Data in Brief 39 (2021) 107663

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

# Comparative analysis of *Phytophthora* genomes data

Rui-Fang Gao[a,b], Jie-Yu Wang[c,d], Ke-Wei Liu[e], Zhi-Wen Wang[f], Diyang Zhang[c], Xiang Zhao[f], Wen-Ying Zhong[f], Wen-Chieh Tsai[g,h,i,*], Zhong-Jian Liu[c,j,k,*], Gui-Ming Zhang[a,b,*]

[a] *Animal & Plant Inspection and Quarantine Technology Center of Shenzhen Customs District P.R. China, Shenzhen 518045, China*
[b] *Shenzhen Key Laboratory for Research & Development on Detection Technology of Alien Pests, Shenzhen Academy of Inspection and Quarantine, Shenzhen 518045, China*
[c] *Key Laboratory of National Forestry and Grassland Administration for Orchid Conservation and Utilization at College of Landscape Architecture, Fujian Agriculture and Forestry University, Fuzhou 350002, China*
[d] *Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China*
[e] *Tsinghua-Berkeley Shenzhen Institute (TBSI), Center for Biotechnology and Biomedicine, Shenzhen Key Laboratory of Gene and Antibody Therapy, State Key Laboratory of Chemical Oncogenomics, State Key Laboratory of Health Sciences and Technology, Institute of Biopharmaceutical and Health Engineering (iBHE), Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China*
[f] *PubBio-Tech Services Corporation, Wuhan 430070, China*
[g] *Orchid Research and Development Center, National Cheng Kung University, Tainan 701, Taiwan*
[h] *Department of Life Sciences, National Cheng Kung University, Tainan 701, Taiwan*
[i] *Institute of Tropical Plant Sciences and Microbiology, National Cheng Kung University, Tainan 701, Taiwan*
[j] *Institute of Vegetable and Flowers, Shandong Academy of Agricultural Sciences, 250100, Jinan, China*
[k] *Zhejiang Institute of Subtropical Crops, Zhejiang Academy of Agricultural Sciences, Wenzhou 325005, China*

ARTICLE INFO

*Article history:*
Received 5 March 2021
Revised 25 November 2021
Accepted 29 November 2021
Available online 2 December 2021

ABSTRACT

The data presented here are related to the article entitled "Comparative analysis of *Phytophthora* genomes reveals oomycete pathogenesis in crops" [1]. These data contain the description of genomic structure of the two plant pathogens, *P. fragariae* and *P. rubi* and characterize several gene families associated with pathogenicity of them: P450, *ACX* gene families, CAZymes and effector. This data presents the

DOI of original article: 10.1016/j.heliyon.2021.e06317
* Corresponding authors
*E-mail addresses:* tsaiwc@mail.ncku.edu.tw (W.-C. Tsai), zjliu@fafu.edu.cn (Z.-J. Liu), zgm2001cn@163.com (G.-M. Zhang).

https://doi.org/10.1016/j.dib.2021.107663

*Keywords:*
*Phytophthora fragariae*
*Phytophthora rubi*
pathogenic gene families

relevant results of two newly sequenced *P. fragariae* and *P. rubi*, so as to provide data for further studies by researchers.


## Specifications Table

| | |
|---|---|
| Subject | Agricultural and Biological Sciences (General) |
| Specific subject area | Plant pathogens, genomic features associated with pathogenicity of *Phytophthora* species |
| Type of data | Table, Chart, Graph, Figure |
| How data were acquired | The whole genome sequences of two *Phytophthora* species were sequenced by Illumina HiSeq 2500 platform. The other available whole genome sequences, protein-coding genes, repeat sequences, gene families were from public datasets. |
| Data format | Raw, Analyzed |
| Parameters for data collection | The genome assembled by ourselves was of good quality with the coverage more than 200, the length of contigs N50 almost 50 kb and the length of scaffold N50 almost 100 kb. The genomic data from public datasets had complete information, well assembled and annotated, and were all published in authoritative journals with more citations. |
| Description of data collection | We sequenced and assembled the genome of using a strategy that combined paired-end and mate-paired libraries. The paired-end libraries of 180 bp, 300 bp and 500 bp, and mate-paired libraries of 3 kb and 5 kb were constructed for each genome. *De novo* assembly was performed using Allpaths-LG. The assembled genome scaffolds were aligned to the most closely related publicly available genomes using MUMmer which included *P. infestans* T30–4, *P. parasitica* INRA-310, *P. sojae* V3.0, *P. nicotianae* and *P. ramorum*. Scaffolds were broken at points where non-contiguous regions of the reference genome were juxtaposed and then re-ordered so that the scaffolds were syntenic with the reference genome. |
| Data source location | The sequenced strains were imported from Westerdijk Fungal Biodiversity Institute, Netherlands. The original genome sequenced data was stored in Shenzhen, Fuzhou and Wuhan. The other genome sequence and data were downloaded from public sources, all freely available. |
| Data accessibility | This article provides the analyzed data. Raw data are deposited in a public repository.<br>The genome assemblies of two *Phytophthora* species were uploaded to NCBI database.<br>Repository name: NCBI database<br>Raw data identification number (SRA):<br>*Phytophthora fragairae*: SRR16352867, SRR16352866, SRR16352865, SRR16352864<br>*Phytophthora rubi*: SRR16351716, SRR16351717<br>The genome assembly of two *Phytophthora* species was uploaded to NCBI database<br>Direct URL to genome assembly:<br>https://www.ncbi.nlm.nih.gov/assembly/GCA_000686205.4 (*Phytophthora fragairae*),<br>https://www.ncbi.nlm.nih.gov/assembly/GCA_000687305.2 (*Phytophthora rubi*) |
| Related research article | R.F. Gao, J.Y.Wang, K.W. Liu, K. Yoshida, Y.Y. Hsiao, Y.X. Shi, K.C. Tsai, Y.Y. Chen, N. Mitsuda, C. K. Liang. *et al.* Comparative analysis of *Phytophthora* genomes reveals oomycete pathogenesis in crops. Heliyon. 2021, 7, e6317. https://doi.org/10.1016/j.heliyon.2021.e06317 |

