2$ (which is actually a vector of 1s). The additive relationship is defined as 1 minus the Rogers' distance of two individuals. The genetic representativeness of two heterotic groups (G_1, G_2), is the average value of Pr_w of all n individuals in L: $\overline{Pr} = \frac{1}{n}\sum_{w=1}^{n} Pr_w$.

Table Supplementary Note e-4. Comparison of overlapping genotypes (OG, %), yield increase expressed as number of years needed to realize this selection gain ((18), ΔSG, years), increase in midparent heterosis (ΔMPH, %, increase in average Rogers' distances (ΔRD, %), and decrease in the ratio of dominance versus additive genetic variance (ΔVC, %) contrasting the heterotic pattern identified based upon the simulated annealing algorithm versus three alternative approaches: Parents were selected based on their general combining ability effects followed by a random clustering of superior lines into heterotic groups (GCA selection). Midparent heterosis values were used as input matrix of the simulated annealing algorithm (Midparent heterosis). Better parent heterosis values were used as input matrix of the simulated annealing algorithm (Better-parent heterosis).

Group size	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36
GCA selection																		
OG	50	50	58	69	75	71	71	81	81	83	82	81	87	86	88	89	87	86
ΔSG	7.5	6.9	6.3	3.4	2.9	3.1	3.2	2.5	2.3	2.5	2.7	2.7	2.3	2.4	2.2	2.1	2.3	2.3
ΔMPH	12	13	11	5	5	5	5	4	5	7	7	6	6	5	5	6	6	5
ΔRD	1	2	6	2	1	1	1	2	3	3	3	2	3	3	3	3	3	3
ΔVC	81	84	83	76	71	44	45	56	53	49	42	43	51	43	39	39	38	38
Midparent heterosis																		
OG	0	50	42	38	35	38	32	34	39	48	52	52	56	52	57	61	62	67
ΔSG	20.4	6.2	7.7	9.0	9.6	8.8	10.1	9.0	8.5	7.6	7.0	7.2	6.6	6.6	5.8	5.8	5.4	4.9
ΔMPH	-9	-10	-9	-9	-10	-9	-9	-8	-8	-7	-7	-6	-6	-7	-6	-5	-5	-4
ΔRD	-15	-10	-8	-14	-15	-12	-12	-10	-10	-10	-10	-10	-9	-9	-8	-8	-7	-6
ΔVC	-2480	-642	-951	-634	-698	-959	-848	-709	-686	-551	-419	-372	-338	-324	-325	-311	-245	-234
Better-parent heterosis																		
OG	0	25	33	0	0	0	0	0	3	10	14	19	19	20	25	34	46	51
ΔSG	7.6	7.0	6.5	20.9	20.2	19.2	16.9	15.6	14.8	13.7	13.1	13.0	13.2	13.1	12.6	11.7	10.3	9.7
ΔMPH	-3	-4	-4	-3	-3	-2	-3	-3	-2	-1	0	0	0	-1	0	1	1	1
ΔRD	-7	-9	-6	-15	-15	-13	-12	-11	-9	-9	-8	-8	-8	-7	-6	-7	-5	-4
ΔVC	88	91	55	63	56	38	-15	-5	8	1	9	15	5	13	14	23	28	26

f. Phenotypic data analyses

The 1,749 genotypes (1,604 single-cross hybrids, 135 parents, 10 commercial varieties) were evaluated in two years at totally 11 environments in Germany. The locations were Adenstedt (Ade), Seligenstadt (Sel), Böhnshausen (Boh), Hohenheim (Hoh), Hadmersleben (Had), and Harzhof (Hhof). The experimental design at each environment consisted of 3 trials. The trials were partially replicated 20x10 (augmented with 10) alpha lattice designs which were connected by 10 replicated common checks (Fig. Supplementary Note f-1). The same seeding rate was used for both parental lines and hybrids. The plot size ranged from 5 m² to 7.4 m². Harvesting was performed mechanically and adjusted to a moisture concentration of 140 g H_2O kg^{-1}. We checked for the presence of neighboring effects due to plant height. We observed, however, absence of an association between grain yield and plant height of the two adjacent plots (e.g., average Pearson moment correlation of r = 0.05 for the three trials at the environment Seligenstadt 2012). Besides this, further data was collected for quality traits (37) and on separate observation plots also for abiotic and biotic stress resistances), Details on the analyses of these traits were published elsewhere (37).

