Computational analysis of the Plasmodiophora brassicae genome: mitochondrial sequence description and metabolic pathway database design

Stéphanie Daval, Arnaud Belcour, Kévin Gazengel, Ludovic Legrand, Jerome Gouzy, Ludovic Cottret, Lionel Lebreton, Yoann Aigu, Christophe Mougel, Maria M. Manzanares-Dauleux

Computational analysis of the *Plasmodiophora brassicae* genome: mitochondrial sequence description and metabolic pathway database design

Daval S.[1a], Belcour A.[1], Gazengel K.[1], Legrand L.[2], Gouzy J.[2], Cottret L.[2], Lebreton L.[1], Aigu Y.[1], Mougel C.[1], Manzanares-Dauleux M.J.[1]

[1] IGEPP, INRA, AGROCAMPUS OUEST, Université Rennes, Domaine de la Motte, Le Rheu, F-35653, France

[2] LIPM, Université de Toulouse, INRA, CNRS, Castanet-Tolosan, France

Key words: clubroot; mitochondrial genome; ClubrootCyc; spermidine

[a] Corresponding author: stephanie.daval@inra.fr

S. Daval and A. Belcour have contributed equally to the work

Abstract

Plasmodiophora brassicae is an obligate biotrophic pathogenic protist responsible for clubroot, a root gall disease of Brassicaceae species. In addition to the reference genome of the *P. brassicae* European e3 isolate and the draft genomes of Canadian or Chinese isolates, we present the genome of eH, a second European isolate. Refinement of the annotation of the eH genome led to the identification of the mitochondrial genome sequence, which was found to be bigger than that of *Spongospora subterranea*, another plant parasitic Plasmodiophorid phylogenetically related to *P. brassicae*. New pathways were also predicted, such as those for the synthesis of spermidine, a polyamine up-regulated in clubbed regions of roots. A *P. brassicae* pathway genome database was created to facilitate the functional study of metabolic pathways in transcriptomics approaches. These available tools can help in our understanding of the regulation of *P. brassicae* metabolism during infection and in response to diverse constraints.

1. Introduction

Clubroot, caused by the obligate telluric biotroph protist *Plasmodiophora brassicae* Wor., is one of the economically most important diseases affecting Brassica crops in the world. The pathogen leads to severe yield losses worldwide (Dixon, 2009). Clubroot development is characterized by the formation of galls on the roots of infected plants. Above-ground symptoms include wilting, stunting, yellowing and premature senescence. Chemical control of the disease is difficult and/or expensive. Agricultural practices, such as crop rotation, and cultivation of resistant Brassica varieties are the main methods for disease control (Diederichsen et al., 2009).

The pathogen has a complex life cycle comprising three stages: survival in soil as spores, root hair infection, and cortical infection. First, resting spores can survive for a long period (up to fifteen years) in the soil. Secondly, germination of haploid resting spores releases primary zoospores which infect root hairs, leading to intracellular haploid primary plasmodia. In the third stage, after nuclear division, secondary zoospores may be released into the soil and penetrate the cortical tissues. The step of released zoospores in the soil before invading the root cortex is not universally accepted among the research community. During this last stage, the pathogen develops into secondary multinucleate diploid plasmodia which cause the hypertrophy and hyperplasia of infected roots into characteristic clubs (Ingram and Tommerup 1972; Bulman and Braselton 2014). Secondary plasmodia finally release a new set of haploid resting spores.

P. brassicae is a member of the order Plasmodiophorid within the eukaryote supergroup Rhizaria. This protist taxon includes other important plant pathogens such as *Spongospora subterranea*, causal agent of powdery scab on potato, and *Spongospora nasturtii*, the causal agent of crook root in watercress (Burki et al., 2010).

