1   Comparative genomic analysis of 31 *Phytophthora* genomes reveal genome plasticity and horizontal

2   gene transfer

3

4   Brent A. Kronmiller[a,b], Nicolas Feau[c], Danyu Shen[d], Javier F. Tabima[e], Shahin S. Ali[f], Andrew D.

5   Armitage[g], Felipe Arredondo[a,b], Bryan A. Bailey[f], Stephanie R. Bollmann[h], Angela Dale[c, i], Richard J.

6   Harrison[j], Kelly Hrywkiw[c], Takao Kasuga[k], Rebecca McDougal[l], Charlotte F. Nellist[j], Preeti Panda[m],

7   Sucheta Tripathy[n], Nari M. Williams[l, o], Wenwu Ye[d], Yuanchao Wang[d], Richard C. Hamelin[c,p,q], Niklaus J.

8   Grunwald[r]

9

10   [a]Center for Quantitative Life Sciences, Oregon State University, Corvallis, OR, USA

11   [b]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, USA

12   [c]Department of Forest and Conservation Sciences, The University of British Columbia, Vancouver, British

13   Columbia, Canada

14   [d]Department of Plant Pathology, Nanjing Agricultural University, Nanjing, Jiangsu, China

15   [e]Department of Biology, Clark University, Worcester, MA, USA

16   [f]Sustainable Perennial Crops Laboratory, Northeast Area, USDA/ARS, Beltsville Agricultural Research

17   Center-West, Beltsville, MD, USA

18   [g]Natural Resources Institute, University of Greenwich, Chatham Maritime, Kent, UK

19   [h]Department of Integrative Biology, Oregon State University, Corvallis, USA

20   [i]SC-New Construction Materials, FPInnovations, Vancouver, BC, V6T 1Z4, Canada

21   [j]NIAB EMR, East Malling, United Kingdom.

22   [k]Crops Pathology and Genetics Research Unit, Agricultural Research Service, Davis, United States

23   Department of Agriculture, CA, USA

24   [l]Scion (Zealand Forest Research Institute), 49 Sala Street, Te Papa Tipu Innovation Park, Private Bag

25   3020, Rotorua, New Zealand

26   [m]The New Zealand Institute for Plant and Food Research Ltd, 74 Gerald Street, Lincoln, 7608, New

27   Zealand

28   [n]CSIR Indian Institute of Chemical Biology, Kolkata, India

29   [o]Department of Pathogen Ecology and Control, Plant and Food Research, Private Bag 1401, Havelock

30   North, New Zealand

31   [p]Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Canada

32   [q]Département des sciences du bois et de la forêt, Faculté de Foresterie et Géographie, Université Laval,

33   Québec, Canada

34   [r]Horticultural Crop Research Unit, United States Department of Agriculture, Agricultural Research

35   Service, Corvallis, Oregon, USA

36

37   Running Head: Comparative genomic analysis of 31 *Phytophthora* genomes

38

39   Address correspondence to Brent Kronmiller, brent.kronmiller@oregonstate.edu

40   Nicolas Feau, Danyu Shen, Javier F. Tabima contributed equally to this work. Author order is

41   alphabetical.

42

43

44   **Abstract**

45   *Phytophthora* species are oomycete plant pathogens that cause great economic and ecological impacts.

46   The *Phytophthora* genus includes over 180 known species, infecting a wide range of plant hosts including

47   crops, trees, and ornamentals. We sequenced 31 individual *Phytophthora* species genomes and 24

48   individual transcriptomes to study genetic relationships across the genus. *De novo* genome assemblies

49   revealed variation in genome sizes, numbers of predicted genes, and in repetitive element content across

50   the *Phytophthora* genus. A genus-wide comparison evaluated orthologous groups of genes. Predicted

51   effector gene counts varied across *Phytophthora* species by effector family, genome size, as well as plant

52   host range. Predicted numbers of apoplastic effectors increased as the host range of *Phytophthora* species

53   increased. Predicted numbers of cytoplasmic effectors also increased with host range but leveled off or

54   decreased in *Phytophthora* species that have enormous host ranges. With extensive sequencing across

55   the *Phytophthora* genus we now have the genomic resources to evaluate horizontal gene transfer events

56   across the oomycetes. Using a machine learning approach to identify horizontally transferred genes with

57   bacterial or fungal origin we identified 44 candidates over 36 *Phytophthora* species genomes. Phylogenetic

58   reconstruction indicates that the transfers of most of these 44 candidates happened in parallel to major

59   advances in the evolution of the oomycetes and Phytophthoras. We conclude that the 31 genomes

60   presented here are essential for investigating genus-wide genomic associations in *Phytophthora*.

61   **Introduction**

62   Members of the *Phytophthora* genus are oomycete plant pathogens that collectively infect a wide range of

63   plants (1, 2) and cause great economic, environmental, and societal impact (3). Oomycetes are

64   morphologically similar to filamentous fungi (4–7) but are classified as stramenopiles, a group that also

65   includes diatoms and brown algae (4, 8). Oomycetes include many plant pathogenic species besides

66   *Phytophthora* including numerous *Pythium* species that cause seed, seedling, root and fruit rots and a

67   broad diversity of obligate biotrophs that cause downy mildew.

68   *Phytophthora* species infect numerous plants including crops, trees, and ornamentals in managed and

69   natural ecosystems. The agent responsible for potato blight, *Phytophthora infestans* triggered the 1840s

70   potato famine (9, 10), and *Phytophthora ramorum* is responsible for sudden oak death in North America

71   and sudden larch death in the UK, which have destroyed millions of trees, in addition to infecting hundreds

72   of additional tree and ornamental species (11–14). Some *Phytophthora* species are relatively host-specific,

73   such as the soybean pathogen *Phytophthora sojae* (15), the strawberry pathogen *Phytophthora fragariae*

74   (16), and the lychee pathogen *Phytophthora litchi* (17). In contrast, others including *Phytophthora*

75   *cinnamomi* (18), *Phytophthora palmivora* (1), and *Phytophthora parasitica* (1), can infect a vast assortment

76   of plant hosts. The mechanistic basis of this large variation in apparent host specificity is currently unknown

77   (10, 19).

78   There are over 180 known *Phytophthora* species (20, 21) phylogenetically falling into 12 or more clades

79   (21–23) and further divided into numerous sub-clades (20). Previous genome sequencing studies have

80 examined species within the *Phytophthora* genus (9, 19, 20, 22, 24–28) with *P. sojae* and *P. ramorum* (29),

81 *P. infestans* (10, 30), and *Phythophthora capsici* (31) serving as key models for the genus. Overall, these

82 studies revealed highly dynamic genomes containing both rapidly evolving and conserved regions.

83 *Phytophthora* species infect plant hosts through the use of two broad classes of secreted effector proteins

84 (5, 32–35). Apoplastic effectors act outside the plant cells and include glycoside hydrolases, necrosis

85 inducing proteins (NLPs) (36), proteases, lipases, lipid-binding proteins, and protease inhibitors (37). Roles

86 of apoplastic effectors include weakening of plant physical and chemical defenses, and a source of nutrition

87 early in infection. In contrast, cytoplasmic effectors enter plant cells, often through the differentiation of

88 specialized structures called haustoria, and include RxLR effectors (29, 38, 39), crinkler effectors (CRN)

89 (40–44) and non-conventionally secreted effectors (45). In oomycetes, as well as other pathogens,

90 cytoplasmic effectors manipulate numerous aspects of host physiology and morphology to promote

91 susceptibility, including suppression of host immunity and programmed cell death (46, 47), and stimulating

92 and inhibiting the release of nutrients (5, 48–51). The genomes of oomycete pathogens sequenced to date

93 include large rapidly evolving gene families encoding these effectors (10, 29, 31). Many of these effector

94 genes, especially those encoding RxLR effectors, display evidence of accelerated evolution due to host-

95 pathogen co-evolutionary conflict (38, 52, 53).

96 Examination of previously sequenced genomes has identified two distinct partitions namely gene-dense,

97 repeat-poor regions and gene-sparse, repeat-rich regions (10, 29–31). Highly conserved housekeeping

98 genes are typically found in gene-dense regions, while rapidly evolving gene families associated with

99 infection are typically found in gene-sparse regions that are transposon-rich (10, 29–31, 54). This

100 arrangement has been labeled "the two-speed genome" (19, 24). It has been hypothesized that transposons

101 in the gene-sparse, transposon-rich regions may contribute to the genomic diversity and possibly to

102 epigenetic variability of expression of genes in those regions including infection-associated genes (55).

103 To investigate phylogenetic relationships, horizontal gene transfer (HGT), effector genomics, and possible

104 mechanisms underlying host ranges across members of the genus *Phytophthora*, we sequenced 31

105 genomes using Illumina short read technology. Our newly sequenced genomes include most species in

106 clade 7 as well as representative species from nine of the phylogenetic clades (56). Several genomes have

107   already been published individually as a result of this project: *P. litchii* (25), *Phytophthora megakarya* and

108   *Phytophthora palmivora* (26), *Phytophthora fragariae* and *Phytophthora rubi* (27), and *Phytophthora*

109   *cactorum* (28, 57). Here we present a combined analysis of the genome sequences of 37 *Phytophthora*

110   species, including the 31 newly sequenced species by our sequencing consortium, resulting in the first

111   large scale comparative genomic study including species from nine *Phytophthora* clades. This work

112   provides insights into genome architecture and evolution in the genus *Phytophthora* as well as novel

113   genomic resources of broad interest.

114   **RESULTS**

115   **Sequencing and assembly of the 31 *Phytophthora* genomes**

116   The 31 *Phytophthora* genomes sequenced by our consortium produced between 16.6M and 44.7M raw

117   reads per genome (Table 1 and Table S1). The sizes of the *Phytophthora* genome assemblies varied

118   greatly, ranging from 37.3 to 107.8 Mb (mean 61.3 ±17.6Mb). Due to variations in the numbers of reads

119   per genome and differences in genome sizes, the quality of assembled genomes varied as well. The

120   number of contigs per assembly ranged from 2,131 to 28,263 (mean 11380 ±7919.2). The more completely

121   assembled genomes benefited from deeper coverage. For example, the *Phytophthora boehmeriae*

122   assembly was composed of 2,866 contigs with an N_{50} of 41,917 bp; this assembly benefitted from 87X

123   coverage as a result of receiving 34.7M reads to cover a 39.7 Mb assembly. In contrast, the *Phytophthora*

124   *lateralis* assembly had only 45X coverage (23.2M reads across a 50.5 Mb assembly size), resulting in

125   28,263 contigs and an N_{50} of 2,396 bp (Table 1). Sequence read length did not seem to make a difference

126   in assembly quality. Two genomes (*P. cactorum* and *P. idaei*) with 250 bp reads (compared to most

127   assemblies with 50 bp reads) had an N_{50} and number of contigs that were not better than the other genomes

128   when compared to genome size (*P. cactorum*: contigs 7,888, N_{50} 15,053 size 56 Mb; *P. idaei*: contigs 7,163,

129   N_{50} 14,461, size 53.5 Mb) (Table 1).

130   To aid with gene calling, and identify active genes, RNA sequencing was conducted for 25 *Phytophthora*

131   species on V8-grown mycelia and either germinated cysts or Plich-grown mycelia for those species that did

132   not readily yield zoospores (58). RNA sequencing produced 12.5M to 36.0M paired-end sequence reads

133   per sample, with resulting transcriptome assemblies averaging 36,189 contigs per assembly with an

134   average N_{50} of 1,822 bp. Most species' transcriptome assemblies had between 20K and 40K contigs, with

135   the exception of *Phytophthora parvispora* that produced more than 109K contigs. TransDecoder, which

136   reduces duplication in *de novo* transcriptome assemblies by identifying candidate coding regions and

137   removing duplicates, was run on the Trinity assemblies and lowered average transcriptome assembly to

138   33,200 contigs and raised average N_{50} to 2,753bp. Transcriptome size differences between species are

139   not correlated with genome size differences with the two largest genome assemblies (*P. megakarya* and

140   *P. palmivora*) both resulting in mid-range transcriptome assembly sizes (Table S1).

141   ***De novo* repeat identification and analysis**

142   Repeats were identified, classified, and masked to prepare genomes for gene prediction. *De novo* repeat

143   prediction identified between 27 and 295 different repeat sub-families per species. The percentage repeat

144   content of the assemblies varied greatly across the *Phytophthora* genomes sequenced. The genome

145   assemblies of *Phytophthora kernoviae*, *P. litchii*, and *Phytophthora agathidicida* contained very low repeat

146   content of 4.15%, 5.76%, and 5.98%, respectively (Fig. 1). On the other end, several genome assemblies

147   had high repeat content: *P. megakarya*, *Phytophthora hibernalis*, and *Phytophthora pinifolia* contained

148   32.94%, 32.46, and 33.00%, respectively.

149   Repeat annotations were classified into types (Class I Retrotransposons, Class II DNA transposons, and

150   other), families, and sub-families, (Fig. 1). Fifteen different DNA transposons were identified across all 31

151   species. Long terminal repeat (LTR) retrotransposons were more diverse across *Phytophthora* species.

152   As many as 57 Copia LTR retrotransposons and 144 Gypsy LTR retrotransposon types were found in each

153   genome assembly.

154   ***Phytophthora* gene prediction and annotation**

155   Gene predictions ranged from 12,391 to 37,283 (mean 19,943, ±5,864) across the newly sequenced

156   genomes.   These gene counts per species are within the general range of previously published

157   *Phytophthora* genomes (*P. sojae*, *P. ramorum*, *P. infestans, P. capsici*, *P. cinnamomi*, and *P. parasitica*)

158   ranging from 16,066 to 26,584 genes per species, considering that we observed higher gene counts in

159   genomes with genome duplications (*P. palmivora* and *P. megakarya* (26)).

160   Some of the previously published *Phytophthora* genome sequences were annotated with the MAKER gene

161   prediction process outlined here to validate the methods. The *P. capsici* genome was reported to contain

162   19,805 predicted genes (31), while we obtained 18,917 predicted genes. The *P. sojae* genome v3.0 was

163   reported to contain 26,584 predicted genes (29), whereas we identified 21,447 MAKER-predicted genes.

164   Therefore, our MAKER pipeline may slightly undercount the gene content compared to other methods.

165   The completeness of the gene sets predicted from the genomes and transcriptomes was assessed by

166   identifying single-copy core orthologs using the BUSCO pipeline with the Alveolata_Stramenopiles

167   database (234 genes) as the reference. BUSCO analysis of the 31 genome assemblies identified 147-231

168   complete genes (mean 204 ±27) of the 234 single copy genes in the database (Fig. 2A). Analysis of the

169   24 transcriptome assemblies identified 18 to 228 complete genes (mean 187 ±54) (Fig. 2B). The *P.*

170   *kernoviae* and *P. lateralis* transcriptomes were outliers, with only 20 and 18 complete genes identified,

171   respectively. When *P. kernoviae* and *P. lateralis* were removed, transcriptome assemblies ranged from

172   142 to 228 complete (single and duplicated) genes (mean 203 ±18). Results of BUSCO analyses run on

173   predicted proteins ranged from 139-222 complete (single and duplicated) genes (mean 192, ±25) of the

174   234 conserved orthologous proteins in the database (Fig. 2C).

175   Predicted proteins from the MAKER gene prediction were functionally annotated by matching to published

176   *Phytophthora*, stramenopile, and fungal proteins. Across the 614,862 proteins predicted in the 31

177   *Phytophthora* species, when aligned to the NCBI and UniProt TrEMBL 294,146 *Phytophthora* proteins

178   produced alignments that passed the cutoff filter. Removing 'Uncharacterized Protein' or similarly

179   uninformative functional annotations yielded 196,652 proteins with functional classifications (mean 6,343.6

180   ± 3,191.7). When aligned to the *Phytophthora*, stramenopile, and fungal sequence databases, 445,458

181   proteins passed the alignment cutoffs, with 304,951 proteins (mean 9,837.1 ± 2523.3) that had informative

182   functional annotations.

183   InterProScan was used to identify domains and motifs in all predicted proteins. Of the 614,862 proteins,

184   173,727 had domains identified. GO terms were assigned from BLASTX alignments between the UniProt

185   BLASTX alignments and the InterProScan predictions, identifying 321,953 proteins with GO terms

186   assigned.

187 **Effector protein identification in *Phytophthora***

188 The cytoplasmic effector identification process predicted a total of 10,354 RxLR effector proteins and 4,415

189 CRN effector proteins from the genomes and transcriptomes of the species sequenced in this study (Fig.

190 3). The numbers of predicted RxLR effector genes differed greatly across genomes. *P. pinifolia* exhibited

191 the lowest number, with 46 predicted RxLR effectors, while *P. megakarya* exhibited the highest number,

192 with 1,183 predicted proteins. We also observed great variation in the number of predicted CRN effectors,

193 with *P. litchii* showing the lowest number of CRN effectors, 27, while *Phytophthora cajani* showed the

194 highest, 274.

195 Our search for apoplastic effectors identified 6,671 glycoside hydrolases, 1,191 NLPs, and 1,046 protease

196 inhibitors across the 31 *Phytophthora* genomes (Fig. 3). The predicted glycoside hydrolase genes range

197 from 139 (*Phytophthora brassicae*) to 386 (*P. palmivora*). The numbers of NLP genes range from 5 (*P.

198 pinifolia*) to 78 (*Phytophthora niederhauserii*). The counts of protease inhibitor genes range from 23

199 (*Phytophthora brassicae*) to 67 (*P. palmivora*).

200 **Orthology clustering of *Phytophthora* proteins**

201 The 715,980 predicted genes from the 31 genome assemblies along with those of *P. sojae, P. ramorum,

202 P. infestans, P. capsici, P. cinnamomi,* and *P. parasitica* were subjected to orthology analysis. In the first

203 step, the *Phytophthora* genes were matched against the pre-computed publicly available orthologous

204 groups. From this analysis, 560,201 genes were assigned to 7,829 unique clusters, leaving 155,779 genes

205 unassigned. In step two, these remaining genes were clustered using OrthoMCL yielding 13,474 additional

206 unique clusters. In total, the 715,980 genes were assigned to 21,303 orthologous clusters.

207 Orthologous groups were assigned functional annotations based on the proteins that composed the group.

208 Of the 21,303 orthologous clusters, 9,806 could be assigned informative functional annotations as defined

209 in the gene annotation section.

210 The numbers of genes present in orthologous groups encompassing all 37 *Phytophthora* species shows a

211 bimodal frequency distribution (Fig. 4), with peaks observed at 1 to 6 genes per group, and 34 to 41 genes

212 per group. We hypothesize this first peak represents rapidly evolving genes that are conserved in only a

213   few of the 37 *Phytophthora* species in this study. This first peak was much smaller when only genes with

214   meaningful annotations were considered, suggesting an enrichment for genes that have previously

215   uncharacterized functions. This peak included both small orthologous groups in which all genes were from

216   the same *Phytophthora* clade, as well as groups consisting of genes from multiple clades. The second

217   large peak was centered at 37 genes per ortholog group, thus representing ortholog groups that have one

218   gene per species. Additional groups include one or a few species missing the orthologous gene or one or

219   a few species with a second copy of an orthologous gene.

Phylogenetic relationship across *Phytophthora* species

221   We reconstructed the phylogenetic relationships of the sequenced species of the genus *Phytophthora* (Fig.

222   5) based on 61 single-copy core orthologous genes shared across 37 species. Predicted genes and amino

223   acid protein sequences used to build the phylogeny are found in Table S2. The RAxML phylogenetic tree

224   clustered these species into phylogenetic clades consistent with previous studies (20, 21, 56, 59, 60), with

225   the exception of *Phytophthora* taxon totara placed into clade 5. The separation of *P*. taxon totara from clade

226   3 containing *P. pluvialis* has been reported previously (59).

Relationship of *Phytophthora* effector gene numbers to plant host range.

228   To examine the assemblies for clues as to genomic basis for the diverse host ranges of the sequenced

229   species, *Phytophthora* species were categorized into plant host ranges as follows. Thirteen species were

230   defined as having 'Narrow', 12 as 'Multiple', six as 'Wide', and six as 'Huge' host range (Fig. 5).   The

231   numbers of genes predicted to encode various families of effectors were plotted for each host range class.

232   Paralogous effectors with greater than 95% nucleotide identity over the full sequence length were counted

233   only once. Counts were plotted for each of two cytoplasmic effector categories (RxLR and CRN) and three

234   apoplastic effector categories (glycoside hydrolases, protease inhibitors, and NLPs) (Fig. 6). Subcategories

235   of the apoplastic effectors glycoside hydrolases and protease inhibitors were individually plotted (Fig. S1 A

236   and B). In each case, numbers of genes predicted to encode each effector sub-category for each species

237   were plotted against host range.

238   Apoplastic effectors show a distinct overall pattern; *Phytophthora* species with smaller host ranges had

239   fewer predicted effector genes, with numbers of predicted effector genes increasing with increased host

240   range (Fig. 6: A, B, C). Predicted cytoplasmic effector genes show a similar pattern, starting with low

241   numbers of predicted genes in *Phytophthora* species with 'Narrow' host ranges and increasing in those with

242   'Multiple' and 'Wide' host ranges.   However, for both the RxLR and CRN effector categories, the numbers

243   of predicted genes per species decreas from the 'Wide' to the 'Huge' host range species (Fig. 6: D and E).

244   Two species were outliers with respect to the number of apoplastic NLP effector genes, *P. pistaciae* in the

245   'Narrow' host category and *P. sojae* in the 'Multiple' host category had many more predicted apoplastic

246   effector genes than the other species in each of their host range categories, respectively. Four species

247   were outliers with respect to the numbers of predicted RxLR effector genes; *P. megakarya* and *P. pistaciae*

248   in the 'Narrow' host category, *P. parvispora*, in the 'Multiple' host category, and *P. palmivora* in the 'Huge'

249   host range category all had many more predicted RxLR genes than the other species in those categories.

250   One outlier was observed in the CRN effector genes; *P. infestans* in the 'Wide' host range category had

251   many more genes than the other species in that category.

252   **Horizontal Gene Transfer**

253   We evaluated all 31 genomes for evidence of HGT. We used machine learning to identify HGT candidates

254   and phylogenetic approaches to validate candidate HGT genes.

255   **Support Vector Machine classifier predicted HGT candidates.** Analysis of the 722,232 transcripts with

256   our support vector machine classifier (SVM) over the 31 genome assemblies and the six previously

257   published *Phytophthora* genomes identified 35,246 HGT candidates. A total of 28,791 of these transcripts

258   that could be regrouped in orthology groups encoding putative transposable elements were discarded,

259   resulting in 6,455 non-TE HGT candidates. The number of candidates predicted ranged from 91 in *P.*

260   *agathidicida* to 233 in *P. megakarya and* 262 in *P. palmivora* (mean 160.13 ±36.68). *P. agathidicida* has

261   one of the lowest number of transcripts annotated (12,923 transcripts), while *P. palmivora* and *P.*

262   *megakarya* have the highest gene content with 37,283 and 33,614 transcripts, respectively. Overall, we

263   identified a significant linear correlation between gene space in each genome and the number of HGT

264   candidates predicted with the SVM classifier (r^2 = 0.45; p < 0.0001).

265 **Phylogenetic filtering of HGT candidates.** The 6,455 non-TE HGT candidates predicted with the SVM

266 classifier were subject to a two-step filtering process to discard false positives. In the first step, we searched

267 for homologs among a database of sequences built from seven clades (including putative fungal and

268 bacterial donors), followed by phylogenetic tree reconstruction with bootstrap analysis. The phylogenetic

269 filter retained 2,214 candidates, among which 1,113 (50.3%) showed a strong phylogenetic discordance

270 and were seen nested within a distantly related clade (Fungi, Bacteria or Amoebozoa) in direct contradiction

271 to the expected phylogenetic relationships of the respective organisms. The 1,110 other candidates left

272 also branched within a clade of fungal or bacterial genes; however, in these cases, placement of the

273 *Phytophthora* transcript with Fungi or Bacteria was caused by the absence of homologs in the intermediate

274 clades Viridiplantae, Alveolata and/or Amoebozoa. Numbers of candidates that passed this filter ranged

275 from 36 in *Phytophthora europea* to 127 in *P. palmivora*; There was a significant correlation between the

276 total number of genes and the number of HGT candidates across the analyzed species (r^2 = 0.49; p <

277 0.0001) (Fig. S2).

278 **Sequence identity filtering of HGT candidates.** The 2,214 HGT candidates retained after the

279 phylogenetic filtering were submitted to a sequence identity discrepancy filter. A total of 1,688 candidates

280 were rejected after the first "identity test", resulting in a "relaxed" set of 526 HGT candidates for which the

281 sequence identity between the candidate HGT sequence in *Phytophthora* and its closest homolog

282 sequences in the putative "donor" species was shorter than the average identity between the two species;

283 an average of 14.6 (±6.2) candidates were retained per *Phytophthora* genome with a maximum of 33 for *P.*

284 *palmivora* and *P. niederhauserii* (Fig. S2). A gene enrichment analysis of this candidate set showed a

285 significant enrichment for GO terms related to oxidoreductase activity and hydrolysis and metabolism of

286 carbohydrates (cutinase activity, carbohydrate metabolic process) and proteins (Table S3).

287 Among the 526 candidates of the relaxed set, 44 passed the second "identity test", constituting a "strict" set

288 of HGT candidates. For 28 of them (56.0%), BLAST search results indicated a strong homology with the

289 clade of the putative donor where the BLAST e-values with species of the putative donor clade were lower

290 than the e-values observed with species from non-donor clades. We then looked at their physical location

291 on their respective scaffold to eliminate potential contaminants; All the candidates were found on scaffolds

292   that had at least two gene models predicted on them. GO term enrichment analysis of this "strict" set

293   indicated significant enrichment for GO terms related to oxidoreductase activity (GO:0055114;

294   GO:0016491; GO: 0008670), and carbohydrates activity and cell wall modification (GO:0000272;

295   GO:0045490; GO: 0042545; GO:0045493; GO:0031176; GO:0030599) (Table 2).

296   **Phylogenetic reevaluation of the strict set of HGT candidates.** The 44 candidates of the "strict" set of

297   HGT candidates were subjected to reevaluation by sampling additional taxa within the oomycetes. Their

298   amino-acid sequences were first clustered into closely related groups of sequences by assigning them to

299   the 21,303 orthologous clusters previously defined (see *Orthology Clustering of Phytophthora Proteins*).

300   This process reduced the set of 44 candidates into 28 orthologous clusters, that were then searched against

301   the sequences of the 31 genome assemblies, five *Phytophthora* species sequenced in previous studies,

302   and 36 oomycete genomes (Table S4). Protein members of six of these clusters had homologs (BLASTp

303   e-values ≤ 1E-025; Table S4) in the set of 21 strongly supported HGT candidates identified in the genome

304   of *P. ramorum*, *P. infestans* and *P. sojae* by Richards et al. (61). Following these searches, we

305   reconstructed maximum likelihood phylogenies for 19 of these candidates that had putative functions

306   related to the modification of compounds of the plant cell wall (e.g., pectin esterase, xylulose reductase,

307   tannase and endo-1,4-beta-xylanase), peptidases, oxidoreductases and putative elicitors such as a NPP1

308   protein and an ABC transporter (Fig. S3). In 15 cases, the HGT candidate was found nested within a group

309   of fungi or bacteria, as expected under the hypothesis of a transfer from one of these groups through a

310   horizontal transfer event; comparative topology analysis of alternative tree hypotheses (expected

311   phylogeny and transfer from an oomycete donor to a fungus or bacterium) using the Shimodaira-Hasegawa

312   test (SH-test) provided support for this observation in 14 cases (Table 3). In four other cases, the topology

313   test was significant for the opposite relationship where a transfer occurred from the oomycetes to fungi or

314   bacteria. For HGT9, taxon sampling was not sufficient to accurately infer with confidence the putative HGT

315   donor and enable tree topology testing (Table 3).

316   The distribution of sequence homologs of the "strict" set of HGT candidates among the oomycete phylogeny

317   was strongly variable (Fig. 7; Table S4). The majority of transfer events to oomycetes appear to have

318   occurred relatively recently; three candidates had strong statistical support for transfer from bacteria or

319   fungi to a common ancestor of the *Phytophthora* genus (HGT5, HGT7 and HGT10), two candidates to the

320   *Phytophthora* and Peronosporales with hemibiotrophic lifestyle (HGT2 and HGT15) and two to

321   Peronosporales with a hemibiotrophic or an obligate biotrophic lifestyle (HGT12 and HGT20) (Fig. 7A). Four

322   of these transfers reached close to gene fixation within the *Phytophthora* genus as they were found in more

323   than 80% of the species surveyed and in the nine *Phytophthora* phylogenetic clades considered (Fig. 7A).

324   However, fixation was not the general rule accompanying recent transfers. For instance, two HGT

325   candidates, with functions related to plant pathogenicity (NPP1 protein and peptidase S9) were unique to

326   clade 8 in *Phytophthora* and did not have a homolog in any other *Phytophthora* clade or oomycete species

327   (HGT5 and HGT7). Several HGT events with strong statistical support (Fig. 7A) appear to have occurred

328   following major lifestyle transitions within the oomycetes, i.e., necrotrophy in the Pythiales (HGT6, HGT13

329   and HGT14) to obligate biotrophy and hemibiotrophy in the Peronosporales (HGT2, HGT12, HGT15 and

330   HGT20) and transition to parasitism with three events trackable to a common ancestor of the

331   Saprolegniales, Pythiales and Peronosporales (HGT1, HGT4 and HGT8). Eight out of these ten genes had

332   homologs (BLASTp e-values from 1E-137 to 1E-012) with pathogenicity, virulence, and effector genes of

333   the Pathogen Host Interaction database (PHI-base; (62)). Transition to the necrotrophic lifestyle involved

334   transfers of genes encoding enzymes potentially involved in redox activity and toxin production (2,4-dienoyl-

335   CoA reductase and phenol acid decarboxylase) while two out of the four genes transfer at the transition to

336   hemibiotrophic lifestyle comprehended have putative functions related to the degradation of the cell wall

337   (xylulose reductase and endo-1,4-beta-xylanase).

