

RESEARCH ARTICLE

Open Access



Comparative genome analysis of *Phyllosticta citricarpa* and *Phyllosticta capitalensis*, two fungi species that share the same host

Carolina Munari Rodrigues^{1†}, Marco Aurélio Takita^{1†}, Nicholas Vinicius Silva², Marcelo Ribeiro-Alves³ and Marcos Antonio Machado^{1*}

Abstract

Background: Citrus are among the most important crops in the world. However, there are many diseases that affect Citrus caused by different pathogens. Citrus also hosts many symbiotic microorganisms in a relationship that may be advantageous for both organisms. The fungi *Phyllosticta citricarpa*, responsible for citrus black spot, and *Phyllosticta capitalensis*, an endophytic species, are examples of closely related species with different behavior in citrus. Both species are always biologically associated and are morphologically very similar, and comparing their genomes could help understanding the different lifestyles. In this study, a comparison was carried to identify genetic differences that could help us to understand the biology of *P. citricarpa* and *P. capitalensis*.

Results: Drafts genomes were assembled with sizes close to 33 Mb for both fungi, carrying 15,206 and 14,797 coding sequences for *P. citricarpa* and *P. capitalensis*, respectively. Even though the functional categories of these coding sequences is similar, enrichment analysis showed that the pathogenic species presents growth and development genes that may be necessary for the pathogenicity of *P. citricarpa*. On the other hand, family expansion analyses showed the plasticity of the genome of these species. Particular families are expanded in the genome of an ancestor of *P. capitalensis* and a recent expansion can also be detected among this species. Additionally, evolution could be driven by environmental cues in *P. citricarpa*.

Conclusions: This work demonstrated genomic differences between *P. citricarpa* and *P. capitalensis*. Although the idea that these differences could explain the different lifestyles of these fungi, we were not able to confirm this hypothesis. Genome evolution seems to be of real importance among the *Phyllosticta* isolates and it is leading to different biological characteristics of these species.

Keywords: Pathogenic species, Endophytic species, Citrus black spot, Functional enrichment analysis, Phylogeny, Gene family expansion

* Correspondence: marcos@ccsm.br

[†]C. M. Rodrigues and M. A. Takita contributed equally to this work.

¹Laboratório de Biotecnologia, Centro APTA Citrus Sylvio Moreira, CP4, Cordeirópolis, SP 13490-970, Brazil

Full list of author information is available at the end of the article



Background

Phyllosticta citricarpa McAlpine (Synonym: *Guignardia citricarpa* Kiely) is the fungus responsible for causing citrus black spot (CBS), which is one of the most important diseases affecting the citrus industry worldwide [1, 2]. This disease was first described in Australia, where it caused considerable losses for sweet orange growers [3]. Since then, CBS has seriously affected citrus crops in countries in Africa, Asia, South America, and North America, especially in Argentina, United States and Brazil [1, 2, 4–6].

CBS affects almost all commercial varieties of citrus with the main symptom associated with this disease being the development of hard spot lesions in the fruit peel [7]. Advanced stages of the disease lead to maturation and early fall of the fruit [5]. In addition, the affected fruits have their appearance depreciated, making them unsuitable for the fresh fruit market, and therefore the costs associated with chemical control of the disease are significant [1, 7–9].

Most of the species in the genus *Phyllosticta* are plant pathogens of a wide range of hosts [10, 11] and although *P. citricarpa* is pathogenic, other endophytic and saprophytic species have also been reported for citrus [12–15]. *P. capitalensis* is among the endophytic species that lives within citrus and other hosts with a wide geographic distribution [10, 16, 17]. *P. capitalensis* is commonly found to be associated with lesions in plants; from an economic point of view, this association can be very negative since *P. capitalensis* is normally confused with the pathogen that is actually responsible for causing the disease. This is an even worse problem if the actual pathogen is a quarantine organism like *P. citricarpa* that has phytosanitary restrictions, being classified as quarantine A1 in the European Union and A2 in the United States [4, 18]. The similarity between *P. citricarpa* and *P. capitalensis* is so close that CBS was erroneously reported in New Zealand [2, 19], with the endophytic species being identified as the species responsible for causing the disease [20]. Due to the high similarity between these species, identifications can only be made at the molecular level [21, 22].

In recent years, the number of works using comparative genomic analysis to understand the genetic basis of the lifestyle of pathogenic and endophytic or symbiont organisms has grown considerably [23–26]. In this study, the genome of the citrus pathogenic species *P. citricarpa* was compared to the genome of *P. capitalensis*, the endophytic species, in order to identify genetic differences that could help understand their different lifestyles.

Results and discussion

Overview of genome sequencing and categorization

The Illumina sequencing of *P. citricarpa* and *P. capitalensis* genomes generated a total of 179,880,616 and

148,831,020 paired-end reads, respectively, with more than 90% showing Phred quality > 20 (Q20) for both species. The CG content was slightly lower for *P. citricarpa* (48.72) compared to *P. capitalensis* (51.43).

A de novo assembly was done for both genomes, generating drafts of 19,143 contigs in 32.6 Mb for *P. citricarpa* and N50 of 3049. The assembly of *P. capitalensis* reads resulted in 11,080 contigs and 33.2 Mb with a N50 of 4925. A Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the BioProject number PRJNA486917.

The contigs of both fungi were used for the prediction of coding sequences using *Botrytis cinerea*, an ascomycete necrotrophic plant pathogen with a broad host range [27] as a reference in Augustus [28]. For *P. citricarpa*, 15,206 proteins were identified, while for *P. capitalensis*, the total was 14,797. These proteins were annotated using blastp [29] to find similarities with subjects in the Protein database of the National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/protein/>) (Additional files 1 and 2). Further characterization of the putative proteins was performed in Blast2GO [30], and when they were distributed in different categories, i.e., cellular component, molecular function, and biological process (level 2).

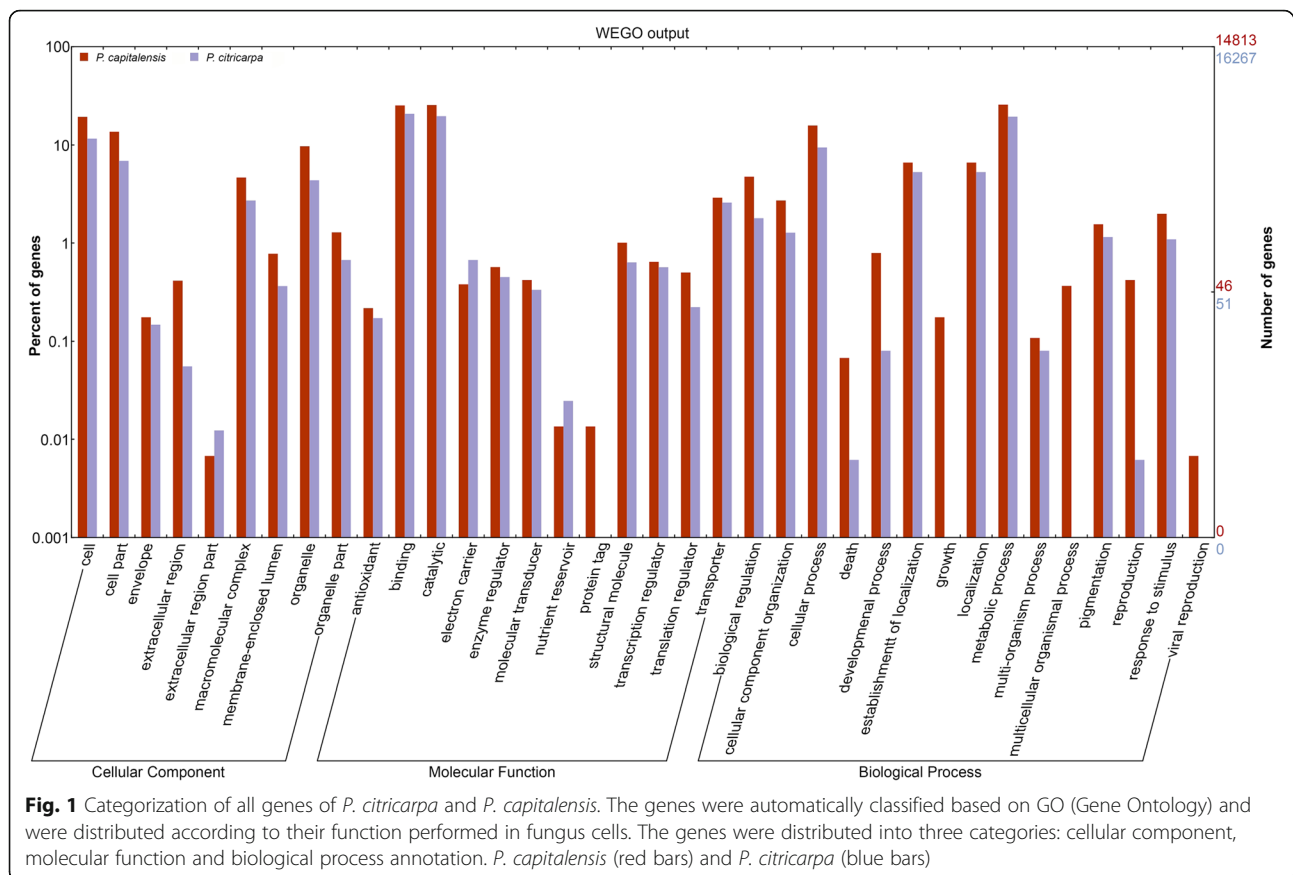
Even though the distribution of the genes of *P. citricarpa* and *P. capitalensis* showed the same pattern for most of the subcategories, with a few additional genes in the latter compared to the former, there were some differences that could be representative for understanding the behavior of both organisms (Fig. 1). In the molecular function category, the binding and catalytic subcategories presented the highest percentage of genes in both genomes while the protein tag subcategory was present only in *P. capitalensis* (Fig. 1).