Several *P. brassicae* genomes are available, including one reference genome from a European isolate (Schwelm et al., 2015), and additional re-sequencing of five Canadian (Rolfe et al., 2016) and one Chinese isolates (Bi et al., 2016). The pathotype classification of these isolates is difficult to compare because they have been characterized using different differential hosts sets. In these previous studies, the main characteristics of the *P. brassicae* genome were described, such as its small size (24.2 - 25.5 Mb) and high gene density. Thirteen chitin synthases were described, suggesting an important role for this gene family in resting spore formation (Schwelm et al., 2015). Secreted proteins, described to play a role in plant pathogenic organisms in suppressing the plant defense responses and modifying the host metabolism, were previously predicted *in silico* in *P. brassicae*, but for most of them, no putative function could be assigned (Schwelm et al., 2015; Rolfe et al., 2016). The *P. brassicae* genome also contains genes potentially involved in host hormone manipulation, such as the auxin-responsive Gretchen Hagen 3, isopentenyl-transferases, methyltransferase and cytokinin oxidase (Schwelm et al., 2015). The clubroot genome also appears to be lacking several metabolic pathways, a characteristic of the eukaryotic biotrophic plant pathogens (Spanu et al., 2010; Baxter et al., 2010; Kemen et al., 2011). The missing genes encode proteins involved in sulfur and nitrogen uptake, as well as the arginine, lysine, thiamine and fatty acid biosynthesis pathways. In addition, only a few carbohydrate active enzymes (CAZymes), involved in the synthesis, metabolism and transport of carbohydrates, were found. Characterization of the *P. brassicae* genome is, however, still incomplete as the above data were from a limited number of genotypes and conditions.

Schwelm et al. (2016) recently summarized the current life stage-specific transcriptomics data related to the molecular events in the life cycle of *P. brassicae*. During spore germination and at the primary zoospore stage, enzymes involved in chitinous cell wall digestion are highly expressed. Several metabolism pathways such as the starch, citrate cycle, pentose phosphate pathway, as well as pyruvate and trehalose metabolism were also very active (Schwelm et al., 2015). Although the molecular mechanisms involved in primary root hair infection are unknown, the secondary phase has been better described. During re-infection by the secondary zoospores in root hairs or the root cortex, genes involved in basal and lipid metabolism were highly expressed (Bi et al., 2016). The plant clubroot symptoms (hyperplasia and hypertrophy of the infected tissues leading to the gall phenotype) were found to result from the deregulation of the host plant primary and secondary metabolism (Ludwig-Müller 2009a; Gravot et al. 2011, 2012; Wagner et al. 2012) as well as from the modification of plant hormone homeostasis (Devos et al., 2006; Siemens et al., 2006; Ludwig-Müller, 2009b; Malinowski et al., 2016) by the pathogen. The main plant hormone pathways modified were cytokinin biosynthesis, auxin homeostasis (Siemens et al., 2006; Ludwig-Muller, 2009b; Schuller et al., 2014), and salicylic acid and jasmonic acid metabolism (Lemarié et al., 2015). However, the mechanisms by which the pathogen may drive plant hormone homeostasis are still not completely deciphered. At the end of the *P. brassicae* life cycle, secondary plasmodia transform into resting spores. In these structures, trehalose could play an important role for their long-term survival capacity (Brodmann et al., 2002). This description of the current transcriptomics data in *P. brassicae* shows that some important steps in *P. brassicae* developmental stages are still not well-characterized at the molecular level.

More information is thus needed, in particular from other clubroot genotypes to improve gene and protein descriptions, thus perhaps allowing the discovery of some of the potentially missing pathways or some specific pathotype features. Therefore, to gain further structural and functional knowledge of the *P. brassica*e genome, the first aim of this study was to generate a de *novo*-sequence from a second European isolate. We report here the genome sequence of the European eH *P. brassicae* isolate belonging to a pathotype widespread in France but different to e3, the reference clubroot isolate. For the first time, the mitochondrial genome of this clubroot pathogen was described. The second aim was to improve genome annotation, to predict new metabolic pathways and to develop ClubrootCyc, a metabolic pathway database for *P. brassicae*. This new tool gives researchers access to a repertoire of *P. brassicae* metabolic and transport pathways as well as to an Omics Viewer to analyze expression data, which can be updated with new data.

