Table S1. The C matrix (after Demuth and Wade 2007) for the joint scaling analyses. Expected contributions of additive (A), dominance (D), Epistatic (AA, AD, and/or DD), cytoplasmic (Cyt), maternal additive (Ma) and maternal dominance (Md) effects are given for each of the 8 cross types from a given line pair.

Generation (dam x sire)	Mean	A	D	AA	AD	DD	Ma	Md	Cyt
P_1	1	1	0	1	0	0	1	0	1
P_2	1	-1	0	1	0	0	-1	0	-1
F_1 (P1 x P2)	1	0	1	0	0	1	1	0	1
rF_1 (P2 x P1)	1	0	1	0	0	1	-1	0	-1
F_2 (F_1 x F_1)	1	0	0.5	0	0	0.25	0	1	1
rF_2 (rF_1 x rF_1)	1	0	0.5	0	0	0.25	0	1	-1
BC_1P1 (P1 x pooled F_1 and rF_1)	1	0.5	0.5	0.25	0.25	0.25	1	0	1
BC_1P2 (P2 x pooled F_1 and rF_1)	1	-0.5	0.5	0.25	0.25	0.25	-1	0	-1

Table S2. Goodness of fit scores for all joint scaling models tested for fitness (= total seed number) for two line pairs from each population set (C↔R = Castelnuovo, Italy and Rödåsen, Sweden; B↔S = Bolsena, Italy and Skuleberget, Sweden). Parameter abbreviations as in Table S1. The simplest model(s) that are sufficient to explain the data are highlighted in grey. If multiple models with the same number of parameters were sufficient to explain the data, models were determined to be significantly better than others by the criterion of at least a 3.84 lower X^2, (i.e. greater likelihood). The best models are marked with an *. The best model for B1↔S1 was chosen from the four equally simple and sufficient models with similar model fit because this model contained the most individually significant parameters (see Table S5).

Model	DF	Goodness of fit (X^2)			
		C1↔R1	C2↔R2	B1↔S1	B2↔S2
Mean+A	6	14.5	28.5	32.3	171.5
Mean+A+Cyt	5	12.8	20.8	29.4	134.7
Mean+A+D	5	6.8*	9.2*	30.4	48.2
Mean+A+D+Cyt	4	4.7	2.5	28.3	32.9
Mean+A+D+Ma	4	5.5	7.9	21.8	4.7*
Mean+A+D+AA	4	5.8	8.6	15.9	42.9
Mean+A+D+AD	4	6.8	9.1	30.4	34.3
Mean+A+D+DD	4	3.5	9.2	17.4	33.2
Mean+A+D+Ma+Cyt	3	4.6	2.4	8.7	3.9
Mean+A+D+Ma+Md	3	5.5	6.1	15.2	3.8
Mean+A+D+AA+Cyt	3	3.8	2.0	13.2	31.1
Mean+A+D+AD+Cyt	3	4.4	2.0	28.1	27.3
Mean+A+D+DD+Cyt	3	1.7	2.5	12.5	25.4
Mean+A+D+AA+Ma	3	4.6	7.2	13.6	4.7
Mean+A+D+AD+Ma	3	4.5	7.4	15.0	4.7
Mean+A+D+DD+Ma	3	2.6	7.8	15.9	3.7
Mean+A+D+AA+AD	3	5.8	8.5	15.7	31.5
Mean+A+D+AA+DD	3	2.7	6.9	15.6	27.7
Mean+A+D+AD+DD	3	3.5	9.1	16.4	26.6
Mean+A+D+Ma+Md+Cyt	2	4.6	0.9	5.5	3.1
Mean+A+D+AA+Ma+Cyt	2	3.7	1.9	4.1	3.9
Mean+A+D+AD+Ma+Cyt	2	3.6	2.0	2.8	3.9
Mean+A+D+DD+Ma+Cyt	2	1.6	2.3	4.7*	2.9
Mean+A+D+AA+Ma+Md	2	1.8	5.6	13.6	1.1
Mean+A+D+AD+Ma+Md	2	4.4	5.6	10.9	3.8
Mean+A+D+DD+Ma+Md	2	1.8	5.6	13.6	1.1
Mean+A+D+AA+AD+Cyt	2	3.5	1.5	13.1	26.0
Mean+A+D+AA+DD+Cyt	2	0.9	0.5	11.9	20.6
Mean+A+D+AD+DD+Cyt	2	1.3	1.9	12.5	22.2
Mean+A+D+AA+AD+Ma	2	3.8	6.8	9.6	4.7
Mean+A+D+AA+DD+Ma	2	1.8	5.6	13.6	1.1
Mean+A+D+AD+DD+Ma	2	1.8	7.4	10.9	3.7
Mean+A+D+AA+AD+DD	2	2.6	6.9	15.1	23.6

