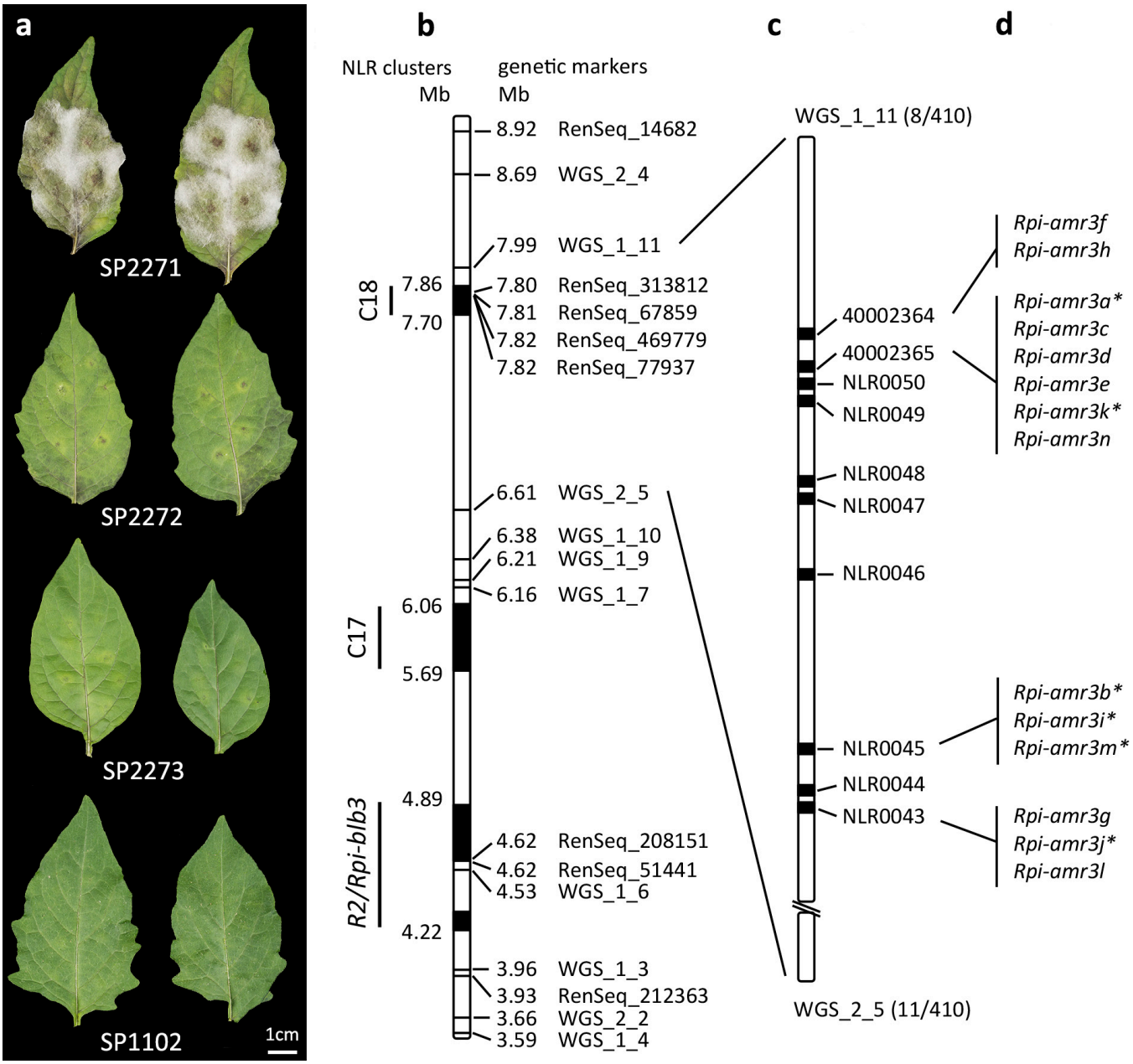
Supplementary Materials

Accelerated cloning of a potato late blight–resistance gene using RenSeq and SMRT sequencing

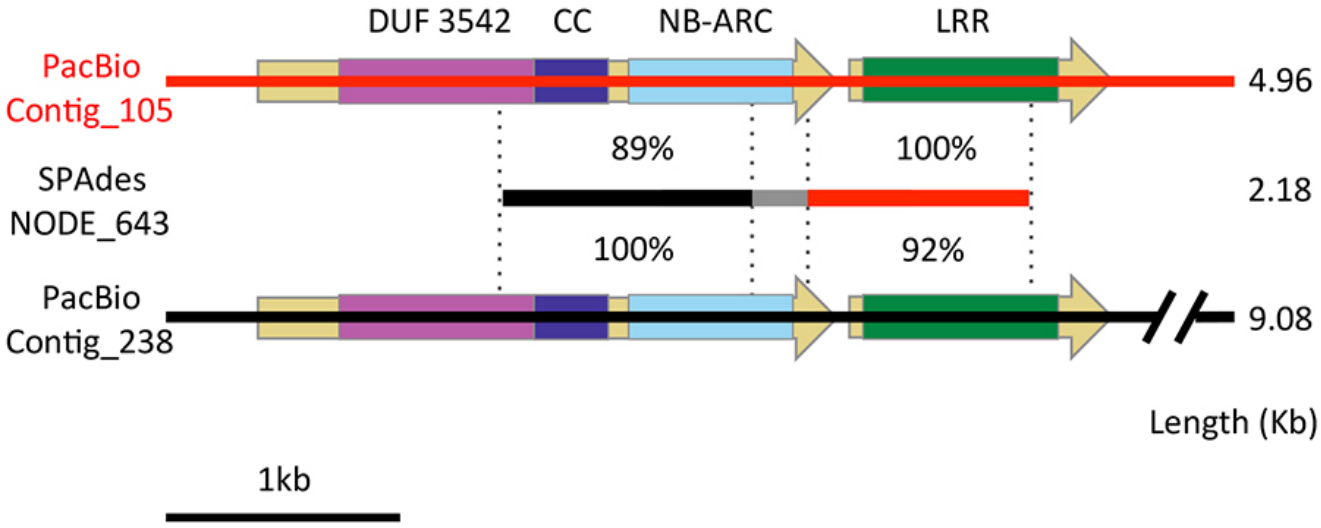
Kamil Witek[1,4], Florian Jupe[1,3,4], Agnieszka I Witek[1], David Baker[2], Matthew D Clark[2], and Jonathan DG Jones[1]

Supplementary Figures



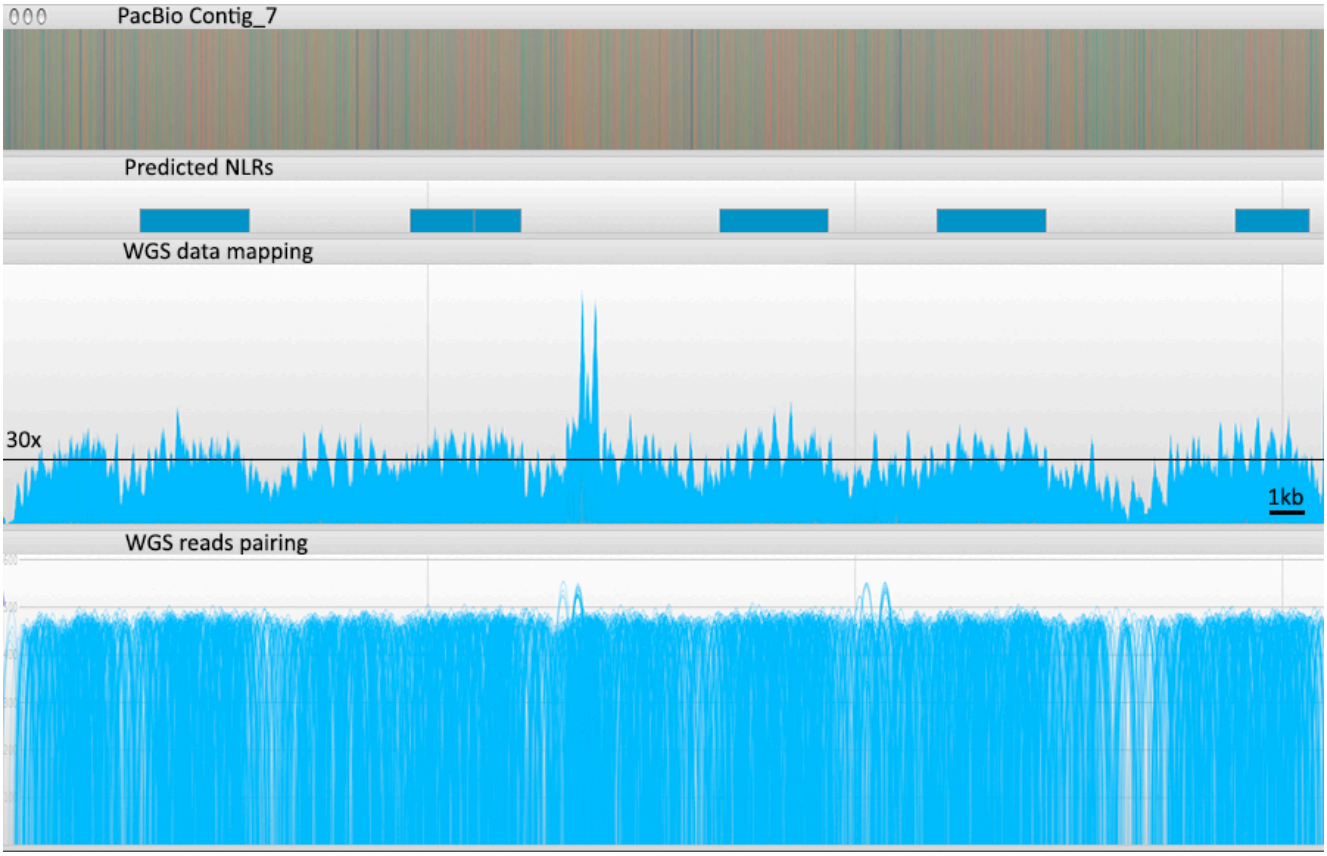
Supplementary Figure 1. Resistance in *S. americanum* accession SP1102 is physically linked to the *R2/Rpi-blb3*, C17 and C18 DM reference genome loci.

