The *Solanum americanum* pangenome and effectoromics reveal new resistance genes against potato late blight

Xiao Lin[1*#], Yuxin Jia[2*], Robert Heal[1], Maxim Prokchorchik[3†], Maria Sindalovskaya[1], Andrea Olave-Achury[1], Moffat Makechemu[1], Sebastian Fairhead[1], Azka Noureen[1], Jung Heo[5], Kamil Witek[1], Matthew Smoker[1], Jodie Taylor[1], Ram-Krishna Shrestha[1], Yoonyoung Lee[3], Chunzhi Zhang[2], Soon Ju Park[5,6], Kee Hoon Sohn[3,4#], Sanwen Huang[2#], Jonathan D. G. Jones[1#]

[1]The Sainsbury Laboratory, University of East Anglia, Norwich Research Park, Norwich, NR4 7UH, UK

[2]Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Area, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China

[3]Department of Life Sciences, Pohang University of Science and Technology, Pohang, Republic of Korea

[4]School of Interdisciplinary Bioscience and Bioengineering, Pohang University of Science and Technology, Pohang 37673, Republic of Korea

[5]Department of Biological Science and Institute of Basic Science, Wonkwang University, Iksan 54538, Republic of Korea

[6]Division of Applied Life Sciences, Gyeongsang National University, Jinju, Republic of Korea

*These authors contributed equally

†Present address: Plant pathology group, The Institute of Crop Science and Resource Conservation, University of Bonn, Nussalle 9, 53115, Bonn, Germany

#Corresponding authors:
Xiao Lin (xiao.lin@tsl.ac.uk)
Kee Hoon Sohn (khsohn@postech.ac.kr)
Sanwen Huang (huangsanwen@caas.cn)
Jonathan D. G. Jones (jonathan.jones@tsl.ac.uk)

1 **Abstract**

2

3 Late blight caused by the oomycete pathogen *Phytophthora infestans* continues to cause major

4 worldwide losses in potato and tomato. Most accessions of *Solanum americanum*, a globally

5 distributed, wild Solanaceae plant, are highly resistant to late blight. We generated high-quality

6 reference genomes of four *S. americanum* accessions, re-sequenced 52 accessions, and we

7 defined variation in the NLR immune receptor genes (the *S. americanum* NLRome). We further

8 screened for variation in recognition of ~315 *P. infestans* RXLR effectors in 52 *S. americanum*

9 accessions. Using these genotypic and phenotypic data, we cloned three novel NLR-encoding

10 genes *Rpi-amr4*, *Rpi-amr16* and *Rpi-amr17*, and determined their corresponding RXLR

11 effector genes *Avramr4* (*PITG_22825*), *Avramr16* (*PITG_02860*) and *Avramr17* (*PITG_04373*)

12 from *P. infestans*. These genomic resources and methodology will support efforts to convert

13 potato into a "nonhost" of late blight and can be applied to diseases of other crops.

14

15

Introduction

Potato is the most consumed non-grain crop worldwide. However, pests and diseases reduce global yields by ~17% [1]. Potato late blight, which is caused by the oomycete pathogen *Phytophthora infestans* [2], triggered the Irish famine in the 1840s, and is still the most damaging disease for global potato production [1].

Plant immunity depends on pathogen recognition by both cell-surface pattern recognition receptors (PRRs) and intracellular immune receptors. Most cloned plant *Resistance (R)* genes encode intracellular nucleotide-binding domain, leucine-rich repeat (NLR) proteins. Many *R* genes against *P. infestans* (*Rpi* genes) were cloned from wild relatives of potato species, such as *R2*, *R3*, *R8*, *Rpi-blb1*, *Rpi-blb2* and *Rpi-vnt1* from *Solanum demissum*, *S. bulbocastanum* and *S. venturii* [3-10]. However, most cloned *Rpi* genes have been overcome by the fast-evolving pathogen.

NLR proteins recognize pathogen effectors, either directly or indirectly. *P. infestans* effectors carry a signal peptide and an RXLR (Arg-X-Leu-Arg, X represents any amino acid)-EER (Glu-Glu-Arg) motif. In the *P. infestans* reference genome (strain T30-4), 563 RXLR effectors were predicted, enabling screens for recognition of these effectors ("effectoromics") in various plant genetic backgrounds [11,12].

Reference genome sequences of potato, tomato, eggplant and pepper have been determined [13-16]. Phased, chromosome-level genome assemblies of heterozygous diploid and tetraploid potatoes are also available [17-19]. Pan-genome studies of crop plants including potato have also emerged, that shed light on the extensive genetic variation in these species [20-23]. To reduce the genomic complexity and sequencing costs, resistance gene enrichment sequencing (RenSeq) and RLP/K enrichment sequencing (RLP/KSeq) technologies were developed to sequence plant *NLR* and *PRR* gene repertoires [24-26]. These methods led to many important applications such as AgRenSeq and the defining of the pan-NLRome of *Arabidopsis* [27,28].

Diploid *Solanum americanum* and its closely related hexaploid descendant *S. nigrum* are both highly resistant to LB. Previously, our group cloned *Rpi-amr1* and *Rpi-amr3* from different resistant *S. americanum* accessions, and their functional homologs were cloned from *S. nigrum*

49   [25,29,30]. Both *Rpi-amr1* and *Rpi-amr3* confer late blight resistance in potato. The cognate

50   effectors AVRamr1 and AVRamr3 were identified by pathogen enrichment sequencing

51   (PenSeq), and effectoromics screening [30,31].

52

53   Here, we sequenced and assembled four high-quality genomes of *S. americanum* and re-

54   sequenced 52 accessions that were collected from seed banks around the world. By manually

55   annotated *NLR* genes we defined the Pan-NLRome of *S. americanum* in combination with

56   additional 16 SMRT-RenSeq assemblies. We also screened 315 *P. infestans* RXLR effectors

57   in 52 *S. americanum* accessions by transient expression through *Agrobacterium* infiltration.

58   These genomic resources and functional data led to the rapid identification of three novel *Rpi*

59   genes *Rpi-amr4*, *Rpi-amr16* and *Rpi-amr17* that are responsible for AVRamr4 (PITG_22825),

60   AVRamr16 (PITG_02860) and AVRamr17 (PITG_04373) recognition. This study unveils a

61   non-host effector-triggered immunity (ETI) interaction between *S. americanum* and *P.*

62   *infestans*, that will enable us to clone more *Rpi* genes from the *S. americanum* gene pool. In

63   combination with novel technologies for potato genome design and engineering [32], this will

64   help develop potatoes with durable late blight resistance.

65

66

67

68 ## Results

69

70 **Genome assembly and gene model prediction of *Solanum americanum***

71

72 *S. americanum* is a globally distributed Solanaceae species, which is resistant to many

73 pathogens including *Phytophthora infestans* and *Ralstonia solanacearum*[25,29,33]. Four *S.*

74 *americanum* accessions SP1102, SP2271, SP2273 and SP2275 were selected for sequencing

75 based on their variation in resistance to late blight (Supplementary Fig. 1). We first generated

76 Illumina paired-end reads to analyze these genomes. K-mer based analysis indicated that the

77 genome sizes of these four *S. americanum* accessions were 1.15-1.31 Gb, with 0.05% to 0.35%

78 heterozygosity (Table 1, Supplementary Fig. 2). We then generated an average of 29.5 Gb (~26

79 fold) PacBio high-fidelity reads for SP1102 and SP2271, respectively, and assembled the reads

80 into 298 and 568 contigs with N50s 82.9 Mb and 55.2 Mb. The ONT platform was used for

81 sequencing SP2273 and SP2275 resulting in ~81.1 Gb (~62 fold) and ~114.5 Gb ONT reads

82 (~95 fold), respectively. The noisy reads were self-corrected and assembled into 304 and 539

83 contigs with N50s 9.6 Mb and 4.8 Mb, respectively. To generate chromosome level assemblies,

84 we further generated ~86.5 Gb, ~81.8 Gb and ~54.8 Gb Hi-C data for SP1102, SP2271 and

85 SP2273, and anchored the contigs into 12 pseudomolecules (Supplementary Fig. 3). The

86 completeness of these assemblies was estimated to be ~98.4% (single-copy and duplicated,

87 Supplementary Fig. 4a) by BUSCO, which indicates the high quality of genome assembly.

88

89   To annotate gene models of SP1102 and SP2271, we generated an average of ~80 Gb

90 transcriptome data (150 bp, paired-end reads) for each accession, which derived from young

91 seedlings, roots, stems, leaves, flowers and fruits. The EVidenceModeler pipeline was applied

92 to annotate the gene models by integrating *ab initio* prediction, homology-based annotation

93 and transcriptome evidence. To predict gene models in SP2273 and SP2275 genomes, we

94 deployed the GeMoMa pipeline to integrate homology-based annotation and transcriptome

95 evidence. In summary, we predicted an average of 34,193 gene models with an average of 98.1%

96 (single-copy and duplicated, Supplementary Fig. 4b, Table 1) BUSCO evaluation for each *S.*

97 *americanum* genome.

98

99 **Genome evolution of *S. americanum***

100   To investigate the evolution of *S. americanum* genomes, we clustered the representative protein

101   sequences from 15 genomes, comprising four *S. americanum*, four potatoes, three tomatoes,

102   four additional Solanaceae species and an outgroup species of *Arabidopsis thaliana*, into

103   33,115 orthogroups, from which we further identified 1,363 single-copy orthogroups. Using a

104   supermatrix method, we inferred the phylogenetic relationships and estimated the divergence

105   time of the 15 genomes. The species tree topology suggests that *S. americanum* is a sister

106   species to the common ancestor of potato and tomato, and they diverged ~14.1 million years

107   ago (Ma, 95% highest posterior density interval: 11.7-17.2 Ma, Fig. 1a), which is consistent

108   with the former report based on plastid sequences [34].