Table S3. Goodness of fit scores for all joint scaling models tested for number of fruits, abbreviations as in Table S2. The simplest model(s) that are sufficient to explain the data are highlighted in grey. If multiple models with the same number of parameters were sufficient to explain the data, models were determined to be significantly better than others by the criterion of at least a 3.84 lower X^2, (i.e. greater likelihood). The best models are marked with an *. The best model for B2↔S2 was chosen from the three equally simple and sufficient models with similar model fit because all model parameters were individually significant for this model only (see Table S5).

Model	DF	Goodness of fit (X^2)			
		C1↔R1	C2↔R2	B1↔S1	B2↔S2
Mean+A	6	23.2	23.8	44.2	46.1
Mean+A+Cyt	5	21.5	12.9	33.7	28.4
Mean+A+D	5	3.7*	13.1	39.3	34.2
Mean+A+D+Cyt	4	2.6	1.5*	30.9	16.2
Mean+A+D+Ma	4	3.4	8.5	36.2	16.4
Mean+A+D+AA	4	1.4	12.4	23.0	22.4
Mean+A+D+AD	4	3.4	9.7	39.3	31.1
Mean+A+D+DD	4	2.9	12.0	33.4	18.1
Mean+A+D+Ma+Cyt	3	2.6	1.4	14.8	11.6
Mean+A+D+Ma+Md	3	0.9	8.4	16.8	12.2
Mean+A+D+AA+Cyt	3	0.4	1.2	13.4	8.6
Mean+A+D+AD+Cyt	3	2.6	1.0	29.8	16.1
Mean+A+D+DD+Cyt	3	1.9	0.7	23.2	4.9
Mean+A+D+AA+Ma	3	1.2	7.9	22.6	6.7*
Mean+A+D+AD+Ma	3	3.3	8.0	35.0	15.3
Mean+A+D+DD+Ma	3	2.7	7.5	32.3	5.0
Mean+A+D+AA+AD	3	1.2	9.0	23.0	19.6
Mean+A+D+AA+DD	3	1.0	12.0	17.6	18.1
Mean+A+D+AD+DD	3	2.7	8.5	33.4	16.7
Mean+A+D+Ma+Md+Cyt	2	0.1	1.4	0.8*	8.8
Mean+A+D+AA+Ma+Cyt	2	0.4	1.2	5.3	3.7
Mean+A+D+AD+Ma+Cyt	2	2.5	0.9	13.7	10.6
Mean+A+D+DD+Ma+Cyt	2	1.9	0.7	12.2	1.5
Mean+A+D+AA+Ma+Md	2	0.8	7.5	16.6	4.9
Mean+A+D+AD+Ma+Md	2	0.9	7.9	16.3	11.6
Mean+A+D+DD+Ma+Md	2	0.8	7.5	16.6	4.9
Mean+A+D+AA+AD+Cyt	2	0.4	0.7	12.4	8.4
Mean+A+D+AA+DD+Cyt	2	0.1	0.7	9.7	4.9
Mean+A+D+AD+DD+Cyt	2	1.9	0.1	22.0	4.9
Mean+A+D+AA+AD+Ma	2	1.2	7.3	22.4	5.8
Mean+A+D+AA+DD+Ma	2	0.8	7.5	16.6	4.9
Mean+A+D+AD+DD+Ma	2	2.6	6.9	31.8	3.5
Mean+A+D+AA+AD+DD	2	0.8	8.5	17.6	16.7

Table S4. Goodness of fit scores for all joint scaling models tested for seed number per fruit, abbreviations as in Table S2. The simplest model(s) that are sufficient to explain the data are highlighted in grey. If multiple models with the same number of parameters were sufficient to explain the data, models were determined to be significantly better than others by the criterion of at least a 3.84 lower X^2 ,(i.e. greater likelihood). The best models are marked with an *. For B1↔S1, the best model was chosen from the four equally simple and sufficient models with similar model fit because this model contained the most individually significant parameters (see Table S5). For B2↔S2, both of these models were equivalently simple and sufficient and for both all model terms were individually significant (see Table S5), so the model with the slightly better fit was chosen as the best model.