For biological processes, the most representative subcategories were metabolic processes, cellular processes, establishment of localization, and localization. The growth, multicellular organismal processes and viral reproduction subcategories were unique for the endophytic fungus and in addition, it also showed a higher percentage of genes in the death and reproduction subcategories (Fig. 1).

BUSCO analysis confirms that the genomes are not complete with 502 and 765 proteins missing for *P. capitalensis* and *P. citricarpa*, respectively (Additional file 3) in a total of 3156 proteins. Other genomes of *Phyllosticta* spp. are much more complete and therefore were used for validating the results obtained with our genomes.

Phylogenetic analysis of *Phyllosticta* species

To verify the evolutionary relationship of different *Phyllosticta* species, we used the two genomes sequenced in



this work together with seven other genomes available in public databases (*P. capitalensis* strains: CBS128856, Gm33, and LGMF01; *P. citricarpa* strains: CBS141350, Gc12, and LGMF06, *P. citribraziliensis* strain CBS100098, and *P. citrichinaensis* strain CBS130529). The phylogenetic tree obtained using the 3185 single copy-ortholog genes (Fig. 2) confirms that *P. citricarpa* and *P. citriasiana* are very closely related and that *P. capitalensis* is distant from this group, confirming the result obtained with ITS, LSU, TEF1, ACT and GPDH sequences alignment [10, 31]. *P. capitalensis* and *P. citribraziliensis* are endophytes in *Citrus* while *P. citricarpa* and *P. citriasiana* are pathogenic. *P. citrichinaensis*, on the other hand, causes minor disease symptoms [32] and therefore, pathogenicity in *Citrus* may be related to the phylogeny of *Phyllosticta* species at least in part since the two pathogenic species cluster together.

Major functional categories enriched in *P. citricarpa* compared with *P. capitalensis*

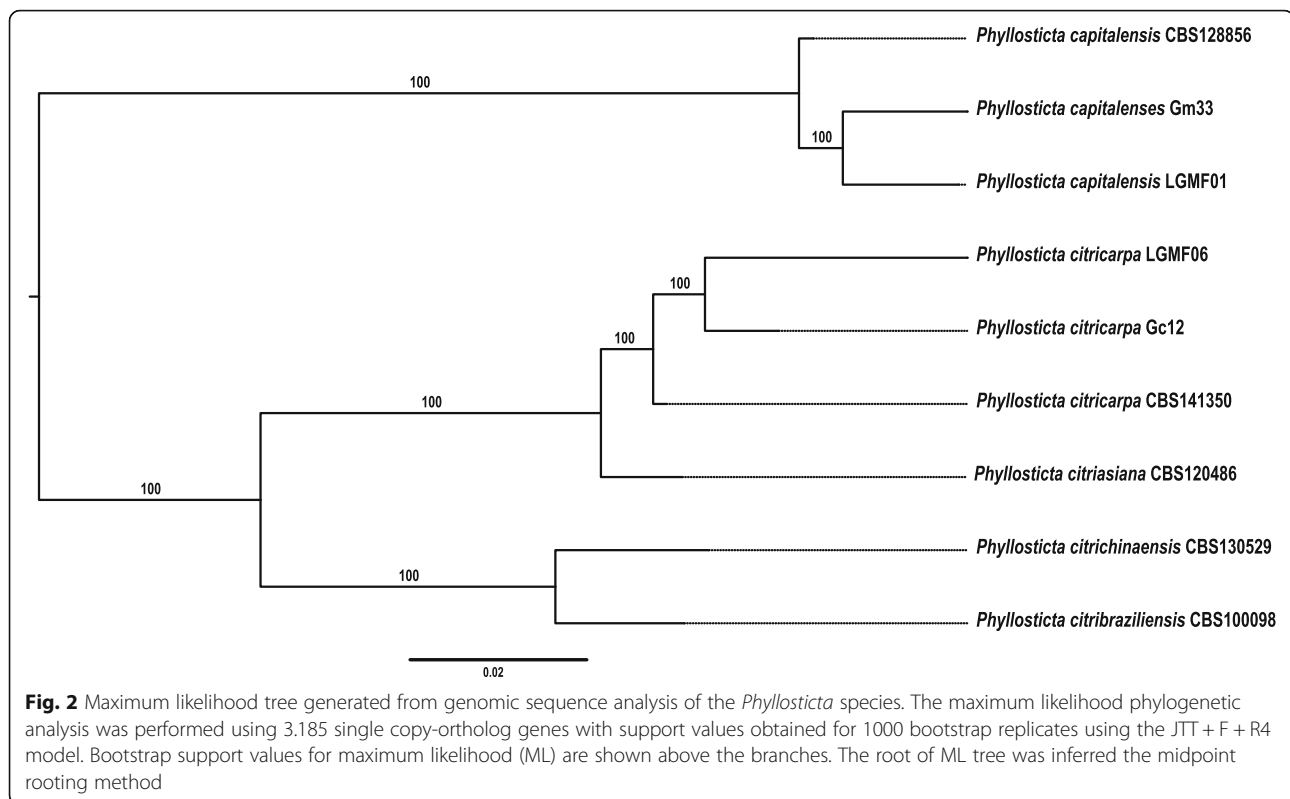
To better evaluate the differences at the genetic level between one pathogenic and one endophytic species, the sequences from *P. citricarpa* and *P. capitalensis* were used in a functional enrichment analysis to verify

potential molecular mechanisms associated with their interaction with citrus (Fig. 3).

It was observed that enrichment analysis resulted in a difference between *P. citricarpa* and *P. capitalensis*. For the Biological Process category, *P. citricarpa* showed a higher number of enriched subcategories, GO:0044848, GO:0051704, GO:0000003, and GO:0032502, which are all related with growth and development. This result may indicate that *P. citricarpa* presents a more complex living compared to *P. capitalensis*, which could result from the distant relationship in evolution. However, even though it is an appealing hypothesis for testing, it would be necessary to evaluate other sequenced genomes that show closer phylogenetic relationship to *P. capitalensis*. Other works reporting the colonization of *P. capitalensis* and *P. citricarpa* in *Citrus* are also missing, in particular those of global gene expression.

Gene families expansion in *Phyllosticta*

We also did an analysis of gene families expansion in the genomes of *Phyllosticta* species. This analysis was done based on the phylogeny with the five different species of *Phyllosticta* used in this work (Fig. 4). The analysis of gene family expansion shows a high



number of families being expanded from ancestral 5 in relation to 17, in the *P. capitalensis* clade (Additional file 4). This seems to be much more related to the species evolution than the lifestyle of the organisms since *P. citribraziliensis*, a *Citrus* endophyte, is not in the same clade of *P. capitalensis*. In addition, a family of HET-domain containing proteins is expanded in all the *P. capitalensis* analyzed, being a more recent event in evolution of the species. HET proteins were identified as important for vegetative incompatibility, in the formation of heterokaryon [33]. These proteins induce cell death avoiding the formation of heterokaryon [34]. Therefore this genomic data make us believe that *P. capitalensis* may be more effective in avoiding heterokaryon formation than the other *Phyllosticta* species.