109

110   Chromosome rearrangement (CR) is an important evolutionary process [35]. The reference-grade

111   genome assemblies enabled us to explore *S. americanum* chromosome evolution. Both protein

112   coding gene-based chromosome syntenic analysis and alignment based on genomic sequences

113   were carried out and generated similar results (Fig. 1b, Supplementary Fig. 5). We observed

114   45 large CRs (> 1 Mb in size), comprising 26 inversion and 19 inter-chromosome translocation

115   events, between *S. americanum* and potato genomes (Fig. 1b). In contrast, 67 large CRs (30

116   inversions and 37 inter-chromosome translocations) were found between *S. americanum* and

117   eggplant (Fig. 1b). Notably, CRs were not evenly distributed across the genome. No CR was

118   identified on chromosome 2 between *S. americanum* and potato, while as many as 11 CRs

119   occurred on chromosome 11.

120

121   **Structural variation between *S. americanum* genomes**

122   Structural variations (SVs), including insertions, deletions, duplications, inversions and

123   translocations, cause and maintain phenotypic diversity [36]. The chromosome assemblies of

124   three *S. americanum* genomes enabled analysis of large SVs (> 1 Mb in size). Using SP1102

125   as the reference, we identified 56 large SVs in SP2271 (Supplementary Fig. 6), impacting ~256

126   Mb of the reference genome. However, only 14 large SVs were identified in SP2273, covering

127   ~54 Mb of the reference genome (Supplementary Fig. 6). The huge differences of SV number

128   between *S. americanum* genomes shed light on the complex history of *S. americanum* species

129   evolution.

130

131   To further characterize SVs within *S. americanum*, we used the SP1102 as the reference and

132   identified 60,849, 23,672 and 18,492 SVs (>40 bp in size) in SP2271, SP2273 and SP2275,

133   ranging from 27.9 to 102.0 Mb in total length (Supplementary Fig. 7). Most SVs (~54.5% on

134 average) located in putative gene regulatory regions (5 kb upstream, downstream or in introns),

135 followed by 44.0% and 1.5% overlapped with intergenic region and exon. SVs might contribute

136 to gene expression variation. We identified 1,837 differentially expressed genes (DEGs, Fold

137 change >= 2 or <= 0.5, P-value < 0.05) in leaves between SP1102 and SP2271, of which, 1,084

138 DEGs might associate with SVs, as exemplified by *sp1102chr11_nlr_6*, an NLR gene whose

139 expression might be impaired by a 286 bp deletion in the promoter region in leaves of SP2271

140 (Supplementary Fig. 8).

141

142 Previous reports indicate that NLR genes in plants are highly diversified between individuals

143 even in the same species [37]. We found that sequence diversity in NLR region is significantly

144 higher than that of non-NLR region (Supplementary Fig. 9a). As an example, the *Rpi-amr3*

145 locus varied greatly among *S. americanum* genomes (Supplementary Fig. 9b), suggesting the

146 NLRome based on high-quality genomes is required for deep understanding the non-host

147 resistance of *S. americanum* species.

148

149

150 **Defining the *S. americanum* pan-NLR repertoire**

151 To understand *NLR* gene diversity, a phylogenetic tree was generated by using the NB-ARC

152 domain of the NLR proteins from SP1102 (Fig. 2a), and the physical positions of these *NLRs*

153 in SP1102 genome were visualized in the physical map (Fig. 2b). We found 71% *NLRs* in the

154 SP1102 genome are in clusters and the rest are singletons (Fig. 2c). Due to the complexity of

155 *NLR* gene clusters, most automatic annotation pipelines produce incorrect gene models[37]. To

156 generate better gene models of the *NLR* genes from *S. americanum* genome. Here, we first used

157 NLR annotator [38] to identify the putative *NLR* genes from SP1102, SP2271, SP2273 and

158 SP2275 reference genomes, revealing 641, 669, 608 and 616 *NLRs* respectively (Fig. 2d). By

159 adopting a definition of *NLR* cluster (3 *NLRs* within 200 kb interval) [39]. To predict the precise

160 *NLR* gene models, we manually annotated 528, 579 and 524 *NLRs* from SP1102, SP2271 and

161 SP2273 genomes by incorporating NLR-annotator results and cDNA sequence data (Fig. 2d,

162 Supplementary dataset 1-3). Next, we examined presence/absence (P/A) polymorphism of

163 NLRs among *S. americanum* accessions. NLRs from 3 other *S. americanum* reference genomes,

164 and from 16 SMRT-RenSeq assemblies, were compared with the SP1102 NLRs. P/A

165 polymorphism is shown in Fig. 2a. We also inspected the expression level of SP1102 NLRs by

166 re-analysing cDNA RenSeq data [25]. The transcripts per million *NLR* transcripts were visualized

167 by a heatmap (Fig. 2a). Many well-known *NLR* genes are relatively highly expressed, such as

168   *ADR1*, *NRG1*, *Rpi-amr3*, *NRC4a*, *NRC2* and *NRC3* (Fig. 2f). Many Solanaceae CNLs require

169   NRC helpers to execute their function and 50% of the *S. americanum* NLRs lie within the NRC

170   superclade [40] (Fig. 2a). To investigate the NRC family, we generated another phylogenetic tree

171   for NRCs. We found NRC1, NRC2, NRC3, NRC4a and NRC6 homologs, two NRC5a copies.

172   Interestingly, the copy number of NRC4b homologs has expanded in the *S. americanum*

173   genome (Fig. 2e). Rpi-amr3 and Rpi-amr1 require NRC2/3/4 and NRC2/3, respectively [29,30].

174   However, *NRC1* is missing in *N. benthamiana* [41]. To test if NRC1 can enable *Rpi-amr1/3*

175   function, we cloned the *SaNRC1* from SP1102 and showed *SaNRC1* enables *Rpi-amr3* but not

176   *Rpi-amr1* function in *N. benthamiana* nrc2/3/4 knockout line (Supplementary Fig. 10). In

177   summary, we generated an NLRome of ~20 *S. americanum* accessions, and manually

178   annotated the *NLRs* from 3 reference genomes. This resource is important for the investigation

179   *NLR* gene evolution, and functional studies of ETI in *S. americanum* and other Solanaceae

180   species.

181

182   **The ETI landscape of the *S. americanum* and *P. infestans* interaction**

183

184   There are 563 predicted RXLR effectors from *P. infestans* reference genome T30-4. In this

185   study, we showed that there are ~550 *NLR* genes in the best-studied *S. americanum* reference

186   genomes (Fig. 2d). To reveal one-to-one effector-receptor interactions and clone more *Rpi*

187   genes, we screened ~315 RXLR effectors on 52 *S. americanum* accessions (Fig. 3). Based on

188   the cDNA PenSeq data, most if not all expressed RXLR effectors were included in this

189   screening [31]. We found that 5 effectors triggered HR on most *S. americanum* accessions (≥

190   50), including effectors from AVRblb2 family. 185 effectors did not trigger HR in any *S.*

191   *americanum* accessions, 71 effectors can only be recognized by less than 5 *S. americanum*

192   accessions. Notably, the rest 54 effectors showed differential recognition by different *S.*

193   *americanum* accessions. AVRamr1 and AVRamr3 are also widely recognized by different *S.*

194   *americanum* accessions (Fig. 3). The 4 reference accessions SP2271, SP2275, SP1102 and

195   SP2273 can recognize 25, 18, 30 and 30 RXLR effectors respectively, of which 5, 3, 7 and 9

196   effectors are specifically recognized in each accession (Supplementary Fig. 11). Notably,

197   accession SP2271 is susceptible to *P. infestans* in the detached leaf assay (DLA), yet resistant

198   to late blight in the field. As expected, SP2271 does not recognize AVRamr1 and AVRamr3.

199   We found premature stop codons in both *Rpi-amr1* and *Rpi-amr3* SP2271 homologs.

200   Intriguingly, 25 RXLR effectors trigger HR in SP2271, these RXLR effector recognitions

201 might contribute to the SP2271 field resistance to late blight. Conceivably, effector
202 recognitions from other *S. americanum* accessions other than SP2271 contribute to stronger
203 resistance against late blight. These results reveal the ETI landscape of *S. americanum* against
204 the late blight pathogen and enable identification of novel *Rpi* genes.

205

206 **Resequencing of *S. americanum* accessions**

207

208 To understand the genetic polymorphism of all *S. americanum* accessions in our collection, we
209 performed PCR-free, 150 bp paired-end sequencing for 52 *S. americanum* accessions at 10x
210 coverage. The reads were mapped to the reference *S. americanum* genome SP1102, and SNPs
211 were called and filtered for further analysis. We constructed a phylogenetic tree using all SNPs
212 on genes. Eggplant, potato, and tomato were used as outgroups (Supplementary Fig. 12a). The
213 structural and inbreeding coefficient values were also analysed (Supplementary Fig. 12b and
214 c). The 52 accessions can be clearly assigned into 4 groups (Supplementary Fig. 12a). Group
215 1 contains SP2271, which is susceptible to late blight in lab conditions. All the 6 members in
216 this group lack *Rpi-amr1* and *Rpi-amr3* based on the effectoromics screening (Fig. 3 and
217 Supplementary Table 2). SP2275 and SP1102 are in Group 2, SP2273 is in Group 3. For Group
218 4, no reference genome is available, but we have SMRT RenSeq assemblies for several of these
219 accessions. Surprisingly, four accessions SP2303, SP2310, SP3393 and SP3051 are not closely
220 related to other groups and are highly heterozygous (Supplementary Fig. 12c), suggesting they
221 may be polyploid species like *S. nigrum*. Two other accessions, SP3052 and SP3376 are also
222 not closely related to the four *S. americanum* groups and might belong to another Solanaceae
223 species. In summary, we re-sequenced 52 accessions and reveal their phylogenetic relationship
224 and population structure. These data could be used for genome-wide association studies
225 (GWAS), and molecular marker development.

