Supporting Information

Guo et al. 10.1073/pnas.0808012106

SI Text: How much variation will selective sweeps remove in *C. rubella*?

We first make a rough estimate of the population recombination rate in *C. rubella*. We cannot directly use the shared polymorphism between *C. rubella* and *C. grandiflora*, since for such variation we do not know whether recombination events have occurred in *C. rubella* or in the ancestral species of *C. rubella* and *C. grandiflora*. We have therefore collected only the segregating sites private to *C. rubella* around the S-locus (see Fig. S4). From this dataset we can estimate (assuming the infinite sites model and using the 4-gamete test) at least 4 recombination events in $<$100 kb to the S-locus on either side (and at least 1 recombination event in the SRK). There is also evidence for recombination events between this core 200-kb region and the genes at distances of 500–1,000 kb on either side. One can translate the minimal number of recombination events into an (arguably very rough) estimator of the scaled recombination rate, rho, which is $\approx$20 for the 200-kb region (1). If we assume an effective population size of rubella of $\approx$10,000, then this corresponds to an estimated rate of 0.5 cM/Mb.

Thus, we would not expect a selective sweep of $s = 10\%$ to sweep variation from a region much more than $r = 0.0005$ (corresponding to $r/s = 0.05$) (see ref. 2). For the S-locus region with the above estimated rate, this corresponds to 0.5 cM or 1 Mb. Note that this is a very strong selective sweep (a sweep of 1% would affect only 100 kb) and that our estimate of the recombination rate is likely to be an underestimate, since not all variation in the region has been surveyed. If recombination rates are similar or higher at other locations across the *C. rubella* genome, sweeps would be expected to affect similar or smaller regions. Therefore, we would need to invoke multiple, localized sweeps to explain the reduction of diversity at several additional loci unlinked to each other and the S locus. We argue that a more parsimonious explanation is that a bottleneck caused the reduction in diversity at both the S locus and unlinked loci.

1. Hudson RR (1987) Estimating the recombination parameter of a finite population model without selection. *Genet Res* 50:245–250.
2. Santiago E, Caballero A (2005) Variation after a selective sweep in a subdivided population. *Genetics* 169:475–483.

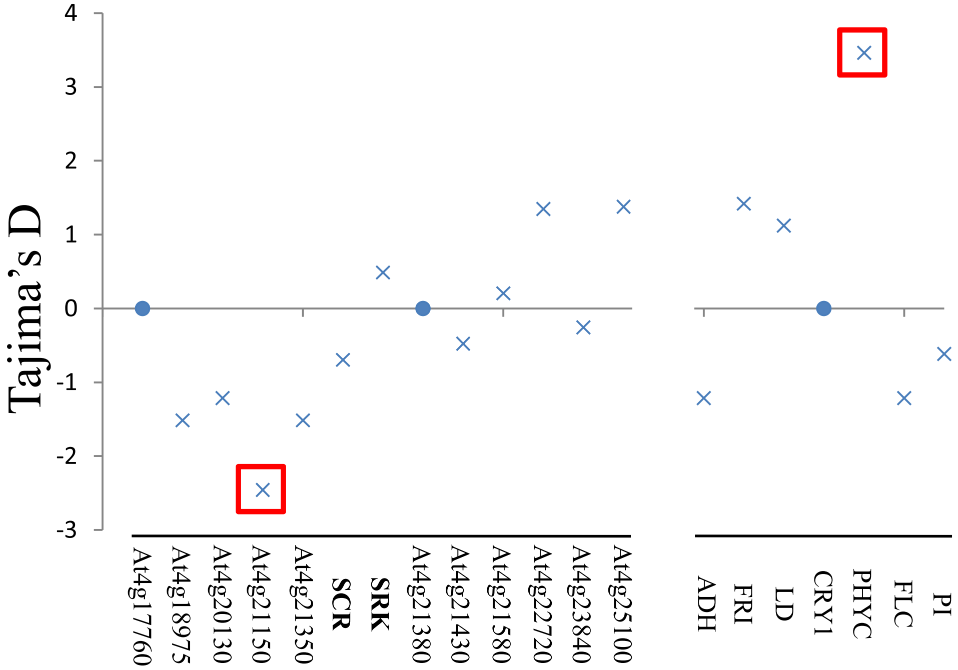


Fig. S1. Tajima's D values in *C. rubella* for 20 gene fragments. Three fragments had no variation (indicated with blue dots), and the 2 with significant values ($P < 0.05$) are boxed in red.