338   Finally, the four significant transfers for the opposite relationship (oomycetes to bacteria or fungi) were all

339   for genes fixed in the *Phytophthora* genus and mapped within the Peronosporales (three candidates) or the

340   Peronosporales and Pythiales (one candidate), suggesting relatively recent transfer events (Fig. 7B).

341   Annotation of these genes indicates that they are potentially involved in the plant-pathogen interaction as

342   they encode proteins involved in protection against plant defensive molecules (tannase and ABC

343   transporter) and the oxidative stress occurring during the plant defense response (quinone oxidoreductase)

344   and remodeling of the plant cell wall (pectinesterase) (Table 3; Table S5).

345   **DISCUSSION**

346   In this comparative genome study of 37 *Phytophthora* spp., we sequenced and assembled 31 genomes *de*

347   *novo*. We investigated these genomes for evidence of horizontal gene transfer, phylogenetic relationships

348   of genome structure and effectors, and association of host ranges. Horizontal gene transfer has been

349   identified as a significant source of variation in connection with the evolution of pathogenicity in

350   *Phytophthora*. So far, genome-wide analyses of HGT impact on oomycete and *Phytophthora* genome

351   evolution has identified putative transfers from fungi (61, 63) and bacteria (64), many of which involve

352   functions related to carbohydrate metabolism and pathogenicity. Our analysis supports these findings,

353   identifying a set of 44 HGT candidates in *Phytophthora* species, associated with enzymes putatively

354   involved in the deconstruction of plant cell wall components, evasion and protection against host defenses

355   (Table 3; Table S5). Our machine learning approach to identify HGT candidates aimed to identify genes

356   likely inherited from bacteria or fungi; validation of these candidates with classical methods based on the

357   identification of topological incongruence in phylogenies and the detection of discrepancies between gene

358   and species distances resulted in a more conservative list of candidates than those previously proposed

359   for *Phytophthora* species (61, 63, 64). Such a stringent approach had the power of rejecting the alternative

360   evolutionary scenario where the gene was present in the last common ancestor of the donor and recipient

361   and was lost in intermediate lineages. Despite such a conservative approach, 30% of HGT candidates

362   identified in a previous study (61) that included only three *Phytophthora* genomes were retrieved in our

363   analysis.

364   An underlying hypothesis related to laterally transferred genes is that they may have functional or ecological

365   roles, allowing the recipient to adapt to a novel lifestyle or to exploit a new ecological niche (65). Using the

366   comprehensive collection of *Phytophthora* genomes sequenced in this study and the oomycetes for which

367   genome assemblies were available, we have been able to assess the extent of distribution of homologs of

368   these candidates across the oomycete phylum. We confirmed that most of the HGTs into the oomycetes

369   have occurred coincident with the emergence of major lifestyle innovations, such as the acquisition of plant

370   parasitism, or biotrophy (obligate or hemibiotrophy). Many candidates were detected in a large majority of

371   the *Phytophthora* genomes sequenced, for example seven candidates (HGT6, HGT11, HGT17, HGT18,

372   HGT19, HGT21 and HGT22) were found in 94% or more of the 36 genomes surveyed. In some instances,

373   homologs have been retained in distinct genetic lineages among the oomycetes, suggesting that these

374   candidate genes may confer a function conserved across lineages with different lifestyles; for example, the

375   endonuclease encoding gene HGT1 was likely transferred before the radiation of the oomycetes and

376   retained in the five oomycete orders surveyed in this study. On the other hand, in some cases the acquisition

377   and/or retention of specific key pathogenicity genes appears to be restricted to some specific clades within

378   *Phytophthora* (e.g., HGT5 and 7, NPP1; Fig. 7), suggesting recent transfers following divergence of

379   *Phytophthora* clades. In cases where a putative HGT gene is present in a limited number of *Phytophthora*

380   species from diverse clades, e.g., HGT13, gene loss by drift in species where there was little benefit may

381   be an explanation. Rapid diversification of the HGT gene under positive selection might make the gene

382   undetectable to our algorithm in some species.

383   Oomycete-derived transfers to other kingdoms have been identified in a few rare instances, usually with

384   limited statistical support (61). With the comprehensive genome sampling of our study, we found strong

385   support for four transfers (out of 44) from oomycetes to either bacteria or fungi, indicating bi-directional

386   exchanges across kingdoms. By providing a source of novel genetic material that can increase the fitness

387   of micro-organisms to their environments or their hosts (66–68), genes transferred horizontally have the

388   potential to be traded back and forth across kingdoms. In the context of an ecological system in which a

389   host plant interacts with a multitude of micro-organisms (microbiota), we can hypothesize that some of the

390   evolutionary innovations that are generated during the co-evolutionary arms race between a pathogen and

391   the host could be shared within the microbiota. For host-associated micro-organisms sharing the same

392   ecological niche, the transfer of genetic material from those that are fit to the shared environment should

393   represent a straightforward mechanism that will drive rapid adaptation of others to this environment (69).

394   We investigated multiple aspects of *Phytophthora* genome structure and how this relates to the genus

395   phylogeny. Genome size, gene amounts, and counts of orthologous genes varied within phylogenetic

396   clades highlighting the great diversity within the *Phytophthora* genus and likely reflect the large observed

397   differences in repeat content, some of which resulted from genome duplication. While some variation in

398   repeat content may be due to differences in repeats collapsed in the genome assembly, repeat types and

399   lengths identified in this study and sequences used for assembly are generally consistent across the

400   sequenced genomes and therefore should collapse in assembly at similar rates. Interestingly, we noted

401   that none of the genomes sequenced in this study approaches the 73% repeat content reported in the

402   Sanger-assembled *P. infestans* genome (10) and is possibly due to the differences in sequencing

403   technologies including sequence longer sequence lengths. Greater repeat content was not necessarily an

404   indicator of large genome assembly size. While the five species with reported repeat content of greater

405   than 30% (*P. pinifolia*, *P. megakarya*, *P. palmivora*, *P. fragariae* and *P. hibernalis*) were within the nine

406   largest assembly sizes, other species with large assembly sizes had less repeat content. For example, *P.*

407   *niederhauserii* had an assembly size of 90Mb but only 20.9% repeat content which was considerably lower

408   than expected when compared to even moderately repetitive *Phytophthora* genomes such as *P. sojae* 40%

409   (29). Genome size estimation using K-mer analysis also shows assemblies are shorter than expected

410   (Table S1). These observations suggest that repeat content may be underestimated in short-read genome

411   assemblies and would expand with improved assembly and may also indicate missing repeat sequences

412   from the *de novo* repeat identification process.

413   The numbers of predicted genes per species, genes per orthologous groups, and effector genes per

414   species were consistent with those previously reported for *Phytophthora* species. Both the numbers of

415   genes and the average sizes of genes were well within the ranges of the six previously sequenced

416   *Phytophthora* species, for example *P. sojae* was shown to have 26K genes with an average size of 1,181

417   bp (29). This supports our observation that the smaller assembly sizes of the *Phytophthora* genomes

418   presented in this study were mainly associated with an overall reduction in repetitive regions while the gene-

419   containing sequences are relatively consistent in size. BUSCO analysis of the sequenced core ortholog

420   content also showed similar results to previous *Phytophthora* studies. Some genome assemblies, including

421   *P. palmivora* that underwent whole genome duplication, had lower single copy gene numbers due to

422   duplications. But overall, censuses of single-copy orthologs showed that both the genome assemblies and

423   gene predictions were quite complete and comprised the majority of genes in each individual species

424   sequenced. This suggests, that while genomes were small due to collapsed repeat regions, the majority

425   of core orthologs were captured in the assemblies and it can be extrapolated that the majority of gene

426   regions are assembled.

427   Effectors are proteins produced by pathogens that assist in host infection. Effectors are considered rapidly

428   evolving genes that are usually conserved among few closely related species and quickly diverge along the

429   phylogeny. Two types of ortholog groups were identified in our *Phytophthora* genus analysis supporting this

430   hypothesis: one group corresponds to well-conserved genes among all *Phytophthora* species responsible

431   for core cellular functions, while a second group includes rapidly evolving gene families likely responsible

432   for host infection and adaptation. We also investigated how the amounts of predicted apoplastic and

433   cytoplasmic effectors related to the host range of each *Phytophthora* species. There were large differences

434   in the numbers of effector genes identified per species. We did not observe a correlation between the

435   number of predicted effectors with phylogenetic relationships or genome size. However, a clear relationship

436   with host range was observed. Species with smaller host ranges had on average fewer predicted effectors

437   than those with larger host ranges. When the *Phytophthora* species in this study were separated into 4

438   host range categories, the distribution of apoplastic effectors increased as the range of infected hosts

439   increased, from narrow to multiple, multiple to wide, and wide to huge. Cytoplasmic effectors showed a

440   similar pattern, however, both the RxLR and CRN effector numbers dropped from wide to huge host ranges.

441   Our study does not include a detailed measurement of gene expression levels of these effector genes

442   during infection of numerous hosts, so there are limitations in how these correlations can be interpreted. In

443   the absence of that information, we speculate that a large diversity of apoplastic effectors may be important

444   for successfully overcoming the apoplastic defenses of a large diversity of host plants.

445   There may be a similar requirement for larger numbers of cytoplasmic effectors, but expression of very

446   large numbers of cytoplasmic effectors may limit host range due to plant immune surveillance mechanisms.

447   Detection of a single cytoplasmic effector by an NLR resistance protein may be sufficient to prevent

448   infection, therefore a limited number of cytoplasmic effectors may result in a greatly expanded host range.

449   Our observation of reduced cytoplasmic effector compliments in huge host range species may also be

450   indicative of cryptic host-specialization within these *Phytophthora* species. Recent work in *P. cactorum*,

451   commonly considered a broad host-range pathogen, has shown genomic signatures of host specificity (57).

452   In this case, high resolution phylogenetics demonstrated that host adaptation was associated with effector

453   gene gain/loss between strawberry and apple infecting clades. Where such cryptic host-adaptation is

454   present, pangenomic analysis may be a useful tool to infer broad or narrow host range and provide insight

455   to associations of effector diversity across *Phytophthora*.

456   **METHODS**

457   **Collection and isolation of a genus-wide *Phytophthora* collection**

458   Mycelium samples were isolated for all *Phytophthora* species in this study, as well as germinated cyst

459   samples for RNA sequencing of a subset of the species. For mycelium tissue, plugs of mycelium grown on

460   standard V8 agar plates were added to a flask with 20% liquid V8 media clarified with calcium carbonate

461   and incubated with shaking at 45 rpm at 25°C for one week. Agar plugs were removed from the mycelium

462   mass and the tissue was ground to a powder with liquid nitrogen, followed by DNA extraction using the

463   methods in Moller et al. (1992) (70) except using 1% CTAB and phenol/chloroform treatment, or total RNA

464   extraction (71) using TRIzol (Invitrogen, Carlsbad, CA) following manufacture's instructions.

465   For *Phytophthora robiniae* and *Phytophthora vignae*, similar to the protocol for *P. sojae* (29) zoospores

466   were produced by repeated washing of 11 day-old V8-200 plates of mycelium with sterile double distilled

467   wather followed by overnight incubation at 14°C. Germinated cysts were produced by exposing collected

468   zoospores to cleared V8 broth for 1 hour. For *P. parvispora*, mycelium mats were grown in liquid V8 for 5

469   days, then the liquid V8 was changed out for soil extract (soil collected with stream water, mixed, and filter-

470   sterilized) and zoospores were collected after three days, followed by germinated cyst induction as above.

471   For *P. cajani*, *P. europaea*, *Phytophthora foliorum*, *P. hibernalis*, *P. pistaciae*, and *Phytophthora uliginosa*

472   species, that did not readily yield zoospores, mycelium was grown in Plich medium (72) for RNA sequencing

473   to compare against V8 medium growth. Known ITS and CoxII sequences for each species were used to

474   confirm species identification before high-throughput sequencing. DNA and RNA quality were checked

475   with electrophoresis (DNA), Bioanalyzer (RNA), and NanoDrop.

476   Isolate P414 of the strawberry crown rot pathogen *P. cactorum* and isolate SCRP371 of the raspberry root

477   rot pathogen *P. idaei* were sequenced at the National Institute of Agricultural Botany at East Malling

478   Research (NIAB EMR). P414 and SCRP371 were isolated from symptomatic strawberry and raspberry

479   plants, respectively. DNA extraction was performed on freeze dried mycelium using a GenElute Plant

480   Genomic DNA Miniprep Kit (Sigma), following the manufacturers protocol with the following modifications:

481   the RNase A digestion step was not performed and samples were eluted using 2x 100 µL elution buffer for

482   P414 and 2x 75 µL for SCRP371. Genomic libraries were prepared using a Nextera XT Library Preparation

483   Kit (Illumina) or TruSeq DNA LT Kit (Illumina) for *P. cactorum* and *P. idaei*, respectively.

484   *P. kernoviae*, *P. lateralis*, *P. cryptogea*, and *P. pinifolia* were collected and isolated as described (73). *P.*

485   *agathidicida*, *P. multivora*, *P. pluvialis*, and *P*. taxon totara were collected and isolated as described (74).

486   **Genome sequencing and assembly**

487   Thirty-one *Phytophthora* species were sequenced by our consortium. Genomes of 21 *Phytophthora*

488   species were sequenced by BGI Genomics (Shenzhen, China) (*P. boehmeriae*, *P. brassicae*, *P. cajani*, *P.*

489   *pini*, *P. europaea*, *P. foliorum*, *P. fragariae*, *P. hibernalis*, *P. litchii*, *P. megakarya*, *Phytophthora melonis*, *P.*

490   niederhauserii, *P. palmivora*, *P. parvispora*, *P. pisi*, *P. pistaciae*, *P. robiniae*, *P. rubi*, *P. syringae*, *P.*

491   *uliginosa*, and *P. vignae*) using 90-bp paired-end reads produced on the Illumina HiSeq2000 platform. Four

492   *Phytophthora* species were sequenced by the University of British Columbia (*P. kernoviae*, *P. lateralis*, *P.*

493   *pinifolia*, and *P. crypogea*), using Illumina HiSeq 2000 100 bp paired-end reads (73). Genomes from four

494   *Phytophthora* species isolated from New Zealand (*P. agathidicida*, *P. multivora*, *P. pluvialis*, and

495   *Phytophthora* taxon totara) were sequenced by Scion (New Zealand Forest Research Institute, Ltd.), using

496   primarily Illumina HiSeq 100 bp paired-end reads (74). Two *Phytophthora* genomes were sequenced by

497   NIAB EMR (*P. idaei* and *P. cactorum*), using 250-bp paired-end reads produced on a MiSeq Benchtop

498   Analyser (Illumina, San Diego, CA, USA). BGI-sequenced genomes were adapter trimmed to remove

499   Illumina adapters and quality trimmed to remove Phred scores of less than Q20 from the ends of reads

500   (75). Genome sequences were assembled with SOAPdenovo2 (76). Several initial assemblies were done

501   to identify an optimal K-mer length for each genome. Gap filling and single base proofreading were

502   conducted with SOAPAligner (77). The University of British Columbia (UBC) genomes were quality trimmed

503   and assembled using ABySS (78) and a range of k-values from 32 to 96 (73). Scion genomes were

504   assembled using SPAdes (79) and contigs were extended using SSPACE (74, 80). Genomes sequenced

505   at NIAB EMR were trimmed and adapters removed using fastq-mcf (81) prior to *de novo* assembly of the

506   data using Velvet (82), at K-mer lengths of 61 and 41 bp for *P. cactorum* and *P. idaei*, respectively.

507   Genome size was estimated using K-mer counts of the raw Illumina sequence. K-mers were counted using

508   Jellyfish count (version 2.2.6, -m 32) (83) Histograms created using Jellyfish hist were plotted using R (84)

509   to identify the apex and boundaries of the single copy K-mer peak. Genome size was calculated by dividing

510   the total of unique K-mers by the mean coverage (peak K-mer frequency).

511   In order to separate mitochondrial genome contigs from the nuclear genome assembly, full length

512   mitochondrial genome sequences were collected from GenBank (85) for the following 9 *Phytophthora*

513   species; *P. andina*, *P. infestans*, *P. ipomoeae*, *P. mirabilis*, *P. parasitica*, *P. phaseoli*, *P. polonica*, *P.*

514   *ramorum*, and *P. sojae*. The consortium's 31 assembled *Phytophthora* genomes were aligned with Blat

515   (86) to identify mitochondrial contigs. Blat alignments were filtered to return alignments greater than 50%

516   of the aligned contig length, greater than 100 bp, and with gaps less than 50% of the contig length. Contigs

517   identified as mitochondrial were removed from the genome assembly and are a part of a different study.

518   **Transcriptome sequencing and assembly**

519   Twenty-four of the genome-sequenced *Phytophthora* species underwent RNA sequencing (*P. brassicae,*

520   *P. cactorum, P. cajani, P. pini, P. europaea, P. foliorum, P. fragariae, P. hibernalis, P. kernoviae, P. lateralis,*

521   *P. litchii, P. megakarya, P. melonis, P. niederhauserii, P. palmivora, P. parvispora, P. pinifolia, P. pisi, P.*

522   *pistaciae, P. robiniae, P. rubi, P. syringae, P. uliginosa,* and *P. vignae*). Two samples, V8-grown mycelia

523   and either Plich-grown mycelia or germinated cysts, were sequenced for each. Twenty-one species were

524   sequenced by the BGI using custom library construction protocol: random hexamer-primer were used to

525   synthesize the first-strand cDNA; second-strand cDNA was synthesized using buffer, dNTPs, RNase H and

526   DNA polymerase I; short fragments were purified with QiaQuick PCR extraction kit resolved with EB buffer

527   and connected with sequencing adaptors. Each *Phytophthora* transcriptome received 90 bp paired end

528   reads. Two of the above species were sequenced by UBC (*P. kernoviae*, and *P. lateralis*). RNA of *P.*

529   *cactorum* was sequenced by NIAB EMR.

530   To create a transcriptome reference for gene predictions, Trinity assemblies (87) were made for each

531   *Phytophthora* species using both RNA sequence samples (--seqType fq --min_contig_length 200). The *de*

532   *novo* transcriptome assemblies were cleaned using the 3-step TransDecoder process (88). A set of

533   predicted protein sequences was made by combining *Phytophthora* protein sequences from GenBank (85)

534   with protein sequences from the six previously sequenced and annotated *Phytophthora* species (*P. sojae*

535   (29), *P. ramorum* (29), *P. infestans* (10), *P. capsici* (31), *P. cinnanomi*

536   (http://genome.jgi.doe.gov/Phyci1/Phyci1.home.html), and *P. parasitica*

537   (https://www.ncbi.nlm.nih.gov/genome/11752?genome_assembly_id=49439)).   Transdecoder.LongOrfs

538   (88) was used to identify the longest open reading frames in the Trinity assembly. BLASTP (89) was used

539   to align the longest ORFs to the set of constructed proteins identified from GenBank (parameters: -

540   max_target_seqs 1 -evalue 1e-5). Finally, Transdecoder.Predict was used to predict the gene structure

541   from the transcriptome assembly.   The resulting cleaned transcriptome assemblies were used in the

542   subsequent gene prediction methods.

543   **De novo repeat identification**

544   Each genome was repeat-masked to create a genome assembly ready for gene prediction as described

545   below.   Repeat elements were *de novo* identified separately by species.   *De novo* predictions were

546   combined along with previously identified *Phytophthora* repeats for species-specific repeat identification.

547   To identify *de novo* discover LTR retrotransposons, LTRharvest (90) and LTR_(91) finder were run on each

548   genome assembly. By species, LTR retrotransposon predictions from both LTRharvest and LTRfinder were

549   condensed   by   coordinates   and   reduced   by   Blat   alignments.   LTR   retrotransposons,   non-LTR

550   retrotransposons, DNA transposons, and other repeat elements were identified following the MAKER

551   'Repeat   Library   Construction-Advanced'   (92)   method

552   (http://weatherby.genetics.utah.edu/MAKER/wiki/index.php/Repeat_Library_Construction-Advanced).

553   This process utilizes the following programs: MITE-Hunter (with default parameters) (93); GenomeTools

554   suffixorator, LTRharvest, LTRdigest (run with 99% and 85% identity) (94); RepeatModeler (95) which calls

555   RECON   (96),   RepeatScout   (97),   TRF   (98),   NSEG   (99)   and   RMBlast

556   (http://www.repeatmasker.org/RMBlast.html); and sequence databases provided by the MAKER 'Repeat

557   Library Construction-Advanced' method.

558   De novo repeat identification was further supplemented by LTR_retriever (100) using results from

559   LTRharvest and LTRfinder. LTR_retriever retrotransposons were classified into sub-families by species.

560 For each assembled genome, predicted repetitive elements identified in the above methods were combined

561 with GIRI RepBase (volume 18, issue 9) (101) *Phytophthora* repeats. Genomes were repeat-masked using

562 RepeatMasker (v 4.0.6, run with the described combined custom *Phytophthora* library and default

563 parameters) (102) and the combined repeat database to create gene prediction-ready genome assemblies.

Gene prediction and annotation in 31 *Phytophthora* species

565 For each species sequenced, gene training models were made with both Augustus (103) and SNAP (104).

566 Augustus was trained using the genome assembly and the set of previously sequenced *Phytophthora*

567 proteins described in the transcriptome sequencing and assembly process in the text above. To train

568 SNAP, for each species BUSCO (105) was run using the Alveolata_Stramenopiles database in genome

569 mode on each genome assembly to identify core orthologs. BUSCO gff files were converted to zff using

570 maker2zff (92), SNAP tools fathom (-categorize 1000, -export 1000), forge, and hmm-assembler were used

571 to create a training HMM. Genes identified as single copy core orthologs were combined and used as the

572 SNAP training set.

573 Several supplementary files were created to run the MAKER gene prediction pipeline. For the 'EST

574 Evidence' section of MAKER the transcriptome result of the three-step TransDecoder process (88) was

575 used as the EST field. In the seven cases where RNA was not sequenced and therefore the TransDecoder

576 transcriptome was not created, the phylogenetically closest species with RNA sequences was used. A

577 concatenation of the gene sequences of previously sequenced *Phytophthora* species and TransDecoder

578 transcriptome assemblies was used for the alt-est field. For each genome assembly, *P. sojae, P. infestans,*

579 and *P. ramorum* gene predictions were combined with five representative TransDecoder-cleaned

580 transcriptome assemblies. To create species-specific alt-est sets the five representative species were

581 selected from within the *Phytophthora* phylogenetic clade (excluding the species of interest). If less than

582 five species received RNA sequencing, nearby clades were used until five transcriptome assemblies were

583 combined. This set of eight gene predictions and transcriptome assemblies was used in the alt-est field.

584 For the 'Protein Homology' section of MAKER all previously identified *Phytophthora* proteins were combined

585 from GenBank together with the six previously sequenced *Phytophthora* species.

586   To predict genes, MAKER was run on the 31 genomes. The repeat-masked genome for each assembly

587   was run using the Augustus and SNAP training models described above and the EST and protein evidence

588   sequence sets as described above.

589   For validation purposes, three sets of BUSCO analyses were run. BUSCO 3.02 was run on the 31 genomic

590   assemblies, on the 24 transcriptome assemblies, and on the 31 sets of predicted proteins. In all cases, the

591   Alveolata_Stramenopiles BUSCO database of 234 single copy orthologs was used.

592   Predicted genes were functionally annotated using BLASTX (89) to align against known proteins. First, all

593   predicted proteins were aligned against all *Phytophthora* species' proteins obtained from the RefSeq non-

594   redundant protein database in NCBI (85, 106) and the UniProt TrEMBL (107) database.   Second, all

595   predicted proteins were aligned against all stramenopile proteins from NCBI RefSeq and UniProt. Third,

596   all predicted proteins were aligned against all Fungi proteins from NCBI RefSeq and UniProt. BLASTX

597   alignments were generated with the following parameter settings: -evalue 1e-5, -max_target-seqs 50 and

598   were further filtered to return only hits that were at least 50% identical for 50% of the length of the subject

599   protein. Protein functional annotation was made from the consensus of the top five protein alignments for

600   each taxonomic classification.   All results per query protein were screened to return proteins with

601   informative functional annotations ranked by the two rounds of alignments.

602   Predicted proteins from MAKER were screened using InterProScan version 5.20-59.0 (108) to identify

603   functional domains. Gene Ontology (GO) (109, 110) terms were obtained from UniProt BLASTX alignments

604   and the InterProScan runs.

605   **Identification of Effector Proteins**

606   The prediction of cytoplasmic effectors of the RxLR and crinkler (CRN) families was performed on the six-

607   frame translations of the *Phytophthora* whole-genome assemblies using the application getorf (EMBOSS

608   suite) (111). We searched for evidence of the presence of the motifs of interest (RxLR+EER motif for RxLR

609   effectors (112), and LxLAK for CRN effectors (10, 42)) in each ORF translation by using a combination of

610   regular expressions using effectR (113).

611   To identify additional effector proteins that may not include one the canonical motifs, thus not recognized

612   by the RxLR or CRN regular expression, we searched against a profile-HMM (114). We built the HMM

613   profile using an intersect of each set of candidate effectors predicted using regular expressions for each

614   sequenced species in the consortium and the previously predicted effectors from the reference genomes

615   of *P. infestans*, *P. ramorum*, and *P. sojae* (10, 29). We searched for additional effectors in all ORF

616   translations against each of HMM profile using the hmmsearch program in HMMER (115) using default

617   threshold parameters. Predicted effectors from the motif method and the HMM method were examined for

618   signal peptides using SignalP 3.0 (116).

619   Putative apoplastic protease inhibitors were annotated by batch BLASTP (E-value < 1e-30) against

620   MEROPS database (117). The glycoside hydrolase proteins were annotated using the carbohydrate-active

621   enzyme database (cazy) annotation web server dbCAN (118). The hidden Markov model (HMM) profile of

622   NLP family (PF05630) was downloaded from Pfam database (119). The hmmsearch program (115) (with

623   default threshold parameters) was used to search for NLP proteins in each genome assembly.

624   **Classification into Orthologous Groups**

625   Predicted proteins from the 31 sequenced *Phytophthora* and from the six previously sequenced

626   *Phytophthora* species were combined into orthologous groups using OrthoMCL (120). Due to the large

627   data set of 37 full genomes, a two-stage process was used. In stage one, proteins were assigned to the

628   online pre-constructed OrthoMCL orthology groups (121). Predicted proteins from the MAKER process

629   were uploaded to the OrthoMCL web site (www.orthomcl.org). This returned a file of proteins assigned to

630   OrthoMCL groups. In stage two, all unassigned proteins were assigned to groups using the stand-alone

631   version of OrthoMCL. All unassigned proteins from all 37 species were combined into a single FASTA

632   sequence file. The protein FASTA file was aligned against itself using BLASTP. The BLASTP output was

633   converted for input into OrthoMCL which was run in 'Mode 4'.

634   **Phylogenetic analysis of single copy orthologs across *Phytophthora* species**

635   To estimate phylogenetic relationships across the 31 genomes and the six previously sequenced

636   *Phytophthora* species, we first identified the single-copy, core orthologous genes shared across all

637   sequenced species. We selected each of the orthology groups that contain exactly one gene from each of

638   the 37 genomes in the orthology construction.

639   We constructed a phylogenetic tree using 61 genes from each species. Each set of orthologous proteins

640   were multiply aligned using MAFFT ver. 7.271 (122, 123). The phylogenetic tree was reconstructed using

641   RAxML (124), using each gene as an independent partition with its own substitution model, bootstrapped

642   1,000 times. Only one tree was calculated using all partitions.

643   **Effector distribution across plant host ranges**

644   The USDA fungal database (https://nt.ars-grin.gov/fungaldatabases) was used to define the number of

645   known plant hosts infected by each *Phytophthora* species considered in this study. With this information,

646   host ranges were classified into four categories defined by the number of host genera containing known

647   hosts: 'Narrow', host species confined to 1 plant host genus (1-3 host species total); 'Multiple', host species

648   confined within 2 to 9 host genera (2 to 32 host species total); 'Wide', host species spanning 16-55 host

649   genera (22 to 119 host species total); and 'Huge', host species spanning 107-327 host genera (163-718

650   host species total).

651   Numbers of predicted cytoplasmic and apoplastic effector genes were plotted for each host range. In order

652   to reduce errors caused by genome assembly artifacts, and to limit counts of functionally identical effector

653   genes, near-identical paralogs were removed from the counts. To identify near-identical paralogs, effector

654   amino acid sequences were aligned to one another using the Smith-Waterman local aligner from the

655   EMBOSS package (111) to identify similarity. Effectors with greater than 95% amino acid identity over the

656   full sequence length were reduced to one representative effector sequence. The resulting reduced set of

657   effector predictions was used for analysis of the relationship of effector repertoires to host range. Host

658   ranges were plotted for each effector type using ggplot2 (125) in R (84).

659   **Horizontal Gene Transfer**

660   We used a two-step process to identify HGT gene candidates in *Phytophthora* genomes. In the first step,

661   we used a SVM classifier to predict HGT candidates. In the second step, we applied two filters to screen

662   out false positive candidates and assess the likelihood that the candidates were acquired through HGT.

663 **Support Vector Machine classifier for prediction of Horizontal Gene Transfer candidates**

664 We hypothesized that DNA sequence-composition features such as G + C content, codon bias and codon

665 usage frequency (126) can be used to identify genes of recent bacterial or fungal origin in *Phytophthora*

666 genomes. We constructed a multiclass SVM; (127) for composition-based analysis of *Phytophthora*

667 protein-coding genes and classification as either *Phytophthora*, bacterial or fungal origin. SVM is well suited

668 for sequence-composition classification because of the availability of SVM libraries that perform well with

669 large data sets with numerous variables and the ability of SVM to minimize unimportant features (128). The

670 SVM algorithm was implemented in a custom Python script using the SVC function, available from *Scikit-*

671 *learn* Python library (128).