2. Materials and methods

2.1. Pathogen material and plant inoculation

The *P. brassicae* isolate used in this study was eH. The "selection" isolate eH (*i.e.*, an isolate displaying a phytopathological pattern similar to a single spore isolate) was kindly provided by Dr. J. Siemens (University of Berlin) and was described by Fähling et al. (2003). Isolate eH was characterized for pathogenicity on differential *Brassica napus* cultivars according to the classification proposed by Somé et al. (1996) and corresponded to the pathotype P1.

To ensure genetic homozygosity of the pathogen DNA material, resting *P. brassicae* spores were used. The isolate was propagated on the universal susceptible host Chinese cabbage (*Brassica rapa* ssp *pekinensis* cv. *Granaat*). Clubroots were collected, washed and surface sterilized to avoid host and bacterial contamination. Galls were homogenized in a blender with sterile water and separated by filtration through layers of cheesecloth. The resting spores were then separated by filtration through 100 µm then 55 µM sieves to remove plant cell debris. After centrifugation at 2500 g for 5 min, the pellet was carefully treated with DNase RQ1 for 10 min at 65°C and centrifuged at 2500 g for 5 min at 4°C (Manzanares-Dauleux et al., 2000).

2.2. DNA and RNA extraction

Prior to nucleic acid extraction for genome sequencing, the purified resting spores were ground in liquid nitrogen using a mortar and a pestle. DNA was purified using the NucleoSpin Plant II Kit (Masherey-Nagel) following the manufacturer's instructions. The DNA quality was verified on an agarose gel and the quantity was estimated with a Nanodrop 2000 (Thermoscientific). The absence of host DNA plant contamination was tested by PCR using Brassica specific cruciferin primers (forward primer 5'-ggccagggtttccgtgat-3' and reverse primer 5'-ccgtcgttgtagaaccattgg-3').

Total RNA was also obtained from purified, non-germinating, resting spores using TRIzol reagent (Invitrogen) following the manufacturer's instructions. RNA purity and quality were assessed with a Bioanalyser 2100 (Agilent) and quantified with a Nanodrop (Agilent). The RNA Integrity Number (RIN) of the purified RNA was 7.0.

2.3. Genome sequencing and assembly

For *de novo* sequencing of the eH genome, different libraries were sequenced using 40 µg DNA with Illumina HiSeq 2500 technology (Genoscreen, Lille, France). Three libraries were sequenced: a paired-end shotgun library with an insert size of 430 bp, a 3 kb insert mate-pair library, and a 5 kb insert mate-pair library. The DNA-shotgun library was prepared with the TruSeq nano DNA kit (Illumina), and the mate-pair libraries were done with the Nextera Mate Pair Sample Prep Kit (Illumina). The sequence quality was assessed with fastqc. Paired-end reads were cleaned (N-trimming) and reads shorter than 50 bp were removed with home-made scripts. The eH genome was *de novo* assembled (86 bases k-mer) using SOAPdenovo. Gaps were trimmed and gap-closed using GapCloser, scaffolds were filtered (500 nt in size) and redundancy was removed. The completeness of genome assembly data sets was analyzed with Benchmarking Universal Single-Copy Orthologs (BUSCO, http://busco.ezlab.org) from eH, e3 and ZJ-1 genome sequences with the EnsemblProtists database.

For scaffold #43, identified as the mitochondrial genome, assembly was performed with appropriate software, such as MITObim (Hahn et al., 2013), NOVOPlasty (Dierckxsens et al., 2017) and MindTheGap (Rizk et al., 2014).

2.4. Transcriptome sequencing and functional annotation

RNA from eH non-germinating resting spores was used for RNA sequencing in order to obtain high quality *P. brassicae* RNA free of RNA from plants or microbes. Sequencing was performed in 1x100 bp Illumina HiSeq2500 (Genoscreen, Lille, France). For this, an oriented library was prepared with the TruSeq Stranded mRNA Sample Prep Kit (Illumina). Briefly, after isolation of the mRNA using oligodT and removing of the rRNA, the mRNA were purified and fragmented. The cDNA were synthesized and used for library's construction with the TrusSeq Stranded mRNA Kit (Illumina). The oriented single end reads were trimmed and cleaned with home-made scripts. All non-redundant generated sequences were processed using Velvet for an intermediate transcriptome assembling, and Gmap for alignment to the eH genome sequence.