226

227 **Cloning of *Rpi-amr4* by GWAS and linkage analysis**

228

229 In our effectoromics screening, we found an effector PITG_22825 that triggers HR in 28/50 *S.*
230 *americanum* accessions (Fig. 3, Supplementary Table 2), including SP1102 and SP2271 but
231 not SP2298 (Fig. 4a). PITG_22825 is an RXLR effector with a signal peptide, RQLR and EER
232 motifs followed by the effector domain (Fig. 4a). This effector had not received attention prior
233 to our cDNA-PenSeq study [31]. To map the gene conferring its recognition, a GWAS analysis
234 was performed to associate this effector responsiveness with linked SNPs. A strong signal was

235   identified on an NLR singleton that locates on Chr01 of SP1102 (Fig. 4b, Fig. 2b). This encodes

236   a CNL that belongs to the Rpi-amr3 phylogenetic clade CNL-13 (Fig. 2a), but not physically

237   in the Rpi-amr3 gene cluster on Chr11 (Fig. 2b). This likely explains another weaker GWAS

238   signal on the *Rpi-amr3* cluster of Chr11 (Fig 2b). Based on cDNA RenSeq data from SP1102,

239   the corresponding *NLR* gene carries an extra exon compared to *Rpi-amr3* (Fig. 4b). To verify

240   this GWAS signal, we performed a bulked segregant analysis and resistance gene enrichment

241   sequencing (BSA-RenSeq) in a segregating F2 population of SP2298 x SP2271

242   (Supplementary Fig. 13). The PITG_22825 responsive gene from SP2271 was mapped to the

243   same position.

244

245   To test the gene function, the ORF of the candidate genes from SP2271 and SP1102 were PCR

246   amplified and cloned into an over-expression binary vector with 35S promoter and Ocs

247   terminator and introduced into *Agrobacterium*. The candidate genes were heterologously

248   expressed alone, or co-expressed with PITG_22825 or AVRamr3-GFP (negative control) in *N.*

249   *benthamiana*. Rpi-amr3-FLAG was used as a positive control. We found that the SP2271 allele

250   (*Rpi-amr4-2271* hereafter) is auto-active in *N. benthamiana*, but when co-expressed with

251   PITG_22825, the HR was faster and stronger compared to the control (Fig. 4c). In contrast, the

252   SP1102 allele (*Rpi-amr4-1102* hereafter) is not auto-active in *N. benthamiana*, HR was

253   triggered when *Rpi-amr4-1102* was co-expressed with PITG_22825, but not with AVRamr3-

254   GFP (Fig. 4c). Rpi-amr3 does recognize AVRamr3-GFP but not recognize PITG_22825 but

255   does recognize AVRamr3-GFP (Fig. 4c). There are only 3 amino-acid differences between

256   *Rpi-amr4-1102* and *Rpi-amr4-2271*, and these differences might cause the auto-activity of the

257   SP2271 allele (Supplementary Fig. 14). We also found *Rpi-amr4* is conserved in the

258   PITG_22825 responsive accessions (Fig. 4f). To verify the function of *Rpi-amr4*, we designed

259   two specific CRISPR-Cas9 sgRNAs and mutated *Rpi-amr4-2271* in SP2271. In total, sixteen

260   CRISPR-Cas9 knockout lines were generated (Fig. 4e and Supplementary Table 4). The

261   genotype and phenotype of two lines are shown in Fig 4e. Wild-type SP2271 can recognize

262   PITG_22825 but the *Rpi-amr4*-mutated lines cannot. The HR phenotype could be

263   complemented when *Rpi-amr4-1102* was co-expressed with PITG_22825 in the knockout lines

264   (Fig. 4e). Therefore, we conclude that *Rpi-amr4* encodes the PITG_22825-recognizing

265   immune receptor, and that *PITG_22825* is *Avramr4*.

266

267   To test if *Rpi-amr4* confers late blight resistance, we transiently expressed *Rpi-amr4-1102* in

268   *N. benthamiana* leaves and inoculated with *P. infestans* zoospores. *Rpi-amr3* was used as a

269   positive control and non-functional *Rpi-amr3a* from SP1102 was used as a negative control

270   (Fig. 4d). This assay shows that *Rpi-amr4-1102* confers resistance against *P. infestans* isolate

271   T30-4. However, the resistance is weaker compared to Rpi-amr3 (Fig. 4d). We also generated

272   stable *Rpi-amr4-1102* transgenic *N. benthamiana* lines. As expected, the T0 transgenic plants

273   gain the PITG_22825 recognition capacity, and they are resistant to two *P. infestans* isolates

274   T30-4 and 88069 (Supplementary Fig. 15).

275

276   In summary, we successfully cloned a novel *Rpi* gene *Rpi-amr4* from *S. americanum* and

277   defined its cognate effector gene *Avramr4* (*PITG_22825*). *Rpi-amr4* confers late blight

278   resistance and may serve as resource for producing late blight resistant potatoes.

279

280

281   **Cloning of *Rpi-amr16* and *Rpi-amr17***

282

283   Although *Rpi-amr4* could be identified using GWAS approach, those effectors only recognized

284   by a few *S. americanum* accessions did not enable a clear GWAS signal. We therefore deployed

285   BSA-RenSeq for cloning two such effector responsive genes.

286

287   PITG_02860 (Fig 5a) targets a host protein NRL1, attenuates plant immunity and increases

288   pathogen virulence [42]. We found PITG_02860 triggers HR in 5/52 *S. americanum* accessions,

289   including SP2271. We tested the F2 population of SP2271 (PITG_02860 responsive, R) x

290   SP2272 (PITG_02860 non-responsive, NR), and found the recognition of PITG_02860

291   segregated according to a 3:1 ratio (104 R and 34 NR, χ^2 (1, $N = 138$) = 0.00966, $p = 0.92169$)

292   (Fig. 5b). gDNA was isolated from the R and NR bulks respectively, RenSeq libraries were

293   constructed and sequenced (250 bp paired-end reads). The reads were mapped to SP2271, and

294   SNPs were called and filtered. Most filtered SNPs located within a 1 Mb region on the top of

295   Chr04 of SP2271 (Fig. 5c). To develop molecular markers for map-based cloning. The whole

296   genome sequencing reads of SP2272 were mapped to SP2271 genome, then SCAR markers

297   were designed to specifically amplify the responsive allele from SP2271, and the 34 NR

298   progenies were then used for genotyping. The candidate gene was mapped to a 295 kb region

299   between markers S42 and S36. Seven *NLRs* reside within the mapping interval, all from the *R3*

300   family (Fig. 5c, Fig. 2a and b). To test these candidate genes, the ORFs from 4 candidate genes

301   (*nlr13*, *nlr14*, *nlr16*, *nlr17*) were cloned into a binary vector under control of 35S promoter and

302   Ocs terminator and transformed into *Agrobacterium* for transient expression. The candidate

303   genes were expressed alone or with PITG_02860 or AVRamr3 in *N. benthamiana* and *N.*

304   *tabacum*. NLR16 and NLR17 are auto-active in *N. benthamian*, but we found co-expression of

305   NLR16 and PITG_02860 activate HR in *N. tabacum* (Fig. 5d). To verify this finding, we

306   generated nlr16 knockout SP2271 lines by CRISPR-Cas9 system. As expected, the knockout

307   lines lost recognition of PITG_02860 (Supplementary Fig. 16). Therefore, we conclude NLR16

308   (Rpi-amr16 hereafter) is the immune receptor for PITG_02860 (AVRamr16).

309

310   A sequence corresponding to PITG_04373 was designed and synthesized based on the PacBio

311   and cDNA pathogen enrichment sequencing (PenSeq) data [31] (Fig. 5e). PITG_04373 triggers

312   HR in only 3/50 *S. americanum* accessions including SP2300, which carries both functional

313   *Rpi-amr1* and *Rpi-amr3* (Fig. 3). To clone the corresponding *Rpi* gene, we first phenotyped a

314   BC1 population of SP2271(NR) x SP2300 (R) by transiently expressed PITG_04373 by

315   agroinfiltration. The BC1 population segregates for PITG_04373 responsiveness with a 1:1

316   ratio (198 R : 182 NR, χ^2 (1, $N = 380$) = 0.67368, $p = 0.41177$) (Fig. 5f). The DNA from R or

317   NR progenies were bulked for BSA-RenSeq and most linked SNPs mapped to SP2271 Chr10.

318   SCAR markers were designed by the same strategy and the candidate gene was ultimately

319   mapped to a 1.447 Mb interval with 8 genes in SP2271 (Fig. 5g). Most candidates belong to

320   the *Rpi-chc1* family, except a *R1* homolog (Fig. 5g). In the absence of a reference genome for

321   SP2300, we generated a SMRT RenSeq assembly for the SP2300 NLRome [29]. The SMRT-

322   RenSeq contigs were mapped to this region of the SP2271 genome, and candidate genes from

323   SP2300 were cloned into a vector with 35S promoter and Ocs terminator for transient assays.

324   Five candidate genes were tested (C18.1, C18.2, C127, C168 and C829) by agroinfiltration

325   with or without the effector. We found that the candidate gene C168 (Rpi-amr17 hereafter) can

326   specifically recognize PITG_04373 after transient expression in *N. benthamiana* (Fig. 5h).