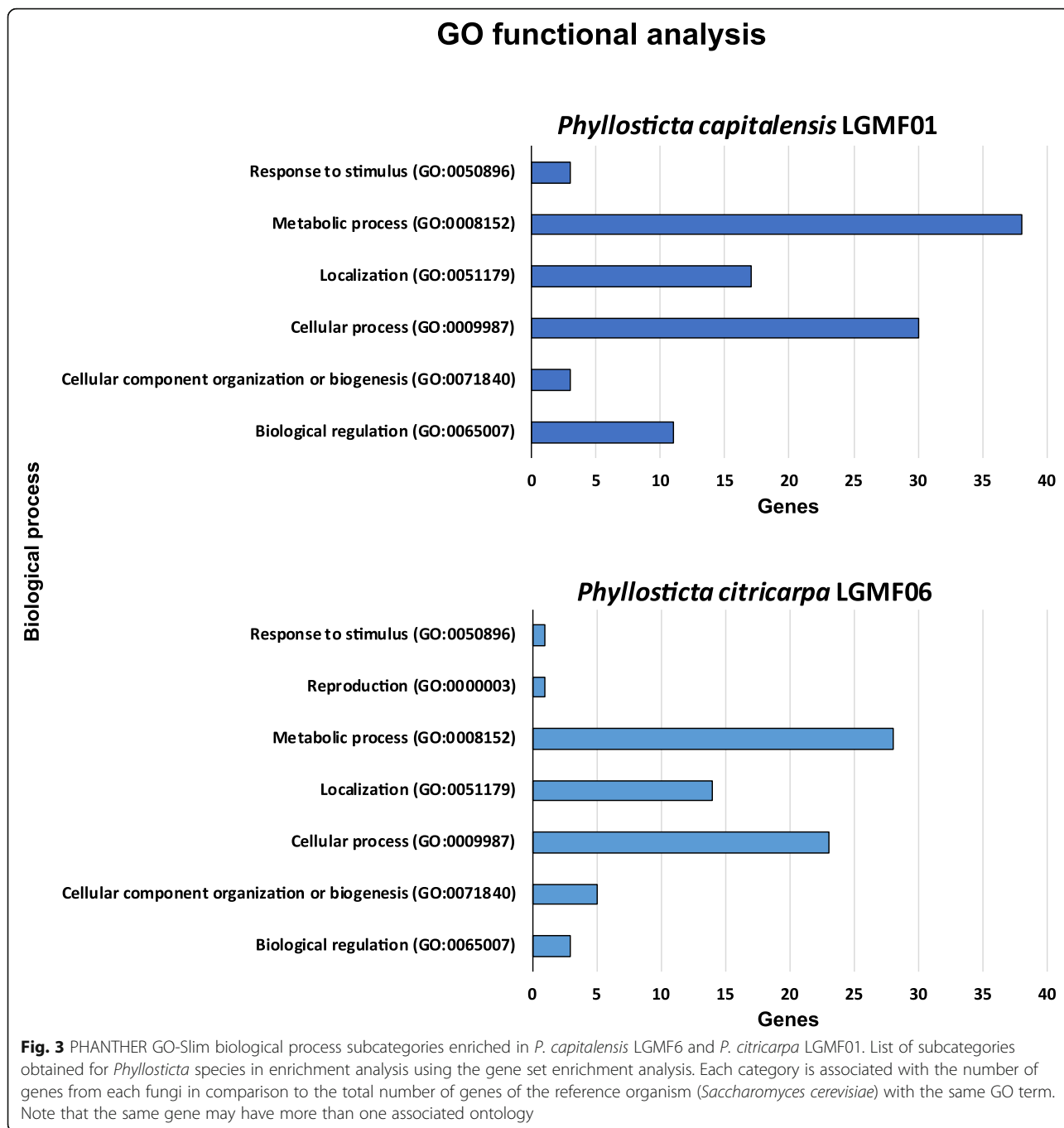
In the other clade, fewer families were expanded and besides, most of them encodes uncharacterized proteins (Fig. 4). However it is worth mentioning the expansion of a family of putative chitin recognition proteins that happens just in the *P. citricarpa* genome reported in this work (LGMF06). LysM domain proteins bind chitin and are important for fungus defense against the host response [35]. Because the expansion occurs just in the isolate sequenced in our lab, we suggest that the environment may have an importance in modulating the genome plasticity in *P. citricarpa*, as known for other filamentous fungi [36].

Comparative genomics

The phylogenetic analysis with all the genes confirmed that indeed *P. citricarpa* and *P. capitalensis* are distant species as previously reported. Therefore, the changes observed in their genomes may result from the evolution of the species, including gain and loss of genes. Even distant, a comparative genomics analysis is appealing because of the different lifestyle and it was carried out to identify genes exclusively found in the genomes of *P. citricarpa* LGMF06 and *P. capitalensis* LGMF01. A reciprocal blastp analysis was carried out using the protein sequences encoded in both genomes. The exclusive proteins were considered those that either had a positive hit with an E-value above 1×10^{-30} or resulted in a “no hit”. These proteins were reciprocally re-evaluated in blastp using the proteomes from *P. citricarpa* CBS141350 and *P. capitalensis* CBS128856 to make sure that the differences were not a problem resulting from the incompleteness of the genomes (Table 1; Additional files 5 and 6). These genes were further evaluated in order to better understand the pathogenicity of *P. citricarpa*.

Genes related to pathogenicity

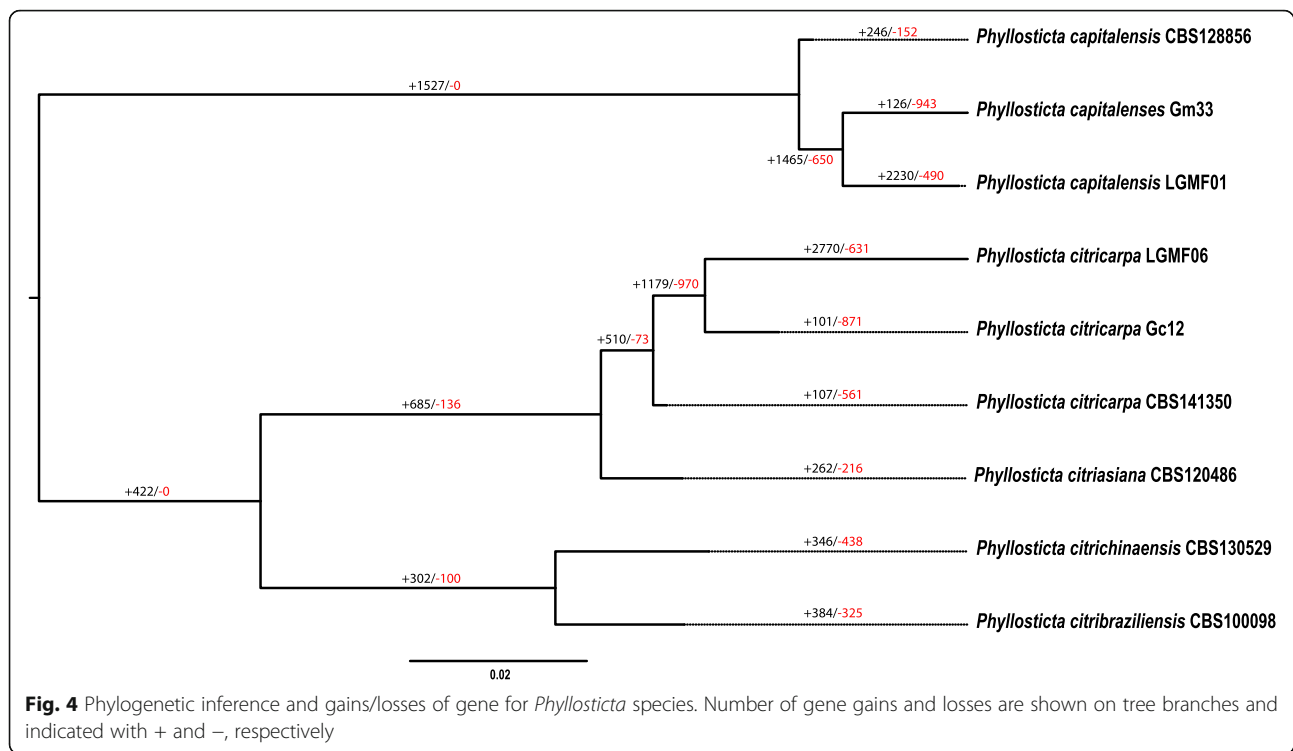
To identify putative genes involved in pathogenicity, we analyzed the *P. citricarpa* and *P. capitalensis*-exclusive genes for similarity with subjects in the pathogen-host interaction gene database (PHI-base).