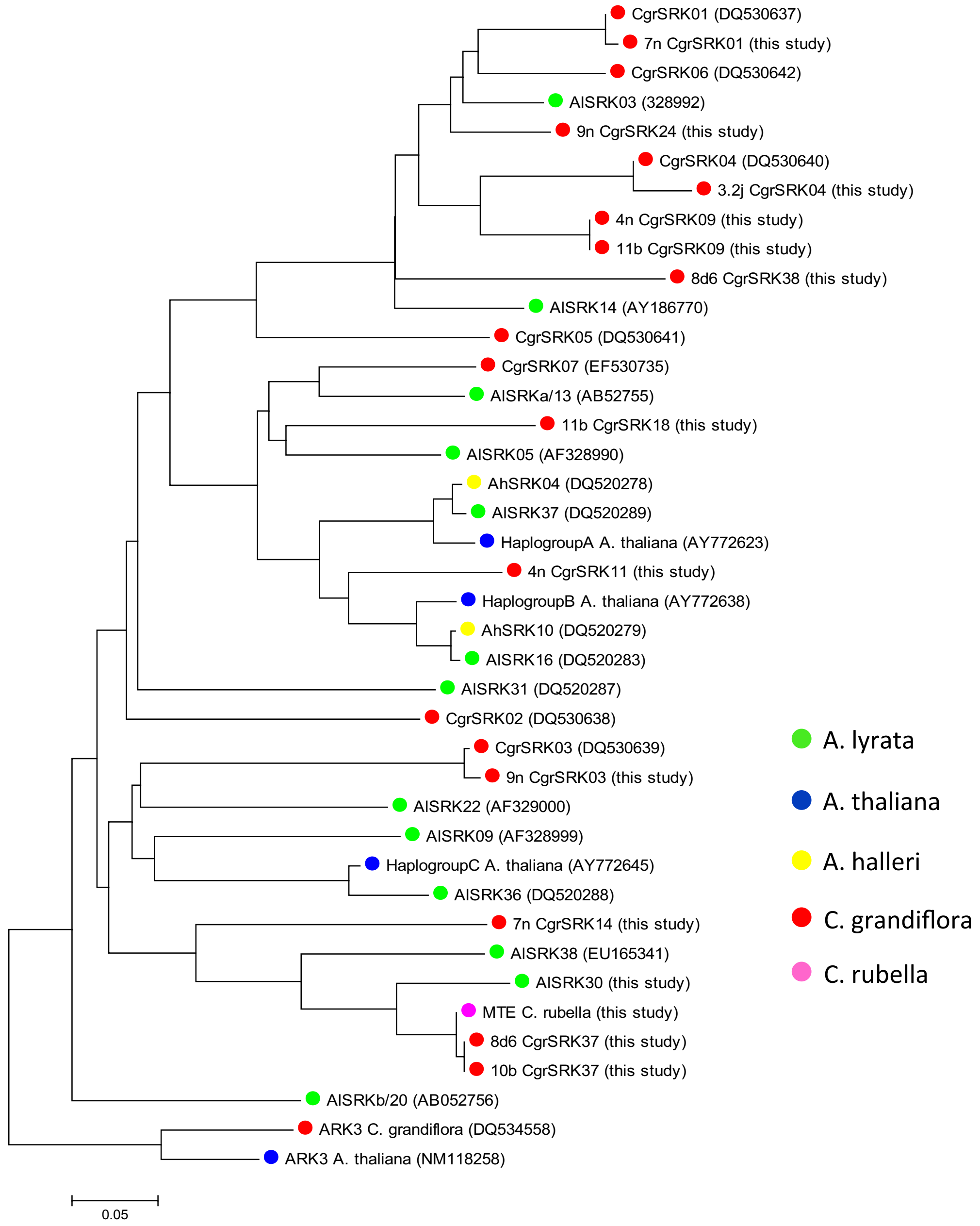
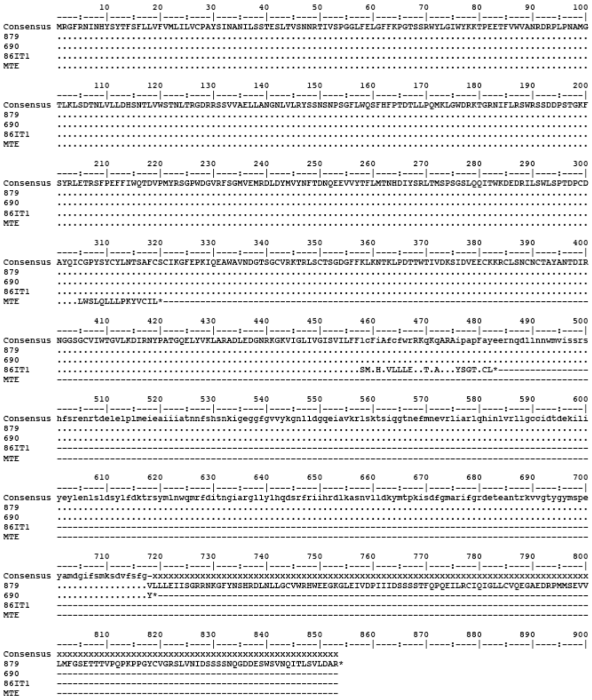