(**a**) 10-week-old *S. americanum* plants were inoculated with *P. infestans* strain 88069. Top to bottom: accession 954750186 (working name SP2271) is susceptible to the tested isolates; accessions 954750174 (SP2272), 954750184 (SP2273) and 944750095 (SP1102) are resistant. Each leaf was inoculated with 4-6 droplets containing 500 zoospores each; photographs were taken 7 dpi. Scale bar indicates 1cm. This experiment was repeated at least three times with similar results (**b**) Bulked segregant analysis coupled with RenSeq enables positioning of *Rpi-amr3* on DM reference genome chromosome 4. 50 susceptible F2 and F3 plants were genotyped with markers derived from Whole Genome Shotgun sequencing (WGS) data. Two flanking markers (WGS_1_4 at 3.59Mb and WGS_2_4 at 8.69Mb), and eight co-segregating with resistance (WGS markers), were identified. Physical positions of NLR genes and markers are given in Mb, based on the reference genome. Solid areas mark NLR cluster locations. (**c**) Genotyping and phenotyping of 405 plants from a selfed backcross population (BC_1F_2) confirmed the location of *Rpi-amr3* within the C18 locus (10 NLRs in reference genome), flanked by markers WGS_2_5 (6.61Mb) and WGS_1_11 (7.99Mb), Numbers in brackets indicate number of recombinants out of all screened plants. (**d**) Sequence comparisons of PacBio-RenSeq derived *de novo* assembly with members of the C18 locus identified four homologous sub groups, designated *Rpi-amr3* candidates a-n. Asterisk marks expressed candidates.



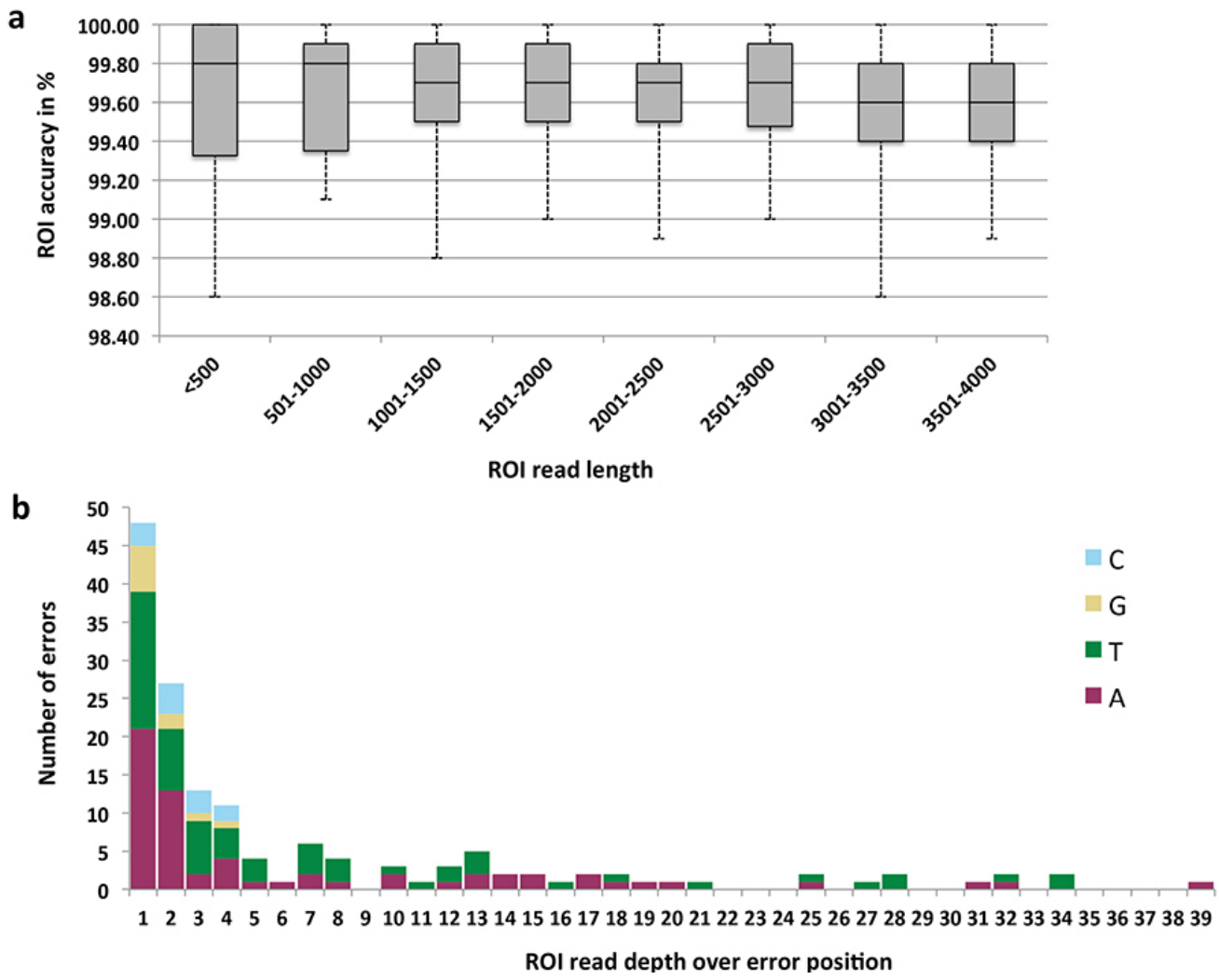
Supplementary Figure 2. An example of chimeric MiSeq assembly of 2x250bp SP1102 RenSeq data.

SPAdes assembled MiSeq sequence contigs were aligned to PacBio derived contigs with high stringency (Geneious). Top (red) and bottom (black) – PacBio contigs, middle – chimeric MiSeq contig with marked regions fitting PacBio contigs (black and red). Above and below SPAdes NODE_643 identity to PacBio Contig_105 and Contig_238 are shown, respectively. Grey area in SPAdes contig marks region of NB-ARC domain with 99% identity to both PacBio contigs (240nt), which is a part of larger fragment (650nt) displaying 96% identity between PacBio contigs. Overall identity between PacBio contigs (ORF region) is 85% on nucleotide level. ORFs are marked with arrows; NLR typical protein domains: coiled-coil (CC; purple box), nucleotide binding site (NB-ARC; blue box) and leucine-rich repeats (LRR; green box). Fused protein domain DUF 3542 is shown as light purple box. Drawn to scale.



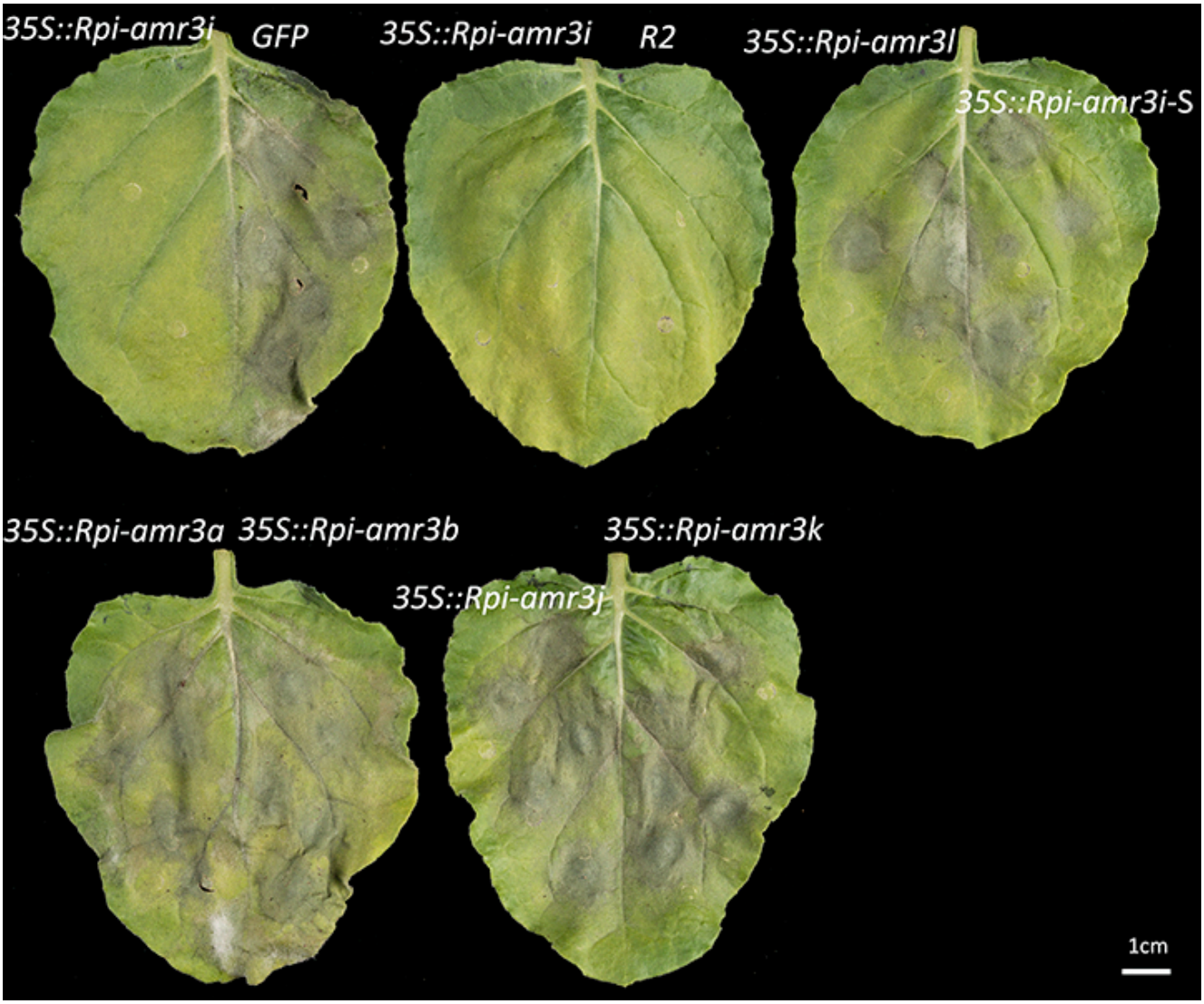
Supplementary Figure 3. Confirmation of continuity and accuracy of the PacBio ROI assembled *Contig_7*.