672 Training sets consisted of 15,000 each of ascomycete, *Phytophthora* and bacterial transcripts, for a total of

673 45,000 transcripts. Ascomycete transcripts were selected by submitting a collection of complete transcript

674 sets predicted from the genomes of representative species of eight main ascomycete classes: *Tuber*

675 *melanosporum* (Pezizomycetes; GCA_000151645) (129), *Arthrobotrys oligosporus* (Orbiliomycetes;

676 GCA_000225545) (130), *Penicillium chrysogenum* (Eurotiomycetes; GCA_000226395) (131),

677 *Leptosphaeria maculans* (Dothideomycetes; GCA_000230375) (132), *Cladonia grayi* (Lecanoromycetes)

678 (133), *Sclerotinia sclerotiorum* (Leotiomycetes; GCA_000146945) (134), *Fusarium graminearum*

679 (Sordariomycetes; GCA_000240135.3) (135), and *Xylona heveae* (Xylonomycetes; GCA_001619985)

680 (136). Potential genes that underwent HGT were discarded from each transcript set by applying the

681 following protocol: 1, transcripts were translated into proteins and clustered using OrthoMCL (coverage and

682 identity of at least 50%, E-value cut-off of 1e-05, inflation parameter = 2.5) (120); 2, one protein from each

683 cluster was then queried against the NCBI nr database (max target sequences = 500); 3, clusters with at

684 least one hit in any other organisms than a fungal taxon were discarded; 4, for each remaining cluster, each

685 protein was queried against nr (max target sequences = 500) and step 3 was re-applied. *Phytophthora*

686 genes were selected by the same process, using transcripts from the following species: *P. syringae* (this

687 study), *P. sojae* (GCA_000149755) (29), *P. ramorum* (GCA_000149735) (29), *P. lateralis*

688 (GCA_000500205) (73), *P. pinifolia* (GCA_000500225) (73), *P. cryptogea* (GCA_000468175) (73), *P.*

689 *infestans* (GCA_000142945) (10), *P. brassicae* (this study), and *P. kernoviae* (GCA_000448265) (73) and

690   eliminating clusters with any protein match other than with an oomycete taxon. Bacterial transcripts were

691   selected following the same filtering approach on 21,096 transcripts retrieved from GenBank (representing

692   23 bacterial classes).

693   Sequence-composition features were used as input vectors to an SVM classifier and the curated training

694   sets (see above) were used as model data. Following a preliminary analysis, codon usage frequency and

695   GC content were selected as the sequence features as they resulted in a higher prediction accuracy than

696   codon bias (0.976 ±0.002 *vs.* 0.973 ±0.004, t = 10.4, p <0.0001; data not shown). This is consistent with

697   the point that codon use frequency is inherently the fusion of both codon usage bias and amino acid

698   composition signals (137). To choose the best kernel for the SVM, we first used principal component

699   analysis (PCA) to explore the relationships among the three different classes (Fig. S4). Radial basis function

700   (rbf) kernel parameters (*C* and *gamma*) were systematically varied to optimize prediction accuracy using a

701   two-dimensional grid where both parameters were chosen from the set $\{10^{-3}, 10^{-2}, …, 10^{6}\}$. All these

702   optimizations were performed with fivefold cross-validation of the training set (randomly withholding one-

703   fifth of the training data as a testing data set; 100 random draws for each pair of parameters tested) (Fig.

704   S4). Accuracy as defined by (TP + TN) / (TP + TN + FP + FN) was used as a measure of the quality of the

705   classification. Best classification accuracy (98.3%) was obtained with rbf kernel parameters of *C* = 1000.0

706   and *gamma* = 1.0 (Fig. S4).

707   ***Phytophthora* transcript classification for Horizontal Gene Transfer**

708   The 618,240 transcripts predicted from the 31 genomes and 103,992 transcripts predicted from five

709   previously sequenced *Phytophthora* species (i.e. *P. sojae*, *P. ramorum*, *P. infestans*, *P. capsici* and *P.*

710   *cinnamomi* [see *Transcriptome sequencing and assembly*]) were submitted to the classifier and sorted into

711   *Phytophthora*-origin, bacterial-origin or fungal-origin classes depending on the probability returned by the

712   classifier for each of these classes. To generate a confidence score, we repeated the training of the

713   classifier 100 times before running the classification on each genome. To maximize the training process of

714   the classifier without increasing computing time and overloading memory we used a random subsample of

715   45,000 transcripts (15,000 genes in each of the three classes) as a training set each time. Preliminarily, we

716   determined the minimum threshold number of bootstrap replicates in which an HGT candidate was found

717   that would minimize the probability that this candidate was a false positive (e.g., misclassifying a

718   *Phytophthora* or a bacteria sequence as deriving from a fungal donor via HGT). This was done by submitting

719   a subsample of 1,500 sequences randomly picked in the training set (500 transcripts in each class) to the

720   classifier with the bootstrap procedure; then, false positive and true positive rates were calculated for

721   incremental values of bootstrap replicates. Based on this test, we determined that the chance of

722   misclassifying a fungal transcript as a *Phytophthora* or a bacterial sequence (i.e. a false positive) was <

723   0.2% (1/500) if it was classified as fungal in at least 89/100 bootstrap replicates; in such case the true

724   positive rate (recall; TP/(TP + FN)) would be 92.3% (Fig. S5A). For the bacterial sequences, this value was

725   ≥79/100 bootstrap replicates; this corresponded to a true positive rate (recall) of 97.5% (Fig. S5B). These

726   two bootstrap replicate thresholds were then used for the identification of HGT candidates in *Phytophthora*

727   species.

728   **Horizontal Gene Transfer candidate false positive filtering**

729   HGT candidates predicted with the SVM classifier were submitted to a phylogenetic filtering step by

730   assessing the congruence of the gene phylogeny with the organism phylogeny. Each candidate transcript

731   was translated into a protein sequence and searched using DIAMOND BLASTp (minimum BLASTP E-value

732   of 1e-03, sequence subject coverage of 50% and sequence query coverage of 50%) (138) for closest

733   homologs against protein-coding sequences downloaded from the NCBI Reference Sequence collection

734   (RefSeq) (106) for *Phytophthora* species. (72,639 sequences) and the following clades: heterokonts

735   (excluding *Phytophthora*; 164,619 sequences), Alveolata (1,527,928 sequences), Amoebozoa (113,408

736   sequences), Viridiplantae (5,556,940 sequences), Fungi (2,912,973 sequences), Archaea (1,830,006

737   sequences) and Bacteria (131,971,793 sequences) clades. Candidates with no hits in the Bacteria or Fungi

738   clades were directly rejected. Protein sequences for the top three DIAMOND BLASTp hits within each of

739   the above clade were retrieved and aligned with the query protein using MAFFT ver. 7.271 (123). Amino-

740   acid sites with a gap in more than one third of the sequences were removed. IQ-TREE was used to

741   determine the best-fitting substitution model and reconstruct a maximum-likelihood tree for each of the

742   alignments (139, 140) with node support assessed by using the ultrafast bootstrap approximation method

743   (141). Each phylogenetic tree was exported in Newick format, automatically rooted with sequences from

744   the Archaea or Bacteria clades and exported into a .png file using the BioPython package Phylo. For

745   facilitating visual examination of trees, png files were gathered into one single pdf catalog with the Python

746   library PyFPDF. Tree nodes were visually inspected to identify phylogenetic discordance (142). Two types

747   of discordance were examined: 1, "complete incongruence", when the HGT candidate sequence clusters

748   with the fungi or bacteria clade with bootstrap support ≥50%, resulting in a phylogenetic tree completely

749   discordant with the expected organism phylogeny i.e. ((((((Phytophthora, Heterokonta), Alveolata),

750   Viridiplantae), (Fungi, Amoebozoa)), Archaea), Bacteria) (143, 144) 2, "missing clades", when the HGT

751   candidate sequence clusters with the fungi or bacteria clade because other clades are missing (i.e. the

752   HGT candidate sequence didn't have orthologs in intermediate clades such as Viriplantae, Amoebozoa and

753   Alveolata).

754   HGT candidates that passed phylogenetic filtering were submitted to a sequence identity discrepancy filter.

755   Assuming a molecular clock, the sequence identity between a pair of orthologous genes should be in the

756   same range as the average sequence identity between their respective species. However, for a pair of

757   sequences related through an HGT event between two species, the proportionality should be broken,

758   leading to an identity discrepancy when compared to the pairwise species identity (145). To identify such

759   discrepancies, we performed a first "identity test": We calculated the nucleotide sequence identity between

760   the candidate HGT sequence in *Phytophthora* and its closest homolog sequences in the putative "donor"

761   species in Bacteria or Fungi. Then, the full transcriptome of the putative "donor" species was downloaded

762   and searched with BLASTN for 1,000 random transcripts from the *Phytophthora* species to identify one-to-

763   one orthologs and plot a distribution of expected nucleotide identity values. In a first "identity test",

764   discrepancies were identified by comparing the observed nucleotide sequence identity found between the

765   HGT candidate in *Phytophthora* and its homolog sequence in the putative donor to the expected distribution

766   using a Wilcoxon sign-rank test; Candidates were rejected if the difference between the observed value

767   (sequence identity between the HGT candidate in *Phytophthora* and its homolog sequence in the putative

768   donor) and the average of the expected distribution was not significant and lower than an arbitrary

769   "discrepancy cutoff" of 8.56 (corresponding to the top quartile of the distribution of the difference between

770   the observed values and the expected values). Proteins retained at this step were included into a "relaxed"

771   list of candidates. To ascertain that the discrepancy was not caused by a high conservation of the gene

772   among the different clades, we validated this list with a second "identity test" that consisted of examining if

773   the nucleotide sequence identity between the HGT candidate and the closest species in the non-donor

774   clades was not significantly higher than the average identity expected between the two species. Only

775   candidates for which the difference in nucleotide sequence identity between the HGT candidate in

776   *Phytophthora* and its homolog sequence in the putative donor was lower than 52% (corresponding to the

777   uppermost quartile of the distribution of the difference between the observed values and the expected

778   values of nucleotide sequence identity between pairs of homologs from the two species) were retained after

779   this stage. When the second "identity test" could not be performed due to the absence of sequence

780   homologs in non-donor clades (i.e. Viriplantae, Amoebozoa and Alveolata), the "discrepancy cutoff" i.e. the

781   difference between the nucleotide sequence identity for the HGT candidate in *Phytophthora* and its homolog

782   sequence in the putative donor and the average of the expected distribution was raised to 80.83% identity

783   corresponding to the 5% upper quantile of the of the distribution of the difference between the observed

784   values and the expected values. Proteins that passed this second filter were kept in a "strict" list of

785   candidates.

786   We assessed if the HGT candidate could have been a bacterial or fungal contaminant mistakenly

787   sequenced and assembled with a genome assembly generated in this study. HGT candidates were

788   considered as a putative "contaminant" candidates if they were the only coding sequences to map to a

789   given scaffold.

790   **Data Availability**

791   Genome assemblies and genome and transcriptome sequences that were created in this study have been

792   deposited in NCBI under BioProjects PRJNA746351, PRJNA714689, and PRJNA702516. *P. cactorum* and

793   *P. idaei* data are available under BioProjects PRJNA383548 and PRJNA391273. Other genomic resources

794   including assembly files, gene/protein predictions and annotations, differential expression analysis, and

795   orthology analysis can be accessed at https://phyto-seq.cqls.oregonstate.edu.

796   **Acknowledgments**

810

811   **References**

812   1.   Erwin DC, Ribeiro OK. 1996. *Phytophthora* diseases worldwide. American Phytopathological
813   Society (APS Press), St. Paul, Minnesota.

814   2.   Judelson HS, Blanco FA. 2005. The spores of *Phytophthora*: weapons of the plant destroyer. Nat
815   Rev Microbiol 3:47–58.

816   3.   Drake B, Jones G. 2017. Public value at risk from *Phytophthora ramorum* and *Phytophthora*
817   *kernoviae* spread in England and Wales. J Environ Manage 191.

818   4.   Gunderson JH, Elwood H, Ingold A, Kindle K, Sogin ML. 1987. Phylogenetic relationships between
819   chlorophytes, chrysophytes, and oomycetes. Proceedings of the National Academy of Sciences
820   84:5823–5827.

821   5.   Jiang RHY, Tyler BM. 2012. Mechanisms and Evolution of Virulence in Oomycetes. Annu Rev
822   Phytopathol 50:295–318.

823   6.   Thines M, Kamoun S. 2010. Oomycete–plant coevolution: recent advances and future prospects.
824   Curr Opin Plant Biol 13:427–433.

825   7.   Thines M. 2014. Phylogeny and evolution of plant pathogenic oomycetes—a global overview. Eur
826   J Plant Pathol 138:431–447.

827   8.   Dick MW. 2001. The Peronosporomycetes, p. 39–72. *In* Systematics and Evolution.

828   9.   TURNER RS. 2005. After the famine: Plant pathology, *Phytophthora infestans*, and the late blight
829   of potatoes, 1845––1960. Historical Studies in the Physical and Biological Sciences 35:341–370.

830   10.   Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, Grabherr M, Kodira CD,
831   Raffaele S, Torto-Alalibo T, Bozkurt TO, Ah-Fong AM v., Alvarado L, Anderson VL, Armstrong MR,
832   Avrova A, Baxter L, Beynon J, Boevink PC, Bollmann SR, Bos JIB, Bulone V, Cai G, Cakir C,
833   Carrington JC, Chawner M, Conti L, Costanzo S, Ewan R, Fahlgren N, Fischbach MA, Fugelstad J,
834   Gilroy EM, Gnerre S, Green PJ, Grenville-Briggs LJ, Griffith J, Grünwald NJ, Horn K, Horner NR, Hu
835   C-H, Huitema E, Jeong D-H, Jones AME, Jones JDG, Jones RW, Karlsson EK, Kunjeti SG, Lamour K,
836   Liu Z, Ma L, MacLean D, Chibucos MC, McDonald H, McWalters J, Meijer HJG, Morgan W, Morris
837   PF, Munro CA, O'Neill K, Ospina-Giraldo M, Pinzón A, Pritchard L, Ramsahoye B, Ren Q, Restrepo
838   S, Roy S, Sadanandom A, Savidor A, Schornack S, Schwartz DC, Schumann UD, Schwessinger B,
839   Seyer L, Sharpe T, Silvar C, Song J, Studholme DJ, Sykes S, Thines M, van de Vondervoort PJI,
840   Phuntumart V, Wawra S, Weide R, Win J, Young C, Zhou S, Fry W, Meyers BC, van West P,
841   Ristaino J, Govers F, Birch PRJ, Whisson SC, Judelson HS, Nusbaum C. 2009. Genome sequence
842   and analysis of the Irish potato famine pathogen Phytophthora infestans. Nature 461:393–398.

843   11.   Goheen EM, Hansen EM, Kanaskie A, McWilliams MG, Osterbauer N, Sutton W. 2002. Sudden
844   Oak Death Caused by *Phytophthora ramorum* in Oregon. Plant Dis 86:441–441.

845   12.   Rizzo DM, Garbelotto M, Hansen EM. 2005. *Phytophthora ramorum*: Integrative Research and
846   Management of an Emerging Pathogen in California and Oregon Forests. Annu Rev Phytopathol
847   43:309–335.

848   13.   Rizzo DM, Garbelotto M, Davidson JM, Slaughter GW, Koike ST. 2002. *Phytophthora ramorum* as
849   the Cause of Extensive Mortality of *Quercus* spp. and *Lithocarpus densiflorus* in California. Plant
850   Dis 86:205–214.

851   14.   GRÜNWALD NJ, GOSS EM, PRESS CM. 2008. *Phytophthora ramorum* : a pathogen with a
852   remarkably wide host range causing sudden oak death on oaks and ramorum blight on woody
853   ornamentals. Mol Plant Pathol 9.

854   15.   TYLER BM. 2007. *Phytophthora sojae*: root rot pathogen of soybean and model oomycete. Mol
855   Plant Pathol 8:1–8.

856   16.   KENNEDY DM, DUNCAN JM, DUGARD PI, TOPHAM PH. 1986. Virulence and aggressiveness of
857   single-zoospore isolates of *Phytophthora fragariae*. Plant Pathol 35:344–354.

858   17.   Kao CW, Leu LS. 1980. Sporangium Germination of *Peronophythora Litchii* , The Causal Organism
859   of Litchi Downy Blight. Mycologia 72:737–748.

860   18.   HARDHAM AR. 2005. *Phytophthora cinnamomi*. Mol Plant Pathol 6:589–604.

861   19.   Raffaele S, Kamoun S. 2012. Genome evolution in filamentous plant pathogens: why bigger can
862   be better. Nat Rev Microbiol 10:417–430.

863   20.   Yang X, Tyler BM, Hong C. 2017. An expanded phylogeny for the genus Phytophthora. IMA
864   Fungus 8:355–384.

865   21.   Kroon LPNM, Brouwer H, de Cock AWAM, Govers F. 2012. The Genus *Phytophthora* Anno 2012.
866   Phytopathology 102:348–364.

867   22.   Jung T, Jung MH, Cacciola SO, Cech T, Bakonyi J, Seress D, Mosca S, Schena L, Seddaiu S, Pane A,
868   di San Lio GM, Maia C, Cravador A, Franceschini A, Scanu B. 2017. Multiple new cryptic
869   pathogenic Phytophthora species from Fagaceae forests in Austria, Italy and Portugal. IMA
870   Fungus 8.

871   23.   Brasier C, Scanu B, Cooke D, Jung T. 2022. *Phytophthora*: an ancient, historic, biologically and
872   structurally cohesive and evolutionarily successful generic concept in need of preservation. IMA
873   Fungus 13:12.

874   24.   Dong S, Raffaele S, Kamoun S. 2015. The two-speed genomes of filamentous pathogens: waltz
875   with plants. Curr Opin Genet Dev 35:57–65.

876   25.   Ye W, Wang Y, Shen D, Li D, Pu T, Jiang Z, Zhang Z, Zheng X, Tyler BM, Wang Y. 2016. Sequencing
877   of the Litchi Downy Blight Pathogen Reveals It Is a *Phytophthora* Species With Downy Mildew-
878   Like Characteristics. Molecular Plant-Microbe Interactions® 29:573–583.

879   26.   Ali SS, Shao J, Lary DJ, Kronmiller BA, Shen D, Strem MD, Amoako-Attah I, Akrofi AY, Begoude
880   BAD, ten Hoopen GM, Coulibaly K, Kebe BI, Melnick RL, Guiltinan MJ, Tyler BM, Meinhardt LW,
881   Bailey BA. 2017. *Phytophthora megakarya* and *Phytophthora palmivora*, Closely Related Causal
882   Agents of Cacao Black Pod Rot, Underwent Increases in Genome Sizes and Gene Numbers by
883   Different Mechanisms. Genome Biol Evol 9:536–557.

884   27.   Tabima JF, Kronmiller BA, Press CM, Tyler BM, Zasada IA, Grünwald NJ. 2017. Whole Genome
885   Sequences of the Raspberry and Strawberry Pathogens *Phytophthora rubi* and *P. fragariae*.
886   Molecular Plant-Microbe Interactions® 30:767–769.

887   28.   Armitage AD, Lysøe E, Nellist CF, Lewis LA, Cano LM, Harrison RJ, Brurberg MB. 2018.
888   Bioinformatic characterisation of the effector repertoire of the strawberry pathogen
889   Phytophthora cactorum. PLoS One 13:e0202305.

890   29.   Tyler BM. 2006. Phytophthora Genome Sequences Uncover Evolutionary Origins and
891   Mechanisms of Pathogenesis. Science (1979) 313:1261–1266.

892   30.   Knaus BJ, Tabima JF, Shakya SK, Judelson HS, Grünwald NJ. 2020. Genome-Wide Increased Copy
893   Number is Associated with Emergence of Dominant Clones of the Irish Potato Famine Pathogen
894   Phytophthora infestans. mBio 11.

895   31.   Lamour KH, Mudge J, Gobena D, Hurtado-Gonzales OP, Schmutz J, Kuo A, Miller NA, Rice BJ,
896   Raffaele S, Cano LM, Bharti AK, Donahoo RS, Finley S, Huitema E, Hulvey J, Platt D, Salamov A,
897   Savidor A, Sharma R, Stam R, Storey D, Thines M, Win J, Haas BJ, Dinwiddie DL, Jenkins J, Knight
898   JR, Affourtit JP, Han CS, Chertkov O, Lindquist EA, Detter C, Grigoriev I v., Kamoun S, Kingsmore
899   SF. 2012. Genome Sequencing and Mapping Reveal Loss of Heterozygosity as a Mechanism for
900   Rapid Adaptation in the Vegetable Pathogen *Phytophthora capsici*. Molecular Plant-Microbe
901   Interactions® 25:1350–1360.

902   32.   McGowan J, Fitzpatrick DA. 2017. Genomic, Network, and Phylogenetic Analysis of the Oomycete
903   Effector Arsenal. mSphere 2.

904   33.   Asai S, Shirasu K. 2015. Plant cells under siege: plant immune system versus pathogen effectors.
905   Curr Opin Plant Biol 28:1–8.

906   34.   Wang W, Jiao F. 2019. Effectors of *Phytophthora* pathogens are powerful weapons for
907   manipulating host immunity. Planta 250:413–425.

908   35.   Wang S, Boevink PC, Welsh L, Zhang R, Whisson SC, Birch PRJ. 2017. Delivery of cytoplasmic and
909   apoplastic effectors from *Phytophthora infestans* haustoria by distinct secretion pathways. New
910   Phytologist 216:205–215.

911   36.   Dong S, Kong G, Qutob D, Yu X, Tang J, Kang J, Dai T, Wang H, Gijzen M, Wang Y. 2012. The NLP
912   Toxin Family in *Phytophthora sojae* Includes Rapidly Evolving Groups That Lack Necrosis-Inducing
913   Activity. Molecular Plant-Microbe Interactions® 25:896–909.

914   37.   Tian M, Win J, Song J, van der Hoorn R, van der Knaap E, Kamoun S. 2007. A *Phytophthora
915   infestans* Cystatin-Like Protein Targets a Novel Tomato Papain-Like Apoplastic Protease. Plant
916   Physiol 143:364–377.

917   38.   Jiang RHY, Tripathy S, Govers F, Tyler BM. 2008. RXLR effector reservoir in two *Phytophthora*
918   species is dominated by a single rapidly evolving superfamily with more than 700 members.
919   Proceedings of the National Academy of Sciences 105.

920   39.   Morgan W, Kamoun S. 2007. RXLR effectors of plant pathogenic oomycetes. Curr Opin Microbiol
921   10:332–338.

922   40.   Zhang M, Li Q, Liu T, Liu L, Shen D, Zhu Y, Liu P, Zhou J-M, Dou D. 2014. Two Cytoplasmic Effectors
923   of *Phytophthora sojae* Regulate Plant Cell Death via Interactions with Plant Catalases. Plant
924   Physiol 167.

925   41.   Schornack S, van Damme M, Bozkurt TO, Cano LM, Smoker M, Thines M, Gaulin E, Kamoun S,
926   Huitema E. 2010. Ancient class of translocated oomycete effectors targets the host nucleus.
927   Proceedings of the National Academy of Sciences 107.

928   42.   Stam R, Jupe J, Howden AJM, Morris JA, Boevink PC, Hedley PE, Huitema E. 2013. Identification
929   and Characterisation CRN Effectors in *Phytophthora capsici* Shows Modularity and Functional
930   Diversity. PLoS One 8.

931   43.   Win J, Kanneganti T-D, Torto-Alalibo T, Kamoun S. 2006. Computational and comparative
932   analyses of 150 full-length cDNA sequences from the oomycete plant pathogen *Phytophthora
933   infestans*. Fungal Genetics and Biology 43:20–33.

934   44.   Torto TA. 2003. EST Mining and Functional Expression Assays Identify Extracellular Effector
935   Proteins From the Plant Pathogen *Phytophthora*. Genome Res 13:1675–1685.

936   45.   Liu T, Song T, Zhang X, Yuan H, Su L, Li W, Xu J, Liu S, Chen L, Chen T, Zhang M, Gu L, Zhang B, Dou
937   D. 2014. Unconventionally secreted effectors of two filamentous pathogens target plant
938   salicylate biosynthesis. Nat Commun 5.

939   46.   Oh S-K, Young C, Lee M, Oliva R, Bozkurt TO, Cano LM, Win J, Bos JIB, Liu H-Y, van Damme M,
940   Morgan W, Choi D, van der Vossen EAG, Vleeshouwers VGAA, Kamoun S. 2009. In Planta
941   Expression Screens of *Phytophthora infestans* RXLR Effectors Reveal Diverse Phenotypes,
942   Including Activation of the *Solanum bulbocastanum* Disease Resistance Protein Rpi-blb2. Plant
943   Cell 21:2928–2947.

944   47.   Wang Q, Han C, Ferreira AO, Yu X, Ye W, Tripathy S, Kale SD, Gu B, Sheng Y, Sui Y, Wang X, Zhang
945   Z, Cheng B, Dong S, Shan W, Zheng X, Dou D, Tyler BM, Wang Y. 2011. Transcriptional
946   Programming and Functional Interactions within the *Phytophthora sojae* RXLR Effector
947   Repertoire   . Plant Cell 23:2064–2086.

948   48.   Torto-Alalibo T, Collmer CW, Gwinn-Giglio M, Lindeberg M, Meng S, Chibucos MC, Tseng T-T,
949   Lomax J, Biehl B, Ireland A, Bird D, Dean RA, Glasner JD, Perna N, Setubal JC, Collmer A, Tyler BM.
950   2010. Unifying Themes in Microbial Associations with Animal and Plant Hosts Described Using the
951   Gene Ontology. Microbiology and Molecular Biology Reviews 74.

952   49.   Wang Y, Tyler BM, Wang Y. 2019. Defense and Counterdefense During Plant-Pathogenic
953   Oomycete Infection. Annu Rev Microbiol 73.

954   50.   Bozkurt TO, Schornack S, Win J, Shindo T, Ilyas M, Oliva R, Cano LM, Jones AME, Huitema E, van
955   der Hoorn RAL, Kamoun S. 2011. *Phytophthora infestans* effector AVRblb2 prevents secretion of
956   a plant immune protease at the haustorial interface. Proceedings of the National Academy of
957   Sciences 108:20832–20837.

958   51.   Caillaud M, Piquerez SJM, Fabro G, Steinbrenner J, Ishaque N, Beynon J, Jones JDG. 2012.
959   Subcellular localization of the *Hpa* RxLR effector repertoire identifies a tonoplast-associated
960   protein HaRxL17 that confers enhanced plant susceptibility. The Plant Journal 69:252–265.

961   52.   Shen D, Liu T, Ye W, Liu L, Liu P, Wu Y, Wang Y, Dou D. 2013. Gene Duplication and Fragment
962   Recombination Drive Functional Diversification of a Superfamily of Cytoplasmic Effectors in
963   *Phytophthora sojae*. PLoS One 8.

964   53.   Ye W, Wang Y, Tyler BM, Wang Y. 2016. Comparative Genomic Analysis among Four
965   Representative Isolates of *Phytophthora sojae* Reveals Genes under Evolutionary Selection. Front
966   Microbiol 7.

967   54.   Gijzen M. 2009. Runaway repeats force expansion of the Phytophthora infestans genome.
968   Genome Biol 10:241.

969   55.   Wang L, Chen H, Li J, Shu H, Zhang X, Wang Y, Tyler BM, Dong S. 2020. Effector gene silencing
970   mediated by histone methylation underpins host adaptation in an oomycete plant pathogen.
971   Nucleic Acids Res 48.

972   56.   Blair JE, Coffey MD, Park S-Y, Geiser DM, Kang S. 2008. A multi-locus phylogeny for Phytophthora
973   utilizing markers derived from complete genome sequences. Fungal Genetics and Biology
974   45:266–277.

975   57.   Nellist CF, Armitage AD, Bates HJ, Sobczyk MK, Luberti M, Lewis LA, Harrison RJ. 2021.
976   Comparative Analysis of Host-Associated Variation in Phytophthora cactorum. Front Microbiol
977   12.

978   58.   van der Lee T, de Witte I, Drenth A, Alfonso C, Govers F. 1997. AFLP Linkage Map of the
979   OomycetePhytophthora infestans. Fungal Genetics and Biology 21:278–291.

980   59.   McCarthy CGP, Fitzpatrick DA. 2017. Phylogenomic Reconstruction of the Oomycete Phylogeny
981   Derived from 37 Genomes. mSphere 2.

982   60.   Martin FN, Blair JE, Coffey MD. 2014. A combined mitochondrial and nuclear multilocus
983   phylogeny of the genus Phytophthora. Fungal Genetics and Biology 66.

984   61.   Richards TA, Soanes DM, Jones MDM, Vasieva O, Leonard G, Paszkiewicz K, Foster PG, Hall N,
985   Talbot NJ. 2011. Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms
986   in the oomycetes. Proceedings of the National Academy of Sciences 108.

987   62.   Urban M, Cuzick A, Seager J, Wood V, Rutherford K, Venkatesh SY, De Silva N, Martinez MC,
988   Pedro H, Yates AD, Hassani-Pak K, Hammond-Kosack KE. 2019. PHI-base: the pathogen–host
989   interactions database. Nucleic Acids Res https://doi.org/10.1093/nar/gkz904.

990   63.   Richards TA, Dacks JB, Jenkinson JM, Thornton CR, Talbot NJ. 2006. Evolution of Filamentous
991   Plant Pathogens: Gene Exchange across Eukaryotic Kingdoms. Current Biology 16.

992   64.   McCarthy CGP, Fitzpatrick DA. 2016. Systematic Search for Evidence of Interdomain Horizontal
993   Gene Transfer from Prokaryotes to Oomycete Lineages. mSphere 1.

994   65.   Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet 9.

995   66.   Husnik F, McCutcheon JP. 2018. Functional horizontal gene transfer from bacteria to eukaryotes.
996   Nat Rev Microbiol 16.

997   67.   Feurtey A, Stukenbrock EH. 2018. Interspecific Gene Exchange as a Driver of Adaptive Evolution
998   in Fungi. Annu Rev Microbiol 72.

999   68.   van Etten J, Bhattacharya D. 2020. Horizontal Gene Transfer in Eukaryotes: Not if, but How
1000   Much? Trends in Genetics 36.