327   Therefore, we conclude that *Rpi-amr17* encodes the PITG_04373 immune receptor from

328   SP2300, and that PITG_04373 is AVRamr17.

329

330   To verify the function of *Rpi-amr17* in SP2300, we designed two sgRNAs targeting *Rpi-amr17*

331   (Supplementary Fig. 17d). SP2300 also carries functional *Rpi-amr1* and *Rpi-amr3* homologs,

332   to test if SP2300 carries additional *Rpi* genes, we also designed one sgRNAs for *Rpi-amr1-*

333   *2300* and *Rpi-amr3*, respectively (Supplementary Fig. 17b and c). Forty transgenic SP2300

334   knockout lines were generated and tested by the three effectors (PpAVRamr1, AVRamr3 and

335   PITG_04373), and we found 13 lines that lack the recognition for all three effectors; two such

336   lines are shown (Supplementary Fig. 17e). We genotyped 2 knockout lines SLJ25603#3 and

337   SLJ25603#17, and confirmed that all *Rpi-amr1-2300*, *Rpi-amr3-2300* and *Rpi-amr17* were

338   knocked out (Supplementary Fig. 1b-d). We also co-expressed *Rpi-amr17* with *PITG_04373*,

339   however the HR phenotype was not restored in this genotype. Interestingly, the triple knockout

340   lines show some susceptibility to *P. infestans* (Supplementary Fig. 17f) compared to the wild-

341   type SP2300, but it is still more resistant compared to SP2271, suggesting that there are

342   additional *Rpi* genes in SP2300.

343

344   In summary, by using BSA-RenSeq, SMRT-RenSeq and map-based cloning strategies, we

345   successfully cloned two novel *Rpi* genes *Rpi-amr16* and *Rpi-amr17* that confer recognition of

346   two RXLR effectors PITG_02860 and PITG_04373.

Discussion

Solanum is the largest genus of the Solanaceae family, comprising more than 1500 species, including many important crop plants such as potato, tomato and eggplant for which extensive genome sequence data are available. However, less is known about wild *Solanum* species. The *Solanum nigrum* complex is composed of many species with different ploidy levels, including *S. nigrum* (6x), *S. scabrum* (6x), *S. villosum* (4x) and *S. americanum* (2x). Some are regarded as weeds but others are consumed as food and medicine in various countries [43]. Importantly, these species carry valuable natural sources of genetic resistance to diseases such as potato late blight and bacterial wilt [25,29,33]. Variation for resistance to other pests and diseases is being investigated but when defining corresponding genes, reference genomes are highly advantageous. In this study, we sequenced and assembled four *S. americanum* genomes, three with Hi-C data that enable chromosome level assembly. These assemblies of *S. americanum* enabled us to gain insights into the interspecies structural variation, genomic evolution, and speciation of cultivated and wild Solanaceae. This understanding of the intraspecies structural variation can guide generation of mapping populations, and will facilitate the understanding of its genetic architecture for resistance to multiple diseases.

In this study, we also generated multi-omics datasets including resequencing data of 52 *S. americanum* accessions, RNAseq, SMRT-RenSeq and cDNA-RenSeq data. The resequencing data enabled us to understand the phylogeny and population structure of these accessions, and indicate that some accessions may belong to other species. Importantly, molecular markers can be rapidly developed by using the reference genomes and the resequencing data for any mapping populations. These data also enabled us to build a *S. americanum* Pan-NLRome to study the evolution and function of the *NLR* genes in *S. americanum*. Automatic annotation of NLR gene models remains a challenge, and therefore we manually annotated all *NLR* genes in the *S. americanum* reference genomes SP1102, SP2271 and SP2273. These gene models will facilitate curation of the *NLR* genes from other Solanaceae species, and will help the cloning and characterization of NLR immune receptors. Notably, consistent with previous findings [40], nearly half of the *S. americanum NLR* genes are in the *NRC* superclade. We found that an *NRC1* homolog that is lacking in the *N. benthamiana* genome, supports the function of Rpi-amr3 but not Rpi-amr1 (Supplementary Fig. 10), thus expanding our knowledge of the NRC network in Solanaceae. We also found that *NRC4b* clade has expanded in *S. americanum* compared to *N.*

380   *benthamiana* and tomato [41]. Whether these *NRC4* paralogs are functionally redundant or

381   divergent requires further study.

382

383   Potato late blight triggered the Irish famine in the 1840s, and remains a global challenge that

384   greatly constrains potato production. To understand effector-triggered immunity (ETI) of *S.*

385   *americanum* to *P. infestans*, we used "effectoromics" to dissect the ETI interactions between

386   them. Using transient expression of effectors, we generated a matrix of 315 RXLR effectors ×

387   52 *S. americanum* accessions. As expected, most RXLR effectors cannot be recognized by

388   most *S. americanum* accessions. Interestingly, AVRamr1 (36/52) and AVRamr3 (43/52)

389   recognitions are widely distributed in *S. americanum* accessions, indicating that *Rpi-amr1* and

390   *Rpi-amr3* play an important role in the late blight resistance of *S. americanum*. This finding is

391   consistent with the conclusions of a Pan-genome ETI study of the interaction between

392   *Arabidopsis* and effectors from diverse strains of *P. syringae* [44]. Some effectors induce cell

393   death in all *S. americanum* accessions, such as effectors in the AVRblb2 family (Fig. 3). This

394   observation is consistent with previous findings that AVRblb2 (PexRD39/40) induces cell

395   death in all tested wild potato species, but not in *N. benthamiana* [45,46]. Conceivably, this non-

396   specific cell death might be the result of the virulence activities of AVRblb2. Some resistant

397   accessions lack AVRamr1 and AVRamr3 recognition, and thus are valuable sources of novel

398   *Rpi* genes. However, isolation of novel *Rpi* genes from resistant accessions that carry *Rpi-amr1*

399   and *Rpi-amr3* requires their removal by either genetic segregation or gene knockout.

400

401   In the current study, this challenge was successfully addressed using effectoromics. Using

402   GWAS or BSA-RenSeq approaches, we cloned three novel *Rpi* genes: *Rpi-amr4*, *Rpi-amr16*

403   and *Rpi-amr17*, as well as their cognate *Avr* genes *Avramr4* (*PITG_22825*), *Avramr16*

404   (*PITG_02860*) and *Avramr17* (*PITG_04373*) (Fig. 4 and Fig. 5). Using the transient expression

405   assay in *N. benthamiana*, we showed that *Rpi-amr4* contributes to *P. infestans* resistance. Using

406   CRISPR-Cas9 in *S. americanum*, we generated *Rpi-amr4* and *Rpi-amr16* knockout SP2271

407   lines respectively, as well as the *Rpi-amr1*/*Rpi-amr3*/*Rpi-amr17* triple knockout SP2300 lines.

408   These enabled us to validate gene function in *S. americanum* and create materials for cloning

409   new *Rpi* genes. *Avramr4*, *Avramr16* and *Avramr17* are present in all tested *P. infestans* isolates,

410   and are up-regulated upon infection [31,47]. Unlike some RXLR effectors that belong to

411   superfamilies such as *Avr2* and *Avrchc1*[12,48], *Avramr4*, *Avramr16* and *Avramr17* are all

412   singletons. AVRamr16 (PITG_02860) was reported to promote host susceptibility by targeting

413   a host protein NLR1 [42], and the virulence function(s) and host targets of AVRamr4 and

414   AVRamr17 remain to be discovered.

415

416   *Rpi-amr4-1102* confers quantitative and incomplete resistance in *N. benthamiana* transient

417   assays compared to the full resistance conferred by *Rpi-amr3*. *Rpi-amr16* was isolated from

418   SP2271- a late blight susceptible accession in detached leaf assays (DLAs). These newly

419   isolated *Rpi* genes thus likely confer weaker late blight resistance compared to the previously

420   cloned *Rpi-amr1* and *Rpi-amr3*, despite showing strong HR when transiently co-expressed

421   with their cognate recognized effectors. However, many effectors are suppressors of host

422   immunity, notably AVRcap1 that can attenuate the function of the helper NLRs NRC2 and

423   NRC3 [49], and this may explain why some *Rpi* genes are weaker than others. Resistance based

424   on recognition of a single effector can be easily overcome by mutations or silencing, as shown

425   by *Avrvnt1* [50,51]. *Rpi-amr4*, *Rpi-amr16* and *Rpi-amr17* belong to *Rpi-amr3*, *R3* and *Rpi-chc1*

426   clades, respectively (Fig. 2a) and *R3* and *Rpi-chc1* are not NRC-dependent. Since resistance

427   is a quantitative phenotype, as illustrated by the fact that most *R* genes are semi-dominant

428   (resistance in heterozygotes is weaker than in homozygotes), and ETI functions at least in part

429   by potentiating PTI [52], then more recognition capacity should strengthen resistance, even if

430   some recognition capacities confer weak resistance alone. Even weak resistance genes can thus

431   be useful, and the three newly cloned *Rpi* genes could be stacked with *Rpi-vnt1*, *Rpi-amr1* and

432   *Rpi-amr3* to quantitatively increase defence activation [52], minimize the risk of *Rpi* gene

433   breakdown, and ultimately help to turn potato into a nonhost plant of late blight as is *S.*

434   *americanum*. We anticipate that additional effector-detecting immune receptor genes can be

435   rapidly cloned using this pipeline.

436

437   Both Rpi-amr1 and Rpi-amr3 can recognize the AVRamr1 and AVRamr3 homologs from other

438   *Phytophthora* pathogens, and *Rpi-amr3* confers resistance against *P. parasitica* and *P.*

439   *palmivora* by detecting their conserved AVRamr3 homologs [29,30]. Conceivably, Rpi-amr4, Rpi-

440   amr16 and Rpi-amr17 might also recognize the effectors from other important *Phytophthora*

441   pathogens and confer resistance against multiple diseases.