Whole Genome Shotgun (WGS) Illumina 100bp paired-end data were mapped to *Contig_7* (Supplementary File 2) using BWA with default settings and then filtered with SAM tools for correctly mapped read pairs. Mapping results were visualized in Savant Genome Browser and inspected manually. No signatures of misassembly or indels larger than a single nucleotide (typically errors in homopolymeric regions) were noticed. Single visible SNPs (up to 50% in nucleotide ratio) in the ORF result from non-specific read mapping from close paralogs. From top to bottom: graphical representation of *Contig_7* reference sequence as assembled from PacBio ROI data; ORF predicted with Geneious R8 (blue bars); WGS coverage data, blue color = perfect coverage, various color bars = SNPs or single indels, black line marks 30x coverage; WGS data read pairing, blue color = proper pairing, no discordant, everted or unmapped pairs present.



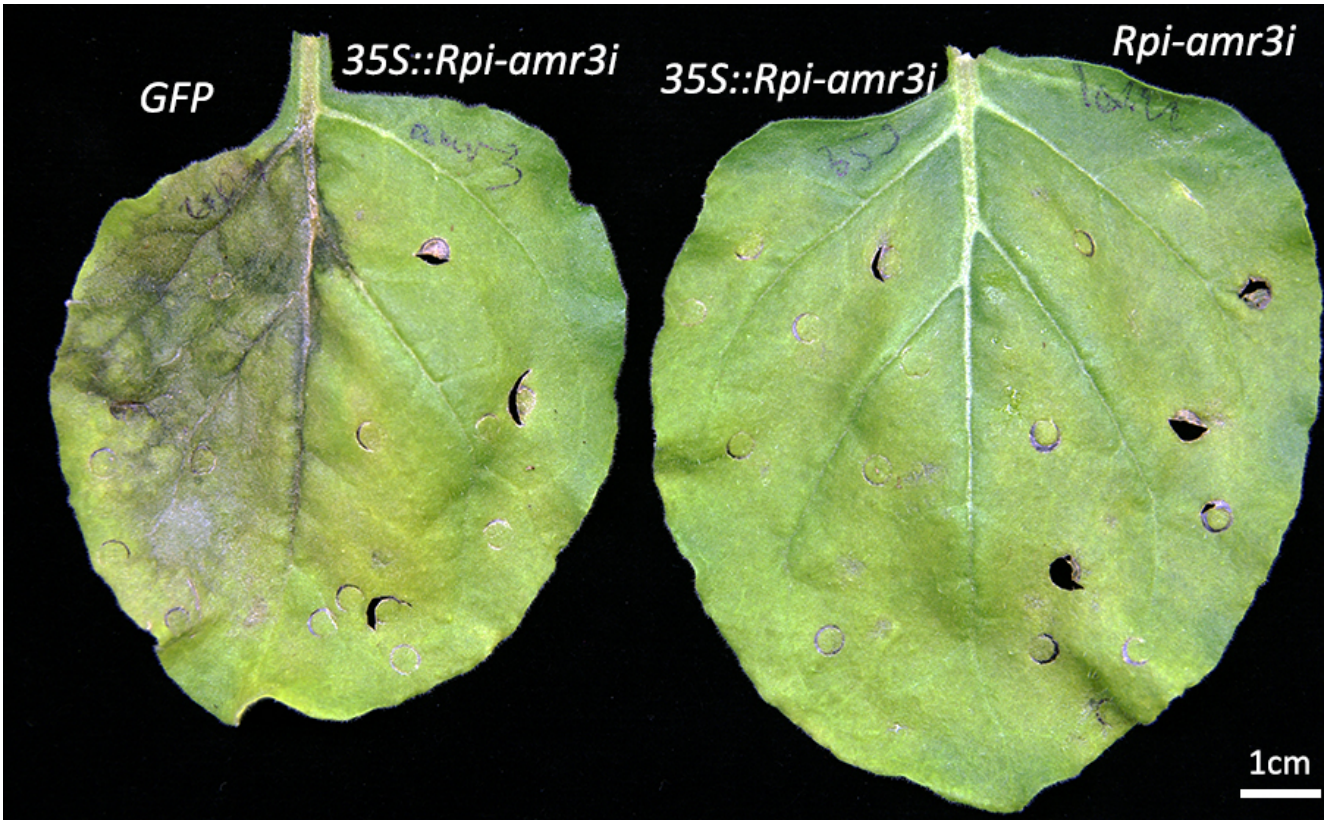
Supplementary Figure 4. Error rate of ROI reads and resulting assemblies.

(a) Box plot showing accuracy of ROI reads divided into bins based on their length (500 bp intervals). ROI reads that were used to *de novo* assemble the C18 clusters of resistant and susceptible parents, were re-aligned to the error corrected contigs. The percentage of accuracy was calculated for each individual ROI read based on the pairwise identity. (b) Bar graph showing total number of errors per ROI read coverage over *de novo* assembled contigs. WGS data was mapped to *de novo* assembled contigs of the C18 cluster and the mapping data was checked for errors. Colors represent the four bases.



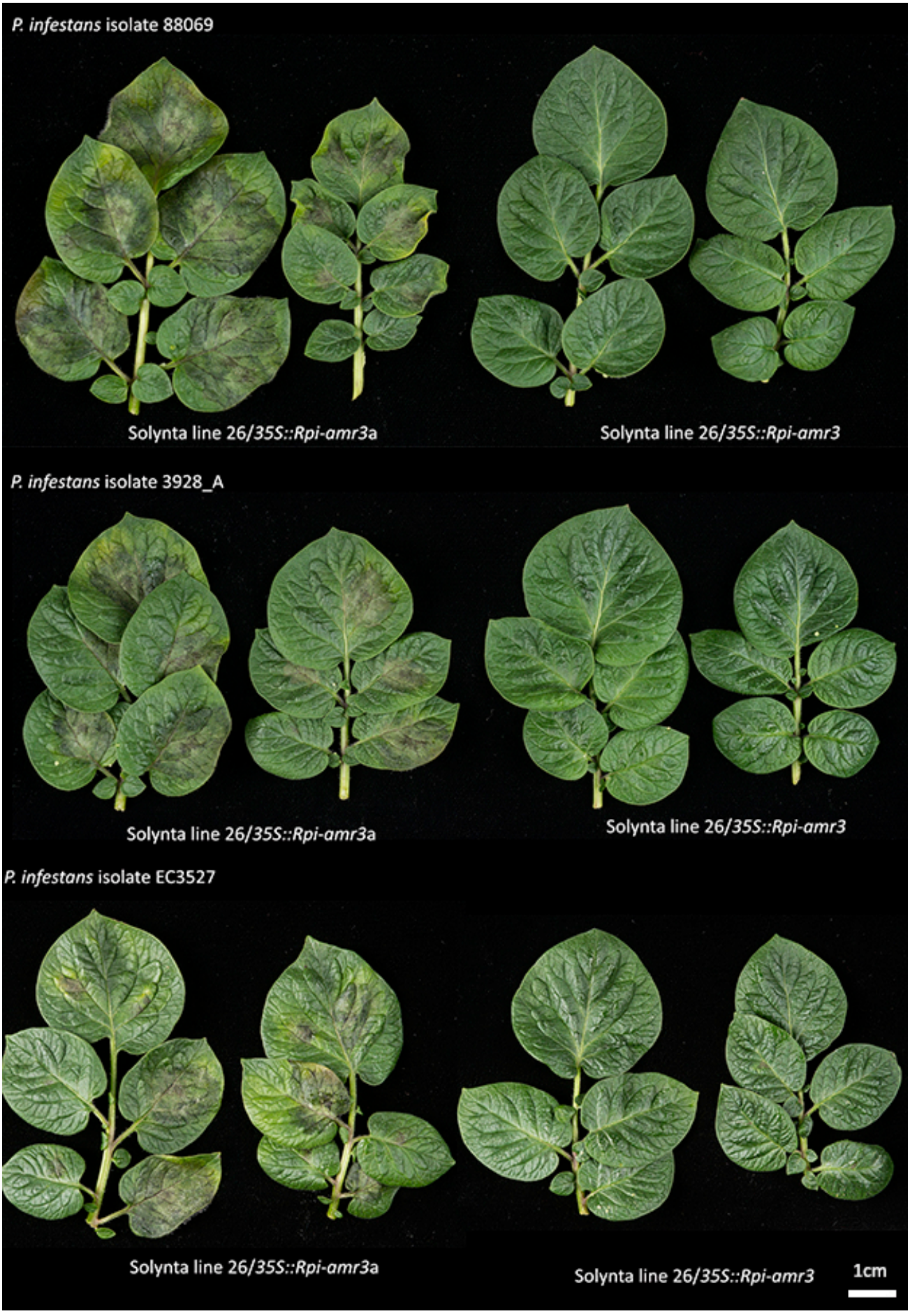
Supplementary Figure 5. Transient complementation assays in *Nicotiana benthamiana* with six *Rpi-amr3* candidates.

Third leaves of *N. benthamiana* plants were infiltrated with the binary vector pICSLUS0003::35S overexpressing either the late blight resistance gene *R2* (positive control), six *Rpi-amr3* candidates, *Rpi-amr3i-S* cloned from the susceptible accession 2271 or GFP (negative control). Leaves are 24 hours later inoculated with *P. infestans* strain 88069. Only leaves infiltrated with *R2* and *Rpi-amr3i* remained infection free, while *P. infestans* was able to proliferate on the remaining *Rpi-amr3* candidates, the susceptible allele of *Rpi-amr3i-S* as well as the GFP control. Scale bar indicates 1cm. These experiments were repeated twice with similar results.