1001 69.   Soucy SM, Huang J, Gogarten JP. 2015. Horizontal gene transfer: building the web of life. Nat Rev
1002   Genet 16.

1003 70.   Möller EM, Bahnweg G, Sandermann H, Geiger HH. 1992. A simple and efficient protocol for
1004   isolation of high molecular weight DNA from filamentous fungi, fruit bodies, and infected plant
1005   tissues. Nucleic Acids Res 20:6115–6116.

1006 71.   Johansen LK, Carrington JC. 2001. Silencing on the Spot. Induction and Suppression of RNA
1007   Silencing in the *Agrobacterium* -Mediated Transient Expression System. Plant Physiol 126:930–
1008   938.

1009 72.   Kamoun S. 1993. Extracellular Protein Elicitors from *Phytophthora* : Host-Specificity and Induction
1010   of Resistance to Bacterial and Fungal Phytopathogens. Molecular Plant-Microbe Interactions 6.

1011 73. Feau N, Taylor G, Dale AL, Dhillon B, Bilodeau GJ, Birol I, Jones SJM, Hamelin RC. 2016. Genome
1012   sequences of six *Phytophthora* species threatening forest ecosystems. Genom Data 10:85–88.

1013 74. Studholme DJ, McDougal RL, Sambles C, Hansen E, Hardy G, Grant M, Ganley RJ, Williams NM.
1014   2016. Genome sequences of six *Phytophthora* species associated with forests in New Zealand.
1015   Genom Data 7:54–56.

1016 75. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads.
1017   EMBnet J 17:10.

1018 76. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y,
1019   Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu S-M, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang
1020   H, Wang J, Lam T-W, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient
1021   short-read de novo assembler. Gigascience 1:18.

1022 77. Gu S, Fang L, Xu X. 2013. Using SOAPaligner for Short Reads Alignment. Curr Protoc
1023   Bioinformatics 44.

1024 78. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. 2009. ABySS: A parallel assembler
1025   for short read sequence data. Genome Res 19:1117–1123.

1026 79. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI,
1027   Pham S, Prjibelski AD, Pyshkin A v., Sirotkin A v., Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.
1028   2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing.
1029   Journal of Computational Biology 19:455–477.

1030 80. Boetzer M, Henkel C v., Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs
1031   using SSPACE. Bioinformatics 27:578–579.

1032 81. Aronesty E. 2013. Comparison of Sequencing Utility Programs. Open Bioinforma J 7:1–8.

1033 82. Zerbino DR, Birney E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn
1034   graphs. Genome Res 18:821–829.

1035 83. Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of
1036   occurrences of k-mers. Bioinformatics 27.

1037 84. R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for
1038   Statistical Computing, Vienna, Austria. https://www.R-project.org/.

1039 85. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2016. GenBank. Nucleic Acids Res
1040   44:D67–D72.

1041 86. Kent WJ. 2002. BLAT---The BLAST-Like Alignment Tool. Genome Res 12:656–664.

1042 87. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
1043   Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren
1044   BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly
1045   from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652.

1046 88. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B,
1047 Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T,
1048 Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. 2013. De novo transcript sequence
1049 reconstruction from RNA-seq using the Trinity platform for reference generation and analysis.
1050 Nat Protoc 8:1494–1512.

1051 89. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J
1052 Mol Biol 215:403–410.

1053 90. Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo
1054 detection of LTR retrotransposons. BMC Bioinformatics 9:18.

1055 91. Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR
1056 retrotransposons. Nucleic Acids Res 35:W265–W268.

1057 92. Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sanchez Alvarado A, Yandell M.
1058 2007. MAKER: An easy-to-use annotation pipeline designed for emerging model organism
1059 genomes. Genome Res 18:188–196.

1060 93. Han Y, Wessler SR. 2010. MITE-Hunter: a program for discovering miniature inverted-repeat
1061 transposable elements from genomic sequences. Nucleic Acids Res 38:e199–e199.

1062 94. Gremme G, Steinbiss S, Kurtz S. 2013. GenomeTools: A Comprehensive Software Library for
1063 Efficient Processing of Structured Genome Annotations. IEEE/ACM Trans Comput Biol Bioinform
1064 10:645–656.

1065 95. Smit A, Hubley R. RepeatModeler Open-1.0. 2008-2015 <http://www.repeatmasker.org>.

1066 96. Bao Z. 2002. Automated De Novo Identification of Repeat Sequence Families in Sequenced
1067 Genomes. Genome Res 12:1269–1276.

1068 97. Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes.
1069 Bioinformatics 21:i351–i358.

1070 98. Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res
1071 27:573–580.

1072 99. Wootton JC, Federhen S. 1996. Methods in Enzymology; Analysis of compositionally biased
1073 regions in sequence databases, p. 554–571. *In* .

1074 100. Ou S, Jiang N. 2018. LTR_retriever: A Highly Accurate and Sensitive Program for Identification of
1075 Long Terminal Repeat Retrotransposons. Plant Physiol 176:1410–1422.

1076 101. Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in
1077 eukaryotic genomes. Mob DNA 6:11.

1078 102. Smit A, Hubley R, Green P. RepeatMasker Open-4.0. 2013-2015 <http://www.repeatmasker.org>.

1079 103. Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that
1080 allows user-defined constraints. Nucleic Acids Res 33:W465–W467.

1081   104.   Korf I. 2004. Gene finding in novel genomes. BMC Bioinformatics 5:59.

1082   105.   Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: Assessing Genome Assembly and Annotation
1083   Completeness, p. 227–245. *In* .

1084   106.   O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-
1085   White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J,
1086   Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS,
1087   Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala
1088   SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully
1089   RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P,
1090   Murphy TD, Pruitt KD. 2016. Reference sequence (RefSeq) database at NCBI: current status,
1091   taxonomic expansion, and functional annotation. Nucleic Acids Res 44.

1092   107.   2019. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47:D506–D515.

1093   108.   Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka
1094   G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014.
1095   InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240.

1096   109.   2019. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res 47:D330–
1097   D338.

1098   110.   Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS,
1099   Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE,
1100   Ringwald M, Rubin GM, Sherlock G. 2000. Gene Ontology: tool for the unification of biology. Nat
1101   Genet 25:25–29.

1102   111.   Rice P, Longden I, Bleasby A. 2000. EMBOSS: The European Molecular Biology Open Software
1103   Suite. Trends in Genetics 16:276–277.

1104   112.   Win J, Morgan W, Bos J, Krasileva K v., Cano LM, Chaparro-Garcia A, Ammar R, Staskawicz BJ,
1105   Kamoun S. 2007. Adaptive Evolution Has Targeted the C-Terminal Domain of the RXLR Effectors
1106   of Plant Pathogenic Oomycetes. Plant Cell 19:2349–2369.

1107   113.   Tabima JF, Grünwald NJ. 2019. *effectR* : An Expandable R Package to Predict Candidate RxLR and
1108   CRN Effectors in Oomycetes Using Motif Searches. Molecular Plant-Microbe Interactions® 32.

1109   114.   Eddy SR. 1998. Profile hidden Markov models. Bioinformatics 14:755–763.

1110   115.   Eddy SR. HMMER www.hmmer.org.

1111   116.   Dyrløv Bendtsen J, Nielsen H, von Heijne G, Brunak S. 2004. Improved Prediction of Signal
1112   Peptides: SignalP 3.0. J Mol Biol 340.

1113   117.   Rawlings ND, Barrett AJ, Bateman A. 2012. MEROPS: the database of proteolytic enzymes, their
1114   substrates and inhibitors. Nucleic Acids Res 40:D343–D350.

1115   118.   Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated
1116   carbohydrate-active enzyme annotation. Nucleic Acids Res 40:W445–W451.

1117   119.   El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar
1118   GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The
1119   Pfam protein families database in 2019. Nucleic Acids Res 47:D427–D432.

1120   120.   Li L. 2003. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. Genome Res
1121   13:2178–2189.

1122   121.   Chen F. 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog
1123   groups. Nucleic Acids Res 34:D363–D368.

1124   122.   Nakamura T, Yamada KD, Tomii K, Katoh K. 2018. Parallelization of MAFFT for large-scale multiple
1125   sequence alignments. Bioinformatics 34:2490–2492.

1126   123.   Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7:
1127   Improvements in Performance and Usability. Mol Biol Evol 30:772–780.

1128   124.   Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
1129   phylogenies. Bioinformatics 30:1312–1313.

1130   125.   H. Wickham. 2016. ggplot2: Elegant Graphics for Data Analysis. . Springer-Verlag.

1131   126.   Sharp PM, Matassi G. 1994. Codon usage and genome evolution. Curr Opin Genet Dev 4:851–
1132   860.

1133   127.   Boser BE, Guyon IM, Vapnik VN. 1992. A training algorithm for optimal margin classifiers, p. 144–
1134   152. *In* Proceedings of the fifth annual workshop on Computational learning theory - COLT '92.

1135   128.   Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P,
1136   Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É.
1137   2011. Scikit-Learn: Machine Learning in Python. J Mach Learn Res 12:2825–2830.

1138   129.   Martin F, Kohler A, Murat C, Balestrini R, Coutinho PM, Jaillon O, Montanini B, Morin E, Noel B,
1139   Percudani R, Porcel B, Rubini A, Amicucci A, Amselem J, Anthouard V, Arcioni S, Artiguenave F,
1140   Aury J-M, Ballario P, Bolchi A, Brenna A, Brun A, Buée M, Cantarel B, Chevalier G, Couloux A, da
1141   Silva C, Denoeud F, Duplessis S, Ghignone S, Hilselberger B, Iotti M, Marçais B, Mello A, Miranda
1142   M, Pacioni G, Quesneville H, Riccioni C, Ruotolo R, Splivallo R, Stocchi V, Tisserant E, Viscomi AR,
1143   Zambonelli A, Zampieri E, Henrissat B, Lebrun M-H, Paolocci F, Bonfante P, Ottonello S, Wincker
1144   P. 2010. Périgord black truffle genome uncovers evolutionary origins and mechanisms of
1145   symbiosis. Nature 464.

1146   130.   Yang J, Wang L, Ji X, Feng Y, Li X, Zou C, Xu J, Ren Y, Mi Q, Wu J, Liu S, Liu Y, Huang X, Wang H, Niu
1147   X, Li J, Liang L, Luo Y, Ji K, Zhou W, Yu Z, Li G, Liu Y, Li L, Qiao M, Feng L, Zhang K-Q. 2011. Genomic
1148   and Proteomic Analyses of the Fungus *Arthrobotrys oligospora* Provide Insights into Nematode-
1149   Trap Formation. PLoS Pathog 7.

1150   131.   van den Berg MA, Albang R, Albermann K, Badger JH, Daran J-M, M Driessen AJ, Garcia-Estrada C,
1151   Fedorova ND, Harris DM, Heijne WHM, Joardar V, W Kiel JAK, Kovalchuk A, Martín JF, Nierman
1152   WC, Nijland JG, Pronk JT, Roubos JA, van der Klei IJ, van Peij NNME, Veenhuis M, von Döhren H,
1153   Wagner C, Wortman J, Bovenberg RAL. 2008. Genome sequencing and analysis of the
1154   filamentous fungus *Penicillium chrysogenum*. Nat Biotechnol 26.

132. Rouxel T, Grandaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, Dominguez V, Anthouard V, Bally P, Bourras S, Cozijnsen AJ, Ciuffetti LM, Degrave A, Dilmaghani A, Duret L, Fudal I, Goodwin SB, Gout L, Glaser N, Linglin J, Kema GHJ, Lapalu N, Lawrence CB, May K, Meyer M, Ollivier B, Poulain J, Schoch CL, Simon A, Spatafora JW, Stachowiak A, Turgeon BG, Tyler BM, Vincent D, Weissenbach J, Amselem J, Quesneville H, Oliver RP, Wincker P, Balesdent M-H, Howlett BJ. 2011. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. Nat Commun 2.

133. Yang J, Wang L, Ji X, Feng Y, Li X, Zou C, Xu J, Ren Y, Mi Q, Wu J, Liu S, Liu Y, Huang X, Wang H, Niu X, Li J, Liang L, Luo Y, Ji K, Zhou W, Yu Z, Li G, Liu Y, Li L, Qiao M, Feng L, Zhang K-Q. 2011. Genomic and Proteomic Analyses of the Fungus *Arthrobotrys oligospora* Provide Insights into Nematode-Trap Formation. PLoS Pathog 7.

134. Amselem J, Cuomo CA, van Kan JAL, Viaud M, Benito EP, Couloux A, Coutinho PM, de Vries RP, Dyer PS, Fillinger S, Fournier E, Gout L, Hahn M, Kohn L, Lapalu N, Plummer KM, Pradier J-M, Quévillon E, Sharon A, Simon A, ten Have A, Tudzynski B, Tudzynski P, Wincker P, Andrew M, Anthouard V, Beever RE, Beffa R, Benoit I, Bouzid O, Brault B, Chen Z, Choquer M, Collémare J, Cotton P, Danchin EG, da Silva C, Gautier A, Giraud C, Giraud T, Gonzalez C, Grossetete S, Güldener U, Henrissat B, Howlett BJ, Kodira C, Kretschmer M, Lappartient A, Leroch M, Levis C, Mauceli E, Neuvéglise C, Oeser B, Pearson M, Poulain J, Poussereau N, Quesneville H, Rascle C, Schumacher J, Ségurens B, Sexton A, Silva E, Sirven C, Soanes DM, Talbot NJ, Templeton M, Yandava C, Yarden O, Zeng Q, Rollins JA, Lebrun M-H, Dickman M. 2011. Genomic Analysis of the Necrotrophic Fungal Pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. PLoS Genet 7.

135. Cuomo CA, Guldener U, Xu J-R, Trail F, Turgeon BG, di Pietro A, Walton JD, Ma L-J, Baker SE, Rep M, Adam G, Antoniw J, Baldwin T, Calvo S, Chang Y-L, DeCaprio D, Gale LR, Gnerre S, Goswami RS, Hammond-Kosack K, Harris LJ, Hilburn K, Kennell JC, Kroken S, Magnuson JK, Mannhaupt G, Mauceli E, Mewes H-W, Mitterbauer R, Muehlbauer G, Munsterkotter M, Nelson D, O'Donnell K, Ouellet T, Qi W, Quesneville H, Roncero MIG, Seong K-Y, Tetko I v., Urban M, Waalwijk C, Ward TJ, Yao J, Birren BW, Kistler HC. 2007. The *Fusarium graminearum* Genome Reveals a Link Between Localized Polymorphism and Pathogen Specialization. Science (1979) 317.

136. Gazis R, Kuo A, Riley R, LaButti K, Lipzen A, Lin J, Amirebrahimi M, Hesse CN, Spatafora JW, Henrissat B, Hainaut M, Grigoriev I v., Hibbett DS. 2016. The genome of *Xylona heveae* provides a window into fungal endophytism. Fungal Biol 120.

137. WU X-M. 2007. The analysis method and progress in the study of codon bias. Hereditas 29:420.

138. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60.

139. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589.

140. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol 32:268–274.

1193   141.   Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the
1194   Ultrafast Bootstrap Approximation. Mol Biol Evol 35:518–522.

1195   142.   Stöver BC, Müller KF. 2010. TreeGraph 2: Combining and visualizing evidence from different
1196   phylogenetic analyses. BMC Bioinformatics 11:7.

1197   143.   Burki F, Roger AJ, Brown MW, Simpson AGB. 2020. The New Tree of Eukaryotes. Trends Ecol Evol
1198   35.

1199   144.   Keeling PJ, Burki F. 2019. Progress towards the Tree of Eukaryotes. Current Biology 29.

1200   145.   Novichkov PS, Omelchenko M v., Gelfand MS, Mironov AA, Wolf YI, Koonin E v. 2004. Genome-
1201   Wide Molecular Clock and Horizontal Gene Transfer in Bacterial Evolution. J Bacteriol 186:6575–
1202   6585.

1203

1204

Table 1. Sequencing and assembly statistics for 31 *Phytophthora* species grouped by clade.

Species	Clade	Genome			Annotations	
		Contigs	N_{50}	Assembly Length	Repeat Percent	Predicted Genes
P. cactorum	1	7888	15053	56443298	19.96	18027
P. idaei	1	7163	14461	53468943	16.19	18038
P. pini	2	2131	42987	38730000	7.31	14019
P. multivora	2	2844	46133	40059192	10.86	13682
P. pluvialis	3	4340	30816	53616150	16.04	16285
P. litchii	4	2543	34546	38200938	5.98	12391
P. palmivora	4	24815	6694	107798747	29.62	37283
P. megakarya	4	24073	7093	101609312	31.94	33614
P. agathidicida	5	3754	19544	37337699	5.76	12923
P. taxon totara	5	4425	30809	55576372	16.56	17619
P. parvispora	7	9906	6820	46825958	8.75	15642
P. pisi	7	7667	15253	58856683	16.80	18953
P. robiniae	7	14865	8754	69938814	25.58	23128
P. niederhauseri	7	26463	4805	90270009	20.96	29587
P. cajani	7	18255	5113	64854085	20.65	19840
P. vignae	7	10330	8363	56137732	17.45	18535
P. melonis	7	11353	15342	73416743	25.93	21276
P. pistaciae	7	10414	10302	63209321	18.56	19423
P. uliginosa	7	8955	10095	57072031	24.16	17226
P. europaea	7	8301	11551	58787065	23.33	17117
P. fragariae	7	8544	20362	76969737	30.81	20448
P. rubi	7	9434	17808	74863594	29.48	23476
P. pinifolia	6	22610	6021	74478861	33.00	23717
P. lateralis	8	28263	2396	50496828	23.39	19503
P. hibernalis	8	6587	21408	71256216	32.46	23578
P. foliorum	8	5320	15800	48973082	19.26	16083
P. brassicae	8	12447	12337	72849437	28.39	26010
P. syringae	8	6572	15987	57045526	21.71	18234
P. cryptogea	8	25944	4730	69446343	17.65	24936
P. boehmeriae	10	2866	41917	39747814	7.83	13325
P. kernoviae	10	13710	5225	42698878	4.15	14322

Table 2. Over-represented GO terms for a set of 44 candidate HGT transcripts found in 36 *Phytophthora* genomes.

GO	Term	# terms in full set[a]	# terms in HGT set[b]	Pr($X=k$)[c]
Biological process				
GO:0055114	obsolete oxidation-reduction process	1543	5	<0.001
GO:0034079	butanediol biosynthetic process	28	5	<0.001
GO:0045493	xylan catabolic process	151	1	<0.01
GO:0008152	Metabolism	3663	3	<0.01
GO:0042545	cell wall modification	493	1	<0.02
GO:0000272	polysaccharide catabolism	635	1	<0.02
GO:0002084	protein depalmitoylation	374	1	<0.02
GO:0006118	electron transport	1389	1	<0.05
GO:0045490	pectin catabolic process	865	1	<0.05
Molecular function				
GO:0016491	oxidoreductase activity	10881	11	<0.001
GO:0000721	(R,R)-butanediol dehydrogenase activity	28	5	<0.001
GO:0003939	L-iditol 2-dehydrogenase activity	11	1	<0.001
GO:0016831	carboxy-lyase activity	201	6	<0.001
GO:0008080	N-acetyltransferase activity	1082	3	<0.001
GO:0008270	zinc ion binding	26982	8	<0.001
GO:0005488	Binding	2200	2	<0.01
GO:0031176	endo-1,4-beta-xylanase activity	111	1	<0.01
GO:0008670	2,4-dienoyl-CoA reductase (NADPH) activity	15	1	<0.01
GO:0004022	alcohol dehydrogenase activity	102	1	<0.01
GO:0030599	pectinesterase activity	498	1	<0.02
GO:0008474	palmitoyl-(protein) hydrolase activity	457	1	<0.02
GO:0051213	dioxygenase activity	460	1	<0.02
GO:0045330	aspartyl esterase activity	423	1	<0.02
GO:0030570	pectate lyase activity	657	1	<0.02
GO:0015267	channel activity	742	1	<0.05
Cellular component				
GO:0005576	extracellular region	4624	2	<0.02

[a]Transcriptome of 35 *Phytophthora* genomes.
[b]44 HGT candidates.
[c]Probability (q-value) of obtaining the same number of transcripts (k) or more by chance as given by a hypergeometric probability distribution.

Table 3. Summary of 19 HGT candidates with strong phylogenetic support identified among *Phytophthora* spp.

	Putative function	Best hit on PHI-base (e-value)[a]	HGT identification[2]	# of *Phyto.* species	# of *Phyto.* clades	Closest clade[3]
Transfer from other groups to oomycetes						
HGT_1	Endonuclease	PHI:5754, endonuclease, *Fusarium graminearum* (3.0e-024)	SVM, IP, D, T	29	8	Fungi
HGT_2	Xylulose reductase	PHI:2256, xylitol dehydrogenase, *Parastagonospora nodorum* (1.0e-128)	SVM, IP, D, T	12	5	Fungi
HGT_15	Zinc-binding dehydrogenase, Polyketide synthase, enoylreductase domain	PHI:8321, gluconate 5-dehydrogenase, *Salmonella enterica* (1.0e-016)	SVM, IP, D, T	21	8	Bacteria
HGT_4	Aquaporin	PHI:7047, water channel protein aquaporin, *Cryptococcus neoformans* (4.0e-012)	SVM, IP, D, T	30	8	Bacteria
HGT_5	NPP1	-	SVM, IP, D, T	4	1	Bacteria
HGT_6	Phenol acid carboxylase	-	SVM, MC, D, T	35	9	Fungi
HGT_7	Peptidase S9	-	SVM, IP, D, T	8	1	Fungi
HGT_8	UDP-N-acetylglucosamine-peptide N-acetylglucosaminyl-transferase	PHI:4921, flagellin glycosyltransferase, *Burkholderia cenocepacia* (1.0e-021)	SVM, IP, D, T	2	2	Bacteria
HGT_9	Alternative oxidase	-	SVM, IP, D	10	4	Fungi
HGT_10	Dioxygenase	-	SVM, MC, D, T	33	9	Fungi
HGT_11	Thioesterase	PHI:4988, sfp-type 4'-phosphopantetheinyl transferase, *Bipolaris maydis* (3.0e-005)	SVM, IP, D, T	35	8	Amoebozoa
HGT_12	Endo-1,4-beta-xylanase GH10	PHI:7912, endo-beta-1,4-xylanase *Phytophthora parasitica* (1.0e-137)	SVM, IP, D, T	33	9	Fungi
HGT_13	4-coumarate CoA ligase	PHI:10606, long-chain-fatty-acid–Co Aligase, *Pseudomonas aeruginosa* (1.0e-032)	SVM, IP, D, T	4	4	Fungi/ Bacteria
HGT_14	2,4-dienoyl-CoA reductase	PHI:8134, 3-Oxoacyl-[acyl-carrier-protein] reductase, *Salmonella enterica* (1.0e-014)	SVM, IP, D	28	7	Bacteria/ Archaea
HGT_20	Ribosomal-protein-alanine acetyltransferase	-	SVM, IP, D	29	9	Bacteria

Transfer from oomycetes to other groups

HGT_3	Quinone oxidoreductase	-	SVM, IP, D, T	23	9	Fungi
HGT_16	Putative tannase	PHI:10222, feruloyl esterase, *Valsa mali* (4.0e-030)	SVM, IP, D, T	33	8	Bacteria
HGT_17	Putative pectinesterase CE8	PHI:278, pectin methylesterase, *Botrytis cinerea* (4.0e-077)	SVM, IP, D, T	34	9	Fungi
HGT_18	ATP-binding Cassette (ABC)	-	SVM, IP, D,	34	9	Fungi

[a] Best BLASTp hit on the Pathogen Host Interaction database PHI-base (Urban et al. 2019); [2]SVM, support vector machine; IP, incongruent phylogeny; MC, missing clades; D, distance; T, Alternate topology test (Shimodaira-Hasegawa-test) significant; [3]as reported in BLASTp analysis; [3]HB, only hemibiotrophic Peronosporales i.e. *Phytophthora*, *Phytopythium* and *Nothophytophthora*.
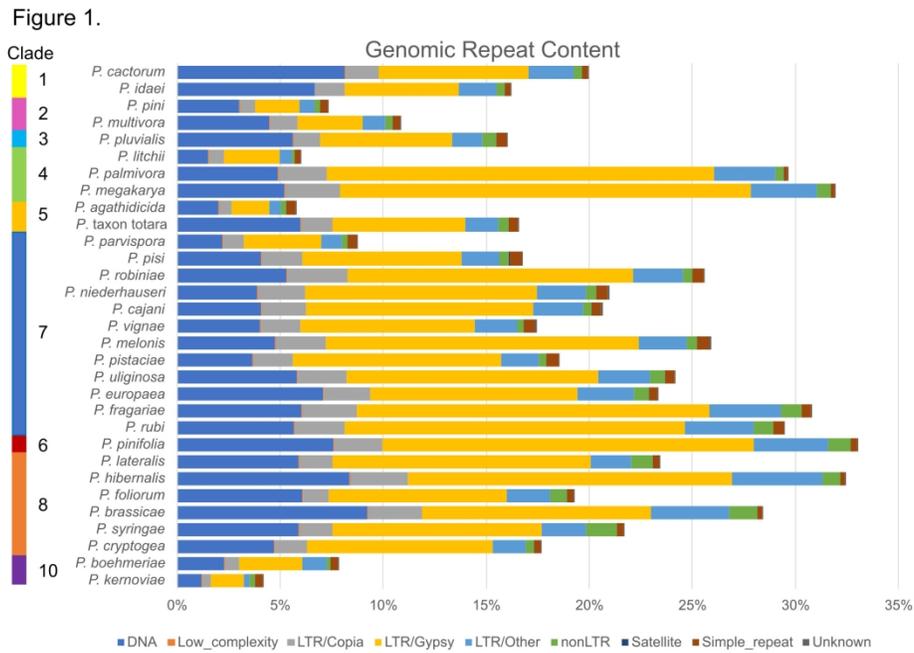
Figure 1. Analysis of 31 *Phytophthora* species shows an abundance of repetitive elements. Species are shown in phylogenetic clade order; clade designations are shown on the left. Repeat content is displayed as percentage of the total genome content. Repeat classifications are shown as colored bar segments.

705x481mm (90 x 90 DPI)
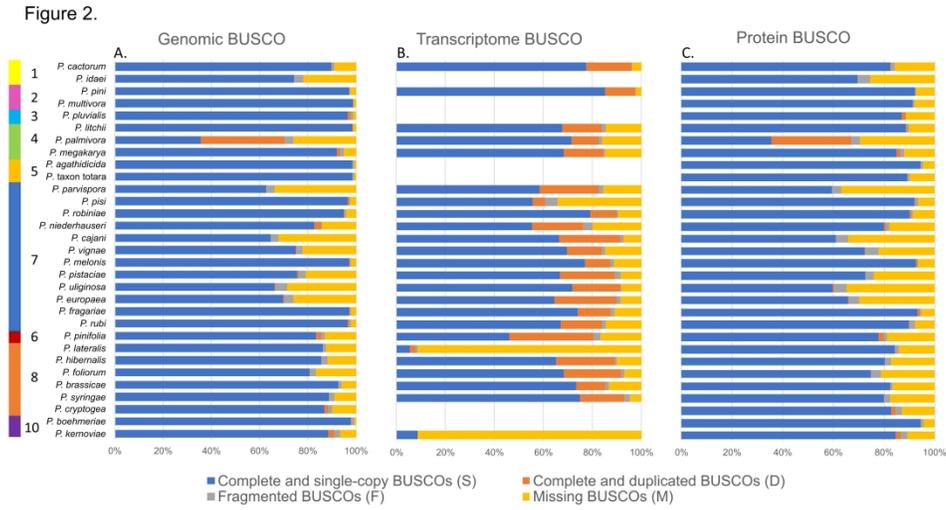
Figure 2.

Figure 2. BUSCO analysis demonstrates completeness of the 31 *Phytophthora* species in this study. A) genomic assembly, B) transcriptome assembly, and C) predicted proteins. Species are shown in phylogenetic clade order and clade designations are shown on the left. For each BUSCO analysis, results from searching 234 single copy orthologs in the Alveolata_Stramenopiles dataset are shown.

987x518mm (90 x 90 DPI)
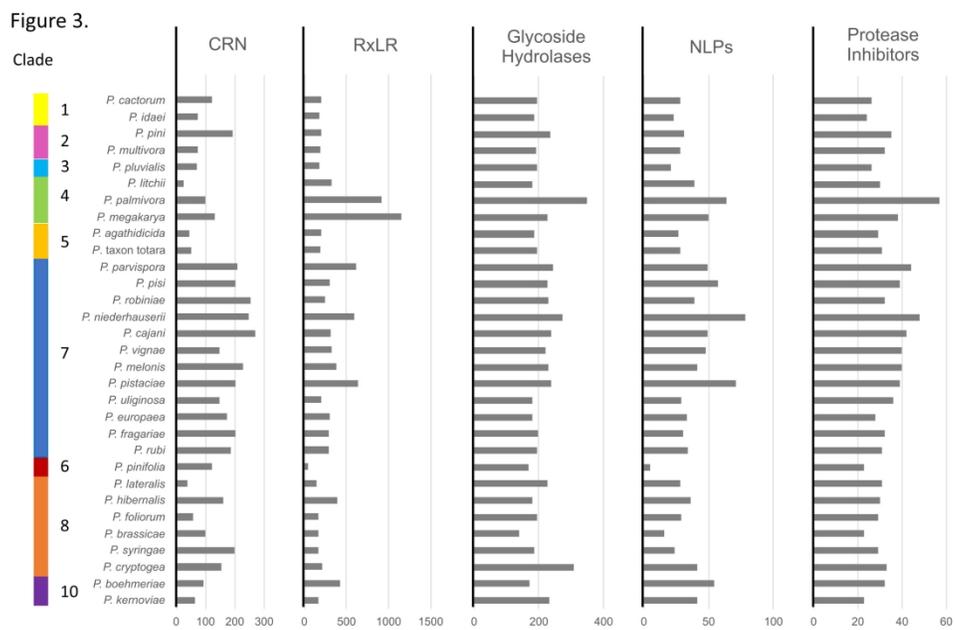
Figure 3. The number of predicted effectors varied across the *Phytophthora* genomes. Bar chart representing amounts of effector genes found in 31 *Phytophthora* species for crinkler (CRN), RxLR, glycoside hydrolases, necrosis inducing proteins (NLPs), and protease inhibitors. Species are shown in phylogenetic clade order, clade designations are shown on the left.