## Value of the Data

- Data provided in the article describe the genome characteristics, the mechanisms of pathogenicity of *Phytophthora* species, which are the most important plant pathogens.
- Data are available for researchers and geneticists to search and compare the valuable information of different functions genes, gene family and protein information of different *Phytophthora* species
- Data provided in this article provide important information on the description of the pathogenic mechanism of *Phytophthora*.

## 1. Data Description

This data article is composed of one table and seven figures. These data contain the characteristic description of genomic structure and pathogenicity of the two pathogens, *P. fragariae* and *P. rubi*, including the distribution of gene density at the genome level and several gene families associated with pathogenicity: P450 [2], *ACX* gene families [3], CAZymes [4] and effector [5]. These results complement the recent published literature on genomic and pathogenicity analyses of several *Phytophthora* pathogens [1].

The GO function of *ACX* motif is shown in Table. The mycelial growth in plate and the infection to plant seedlings of two *Phytophthora* strains is present in Figs. 1. 2 is the distribution of polymorphisms in *P. fragariae* and *P. rubi* according to local gene density. The structure of 19 predicted Acyl-coenzyme A oxidase gene families (*ACX*) motifs and the seven clades of these families is shown in Figs. 3 and 4 respectively. Fig. 5 contains the information about pathogenic types and structural variety of two kinds of effectors of RxLRs and CRN in *Phytophthora* species. the distribution of each type of CAZyme in six *Phytophthora* species and *Plasmopara halstedii*, including GHs, GTs, CEs, PLs, AAs and CBMs, is shown in Fig. 6. Fig. 7 is the phylogenetic tree constructed based on the P450 genes in *P. fragariae, P. rubi* and other relative species.

## 2. Experimental Design, Materials and Methods

### 2.1. Materials availability

*P. fragariae* and *P. rubi* are highly similar in morphology and physiology, but infect different hosts. Sequencing the genomes of these species will advance our understanding of the genomic mechanisms underlying host adaptation and knowledge of molecular mechanisms of plant pathogenicity [6]. The *P. fragariae* and *P. rubi* strains were imported from Westerdijk Fungal Biodiversity Institute (strain numbers 309.62 and 109,892, respectively). *P. fragariae* 309.62 was isolated from *Fragaria* fruit by C.J. Hickman in Scotland, and *P. rubi* 109,892 was isolated by C. Brasier from raspberry roots in Scotland. We induced the production of motile zoospores in sterile water for over 5 h at 13–14 °C. Single zoospores were picked by inoculating needles under a microscope, followed by germination in V8 plates at 18 °C. After 48 h, the mycelia began to form. On the plate, the mycelia of *P. fragariae* were concentric and round with dense mycelia, while the mycelia of *P. rubi* grew uniformly and was relatively loose, as shown in Fig. 1a. The two pathogens had different hosts, and were inoculated on *Dendrobium candidum* seedlings to observe the pathogenicity. Four different inoculation methods were performed using fresh mycelium blocks to discover the pathogenic characteristics respectively, inoculated into growth medium without antibiotics, smeared on the leaves, needle inoculated into the leaves and inoculated into growth medium containing double antibiotics (ampicillin and streptomycin). After 30 days, the difference in pathogenicity during the observation period is shown in Fig. 1b.