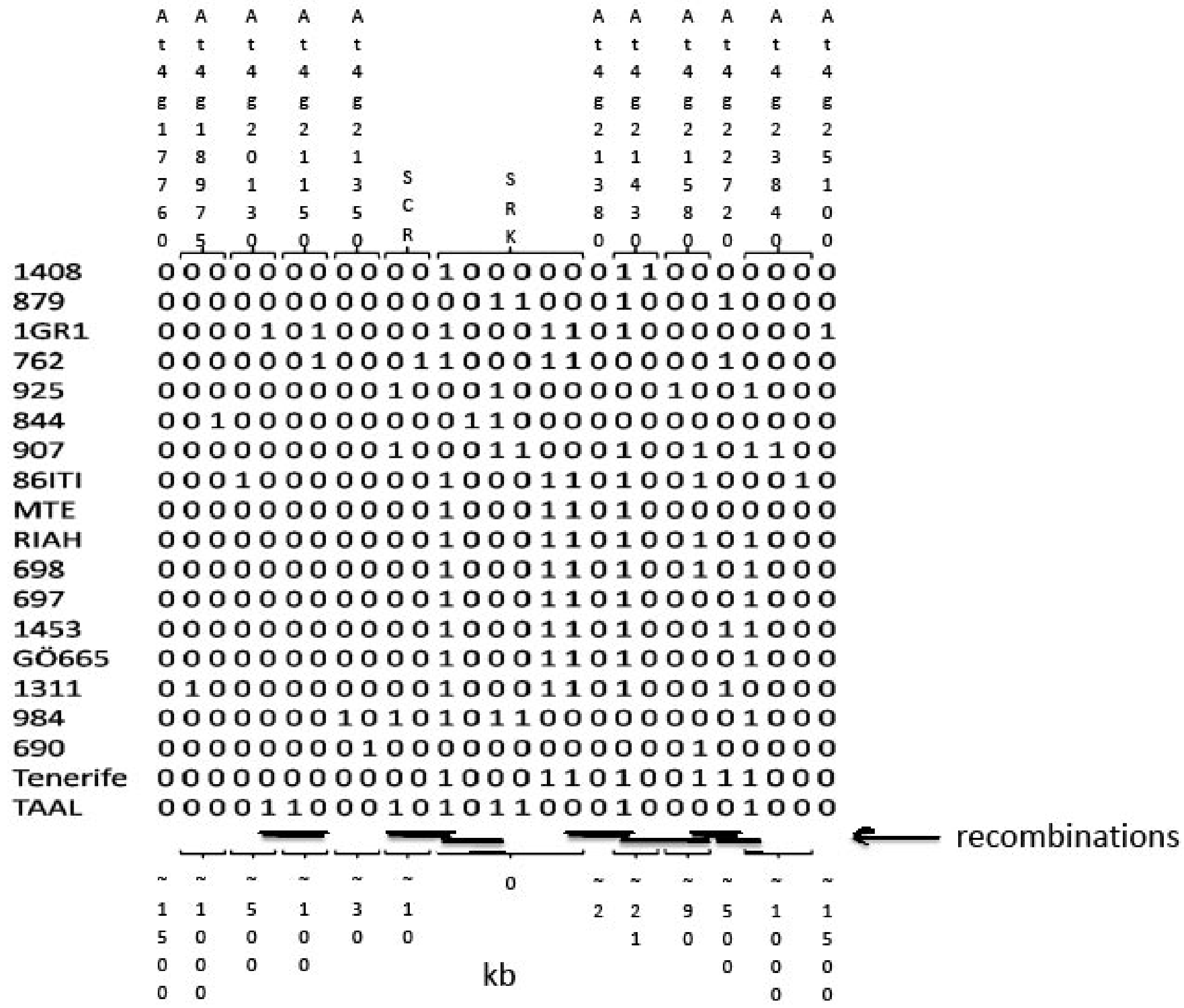