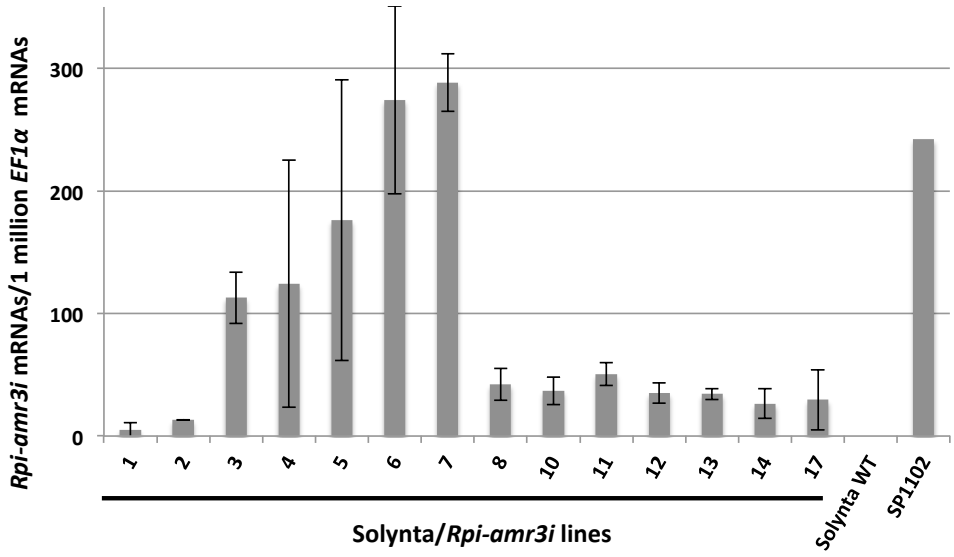
Supplementary Figure 6. Genomic construct with *Rpi-amr3i* confers resistance against *P. infestan*s in a transient complementation assay in *N. benthamiana.*

The *Rpi-amr3i* construct with native promoter and terminator, restricts *P. infestans* growth to the same level as under control of the *35S* promoter. A vector overexpressing GFP was used as a negative control. The experiment was performed as described previously; photographs were taken after 6 days. Scale bar indicates 1cm. This experiment was repeated twice with similar results.


P. infestans isolate 88069
Solynta line 26/*35S::Rpi-amr3*a
Solynta line 26/*35S::Rpi-amr3*
P. infestans isolate 3928_A
Solynta line 26/*35S::Rpi-amr3*a
Solynta line 26/*35S::Rpi-amr3*
P. infestans isolate EC3527
Solynta line 26/*35S::Rpi-amr3*a
Solynta line 26/*35S::Rpi-amr3*
1cm

Supplementary Figure 7. Stable transgenic plants carrying *Rpi-amr3i* under the regulation of a *35S* promoter display resistance to all tested isolates.

Transgenic diploid potato "Line 26" (Solynta B.V.) that express *Rpi-amr3i* under the *35S* promoter are resistant to *P. infestans* isolates 88069 (upper), 06_3928A (middle) and EC3527 (bottom; right panel). The transgenic line displays no to weak HR at the place of inoculation. In contrast, transgenic plants carrying the non-functional candidate *Rpi-amr3a* (left panel) showed large necrotic lesions and sporulation. Each leaflet was inoculated with a droplet containing 500 spores; photographs were taken 6 dpi. Scale bar indicates 1cm. These experiments were repeated twice with similar results.



Supplementary Figure 8. Relative levels of *Rpi-amr3i* expression under native regulatory elements in transgenic potato plants.

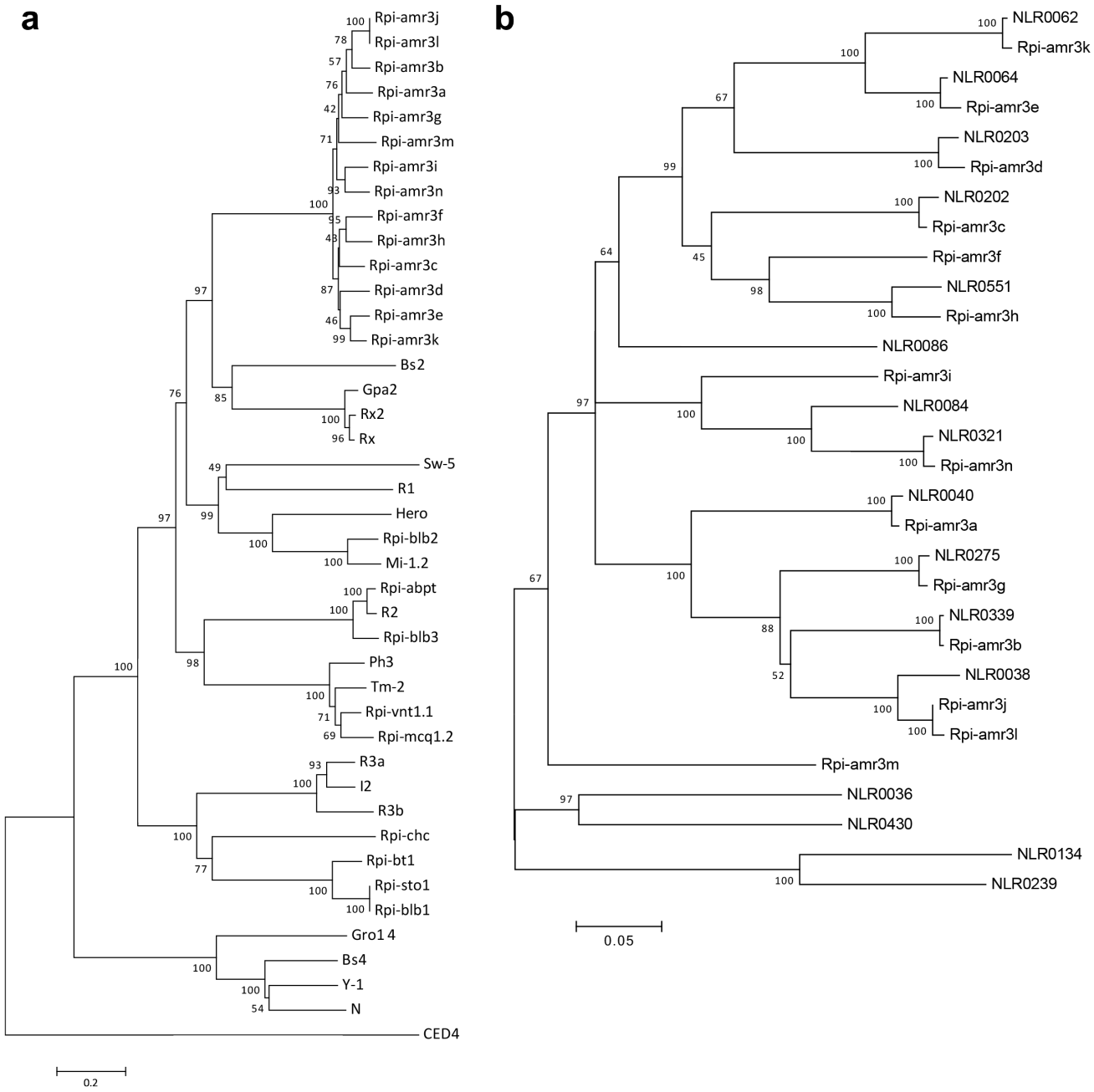
Relative copy number of *Rpi-amr3i* mRNAs per 1 million copies of *EF1* mRNA internal control. Expression levels similar to wild-type *Rpi-amr3i* mRNA (lines 3-7) correspond to full resistance. Level of expression was measured in fully grown leaves from 10 and 14 week-old plants for Solynta transgenic and WT lines and in 12-week-old SP1102. Errors bars show standard deviation based on two time points. Primers show high specificity, as no amplification was observed in Solynta WT plants.

>*Rpi-amr3i* protein sequence

MAAYSAVISLLQTLIDQQNISELFHGHTAQTLDSLHTTAEYFQHVLENITRFDSEKIKSLEEKIRVVVSYAEDVVAMKIS
QIIIGSSWTFGILQHQDLLPLVEKMDTTKKQVMDILSHDDDQILELTAGDSLIGTSSTTYPMLEDDIVQGIDDDLEIIVK
RLTGPPRDLDVVTITGMGGIGKTTLARKAYDHLTIRYHFDILVWITISQEFRCRNVLLEALHCISKSTDIVNTKDYDKKD
DNELADIVQKKLKGPRYLVVVDDIWSRDVWDSIRGIFPNYNNGSRILLTTRENEVAMYANTCSPHEMSLLSLENGWRLLC
DKVFGPKHDHPPELEEIGKEIVEKCQGLPLTISVIAGHVSKMPRTLECWKDVARTLSEIISSHPDNCLGVLGLSYHHLPN
HLKPCFLSMSSFPEDFQVETRRLIYLWIAEGFIRTCENGKSLEEVAVDYLEDLISRNLIQARKRRFNGEIKACGIHDLLR
EFCLIEAEITKHMHVERTYPTLPTQKNNVRRFSFQTKFYSVDDCNKLLPPVARSIYFFSQLDLPVVPYKRYLRCCLPIHR
DDRIIHDFYSRFNLLRVLVISKTNEYFESFPLVITKLFHLRYLQVRFLGDIPESISNLQNLQTLICSGGTLPGKIWMMKN
LRYISIIGNKVTYLPSPRTESLVNLEEFSVLCYRSCTKEVISGIPNLKRLTIDVLSSINNYFPNGLIDMSSLTKLEAFKC
NRCLYSNFNSSVIPTSLKDFVFPTSLKRLSLNYYASHFFWEEISSTIIMLPNLEELKLKDCRSDEYDEWSLSDKDKFKSL
KLLVLTDIFFDRWEATSDNFPNLKRLVLNKCDLEIPSDFGEICTLESIELHDCSTSAEDSAREIEQEQEEMGNNILKVYI
HGSRSKF

Supplementary Figure 9. Amino acid sequence of *Rpi-amr3i* with highlighted conserved domains.

Conserved domains within the sequence are highlighted; coiled-coil, purple; NB-ARC, blue; leucine-rich repeats, green.



Supplementary Figure 10. *Rpi-amr3i* establishes a new branch in the phylogeny of cloned functional *Solanaceae R* genes.

(a) The *Rpi-amr3* gene family forms a separate branch among previously cloned *R* genes. *Gpa2/Rx* represent the closest clade, members share however less than 35% amino-acid sequence identity. This maximum likelihood tree is based on the alignment of the full length amino-acid sequences of various functional Solanaceae NLR resistance genes, and *Caenorhabditis elegans* protein CED4 as outgroup. The numbers at nodes represent their bootstrap support (% support out of 100 bootstraps). (b) The phylogenetic tree of C18 cluster of R and S parents was constructed as described above, using all NLRs mapping to this cluster. C18 cluster members of resistant parent are named Rpi-amr3a-n, names of

paralogs from susceptible parent start with NLR followed by number (see Supplementary File 7 for full NLR list).