Model	DF	Goodness of fit (X^2)			
		C1↔R1	C2↔R2	B1↔S1	B2↔S2
Mean+A	6	15.5	14.9	26.1	112.4
Mean+A+Cyt	5	11.2	14.7	15.6	111.9
Mean+A+D	5	14.6	9.7*	26.0	31.4
Mean+A+D+Cyt	4	9.4*	9.6	14.5	30.9
Mean+A+D+Ma	4	13.5	9.7	12.0	14.4
Mean+A+D+AA	4	13.8	9.7	25.7	29.1
Mean+A+D+AD	4	14.2	7.7	24.8	27.6
Mean+A+D+DD	4	14.2	7.7	21.7	31.4
Mean+A+D+Ma+Cyt	3	9.3	9.4	10.1	7.4
Mean+A+D+Ma+Md	3	9.7	8.4	8.0	6.9*
Mean+A+D+AA+Cyt	3	8.7	9.6	14.4	28.2
Mean+A+D+AD+Cyt	3	9.4	6.9	14.5	27.6
Mean+A+D+DD+Cyt	3	8.6	7.7	13.5	30.8
Mean+A+D+AA+Ma	3	13.0	9.7	10.7	9.6
Mean+A+D+AD+Ma	3	13.5	7.0	9.8	14.2
Mean+A+D+DD+Ma	3	12.5	7.7	12.0	13.4
Mean+A+D+AA+AD	3	13.3	7.7	24.6	24.4
Mean+A+D+AA+DD	3	6.8	4.1	13.5	23.9
Mean+A+D+AD+DD	3	13.8	5.2	21.4	27.3
Mean+A+D+Ma+Md+Cyt	2	4.5	8.2	6.1	0.4
Mean+A+D+AA+Ma+Cyt	2	8.6	9.4	8.9	2.9
Mean+A+D+AD+Ma+Cyt	2	9.3	6.7	7.9	7.1
Mean+A+D+DD+Ma+Cyt	2	8.6	7.4	10.1	6.4
Mean+A+D+AA+Ma+Md	2	5.4	3.9	5.9	6.8
Mean+A+D+AD+Ma+Md	2	9.6	5.3	4.5	6.6
Mean+A+D+DD+Ma+Md	2	5.4	3.9	5.9	6.8
Mean+A+D+AA+AD+Cyt	2	8.7	6.9	14.4	24.4
Mean+A+D+AA+DD+Cyt	2	0.2	4.1	6.5	23.3
Mean+A+D+AD+DD+Cyt	2	8.6	4.6	13.3	27.3
Mean+A+D+AA+AD+Ma	2	13.0	7.0	8.1	9.4
Mean+A+D+AA+DD+Ma	2	5.4	3.9	5.9*	6.8
Mean+A+D+AD+DD+Ma	2	12.4	4.8	9.8	13.2
Mean+A+D+AA+AD+DD	2	6.6	0.7	13.5	20.6

Table S5. Alternative results for cases where multiple models with similar model fit were sufficient to explain the data, abbreviations as in Table S2. Table entries are *t* values and significance levels with 7 degrees of freedom are: *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$, † $0.05 < P < 0.10$, ^{ns} $P > 0.10$. Blank cells in a column indicate that the corresponding parameter was not part of the model.

Param.	Number of fruits		Seed number per fruit				Multiplicative fitness		
	B2↔S2	B2↔S2	B1↔S1	B1↔S1	B1↔S1	B2↔S2	B1↔S1	B1↔S1	B1↔S1
Mean	38.47***	38.51***	6.35***	37.31***	35.02***	45.59***	31.70***	12.06***	32.30***
A	8.97***	7.66***	-2.18†	-3.60**	-2.18†	3.01*	3.11*	3.41*	2.05†
D	1.63^{ns}	1.69^{ns}	0.93^{ns}	-2.33†	1.86^{ns}	-9.02***	-2.04†	0.71^{ns}	2.06†
AA			1.43^{ns}					2.14†	
AD				1.88^{ns}					-2.41*
DD	-3.36*	-3.37*			1.43^{ns}				
Ma		-3.61**	2.76*	4.36**	2.76*	-4.85**	3.75**	3.01*	5.02**
Md			-2.19†	-2.31†	2.46*		-1.77 ^{ns}		
Cyt	-3.63**					2.66*	3.10*	-3.09*	-3.48*

Table S6. Frequency of progeny exhibiting stunted phenotype (%) by cross type, abbreviations as in Table S2.