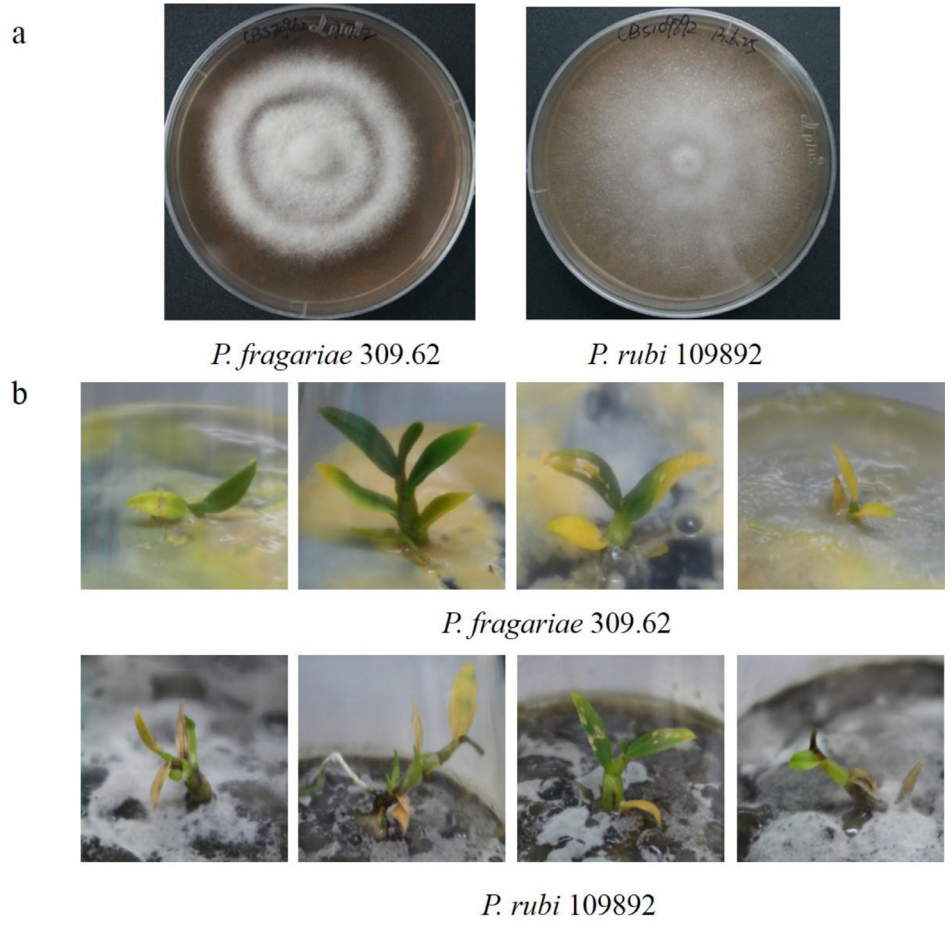


**Fig. 1.** The growth morphology in plate and the infection to plant seedlings of two *Phytophthora* strains. (a) Pictures of mycelial growth on V8 medium *P. fragariae* 309.62 (left) and *P. rubi* 109,892 (right). (b) The infected plants 30 days post inoculation (dpi). The four pictures in top are the infection to *Dendrobium officinale* seedlings of *P. fragariae* 309.62 under different inoculation conditions. From left to right, inoculated into growth medium without antibiotics, smeared on the leaves, needle inoculated into the leaves and inoculated into growth medium containing double antibiotics (ampicillin and streptomycin).The four pictures in bottom are the infection of *P. rubi* 109,892 under the same inoculation condition with *P. fragariae* 309.62 respectively.

### *2.2. The evolutionary rate at the genome level*

A fundamental insight from analyses of filamentous plant pathogen genomes is that genes located in repeat-rich regions tend to evolve more rapidly than those in the rest of the genome [7,8]. We calculated the distribution of gene evolution rate at the genomic level to verify the conclusions of previous studies. This pattern could be demonstrated through comparative genomics of closely related species. Blastp comparison was used to find out the mutually optimal gene pairs in the whole genome of *P. fragariae* and *P. rubi*. The *dn*/*ds* values calculated using CODEML in the PAML package are shown in each branch [9]. The gene pairs with *dn*/*ds* <1 are selected and mapped in Fig. 2.

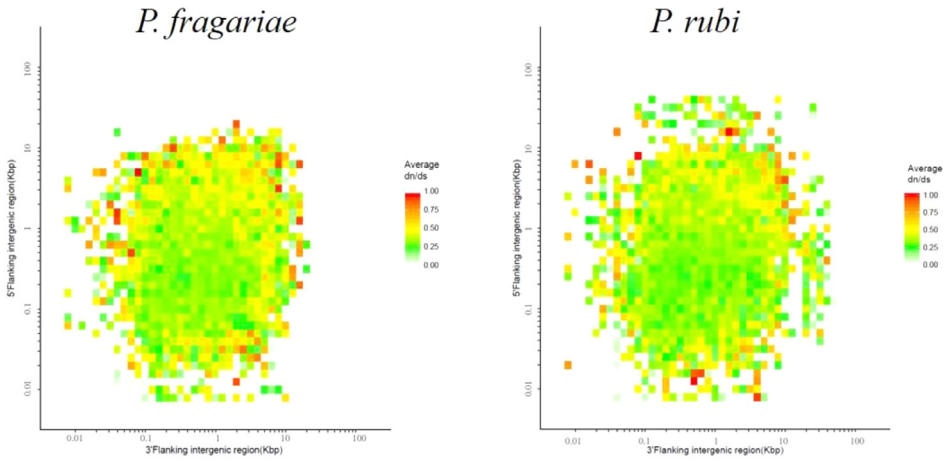


**Fig. 2.** Distribution of polymorphisms in *P. fragariae* and *P. rubi* according to local gene density. The horizontal axis is the length of the 3' and 5′ intergenic regions of this gene, and each point represents the average *dn*/*ds* value of the interval length (bin). The average values of *dn*/*ds* ratio (nonsynoymous to synonymous substitution rate) associated with genes in each bin are shown as a color-coded heat map.