Supplementary Tables

Supplementary Table 1. Details of *S. americanum* accessions used in this study and segregation ratio for *P. infestans* in F2 populations.

Accession	Working name	Species	Place of origin	Source	Phenotype	*Rpi* genes based on F2 segregation pattern
954750186	SP2271	*S. americanum*	Brazil	RU	Susceptible	
954750184	SP2273	*S. americanum* var. Patulum	unknown	RU	Resistant	1
954750174	SP2272	*S. americanum*	unknown	RU	Resistant	1
A14750006		*S. americanum sensu lato*	unknown	RU	Resistant	1
944750095	SP1102	*S. americanum sensu lato*	Mexico	RU	Resistant	2 or more
A54750014		*S. americanum sensu lato*	unknown	RU	Resistant	2 or more
SOLA 140		*S. americanum*	Cuba	IPK	Resistant	2 or more
SOLA 424		*S. americanum*	Middle America	IPK	Resistant	2 or more
SOLA 428		*S. americanum*	Middle America	IPK	Resistant	2 or more
SOLA 432		*S. americanum*	Middle America	IPK	Resistant	2 or more
10145		*S. americanum*	El Salvador	NHM	Resistant	1
Wang 2059		*S. americanum*	China	NHM	Resistant	1
Wang 2058		*S. americanum*	China	NHM	Resistant	1

RU - Radboud University, Nijmegen, The Netherlands
IPK - IPK Gatersleben, Germany
NHM - Natural History Museum, London, United Kingdom

Supplementary Table 2. The NLR complement of SP2271 and SP1102.

Class	Full		Partial		Total	
	2271	**1102**	**2271**	**1102**	**2271**	**1102**
TNL	71	44	29	45	100	89
CNL	330	278	198	222	528	500
Other*	-	-	18	26	18	26
Total	401	322	245	293	646	615

*** No motifs characteristic for CC or TIR domains were detected with MAST pipeline**

Supplementary Table 3. Comparison between MiSeq and PacBio read assemblies for SP1102

		NLR				
SP1102	# contigs	total	complete	partial	unique	chimeras
PacBio	775	615	322	293	327	-
CLC	52,145	584	124	460	1	-
SPAdes	30,314	742	216	526	4	5

Supplementary Table 4. Non-canonical protein domains fused to NLRs.

NLR ID	Fused domain ID	Superfamily	Hit start (amino acid)	Hit stop (amino acid)	e-value
RDC0008NLR0042	DUF659 super family	cl04853	277	376	1.20E-25
RDC0008NLR0047	PspA_IM30	pfam04012	31	199	2.12E-03
RDC0008NLR0077	PTPc super family	cl21483	6	62	9.61E-04
RDC0008NLR0080	ASF1_hist_chap super family	cl22451	55	182	1.22E-03
RDC0008NLR0088	DUF3542	pfam12061	94	402	4.37E-68
RDC0008NLR0089	DUF3542	pfam12061	91	471	2.61E-114
RDC0008NLR0124	DUF3542	pfam12061	94	458	1.33E-109
RDC0008NLR0125	DUF3542	pfam12061	94	475	1.37E-108
RDC0008NLR0148	DUF3542	pfam12061	95	470	3.16E-112
RDC0008NLR0149	DUF3542	pfam12061	91	471	9.07E-108
RDC0008NLR0157	ASF1_hist_chap super family	cl22451	234	315	9.19E-03
RDC0008NLR0188	DUF3542	pfam12061	91	467	5.73E-115
RDC0008NLR0218	DUF3542	pfam12061	169	352	8.88E-05
RDC0008NLR0220	DUF3542	pfam12061	87	463	0.00E+00
RDC0008NLR0222	DUF3542	pfam12061	390	783	1.82E-99
RDC0008NLR0265	FlaC super family	cl23430	68	201	1.31E-04
RDC0008NLR0274	DUF3542	pfam12061	91	466	6.15E-111
RDC0008NLR0276	DUF3542	pfam12061	121	478	1.18E-53
RDC0008NLR0281	PHA03151 super family	cl14512	223	358	1.97E-04
RDC0008NLR0312	DUF3542	pfam12061	83	481	0.00E+00
RDC0008NLR0323	DUF3542	pfam12061	242	501	1.03E-05
RDC0008NLR0329	DUF3542	pfam12061	104	432	8.71E-04
RDC0008NLR0344	DUF3542	pfam12061	63	461	4.21E-05
RDC0008NLR0366	DUF3542	pfam12061	91	470	5.37E-116
RDC0008NLR0419	RNA_pol_3_Rpc31 super family	cl13200	135	235	2.44E-03
RDC0008NLR0443	DUF3542	pfam12061	229	495	9.70E-05
RDC0008NLR0504	DUF3542	pfam12061	113	487	2.67E-05
RDC0008NLR0527	DUF3542	pfam12061	95	466	4.50E-30
RDC0008NLR0548	DUF3542	pfam12061	540	814	1.35E-80
RDC0008NLR0553	HGD-D super family	cl21559	255	396	5.23E-03
RDC0008NLR0578	DUF3542	pfam12061	22	352	5.02E-170

RDC0008NLR0612	DUF3542	pfam12061	116	503	5.57E-164
RDC0008NLR0618	DUF3542	pfam12061	114	505	3.10E-172
RDC0008NLR0629	DUF3542	pfam12061	73	478	4.38E-156
RDC0008NLR0631	DUF3542	pfam12061	112	504	5.95E-172
RDC0008NLR0636	DUF3542	pfam12061	93	498	1.49E-175
RDC0008NLR0681	DUF3542	pfam12061	87	473	5.83E-170

Supplementary Table 5. cDNA reads count for full length *Rpi-amr3* candidates

Gene	Reads count
Rpi-amr3a	460
Rpi-amr3b	5196
Rpi-amr3c	67
Rpi-amr3d	11
Rpi-amr3e	428
Rpi-amr3f	48
Rpi-amr3g	0
Rpi-amr3h	0
Rpi-amr3i	9032
Rpi-amr3j	1282
Rpi-amr3k	510
Rpi-amr3l	1538
Rpi-amr3m	68
Rpi-amr3n	46

Supplementary Table 6. Response of transgenic potato plants containing *Rpi-amr3i* against a range of *P. infestans* isolates

Isolate	Coutry of origin	*Rpi-amr3* phenotype
88069	The Netherlands	Resistant
06_3928A	United Kingdom	Resistant
MP324	Poland	Resistant
EC1	Ecuador	Resistant
EC3527	Ecuador	Resistant
EC3626	Ecuador	Resistant

Supplementary Table 7. Markers details.

For marker sequence see Supplementary File 1.

Marker name	Restriction enzyme	Primer	Primer sequence 5' – 3'
RenSeq			
c14682	NA	NA	NA
c469779	NA	NA	NA
c77937	NA	NA	NA
c67859	NA	NA	NA
c208151	NA	NA	NA
c51441	NA	NA	NA
c313812	NA	NA	NA
c212363	NA	NA	NA
WGS			
R2l_1_11	ClaI	R2l_1_11_F	CTTTCGATATAGCATGTTTAAGATTACATGA
		R2l_1_11_R	TAGCTGATCCTGAGAAGGTTAGACTA
R2l_1_10	HpaI	R2l_1_10_F	CACGTGAACCAGGTGATTCGAAATG
		R2l_1_10_R	GAACTATATTAGTGAAGGTTCAGTAGTGC
R2l_1_9	HaeIII	R2l_1_9_F	CGTACTCATGTTAGATCCTCCAAAAAATG
		R2l_1_9_R	TGATTTTGACTCACTCGCTGTGGATGA
R2l_1_7	HaeII	R2l_1_7_F	CAAATCTGACTCTGCAATAGGAATTGAC
		R2l_1_7_R	GTAATTGTCTATGAGGAGAGGGGGTT
R2l_1_6	ClaI	R2l_1_6_F	AATCTGGTTCCGAAGAGAAGCCACTTAA
		R2l_1_6_R	TCTGGAAAATAGGATGGGTAATTTACGAAGA
R2l_1_3	PCR	R2l_1_3_F	GCATCAACTCTTTTAGTACTAATTTGGTCTG
		R2l_1_3_R	CACTGATTCTTAACATGCATATTTAAGGAGA
R2l_1_4	TaqI	R2l_1_4_F	CCATCATCTCAAGGATTCTCAAGCTAG
		R2l_1_4_R	GAGTTACATCAATGAAGTGTCCGTTTTC
R2l_2_2	AseI	R2l_2_2_F	GGCGAATGGTCACCTGAAGAAGATAT
		R2l_2_2_R	AGGATCCGACTATCTAAAAGGTACTCTA
R2l_2_4	HinfI	R2l_2_4_F	GAAAATGTAAACAGCAAATAATCATGCTACC
		R2l_2_4_R	GTAAGAACATATCCCTTATGTCCAACCA
R2l_2_5	PCR	R2l_1_5_F	ACTGTGTCATCAGCAAATATGTACAGTTG
		R2l_1_5_R	ATGGGAAGGTCCTAGAGCTTTTGCAC

Supplementary Table 8. Primers used to clone candidate *Rpi-amr3* genes

Gene	Primer	Sequence
Rpi-amr3a	KW_35S_amr3a_F	**GGCTTAAU**GTCCTTGCATATCCTGTTTCCAATAATCC
	KW_35S_amr3a_R	**GGTTTAAU**CTCTGCATCTGAAAGATAACAAGTACAACTT
Rpi-amr3b	KW_35S_amr3b_F	**GGTTTAAU**TTTAATAATGTGAAGAATCGAACAACTTGT
	KW_35S_amr3b_R	**GGCTTAAU**TGTCCTTACATATCCTGTTTCCAATAATCC
Rpi-amr3i	KW_35S_amr3i_F	**GGCTTAAU**TCTTCAATTAAATCCTATCCACTCCTCATC
	KW_35S_amr3i_R	**GGTTTAAU**CGAAATATGCTTCCATTTTCCTGCCTATGC
	KW_genomic_amr3i_F	**GGCTTAAU**GTCCATATGTGGAAGCTACTCTCTTTGTCCA
	KW_genomic_amr3i_R	**GGTTTAAU**TCTCCAAAATGGTCACCAAAACAAGTGCCA
Rpi-amr3j	KW_35S_amr3j_F	**GGCTTAAU**CTATCCACTCCTTACATACCATCAATATTC
	KW_35S_amr3j_R	**GGTTTAAU**GCGAAGAATCGAAGAATGTCTTGGAGAGAT
Rpi-amr3k	KW_35S_amr3k_F	**GGCTTAAU**AATGGAACGATATTCAGATTGACTTACCAAA
	KW_35S_amr3k_R	**GGTTTAAU**GTGAAGAATCGAACAACTTTTTGGAGAGAT
Rpi-amr3l	KW_35S_amr3l_F	**GGCTTAAU**CTTGTTCATCATTTTTGAAAATTAAT
	KW_35S_amr3l_R	**GGTTTAAU**TTGGGGTGTGAGTTAGGTCCAAGACTTAAT

In bold - extension for USER cloning

Normal font - gene specific sequence

35S - primers to clone gene into expression vector under 35S promoter and OCS terminator

genomic - primers to clone gene into expression vector under the control of native regulatory elements

Supplementary Table 9. Reference NLR genes for phylogenetic studies.