442

443   In summary, our study provides valuable genomic and genetics tools that accelerate the path

444   to understanding and achieving durable resistance against potato late blight and other plant

445   diseases, and make *S. americanum* an excellent model plant to study molecular plant-microbe

446   interactions and plant immunity.

Author contributions:

X.L. Y.J., K.H.S, S.H. and J.D.G.J. conceived and designed the project. X.L. and Y.J. wrote the first draft with inputs from all the authors. X.L., Y.J., K.H.S., S.H. and J.D.G.J. reviewed and edited the manuscript. Y.J., M.P., X.L. and R.K.S. performed the bioinformatics analyses. X.L. and M.M., performed the effectoromics screening. X.L., R.H., M.S., A.C.O.A., S.F. and A.N. contributed to clone and characterize *Rpi-amr4*, *Rpi-amr16* and *Rpi-amr17*. M.Smoker and J.T. performed the plant transformation. K.W., Y.L., C.Z., S.J.P., K.H.S., S.H. and J.D.G.J. contributed resources.

Acknowledgements

474

Competing interests:

476   The authors declare no competing interests.

References

1. Savary, S. *et al.* The global burden of pathogens and pests on major food crops. *Nat Ecol Evol* **3,** 430–439 (2019).

2. Fry, W. *Phytophthora infestans*: the plant (and *R* gene) destroyer. *Molecular Plant Pathology* **9,** 385–402 (2008).

3. Park, T.-H. *et al.* The late blight resistance locus *Rpi-bib3* from *Solanum bulbocastanum* belongs to a major late blight *R* gene cluster on chromosome 4 of potato. *MPMI* **18,** 722–729 (2005).

4. Huang, S. *et al.* Comparative genomics enabled the isolation of the *R3a* late blight resistance gene in potato. *The Plant Journal* **42,** 251–261 (2005).

5. Vossen, J. H. *et al.* The *Solanum demissum R8* late blight resistance gene is an *Sw-5* homologue that has been deployed worldwide in late blight resistant varieties. *Theor Appl Genet* **129,** 1785–1796 (2016).

6. Song, J. *et al.* Gene *RB* cloned from *Solanum bulbocastanum* confers broad spectrum resistance to potato late blight. *Proceedings of the National Academy of Sciences* **100,** 9128–9133 (2003).

7. van der Vossen, E. *et al.* An ancient *R* gene from the wild potato species *Solanum bulbocastanum* confers broad-spectrum resistance to *Phytophthora infestans* in cultivated potato and tomato. *The Plant Journal* **36,** 867–882 (2003).

8. Vossen, E. A. G. *et al.* The *Rpi-blb2* gene from S*olanum bulbocastanum* is an *Mi-1* gene homolog conferring broad-spectrum late blight resistance in potato. *The Plant Journal* **44,** 208–222 (2005).

9. Foster, S. J. *et al. Rpi-vnt1.1*, a *Tm-22* homolog from *Solanum venturii*, confers resistance to potato late blight. *MPMI* **22,** 589–600 (2009).

10. Pel, M. A. *et al.* Mapping and cloning of late blight resistance genes from *Solanum venturii* using an interspecific candidate gene approach. *MPMI* **22,** 601–615 (2009).

11. Haas, B. J. *et al.* Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* **461,** 393–398 (2009).

12. Vleeshouwers, V. G. A. A. *et al.* Understanding and exploiting late blight resistance in the age of effectors. *Annu. Rev. Phytopathol.* **49,** 507–531 (2011).

13. Xu, X. *et al.* Genome sequence and analysis of the tuber crop potato. *Nature* **475,** 189–195 (2011).

14. Sato, S. *et al.* The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485,** 635–641 (2012).

15. Wei, Q. *et al.* A high-quality chromosome-level genome assembly reveals genetics for important traits in eggplant. *Horticulture Research* **7,** 153 (2020).

16. Kim, S. *et al.* Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nature Genetics* **46,** 270–278 (2014).

17. Zhou, Q. *et al.* Haplotype-resolved genome analyses of a heterozygous diploid potato. *Nature Genetics* **52,** 1018–1023 (2020).

18. Sun, H. *et al.* Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. *Nature Genetics* **54,** 342–348 (2022).

19. Hoopes, G. *et al.* Phased, chromosome-scale genome assemblies of tetraploid potato reveal a complex genome, transcriptome, and predicted proteome landscape underpinning genetic diversity. *Molecular Plant* 1–17 (2022). doi:10.1016/j.molp.2022.01.003

20. Zhao, Q. *et al.* Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nature Genetics* **50,** 278–284 (2018).

21. Liu, Y. *et al.* Pan-genome of wild and cultivated soybeans. *Cell* **182,** 162–176.e13 (2020).

22. Gao, L. *et al.* The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nature Genetics* **51,** 1044–1051 (2019).

23. Tang, D. *et al.* Genome evolution and diversity of wild and cultivated potatoes. *Nature* **606,** 535–541 (2022).

24. 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. *The Plant Journal* **76,** 530–544 (2013).

25. Witek, K. *et al.* Accelerated cloning of a potato late blight–resistance gene using RenSeq and SMRT sequencing. *Nature Biotechnology* **34,** 656–660 (2016).

26. Lin, X. *et al.* RLP/K enrichment sequencing; a novel method to identify receptor-like protein (RLP) and receptor-like kinase (RLK) genes. *New Phytol* **227,** 1264–1276 (2020).

27. Arora, S. *et al.* Resistance gene cloning from a wild crop relative by sequence capture and association genetics. *Nature Biotechnology* **37,** 139–143 (2019).

28. Van de Weyer, A.-L. *et al.* A species-wide inventory of *NLR* genes and alleles in *Arabidopsis thaliana*. *Cell* **178,** 1260–1272.e14 (2019).

29. Witek, K. *et al.* A complex resistance locus in *Solanum americanum* recognizes a conserved *Phytophthora* effector. *NPLANTS* **7,** 198–208 (2021).

30. Lin, X. *et al.* A potato late blight resistance gene protects against multiple *Phytophthora* species by recognizing a broadly conserved RXLR-WY effector. *Molecular Plant* 1–41 (2022). doi:10.1016/j.molp.2022.07.012

31. Lin, X. *et al.* Identification of Avramr1 from *Phytophthora infestans* using long read and cDNA pathogen-enrichment sequencing (PenSeq). *Molecular Plant Pathology* **18,** 547–11 (2020).

32. Zhang, C. *et al.* Genome design of hybrid potato. *Cell* **184,** 3873–3883.e12 (2021).

33. Moon, H. *et al.* Identification of RipAZ1 as an avirulence determinant of *Ralstonia solanacearum* in *Solanum americanum*. *Molecular Plant Pathology* **22,** 317–333 (2021).

34. Särkinen, T., Bohs, L., Olmstead, R. G. & Knapp, S. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evol Biol* **13,** 1–15 (2013).

35. Rieseberg, L.2001. Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution* **16,** (2001).

36. Yuan, Y., Bayer, P. E., Batley, J. & Edwards, D. Current status of structural variation studies in plants. *Plant Biotechnology J* **19,** 2153–2163 (2021).

37. Barragan, A. C. & Weigel, D. Plant NLR diversity: the known unknowns of pan-NLRomes. *The Plant Cell* **33,** 814–831 (2021).

38. Steuernagel, B. *et al.* The NLR-Annotator tool enables annotation of the intracellular immune receptor repertoire. *PLANT PHYSIOLOGY* **183,** 468–482 (2020).

39. Holub, E. B. The arms race is ancient history in Arabidopsis, the wildflower. *Nat Rev Genet* **2,** 516–527 (2001).

40. Wu, C.-H. *et al.* NLR network mediates immunity to diverse plant pathogens. *Proceedings of the National Academy of Sciences* **114,** 8113–8118 (2017).

41. Adachi, H. *et al.* An atypical NLR protein modulates the NRC immune receptor network. *BioRxiv* (2021). doi:10.1101/2021.11.15.468391v1

42. Yang, L. *et al.* Potato NPH3/RPT2-like protein StNRL1, targeted by a *Phytophthora infestans* RXLR effector, is a susceptibility factor. *PLANT PHYSIOLOGY* **171,** 645–657 (2016).

575   43.   Sangija, F., Martin, H. & Matemu, A. African nightshades (*Solanum nigrum* complex):
576   The potential contribution to human nutrition and livelihoods in sub-Saharan Africa.
577   *Compr Rev Food Sci Food Saf* **20,** 3284–3318 (2021).
578   44.   Laflamme, B. *et al.* The pan-genome effector-triggered immunity landscape of a host-
579   pathogen interaction. *Science* **367,** 763–768 (2020).
580   45.   Rietman, H. Putting the *Phytophthora infestans* genome sequence at work: multiple
581   novel avirulence and potato resistance gene candidates revealed. (Doctoral thesis, 2011).
582   46.   Oh, S.-K. *et al.* In planta expression screens of *Phytophthora infestans* RXLR effectors
583   reveal diverse phenotypes, including activation of the *Solanum bulbocastanum* disease
584   resistance protein *Rpi-blb2*. *The Plant Cell* **21,** 2928–2947 (2009).
585   47.   Cooke, D. E. L. *et al.* Genome analyses of an aggressive and invasive lineage of the Irish
586   potato famine pathogen. *PLoS Pathog* **8,** e1002940 (2012).
587   48.   Monino-Lopez, D. *et al.* Allelic variants of the NLR protein Rpi-chc1 differentially
588   recognize members of the *Phytophthora infestans* PexRD12/31 effector superfamily
589   through the leucine-rich repeat domain. *The Plant Journal* **107,** 182–197 (2021).
590   49.   Derevnina, L. *et al.* Plant pathogens convergently evolved to counteract redundant nodes
591   of an NLR immune receptor network. *PLOS Biology* **19,** e3001136 (2021).
592   50.   Vleeshouwers, V. G. A. A. & Oliver, R. P. Effectors as tools in disease resistance
593   breeding against biotrophic, hemibiotrophic, and necrotrophic plant pathogens. *MPMI*
594   **27,** 196–206 (2014).
595   51.   Pais, M. *et al.* Gene expression polymorphism underpins evasion of host immunity in
596   an asexual lineage of the Irish potato famine pathogen. *BMC Evol Biol* **18,** 1–11 (2018).
597   52.   Ngou, B. P. M., Ahn, H. K., Ding, P. & Jones, J. D. G. Mutual potentiation of plant
598   immunity by cell-surface and intracellular receptors. *Nature* **592,** 110–115 (2021).
599
600
601
602