**Table 1**
The GO function of *ACX* motif.

| motif name | GO term | p-value | GO name | Functional description |
|---|---|---|---|---|
| motif 6 | GO:0,010,467 | 8.05E-06 | BP | gene expression |
| motif 7 | GO:0,005,739 | 2.45E-06 | CC | mitochondrion |
| motif 8 | GO:0,006,414 | 4.03E-06 | BP | translational elongation |
| motif 15 | GO:0,044,446 | 7.70E-06 | CC | intracellular organelle part |
| | GO:0,043,234 | 2.54E-05 | CC | protein complex |
| motif 17 | GO:0,009,225 | 4.03E-06 | BP | nucleotide-sugar metabolic process |
| motif 19 | GO:0,008,289 | 3.68E-06 | MF | lipid binding |
| | GO:0,005,737 | 1.56E-05 | CC | cytoplasm |

### 2.3. ACX gene families

Together with the CoA synthetase, *ACX* is acting in the $\beta$-oxidation pathway involved in the suppression of plant resistance via signaling molecules, such as jasmonic acid. We analyzed the structure of *ACX* gene family in order to find out whether it has pathogenic characteristics in *Phytophthora*. The motif was predicted by the motif prediction software MEME, the motif distribution characteristics of each clade sequence were analyzed, and the motif function was analyzed by GOMO. The phylogenetic analysis shows the *ACX* gene family for *Phytophthora* and its relative species could be divided into seven clades according to the predicted 19 motifs. The *ACX* gene families in Fig. 2 are clustered into seven categories, but none of the species is unique to a category. As shown in Fig. 2, all seven categories (A, B, C, D, E, F, and G) contain Multiple species (Figs. 3 and 4). The enriched GO function of *ACX* motif of motif 6,7,8,15,17 and 19 is shown in Table 1.

### 2.4. The variety of RxLR and CRN

PHI-base is an expertly curated source for genes which have been proven to affect the pathogen-host interaction [10]. 4017 genes of *P. fragariae* are identified by PHI-base, among them 423 are annotated as 'effector' in other pathogen species. For *P. rubi*, 3710 genes are identified,

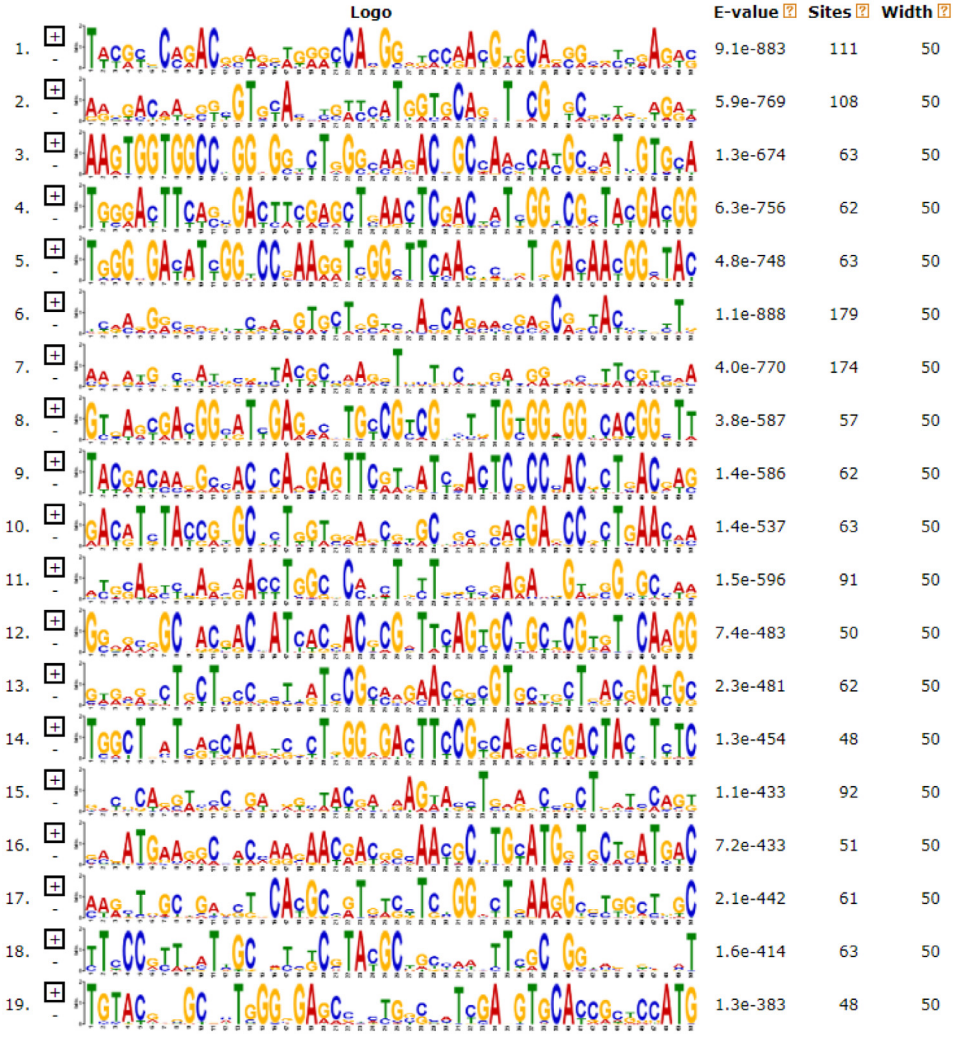


**Fig. 3.** 19 predicted Acyl-coenzyme A oxidase gene families (*ACX*) motifs by Multiple Em for Motif Elicitation (MEME). The smaller the E-value, the more significant the motif is. Sites is the peak number supporting the motif, and width is the length of the motif.