Of the 2896 gene sequences of *P. citricarpa* (Additional file 5), 123 were identified as having similarity to putative PHI proteins (Additional file 7). For *P. capitalensis* this total was 210 out of 4164 exclusive genes (Additional files 6 and 8).

Among the 123 genes of *P. citricarpa* and 210 of *P. capitalensis*, 18 genes are common to both fungi (Fig. 5; Additional file 9). Therefore, 70 genes were unique in the pathogenic species, while 95 were exclusive to the citrus endophytic species (Fig. 5; Additional file 9).

Among the PHI exclusive genes of *P. citricarpa*, genes involved in fungus infection, toxin synthesis, effector protein secretion system (important for the pathogenicity of pathogenic organisms), signaling, detoxification systems, virulence factors and transcription factors were identified (Additional file 5). Because the role of these genes in pathogenicity and virulence is already known for other interactions, we assume that they are important in the development of CBS in citrus [37].



P. capitalensis also presents genes involved in fungal infection, cell wall degradation, growth, cell death regulation and signaling (Additional file 6). Although this fungus is an endophyte in citrus and a weak plant pathogen with a worldwide distribution presently known from 70 plant families [17], it is the causal agent of the leaf spot in the orchid *Bifrenaria harrisoniae* [14], which explains the presence of pathogenicity genes in the genome of this fungus.

The molecular components of *Phyllosticta* spp. pathogenicity in citrus are poorly known. In fact to our knowledge there is one report of identification of putative proteins involved in pathogenicity of *P. citricarpa* in citrus [38], in which the authors suggest a possible major role of pectinases for this organism. Therefore, this work opens new perspectives in understanding the pathogenicity of these *Phyllosticta* species.

Analysis of Carbohydrate-Active Enzymes (CAZymes) in *P. citricarpa* and *P. capitalensis*.

The function of CAZymes is breakdown, biosynthesis or modification of glycoconjugates, oligo- and polysaccharides.

Table 1 Number of exclusive proteins for *P. citricarpa* and *P. capitalensis* after analyses by blastp

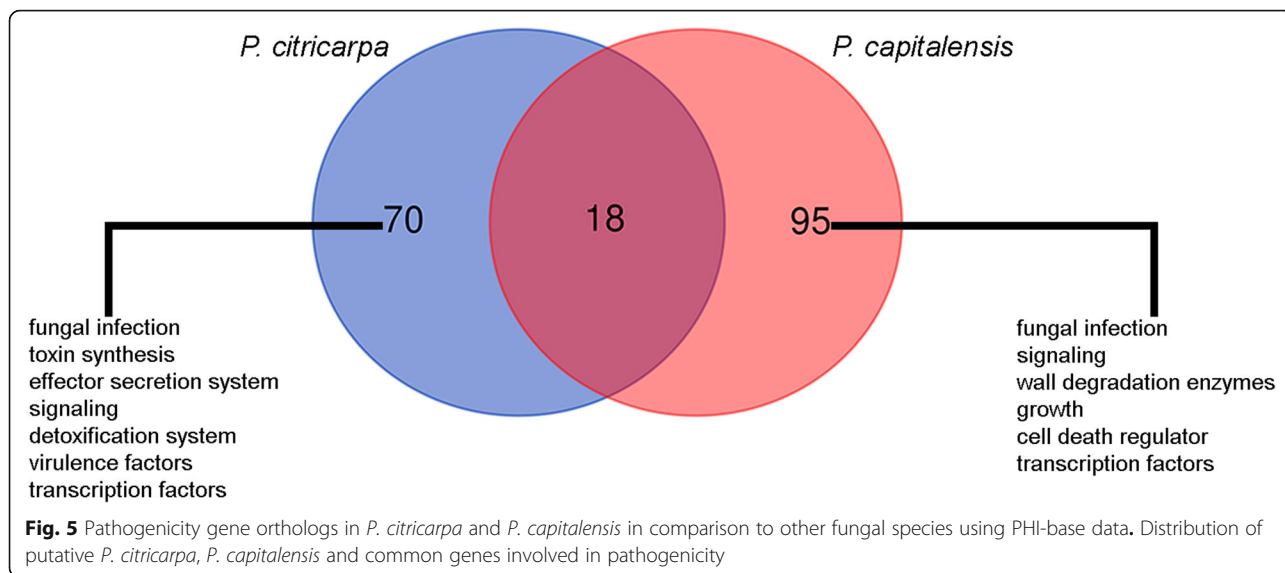
Total proteins of the genome		Unique proteins	
<i>P. citricarpa</i>	<i>P. capitalensis</i>	<i>P. citricarpa</i>	<i>P. capitalensis</i>
15206	14797	2896	4164

These enzymes are produced by phytopathogens and play a central role in the breakdown and synthesis of plant cell walls during host-pathogen interactions [39]. Due to the importance of CAZymes in fungal pathogenicity, we examined the presence of these enzymes in both fungi, comparing and evaluating the exclusive genes of each species.

The search for CAZymes revealed 23 in *P. citricarpa*, which were distributed across CBMs (carbohydrate-binding modules), CEs (carbohydrate esterases), GHs (glycoside hydrolases), GTs (glycosyl transferases) and AAs (auxiliary activities) families (Fig. 6, Table 2). It is also noted there was diversity within each family (and subfamily), mainly GTs and GHs (Table 2).

The diversity within the families in the pathogenic species was also higher in relation to endophytic species (Fig. 6, Table 2). *P. capitalensis* presents 44 genes coding for CAZymes distributed among CBM, CE, GH, GT and AA families, the latter being the only one to present more genes in this species (Fig. 6, Table 2).

We analyzed which of these CAZymes may be secreted by *P. citricarpa* and *P. capitalensis*. Of the 128 enzymes of the pathogenic species, 4 are possibly secreted: 1 of the GT family (34), and 3 of the GH family (16, 17, and 37) (Additional file 10). The endophytic species presented 1 GBM (63) secreted, 3 enzymes of the CE family (4, 5, and 6) and 5 proteins of the family GH (16, 18, 3, 43, and 55), totaling 9 CAZymes secreted (Additional file 10). The results show that *P. capitalensis* presents a greater amount of these enzymes in relation



to *P. citricarpa*, besides presenting a higher diversity of representatives of the different CAZymes families. Since these two species present high similarity and identification can only be made at the molecular level [21, 22], the divergency between the two species regarding CAZymes may be an important distinguishing characteristic, allowing the development of new markers for differencing the two species, like antibodies for instance.

Most of the enzymes secreted by *P. citricarpa* are from the GH family. These are involved in hydrolysis of glycosidic bonds between or within carbohydrate molecules [39].