603 **Figures and Tables:**

604 **Table 1:**

| Genome features	SP1102	SP2271	SP2273	SP2275
Sequencing method	CCS, Hi-C	CCS, Hi-C	ONT, Illumina, Hi-C	ONT, Illumina
Estimated genome size	1.15 Gb	1.21 Gb	1.31 Gb	1.21 Gb
Heterozygosity	0.34%	0.35%	0.05%	0.06%
Total assembly	1.07 Gb	1.13 Gb	1.02 Gb	1.02 Gb
Contigs	299	568	294	550
Contig N50	82.9 Mb	55.2 Mb	10.3 Mb	4.8 Mb
Contig L50	7	8	30	66
Chromosomes	12	12	12	-
Anchor rate	98.43%	97.60%	99.31%	-
BUSCO for assembly	98.40%	98.30%	98.40%	98.40%
Gene models	35,654	36,073	31,976	33,051
Transcripts	35,654	36,073	42,391	44,027
BUSCO for annotation	97.70%	97.90%	98.60%	98.00%

605

606

607
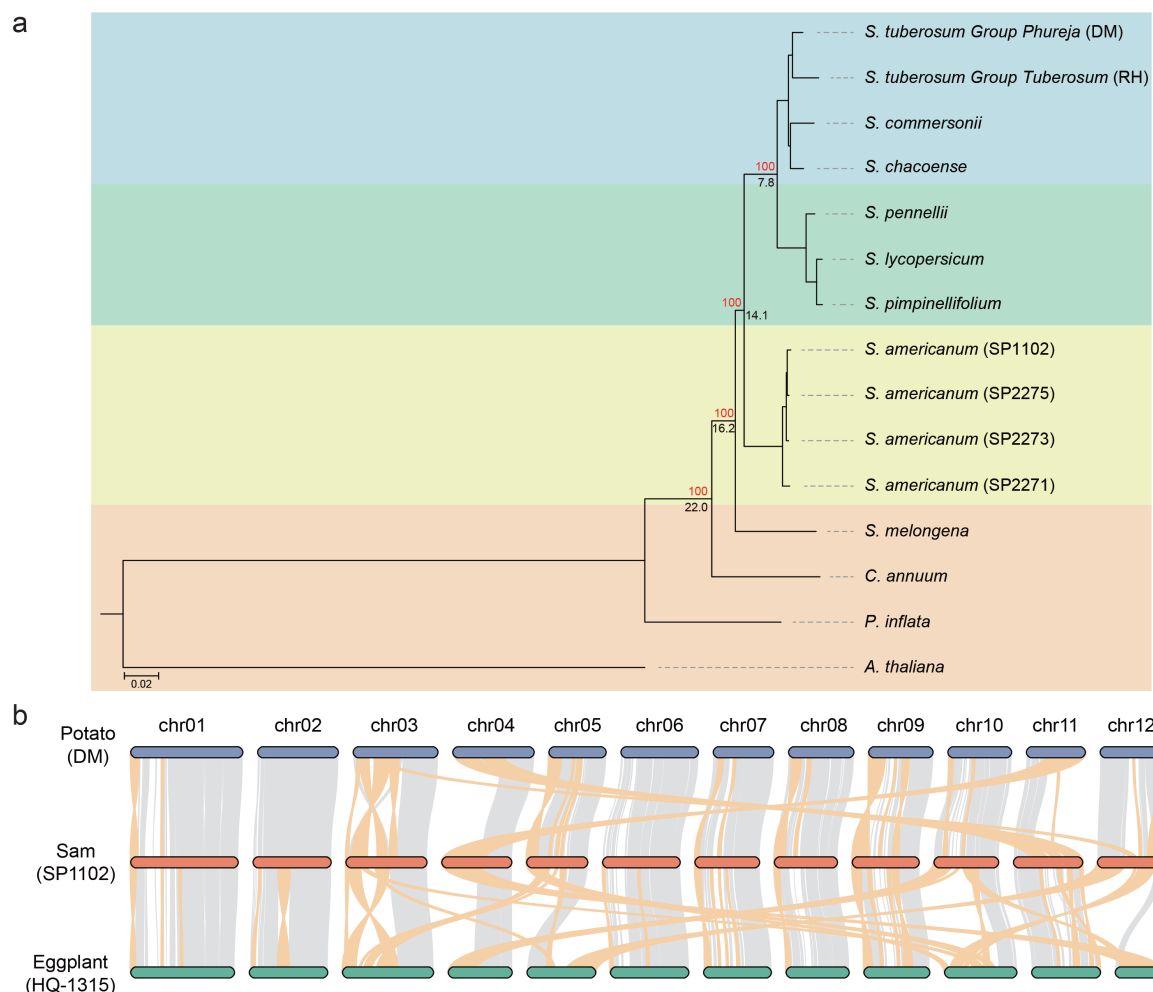
Figure 1. Genome evolution of *Solanum americanum*.

(a) Phylogenetic relationship of *Solanum americanum* and neighbouring species. The red number indicates bootstrap of each node. The black number denotes estimated divergence time (million years ago).
(b) Genome synteny of *Solanum americanum*, potato and eggplant. Ribbons between chromosomes show syntenic regions; Large chromosome rearrangements (> 1 Mb in size) are marked as orange.
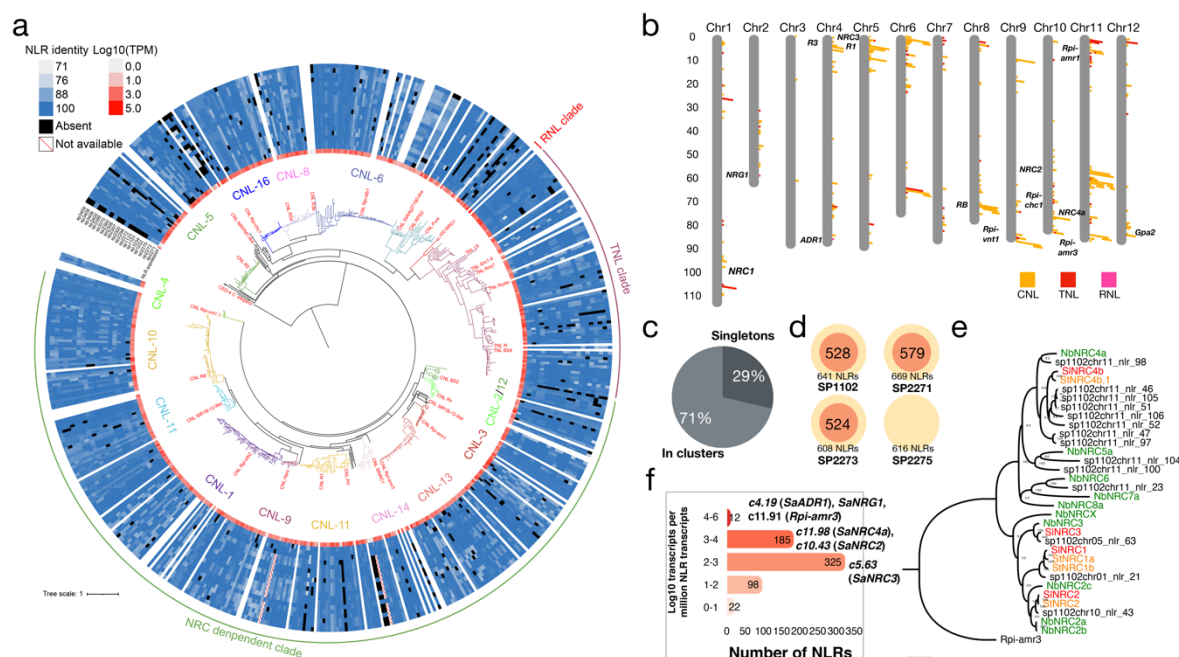
Figure 2. Pan-NLRome of *Solanum americanum*

(a) The NB-ARC domains of *S. americanum* SP1102 were predicted by NLR-annotator and used to generate the maximum-likelihood tree by IQ-TREE with JTT+F+R9 model. Known NLR proteins from Solanaceae species are included (highlighted by red). The NLRs are classified into different groups based on a previous report [29]. The RNL, TNL and NRC-dependent CNL clade are shown. CED-4 from *Caenorhabditis elegans* were used as the outgroup. The expression profile is shown by a heatmap (white to red) based on the cDNA RenSeq data of SP1102. The P/A polymorphism of NLRs from the three other *S. americanum* genomes and SMRT-RenSeq assemblies of 16 additional accessions are shown by the heatmap (white to blue), the absent NLRs are shown by black blocks.

(b) The physical map of *NLR* genes in SP1102 genome. CNLs are shown by yellow blocks, TNLs are red, and RNLs are pink blocks. Some functionally characterized *NLR* clusters are noted on this map.

(c) The proportion of NLR singletons and NLRs in clusters.

(d) Number of manually curated *NLR* genes, and the *NLR* number that predicted by NLR-annotator (manual curation of SP2275 was not performed).