Fig. S2. Phylogenetic tree of *SRK* alleles in *Arabidopsis* and *Capsella*. GenBank accession numbers in parentheses.



```
                10        20        30        40        50        60        70        80        90       100
          ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|
Consensus MRGFRNINHYSYTFSFLLVFVMLILVCPAYSINANILSSTESLTVSNNRTIVSPGGLFELGFFKPGTSSRWYLGIWYKKTPEETFVWVANRDRPLPNAMG
879       ....................................................................................................
690       ....................................................................................................
86IT1     ....................................................................................................
MTE       ....................................................................................................

               110       120       130       140       150       160       170       180       190       200
          ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|
Consensus TLKLSDTNLVLLDHSNTLVWSTNLTRGDRRSSVVAELLANGNLVLRYSSNSNPSGFLWQSFHFPTDTLLPQMKLGWDRKTGRNIFLRSWRSSDDPSTGKF
879       ....................................................................................................
690       ....................................................................................................
86IT1     ....................................................................................................
MTE       ....................................................................................................

               210       220       230       240       250       260       270       280       290       300
          ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|
Consensus SYRLETRSFPEFFIWQTDVPMYRSGPWDGVRFSGMVEMRDLDYMVYNFTDNQEEVVYTFLMTNHDIYSRLTMSPSGSLQQITWKDEDRILSWLSPTDPCD
879       ....................................................................................................
690       ....................................................................................................
86IT1     ....................................................................................................
MTE       ....................................................................................................

               310       320       330       340       350       360       370       380       390       400
          ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|
Consensus AYQICGPYSYCYLNTSAFCSCIKGFEPKIQEAWAVNDGTSGCVRKTRLSCTSGDGFFKLKNTKLPDTTWTIVDKSIDVEECKKRCLSNCNCTAYANTDIR
879       ....................................................................................................
690       ....................................................................................................
86IT1     ....................................................................................................
MTE       ....LWSLQLLLPKYVCIL*--------------------------------------------------------------------------------

               410       420       430       440       450       460       470       480       490       500
          ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|
Consensus NGGSGCVIWTGVLKDIRNYPATGQELYVKLARADLEDGNRKGKVIGLIVGISVILFFlcFiAfcfwrRKqKqARAipapFayeernqdllnnwmvissrs
879       ....................................................................................................
690       ....................................................................................................
86IT1     .........................................................SM.H.VLLLE..T.A...YSGT.CL*-----------------
MTE       ----------------------------------------------------------------------------------------------------

               510       520       530       540       550       560       570       580       590       600
          ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|
Consensus hfsrenrtdelelplmeieaiiiatnnfshsnkigeggfgvvykgnlldgqeiavkrlsktsiqgtnefmnevrliarlqhinlvrllgccidtdekili
879       ....................................................................................................
690       ....................................................................................................
86IT1     ----------------------------------------------------------------------------------------------------
MTE       ----------------------------------------------------------------------------------------------------

               610       620       630       640       650       660       670       680       690       700
          ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|
Consensus yeylenlsldsylfdktrsymlnwqmrfditngiargllylhqdsrfriihrdlkasnvlldkymtpkisdfgmarifgrdeteantrkvvgtygymspe
879       ....................................................................................................
690       ....................................................................................................
86IT1     ----------------------------------------------------------------------------------------------------
MTE       ----------------------------------------------------------------------------------------------------

               710       720       730       740       750       760       770       780       790       800
          ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|
Consensus yamdgifsmksdvfsfg-xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
879       .................VLLLEIISGRRNKGFYNSHRDLNLLGCVWRHWEEGKGLEIVDPIIIDSSSSTFQPQEILRCIQIGLLCVQEGAEDRPMMSEVV
690       .................Y*---------------------------------------------------------------------------------
86IT1     ----------------------------------------------------------------------------------------------------
MTE       ----------------------------------------------------------------------------------------------------

               810       820       830       840       850       860       870       880       890       900
          ----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|----:----|
Consensus xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
879       LMFGSETTTVPQPKPPGYCVGRSLVNIDSSSSNQGDDESWSVNQITLSVLDAR*
690       -----------------------------------------------------
86IT1     -----------------------------------------------------
MTE       -----------------------------------------------------
```