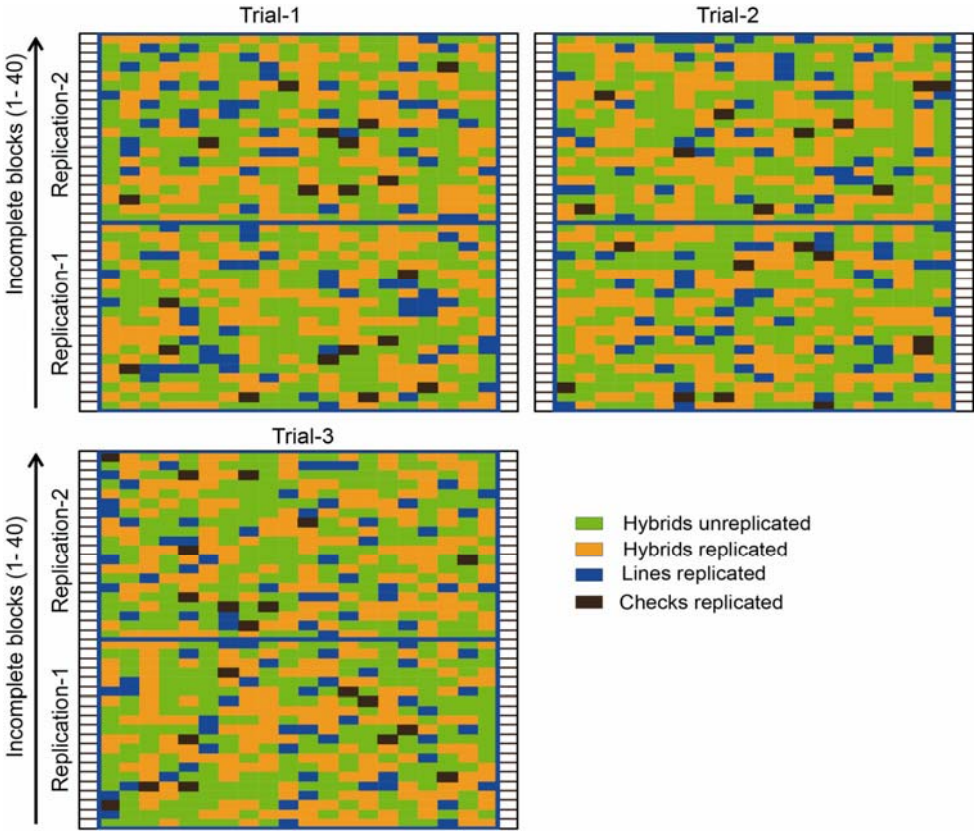

Fig. Supplementary Note f-1. Graphical illustration of the experimental field design used at Seligenstadt in 2012.

We applied a two-step procedure proposed by Möhring and Piepho (49) for the analyses of the grain yield data across environments using the standard error as weighing factor. In the first step, we used a mixed model procedure for the analyses of individual environments modelling effects for genotypes, trials, replications nested within trials, and blocks nested within trials and replications (37). Best linear unbiased estimates (BLUEs) of

genotypes were calculated. We inspected the pairwise correlations among BLUEs of the 1,749 genotypes evaluated in the eleven environments and observed absence of grouping into distinct mega-environments (Fig. Supplementary Note f-2).