		Population pair			
Generation (dam x sire)	n	C1↔R1	C2↔R2	B1↔S1	B2↔S2
P_1	20	0	0	0	0
P_2	20	0	0	0	0
F_1 (P1 x P2)	20	0	0	0	100
rF_1 (P2 x P1)	20	0	0	45	45
F_2 (F_1 x F_1)	80	0	0	0	28
rF_2 (rF_1 x rF_1)	80	0	0	17	15
BC_1P1 (P1 x pooled F_1 and rF_1)	80	0	0	8	36
BC_1P2 (P2 x pooled F_1 and rF_1)	80	0	0	15	8

Table S7. Means and variances used in joint scaling analyses of fitness and fitness components, abbreviations as in Table S2.

		Number of fruits		Seed number per fruit		Fitness (number of fruits × seed number per fruit)	
Line pair	Cross type	Mean	Var.	Mean	Var.	Mean	Var.
C1↔R1	P1	116.79	5.99	46.53	1.69	5409.16	21361.95
	P2	73.11	12.02	46.61	1.20	3385.78	24140.34
	F_1	109.53	16.34	42.89	4.71	4610.11	43067.70
	rF_1	108.26	16.63	45.83	1.28	5036.94	61157.68
	F_2	106.46	4.35	43.90	0.69	4655.23	14607.44
	rF_2	103.83	4.36	46.39	0.51	4819.61	15585.12
	BC_1P1	114.15	3.32	47.09	0.71	5361.19	13796.17
	BC_1P2	90.85	3.20	47.78	0.58	4327.58	10541.67
C2↔R2	P1	125.50	11.11	44.35	1.76	5555.10	43237.02
	P2	56.15	11.54	49.10	2.06	2751.65	32711.46
	F_1	99.15	11.18	48.90	2.14	4863.00	58501.53
	rF_1	105.40	13.40	50.20	1.50	5296.45	52984.20
	F_2	94.26	4.57	47.56	0.65	4492.09	15953.16
	rF_2	101.71	3.44	48.14	0.65	4887.75	14243.67
	BC_1P1	112.73	2.69	46.34	0.44	5210.29	9732.15
	BC_1P2	84.46	3.34	46.01	0.74	3859.05	9892.64
B1↔S1	P1	148.00	20.51	42.60	4.82	6191.65	74988.34
	P2	82.70	14.90	41.80	1.21	3447.60	33337.97
	F_1	109.21	16.31	43.37	1.34	4720.84	37596.90
	rF_1	98.60	54.80	32.95	7.99	3537.35	216886.24
	F_2	91.03	8.24	42.80	1.45	3783.40	19841.04
	rF_2	109.11	12.23	40.49	1.47	4523.23	37003.57
	BC_1P1	127.72	12.20	37.80	2.90	4817.01	50874.55
	BC_1P2	95.56	13.15	37.95	1.35	3667.14	24535.51
B2↔S2	P1	139.80	20.25	47.40	4.41	6594.35	118200.27
	P2	89.42	15.41	46.05	1.61	4101.32	41182.22
	F_1	76.05	34.57	25.63	3.19	1997.95	36664.99
	rF_1	108.60	45.47	36.20	4.71	4007.55	140035.16
	F_2	109.19	16.59	38.40	2.16	4305.91	49349.14
	rF_2	118.86	10.70	33.21	1.97	4044.80	44575.59
	BC_1P1	123.19	18.57	37.55	1.72	4792.28	68302.76
	BC_1P2	105.26	9.31	40.73	1.28	4293.51	26287.50