(e) Phylogeny of the NRC helper NLR family. The NRC homologs from potato, tomato and *N. benthamiana* are marked in yellow, red and green respectively. The *NRC* genes from *S. americanum* are black.

(f) The log10 transcript per million NLR transcripts are classified into 5 groups, and some homologs of known *R* genes are noted. The NLR IDs are shown in Supplementary Table 1.

Figure 3. ETI landscape of *S. americanum* and *P. infestans*

315 RXLR effectors were transiently expressed in 52 *S. americanum* accessions. The HR index (strong HR = 2, weak HR = 1, no HR = 0, NA = not available) was used for the heatmap. These effectors were screened on *N. benthamiana*[30], their recognitions are included in this heatmap. Empty PVX vector was used as negative control, co-expression of Rpi-amr3::HF and AVRamr3::GFP was used as positive control. The *S. americanum* accessions were ordered based on the phylogenetic tree, SP3400 is not included in this tree. The *S. americanum* accessions were classified into 4 groups (grey or yellow shading). The 4 reference accessions are marked by red arrows. The effectors were ordered based on the total HR index. For each effector, the numbers of responsive *S. americanum* accessions are visualized by a bar chart on the top of the heatmap. For each *S. americanum* accession, the numbers of recognized effectors are visualized by a bar chart on the right of the heatmap. Some RXLR effectors previously characterised or mentioned in this study are indicated by grey arrows.
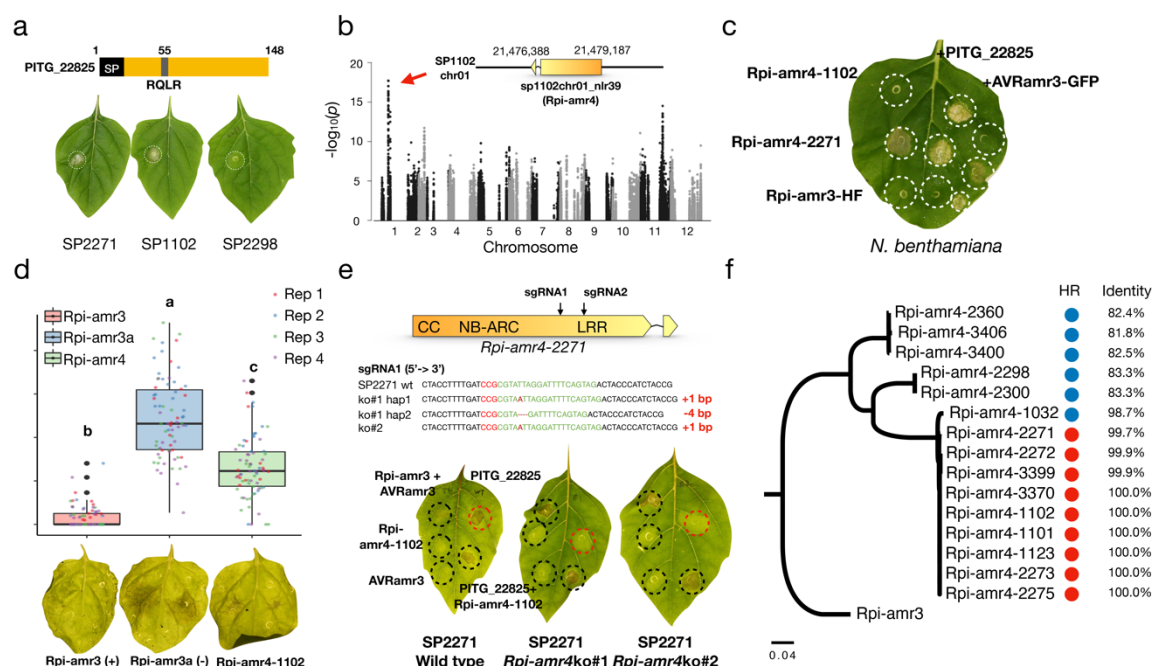
Figure 4. Identification and characterisation of *Rpi-amr4* that recognizes PITG_22825

(a) PITG_22825 is an RXLR effector with a signal peptide (SP), an RQLR-EER motif and an effector domain. The effector without SP was cloned into a vector under control of 35S promoter and transformed into *Agrobacterium tumefaciens* strain GV3101-pMP90 for transient expression through agroinfiltration, and triggers cell death on *S. americanum* SP2271 and SP1102, but not SP2298.

(b) Manhattan plot of the GWAS of PITG_22825 recognition. Resequencing reads from 52 *S. am* accessions were mapped to the *S. americanum* SP1102 genome, the SNPs on *NLR* genes were called and filtered for the GWAS. SNPs significantly associated with PITG_22825 recognition are located on the NLR singleton sp1102chr01_nlr39 (red arrow).

(c) HR assay of candidate genes. *Rpi-amr4-1102* and *Rpi-amr4-2271* were cloned into a vector under control of 35S promoter, then transformed into *Agrobacterium tumefaciens* strain GV3101-pMP90. They were expressed alone, or co-expressed with either 35S::PITG_22825 or 35S::AVRamr3-GFP constructs in *N. benthamiana* leaves. Rpi-amr4-2271 is auto-active in *N. benthamiana*, but when co-expressed with PITG_22825, the HR was stronger. Rpi-amr4-1102 specifically recognize PITG_22825 but not AVRamr3. Rpi-amr3 was used as a control. OD_{600} = 0.5. Four leaves from two plants were used for each experiment, and three biological replicates were performed with same results.

(d) Detached leaf assay (DLA) of *Rpi-amr4*. *Rpi-amr4* (green), *Rpi-amr3* (positive control, red) and *Rpi-amr3a* (a non-functional *Rpi-amr3* paralogue, negative control, blue) were transiently expressed in *N. benthamiana*, OD_{600} = 0.5. The zoospores (300 zoospores in a 10 μL droplet) from *P. infestans* strain T30-4 were used to inoculate the leaves 1 day after the infiltration. A calliper was used to measure the lesion size at 6 days after inoculation. Four biological replicates were performed, all datapoints (74 datapoints/treatment) were visualized as a box plot using R. The outliers are indicated by black dots. Statistical differences were analysed by one-way ANOVA with Tukey's HSD test ($P < 0.001$). The representative leaves of this assay are shown.

(e) *Rpi-amr4* knockout lines lose the PITG_22825 recognition capacity. Two sgRNAs (black arrows) were designed to knockout *Rpi-amr4*-2271 in SP2271. The genotype of two knockout lines is shown. The two knockout lines failed to recognize PITG_22825, but the HR could be complemented by co-expressed PITG_22825 with *Rpi-amr4*-1102. Wild-type SP2271 was used as control plants. *Rpi-amr3* and AVRamr3 were used as control for the agroinfiltration. OD_{600} = 0.5, four-weeks SP2271 plants were used for agroinfiltration, the photos were taken 4 days after infiltration.

(f) Phylogeny of Rpi-amr4 homologs in different *S. americanum* accessions. Rpi-amr3 was used as an outgroup. The PITG_22825-mediated HR are shown by red (HR) or blue (no HR) circles. Percent identity of amino acid sequence relative to Rpi-amr4-1102 is shown. The scale bars represent the number of amino acid substitutions per site.
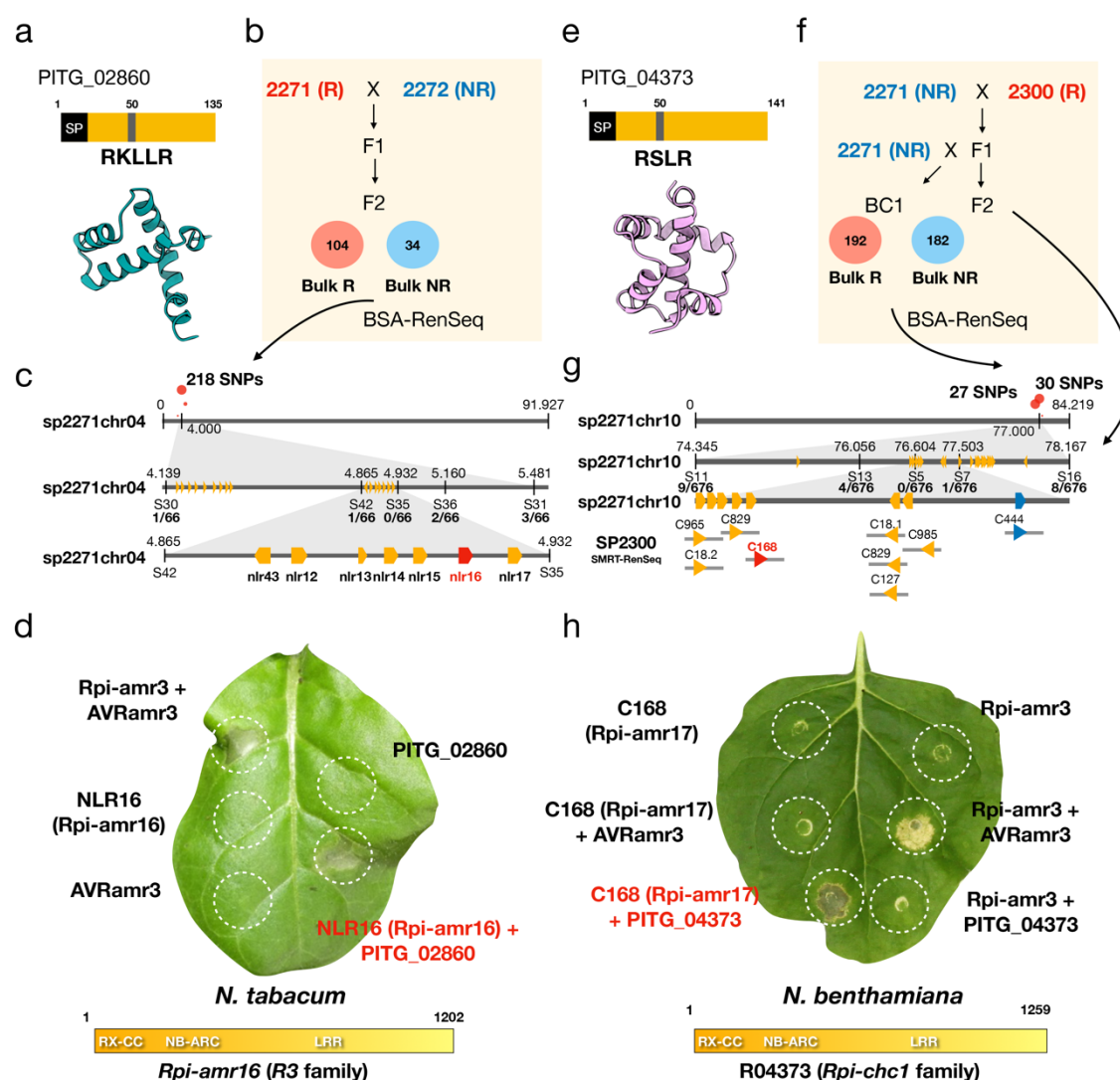
Figure 5. Identification of *Rpi-amr16* and *Rpi-amr17* that recognize PITG_02860 and PITG_04373.