705x481mm (90 x 90 DPI)

Figure 4.
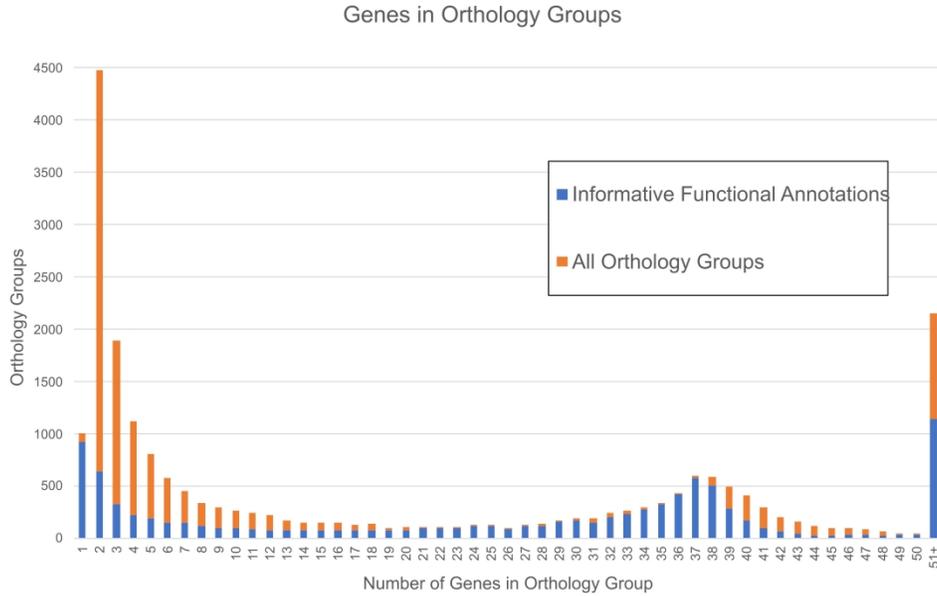
Figure 4. Distribution of genes in *Phytophthora* orthology shows a bimodal frequency distribution, highlighting genes that are conserved in only a few genomes and orthology groups that have one gene per species. Amounts of genes assigned to orthology cluster with OrthoMCL (118) are shown. The 31 sequenced *Phytophthora* and six additional previously sequenced *Phytophthora* genomes (*P. capsici*, *P. cinnanomi*, *P. infestans*, *P. parasitica*, *P. ramorum*, *P. sojae*) are included. 'All Orthology Groups' (orange) show all genes assigned into orthology groups. 'Informative Functional Annotations' show genes assigned into orthology groups that have useful functional definitions and exclude genes labeled as 'uncharacterized', 'hypothetical', or similar.

846x623mm (90 x 90 DPI)

Figure 5.
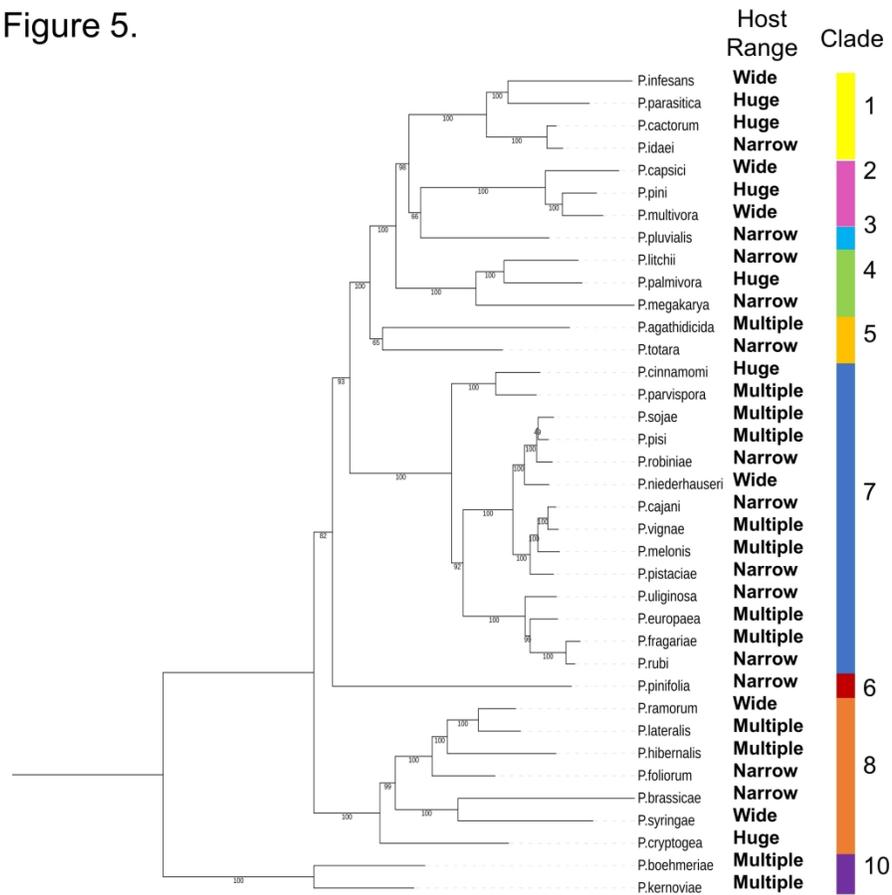
Figure 5. Phylogenetic relationships of the 31 sequenced *Phytophthora* spp. and six additional previously sequenced *Phytophthora* spp. Sixty-one single-copy core orthologous proteins shared across 37 species were used to create a RAxML (122) phylogenetic tree using each gene as an independent partition with its own substitution model and bootstrapped 1,000 times. Ranges of infected hosts are shown next to the phylogenetic tree species defined as 'Narrow', host species confined to 1 plant host genus; 'Multiple', host species confined within 2 to 9 host genera; 'Wide', host species spanning 16-55 host genera; and 'Huge', host species spanning 107-327 host genera. Clade assignments are shown on the right.

812x762mm (96 x 96 DPI)
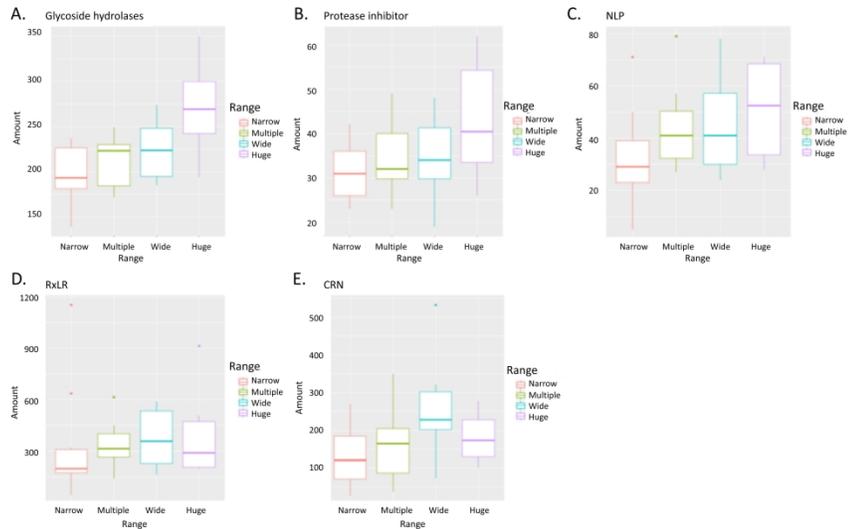
Figure 6. Generally, *Phytophthora* with larger host range showed a greater predicted number of effector genes. Box plots showing amounts of effectors found per *Phytophthora* species, categorized into 'Narrow' (1 plant genera), 'Multiple' (2-9 plant genus), 'Wide' (16-55 host genera), and 'Huge' (107-327 host genera) host ranges. Effectors glycoside hydrolases (A), protease inhibitors (B), NLPs (C), RxLRs (D), and CRNs (E) are shown. Near identical paralogs were removed; proteins with greater than 95% amino acid identity over the full sequence length were reduced to one representative effector sequence. Statistically significant classifications were seen between the Narrow-Wide and Narrow-Huge comparisons in glycoside hydrolases (A).

1016x571mm (90 x 90 DPI)

Figure 7.
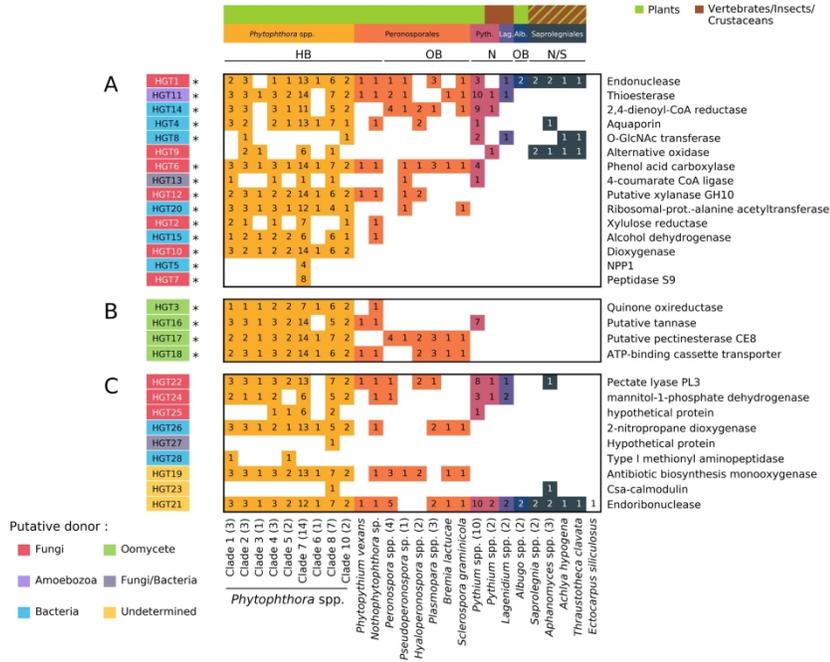
Figure 7. Conservation level of the 44 *Phytophthora* HGT candidates in oomycetes. A and B, set of 19 HGT candidates for which a maximum likelihood phylogeny was reconstructed and alternate tree topologies were tested with the Shimodaira-Hasegawa test (an asterisk indicate significant topological difference [P < 0.05] between the constrained alternate topology and the observed topology; A, HGT to oomycetes; B, HGT for the opposite relationships i.e. transfers from oomycetes to fungi or bacteria; C, HGT candidates with no maximum likelihood phylogeny support. For each HGT candidate the number of sequence homologs identified among 37 *Phytophthora* and 30 oomycetes transcriptomes (identified by reciprocal DIAMOND BLASTp, minimum E-value of 1e-03, sequence subject coverage of 50% and sequence query coverage of 50%; a dash indicates the absence of a one-to-one ortholog) is reported. The filamentous brown alga *Ectocarpus siliculosus* (Ectocarpales, Ectocarpaceae) was used as an outgroup. Putative functions are indicated on the right. Top rows: Pyth., Pythiales; Lag., Lagenidiales; Alb., Albuginales; HB, hemibiotrophic lifestyle; OB, obligate biotrophic; S, saprotrophic; N, necrotrophic. Species names, *Phytophthora* clade names and group of species names are indicated on the bottom; numbers between brackets indicate the number of species considered in a group.

711x558mm (96 x 96 DPI)
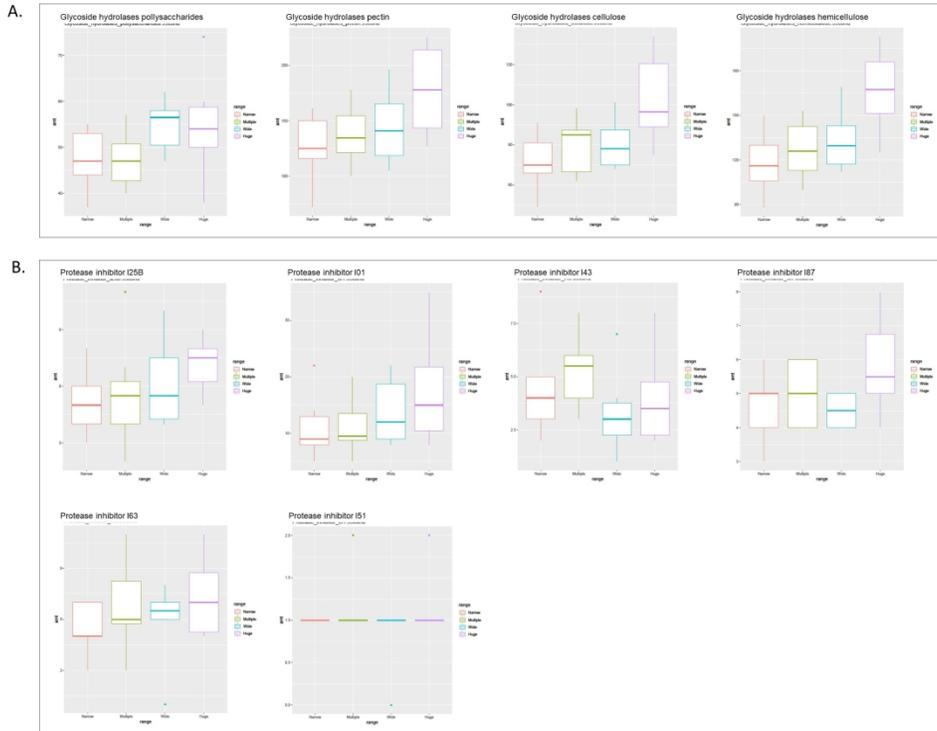
Figure S1.

Figure S1. Box plots showing amounts of glycoside hydrolases, sorted by substrate, in *Phytophthora* species and categorized into host ranges 'Narrow' (1 plant genera), 'Multiple' (2-9 plant genus), 'Wide' (16-55 host genera), and 'Huge' (107-327 host genera). Glycoside hydrolases pollysaccharides, pectins, cellulose, and hemicellulose are shown (A). For protease inhibitors, I01, I25B, I43, I51, I63, and I87 are shown (B). Near identical paralogs were removed; proteins with greater than 95% amino acid identity over the full sequence length were reduced to one representative effector sequence.

762x609mm (96 x 96 DPI)
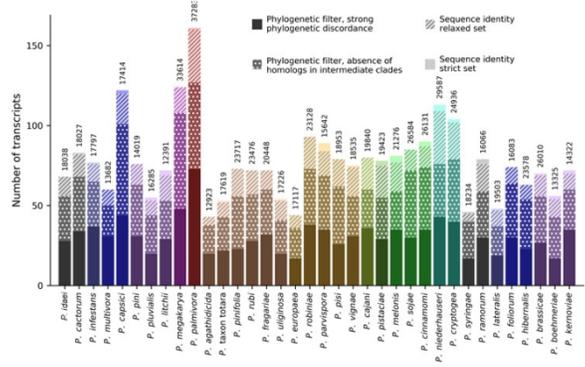
Figure S2. Number of HGT candidates (transcripts) retained for 37 *Phytophthora* genomes after false positive filtering of candidates initially identified with a Support Vector Machine classifier with two "phylogenetic" and "identity" tests. The total number of transcripts predicted for each Phytophthora genome is indicated above bars.

855x481mm (38 x 38 DPI)

Figure S3. Horizontal gene transfer for 19 candidate genes in *Phytophthora* spp. A, Maximum likelihood phylogeny showing relationships between candidate HGT gene(s) (indicated by three asterisks) and closest relative sequences found within seven clades (Heterokonta, Alveolata, Viridiplantae, Fungi, Amoebozoa, Archaea and Bacteria); B. expected phylogeny according to Burki et al. (2020) and Keeling et al. (2019). C and D, sequence identity identity between the candidate HGT sequence in *Phytophthora* and its closest homolog sequences in the putative "donor" species; the distribution represent the expected nucleotide identity values for the two same species calculated from a random sample of 1000 sequences; D, the arrow indicates the sequence identity between the candidate HGT sequence and the closest species in the non-donor clades; the distribution represents the expected nucleotide identity values for the two same from a random sample of 1000 sequences. Bottom table, Shimodaira-Hasegawa test (SH-test) between the observed ML tree and alternative tree topologies.

Burki, Fabien et al. 2020. The New Tree of Eukaryotes. Trends in Ecology & Evolution, 35: 43 – 55.

Keeling, Patrick J. et al. 2019 Progress towards the Tree of Eukaryotes. Current Biology, 29: R808 - R817.

A - Observed phylogeny
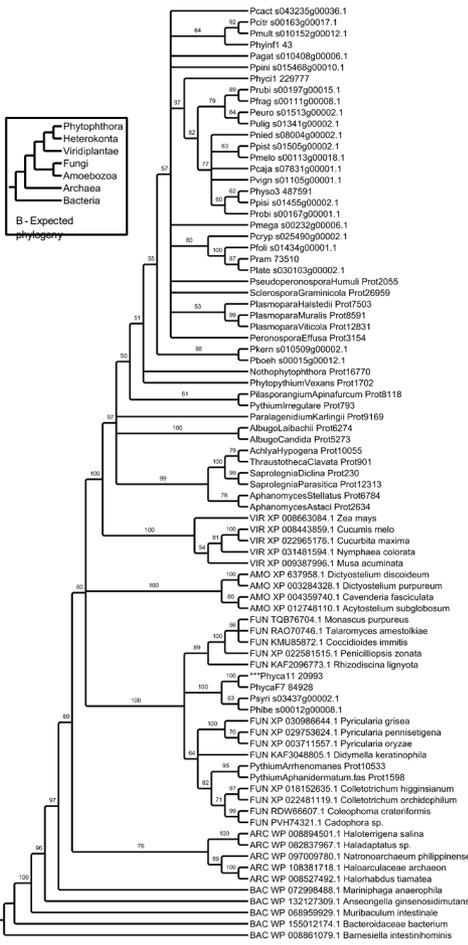
C - Colletotrichum higginsianum (Fungi)
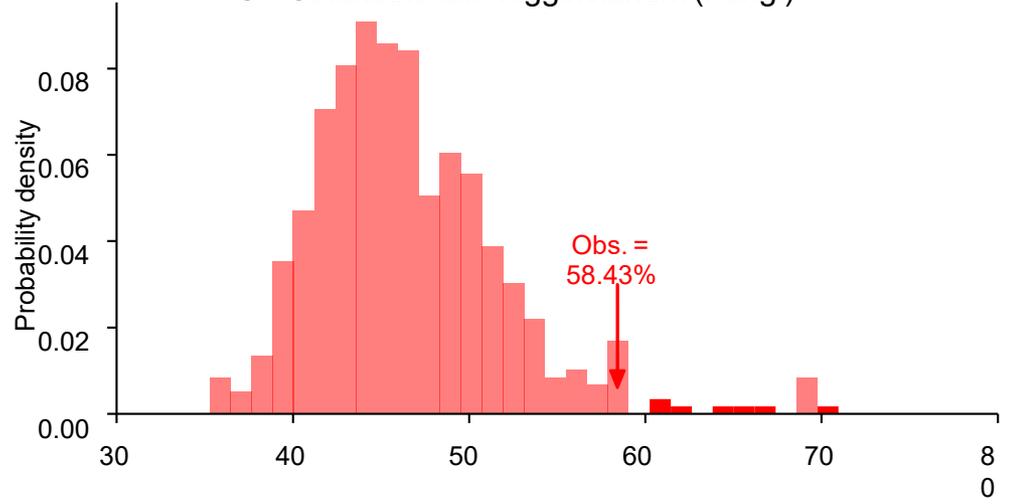
D - Cucumis melo (Viridiplantae)
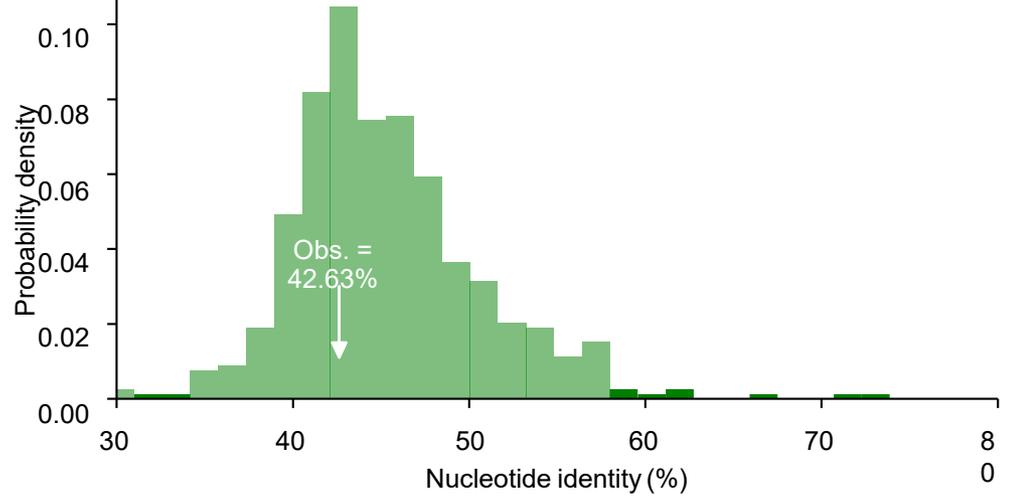
Topology	-Ln	D(-Ln)	P value
Observed topology	242536389999	0	ns
Expected topology	262060664848	564774488	< 0.0001
Fungi branched with the inherited H-center taxa (H-center donor)	242408081616	667722	< 0.0001

Fig. S3_1 - HGT_1 (Endonuclease)

A - Observed phylogeny

B - Expected phylogeny

Phytophthora
Heterokonta
Viridiplantae
Fungi
Bacteria

100 Phytophthora nicotianae
59 Phytophthora parasitica
95 Phytophthora cactorum
74 Phytophthora infestans
51 Phytophthora megakarya
73 ***Pvign s05795g00002.1
95 Phytophthora sojae
Phytophthora fragariae
Phytophthora kernoviae
Nothophytophthora sp. Chile5
94 84 FUN EOD43657.1 Neofusicoccum parvum
67 FUN KAF4301997.1 Botryosphaeria dothidea
88 FUN EKG18521.1 Macrophomina phaseolina
94 FUN Lasiodiplodia theobromae
FUN OMP82594.1 Diplodia seriata
VIR XP 023877611.1 Quercus suber
99 VIR KAF1854692.1 Lupinus albus
61 66 FUN XP 024689256.1 Aspergillus campestris
FUN XP 016239383.1 Exophiala spinifera
50 100 FUN CEG03653.1 Fusarium acuminatum
66 FUN KAF4461154.1 Fusarium albosuccineum
94 52 FUN OLN85435.1 Colletotrichum chlorophyti
FUN XP 003714060.1 Pyricularia oryzae
100 FUN EPE09563.1 Ophiostoma piceae
FUN KFY59188.1 Pseudogymnoascus sp
100 94 FUN KAF3913882.1 Drechslerella brochopaga
FUN KAF3192874.1 Arthrobotrys oligospora
99 BAC RJS88955.1 Candidatus Bathyarchaeota archaeon
BAC WP 123619211.1 Halorubrum sp. CSM 61
BAC WP 156587500.1 Halorubrum sp. CBA1125

C - Diplodia seriata (Fungi)

Obs. = 59.95%
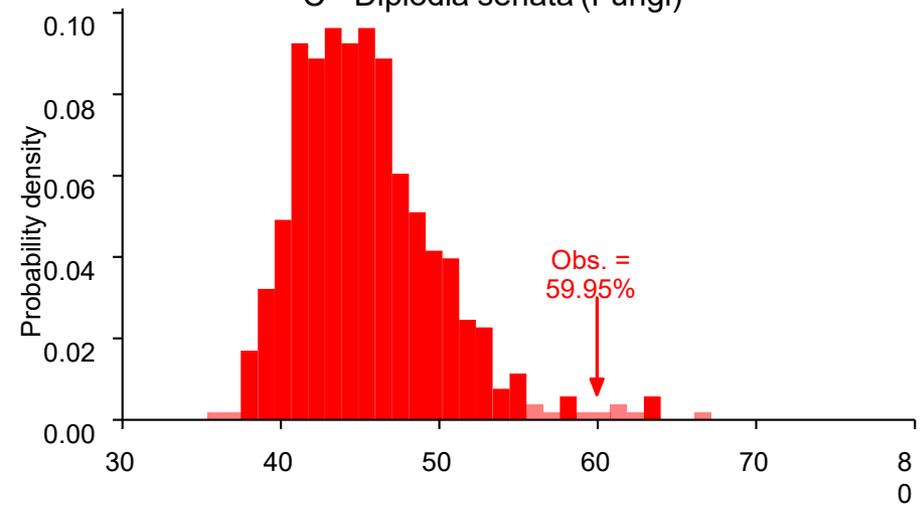
D - Halorubrum sp. (Bacteria)

Obs. = 37.90%

Nucleotide identity (%)

Topology	-Ln	D(-Ln)	P-value
Observed topology	8800.338866	0	ns
Expected topology	8899.04444	100.557	<0.0001
Fungi branched with Heterokonta (Heterokonta donor)	8877.88800	77.49944	<0.0001
Heterokonta branched with Fungi (Fungi as donor)	8811.881133	11.442277	ns

Fig. S3_2 - HGT_2 (Xylulose reductase)
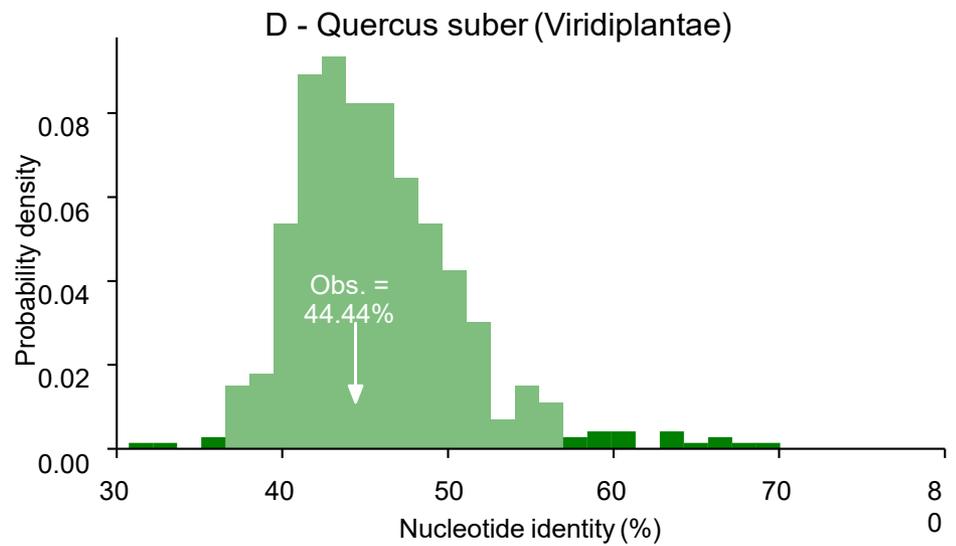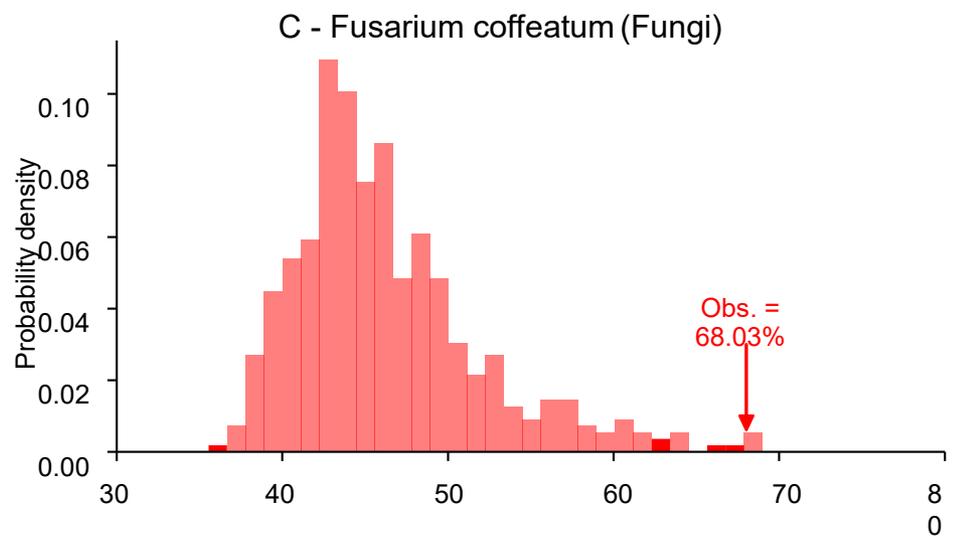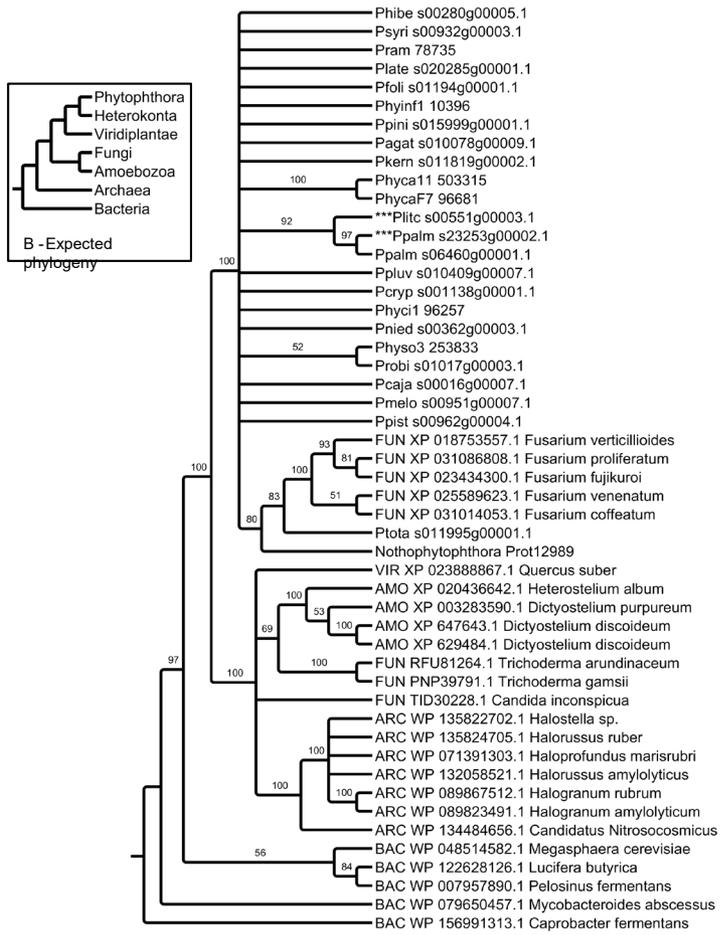
A - Observed phylogeny