Fig. S3. Alignment of SRK amino acid sequences from *C. rubella*. Dots indicate identical amino acid as consensus, asterisk indicate stop codons.



1408	000000000001000000110000000
879	000000000000011000100010000
1GR1	000010100001000110100000001
762	000000100011000110000010000
925	000000000100010000001001000
844	001000000000110000000000000
907	000000000100011000100101100
86ITI	000100000001000110100100010
MTE	000000000001000110100000000
RIAH	000000000001000110100101000
698	000000000001000110100101000
697	000000000001000110100001000
1453	000000000001000110100011000
GÖ665	000000000001000110100001000
1311	010000000001000110100010000
984	000000010101011000000001000
690	000000001000000000000100000
Tenerife	000000000001000110100111000
TAAL	000011000101011000100001000

Fig. S4. Segregating sites private to *C. rubella*. 0 denotes the ancestral variant, 1 the derived variant. On top are the locus names (for genes with most variation shared with *C. grandiflora*), on bottom is the distance to the S-locus in kilobases, and to the left the accession names. Black horizontal bars indicate a recombination event as inferred by the 4-gamete test (DNAsp).

Table S1. Samples studied

Species	Accessions	Origin	Latitude/Longitude
C. grandiflora	3	Greece	+39.40/+19.46
	4	Greece	+39.45/+19.52
	7	Greece	+38.46/+20.59
	8	Greece	+38.52/+20.59
	9	Greece	+39.53/+20.44
	10	Greece	+39.52/+20.46
	11	Greece	+39.52/+20.42
C. rubella	690	Spain	+36.15/−5.58
	697	Itaky	?
	698	Italy	?
	762	Greece	+37.58/+23.43
	844	Greece	?
	879	Greece	+35.29/+24.42
	907	Greece	+39.40/+19.48
	925	Greece	+39.40/+20.51
	984	Spain	+39.30/+3.00
	1,204	Spain (Tenerife)	+28.19/−16.34
	1,207	Spain (Tenerife)	+28.19/−16.34
	1,208	Spain (Tenerife)	+28.19/−16.34
	1,209	Spain (Tenerife)	+28.19/−16.34
	1,215	Spain (Tenerife)	+28.19/−16.19
	1,311	France	+42.53/−0.06
	1,377	Argintina	−34.40/−58.30
	1,408	Greece	+35.29/+24.42
	1,453	Italy	+43.28/+11.02
	1,482	Australia	−31,56/+115.50
	MTE	Italy	?
	RIAH	Italy	?
	86IT1	Sorrento, Italy	+40,62/+14,37
	GÖ665 (Accession 665)	Botanical Garden of Göttingen	?
	1GR1	Samos, Greece	+37,78/+26,83
	TAAL-1	Taguemont, Algeria	?