Gene name	Reference species	GenBank/patent number	Protein class
CED-4	*Caenorhabditis elegans*	NP_001021202	outgroup
ADR1	*Arabidopsis thaliana*	NP_174620	CNL
Bs4	*Solanum lycopersicum*	AY438027	TNL
Bs2	*Capsicum chacoense*	AF202179	CNL
Gpa2	*Solanum tuberosum*	AAF04603	CNL
Gro1-4	*Solanum tuberosum*	AAP44390	TNL
Hero	*Solanum lycopersicum*	AX337980	CNL
I-2	*Solanum lycopersicum*	AF118127	CNL
L6	*Linum usitatissimum*	AAD25968	TNL
M	*Linum usitatissimum*	AAB47618	TNL
Mi-1.2	*Solanum lycopersicum*	AF039682	CNL
N	*Nicotiana glutinosa*	Q40392	TNL
NRG1	*Nicotiana benthamiana*	AAY54606	CNL
NRC1	*Solanum lycopersicum*	DQ304484	CNL
Ph3	*Solanum lycopersicum*	KJ563933	CNL
Prf	*Solanum lycopersicum*	AAC49408	CNL
RPS2	*Arabidopsis thaliana*	AAM90881	CNL
R1	*Solanum demissum*	AF447489	CNL
R2	*Solanum demissum*	FJ536325	CNL
R3a	*Solanum demissum*	AY849382	CNL
R3b	*Solanum demissum*	JF900492	CNL
Rpi-chc1	*Solanum chacoense*	WO2011034433	CNL
Rpi-blb1	*Solanum bulbocastanum*	FB764493	CNL
Rpi-blb2	*Solanum bulbocastanum*	DQ122125	CNL
Rpi-blb3	*Solanum bulbocastanum*	FJ536326	CNL
Rpi-vnt1.1	*Solanum venturii*	FJ423044	CNL
Rpi-mcq1	*Solanum mochiquense*	WO2009013468	CNL
Rx	*Solanum tuberosum*	AJ011801	CNL
Rx2	*Solanum acaule*	CAB56299	CNL
Ry-1	*Solanum tuberosum*	CAC82812	TNL
Sw-5	*Solanum lycopersicum*	AY007366	CNL
Tm2	*Solanum lycopersicum*	AY742887	CNL

Supplementary Data and Methods

Plant material

Seeds of *Solanum* accessions were obtained from seed banks as detailed in Supplementary Table 1. Accessions 944750095, A54750014 and A14750006 were originally described as the hexaploid *S. nigrum* in the seed database at Radboud University (Nijmegen, The Netherlands). Flow cytometric analyses that we carried out identified these however as diploid *S. americanum*. This is similar to earlier misidentifications between these very similar species[22]. *Solanum* taxonomy is not the focus of this paper, and all accessions used here are regarded as belonging to *S. americanum sensu lato*. *S. americanum* plants used in this study were grown in our greenhouse facilities.

DNA and RNA extraction

RenSeq experiments were conducted on gDNA freshly extracted from young leaves (both MiSeq and PacBio protocols) using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. For the cDNA RenSeq experiment, RNA was extracted using TRI-reagent (Sigma- Aldrich, MO, USA) and Directzol RNA Mini-prep kit (Zymo Research, CA, USA), following manufacturer's recommendations. First-strand cDNA was made using a mix of oligo-dT and random hexamer primers and First-Strand Superscript II (Sigma-Aldrich, MO, USA). The second strand was made as described elsewhere[21].

Quantitative-RT_PCR

For quantitative Reverse Transcription PCR, first strand cDNA was made as described above from 2 µg of starting RNA. The cDNA (1µl of 1:5 dilution) was used to quantify *Rpi-amr3i* expression using a SYBR Green quantitative PCR kit (Sigma- Aldrich, MO, USA) and gene specific primers (F: CTGAGGATTCTGCACGAGAGATTG, R: TCATCATAACTTCAAGGAGGTAAG). Expression was quantified to the internal control *elongation factor-1 alpha* (*EF1α*), using primers described by[23]. Measurement for each plant and type of primers was performed in three technical replicates. Themocycling and intensity detection was carried out with Biorad CFX96 cycler and data extraction with CFX Manager software.

Phytophthora infestans strains, infections and DLA assays.

P. infestans isolate MP324[24] was provided by Jadwiga Śliwka from the *Plant Breeding and Acclimatization Institute – National Research Institute*, Mlochow Centre (Poland) and the remaining isolates were maintained in our laboratory as described previously[25]. Detached leaf assays (DLA) on leaves of *S. americanum* and stable transgenic *S. tuberosum* Solynta plants (see below) were carried out as described by[25] with minor modifications. Three leaves from 8 to 12-week-old plants were used and infected with 6 to 8 10 µl droplets of a freshly prepared and 3 hour cold-incubated zoospore suspension (50,000 zoospores ml–1) from 10-day old rye agar (RSA[26]) plate. Inoculated leaves were incubated for 6 to 12 days under controlled environmental conditions (16°C, 16 h of light and 8 h of dark) before phenotypes were scored, as described for *S. nigrum* by[9,27].

Whole Genome Shotgun (WGS) sequencing

WGS sequencing was outsourced to BGI (www.genomics.cn). Fresh leaves from young plants were shipped on dry ice to BGI, where gDNA isolation, library construction, Illumina HiSeq sequencing and initial quality control (trimmed adapters and removed reads which contain more than 50% bases with quality value <=5) were carried out. Each accession was sequenced to around 30x coverage, resulting in 346.15M and 379.53M reads for R an S parent, respectively. Reads after initial QC were deposited in ENA under run accession numbers ERR966154-56 and ERR966142-45 for the resistant and susceptible parent, respectively.

RenSeq library preparation and enrichment

For the Illumina MiSeq RenSeq experiment, 2 µg of gDNA were fragmented with the Covaris sonicator (Covaris Inc. MA, USA) using preset *1 kb* settings. Fragments were size selected for fragments longer than 500bp using AmPureXP beads (Beckman Coulter, CA, USA) with 1:0.55 ratio of sheared DNA to AMPureXP beads. Double-stranded cDNA prepared as described above was used directly for library preparation, without shearing or size selection. Illumina MiSeq gDNA and cDNA libraries were prepared using the NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs, Inc., Ipswich, MA, USA) following Manufacturer's instructions. Target capture was carried out using a custom MYcroarray MYbaits kit (MI, USA) and the corresponding protocol. The bait library was

designed as described[8] and comprises 20,000 unique 120-mer oligos. 10 µl of enriched library was PCR amplified (KAPA HiFi enzyme) up to a quantity of 1 µg in 50 µl reaction volumes using Illumina P5 and P7 primers (Illumina Inc. CA, USA). MiSeq 250-bp paired end (PE) sequencing was carried out at The Genome Analysis Center (TGAC, Norwich Research Park, UK), and generated 2,454,724, 2,824,501, 4,697,915 and 1,675,385 PE reads. Raw reads were deposited at ENA under run accession numbers ERR966152, ERR965967, ERR966146-47 and ERR966153 for resistant parent (R), susceptible parent (S), bulked susceptible (BS) and R parent cDNA, respectively.

A detailed protocol for the SMRT RenSeq library preparation and enrichment that we describe in this manuscript is deposited at Protocol Exchange website (http://www.nature.com/protocolexchange/, Jonathan Jones Lab, SMRT RenSeq Protocol). Briefly, we Covaris sheared and pre-selected gDNA fragments longer than 2 kb using AMPureXP beads in a ratio 1:0.4, and generated sequencing libraries with the NEBNext Ultra DNA Library Prep Kit for Illumina (NEB, MA, USA). NLR gene fragments were then captured using the custom MYcroarray MYbaits bait library. Single reactions were carried out in one-half reaction volume per each sample. Enriched libraries were PCR amplified to 10-20 µg (P5 and P7 primers, KAPA HiFi enzyme) and subsequently size-selected with the BluePippin system (Sage Science, MA, USA) prior PacBio RSII SMRT sequencing at The Genome Analysis Center (TGAC, Norwich Research Park, UK). DNA was size selected using Blue Pippin for 1.5 kb, 2.5 kb and longer fragments from SP1102 (R parent) libraries, and 3 kb and longer fragments from SP2271 libraries (S parent). Samples were sequenced with P5-C3 chemistry and P6-C4 chemistry for R and S parents, respectively. The samples were sequenced on one (1.5 kb), two (2.5 kb) and three (3-4 kb) individual SMRT cells. Reads of Insert (ROI, >3 full passes and >90% accuracy settings using the SMRT Portal software), trimmed by 65 nt at both ends to remove Illumina adaptors, were deposited at ENA under run accession numbers ERR966169, ERR966170 and ERR966123 for 1.5kb R, 2.5kb R and S parents respectively.

Illumina data processing

Quality control, mapping and SNP calling of Illumina Inc. (San Diego, CA, USA) paired-end sequences (both RenSeq MiSeq and WGS data) were carried out using tools embedded in The Sainsbury Laboratory (TSL) customized Galaxy platform[28], if not noted otherwise.