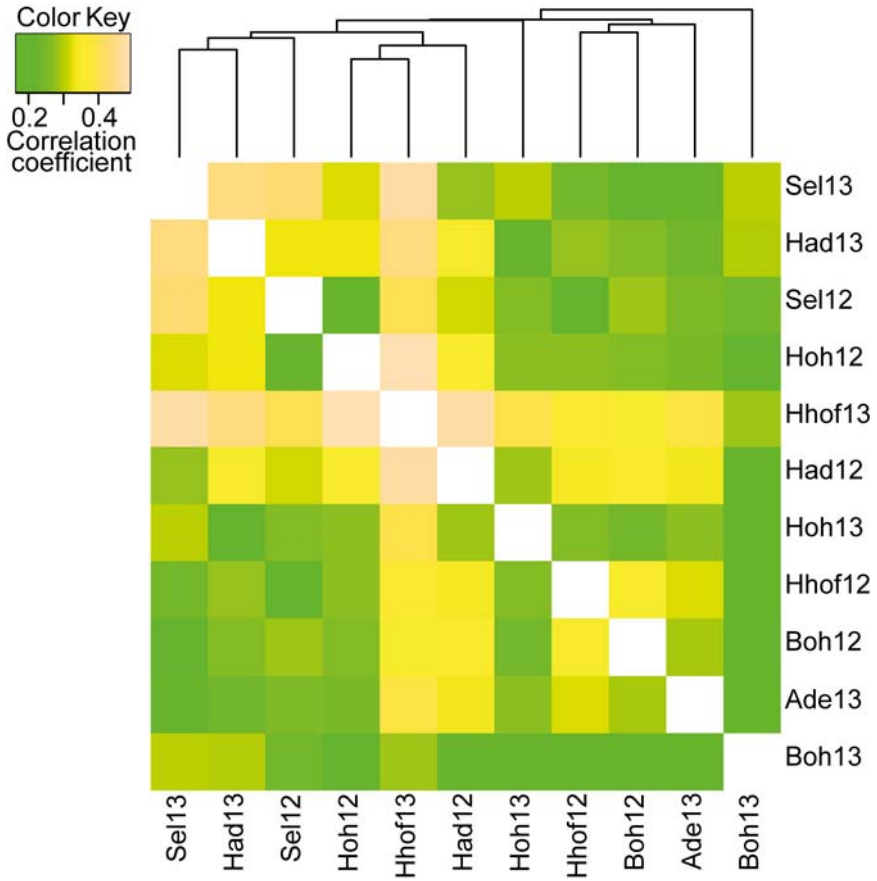

Fig. Supplementary Note f-2. Correlations among BLUEs of the 1,749 genotypes evaluated in eleven environments.

For the analyses across environments, we applied the following linear mixed model based on BLUEs of individual environments:

Grain yield ~ Group + Environment + Lines + Hybrids + Lines:Environment + Hybrids:Environment.

All effects except the group effect were treated as random. The total variance of hybrids was further decomposed into variance due to general combining ability effects (GCA) of males and females, and variance due to specific combining ability (SCA) of crosses:

Grain yield ~ Group + Environment + GCA_{males} + $GCA_{females}$ + SCA + GCA_{males}:Environment + $GCA_{females}$:Environment + SCA:Environment.

We assumed that the variance due to GCA is same for both males and females and estimated the variance components by the REML method using the software ASReml-R 3.0 (45). Significance of the variance component estimates were tested by model comparison with likelihood ratio test (50). Broad-sense heritability was calculated as the ratio of genotypic to phenotypic variance, $H^2 = \sigma_G^2/(\sigma_G^2 + \sigma_{G\times E}^2/E + \sigma_e^2/(E \times R))$, where E refers to the

number of environments, R is the average number of replications per entry at a location, and σ_e^2 refers to pooled error variance. In addition, we assumed fixed genotypic effects to obtain BLUEs of the genotypic values of hybrids, parents, and commercial varieties.

Commercial heterosis was calculated for each hybrid as CH = F1 - Check, where F1 denotes the grain yield performance of the hybrid and check refers to the grain yield of the best performing commercial variety. Midparent heterosis was calculated as MPH = F1 – (P1 + P2)/2, where P1 and P2 denotes the performance of the parental lines of the respective hybrid. We used the least significant difference (LSD) at an alpha level of 5% to test whether a hybrid outperformed the best commercial check.

g. Metabolite profiling

We sampled for each of the 135 parental lines 10 flag leafs per replicate at three environments at the time when >60% of the genotypes had reached BBCH-69 (39). Flag leaf samples were cut off, bulked, shock frozen in liquid nitrogen and further transported on dry ice. All plots were sampled during 9-11 am within 120 min. Prior extraction samples were freeze dried and homogenized by means of a ball mill (MM200, Retsch, Haan, Germany). Polar leaf metabolites were extracted twice with 500 µl/20 mg DW of 70% MeOH containing 50 µM 13C6/D7 glucose as an internal standard (Sigma-Aldrich, Munich, Germany). Supernatants were collected by centrifugation at 20,000g, 4°C for 20 min and combined. For phase separation 1,000 µl of water and 500 µl of chloroform were added. In order to capture systematic shifts during extraction and measurement we created a mixed sample consisting of equal amounts of all samples. The measurement of polar flag leaf extracts followed the protocol of Lippmann et al. (40). Data acquisition and processing was performed with MassLynx 4.1 software (Waters, Milford, MA, US).