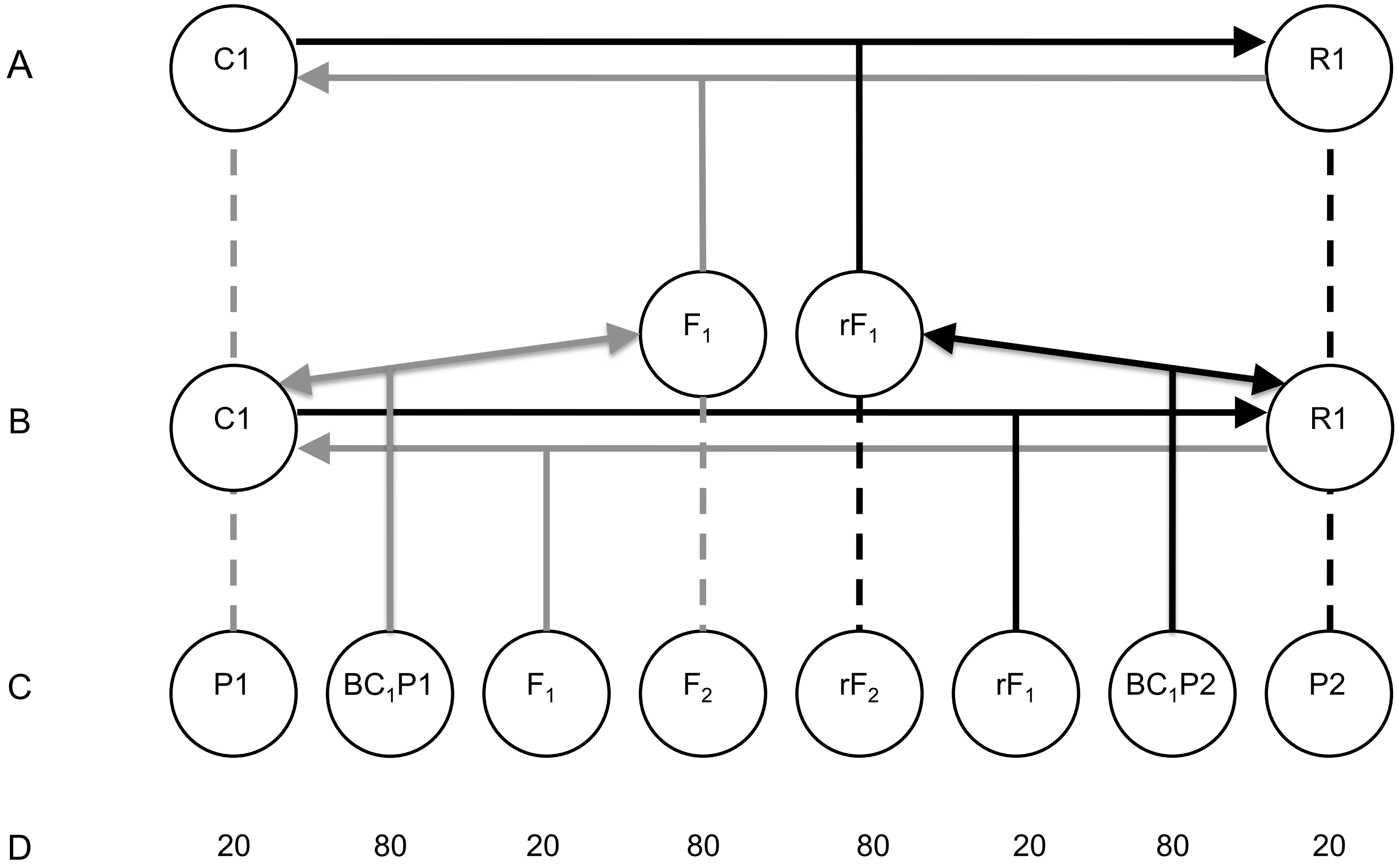

Figure S1. Line cross design (one example line pair shown). Arrows indicate direction of the cross, with vertical lines to the progeny. Dashed vertical lines indicate progeny derived from selfing. Black and grey lines represent R1 and C1 maternal cytotypes, respectively. A. Parental lines were selfed, and initial crosses performed to make reciprocal F_1 lines. B. Parental lines and reciprocal F_2 were generated by selfing, and reciprocal F_1 and both backcrosses were generated by crossing. C. Lines used to estimate fitness and fitness components for joint scaling analyses. D. Sample size (number of plants) for each line in C.