Genetic Mapping of *Rpi-amr3*

Pathogen inoculations on leaves of young F_2 plants (F1 SP2271 x SP1102) revealed 99 resistant and 6 susceptible plants (15:1 segregation; χ^2=0.02 p=0.95), suggesting two unlinked *Rpi* genes. These genes were genetically separated by self-pollination of resistant F_2 plants, and four F_3 populations were selected that segregated 3:1 for disease resistance. We hypothesized that resistance is conferred by a NLR gene, and conducted a multiplexed NLR capture experiment on DNA of the resistant (R) and susceptible (S) parents, as well as bulked DNA of the 50 most susceptible F_2 and F_3 plants (bulked susceptible, BS). To position the *Rpi-amr3* resistance locus on the *S. tuberosum* DM[29] reference chromosomes, we performed SNP calling as described by[4] with minor modifications. We *de novo* assembled quality controlled R reads (Q>20, no unambiguous nucleotides allowed) using SPAdes[16] with the multiple k-mer option (command line options: -k 21,33,55,77,99,127 --pe1-1 --pe1-2 –careful). We mapped R, S and BS reads onto the assembled R contigs (using BWA[30] with; default settings), and called polymorphisms using SAM tools[30]. To identify linked polymorphisms, we looked for SNPs between R and S parent (100% alternate allele in S reads mapping results to R assembly) with coverage above 50, and absent (>95% alternate allele) from BS reads mapping to R assembly. Positive contigs were subsequently used in BLAST searches[31] against the DM reference genome (>80% identity >1kb) to identify the most likely position, see Supplementary Table 7 and Supplementary File 1 for marker details. This analysis allowed us to position the underlying resistance gene on Ch 4: 3.5-8.5Mb. This region carries three NLR cluster in the potato and tomato reference genomes, the *R2/Rpi-blb3* cluster and the uncharacterized cluster C17 and C18, see Supplementary Figure 1b for details.

To fine-map the resistance locus, and to separate the three candidate NLR gene clusters (R2/Rpi-blb3, C17 and C18) based on the reference DM, we developed Cleaved Amplified Polymorphic Sequences (CAPS[32]) markers between them. We *de novo* assembled WGS data with CLC Assembly Cell (www.clcbio.com) using default settings, and anchored the assembled contigs to the reference DM genome using BLAST (>80% identity, >2kb) and selected those flanking each of the three clusters. We mapped R and S WGS reads to the selected contigs using BWA with default settings and called homozygous (100%) polymorphisms between them as described in[4]. Polymorphic positions and sequences

around them (5 nt each side) were manually analyzed in Geneious R8 for differential restriction enzyme recognition sites. Primers were designed manually to flank the predicted polymorphic positions on the selected contigs (Supplementary Table 7 and Supplementary File 1 for marker details) using Geneious R8 and amplified from R, S and BS gDNA in 25 µl PCR reactions (35 cycles at 58°C), using homemade Taq polymerase and digested with the appropriate restriction enzyme (New England Biolabs, Inc., Ipswich, MA, USA) for 2 h at the required temperature. Digestion products were visualized on 1.5% agarose gels.

Two markers, WGS_1_4 (3.59 Mb) and WGS_2_4 (8.69 Mb) recombined with resistance, while the remaining markers 'co-segregated with resistance' in screens on the initial 50 BS plants (Supplementary Figure 1 b). To resolve the complex physical NLR cluster structure with more recombination events, we created a larger mapping population by backcrossing the resistant F_1 plant to the susceptible SP2271 (female parent). Eight resistant BC_1F_1 plants (heterozygous at WGS_1_4 and WGS_2_4) were self-pollinated and between 60-100 plants of eight BC_1F_2 populations were screened with *P. infestans* isolate 06_3928A. Two populations (SP3534 and SP3543) segregated 3:1 for resistance and 210 and 195 plants respectively, were phenotyped with 06_3928A. Genotyping identified 41 new recombinants between WGS_1_11 and WGS_2_5, which were further genotyped with marker WGS_2_5, revealing recombinants between the *R2/Rpi-blb3* and C18 clusters (Supplementary Figure 1b and 1c), and confirming that the candidate gene *Rpi-amr3i* is a member of cluster C18 (Supplementary Figure 1c).

PacBio ROI analysis and assembly

All PacBio ROI data were processed and analyzed using Geneious R8 (www.geneious.com). Raw reads were processed using Pacific Biosciences SMRT Portal (http://www.pacificbiosciences.com) to generate ROI reads (>3 full passes and >90% accuracy). No further quality control steps were performed on these reads, except that we trimmed Illumina adapter sequences (65 nt at each end). To analyze enrichment efficiency we used the NLR-specific MAST pipeline on six-frame translations of all ROIs and subsequently the NLR-parser[10,15].

ROI reads were assembled using the native Geneious R8 assembler allowing for 1% mismatches, 1% gaps (no longer than 3 bp) and minimal read overlap (>100 nt with 98%

identity). As a valid assembly we considered contigs supported by at least five reads with minimum two fold average coverage. We found this assembler very useful and user-friendly, and flexible enough to handle PacBio specific read errors that consist of either incorrect bases or short indels. It is a "greedy" algorithm which is similar to that used in multiple sequence alignment and uses a blast-like approach to find best matching reads which are then merged into a contig, followed by fine tuning and heuristics to improve assembly[33].

All contigs were run through the MAST and NLR-parser[10,15] to select those harboring NLR sequences. To validate the assembly we mapped WGS 100bp PE reads to NLR selected contigs using BWA and inspected manually for chimeric assemblies with discordant read mapping using the Savant Genome Browser (www.genomesavant.com). However, we noticed that within regions of low coverage (especially less than 3x – see below), the homopolymeric nucleotide count was not always correct, especially towards the less covered contig ends where correction of sequencing errors was not supported. In these cases, we identified the correct nucleotide count by mapping WGS and MiSeq RenSeq 250bp PE reads using BWA default settings and called homozygous polymorphism as described in[4]. PacBio assembly errors were then corrected manually in Geneious R8, according to the results from MiSeq RenSeq reads mapping.

All assembled and annotated NLR genes from this study received ID's following a convention previously suggested in[4,8]. Susceptible line SP2271 has prefix RDC0008 followed by NLR number; e.g. RDC0008NLR0001. As we discuss only one NLR family in this paper, on the phylogenetic trees we omit RDC0008 prefix to simplify naming.

Comparative analysis between RenSeq PacBio and MiSeq assemblies

While we use MiSeq RenSeq data (250 bp) in this study to identify SNPs in the resistant parent (SP1102) that are depleted in susceptible plants (SP2271), we are in the position to test our claim that this short-read data is not suitable to assemble the repetitive NLR complement. We assembled MiSeq R parent data using SPAdes[16] and CLC (www.clcbio.com) (see **Illumina data processing and Genetic Mapping of *Rpi-amr3*)**, and PacBio ROI with the Geneious R8 assembly option (www.geneious.com). All contigs were subsequently analyzed for the presence of NLR sequences using the MAST and NLR-parser pipeline [15] (Supplementary Table 3). *De novo* assembly of PacBio ROIs resulted in

615 NLR encoding contigs with 322 complete and 293 partial sequences. We used these contigs to test the top-level performance of the two short-read assemblers. SPAdes produced 742 NLR contigs, of which 216 were annotated as complete by our NLR-parser pipeline, whereas CLC generated 124 complete contigs. Local alignments were carried out using blastn to identify SPAdes and CLC contigs that are covered over 99% at an identity 98% to a PacBio assembled contig (main text). The reciprocal search identified four SPAdes and one CLC contig without similar sequence in the PacBio pool. We further searched for the occurrence of chimeric assemblies as contigs that share stretches with over 98% identity with more than one PacBio derived sequence (main text). Interestingly the NLRs assembled from MiSeq data share only 51% (SPAdes) and 66% (CLC) of the target genes, while 49% and 34% are uniquely assembled sequences. The CC-NB-LRR:TIR-NB-LRR ratio amongst uniquely identified NLRs was 2.7:1 for CLC and 4.4:1 for SPADES, not showing a specific bias or enrichment. We computed the additional flanking region for the 5' sequence of all contigs as the distance between nucleotide 1 and the start of the first NLR specific motif (NLR-parser result).

Analysis of ROI reads and *de novo* assembled contigs accuracy.

We selected the C18 cluster contigs from the resistant and susceptible parental lines, to analyze the error rate in ROIs and in *de novo* assembly of ROI data. To correct errors in the assembly we mapped Illumina WGS reads and MiSeq-RenSeq using BWA with default settings, and then filtered with SAM tools for correctly mapped read pairs. Mapping data were visualized in Savant Genome Browser and in total 276,111bp were analyzed for errors. We considered only positions with 20-40x coverage of WGS data and 150-500x coverage of MiSeq-RenSeq data (available in coding sequences mostly), while higher coverage (less than 5% of analyzed contigs) suggested mismapping from close paralogues or presence of transposons. Positions that had 95% alternative allele compared to reference, were considered errors. We then used this information to manually correct errors in the contigs using Geneious software.

These manually corrected contigs were further used to analyze accuracy of ROI reads used for *de novo* assembly of C18. We aligned ROI using blastn and calculated accuracy of ROI from percentage of pairwise identity.

Cloning of the candidate ORFs and transient complementation assay

To facilitate cloning of the candidate *Rpi-amr3* coding sequences we designed PCR primers flanking each predicted ORF from six selected PacBio contigs. All primers were supplemented with specific 5' and 3' extensions to make them compatible with custom USER expression vectors[34] used in this study (Supplementary Table 8). Candidate genes were PCR amplified from R parent gDNA in 50µl PCR reactions (35 cycles with annealing at 60°C and 5 min extension at 68°C) using Platinum Pfx DNA Polymerase (Life Technologies, Carlsbad, CA, USA). 30 ng of purified PCR product was hybridized with 30ng pICSLUS0003::35S (see Supplementary File 6 for an annotated vector sequence) in the presence of 1 µl of USER enzyme mix (New England Biolabs, Inc., MA, USA). All constructs were verified by DNA sequencing. Plasmids containing the candidate genes were transformed into *Agrobacterium* strain Agl1 and used for transient complementation assays as well as stable potato transformations. To create a genomic construct of *Rpi-amr3i* under its native regulatory elements, we PCR amplified the whole contig (5352 bp, Figure 2c) that was assembled from PacBio reads. Although the ends were from single reads, we were able to amplify the full contig sequence. PCR amplicons were hybridized into USER-vector pICSLUS0001 (see Supplementary File 7 for annotated vector sequence), lacking 35S promoter and OCS terminator.