To achieve homoscedasticity of the residuals of metabolites, the Box-Cox power transformation was applied. After outlier tests (51), we used a one-step model to estimate the genetic variance components of lines as well as the variance of genotype × environment interactions. Significance of variance component estimates were tested by model comparison with likelihood ratio tests where the halved P values were used as an approximation (50). A total of 34 metabolic traits were included for further analyses exhibiting significant ($P < 0.05$) genetic variances (Table S5). Using the variance components, we estimated the heritability on an entry-mean basis. In addition, we assumed fixed genetic effects and estimated the best linear unbiased estimates (BLUEs) of lines. The data analyses of the metabolic traits were performed using the software ASReml-R 3.0 (45).

h. Statistical methods for genomic prediction

Genomic best linear unbiased prediction model (G-BLUP) including additive, dominant and epistatic effects

Let n be the number of genotypes. The G-BLUP model has the following form:

$$y = 1_n\mu + g_a + g_d + g_{aa} + g_{ad} + g_{dd} + e,$$

where y is the vector of phenotypic records, 1_n is an n-dimensional vector of ones, μ refers to the mean. The total genotypic values are decomposed into five parts: additive (g_a), dominance (g_d), additive × additive (g_{aa}), additive × dominance (g_{ad}), and dominance × dominance effects (g_{dd}). Please note that g_{ad} represents a composite of additive × dominance as well as of dominance × additive effects. We assumed that μ is a fixed parameter, $e \sim N(0, I\sigma_e^2)$, $g_a \sim N(0, G_a\sigma_a^2)$, $g_d \sim N(0, G_d\sigma_d^2)$, $g_{aa} \sim N(0, G_{aa}\sigma_{aa}^2)$, $g_{ad} \sim N(0, G_{ad}\sigma_{ad}^2)$, and $g_{dd} \sim N(0, G_{dd}\sigma_{dd}^2)$, where the matrices G_a, G_d, G_{aa}, G_{ad}, and G_{dd} are the relationship matrices corresponding to additive, dominance, and epistatic genotypic values. We further assumed that all other possible covariance terms are zero.

The different relationship matrices were calculated as follows: Let $X = (x_{ij})$ be the $n \times p$ matrix of SNP markers, where x_{ij} equals the number of a chosen allele at the j-th locus for the i-th genotype (so $x_{ij} = 0, 1$ or 2). Let p_j be the allele frequency of the j-th marker. We defined the additive design matrix $W = (w_{ij})$ by setting $w_{ij} = x_{ij} - 2p_j$. Then the additive relationship matrix is $G_a = \frac{WW'}{2\sum_{k=1}^{p} p_k(1-p_k)}$, which is the same as in the standard G-BLUP model (41). Next, we defined the $n \times p$ dominance design matrix $D = (d_{ij})$ as follows (52):

$$d_{ij} = \begin{cases} -\dfrac{2p_{12}p_{22}}{\theta}, & \text{if } x_{ij} = 0 \\ \dfrac{4p_{11}p_{22}}{\theta}, & \text{if } x_{ij} = 1 \\ -\dfrac{2p_{11}p_{12}}{\theta}, & \text{if } x_{ij} = 2 \end{cases}$$

where $\theta = p_{11} + p_{22} - (p_{11} - p_{22})^2$. Then the dominance relationship matrix is $G_d = nDD'/\mathrm{tr}(DD')$. Here tr denotes the trace of a matrix, i.e., the sum of all diagonal elements. The epistatic relationship matrices were defined as follows: $G_{aa} = G_a \# G_a$, $G_{ad} = G_a \# G_d$, and $G_{dd} = G_d \# G_d$, where # denotes the Hadamard product (element-wise product) of matrices. The above model considered additive (A), dominance (D) and all first-order digenic epistatic effects (AA, AD and DD). We also considered reduced models which only include A (which is the same as the standard G-BLUP), A+AA, A+D+AA, or A+D+AA+AD.