Table S2. Genes sequenced and PCR primers used. (a) the primers are designed based on *A. thaliana* sequence; (b) designed based on *C. rubella* BAC sequence. Bold primers were used only for sequencing

No.	Gene	Primers	Fragment size, bp	Distance from SRK (*A. thaliana*—*C. rubella*, kb)
1	At4 g17760*	G16211 ATT TGC CAG CTC GAT AAT GTT C G16212 GGT ATC ACA ATG AAA GGC AAG G	978	1,496–?
2	At4 g18975*	G16020 CTG GGA AAT TGG ACA GAG GTT A G16021 TCT TCT TGG TTG AGT TCC CTG A	990	990–?
3	At4 g20130*	G15788 TCA CTT CTG GCG GCT CTA TG G15789 TGC GAG AAT CTC CAG AGA AC	958	504–?
4	At4 g21150*	G15708 GCA CCA GCT CTT AGT GTG AAG C G15709 GCT CGA CCT GTG ACT CAT GTT T	867	100–?
5	At4 g21350 (U-box)[†]	*C.grandiflora* G14552 GTC ATC CTC CAA TCG GGT CAT ACT T G14553 CTC ATT ATC TTC CAA ACC AAA GCA T *C. rubella* G18406 TTT CGG GTT CTC TGC TCT GT G18407 TTC CAT TTT CAG CAC CAA ATC	943/1,283	27–32
6	SCR[†]	G15065 CGT CAA TGT CGT TCA TTT GGT T G15066 TGT TGA AAA ATA GCC ACG GTT C	857	2–14
7	At4 g21370 (SRK)[†]	*C.grandiflora*–General primers for S-domain 13 seq2F GATGGAMTCCGRTTTAKYGGCAT SLGR ATCTGACATAAAGATCTTGACC ClassB1000-F1 AATGGCAATTTCGTGCT ClassB600-R1 CCCGACCCACCATCTTGG ClassA3.1000-F1 GGCAATTTCGTGCTAAGG SLGR ATCTGACATAAAGATCTTGACC *C.grandiflora*–*C. rubella* allele primers SRKrub5endF–TCCTGCGTACTCGATCAATG SRKrub3endR–AGTACAGCTCAACCGCGTC SRKrub3endF–CGAGGATCGGATTCTGTC SRKrubex2R–CGGAATAGCTCTTGCTTGC SRKrubex2F–GTAATAGGTCTGATCGTAGG SRKrubex5R–TTCCAAAGATCCTTGCCATC SRKrubex5F–TGCGATTTGATATCACCAATG SRKrubex7R–AAGGACCGAGAGGGTG *C. rubella* Fragment1 G11587 CTG CTG GGA AGT TGA AAA TAA ATA A G11588 GAA TAG CTC TTG CTT GCT TCT GTT T Fragment2 G11589 CCA CAG GTC AAG AAC TTT ATG TCA A G11590 TAC CTA CCA TCA AGA GCG ATA AAC T G11793 TCA GTG GTT GCA GAG CTT CTT G G11794 ACC AGT CCG GCA GTC ACA ATA TG G11795 TTG GTC AAC AAC ACT GTC ATA TTC G11796 TCT CTT GTG GTT GGA ATG TTG ATG G11941 GCC AAT TTA GCA TAT AGC TTC GAG G12031 TCA GCG GAA GTT CTA GTT CGT CTG T G12182 ATC GCC AGC AAA ATA TAC AAT ATA A G12496 CTT GGA AAA TTT AAG CCT CGA TTC G12497 ATG TGA GGT GCT GAA AAG TGA CTC	600, 800, 900, 970, 1,290, 960, 850, 4,103, 2,263, 2,673	0–0
8	At4 g21380 (ARK3)[†]	G15576 GGT TGC AAA CAG AGA CCA TCC T G15671 TCC ATA TCA CAC AAC CTG ACC C	930	2–2
9	At4 g21430[†]	G15672 CAA TCT AGT AGT AGT GAC AGT C G15673 ATT TAC TCT CCC TCC CCG TTT C	969	21–22
10	At4 g21580*	G15710 TGC TCT TCT TTC TGG AGG AGG T G15711 CGA CTT CCC TTA CAA CAA CTG C	897	89–?
11	At4 g22720*	G15790 GCA CAT CAC CAT CTT GAC CA G15791 GCG AAA TCG TTG TGT AAA GG	989	554–?
12	At4 g23840*	G16022 GTT ACG GAT GCT GGT ATG AAG C G16023 CAC ATG CTA CAA ACC CAA CAG A	999	1013–?
13	At4 g25100*	G16213 GGA AAA CAT CAC AGA GCT TAC G G16214 CAA GTC TGG CAC TTA CAG CTT C	912	1510–?
14	At4 g00650 (FRI)*	G-9905 TGA AGG AGG ATT AGC TGT GGC G-9906 TCG TCT CTT TGA CTA GGA AAG	955	11112.5 kb–?
15	At1 g77120*	G16230 GTG ATG GAG GAA GTG GAG GTT G16456 TTG ATC TCT GAG AAT GGC ACT G G16231 TGA TCT TCT CAG GAT CAA CAC C	1405	-
16	AT5G10140 (FLC)*	G10287 TCT TAA AGC CTT GGT AAT ACA AAC A G11096 AAG ATC ACC ATG TTC AGC AAT CA	915	-
17	At5 g35840 (PHYC)*	G-9910 TGG TGT ATG CTT CAT TGG ACA G-10442 TGC ACT TTA CGT TCT ATG CTG A	949	-
18	matK*	Fragment1 G-20636 GTT GCC CGG GAT TCG AA G-20639 CCT TTC ATA CAT TAT GTT AG G-20637 ATT AGG GCA TCC CAT TAG TA G-20638 TTA ACA CAA CAA AGA CGA AG Fragment2 G-20641 GTA ATA TTT CCA TTT ATG CAT C G-20642 GAG TAC TCG GCT TTT AAG TG G-20640 GTA CTT GCT CAT GAT CAT GG G-20643 GAT TCT GTT CAT ACA TTC GCA	2571, 1,004, 1,612,	-