Transient complementation assays were carried out as described by[34,35] and photographs were taken at 6 dpi.

Stable transgenic potato plants

We created stable transgenic plants with constructs carrying *Rpi-amr3* candidates under 35S and *Rpi-amr3i* under the control of its native regulatory elements in the diploid homozygous Solynta Research line nr 26 (www.solynta.com) as described[36]. For *Rpi-amr3i*, 14 (native promoter) and 27 (35S promoter) plants were positively tested for transgene presence (PCR) and further phenotyped with *P. infestans* isolates 88069 and 06_3928A. We observed 5 (35%) and 5 (19%) fully resistant plants with native and 35S promoter constructs, respectively, while the remaining plants had intermediate levels of resistance. We further measured levels of *Rpi-amr3i* expression in plants with native regulatory elements (Supplementary Figure 8) and observed a link between number of

mRNA copies and level of resistance. Plants with gene expression similar to R parent showed full resistance (lines 3-7), while remaining had intermediate phenotypes. One fully resistant plant of each type (number 7 for *Rpi-amr3i* under native regulatory elements) was selected and tested with multiple *P. infestans* isolates, see Figure 3b, Supplementary Table 5 and Supplementary Figure 7. For the remaining *Rpi-amr3* candidate constructs, we PCR screened 10-15 transgenic plants and further phenotyped all selected (7-10 per construct) in DLA assays with 88069 and 06_3928A *P. infestans* isolates. All plants were susceptible (data not shown).

Phylogenetic analyses

We constructed a phylogenetic tree of SP2271 NLRs as described by[8,10] with minor modifications. *De novo* assembled contigs with PacBio ROI data containing multiple NLRs were manually split into single NLR genes (Supplementary File 4) in Geneious R8 using ORF-finding function and positional MAST NLR-motif information. For the tree construction we used only complete NLRs determined by the NLR-parser[15] based on the MAST result (Supplementary File 5). To identify the NB-ARC domain sequences used for the phylogenetic analysis, amino acid sequences of the NB-ARC domain of reference *R* genes (reported in Supplementary Table 9), were used to search in a Blastx analysis with an expected value of <1e−3 and predicted NB-ARC domains were verified with MAST NLR-motif[10,15] (Supplementary File 3). All the sequences were aligned using ClustalW 1.74[37]. Evolutionary analyses were conducted using MEGA6[38] using Poisson (G+I) model with 100 bootstraps.

Detection of fused protein domains

To detect fusions between a canonical NLR sequence and a different protein, we took advantage of the assembled NLR flanking regions and analysed all 2271 NLRs (Supplementary File 4) with AUGUSTUS[39] to predict potential protein coding sequences. Translated protein sequences were subjected to BLAST Conserved Domains search using the online tool http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi [40] and CDD v3.14 database with default settings (e-value 0.01 and suppression of weak overlapping hits). NLR loci with potential fusion domains were manually annotated in Geneious R8 using the ORF-finding function, AUGUSTUS gene prediction information and positional MAST NLR-

motif information (Supplementary File 5) to predict potential start and stop codons. Manually verified ORFs were again subjected to BLAST Conserved Domains search as described above to verify fused domains (Supplementary Table 4). We observed potential additional fused domains adjacent to annotated ORFs, however, due to lack of RNA sequencing data we do not report them in this paper as fused domains. Detailed analysis of NLR complement of these accessions will be reported elsewhere.

List of Supplementary Files.

Supplementary File 1. Fasta-file containing all marker sequences used to position and map the *Rpi-amr3* gene.

Supplementary File 2. Fasta file containing *Contig_7.*

Supplementary File 3. Fasta-file containing NB-ARC domains of all genes used for phylogenetic tree construction.

Supplementary File 4. Fasta file containing complete NLRs of SP2271 accession annotated with MAST search output.

Supplementary File 5. MAST search result for SP2271 NLRs loci.

Supplementary File 6. Annotated sequence of the pICSLUS0003 vector in GeneBank format.

Supplementary File 7. Annotated sequence of the pICSLUS0001 vector in GeneBank format.

References

1. Jones, J. D. G. *et al.* Elevating crop disease resistance with cloned genes. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **369,** 20130087–20130087 (2014).
2. Haverkort, A. J. *et al.* Societal Costs of Late Blight in Potato and Prospects of Durable Resistance Through Cisgenic Modification. *Potato Res.* **51,** 47–57 (2008).
3. Rodewald, J. & Trognitz, B. Solanumresistance genes against Phytophthora infestansand their corresponding avirulence genes. *Molecular Plant Pathology* **14,** 740–757 (2013).
4. Jupe, F. *et al.* Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J.* **76,** 530–544 (2013).
5. Dangl, J. L. & Jones, J. D. Plant pathogens and integrated defence responses to infection. *Nature* **411,** 826–833 (2001).
6. Jones, J. D. G. & Dangl, J. L. The plant immune system. *Nature* **444,** 323–329 (2006).
7. Dangl, J. L., Horvath, D. M. & Staskawicz, B. J. Pivoting the plant immune system from dissection to deployment. *Science* **341,** 746–751 (2013).
8. Andolfo, G. *et al.* Defining the full tomato NB-LRR resistance gene repertoire using genomic and cDNA RenSeq. **14,** 1–12 (2014).

9. Lebecka, R. Host–pathogen interaction between Phytophthora infestans and Solanum nigrum, S. villosum, and S. scabrum. *Eur J Plant Pathol* **120,** 233–240 (2007).
10. Jupe, F. *et al.* Identification and localisation of the NB-LRR gene family within the potato genome. *BMC Genomics* **13,** 75 (2012).
11. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323,** 133–138 (2009).
12. Sharon, D., Tilgner, H., Grubert, F. & Snyder, M. A single-molecule long-read survey of the human transcriptome. *Nat Biotechnol* **31,** 1009–1014 (2013).
13. Wang, M. *et al.* PacBio-LITS: a large-insert targeted sequencing method for characterization of human disease-associated chromosomal structural variations. *BMC Genomics* **16,** 214 (2015).
14. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucl. Acids Res.* **37,** W202–8 (2009).
15. Steuernagel, B., Jupe, F., Witek, K., Jones, J. D. G. & Wulff, B. B. H. NLR-parser: rapid annotation of plant NLR complements. *Bioinformatics* **31,** 1665–1667 (2015).
16. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19,** 455–477 (2012).
17. Jiao, X. A Benchmark Study on Error Assessment and Quality Control of CCS Reads Derived from the PacBio RS. *J Data Mining in Genom Proteomics* **04,** 1–12 (2013).
18. Cesari, S., Bernoux, M., Moncuquet, P., Kroj, T. & Dodds, P. N. A novel conserved mechanism for plant NLR protein pairs: the 'integrated decoy' hypothesis. *Front Plant Sci* **5,** 606 (2014).
19. Narusaka, M. *et al.* RRS1 and RPS4 provide a dual Resistance-gene system against fungal and bacterial pathogens. *Plant J.* **60,** 218–226 (2009).
20. Sarris, P. F. *et al.* A Plant Immune Receptor Detects Pathogen Effectors that Target WRKY Transcription Factors. *Cell* **161,** 1089–1100 (2015).
21. Rallapalli, G. *et al.* EXPRSS: an Illumina based high-throughput expression-profiling method to reveal transcriptional dynamics. *BMC Genomics* **15,** 341 (2014).
22. Manoko, M. L. K., van den Berg, R. G., Feron, R. M. C., van der Weerden, G. M. & Mariani, C. AFLP markers support separation of Solanum nodiflorum from Solanum americanum sensu stricto (Solanaceae). *Plant Syst Evol* **267,** 1–11 (2007).
23. Baebler, S. *et al.* Salicylic acid is an indispensable component of the Ny-1 resistance-gene-mediated response against Potato virus Y infection in potato. *J. Exp. Bot.* ert447 (2014). doi:10.1093/jxb/ert447
24. Śliwka, J. *et al.* The novel, major locus Rpi-phu1 for late blight resistance maps to potato chromosome IX and is not correlated with long vegetation period. *Theor Appl Genet* **113,** 685–695 (2006).
25. Foster, S. J. *et al.* Rpi-vnt1.1, a Tm-2(2) homolog from Solanum venturii, confers resistance to potato late blight. *Mol. Plant Microbe Interact.* **22,** 589–600 (2009).
26. Caten, C. E. & Jinks, J. L. Spontaneous variability of single isolates of Phytophthora infestans. I. Cultural variation. *Canadian Journal of Botany* (1968). doi:10.1139/b68-055
27. Lebecka, R. Inheritance of resistance in Solanum nigrum to Phytophthora infestans. *Eur J Plant Pathol* **124,** 345–348 (2009).
28. MacLean, D. & Kamoun, S. Big data in small places. *Nat Biotechnol* **30,** 33–34 (2012).
29. Xu, X. *et al.* Genome sequence and analysis of the tuber crop potato. *Nature* **475,** 189–195 (2011).
30. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
31. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
32. Konieczny, A. & Ausubel, F. M. A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. *Plant J.* **4,** 403–410 (1993).
33. Geneious 8.0. 1–213 (2014).
34. Lokossou, A. A., Park, T. & van Arkel, G. Exploiting knowledge of R/Avr genes to rapidly clone a new LZ-NBS-LRR family of late blight resistance genes from potato linkage group IV.

Molecular Plant- … **22,** 630–641 (2009).
35. Saunders, D. G. O. *et al.* Host protein BSL1 associates with Phytophthora infestans RXLR effector AVR2 and the Solanum demissum Immune receptor R2 to mediate disease resistance. *Plant Cell* **24,** 3420–3434 (2012).
36. Kumar, A., Taylor, M. A., Arif, S. A. M. & Davies, H. V. Potato plants expressing antisense and sense S-adenosylmethionine decarboxylase (SAMDC) transgenes show altered levels of polyamines and ethylene: antisense plants display abnormal phenotypes. *The Plant Journal* **9,** 147–158 (1996).
37. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22,** 4673–4680 (1994).
38. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30,** 2725–2729 (2013).
39. Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucl. Acids Res.* **32,** W309–W312 (2004).
40. Marchler-Bauer, A. *et al.* CDD: NCBI's conserved domain database. *Nucl. Acids Res.* **43,** D222–6 (2015).