(a) PITG_02860 is an RXLR effector from *P. infestans*, an illustration and AlphaFold predicted structure are shown.

(b) An F2 population was generated from a cross between SP2271 (Responds to PITG_02860, R) and SP2272 (No response to PITG_02860, NR). Within this population PITG_02860 responsiveness segregates in a ratio indicating a single dominant gene is responsible. The responsive bulk (Bulk R) and non-responsive bulk (Bulk NR) were used for BSA-RenSeq.

(c) The BSA-RenSeq analysis led to the identification of 218 informative SNPs (red dots) on the top of Chr04 of SP2271. The grey bar represents the chromosome. The physical positions (MB) are shown by number. Five molecular markers (S30, S42, S35, S36 and S31) were designed and used for the map-based cloning. Number of recombination events/total tested gametes are shown below the marker name. the physical positions are indicated in Mb.

(d) HR assay of the candidate PITG_02860 receptor. The candidates were cloned into a vector with 35S promoter and transformed into *Agrobacterium tumefaciens* strain GV3101-pMP90. They were expressed alone, or co-expressed with PVX::PITG_02860 in *N. tabacum* leaves, Rpi-amr3-HF and AVRamr3-GFP were used as controls. NLR16 turned out to be the PITG_02860 receptor (*Rpi-amr16* hereafter). OD$_{600}$ = 0.5, four-week *N. tabacum* plants were used for agroinfiltration, the photos were taken at 4 days after infiltration. Four leaves from two plants were used for each experiment, and three biological replicates were performed with same results. *Rpi-amr16* encodes a 1202 aa protein, and it belongs to the *R3* family.

(e) PITG_04373 is an RXLR effector from *P. infestans*, the illustration and predicted structure are shown.

707   (f) Both backcross (BC1) and F2 populations were generated from SP2271 and SP2300. The BC1 population of
708   192 responsive plants and 182 non-responsive progenies were bulked for BSA_RenSeq. The F2 population were
709   used for fine mapping.

710   (g) BSA-RenSeq led to the identification informative SNPs (red dots) on the top of Chr10 of SP2271. The grey
711   bar represents the chromosome. The physical positions (MB) are shown by number. Five molecular markers (S11,
712   S13, S5, S7 and S16) were designed and used for fine mapping. Number of recombination gametes/ total tested
713   gametes are shown below the marker name. SMRT_RenSeq assemblies of SP2300 were mapped to SP2271
714   genome, and nine genes which mapped to the mapping interval were identified as candidate genes of PITG_04373
715   receptor. *C444* belongs to *R1* family, all other candidate genes belong to *Rpi-chc1* family.

716   (h) HR assay of the PITG_04373 receptor candidates. Genes were cloned into a vector with 35S promoter and
717   transformed into *Agrobacterium tumefaciens* strain GV3101-pMP90. They were expressed alone, or co-expressed
718   with 35S::PITG_04373 in *N. benthamiana* leaves, Rpi-amr3-HF and AVRamr3-GFP were used as controls. C168
719   turned out to be the PITG_04373 receptor (Rpi-amr17 hereafter). $OD_{600} = 0.5$, four-weeks *N. benthamiana* plants
720   were used for agroinfiltration, photos were taken at 4 days after infiltration. Four leaves from two plants were
721   used for each experiment, and three biological replicates were performed with same results. *Rpi-amr17* encodes
722   a 1259 aa protein, and it belongs to the *Rpi-chc1* family.

Supplementary Information:

Supplementary Figure 1. Phenotypes of four *S. americanum* accessions at the flowering stage.

Supplementary Figure 2. K-mer distribution of four *S. americanum* genomes.

Supplementary Figure 3. Hi-C interaction maps of SP1102, SP2271 and SP2273.

Supplementary Figure 4. BUSCO evaluation of *S. americanum* genome assemblies and gene model predictions.

Supplementary Figure 5. Genome alignment between *S. americanum* and neighbouring species. (a). Genome alignment between *S. americanum* (SP1102) and potato (DM). (b). Genome alignment between *S. americanum* (SP1102) and eggplant (HQ1315).

Supplementary Figure 6. Large structural variations between *S. americanum* genomes. Structural variations with length < 1 Mb were not shown in the figure.

Supplementary Figure 7. Structural variations among *S. americanum* genomes. (a). SV length distribution. (b). Number of SVs in each accession. (c). SV length in each accession.

Supplementary Figure 8. SVs might contribute to gene expression variations. (a). Number of SVs overlapped with genomic features. (b). An example of different expressed gene (DEG) whose expression might associate with the SV in its promoter region. Upper panel: The gene model of *sp1102chr11_nlr_6* with a deletion in its promoter region. Bottom panel: the volcano plot of DEGs in leaves between SP1102 and SP2271. Red and blue dots stand for DEGs (*P*-value < 0.05) with fold change <= 0.5 or >= 2, respectively; Green dots denote non-significantly expressed genes. The purple star: the expression of *sp1102chr11_nlr_6*.

Supplementary Figure 9. Sequence diversity in NLR region. (a). The comparison of sequence diversity between NLR regions and non-NLR regions. Wilcoxon rank-sum test was used to assess the diversity index between NLR regions and non-NLR regions. (b). Synteny plot of *Rpi-amr3* locus indicates the sequence diversity of NLR regions.

Supplementary Figure 10. SaNRC1 supports the function of *Rpi-amr3* but not *Rpi-amr1*; SaNRC2 supports the function of both Rpi-amr3 and Rpi-amr1. (a). HR assay in nrc2.3.4 knockout *N. benthamiana*; (b). HR assay in wild-type *N. benthamiana.* $OD_{600} = 0.5$.

Supplementary Figure 11. Venn plot for effector recognition profiles in the four *S. americanum* accessions with reference genomes.

Supplementary Figure 12. Phylogeny, population structure and inbreeding coefficient of all *S. americanum* accessions in this study.

754   **Supplementary Figure 13**. BSA-RenSeq for Rpi-amr4 in F2 population of SP2271 x SP2298.

755   (a). Pipeline of the BSA-RenSeq; (b and c) Informative SNPs from the BSA-RenSeq are

756   located on a NLR singleton sp1102chr01_nlr39.

757   **Supplementary Figure 14**. Alignment of Rpi-amr4-1102 and Rpi-amr4-2271 proteins.

758   **Supplementary Figure 15**. PITG_22825 agroinfiltration and disease test on the *Rpi-amr4*

759   transgenic *N. benthamiana*.

760   **Supplementary Figure 16**. *Rpi-amr16* knockout SP2271 lines. (a). The constructs used for

761   the CRISPR-Cas9 knockout experiment; (b). Genotyping data of three *Rpi-amr16* lines

762   SLJ25598#1, SLJ25598#2 and SLJ25598#3. (c). Phenotype of three selected knockout lines

763   by agroinfiltration. Rpi-amr3 and AVRamr3 were used as controls.

764   **Supplementary Figure 17**. *Rpi-amr1*/*Rpi-amr3*/*Rpi-amr17* triple knockout SP2300 lines. (a).

765   The constructs used for the CRISPR-Cas9 knockout experiment; (b, c and d). Genotyping the

766   *Rpi-amr1-2300* and *Rpi-amr3-2300* and *Rpi-amr17* in two knockout lines SLJ25603#3 and

767   SLJ25603#17. (e). Phenotype of the triple knockout lines after expression of PpAVRamr1,

768   AVRamr3 and PITG_04373. Co-expression of Rpi-amr17 and PITG_04373 was used as a

769   control. (f). Disease test on the two triple knockout lines, *P. infestans* isolates T30-4 and 88069

770   was used in this assay. SP2300 and SP2271 were used as controls.

771

772   **Supplementary Dataset 1.** ORFs of the manually annotated *NLRs* from SP1102.

773   **Supplementary Dataset 2.** CDSs of the manually annotated *NLRs* from SP1102.

774   **Supplementary Dataset 3.** Proteins of the manually annotated *NLRs* from SP1102.

775   **Supplementary Table 1.** NLR physical position, P/A polymorphism and cDNA RenSeq data.

776   **Supplementary Table 2.** Effectoromics screening scoring table.

777   **Supplementary Table 3.** Primers and molecular markers in this study.

778   **Supplementary Table 4.** Genotyping data of *Rpi-amr4* knockout SP2271 lines.