Table S3. Summary of diversity parameters

Locus	Distance from SRK in *A. thaliana* and *C. rubella* (bp)	Alleles sequenced (rubella/ grandiflora)		Exons only π	Introns only π (bp)	Tajima's D	Singletons	Informative sites	Fixed differences	Shared differences	Private polymorphisms- rubella	Private polymorphisms– grandiflora		
				Total (bp)	Syn	Non-syn								
At4 g17760	1496	25/14	rubella	0 (493)	0	0	0 (456)	—	0	0	5	0	0	66
grandiflora	0.02046 (570)	0.07365	0.00507	0.07336 (477)	1.16	5	65					0	0	66
At4 g18975	990-?	25/14	rubella	0 (460)	0	0	0.00042 (384)	−1.51	2	0	3	0	2	32
grandiflora	0.0056 (550)	0.02285	0.00091	0.01856 (445)	−1.13	29	9					0	2	32
S2 (At4 g20130)	504-?	25/14	rubella	0.00019 (423)	0.00093	0	0.00032 (482)	−1.21	1	1	2	0	1	43
grandiflora	0.01385 (443)	0.06006	0.00271	0.02583 (564)	0.59	10	36					0	1	43
At4 g21150	100-?	25/14	rubella	0.00124 (375)	0.00185	0.00108	0.00278 (409)	−2.46***	18	1	0	13	6	34
grandiflora	0.00754 (420)	0.02218	0.00336	0.01132 (438)	−1.06	18	14					13	6	34
Ubox (At4 g21350)	27–32	25/14	rubella	0.00017 (936)	0.00069	0	No intron	−1.51	2	0	0	0	2	44
grandiflora	0.01493 (955)	0.05234	0.00277	No intron	−0.36	20	28					0	2	44
SCR	2–14	25/1	rubella	0 (67)	0	0	0.00049 (736)	−0.69	1	1	1*	—	—	—
grandiflora	only 1 sequence											—	—	—
SRK(S domain)	0	25/12	rubella	0.00029 (1306)	0.00126	0	No intron	0.72	0	1	≈6†	0†	≈7†	≈5†
rubella all SRK					0.49	2	5					0†	≈7†	≈5†
grandiflora	0.4153 (717)	0.99986	0.32362	No intron	−0.77	79	236					0†	≈7†	≈5†
ARK3 (At4 g21380)	2	25/14	rubella	0 (839)	0	0	No intron	—	0	0	4	0	0	65
grandiflora	0.01956 (931)	0.05860	0.00858	No intron	−1.10	40	31					0	0	65
B160 (At4 g21430)	21–22	25/14	rubella	0 (537)	0	0	0.00153 (271)	−0.48	1	1	1	0	2	20
grandiflora	0.00333 (673)	0.00561	0.00271	0.00486 (282)	−1.86*	21	3					0	2	20
At4 g21580	89-?	25/14	rubella	0 (463)	0	0	0.00217 (268)	0.21	1	1	1	0	1	52
grandiflora	0.00991 (463)	0.03871	0.00123	0.06153 (357)	0.76	6	53					0	1	52
S12 (At4 g22720)	554-?	25/14	rubella	0.00065 (736)	0	0.00085	0 (182)	1.35	0	1	0	0	1	42
grandiflora	0.01405 (802)	0.05583	0.001860	0.04676 (183)	0.29	13	40					0	1	42