B - Expected phylogeny

C - Fusarium coffeatum (Fungi)

Obs. = 68.03%

D - Quercus suber (Viridiplantae)

Obs. = 44.44%

Nucleotide identity (%)

Topology	-Ln	D(-Ln)	P-value
Observed topology	131262657474	0	ns
Expected topology	131438478722	2222114477	<00001
alternated-best-known-archaea-and-without-the-Fungi (Fungi domain donor)	131363767929	140221177	<00001

Fig. S3_3 - HGT_3 (Quinone oxidoreductase)

A - Observed phylogeny
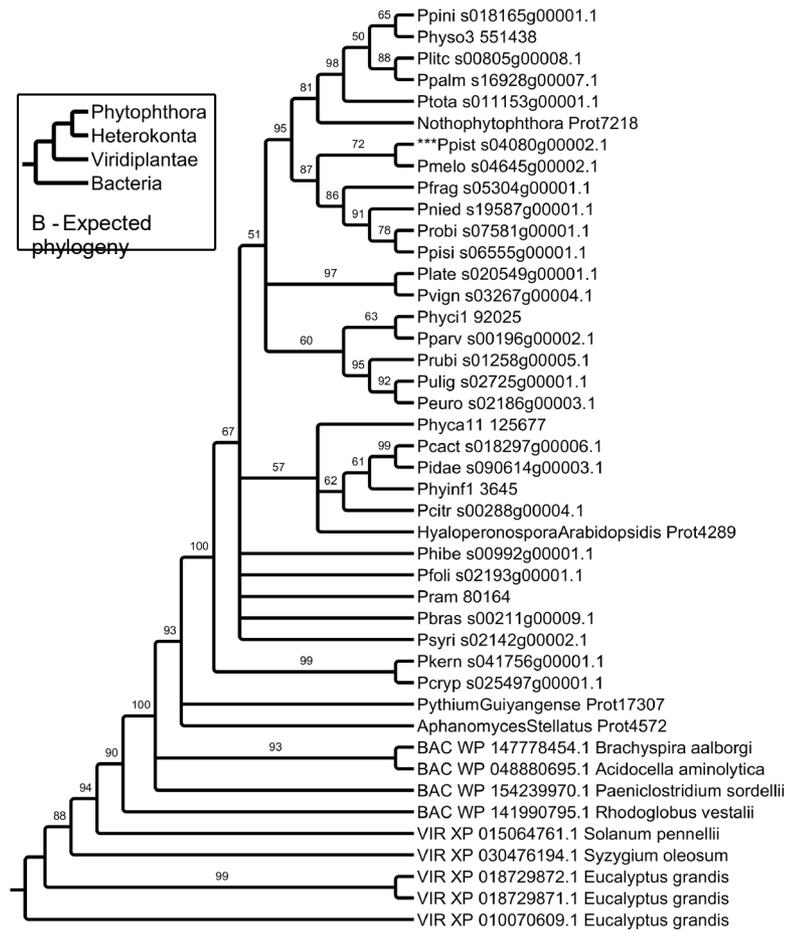
C - Acidocella aminolytica (Bacteria)

Obs. = 57.18%
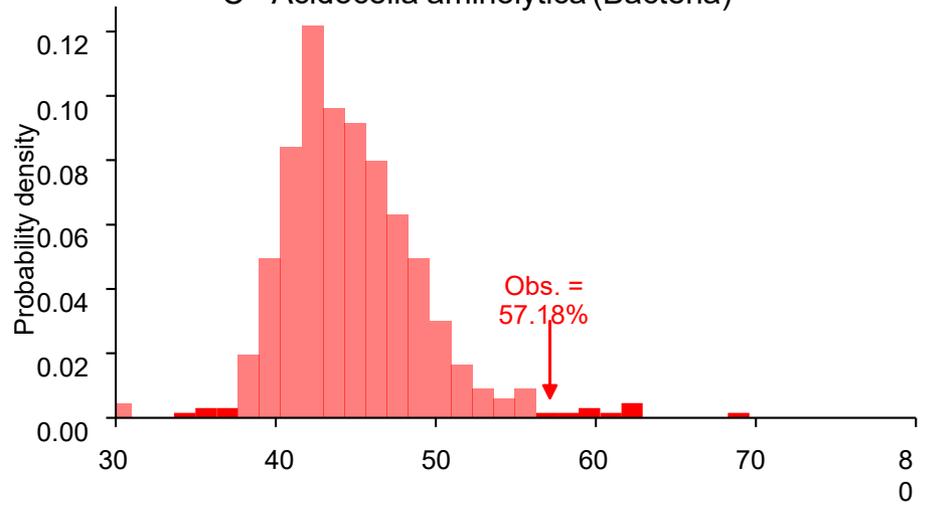
D - Solanum pennellii (Viridiplantae)

Obs. = 41.32%

Nucleotide identity (%)
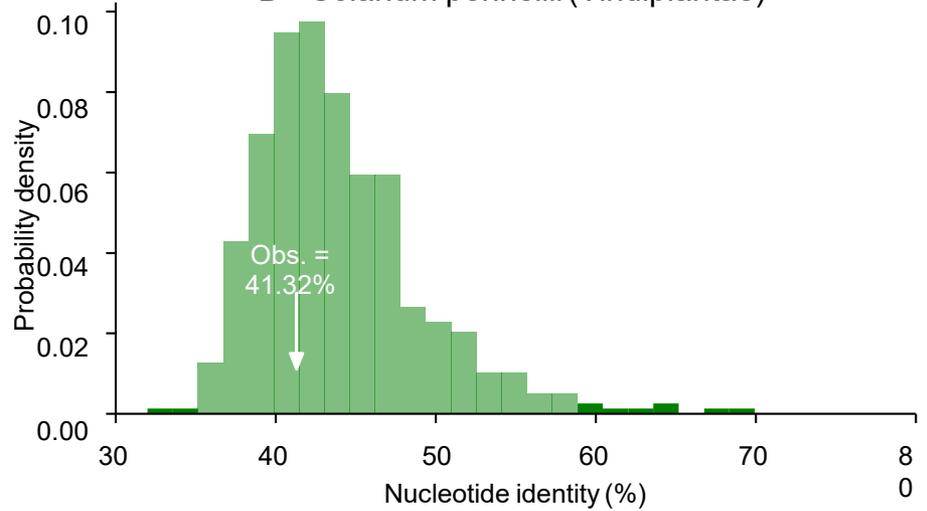
Topology	-Ln	D(-Ln)	P-value
Observed topology	7712.36955	0	ns
Expected topology	7720.0003	9.66077	<0.0001
Bacteria nested within Heterokonta (Host donor)	7719.2774	6.68736	<0.0001
Heterokonta nested within Bacteria (Bacteria donor)	7712.4229	-0.33	ns

Fig. S3_4 - HGT_4 (Aquaporin)

A - Observed phylogeny

B - Expected phylogeny

C - Solirhodobacter olei (Bacteria)

Obs. = 60.95%

D - Capsella rubella (Viridiplantae)

Obs. = 43.95%

Nucleotide identity (%)

Topology	-Ln	D(-Ln)	P-value
Observed topology	10160.8668222	0	ns
Expected topology	10704.2423688	-74.1	<0.0001
Basal eukaryote clade within Heterokonta (Host) donor)	10704.3435688	7455.3866	<0.0001

Fig. S3_5 - HGT_5 (NPP1)

A - Observed phylogeny

Pram 75459
***Pcryp s008723g00001.1
Phycl1 230395
Pcryp s018607g00001.1
Pcitr s00088g00002.1
Physo3 343772
Nothophytophthora Prot24359
****Pkern s010389g00002.1
Pcact s031427g00001.1
Pkern s013657g00002.1
***Pboeh s00136g00008.1
Pmega s00225g00005.1
Ppalm s09543g00002.1
Plitc s00395g00002.1
Ppini s006884g00002.1
Pparv s03010g00001.1
Pulig s01212g00005.1
Peuro s01416g00002.1
Prubi s03370g00002.1
Pfrag s00269g00004.1
Ptota s010166g00006.1
Pram 85165
Plate s016533g00001.1
Pfoli s00065g00008.1
Pbras s00320g00007.1
Psyri s00257g00005.1
Phibe s00484g00001.1
Pvign s05117g00001.1
***Pmelo s01835g00001.1
Ppisi s00683g00002.1
Pnied s09131g00002.1
Pcaja s04371g00002.1
***Ppist s01364g00003.1
Physo3 286364
Probi s00785g00004.1
Phyinf1 11216
Pcryp s021085g00002.1
Phycl1 99184
Pizae s117391g00004.1
Bremial.actucae Prot8981
PhytopythiumVexans Prot1330
PlasmoparaHalstedii Prot13134
STR XP 024585089.1 Plasmopara halstedii
PlasmoparaMuralis Prot8822
PlasmoparaViticola Prot12988
Pmult s010554g00002.1
PhycaF7 98661
PhycaT1 505902
Pcitr s00613g00005.1
Nothophytophthora Prot4494
Pseudoperonospora Humuli Prot928
SclerosporaGraminicola Prot32442
PilasporangiumApinafurcum Prot3109
GlobisporangiumSplendens Prot4035
GlobisporangiumUltimumBR650 Prot6283
GlobisporangiumUltimumBR144 Prot14417
HyaloperonosporaArabidopsidis Prot6212
HyaloperonosporaArabidopsidisEmoy2 Prot11754
FUN XP 016261676.1 Exophiala oligosperma
FUN CCG84751 Taphrina deformans
FUN OCF40846 Kwoniella heveanensis
FUN XP 006668755.1 Cordyceps militaris
FUN XP 008603183.1 Beauveria bassiana
FUN XP 007682001.1 Baudoinia panamericana
FUN XP 003042417.1 Nectria
FUN APP94188 Brettanomyces
FUN XP 023437628.1 Fusarium fujikuroi
FUN XP 031086043.1 Fusarium proliferatum
FUN XP 022480310.1 Colletotrichum orchidophilum
FUN XP 018151969.1 Colletotrichum higginsianum
FUN XP 008094542.1 Colletotrichum graminicola
FUN XP 009656399.1 Verticillium dahliae
FUN XP 028496087.1 Verticillium nonalfalfae
FUN KAF9290506 Mortierella alpina
FUN SPO32604 Ustilago trichophora
FUN XP 007303961.1 Stereum hirsutum
FUN XP 007303961 Stereum hirsutum
FUN QQ079822 Schizophyllum commune
FUN OWZ72201 Cryptococcus neoformans var grubii
FUN RLV85099 Meyerozyma sp.
FUN PRQ77049 Rhodotorula toruloides
FUN SCV67940 Microbotryum intermedium
FUN SGY37645 Microbotryum silenes-dioicae
FUN KAF1797046 Mucor lusitanicus
FUN THH04906 Bondarzewia mesenterica
BAC WP 148223465.1 Segniliparus rotundus
BAC WP 123651874.1 Kluyvera ascorbata
BAC WP 081918546.1 Kluyvera ascorbata
BAC WP 142005843.1 Barrientosiimonas humi

B - Expected phylogeny

Phytophthora
Heterokonta
Viridiplantae
Fungi
Bacteria

C - Exophiala oligosperma (Fungi)

Probability density

Obs. = 75.18%

D - Segniliparus rotundus (Bacteria)

Probability density

Obs. = 51.47%

Nucleotide identity (%)

Topology	-Ln	D(-Ln)	P value
Observed topology	8622.13	0	ns
Bacteria and Fungi monophyletic	8866.96	440.83	<0.001
Fungi branch within Heterokonta (Heterokonta donor)	8839.15	133.02	<0.001
Heterokonta branch within Fungi (Fungi donor)	8626.28	155.52	ns

Fig. S3_6 - HGT_6 (Phenol acid carboxylase)

A - Observed phylogeny

B - Expected phylogeny

C - Diaporthe ampelina (Fungi)

Obs. = 81.40%

D - Quercus suber (Plant)

Obs. = 44.76%

Nucleotide identity (%)

Topology	-Ln	D(-Ln)	P-value
Observed topology	8618.2839	0	ns
BasidiomC-H-method-monophyletic	8626.1239	7.8400	<0.0001
Fungi branch-within-H-within-H-ortholog(H-med)(donor)	8623.8686	5.6986	<0.0001
Hetero-H-bentaoloxocan-taoda-within-H-within-Fwithgi(Fungia)Fd.comgoal(donor)	8622.5683	4.2933	<0.0001

Fig. S3_7 - HGT_7 (Peptidase S9)

A - Observed phylogeny

B - Expected phylogeny

Phytophthora
Heterokonta
Viridiplantae
Bacteria

100 — ***Phyca11 122662
100 — Pkern s011086g00003.1
58 — LagenidiumGiganteum Prot2839
99 — PythiumIrregulare Prot4587
GlobisporangiumSplendens Prot11853
99 — AchlyaHypogena Prot12409
ThraustothecaClavata Prot9823
100 — BAC ALF56607.1 Nostoc piscinale
100 — BAC WP 036386961.1 Microcystis aeruginosa
BAC WP 123230484.1 Microcystis aeruginosa
BAC WP 174718070.1 Microcystis viridis
54 — BAC WP 190721123.1 Microcystis flos aquae
BAC WP 004268994.1 Microcystis aeruginosa
57 — BAC NDC31873.1 Bacteroidetes bacterium
81 — BAC OQX03496.1 Desulfobacteraceae bacterium
BAC MSO93995.1 Rhodospirillales bacterium
67 — BAC HCB15041.1 Gammaproteobacteria bacterium
99 — BAC MBF0304806.1 Alphaproteobacteria bacterium
100 — BAC WP 083386651.1 Phaeospirillum fulvum
BAC WP 021131550.1 Phaeospirillum fulvum
100 — VIR XP 024526762.1 Selaginella moellendorffii
VIR XP 024543070.1 Selaginella moellendorffii
99 — VIR XP 010674215.1 Beta vulgaris
VIR XP 021855358.1 Spinacia oleracea
VIR XP 028946009.1 Malus domestica
100 — VIR XP 027189986.1 Cicer arietinum
VIR XP 024637516.1 Medicago truncatula
VIR XP 028770706.1 Prosopis alba

94
90
87
100

C - Microcystis aeruginosa (Bacteria)
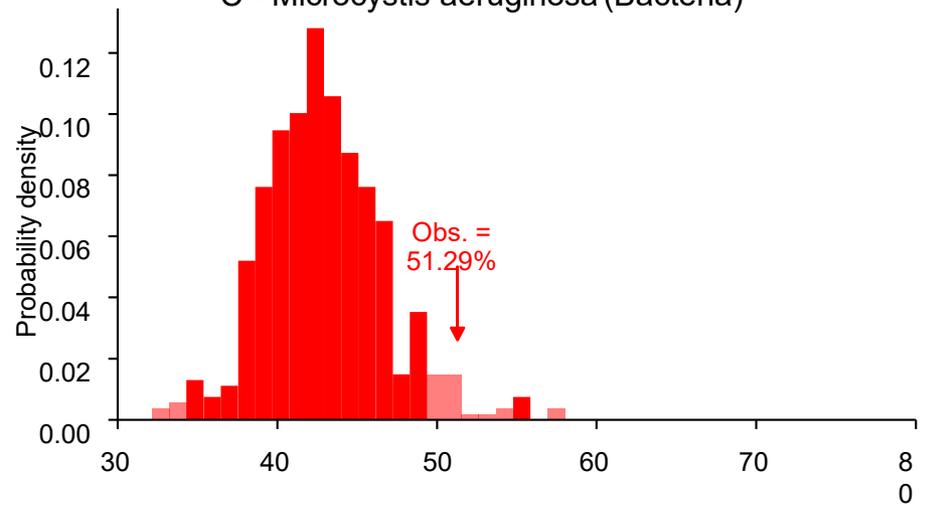
Obs. = 51.29%

D - Prosopis alba (Plant)
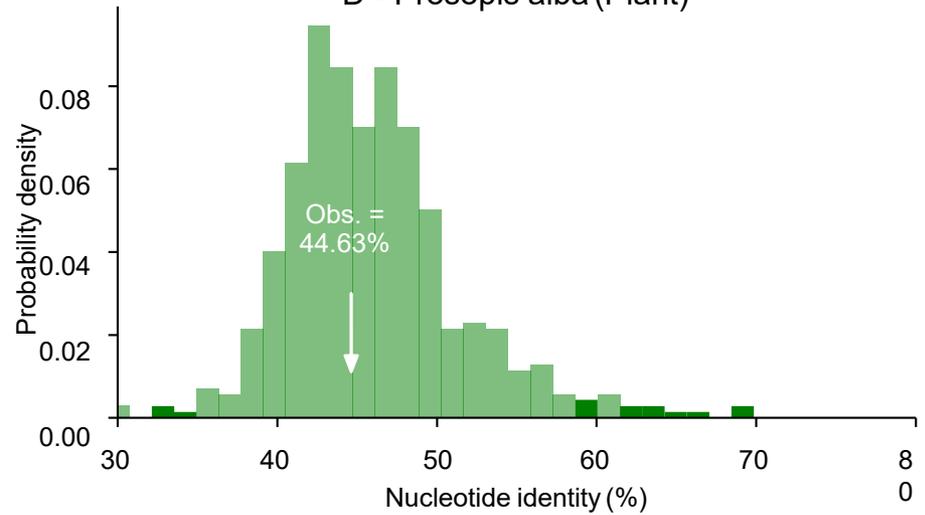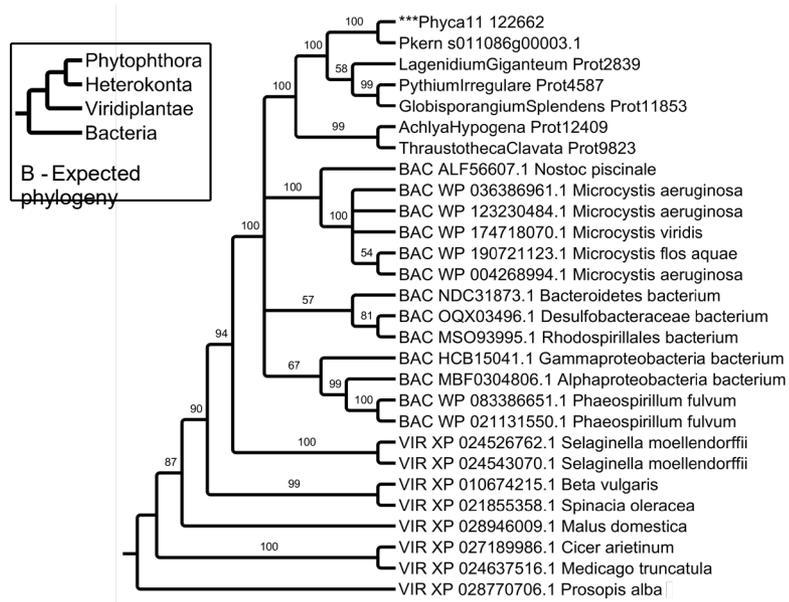
Obs. = 44.63%

Nucleotide identity (%)

Fig. S3_8 - HGT_8 (UDP-N-acetylglucosamine-peptide N-acetylglucosaminyltransferase)
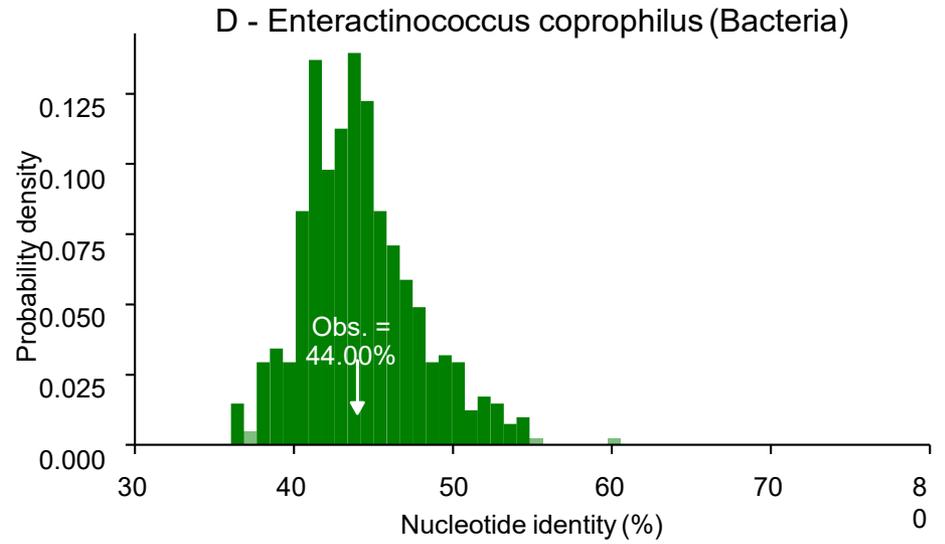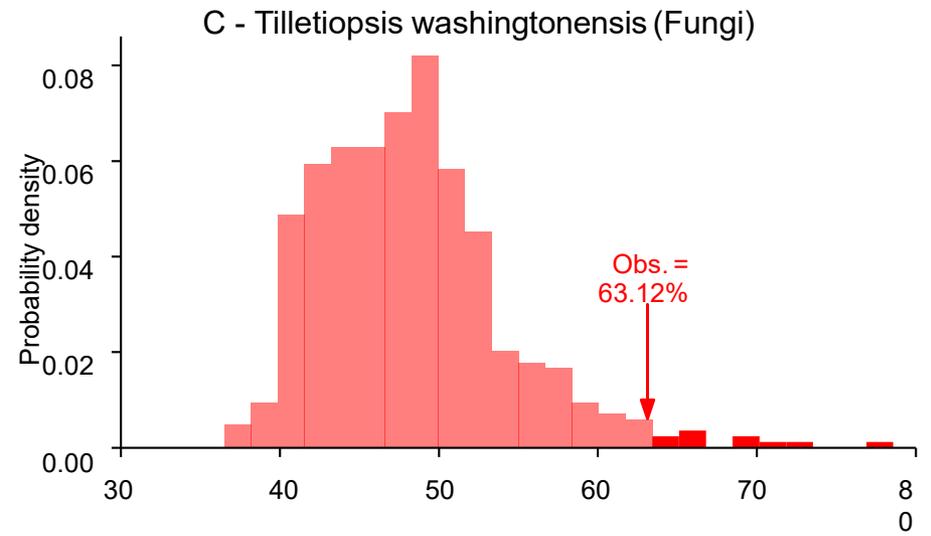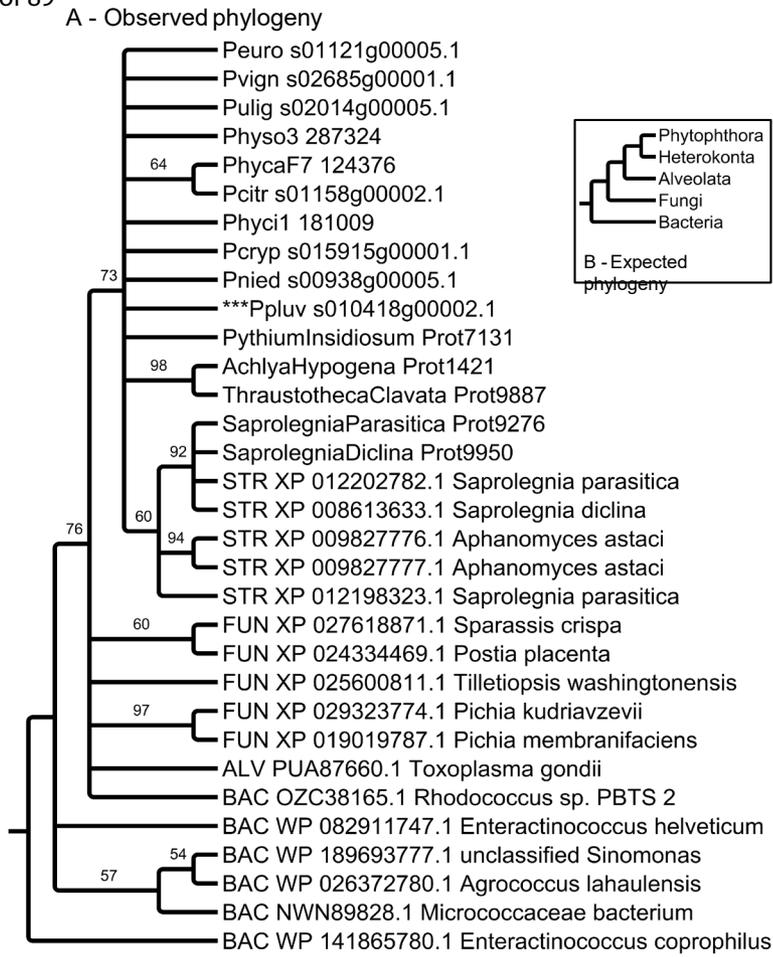
A - Observed phylogeny

B - Expected phylogeny

C - Tilletiopsis washingtonensis (Fungi)

Obs. = 63.12%

D - Enteractinococcus coprophilus (Bacteria)

Obs. = 44.00%

Nucleotide identity (%)

Topology	-Ln	D(-Ln)	P-value
Observed topology	14955.06533	0	ns
Expected topology	14956.44077	1.33244	ns
Fungi branched within Heterokonta (Heterokonta donor)	14996.22	1.55399	ns
Heterokonta branched within Fungi (Fungi donor)	14956.60086	-5.23	ns

Fig. S3_9 - HGT_9 (Alternative oxidase)

A - Observed phylogeny

B - Expected phylogeny

Phytophthora
Heterokonta
Viridiplantae
Fungi
Bacteria

C - Lachnellula hyalina (Fungi)

Probability density

Obs. = 70.03%

Nucleotide identity (%)

D - Auxenochlorella protot11thecoides (Viridiplantae)

Probability density

Obs. = 57.0%

Nucleotide identity (%)

Phylogeny tip labels:

99 Pboeh s00375g00012.1
Pkern s009935g00001.1
Psyri s00321g00008.1
Pcryp s015928g00007.1
57 Pfoli s01232g00004.1
Phibe s01183g00006.1
85 Plate s004123g00001.1
Pram 51406
85 Pparv s04150g00002.1
Phyci1 344036
Ppist s03658g00001.1
Physo3 492989
Probi s07640g00001.1
Ppisi s03365g00008.1
91 ***Pmelo s02638g00003.1
75 Pvign s06826g00001.1
Pcaja s01961g00003.1
Pnied s02909g00003.1
94 Pfrag s00579g00005.1
80 Prubi s00256g00002.1
71 Peuro s02676g00002.1
Pulig s02774g00002.1
Ptota s010236g00023.1
99 Plitc s00178g00001.1
Ppalm s10085g00002.1
Pagat s011532g00004.1
Ppluv s010174g00014.1
95 Pmult s010056g00024.1
Pcitr s01618g00002.1
Phyinf1 12601
100 Pcact s009904g00013.1
Pidae s014492g00009.1
Ppini s001349g00002.1
FUN XP 024719843.1 Amorphotheca resinae
FUN XP 018073374.1 Phialocephala scopiformis
FUN XP 024728127.1 Hyaloscypha bicolor
FUN XP 031007530.1 Lachnellula hyalina
99 FUN XP 001560992.1 Botrytis cinerea
FUN XP 001594914.1 Sclerotinia sclerotiorum
83 FUN KAE9985675.1 Venturia
FUN XP 001910208.1 Podospora
59 FUN CRL27420.1 Penicillium
FUN XP 017991383.1 Malassezia
FUN KAF2090385.1 Saccharata
FUN KAF9787125.1 Thelephora
73 FUN XP 007860659.1 Gloeophyllum
65 FUN RDX50572.1 Polyporus
FUN OCB89003.1 Sanghuangporus
FUN KAF8481614.1 Russula
99 FUN CCO30102.1 Rhizoctonia
FUN KAF8600903.1 Ceratobasidium
VIR XP 011400335.1 Auxenochlorella protothecoides
FUN XP 024512514.1 Cryptococcus
FUN CED83872.1 Phaffia
VIR XP 023877672.1 Quercus suber
BAC WP 155412368.1 Rhodospirillaceae bacterium
BAC WP 123689369.1 Stella humosa

Topology	-Ln	D(-Ln)	P-value
Observed topology	1218278773636	0	ns
Expected topology	131235255343	337777.99	< 0.0001
Fungi branched with the H. wall... H. centered (H-centered donor)	1219293375	56663.88	< 0.0001
...branched with the Fungi right (Fungi donor)	12182818	-4.47	ns

Fig. S3_10 - HGT_10 (Dioxygenase)

A - Observed phylogeny

B - Expected phylogeny

C - Heterostelium album (Amobea)

D - Solanum tuberosum (Viridiplantae)

Topology	-Ln	D(-Ln)	P-value
Observed topology	22086897788	0	ns
Expected topology	20205959591	5661144	<0.0001
Amoeba clade within Heterokont (Host donor)	20205959591	6622011	<0.0001
Heterokont clade within Amoeboa (Amoeboa donor)	20208976589	0000001	ns

Fig. S3_11 - HGT_11 (Thioesterase)

A - Observed phylogeny

B - Expected phylogeny

C - Baudoinia panamericana (Fungi)

Obs. = 55.27%

D -

Helianthus annuus (Viridiplantae)