The fact that CAZymes are responsible for the breakdown of cell wall components suggests they are strictly related to the successful infection process of the fungus. The number of CAZymes is higher in

pathogenic fungi [26]. These results are in agreement with those found by [38] where it was verified that *P. citricarpa* produces more endoglucanases in relation to *P. capitalensis*, in addition to producing more amylases and pectinases. The authors concluded that these differences could be associated with differences in pathogenicity for citrus plants. Similar to classical necrotrophic fungi, such as *Botrytis cinerea* and *Sclerotinia sclerotiorum*, the *P. citricarpa* genome is predicted to encode a large number of CAZymes involved in plant cell wall degradation [40, 41]. However, because *P. capitalensis* has more CAZymes in its genome compared to *P. citricarpa*, we could not do any association of the number of these enzymes with the lifestyle of the fungi. But the possibly

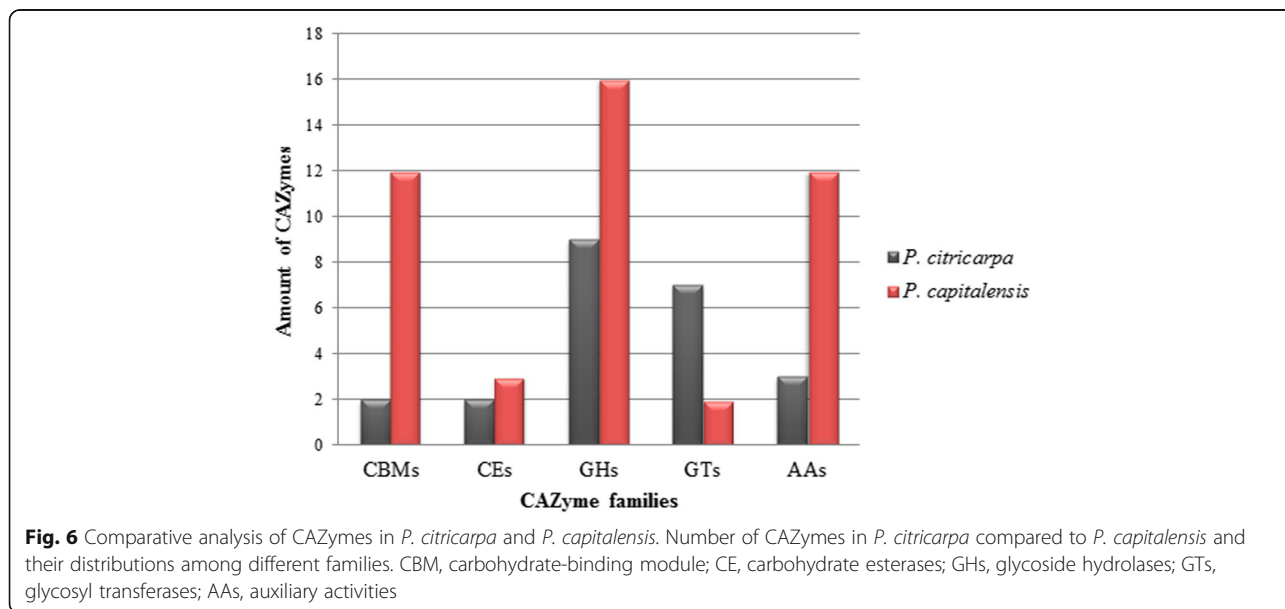


Table 2 Overview of CAZyme and number of genes in each CAZyme category. Color tones differentiate families from enzymes (Continued)

p.capi. LGMF01_12968	AA1_3
p.capi. LGMF01_0108	AA11
p.capi. LGMF01_10555	AA3_2
p.capi. LGMF01_10630	AA3_2
p.capi. LGMF01_10952	AA3_2
p.capi. LGMF01_11995	AA3_2
p.capi. LGMF01_4648	AA3_2
p.capi. LGMF01_7130	AA3_2
p.capi. LGMF01_9145	AA3_2
p.capi. LGMF01_9405	AA3_2
p.capi. LGMF01_13226	AA7
p.capi. LGMF01_12450	CBM1
p.capi. LGMF01_6291	CBM1
p.capi. LGMF01_8576	CBM1
p.capi. LGMF01_13173	CBM13
p.capi. LGMF01_9957	CBM13
p.capi. LGMF01_9957	CBM20
p.capi. LGMF01_13173	CBM32
p.capi. LGMF01_7886	CBM48
p.capi. LGMF01_10165	CBM50
p.capi. LGMF01_0072	CBM50
p.capi. LGMF01_13173	CBM57
p.capi. LGMF01_3297	CBM63
p.capi. LGMF01_14357	CE4
p.capi. LGMF01_1089	CE5
p.capi. LGMF01_9188	CE6
p.capi. LGMF01_6291	GH10
p.capi. LGMF01_13896	GH16
p.capi. LGMF01_8576	GH16
p.capi. LGMF01_11414	GH17
p.capi. LGMF01_6680	GH18
p.capi. LGMF01_9957	GH18
p.capi. LGMF01_9957	GH23
p.capi. LGMF01_12450	GH3
p.capi. LGMF01_13207	GH3
p.capi. LGMF01_9122	GH3
p.capi. LGMF01_9188	GH43_35
p.capi. LGMF01_3143	GH5_7
p.capi. LGMF01_10165	GH55
p.capi. LGMF01_10165	GH71
p.capi. LGMF01_3563	GH76
p.capi. LGMF01_5203	GT1
p.capi. LGMF01_8703	GT2

Table 2 Overview of CAZyme and number of genes in each CAZyme category. Color tones differentiate families from enzymes

ID	<i>P. citricarpa</i> (aa)	CAZymes Family
p.citric. LGMF06_1852		AA10
p.citric. LGMF06_14427		AA5
p.citric. LGMF06_1084		AA9
p.citric. LGMF06_1084		CBM1
p.citric. LGMF06_14427		CBM18
p.citric. LGMF06_14427		CE4
p.citric. LGMF06_13085		CE9
p.citric. LGMF06_2729		GH16
p.citric. LGMF06_0566		GH17
p.citric. LGMF06_14427		GH19
p.citric. LGMF06_12138		GH30_7
p.citric. LGMF06_14013		GH31
p.citric. LGMF06_12570		GH37
p.citric. LGMF06_1320		GH4
p.citric. LGMF06_13007		GH4
p.citric. LGMF06_6360		GH4
p.citric. LGMF06_10292		GT2
p.citric. LGMF06_1320		GT2
p.citric. LGMF06_13007		GT2
p.citric. LGMF06_15124		GT2
p.citric. LGMF06_4508		GT2
p.citric. LGMF06_6360		GT2
p.citric. LGMF06_3350		GT34
ID	<i>P. capitalensis</i> (aa)	CAZymes Family
p.capi. LGMF01_10165		AA5_1

secreted CAZymes of *P. citricarpa* is different from the ones from *P. capitalensis*, which could have an influence in the pathogenicity of the fungus.

All of our analyses could not give enough support to state that the different lifestyle could be explained by the genome differences observed between *P. citricarpa* and *P. capitalensis*. Indeed there is a debate on whether endophytes are pathogenic or not, or may turn pathogenic under certain conditions [42].

Conclusion

There are many different genomes sequenced for *Phyllosticta* species. In this work we presented the partial genomes of two isolates, one from *P. citricarpa* and another one from *P. capitalensis*, which present different lifestyles and are also distantly related in evolution. Different features were identified in each genome and could be used for understanding a little more about the biology on the different fungi. Nevertheless some of these

features seem to be related to the phylogeny of the group and it would be very interesting to have more genomes available from other closely related species in special to *P. capitalensis* to confirm these findings. However, our analyses showed interesting trends like the HET-domain containing proteins expansion in all the *P. capitalensis* strains or the expansion of a family of putative chitin recognition proteins in the *P. citricarpa* genome reported in this work (LGMF06). These findings seem to be related to the evolution of the species detected through the genome analyses but drivers are acting in the genome of the different species that could leads to the formation of new species in the future.

Methods

Fungal isolates, culture condition and DNA isolation

The fungi isolates used for genome sequencing are from the biological bank of the Laboratory of Microorganisms (LabGeM) at the Federal University of Paraná, Curitiba, Paraná, Brazil. The pathogenic species *P. citricarpa* (LGMF06) was isolated from CBS lesions in fruits from a plant grown in the State of São Paulo, Brazil, and tested for pathogenicity according to [43]. The endophytic species *P. capitalensis* (LGMF01) was isolated from asymptomatic fruits from a plant also grown in the State of São Paulo, Brazil. These fungi were grown in liquid Citrus Medium Fabris-Nishimura (CFN) [44] and incubated under agitation (65 rpm) at 25 °C for 7 days in the dark. The mycelium was removed from the medium, washed with distilled water, macerated under liquid nitrogen and stored at -80 °C for later use in the DNA extraction procedure [12].

Genome sequencing, assembly and annotation

A total of 10 µg of DNA from each fungus (*P. citricarpa* and *P. capitalensis*) was sent to Macrogen Inc. (South Korea) for sequencing using the HiSeq 2000 platform (Illumina Inc.). All procedures were performed according to Illumina's protocols. Paired-end fragments were generated with 101 base pairs (bp). One lane was used per library. The results were sent in fastq format. De novo assembly was carried out using CLC Genomics Workbench software (<https://www.qiagenbioinformatics.com/products/clc-genomics-workbench/>) with default parameters.