among them 368 are effectors and 16 are having identity higher than 70%. RxLR and CRN genes, which dominate as effectors in some other *Phytophthora* species, play no or very minor role in *P. fragariae* and *P. rubi*. RxLR is named after an N-terminal motif (x is any amino acid) involved in uptake by host cells [11]. CRNs are named after a "crinkling" or necrotic phenotype that occurs when several of these proteins are overexpressed in plants [12]. Blastp comparison was used to find out the mutually optimal gene pairs in the whole genome of *P. fragariae* and *P. rubi*. Selection conditions of comparison results: E-value $< 1e^{-5}$, Identity >95% and Coverage >0.7, the similarity of two sequences is greater than 95% and the length of similar fragments accounts for more than 80% of the predicted effect factor and the sequence length of PHI database). We used MEGA to construct the gene tree and used MEME to identify the motifs. Then, we summed the pathogenic and structural characterics of verified RxLR and CRN genes families in each *Phy-*

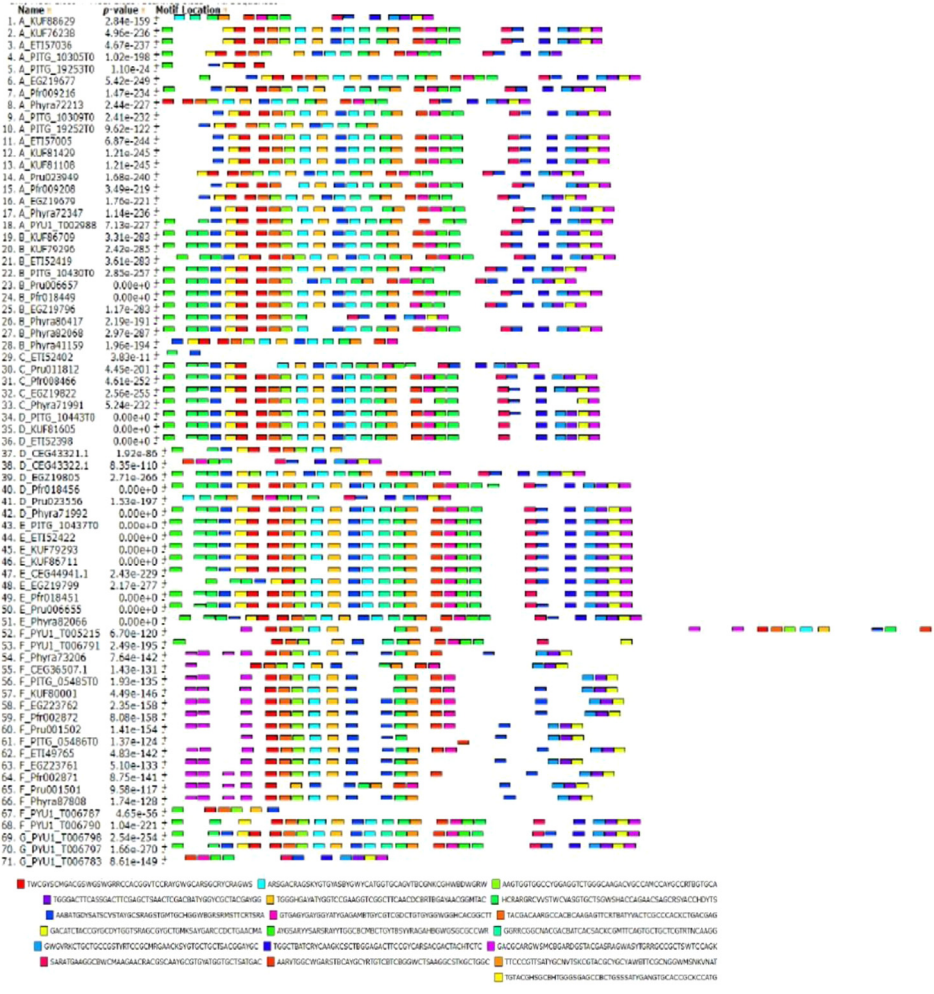


**Fig. 4.** Seven clades of 19 predicted Acyl-coenzyme A oxidase gene families (*ACX*) motifs. There are seven clades (A, B, C, D, E, F, and G) of 71 *ACX* sequences which are classified based on the distribution of 19 predicted *ACX* motifs.

*tophthora* species, including the motif distribution, the corresponding host and the disease to refer.