The G-BLUP model was also used for the metabolome-based hybrid prediction and joint genome- and metabolome-based hybrid prediction. The metabolome-based hybrid prediction corresponds to the above outlined model exclusively focusing on the additive part:

$$y = 1_n\mu + g_{aM} + e,$$

with the additive relationship matrix defined as one minus the Euclidean distance matrix calculated by the metabolomic profiles of each parental line.

For joint genome- and metabolome-based hybrid prediction the model is:

$$y = 1_n\mu + g_{aM} + g_{aS} + g_{dS} + e,$$

with g_{aM} denoting the additive effects estimated based on metabolomic profiles, and g_{aS} and g_{dS} are the additive and dominance effects of SNP markers, respectively. Standard deviations of the prediction accuracies were estimated using a bootstrap procedure. All the above models were implemented using the R package BGLR (53).

Bayesian model (Bayes Cπ)

The Bayes Cπ approach has been previously described for additive effects (42), extended towards additive and dominance effects, and implemented for hybrid wheat prediction (20). The model for Bayes Cπ including additive and dominance effects is defined as:

$$Y = 1_n\mu + Z_A\delta_a a + Z_D\delta_d d + e,$$

where y is the vector of phenotypic records, 1_n is an n-dimensional vector of ones and n is the number of hybrids, μ refers to the mean, Z_A and Z_D are n × m design matrices for the additive and dominance effects of the markers, where m refers to the number of markers. The elements of the element of Z_A is 0, 1, 2, and 0, 1 for Z_D. While $a = (a_1, a_{2,} \dots a_m)^T$ and $d = (d_1, d_{2,} \dots d_m)^T$ are vectors of length m, a_i and d_i denoted the additive and dominance effects for i-th marker. $e = (e_1, e_{2,} \dots e_n)^T$ is a vector of length n, and e_j is the residual for j-th hybrid. The indicator parameters δ_a and δ_d before the marker effect are 1 or 0 denoting whether the marker effect is included or discarded in the model, respectively.

The prior distribution for the marker effects are $a_i \sim N(0, \sigma_a^2)$, $d_i \sim N(\mu_d, \sigma_d^2)$, with μ_d having a prior distribution $N(\gamma, \sigma_d^2/m_p)$, while γ represent the anticipated size of dominance effects, and m_p is the anticipated number of markers that has contribute to the dominance effect. For the residual e, we assumed a prior distribution $e \sim N(0, \sigma_e^2)$. The prior distribution of variances are all assumed to has scaled inverse Chi-square distributions $\sigma_a^2 \sim \nu_a S_a^2 \chi_{\nu_a}^{-2}$, $\sigma_d^2 \sim \nu_d S_d^2 \chi_{\nu_d}^{-2}$, and $\sigma_e^2 \sim \nu_e S_e^2 \chi_{\nu_e}^{-2}$, respectively. The indicator parameter δ_g $(g = a\ or\ d)$, has a prior distribution $\delta_g \sim \begin{cases} 0, with\ probability\ \pi_g \\ 1, with\ probability\ 1-\pi_g \end{cases}$, while the parameter π_g $(g = a\ or\ d)$ has its own prior uniform distribution $\pi_g \sim U(0,1)$. All the unknown parameters are random draws from the full conditional densities by a special Markov Chain Monte Carlo (MCMC) algorithm called Gibbs sampling. The detail of MCMC algorithm for Bayes Cπ has been outlined in detail elsewhere (20). For the convenience of reader, we give a short summary of the Gibbs sampling with the full conditional distribution of all unknown parameters used in the algorithm.

Gibbs sampling:

Step 1: Sample the overall mean: $\mu \sim N(\frac{1_n^T(y - Z_A\delta_a a - Z_D\delta_d d)}{n}, \frac{\sigma_e^2}{n})$.

Step 2: Sample the variance of residual and additive effects from inverted-chi-square distributions $\sigma_e^2 \sim (e^T e + \nu_e S_e^2)\chi_{\nu_e+n}^{-2}$ and $\sigma_a^2 \sim (a^T a + \nu_a S_a^2)\chi_{\nu_a+m}^{-2}$, respectively.