At4 g23840	1013-?	25/14	rubella	0.00022 (365)	0	0.00029	0.00146 (435)	−0.25	1	2	0	0	3	26
grandiflora	0.00698 (556)	0.01764	0.00339	0.00733 (468)	−1.07	18	11					0	3	26
At4 g25100	1510-?	25/14	rubella	0.00494 (380)	0.01594	0.00184	0.00434 (457)	1.38	2	8	0	6	3	53
grandiflora	0.00796 (470)	0.03083	0.00158	0.02630 (488)	−0.32	28	31					6	3	53
Adh (At1 g77120)	Chr. 1 in A. thaliana	25/14	rubella	0.00029 (808)	0.00128	0	0 (467)	−1.21	1	1	1	0	2	25
grandiflora	0.00333 (967)	0.01478	0	0.01594 (458)	−0.09	10	21					0	2	25
FRI (At4 g00650)	Chr. 4 in A. thaliana		rubella	0.00260 (148)	0.01061	0	0.01163 (752)	1.42	1	23	0	14	5	31
			grandiflora	0.01571 (148)	0.04304	0.00726	0.1982 (780)	−0.14	19	30				
PHYC (At5 g35840)	Chr. 5 in A. thaliana		rubella	0.01241 (683)	0.03273	0.00694	0.01131 (141)	3.46***	0	19	0	18	0	18
			grandiflora	0.01746 (683)	0.05508	0.00737	0.01347 (141)	0.61	12	25				
FLC (At5 g10140)	Chr. 5 in A. thaliana		rubella	0.00039 (204)	0	0.000531	0.00024 (645)	−1.21	1	1	2	0	2	14
			grandiflora	0.00219 (204)	0.00475	0.00143	0.00737 (647)	−1.07	12	6				
LD (At4 g02560)	Chr. 4 in A. thaliana		rubella	0	0	0	0.00177 (270)	1.12	0	1	—	—	—	—
PI (At5 g20240)	Chr. 5 in A. thaliana		rubella	—	—	—	0.00219	−0.62	1	9	—	—	—	—
CRY (At4 g08920)	Chr. 4 in A. thaliana		rubella	0	0	0	0	—	0	0	—	—	—	—
matK	chloroplast		rubella	0.00028 (1489)	0	0.00036	0 (796)	−0.48	1	1	0	0	2	4
			grandiflora	0.00058 (1489)	0.00086	0.00050	0.00147 (796)	0.21	1	3				

*Only 224 overlapping basepairs, with 1 difference between *C. rubella* and *C. grandiflora*. *C. rubella* has no polymorphisms in this region.

†Only 2 C. grandiflora sequences as well as ≈420 bp missing from one of the C. grandiflora sequences.

Table S4. ***ADH*** **diversity analyses in different species**

No	Species	N	Length (bp)	S	π	θ_W
1	*A. thaliana*	19	1279	36	0.00814	0.00812
2	*A. lyrata*	11	1284	18	0.00355	0.00487
3	*C. rubella*	25	1262	2	0.00018	0.00042
4	*C. grandiflora*	14	1277	33	0.00845	0.00833
5	*C. bursa-pastoris* A	4	627	12	0.00970	0.01046
6	*C. bursa-pastoris* B	4	625	1	0.00080	0.00087