### *2.5. CAZymes*

Although the previous study have done some analysis about cazymes, the comparisons among these *Phytophthora* species have not been made. We performed a scan of the seven *Phytophthora* genomes and allied *Plasmopara halstedii* against the CAZymes database (http://www.cazy.org/), and the statistically comparison was re-conducted in Fig. 6. We used run_dbCAN.py (https://github.com/linnabrown/run_dbcan) to identify the carbohydrate-active enzyme in each genome and used the Circos software to draw the distribution figure. It proves that the total copy numbers of cazymes homologs are *P. sojae* (470), *P. nicotianae* (462), *P. parasitica* (449), *P. fragariae* (396), *P. ramorum* (382), *P. rubi* (371) and *P. infestans* (364). We then analyze the

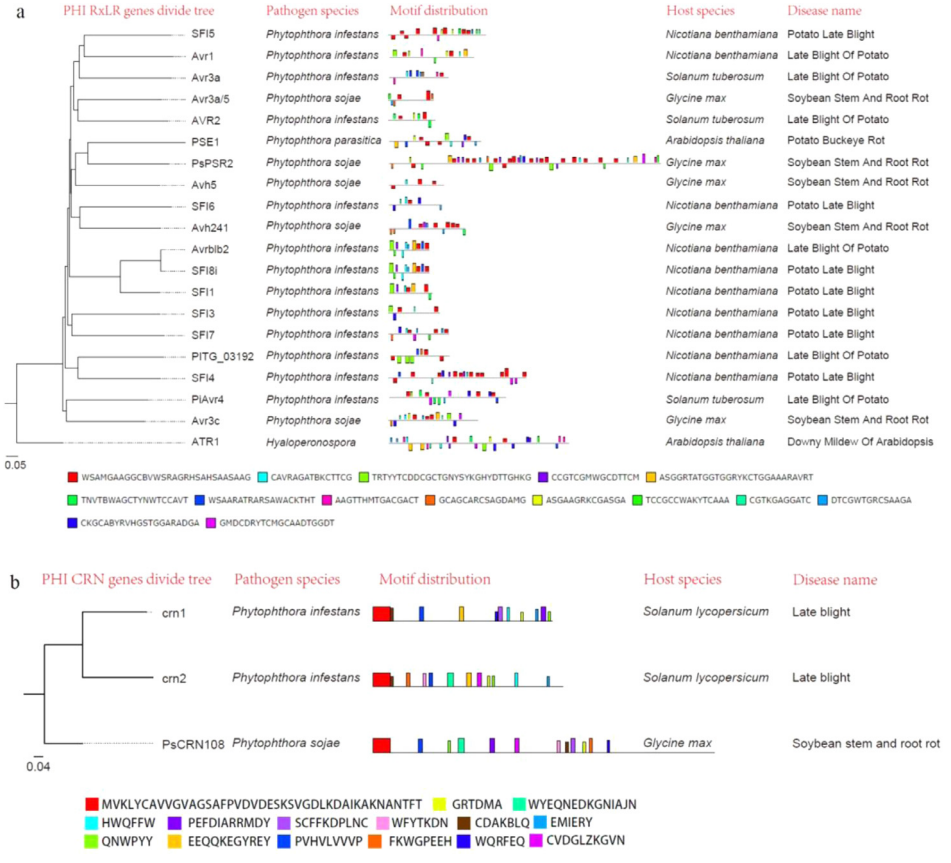


**Fig. 5.** The pathogenic types and structural variety of RxLRs (a) and CRN (b) in *Phytophthora* species. 20 RxLRs genes could be matched in PHI-base, and only 3 CRN genes were found. The motif distribution, host species and disease name are showed in the figure, the colored rectangles mean the different motifs.

similarity and dissimilarity between the cazyme family distributions between *P. infestans* and *P. fragariae* as well as *P. rubi*.

### *2.6. P450 analysis*

As a key mechanism in evading host plant defense, the cytochrome P450 genes are important for plant pathogen *Phytophthora* species. We downloaded 1088 P450 gene sequences belonging to all available *Phytophthora* species from NCBI, and blasted the *P. fragariae* and *P. rubi* genomes against these query collections. Only those with matched identity$\geq$40%, evalue$\leq$1e$^{-10}$ and score $\geq$ 200 were retained as candidates. We further annotated these candidates using Pfam, and genes containing the PF00067 domain were deemed as authentic P450 genes. After the multiple sequence alignment by muscle, Treebest was used to construct the P450 gene tree. As can be seen from the phylogenetic tree we constructed based on the P450 genes in these *Phytophthora* species (Fig. 7), in most of the cases, the genes in *P. fragariae* and *P. rubi* appear in pair on the tree. However, there are some exceptions.