The prediction of open read frames (ORFs) was conducted using Augustus software v3.2 [32] with *Botrytis cinerea* (*Botryotinia fuckeliana*) as the model species. The annotation and categorization of ORFs were automatically carried out using Blast2GO with an E-value cut-off $\leq 10^{-6}$ [34]. The tool called WEGO (Web Gene Ontology Annotation Plot) was used for plotting GO annotation GO [45]. The completeness of the gene

prediction was assessed using BUSCO v3 [46] software package and pezizomycotina_odb9 gene set.

Phylogenomic reconstruction

To reconstruct a phylogenetic hypothesis for nine *Phyllosticta* species (*P. capitalensis* strains: CBS128856, Gm33, LGMF01; *P. citricarpa* strains: CBS141350, Gc12, LGMF06, *P. citribraziliensis* strain CBS100098, *P. citrichinaensis* strain CBS130529), in total, we used 3.185 single copy-ortholog genes identified with OrthoFinder [47]. Proteins sequences were aligned using MAFFT v7.271 [48] and trimmed using TrimAL v1.4 [49]. The Maximum Likelihood (ML) phylogenetic analysis was performed using IQtree v1.6.5 [50], with support was assessed via 1000 bootstrap replicates using the JTT + F + R4 model. The model of nucleotide sequence evolution was inferred using software ModelFinder [51] with both the Akaike information criterion (AIC) [52]. The root of ML tree was inferred the midpoint rooting method [53].

Gene set enrichment analysis

To realize a functional enrichment analysis of *P. citricarpa* and *P. capitalensis*, protein sequences was submitted to a search of orthologous groups based on reciprocal best hits between each fungus and the yeast reference organism *Saccharomyces cerevisiae*. The search of orthologous groups was performed using the Orthofinder v2.2.7 software. Subsequently, we have selected groups present only in the genomes of *P. capitalensis*/*Saccharomyces* and *P. citricarpa*/*Saccharomyces*. The set of genes retrieved was submitted for GO enrichment analysis in the PANTHER classification system [54] and *Saccharomyces* Genome database (SGD) [55].

Expansion and retraction of gene families

We used the maximum-likelihood approach implemented in BadiRate v1.35 [56] to estimate the gene birth, death, and innovation rates in the *Phyllosticta* gene families during the evolution. In brief, we first inferred orthologous groups based on reciprocal best hits within and between gene families of nine *Phyllosticta* genomes using the Orthofinder v2.2.7 software [47]. Single-copy orthologs and lineage-specific gene families were extracted using in-house Perl script. The ML tree previously inferred from the 3.185 single copy orthologs and orthologous gene groups were utilized by the program as input and with the following parameters: -anc 1 -bmodel FR -model BDI -ep ML. For each orthologous group, gene gain and loss events were counted from the number of members at internal nodes inferred by maximum likelihood under the BDI stochastic model [57], assuming that each branch has its own specific turnover rates.

Comparative genomics

Comparative genomics analysis between *P. citricarpa* and *P. capitalensis* were conducted using only exclusive sequences for each species. These sequences were obtained after different rounds of BLAST [29]. Initially, a blastp analysis was carried out with unique coding sequences extracted from the genome of each organism using Augustus against the proteins from the genomes of the other species. The results were filtered in order to just get hits presenting E values $>1e-50$. These subsets of coding sequences were then evaluated reciprocally against proteomes available at the JGI (*Phyllosticta citricarpa*, CBS 127454 v 1.0; *Phyllosticta capitalensis*, CBS 128856, v 1.0) using blastp. Once again an E value filter ($1e-50$) was used and only the sequences showing results below the cutoff were kept for further analyses, being considered as the exclusive sets for each species.

PHI-base analyzes of the exclusive genes of *P. citricarpa* and *P. capitalensis*

The *P. citricarpa* and *P. capitalensis*, exclusive gene sequences for each, were used to investigate whether they had been verified to be pathogenic genes using annotations from the PHI-base using blastp (e-value <0.001) [58].

Analysis of carbohydrate-active (CAZy) enzymes in *P. citricarpa* and *P. capitalensis*

The CAZymes are involved in plant cell wall degradation. For this reason, the genes of both fungi, pathogenic and endophytic, were subjected to CAZY annotation using dbCAN [59], which is based on the CAZY (Carbohydrate-Active Enzyme) database classification [39].

Secretome

To verify which proteins are secreted the following pipeline was used: the secreted proteins were identified using SignalP 3.0 [60], which detects the presence and location of signal peptide cleavage sites in proteins. In addition, we also used SecretomeP 1.0f [61] to predict protein secretion by nonclassical pathways. Afterwards, TMHMM [62] was used to identify the sequences that have transmembrane domains that were excluded from further analyses. Therefore, the sequences presenting signal peptide and no transmembrane domains were selected. These sequences were then submitted to Cello 2.5 [63] analysis to predict cellular localization of the sequences.

Additional files

Additional file 1: Sequences of *P. citricarpa* proteins. The proteins obtained after de novo assembly of *P. citricarpa* genome were used for

coding sequences prediction using *Botrytis cinerea* as reference in Augustus, when 15,206 proteins were obtained. (TXT 5151 kb)

Additional file 2: Sequences of *P. capitalensis* transcripts. The proteins obtained after de novo assembly of *P. capitalensis* were used for proteins prediction using *Botrytis cinerea* as reference in Augustus, when 14,797 coding sequences were obtained. (TXT 5895 kb)

Additional file 3: BUSCO assessment of genomes from Phyllosticta. The genome-level benchmarking value of *P. capitalensis* LGMF01 was C: 65.2% (containing S: 65.1%, D: 0.1%, F: 23.9%, M: 10.9%, n: 3156) and *P. citricarpa* LGMF06 was C: 46.9% (containing: S: 46.9%, D: 0.0%, F: 33.7%, M: 19.4%, n: 3.156). The corresponding protein-level benchmarking value was C: 57.5% (including S: 57.4%, D: 0.1%, F: 26.5%, M: 16.0%, n: 3.156) and C: 40.0% (including S: 40.0%, D: 0.0%, F: 35.7%, M: 24.3%, n: 3156). Light-blue: complete (C) and single-copy (S) genes; dark-blue: complete and duplicated genes (D); yellow: fragmented genes (F); red: missing genes (M) and n: total BUSCO groups for searching. *P. citriasiatica* was used with genome control. (PDF 592 kb)

Additional file 4: Comparative analysis of gene gain and loss rates. Comparative analysis of gene gain and loss rates is represented in relation to different branches from the pathogenic and endophytic Phyllosticta species. (XLSX 41 kb)

Additional file 5: Unique sequences of proteins of *P. citricarpa*. Unique proteins identified after comparison of the *P. citricarpa* and *P. capitalensis* sequences by Blastp. (FA 654 kb)

Additional file 6: Unique proteins of genes of *P. capitalensis*. Unique proteins identified after comparison of the *P. capitalensis* and *P. citricarpa* sequences by Blastp. (FA 992 kb)

Additional file 7: Pathogenicity proteins of *P. citricarpa*. List of *P. citricarpa* exclusive proteins displaying similarity with subjects in the pathogen-host interaction gene database (PHI-base). (XLSX 18 kb)

Additional file 8: Pathogenicity proteins of *P. capitalensis*. List of *P. capitalensis* exclusive proteins displaying similarity with subjects in the pathogen-host interaction gene database (PHI-base). (XLSX 22 kb)

Additional file 9: Comparison of the pathogenicity proteins of *P. citricarpa* and *P. capitalensis*. List of exclusive and common pathogenicity proteins identified after comparison of *P. citricarpa* and *P. capitalensis*. (XLSX 18 kb)

Additional file 10: Secreted CAZymes by *P. citricarpa* and *P. capitalensis*. List of secreted CAZymes identified among the exclusive proteins of each species. (XLSX 9 kb)

Abbreviations

AA: Auxillary activities; ABC: ATP-binding cassette; acetyl-CoA: Acetyl-coenzyme A; CAZymes: Carbohydrate-active enzymes; CBMs: Carbohydrate-binding modules; CBS: Citrus black spot; CE: Carbohydrate esterases; CFN: Citrus medium fabris-nishimura; GHs: Glycoside hydrolases; GO: Gene ontology; GTs: Glycosyl transferases; MFS: The major facilitator superfamily; ORFs: Open read frames; PHI-base: Pathogen-host interaction gene database; ROS: Reactive oxygen species; WEGO: Web gene ontology annotation plot

Acknowledgements

The authors thank Inaiara de Souza Pacheco for performing the DNA extraction of the fungi *P. citricarpa* and *P. capitalensis* for the execution of this work. We also thank Dr. Chirlei Glienke for ceding the isolates of these fungi to this work.