Obs. = 45.51%

Nucleotide identity (%)

Topology	-Ln	D(-Ln)	P-value
Observed topology	2421.323636	0	ns
Expected topology	2434.20747	14.9771100	<0.0001
Fungi branched within Heterokonta (Heterokont donor)	2460.70784	44.844468	<0.0001
Heterokont branched within Fungi (Fungal donor)	2434.61799	22.368422	<0.0001

Fig. S3_12 - HGT_12 (Putative xylanase)

A - Observed phylogeny

64 — BAC WP 149187161.1 Aquicoccus porphyridii
69 — ARC WP 147438470.1 Halobellus sp.
63 — BAC WP 083815113.1 Ktedonobacter racemifer
84 — PIP45956.1 Candidatus Colwellbacteria
86 — BAC WP 155408441.1 Halieaceae bacterium
— BAC WP 146917501.1 Baekduia soli
— FUN XP 007264020.1 Fomitiporia mediterranea
— FUN XP 018752158.1 Fusarium verticillioides
— FUN XP 016600429.1 Penicillium expansum
75 — ***Pnied s06997g00001.1
68 — PythiumBrassicum Prot10484
96 — Pkern s041900g00001.1
— PeronosporaTabacina Prot9611
78 — Pidae s016597g00006.1
— Ppalm s17415g00002.1
96 — 100 — WP 201305092.1 Streptomyces sp.
— WP 189301224.1 Streptomyces albospinus
— WP 190142985.1 Streptomyces glebosus
— PYP29435.1 Gemmatimonadetes bacterium
80 — MBA2449534.1 Chloroflexi bacterium
— HDU09229.1 Blastocatellia bacterium
— FUN XP 008593382.1 Beauveria bassiana
100 — VIR XP 024958155.1 Citrus sinensis
52 — VIR XP 007027424.2 Theobroma cacao
— VIR XP 023914150.1 Quercus suber
— VIR XP 018723048.1 Eucalyptus grandis
— VIR XP 020264712.1 Asparagus officinalis

B - Expected phylogeny

Phytophthora
Heterokonta
Viridiplantae
Fungi
Archaea
Bacteria

C - Halieaceae bacterium (Bacteria)

Obs. = 52.42%

D - Citrus sinensis (Viridiplantae)

Obs. = 42.71%

Nucleotide identity (%)

Topology	-Ln	D(-Ln)	P-value
Observed topology	7722663344	0	ns
Expected topology	74981668866	9900552	<00001
Only Basidio with Heterokonta	74981225688	8882244	<00001
Only Fungi with Heterokonta	7788222211	9868697711	<00001
Fungi Branched with Heterokonta (Host donor)	7777556900	4499556	<00001
Heterokonta branched with the Fungi Branch (Fungi Broad donor)	7722991199	-2.84	ns

Fig. S3_13 - HGT_13 (4-coumarate CoA ligase)

A - Observed phylogeny
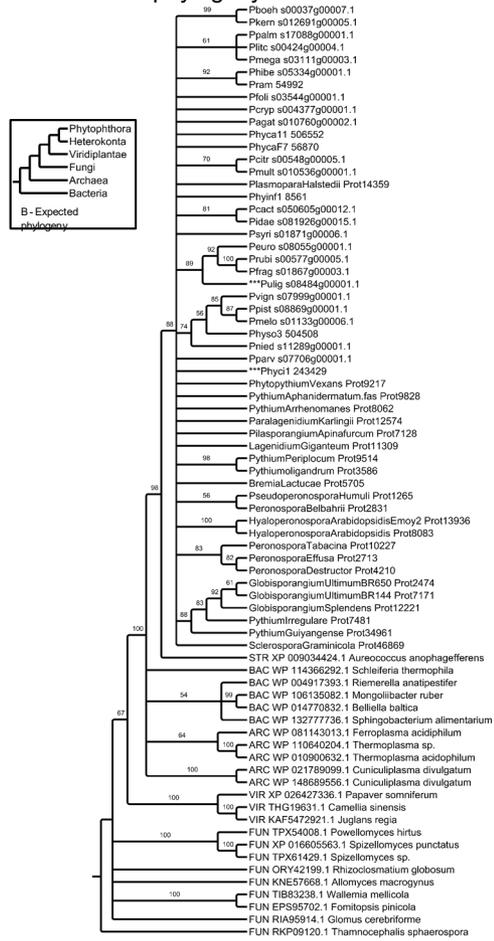
C - Schleiferia thermophila (Bacteria)

D - Juglans regia (Viridiplantae)

Topology	-Ln	D(-Ln)	P-value
Observed topology	141416169464	0	ns
Expected topology	142620607272	144440088	<0.00011
Bacteria/Archaea/Porch reached with fine H with rich H detail (Host donor)	144656651414	3848655	<0.00011
Host H host broken host fixed with H the Eu within for P Bac host reach for Bac donor And Bac donor Porch donor)	141143131272	1466633	ns

Fig. S3_14 - HGT_14 (2,4-dienoyl-CoA reductase)

A - Observed phylogeny

```
              66
            61  ┌ Pfoli s03988g00001.1
           61 ├ Phibe s02040g00004.1
              ├ Pram 72095
              └ Pram 72356
              ┌ Pcryp s021621g00001.1
           57 ├ ***Pmult s010394g00010.1
              ├ Pcitr s01229g00003.1
           87 ├ Pkern s014008g00003.1
              ├ Pboeh s00176g00013.1
          100 ├ PhycaF7 6033
              └ Phyca11 8134
           84 ┌ ***Ppist s02908g00001.1
           89 ├ ***Pmelo s00034g00022.1
           93 ├ Pcaja s04411g00002.1
              └ Pvign s02642g00003.1
           69 ┌ Physo3 352545
              ├ Pnied s08018g00002.1
           74 ├ Probi s01225g00004.1
              └ Ppisi s05071g00002.1
              ***Pparv s02158g00001.1
           96 ***Phyci1 238691
              ┌ Peuro s00811g00004.1
           79 98 ├ Pfrag s00293g00002.1
           99 ├ Prubi s00429g00008.1
              └ Pulig s02906g00003.1
           95 ┌ Nothophytophthora Prot15865
              └ Nothophytophthora Prot5154
           87 ┌ ***Pcact s014044g00003.1
              ├ Pidae s014035g00006.1
           90 Phyinf1 2885
           69 ┌ Ppalm s16930g00002.1
              ├ Pmega s00583g00003.1
           51 ├ Ppluv s010288g00007.1
              └ Ptota s010626g00003.1
          100 BAC WP 117352640.1 Blastomonas fulva
              BAC WP 110264883.1 Sinimarinibacterium flocculans
           81 BAC WP 107941482.1 Stenotrophobium rhamnosiphilum
              BAC WP 093285231.1 Solimonas aquatica
              BAC WP 129520424.1 Agromyces albus
          100 VIR XP 005848280.1 Chlorella variabilis
          100 VIR XP 024995007.1 Cynara cardunculus
          100 VIR XP 019164122.1 Ipomoea nil
           59 95 VIR XP 021737309.1 Chenopodium quinoa
           58 100 VIR XP 027104459.1 Coffea arabica
           79 VIR XP 027163453.1 Coffea eugenioides
              VIR XP 026407045.1 Papaver somniferum
          100 FUN XP 013259305.1 Exophiala aquamarina
           92 FUN XP 007732837.1 Capronia epimyces
           78 FUN XP 022506799.1 Fonsecaea monophora
              FUN XP 016227306.1 Exophiala mesophila
          100 ARC WP 076430257.1 Haladaptatus litoreus
              ARC WP 103993201.1 Halobellus limi
           78 ARC WP 141465529.1 Natronolimnobius sp.
              ARC WP 076608248.1 Natronorubrum thiooxidans
              ARC WP 082837900.1 Haladaptatus sp.
              ARC WP 008318032.1 Haloferax mucosum
```

B - Expected phylogeny

```
┌ Phytophthora
├ Heterokonta
├ Viridiplantae
├ Fungi
├ Archaea
└ Bacteria
```

C - Sinimarinibacterium flocculans (Bacteria)

Obs. = 68.66%

D - Cynara cardunculus (Viridiplantae)

Obs. = 42.39%

Nucleotide identity (%)

Topology	-Ln	D(-Ln)	P-value
Observed topology	12035.04	0	ns
Expected topology	12179.12	444.08	<0.0001
Bacterial clade with the whole dataset (Horizontal transfer) (host donor)	12039.52	377.48	<0.0001
Host clade with the whole dataset (Vertical inheritance) (Bacterial donor)	12035.66	-1.55	ns

Fig. S3_15 - HGT_15 (Alcohol dehydrogenase)

A - Observed phylogeny

B - Expected phylogeny

Phytophthora
Heterokonta
Viridiplantae
Fungi
Bacteria
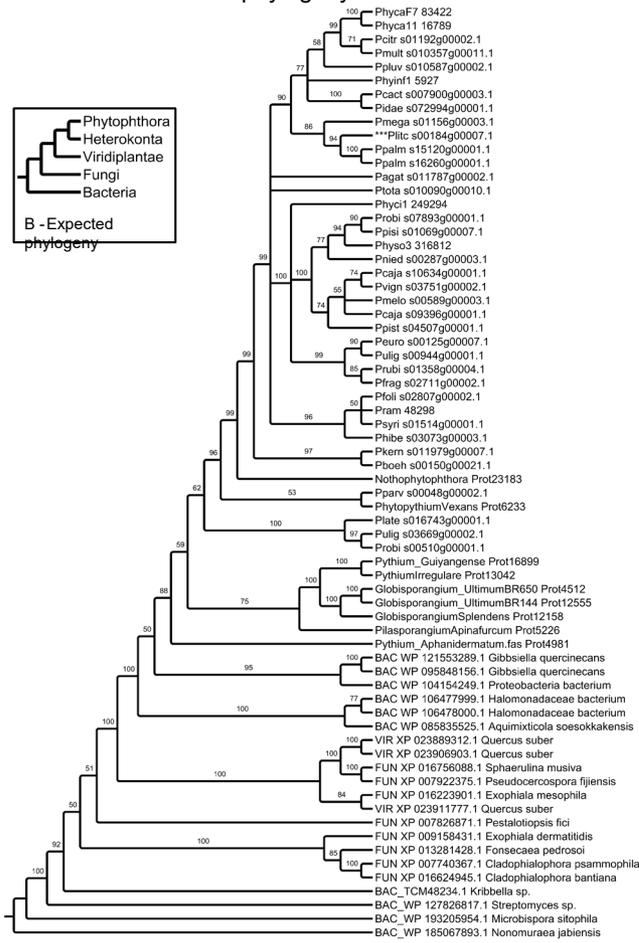
C - Halomonadaceae bacterium (Bacteria)
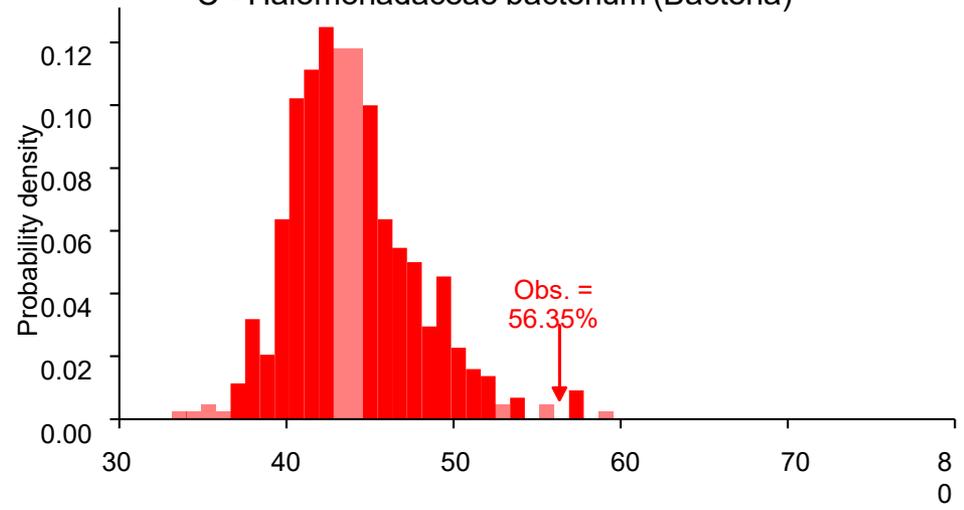
Probability density

Obs. = 56.35%

Nucleotide identity (%)

D - Quercus suber (Viridiplantae)

Probability density

Obs. = 42.73%

Nucleotide identity (%)
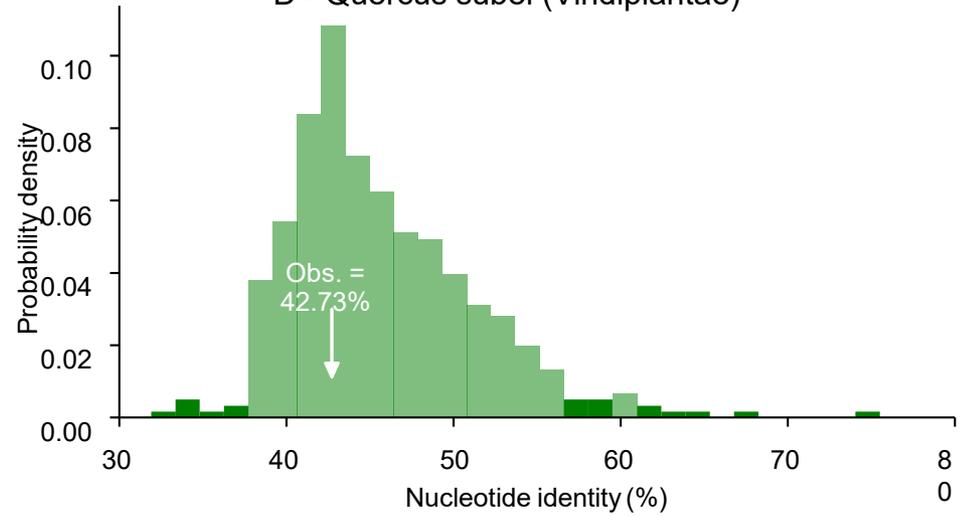
Topology	-Ln	D(-Ln)	P-value
Observed phylogeny	331991511	0	ns
Expected phylogeny	320300882323	142230055	<00001
Bacteria clade within Heterokonta (Host donor)	320210447474	99957	<00001
Heterokonta clade within Bacteria (Bacteria donor)	323524647576	334495558	<00001

Fig. S3_16 - HGT_16 (Putative tannase)

A - Observed phylogeny

C - Aureobasidium pullulans (Fungi)

D - Papaver somniferum (Viridiplantae)

Topology	-Ln	D(-Ln)	P-value
Observed topology	2920.50459099	0	ns
Expected topology	2929.99656353	23.55444	<0.001
Heterokonta and embedded within Fungi (Fungi donor)	2920.80787494	33.84400	<0.001
Fungi branch embedded within Heterokonta (Heterokonta donor)	2990.77558	22.17799	<0.065

Fig. S3_17 - HGT_17 (Putative pectinesterase)

A - Observed phylogeny

B - Expected phylogeny

C - Fusarium coffeatum (Fungi)

Obs. = 65.98%

Probability density
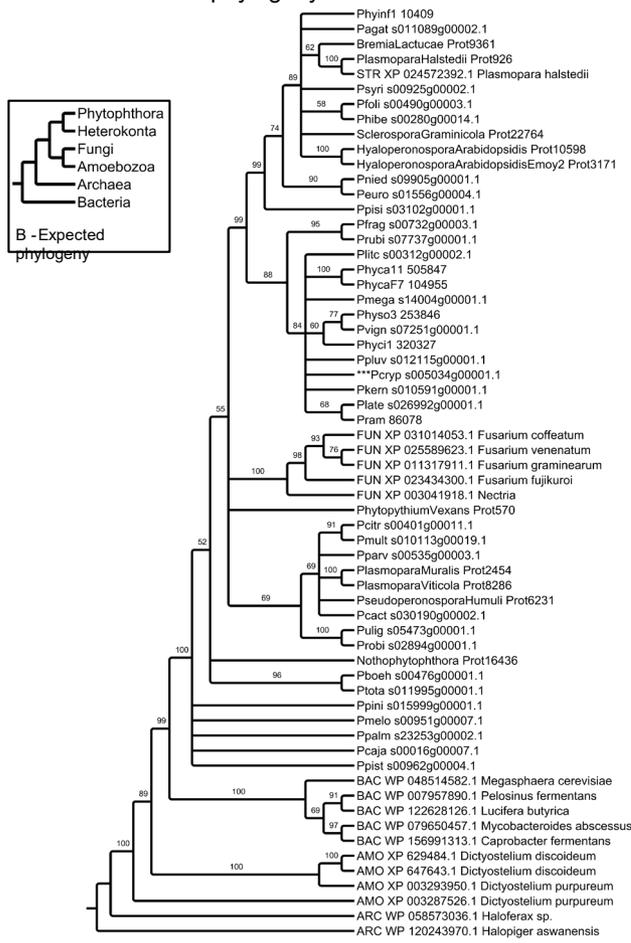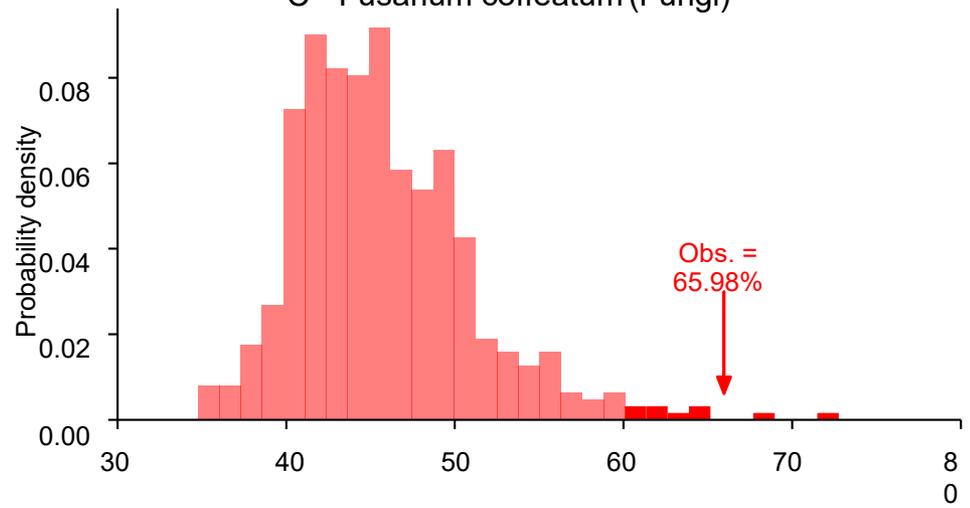
D - Dictyostelium purpureum (Viridiplantae)
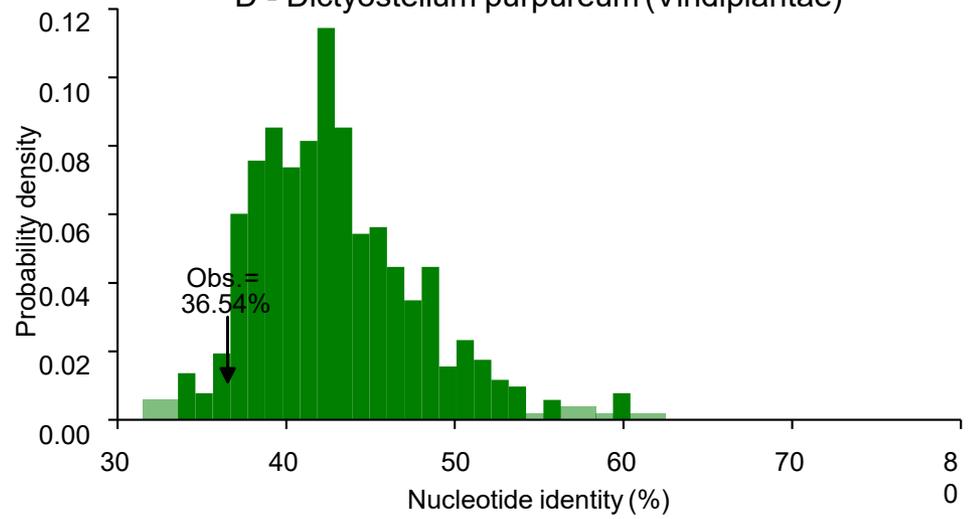
Obs. = 36.54%

Probability density

Nucleotide identity (%)

Topology	-Ln	D(-Ln)	P-value
Observed topology	1618.1627	0	ns
Expected topology	1619.5922	1.4429.86	<0.001
Heterokonta obtained within Fungi (Fungi donor)	1711.3202	33.1575	<0.001
Fungi obtained within Heterokonta (Heterokont donor)	1618.2645	1.0018	ns

Fig. S3_18 - HGT_18 (ATP-binding Cassette [ABC])

A - Observed phylogeny

C - Friedmanniella flava (Bacteria)

Obs. = 56.84%
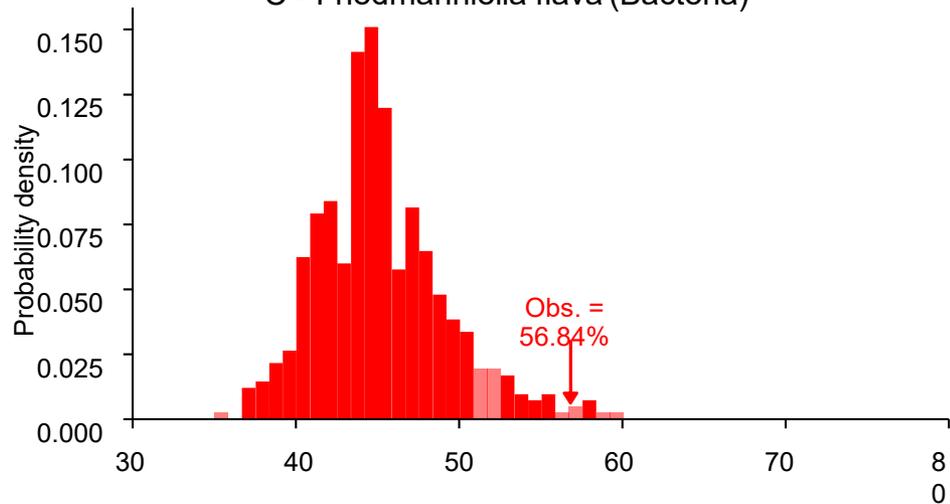
D - Ostreococcus lucimarinus (Viridiplantae)
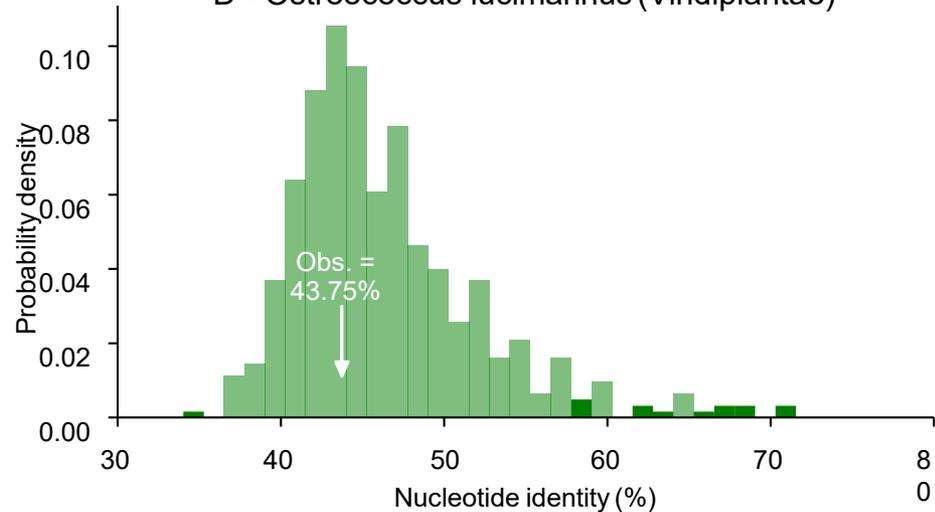
Obs. = 43.75%

Nucleotide identity (%)

Topology	-Ln	D(-Ln)	P value
Observed phylogeny	44633.7552	0	ns
Expected phylogeny	44877.2684	3.55.322	<0.005
Heterokonta clade with the B within the B (Bacteria as donor)	44843.3633	-6.31	ns
Bacteria clade within the H within host (Host as donor)	44877.6107	3.55.66	<0.001

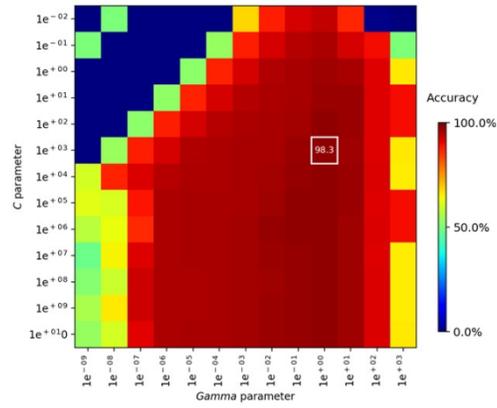Fig. S3_19 - HGT_20 (Ribosomal-protein-alanine acetyltransferase)

Figure S4.

Figure S4. Identification of the radial basis function (rbf) kernel parameters (*C* and *gamma*) resulting in the highest accuracy in classifying Phytophthora, bacteria and fungal transcripts with a Support Vector Machine (SVM) classifier based on Sequence-composition features (GC-content, codon usage). Accuracy values were obtained with a 5-fold cross-validation of the SVM classifier. The value in white represents the highest accuracy obtained.

855x481mm (38 x 38 DPI)

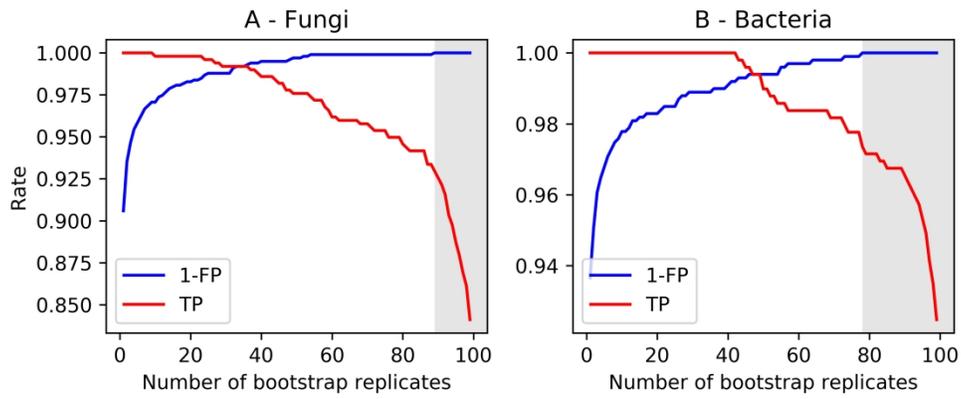Figure S5. Relationships between the false positive rate (FP) and the true positive rate (TP) for HGT candidate discovery with a Support Vector Machine classifier and the number of bootstrap replicate datasets in which the putative HGT candidate was found. The grey area corresponds to the number of bootstrap replicates where the false positive rate reaches the null value.

452x193mm (236 x 236 DPI)

Supplementary Table 1. Full assembly and annotation statistics for 31 Phytophthora genomes.

Species	Group	Clade	Genome					Sequence Reads
			Sequence Reads	Contigs	N50	Assembly Length	K-mer size estimate	
P. cactorum	EMR	1	NA[a]	7,888	15,053	56,443,298	NA[a]	28,693,892
P. idaei	EMR	1	NA[a]	7,163	14,461	53,468,943	NA[a]	NA[b]
P. pini	USDA-D	2	29,016,981	2,131	42,987	38,730,000	53,447,651	24,861,861
P. multivora	Scion	2	NA[a]	2,844	46,133	40,059,192	NA[a]	NA[b]
P. pluvialis	Scion	3	NA[a]	4,340	30,816	53,616,150	NA[a]	NA[b]
P. litchii	NJAU	4	17,809,306	2,543	34,546	38,200,938	58,822,528	20,209,114
P. palmivora	USDA-B	4	44,740,938	24,815	6,694	107,798,747	178,336,461	21,955,557
P. megakarya	USDA-B	4	29,074,100	24,073	7,093	101,609,312	170,257,105	19,683,391
P. agathidicida	Scion	5	NA[a]	3,754	19,544	37,337,699	NA[a]	NA[b]
P. taxon totara	Scion	5	NA[a]	4,425	30,809	55,576,372	NA[a]	NA[b]
P. parvispora	OSU	7	21,793,750	9,906	6,820	46,825,958	72,690,845	12,813,391
P. pisi	Fribourg	7	34,621,167	7,667	15,253	58,856,683	105,905,508	18,243,326
P. niederhauseri	Fribourg	7	39,377,584	26,463	4,805	90,270,009	126,036,921	20,314,715
P. robiniae	NJAU	7	34,689,250	14,865	8,754	69,938,814	107,714,597	12,626,895
P. cajani	OSU	7	26,013,994	18,255	5,113	64,854,085	108,588,397	12,589,952
P. vignae	OSU	7	25,708,667	10,330	8,363	56,137,732	85,443,060	12,598,323
P. melonis	NJAU	7	31,728,667	11,353	15,342	73,416,743	106,514,211	22,052,362
P. pistaciae	OSU	7	41,390,989	10,414	10,302	63,209,321	80,941,143	12,493,625
P. europaea	OSU	7	22,034,445	8,301	11,551	58,787,065	71,680,138	12,832,217
P. uliginosa	OSU	7	21,317,084	8,955	10,095	57,072,031	72,763,188	12,712,782
P. fragariae	USDA-G	7	32,666,667	8,544	20,362	76,969,737	100,354,493	13,274,875
P. rubi	USDA-G	7	38,666,667	9,434	17,808	74,863,594	119,201,141	13,035,132
P. pinifolia	UBC	6	NA[a]	22,610	6,021	74,478,861	NA[a]	36,047,200
P. lateralis	UBC	8	23,158,308	28,263	2,396	50,496,828	NA[a]	23,158,308
P. hibernalis	OSU	8	26,010,981	6,587	21,408	71,256,216	86,155,007	12,973,692
P. foliorum	OSU	8	16,677,112	5,320	15,800	48,973,082	56,233,540	12,894,827
P. brassicae	Fribourg	8	31,477,833	12,447	12,337	72,849,437	105,501,807	20,290,544
P. syringae	USDA-G	8	19,410,417	6,572	15,987	57,045,526	66,661,136	12,799,213
P. cryptogea	UBC	8	NA[a]	25,944	4,730	69,446,343	NA[a]	NA[b]
P. boehmeriae	NJAU	10	34,703,473	2,866	41,917	39,747,814	NA[a]	NA[b]
P. kernoviae	UBC	10	21,178,252	13,710	5,225	42,698,878	NA[a]	21,178,252

[a]Genome assemblies were completed in a different study.