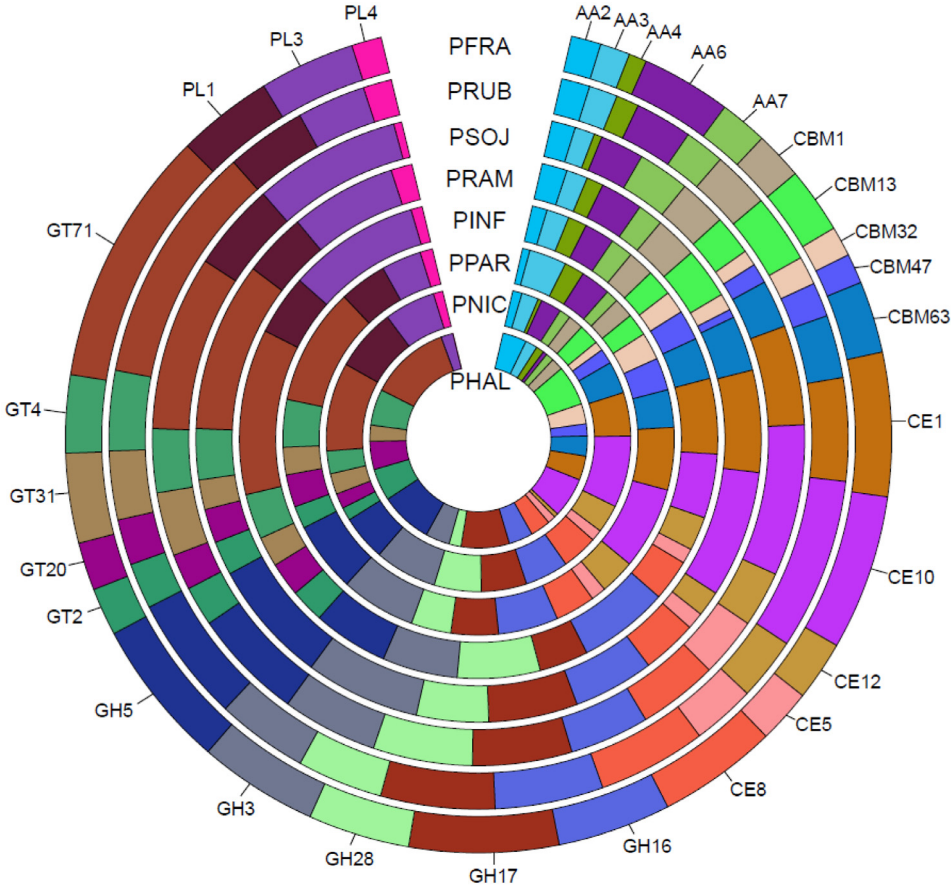


**Fig. 6.** The distribution of each type of CAZyme. Glycoside Hydrolases (GHs), Glycosyl Transferases (GTs), Carbohydrate Esterases (CEs), Polysaccharide Lyases (PLs), Auxiliary Activities (AAs) and Carbohydrate-Binding Modules (CBMs). *P. fragariae* (PFRA), *P. rubi* (PRUB), *P. sojae* (PSOJ), *P. ramorum* (PRAM), *P. infestans* (PINF), *P. parasitica* (PPAR), *P. nicotianae* (PNIC), *Plasmopara halstedii* (PHAL).

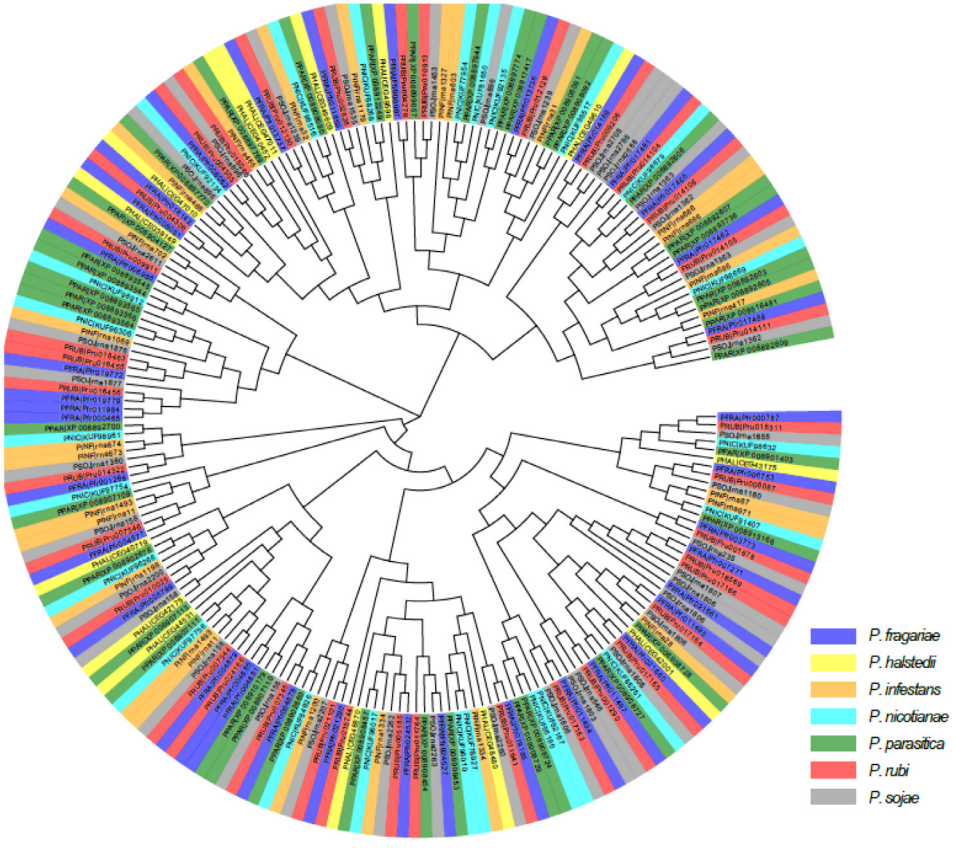


**Fig. 7.** Phylogenetic tree constructed based on the P450 genes in *P. fragariae, P. rubi* and other relative species, the genes of each species marked with a special color.