Authors' contributions

MAM planned and supervised the study. CMR and MAT contributed to the design and execution of the analysis. CMR conducted collection, processing of samples, functional analysis and drafted the manuscript. CMR, NVS, MRA and MAT contributed to the genomic analysis, categorization and annotations of the sequences of *P. citricarpa* and *P. capitalensis*. CMR, NVS, MRA, MAT and MAM contributed to the interpretation of the data and provided intellectual input. CMR, NVS, MRA, MAT and MAM revised the manuscript. All authors read and approved the final manuscript.

Funding

The present study was conducted with the financial support of Conselho Nacional de Pesquisa de Desenvolvimento (CNPq) (560443/2010–2 and 163021/2011–2), which provided a scholarship to CMR and financial support and Instituto Nacional de Ciência e Tecnologia (INCT) de Genômica para Melhoramento de Citros, with the financial support of FAPESP (08/57909–2 and 14/50880–0) and Conselho Nacional de Pesquisa de Desenvolvimento (CNPq) (573848/2008–4 and 465440/2014–2). This study was also financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. MAM and MAT are CNPq fellows. These funding agencies played no role in the design of the study, in the collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

This Whole Genome Shotgun project has been deposited under Bioproject accession PRJNA486917.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Laboratório de Biotecnologia, Centro APTA Citros Sylvio Moreira, CP4, Cordeirópolis, SP 13490-970, Brazil. ²Laboratório de Genômica e BioEnergia, Universidade Estadual de Campinas, Campinas, SP 13083-970, Brazil. ³Laboratório de Pesquisa Clínica em DST/AIDS, Instituto Nacional de Infectologia Evandro Chagas, Fundação Oswaldo Cruz, Rio de Janeiro, RJ 21040900, Brazil.

Received: 14 September 2018 Accepted: 17 June 2019

Published online: 05 July 2019

References

- Kotzé JM. Epidemiology and control of Citrus black spot in South Africa. *Plant Dis.* 1981;65:945. <https://doi.org/10.1094/PD-65-945>.
- Yonow T, Hattingh V, de Villiers M. CLIMEX modelling of the potential global distribution of the citrus black spot disease caused by *Guignardia citricarpa* and the risk posed to Europe. *Crop Prot.* 2013;44:18–28. <https://doi.org/10.1016/J.CROPRO.2012.10.006>.
- Benson AH. Black spot of the orange. *Agric Gaz NSW.* 1895;4:249–52.
- Schubert TS, Dewdney MM, Peres NA, Palm ME, Jeyaprakash A, Sutton B, et al. First report of *Guignardia citricarpa* associated with Citrus black spot on sweet Orange (*Citrus sinensis*) in North America. *Plant Dis.* 2012;96:1225. <https://doi.org/10.1094/PDIS-01-12-0101-PDN>.
- Brentu FC, Oduro KA, Offei SK, Odamtten GT, Vicent A, Peres NA, et al. Crop loss, aetiology, and epidemiology of citrus black spot in Ghana. *Eur J Plant Pathol.* 2012;133:657–70. <https://doi.org/10.1007/s10658-012-9944-1>.
- Wickert E, de Goes A, de Souza A, Lemos EG de M. Genetic diversity and population differentiation of the causal agent of Citrus black spot in Brazil. *Sci World J.* 2012;2012:1–14. <https://doi.org/10.1100/2012/368286>.
- Kotzé JM. Black spot. In: Timmer LW, Garnsey SM, Graham J, editors. *Compendium of Citrus diseases*; 2000. p. 23–5.
- Martínez-Minaya J, Conesa D, López-Quílez A, Vicent A. Climatic distribution of citrus black spot caused by *Phyllosticta citricarpa*. A historical analysis of disease spread in South Africa. *Eur J Plant Pathol.* 2015;143:69–83. <https://doi.org/10.1007/s10658-015-0666-z>.
- Silva Junior GJ, Scapim M d S, Silva FP, Silva ARP, Behlau F, Ramos HH. Spray volume and fungicide rates for citrus black spot control based on tree canopy volume. *Crop Prot.* 2016;85:38–45. <https://doi.org/10.1016/J.CROPRO.2016.03.014>.
- Glienke C, Pereira OL, Stringari D, Fabris J, Kava-Cordeiro V, Galli-Terasawa L, et al. Endophytic and pathogenic *Phyllosticta* species, with reference to those associated with Citrus black spot. *Persoonia.* 2011;26:47–56. <https://doi.org/10.3767/003158511X569169>.
- Wong MH, Crous PW, Henderson J, Groenewald JZ, Drenth A. *Phyllosticta* species associated with freckle disease of banana. *Fungal Divers.* 2012;56(1): 173–87. <https://doi.org/10.1007/s13225-012-0182-9>.
- Glienke-Blanco C, Aguilar-Vildoso CI, Vieira MLC, Barroso PAV, Azevedo JL. Genetic variability in the endophytic fungus *Guignardia citricarpa* isolated from citrus plants. *Genet Mol Biol.* 2002;25:251–5. <https://doi.org/10.1590/S1415-47572002000200021>.
- Okane I, Nakagiri A, Ito T, Lumyong S. Extensive host range of an endophytic fungus, *Guignardia endophyllicola* (anamorph: *Phyllosticta capitalensis*). *Mycoscience.* 2003;44:353–63. <https://doi.org/10.1007/S10267-003-0128-X>.
- Silva M, Pereira OL, Braga IF, Lelis SM. Leaf and pseudobulb diseases on *Bifrenaria harrissoniae* (Orchidaceae) caused by *Phyllosticta capitalensis* in Brazil. *Aust Plant Dis Notes.* 2008;3:53–6. <https://doi.org/10.1007/bf03211238>.
- Huang WY, Cai YZ, Hyde KD, Corke HSM. Biodiversity of endophytic fungi associated with 29 traditional Chinese medicinal plants. *Fungal Divers.* 2008; 33:61–75.
- Baayen RP, Bonants PJM, Verkley G, Carroll GC, van der Aa HA, de Weerd M, et al. Nonpathogenic isolates of the Citrus black spot fungus, *Guignardia citricarpa*, identified as a cosmopolitan endophyte of Woody plants, *G. mangiferae* (*Phyllosticta capitalensis*). *Phytopathology.* 2002;92:464–77. <https://doi.org/10.1094/PHYTO.2002.92.5.464>.
- Wikee S, Lombard L, Crous PW, Nakashima C, Motohashi K, Chukeatirote E, et al. *Phyllosticta capitalensis*, a widespread endophyte of plants. *Fungal Divers.* 2013;60:91–105. <https://doi.org/10.1007/s13225-013-0235-8>.
- EC. Council directive 2000/29/EC of 8 May 2000 on protective measures against the introduction into the community of organisms harmful to plants and plant products and against their spread within the community. *Off J Eur Communities L.* 2000;169:1–112.
- Paul I, van Jaarsveld AS, Korsten L, Hattingh V. The potential global geographical distribution of Citrus black spot caused by *Guignardia citricarpa* (Kiely): likelihood of disease establishment in the European Union. *Crop Prot.* 2005;24:297–308. <https://doi.org/10.1016/J.CROPRO.2004.08.003>.
- Everett KR, Rees-George J. Reclassification of an isolate of *Guignardia citricarpa* from New Zealand as *Guignardia mangiferae* by sequence analysis. *Plant Pathol.* 2006;55:194–9. <https://doi.org/10.1111/j.1365-3059.2006.01334.x>.
- Bonants PJM, Carroll GC, de Weerd M, van Brouwershaven IR, Baayen RP. Development and validation of a fast PCR-based detection method for pathogenic isolates of the Citrus black spot fungus, *Guignardia citricarpa*. *Eur J Plant Pathol.* 2003;109:503–13. <https://doi.org/10.1023/A:1024219629669>.
- van Gent-Pelzer MPE, van Brouwershaven IR, Kox LFF, Bonants PJM. A TaqMan PCR method for routine diagnosis of the quarantine fungus *Guignardia citricarpa* on Citrus fruit. *J Phytopathol.* 2007;155:357–63. <https://doi.org/10.1111/j.1439-0434.2007.01244.x>.
- Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature.* 2010;464:367–73. <https://doi.org/10.1038/nature08850>.
- de Wit PJGM, van der Burgt A, Ökmen B, Stergiopoulos I, Abd-El Salam KA, Aerts AL, et al. The genomes of the fungal plant pathogens *Cladosporium fulvum* and *Dothistroma septosporum* reveal adaptation to different hosts and lifestyles but also signatures of common ancestry. *PLoS Genet.* 2012;8: e1003088. <https://doi.org/10.1371/journal.pgen.1003088>.
- Zaluga J, Stragier P, Baeyen S, Haegeman A, Van Vaerenbergh J, Maes M, et al. Comparative genome analysis of pathogenic and non-pathogenic *Clavibacter* strains reveals adaptations to their lifestyle. *BMC Genomics.* 2014;15:392. <https://doi.org/10.1186/1471-2164-15-392>.
- Zhao Z, Liu H, Wang C, Xu J-R. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics.* 2013;14:274. <https://doi.org/10.1186/1471-2164-14-274>.
- WILLIAMSON B, TUDZYNSKI B, TUDZYNSKI P, JAL VANKAN. *Botrytis cinerea*: the cause of grey mould disease. *Mol Plant Pathol.* 2007;8:561–80. <https://doi.org/10.1111/j.1364-3703.2007.00417.x>.
- Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 2005; 33(Web Server):W465–7. <https://doi.org/10.1093/nar/gki458>.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional

- genomics research. *Bioinformatics*. 2005;21:3674–6. <https://doi.org/10.1093/bioinformatics/bti610>.
31. Wiike S, Lombard L, Nakashima C, Motohashi K, Chuakeatirote E, Cheewangkoon R, et al. A phylogenetic re-evaluation of *Phyllosticta* (Botryosphaerales). *Stud Mycol*. 2013;76:1–29. <https://doi.org/10.3114/SIM0019>.
 32. Wang X, Chen G, Huang F, Zhang J, Hyde KD, Li H. *Phyllosticta* species associated with citrus diseases in China. *Fungal Divers*. 2012;52:209–24. <https://doi.org/10.1007/s13225-011-0140-y>.
 33. Bégueret J, Turcq B, Clavé C. Vegetative incompatibility in filamentous fungi: het genes begin to talk. *Trends Genet*. 1994;10:441–6. [https://doi.org/10.1016/0168-9525\(94\)90115-5](https://doi.org/10.1016/0168-9525(94)90115-5).
 34. Paoletti M, Clavé C. The fungus-specific HET domain mediates programmed cell death in *Podospora anserina*. *Eukaryot Cell*. 2007;6:2001–8. <https://doi.org/10.1128/EC.00129-07>.
 35. Cen K, Li B, Lu Y, Zhang S, Wang C. Divergent LysM effectors contribute to the virulence of *Beauveria bassiana* by evasion of insect immune defenses. *PLoS Pathog*. 2017;13:e1006604. <https://doi.org/10.1371/journal.ppat.1006604>.
 36. Dong Y, Li Y, Qi Z, Zheng X, Zhang Z. Genome plasticity in filamentous plant pathogens contributes to the emergence of novel effectors and their cellular processes in the host. *Curr Genet*. 2016;6:247–51. <https://doi.org/10.1007/s00294-015-0509-7>.
 37. Baldwin TK, Winnenburger R, Urban M, Rawlings C, Koehler J, Hammond-Kosack KE. The pathogen-host interactions database (PHI-base) provides insights into generic and novel themes of pathogenicity. *Mol Plant Microbe Interact*. 2006;19:1451–62. <https://doi.org/10.1094/MPMI-19-1451>.
 38. Romão AS, Spósito MB, Andreote FD, Azevedo JL, Araújo WL. Enzymatic differences between the endophyte *Guignardia mangiferae* (Botryosphaerales) and the citrus pathogen *G. citricarpa*. *Genet Mol Res*. 2011;10:243–52. <https://doi.org/10.4238/vol10-1gmr952>.
 39. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The carbohydrate-active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res*. 2009;37(Database):D233–8. <https://doi.org/10.1093/nar/gkn663>.
 40. Heard S, Brown NA, Hammond-Kosack K. An interspecies comparative analysis of the predicted Secretomes of the necrotrophic plant pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS One*. 2015;10:e0130534. <https://doi.org/10.1371/journal.pone.0130534>.
 41. Brown NA, Antoniw J, Hammond-Kosack KE. The predicted Secretome of the plant pathogenic fungus *Fusarium graminearum*: a refined comparative analysis. *PLoS One*. 2012;7:e33731. <https://doi.org/10.1371/journal.pone.0033731>.
 42. Fesel PH, Zuccaro A. Dissecting endophytic lifestyle along the parasitism/mutualism continuum in *Arabidopsis*. *Curr Opin Microbiol*. 2016;32:103–12. <https://doi.org/10.1016/j.mib.2016.05.008>.
 43. Baldassari RB, Wickert E, de Goes A. Pathogenicity, colony morphology and diversity of isolates of *Guignardia citricarpa* and *G. mangiferae* isolated from Citrus spp. *Eur J Plant Pathol*. 2008;120:103–10. <https://doi.org/10.1007/s10658-007-9182-0>.
 44. Figueiredo JG, Goulin EH, Tanaka F, Stringari D, Kava-Cordeiro V, Galli-Terasawa LV, et al. Agrobacterium tumefaciens-mediated transformation of *Guignardia citricarpa*. *J Microbiol Methods*. 2010;80:143–7. <https://doi.org/10.1016/j.mimet.2009.11.014>.
 45. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, et al. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res*. 2006;34(Web Server):W293–7. <https://doi.org/10.1093/nar/gkl031>.
 46. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31:3210–2. <https://doi.org/10.1093/bioinformatics/btv351>.
 47. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol*. 2015;16:157. <https://doi.org/10.1186/s13059-015-0721-2>.
 48. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80. <https://doi.org/10.1093/molbev/mst010>.
 49. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25:1972–3. <https://doi.org/10.1093/bioinformatics/btp348>.
 50. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32:268–74. <https://doi.org/10.1093/molbev/msu300>.
 51. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14:587–9. <https://doi.org/10.1038/nmeth.4285>.
 52. Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control*. 1974;19:716–23. <https://doi.org/10.1109/TAC.1974.1100705>.
 53. Farris JS. Estimating phylogenetic trees from distance matrices. *Am Nat*. 1972;106:645–68. <https://doi.org/10.1086/282802>.
 54. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, et al. PANTHER version 11: expanded annotation data from gene ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res*. 2017;45:D183–9. <https://doi.org/10.1093/nar/gkw1138>.
 55. Cherry J, Adler C, Ball C, Chervitz SA, Dwight SS, Hester ET, et al. SGD: *Saccharomyces* genome database. *Nucleic Acids Res*. 1998;26:73–9. <https://doi.org/10.1093/nar/26.1.73>.
 56. Librado P, Vieira FG, Rozas J. BadiRate: estimating family turnover rates by likelihood-based methods. *Bioinformatics*. 2012; 28(2):279–81. <https://doi.org/10.1093/bioinformatics/btr623>.
 57. Hahn MW, De Bie T, Stajich JE, Nguyen C, Cristianini N. Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res*. 2005;15:1153–60. <https://doi.org/10.1101/gr.3567505>.
 58. Winnenburger R, Urban M, Beacham A, Baldwin TK, Holland S, Lindeberg M, et al. PHI-base update: additions to the pathogen host interaction database. *Nucleic Acids Res*. 2007;36(Database):D572–6. <https://doi.org/10.1093/nar/gkm858>.
 59. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 2012; 40:W445–51. <https://doi.org/10.1093/nar/gks479>.
 60. Dyrlov Bendtsen J, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol*. 2004;340:783–95. <https://doi.org/10.1016/j.jmb.2004.05.028>.
 61. Bendtsen JD, Jensen LJ, Blom N, von Heijne G, Brunak S. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng Des Sel*. 2004;17:349–56. <https://doi.org/10.1093/protein/gzh037>.
 62. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J Mol Biol*. 2001;305:567–80. <https://doi.org/10.1006/JMBI.2000.4315>.
 63. Yu C, Lin C, Hwang J. Predicting subcellular localization of proteins for gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci*. 2004;13:1402–6. <https://doi.org/10.1110/PS.03479604>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