Step 3: Sample the expected mean and variance of the dominance effects $\mu_d \sim N(\frac{1_m d + \gamma m_p}{m+m_p}, \frac{\sigma_d^2}{m+m_p})$, and $\sigma_d^2 \sim [(d-\mu_d)^T(d-\mu_d) + \nu_d S_d^2]\chi_{\nu_d+m}^{-2}$, respectively.

Step 4: Sample the additive effects from full conditional distribution $a_i \sim N(\tilde{a}_i, \frac{\sigma_e^2}{\tilde{\theta}_i})$ with $\tilde{a}_i = \frac{Z_{A_i}^T(Z_{A_i}a_i+e)}{\tilde{\theta}_i}$ and $\tilde{\theta}_i = Z_{A_i}^T Z_{A_i} + \frac{\sigma_e^2}{\sigma_a^2}$, while Z_{A_i} refers to the i-th column of Z_A. The additive effect a_i was accepted with probability $\frac{1-\pi_a}{1-\pi_a+p_{a_i}\pi_a}$. Here, p_{a_i} was the ratio of likelihood with $\delta_{a_i} = 0$ and $\delta_{a_i} = 1$.

Step 5: Sample the dominance effects from the full conditional distribution $d_i \sim N(\tilde{d}_i, \frac{\sigma_e^2}{\tilde{\eta}_i})$ with $\tilde{d}_i = \frac{Z_{D_i}^T(Z_{D_i}d_i+e)+\mu_d\frac{\sigma_e^2}{\sigma_d^2}}{\tilde{\eta}_i}$ and $\tilde{\eta}_i = Z_{D_i}^T Z_{D_i} + \frac{\sigma_e^2}{\sigma_d^2}$. Z_{D_i} refers to the i-th column of Z_D and d_i was only accepted with probability $\frac{1-\pi_d}{1-\pi_d+p_{d_i}\pi_d}$. Here, p_{d_i} was the ratio of likelihood for $\delta_{d_i} = 0$ and $\delta_{d_i} = 1$.

Step 5: Sample the π_g $(g = a\ or\ d)$ used for the next iteration with a Beta distribution $\pi_g \sim Beta(1, m - \delta_g^T\delta_g + 1, \delta_g^T\delta_g + 1)$.

The above sampling process was repeated 20,000 times, the first 2,000 results were used as burn in. The estimates of μ, a, d, δ_a, and δ_d are denoted as $\hat{\mu}$, $\hat{a}$, $\hat{d}$, $\hat{\delta}_a$, and $\hat{\delta}_d$. The performance of the hybrids in the test set was predicted as: $Y_t = 1_n\hat{\mu} + Z_{At}\hat{\delta}_a\hat{a} + Z_{Dt}\hat{\delta}_d\hat{d}$, while the Z_{At} and Z_{Dt} are design matrices for the additive and dominance effects of the hybrids from the test set.

References

43. Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423-447.
44. Butler D, Cullis B, Gilmour A, Gogel B (2009) *ASREML-R, Reference Manual, Version 3* (Queensland Department of Primary Industries and Fisheries, Brisbane).
45. Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by Simulated Annealing. *Science* 220(4598):671-680.
46. Maurer H, Melchinger A, Frisch M (2008) Population genetic simulation and data analysis with Plabsoft. *Euphytica* 161(1-2):133-139.
47. Druet T, Macleod IM, Hayes BJ (2014) Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity* 112(1):39-47.
48. Möhring J, Piepho H-P (2009) Comparison of Weighting in Two-Stage Analysis of Plant Breeding Trials. *Crop Sci.* 49(6):1977-1988.
49. Stram DO, Lee JW (1994) Variance components testing in the longitudinal mixed effects model. *Biometrics* 50(4):1171-1177.
50. Anscombe FJ, Tukey JW (1963) The Examination and Analysis of Residuals. *Technometrics* 5(2):141-160.
51. Álvarez-Castro JM, Carlborg Ö (2007) A Unified Model for Functional and Statistical Epistasis and Its Application in Quantitative Trait Loci Analysis. *Genetics* 176(2):1151-1167.
52. Pérez P, de los Campos G (2014) Genome-Wide Regression & Prediction with the BGLR Statistical Package. *Genetics* 198:483-495.