## Ethics Statement

This data does not include human studies and animal experiments, has not been published previously, that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder.

## CrediT Author Statement

**Rui-Fang Gao:** Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing - Original Draft **Jie-Yu Wang:** Formal analysis, Data Curation, **Ke-Wei Liu:** Investigation, Data Curation **Zhi-Wen Wang:** Software, Investigation, Data Curation **Diyang Zhang:** Investigation, Data Curation, Visualization **Xiang Zhao:** Software, Investigation, Data Curation **Wen-Ying Zhong:** Software, Investigation **Wen-Chieh Tsai:** Methodology, Supervision **Zhong-Jian Liu:** Conceptualization, Visualization, Supervision, Project administration **Gui-Ming Zhang:** Conceptualization, Resources, Visualization, Project administration

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

## Acknowledgments

This work was supported by National Key Technology Research and Development Programme of China (No. 2012BAK11B06) and National Key R&D Program of China (No. 2016YFF0203204) for G.-M. Z. National Key R&D Program of China (No. 2019YFD1000400), National Natural Science Foundation of China [No. 31870199] and the Key Laboratory of National Forestry and Grassland Administration for Orchid Conservation and Utilization Construction Funds of Fujian, China (Nos. 115/118990050 and 115/KJG18016A) for Z.-J.L.

## References

[1] R. Gao, J. Wang, K. Liu, K. Yoshida, Y. Hsiao, Y. Shi, K. Tsai, Y. Chen, N. Mitsuda, C. Liang, et al., Comparative analysis of *Phytophthora* genomes reveals oomycete pathogenesis in crops, Heliyon 7 (2021) e6317, doi:10.1016/j.heliyon.2021.e06317.
[2] A.E. Desjardins, T.M. Hohn, Mycotoxins in Plant Pathogenesis, Mol. Plant Microbe In. 10 (1997) 147–152, doi:10.1094/MPMI.1997.10.2.147.
[3] B.L. Wang, J. Liu, Z.D. Tian, B.T. Song, C.H. Xie, cDNA microarray analysis of metabolism-related genes in inoculated potato leaves expressing moderate quantitative resistance to *Phytophthora infestans*, J. Hortic. Sci. Biotech. 83 (2008) 419–426, doi:10.1080/14620316.2008.11512401.
[4] M.D. Ospina-Giraldo, J.G. Griffith, E.W. Laird, C. Mingora, The CAZyome of *Phytophthora* spp.: a comprehensive analysis of the gene complement coding for carbohydrate-active enzymes in species of the genus *Phytophthora*, BMC Genomics 11 (2010) 525, doi:10.1186/1471-2164-11-525.
[5] S. Kamoun, Groovy times: filamentous pathogen effectors revealed, Curr. Opin. Plant Biol. 10 (2007) 358–365, doi:10.1016/j.pbi.2007.04.017.
[6] W.A. Man In, T. Veld, Gene flow analysis demonstrates that *Phytophthora fragariae* var. *rubi* constitutes a distinct species, *Phytophthora rubi* comb, nov. Mycologia 99 (2007) 222–226, doi:10.1080/15572536.2007.11832581.
[7] S. Raffaele, S. Kamoun, Genome evolution in filamentous plant pathogens: why bigger can be better, Nat. Rev. Microbiol. 10 (2012) 417–430, doi:10.1038/nrmicro2790.
[8] D. Croll, B.A. Mcdonald, J. Heitman, The accessory genome as a cradle for adaptive evolution in pathogens, PLOS Pathog 8 (2012) e1002608, doi:10.1371/journal.ppat.1002608.

[9] C. Zhang, J. Wang, W. Xie, G. Zhou, L.Q. Zhang, Dynamic programming procedure for searching optimal models to estimate substitution rates based on the maximum-likelihood method, P. Natl. Acad. Sci. USA 108 (2011) 7860–7865, doi:10.1073/pnas.1018621108.

[10] M. Urban, R. Pant, A. Raghunath, A.G. Irvine, H. Pedro, K.E. Hammondkosack, The pathogen-host interactions database (PHI-base): additions and future developments, Nucleic Acids Res 43 (2015) D645, doi:10.1093/nar/gku1165.

[11] R.H. Jiang, S. Tripathy, F. Govers, B.M. Tyler, RXLR effector reservoir in two *Phytophthora* species is dominated by a single rapidly evolving superfamily with more than 700 members, P. Natl. Acad. Sci. USA 105 (2008) 4874–4879, doi:10.1073/pnas.0709303105.

[12] T.M. Amaro, G.J. Thilliez, G.B. Motion, E. Huitema, A perspective on CRN proteins in the genomics age: evolution, classification, delivery and function revisited, Front Plant Sci 8 (2017) 99, doi:10.3389/fpls.2017.00099.