[b]RNA was not sequenced.

Transcriptome					Annotations		
Assembled Contigs	Assembled N50	Contigs after TransDecoder	N50 after TransDecoder	Assembly Length	Repeat Percent	Predicted Genes	Strain
28,465	1,890	25,690	2,595	47,486,039	19.96	18,027	P414
NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	16.19	18,038	SCRP371
20,548	2,695	22,766	3,818	59,507,934	7.31	14,019	Unknown
NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	10.86	13,682	NZFS 3378
NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	16.04	16,285	LC9-1
28,496	2,139	28,596	3,087	60,642,271	5.98	12,391	SHS3
30,410	1,745	26,805	2,443	45,279,452	29.62	37,283	sbr112.9
31,088	1,802	28,293	2,614	49,914,387	31.94	33,614	zdho120
NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	5.76	12,923	NZFS 3772
NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	16.56	17,619	NZFS 3642
108,991	833	65,896	1,572	70,613,891	8.75	15,642	P8494;23
44,946	885	32,997	1,179	31,197,744	16.80	18,953	Unknown
75,089	1,177	52,635	1,476	59,497,972	20.96	29,587	Unknown
38,510	1,432	33,462	2,346	52,628,916	25.58	23,128	Unknown
33,359	2,139	36,595	2,878	76,345,822	20.65	19,840	P3105;27
41,393	1,596	37,102	2,552	62,795,853	17.45	18,535	P3019;24
28,272	1,873	26,842	2,618	48,278,380	25.93	21,276	Unknown
34,064	2,320	39,208	3,366	91,860,426	18.56	19,423	P6196;25
38,785	2,066	36,606	3,233	76,609,174	23.33	17,117	P10324;31
31,138	2,115	34,305	3,184	74,176,007	24.16	17,226	P10413;26
29,266	1,848	28,012	2,725	52,169,825	30.81	20,448	CBS 209.46
34,446	1,774	31,999	2,677	57,725,263	29.48	23,476	pd0101050015038
38,654	2,461	47,683	3,693	124,451,965	33.00	23,717	CBS 122922
13,645	670	8,094	1,690	10,528,234	23.39	19,503	CBS 168.42
36,774	2,899	49,849	4,759	150,938,535	32.46	23,578	P0647;28
27,112	2,358	32,393	3,607	80,179,397	19.26	16,083	P10974;29
35,153	1,485	27,252	2,207	42,057,210	28.39	26,010	Unknown
29,481	2,735	35,799	4,253	99,905,260	21.71	18,234	PSY-09-046
NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	17.65	24,936	CBS 418.71
NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	7.83	13,325	Unknown
10,461	791	7,931	1,499	9,532,089	4.15	14,322	CBS 122049

Host
Strawberry
Unknown
Unknown
Idesia polycarpa
Unknown
litchi
Theobroma cacao
Unknown
Agathis australis
Podocarpus totara
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Strawberry
Raspberry
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown

Supplementary Table 3. Over-represented GO terms for a set of 551 candidate HGT transcripts found in 35 *Phytophthora* genomes.

GO	Term	# terms in full set[1]	# terms in HGT set[2]	Pr(X=k)[3]
Biological process				
GO:0034079	butanediol biosynthetic process	28	6	<0.001
GO:0009405	pathogenesis	1538	6	<0.001
GO:0055114	obsolete oxidation-reduction process	1543	4	<0.001
GO:0046210	nitric oxide catabolic process	34	1	<0.05
GO:0051301	cell division	817	2	<0.05
GO:0000462	maturation of SSU-rRNA from tricistronic rRNA transcript	431	2	<0.01
GO:0071500	cellular response to nitrosative stress	54	1	<0.05
GO:0048285	organelle fission	64	1	<0.05
GO:0009245	lipid A biosynthesis	116	1	<0.05
GO:0051409	response to nitrosative stress	102	1	<0.05
GO:0045493	xylan catabolic process	151	1	<0.05
GO:0046514	ceramide catabolic process	174	1	<0.05
GO:0016024	CDP-diacylglycerol biosynthetic process	162	1	<0.05
GO:0009072	aromatic amino acid family metabolic process	198	1	<0.05
Molecular function				
GO:0000721	(R,R)-butanediol dehydrogenase activity	28	6	<0.001
GO:0016831	carboxy-lyase activity	201	18	<0.001
GO:0016491	oxidoreductase activity	10,881	14	<0.001
GO:0008080	N-acetyltransferase activity	1,082	4	<0.001
GO:0050525	cutinase activity	110	2	<0.001
GO:0015267	channel activity	742	3	<0.001
GO:0016787	hydrolase activity	8,099	6	<0.01
GO:0003724	RNA helicase activity	505	2	<0.01
GO:0008759	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase activity	61	1	<0.05
GO:0005344	oxygen transporter activity	61	1	<0.05
GO:0004838	tyrosine transaminase activity	72	1	<0.05
GO:0097573	glutathione oxidoreductase activity	82	1	<0.05
GO:0031176	endo-1,4-beta-xylanase activity	111	1	<0.05
GO:0008941	nitric oxide dioxygenase activity	109	1	<0.05
GO:0004731	purine-nucleoside phosphorylase activity	110	1	<0.05
GO:0019825	oxygen binding	118	1	<0.05
GO:0004300	enoyl-CoA hydratase activity	151	1	<0.05
GO:0010181	FMN binding	1595	2	<0.05
GO:0102121	ceramidase activity	162	1	<0.05
GO:0017040	N-acylsphingosine amidohydrolase activity	177	1	<0.05

| GO:0004605 | phosphatidate cytidylyltransferase activity | 244 | 1 | <0.05 |
| GO:0004318 | enoyl-[acyl-carrier protein] reductase (NADH) activity | 289 | 1 | <0.05 |

Cellular component

GO:0005576	extracellular region	4624	6	<0.001
GO:0005835	fatty acid synthase complex	271	1	<0.05
GO:0032040	small-subunit processome	704	2	<0.05
GO:0032153	cell division site	94	1	<0.05
GO:0005730	nucleolus	2223	2	<0.05

[1]Transcriptome of 35 *Phytophthora* genomes.

[2]551 HGT candidates.

[3]Probability (Q-value) of obtaining the same number of transcripts (k) or more by chance as given by a hypergeometric probability distribution.

Table S4. Putative pathogenicity functions inferred for 28 HGT candidates identified among *Phytophthora* spp.

	Ortho group	Putative function	Best hit on PHI-base (E-value) - Reference	Notes - Other reference(s)
HGT_1	OG5_134265	Endonuclease	PHI:5754, endonuclease, *Fusarium graminearum* (3.0E-024) - Yun et al. (2015)	-
HGT_2	OG5_126928_1	Xylulose reductase	PHI:2256, xylitol dehydrogenase, *Parastagonospora nodorum* (1.0E-128) - Lowe et al. (2008)	Oxidation of xylitol to xylulose; alters plant defense responses against aphids - MacWilliams et al. (2020) L-arabinose metabolism, degradation of hemicellulose - Dimarogona and Topakas (2016)
HGT_3	OG5_130201	Quinone oxidoreductase	No hit	Protection against oxidative stress in entomopathogenic fungi (Pedrini et al. 2015); Fungal protection against destructive host-produced quinones (Petrasch et al. 2019)
HGT_4	OG5_126615	Aquaporin	PHI:7047, water channel protein aquaporin, *Cryptococcus neoformans* (4.0E-012) - Meyers et al. (2017)	-
HGT_5	OG5_133138	NPP1	No hit	Necrosis-inducing *Phytophthora* protein - Seidl and Van den Ackerveken (2019)
HGT_6	OG5_177086	Phenol acid carboxylase	No hit	Upregulated expression during *Phytophthora cactorum* x strawberry interaction - Chen et al. (2011)
HGT_7	OG5_152798	Peptidase S9	No hit	Abundant in fungal plant pathogens - Muszewska et al. (2017)
HGT_8	OG5_144202	UDP-N-acetylglucosamine-peptide N-acetylglucosaminyltransferase	PHI:4921, flagellin glycosyltransferase, *Burkholderia cenocepacia* (1.0E-021) - Khodai-Kalaki et al. (2015)	-
HGT_9	OG5_140249	Alternative oxidase	No hit	-
HGT_10	OG5_139476	Dioxygenase	No hit	-

HGT_11	OG5_130394	Thioesterase	PHI:4988, sfp-type 4'-phosphopantetheinyl transferase, *Bipolaris maydis* (3.0E-005) - Zainudin et al. (2015)	-
HGT_12	OG5_135408	Endo-1,4-beta-xylanase GH10	PHI:7912, endo-beta-1,4-xylanase *Phytophthora parasitica* (1.0E-137) - Lai and Liou, (2018)	-
HGT_13	OG5_126609	4-coumarate CoA ligase	PHI:10606, long-chain-fatty-acid–Co Aligase, *Pseudomonas aeruginosa* (1.0E-032) - Pan et al. (2020)	-
HGT_14	OG5_135743	2,4-dienoyl-CoA reductase	PHI:8134, 3-Oxoacyl-[acyl-carrier-protein] reductase, *Salmonella enterica* (1.0e-014) - Kwan et al. (2018)	-
HGT_15	OG5_126928_2	Zinc-binding dehydrogenase, Polyketide synthase, enoylreductase domain	PHI:8321, gluconate 5-dehydrogenase, *Salmonella enterica* (1.0E-016) - Alves Batista et al. (2018)	-
HGT_16	OG5_135419	Putative tannase	PHI:10222, feruloyl esterase, *Valsa mali* (4.0E-030) - Xu et al. (2018)	-
HGT_17	OG5_133550	Putative pectinesterase CE8	PHI:278, pectin methylesterase, *Botrytis cinerea* (4.0E-077) - Valette-Collet et al. (2003)	-
HGT_18	OG5_223601	ATP-binding Cassette (ABC)	No hit	Putative virulence factor - Stergiopoulos et al. (2003); Zeng and Charkowski, (2021)
HGT_19	OG5_161673	Antibiotic biosynthesis monooxygenase	No hit	-
HGT_20	OG5_132142	Ribosomal-protein-alanine acetyltransferase	No hit	-
HGT_21	OG5_136477	Endoribonuclease	No hit	-
HGT_22	OG5_132299	Pectate lyase PL3	PHI :4620, pectate lyase, *Phytophthora capsici* (1.0E-029) – Fu et al. (2015)	-
HGT_23	OG5_203245	Csa-calmodulin	No hit	-

HGT_24	OG5_126661	Mannitol-1-phosphate dehydrogenase	PHI:9524, Alcohol dehydrogenase, *Candida albicans* (4.0E-021)- Song et al. (2019)	-
HGT_25	OG5_137179	hypothetical protein	No hit	-
HGT_26	OG5_129399	2-nitropropane dioxygenase	No hit	Triggers expression of type III secretion system in *Ralstonia solanacearum* to induce pathogenicity on tobacco - Zhang et al. (2017)
HGT_27	ORTHOMCL13162	hypothetical protein	No hit	-
HGT_28	OG5_126721	type I methionyl aminopeptidase	PHI:9334, effector protein, *Citrobacter rodentium* (4.0E-020) - Xia et al. (2019)	-

References

Alves Batista DF, de Freitas Neto OC, Maria de Almeida A, Maboni G, de Carvalho TF, de Carvalho TP, Barrow PA, Berchieri AJ. 2018. Evaluation of pathogenicity of *Salmonella gallinarum* strains harbouring deletions in genes whose orthologues are conserved pseudogenes in *S. pullorum*. *PLoS ONE* **13**: e0200585.

Chen X, Klemsdal SS, Brurberg MB. 2011. Identification and analysis of *Phytophthora cactorum* genes up-regulated during cyst germination and strawberry infection. *Current Genetics* **57**:297-315.

Dimarogona M, Topakas E. 2016. Chapter 12 - Regulation and heterologous expression of lignocellulosic enzymes in *Aspergillus*. In: *New and future developments in microbial biotechnology and bioengineering*. Elsevier, pp 171–190.

Fu L, Zhu C, Ding X, Yang X, Morris PF, Tyler BM, Zhang X. 2015. Characterization of cell-death-inducing members of the pectate lyase gene family in *Phytophthora capsici* and their contributions to infection of pepper. *Molecular Plant-Microbe Interactions* **28**: 766-775.

Khodai-Kalaki M, Andrade A, Mohamed YF, Valvano M. 2015. *Burkholderia cenocepacia* lipopolysaccharide modification and flagellin glycosylation affect virulence but not innate immune recognition in plants. *mBio* **6**: e00679-15.

Kwan G, Plagenz B, Cowles K, Pisithkul T, Amador-Noguez D, Barak JD. 2018. Few differences in metabolic network use found between *Salmonella enterica* colonization of plants and typhoidal mice. *Frontiers in Microbiology* **9**: 695.

Lai MY, Liou RF. 2018. Two genes encoding GH10 xylanases are essential for the virulence of the oomycete plant pathogen *Phytophthora parasitica. Current Genetics* **64**: 931–943.

Lowe RG, Lord M, Rybak K, Trengove RD, Oliver RP, Solomon PS. 2008. A metabolomic approach to dissecting osmotic stress in the wheat pathogen *Stagonospora nodorum*. *Fungal Genetics and Biology* **45**: 1479-1486.

MacWilliams JR, Dingwall S, Chesnais Q, Sugio A, Kaloshian I. 2020. AcDCXR is a cowpea aphid effector with putative roles in altering host immunity and physiology. *Frontiers in Plant Science* **11**: 605.

Meyers GL, Jung KW, Bang S, Kim J, Kim S, Hong J, Cheong E, Kim KH, Bahn YS. 2017. The water channel protein aquaporin 1 regulates cellular metabolism and competitive fitness in a global fungal pathogen *Cryptococcus neoformans*. *Environmental Microbiology Reports* **9**: 268-278.

Muszewska A, Stepniewska-Dziubinska MM, Steczkiewicz K, Pawlowska J, Dziedzic A, Ginalski K. 2017. Fungal lifestyle reflected in serine protease repertoire. *Scientific Reports* **7**: 9147.

Pan X, Fan Z, Chen L, Liu C, Bai F, Wei Y, Tian Z, Dong Y, Shi J, Chen H, Jin Y, Cheng Z, Jin S, Lin J, Wu W. 2020. PvrA is a novel regulator that contributes to *Pseudomonas aeruginosa* pathogenesis by controlling bacterial utilization of long chain fatty acids. *Nucleic Acids Research* **48**: 5967–5985.

Pedrini N, Ortiz-Urquiza A, Huarte-Bonnet C, Fan Y, Juárez MP, Keyhani NO. 2015. Tenebrionid secretions and a fungal benzoquinone oxidoreductase form competing components of an arms race between a host and pathogen. *Proceedings of the National Academy of Sciences of the United States of America* **112**: E3651–E3660.

Petrasch S, Silva CJ, Mesquida-Pesci SD, Gallegos K, van den Abeele C, Papin V, Fernandez-Acero FJ, Knapp SJ, Blanco-Ulate B. 2019. Infection strategies deployed by Botrytis cinerea, *Fusarium acuminatum*, and *Rhizopus stolonifer* as a function of tomato fruit ripening stage. *Frontiers in Plant Science* **10**: 223.

Seidl MF, Van den Ackerveken GV. 2019. Activity and phylogenetics of the broadly occurring family of microbial Nep1-like proteins. *Annual Review of Phytopathology* **57**: 367–386.

Song Y, Li S, Zhao Y, Zhang Y, Lv Y, Jiang Y, Wang Y, Li D, Zhang H. 2019. ADH1 promotes *Candida albicans* pathogenicity by stimulating oxidative phosphorylation. *International Journal of Medical Microbiology* **309**: 151330.

Stergiopoulos I, Zwiers LH, De Waard MA. 2003. The ABC transporter MgAtr4 is a virulence factor of *Mycosphaerella graminicola* that affects colonization of substomatal cavities in wheat leaves. *Molecular Plant-Microbe Interactions* **16**: 689-698.

Valette-Collet O, Cimerman A, Reignault P, Levis C, Boccara M. 2003. Disruption of *Botrytis cinerea* pectin methylesterase gene bcpme1 reduces virulence on several host plants. *Molecular Plant-Microbe Interactions* **16**: 360-367.

Xia X, Liu Y, Hodgson A, Xu D, Guo W, Yu H, She W, Zhou C, Lan L, Fu K, Vallance BA, Wan F. 2019. EspF is crucial for *Citrobacter rodentium*-induced tight junction disruption and lethality in immunocompromised animals. *PLoS Pathogens* **15**: e1007898.

Xu M, Gao X, Chen J, Yin Z, Feng H, Huang L. 2018. The feruloyl esterase genes are required for full pathogenicity of the apple tree canker pathogen *Valsa mali*. *Molecular Plant Pathology* **19**: 1353-1363.

Yun Y, Liu Z, Yin Y, Jiang J, Chen Y, Xu JR, Ma Z. 2015. Functional analysis of the *Fusarium graminearum* phosphatome. *The New Phytologist* **207**:119-134.

Zainudin NA, Condon B, De Bruyne L, Van Poucke C, Bi Q, Li W, Höfte M, Turgeon BG. 2015. Virulence, Host-Selective toxin production, and development of three *Cochliobolus* phytopathogens lacking the Sfp-Type 4'-Phosphopantetheinyl transferase Ppt1. *Molecular Plant-Microbe Interactions* **28**: 1130-1141.

Zeng Y, Charkowski AO. 2021.The role of ATP-binding cassette transporters in bacterial phytopathogenesis. *Phytopathology*, **111**: 600-610.

Zhang W, Li J, Tang Y, Chen K, Shi X, Ohnishi K, Zhang Y. 2017. Involvement of NpdA, a putative 2-nitropropane dioxygenase, in the T3SS expression and full virulence in *Ralstonia solanacearum* OE1-1. *Frontiers in Microbiology* **8**: 1990.

Supplementary Table 5. Putative pathogenicity functions inferred for 28 HGT candidates identified among *Phytophthora* spp.

	Ortho group	Putative function	Best hit on PHI-base (E-value) - Reference	Notes - Other reference(s)
HGT_1	OG5_134265	Endonuclease	PHI:5754, endonuclease, *Fusarium graminearum* (3.0E-024) - Yun et al. (2015)	-
HGT_2	OG5_126928_1	Xylulose reductase	PHI:2256, xylitol dehydrogenase, *Parastagonospora nodorum* (1.0E-128) - Lowe et al. (2008)	Oxidation of xylitol to xylulose; alters plant defense responses against aphids - MacWilliams et al. (2020) L-arabinose metabolism, degradation of hemicellulose - Dimarogona and Topakas (2016)
HGT_3	OG5_130201	Quinone oxidoreductase	No hit	Protection against oxidative stress in entomopathogenic fungi (Pedrini et al. 2015); Fungal protection against destructive host-produced quinones (Petrasch et al. 2019)
HGT_4	OG5_126615	Aquaporin	PHI:7047, water channel protein aquaporin, *Cryptococcus neoformans* (4.0E-012) - Meyers et al. (2017)	-
HGT_5	OG5_133138	NPP1	No hit	Necrosis-inducing *Phytophthora* protein - Seidl and Van den Ackerveken (2019)
HGT_6	OG5_177086	Phenol acid carboxylase	No hit	Upregulated expression during *Phytophthora cactorum* x strawberry interaction - Chen et al. (2011)
HGT_7	OG5_152798	Peptidase S9	No hit	Abundant in fungal plant pathogens - Muszewska et al. (2017)
HGT_8	OG5_144202	UDP-N-acetylglucosamine-peptide N-acetylglucosaminyltransferase	PHI:4921, flagellin glycosyltransferase, *Burkholderia cenocepacia* (1.0E-021) - Khodai-Kalaki et al. (2015)	-
HGT_9	OG5_140249	Alternative oxidase	No hit	-
HGT_10	OG5_139476	Dioxygenase	No hit	-
HGT_11	OG5_130394	Thioesterase	PHI:4988, sfp-type 4'-phosphopantetheinyl transferase, *Bipolaris maydis* (3.0E-005) - Zainudin et al. (2015)	-
HGT_12	OG5_135408	Endo-1,4-beta-xylanase	PHI:7912, endo-beta-1,4-xylanase *Phytophthora parasitica* (1.0E-137) - Lai and Liou, (2018)	-
HGT_13	OG5_126609	4-coumarate CoA ligase	PHI:10606, long-chain-fatty-acid–Co Aligase, *Pseudomonas aeruginosa* (1.0E-032) - Pan et al. (2020)	-
HGT_14	OG5_135743	2,4-dienoyl-CoA reductase	PHI:8134, 3-Oxoacyl-[acyl-carrier-protein] reductase, *Salmonella enterica* (1.0e-014) - Kwan et al. (2018)	-
HGT_15	OG5_126928_2	Zinc-binding dehydrogenase, Polyketide synthase, enoylreductase domain	PHI:8321, gluconate 5-dehydrogenase, *Salmonella enterica* (1.0E-016) - Alves Batista et al. (2018)	-
HGT_16	OG5_135419	Putative tannase	PHI:10222, feruloyl esterase, *Valsa mali* (4.0E-030) - Xu et al. (2018)	-
HGT_17	OG5_133550	Putative pectinesterase	PHI:278, pectin methylesterase, *Botrytis cinerea* (4.0E-077) - Valette-Collet et al. (2003)	-
HGT_18	OG5_223601	ATP-binding Cassette (ABC)	No hit	Putative virulence factor - Stergiopoulos et al. (2003); Zeng and Charkowski, (2021)
HGT_19	OG5_161673	Antibiotic biosynthesis monooxygenase	No hit	-
HGT_20	OG5_132142	Ribosomal-protein-alanine acetyltransferase	No hit	-
HGT_21	OG5_136477	Endoribonuclease	No hit	-
HGT_22	OG5_132299	Pectate lyase	PHI :4620, pectate lyase, *Phytophthora capsici* (1.0E-029) – Fu et al. (2015)	-
HGT_23	OG5_203245	Csa-calmodulin	No hit	-
HGT_24	OG5_126661	Mannitol-1-phosphate dehydrogenase	PHI:9524, Alcohol dehydrogenase, *Candida albicans* (4.0E-021)- Song et al. (2019)	-
HGT_25	OG5_137179	hypothetical protein	No hit	-
HGT_26	OG5_129399	2-nitropropane dioxygenase	No hit	Triggers expression of type III secretion system in *Ralstonia solanacearum* to induce pathogenicity on tobacco - Zhang et al. (2017)
HGT_27	ORTHOMCL13162	hypothetical protein	No hit	-
HGT_28	OG5_126721	type I methionyl aminopeptidase	PHI:9334, effector protein, *Citrobacter rodentium* (4.0E-020) - Xia et al. (2019)	-

References

Alves Batista DF, de Freitas Neto OC, Maria de Almeida A, Maboni G, de Carvalho TF, de Carvalho TP, Barrow PA, Berchieri AJ. 2018. Evaluation of pathogenicity of *Salmonella gallinarum* strains harbouring deletions in genes whose orthologues are conserved pseudogenes in *S. pullorum*. *PLoS ONE* **13**: e0200585.

Chen X, Klemsdal SS, Brurberg MB. 2011. Identification and analysis of *Phytophthora cactorum* genes up-regulated during cyst germination and strawberry infection. *Current Genetics* **57**:297-315.

Dimarogona M, Topakas E. 2016. Chapter 12 - Regulation and heterologous expression of lignocellulosic enzymes in *Aspergillus*. In: *New and future developments in microbial biotechnology and bioengineering*. Elsevier, pp 171–190.

Fu L, Zhu C, Ding X, Yang X, Morris PF, Tyler BM, Zhang X. 2015. Characterization of cell-death-inducing members of the pectate lyase gene family in *Phytophthora capsici* and their contributions to infection of pepper. *Molecular Plant-Microbe Interactions* **28**: 766-775.

Khodai-Kalaki M, Andrade A, Mohamed YF, Valvano M. 2015. *Burkholderia cenocepacia* lipopolysaccharide modification and flagellin glycosylation affect virulence but not innate immune recognition in plants. *mBio* **6**: e00679-15.

Kwan G, Plagenz B, Cowles K, Pisithkul T, Amador-Noguez D, Barak JD. 2018. Few differences in metabolic network use found between *Salmonella enterica* colonization of plants and typhoidal mice. *Frontiers in Microbiology* **9**: 695.

Lai MY, Liou RF. 2018. Two genes encoding GH10 xylanases are essential for the virulence of the oomycete plant pathogen *Phytophthora parasitica. Current Genetics* **64**: 931–943.

Lowe RG, Lord M, Rybak K, Trengove RD, Oliver RP, Solomon PS. 2008. A metabolomic approach to dissecting osmotic stress in the wheat pathogen *Stagonospora nodorum*. *Fungal Genetics and Biology* **45**: 1479-1486.

MacWilliams JR, Dingwall S, Chesnais Q, Sugio A, Kaloshian I. 2020. AcDCXR is a cowpea aphid effector with putative roles in altering host immunity and physiology. *Frontiers in Plant Science* **11**: 605.

Meyers GL, Jung KW, Bang S, Kim J, Kim S, Hong J, Cheong E, Kim KH, Bahn YS. 2017. The water channel protein aquaporin 1 regulates cellular metabolism and competitive fitness in a global fungal pathogen *Cryptococcus neoformans*. *Environmental Microbiology Reports* **9**: 268-278.

Muszewska A, Stepniewska-Dziubinska MM, Steczkiewicz K, Pawlowska J, Dziedzic A, Ginalski K. 2017. Fungal lifestyle reflected in serine protease repertoire. *Scientific Reports* **7**: 9147.

Pan X, Fan Z, Chen L, Liu C, Bai F, Wei Y, Tian Z, Dong Y, Shi J, Chen H, Jin Y, Cheng Z, Jin S, Lin J, Wu W. 2020. PvrA is a novel regulator that contributes to *Pseudomonas aeruginosa* pathogenesis by controlling bacterial utilization of long chain fatty acids. *Nucleic Acids Research* **48**: 5967–5985.

Pedrini N, Ortiz-Urquiza A, Huarte-Bonnet C, Fan Y, Juárez MP, Keyhani NO. 2015. Tenebrionid secretions and a fungal benzoquinone oxidoreductase form competing components of an arms race between a host and pathogen. *Proceedings of the National Academy of Sciences of the United States of America* **112**: E3651–E3660.

Petrasch S, Silva CJ, Mesquida-Pesci SD, Gallegos K, van den Abeele C, Papin V, Fernandez-Acero FJ, Knapp SJ, Blanco-Ulate B. 2019. Infection strategies deployed by Botrytis cinerea, *Fusarium acuminatum*, and *Rhizopus stolonifer* as a function of tomato fruit ripening stage. *Frontiers in Plant Science* **10**: 223.

Seidl MF, Van den Ackerveken GV. 2019. Activity and phylogenetics of the broadly occurring family of microbial Nep1-like proteins. *Annual Review of Phytopathology* **57**: 367–386.

Song Y, Li S, Zhao Y, Zhang Y, Lv Y, Jiang Y, Wang Y, Li D, Zhang H. 2019. ADH1 promotes *Candida albicans* pathogenicity by stimulating oxidative phosphorylation. *International Journal of Medical Microbiology* **309**: 151330.

Stergiopoulos I, Zwiers LH, De Waard MA. 2003. The ABC transporter MgAtr4 is a virulence factor of *Mycosphaerella graminicola* that affects colonization of substomatal cavities in wheat leaves. *Molecular Plant-Microbe Interactions* **16**: 689-698.

Valette-Collet O, Cimerman A, Reignault P, Levis C, Boccara M. 2003. Disruption of *Botrytis cinerea* pectin methylesterase gene bcpme1 reduces virulence on several host plants. *Molecular Plant-Microbe Interactions* **16**: 360-367.

Xia X, Liu Y, Hodgson A, Xu D, Guo W, Yu H, She W, Zhou C, Lan L, Fu K, Vallance BA, Wan F. 2019. EspF is crucial for *Citrobacter rodentium*-induced tight junction disruption and lethality in immunocompromised animals. *PLoS Pathogens* **15**: e1007898.

Xu M, Gao X, Chen J, Yin Z, Feng H, Huang L. 2018. The feruloyl esterase genes are required for full pathogenicity of the apple tree canker pathogen *Valsa mali*. *Molecular Plant Pathology* **19**: 1353-1363.

Yun Y, Liu Z, Yin Y, Jiang J, Chen Y, Xu JR, Ma Z. 2015. Functional analysis of the *Fusarium graminearum* phosphatome. *The New Phytologist* **207**:119-134.

Zainudin NA, Condon B, De Bruyne L, Van Poucke C, Bi Q, Li W, Höfte M, Turgeon BG. 2015. Virulence, Host-Selective toxin production, and development of three *Cochliobolus* phytopathogens lacking the Sfp-Type 4'-Phosphopantetheinyl transferase Ppt1. *Molecular Plant-Microbe Interactions* **28**: 1130-1141.

Zeng Y, Charkowski AO. 2021.The role of ATP-binding cassette transporters in bacterial phytopathogenesis. *Phytopathology*, **111**: 600-610.

Zhang W, Li J, Tang Y, Chen K, Shi X, Ohnishi K, Zhang Y. 2017. Involvement of NpdA, a putative 2-nitropropane dioxygenase, in the T3SS expression and full virulence in *Ralstonia solanacearum* OE1-1. *Frontiers in Microbiology* **8**: 1990.