Genome annotation was conducted using EuGene prediction software (Foissac et al., 2008) trained on *P. brassicae* data. The annotation pipeline included in EuGene integrates various sources of evidence, such as high throughput strand-specific RNA-Seq data, intrinsic information provided by coding potential (Interpolated Markov Models), stop and start codon analysis (using a dedicated RBS alignment tool), similarities with known proteins (SwissProt), high quality CDS predictions and ncRNA predictions (tRNAscan-SE, RNAmmer and Rfam-scan softwares). Putative transcript function and Gene Ontology (GO) analyses were performed using Blast2GO (Conesa et al., 2005) with default settings. For more detailed genome annotation and to retrieve information (Enzyme Commission numbers (EC), GO terms, InterPro and pathways), several external databases were interrogated: Reactome (Fabregat et al., 2016), KEGG (Kanehisa et al., 2017), MetaCyc (Caspi et al., 2016), EupathDB (Aurrecoechea et al., 2017), Uniprot (The UniProt Consortium, 2017), Gene Ontology (Ashburner et al., 2000; The Gene Ontology Consortium, 2017), and ECDomainMiner to link known Pfam domains to EC numbers (Alborzi et al., 2017). Carbohydrate-active enzymes (CAZymes) were predicted using the dbCAN pipeline (Yin et al., 2012). SignalP (Petersen et al., 2011) was used to detect secretion signals in protein sequences. The protein sequences predicted to have a secretion signal by SignalP, were then subjected to transmembrane helix prediction with TMHMM (Krogh et al., 2001). Blastn and Blastp were used to search for already known sequences (Altschul et al., 1997). This was done for the *P. brassicae:* i) serine protease (Feng et al., 2010), ii) methyltransferase

(Ludwig-Müller et al., 2015), iii) genes detected by dot plot and real time PCR in the Led09 isolate (Feng et al., 2013), iv) e3 genome (Schwelm et al., 2015), v) genes associated with lipid droplets (Bi et al., 2016), and vi) the mitochondrial genome of *S. subterranea* (Gutiérrez et al., 2014). To visualize the data, circular representations of scaffolds were created using Circos (Krzywinski et al., 2009).

For the othology analyses, the proteomes of *P. brassicae* eH and e3 were compared using Orthofinder version 2.2.7 (Emms and Kelly, 2015). The e3 proteome was downloaded from the NCBI database (GCA_001049375). No proteome was found for the ZJ-1 isolate.

2.5. Analyses of the mitochondrial genome

Gene annotation of scaffold #43 was performed with the automated MFannot tool (Burger et al., 2013) that predicts group I and group II introns, tRNAs, RNase P-RNA, and 5S ribosomal RNA (rRNA). The annotation was verified with Blastn and Blastp by quering the NCBI databases.

For synteny analyses, the five mitochondrial genomes of *Spongospora subterranea* (NC_034004), *Bigelowiella natans*, (HQ840955), *Lotharella oceanica* (KT806043), *Paracercomonas marina*, (NC_026310) and *P. brassicae* strain ZJ-1 (assembly ASM209382v1) were downloaded from the NCBI database. The annotations came from the downloaded files except for the ZJ-1 strain, for which an annotation by MFannot was performed to obtain the list of genes. No mitochondrial genome was found in *P. brassicae* e3 (GCA_001049375) after running Blast against it with the mitochondrial genome found in the eH isolate. The mitochondrial genome synteny figure was generated using the SimpleSynteny alignment tool version 1.4 with default settings (Veltri et al., 2016).

For the phylogeny, 20 COX1 protein sequences were used. When available, the sequences were extracted from the mitochondrial genome resources. If not, the *COX1* sequences of *Assulina muscorum* (HM641252), *Assulina seminulum* (HM641250), *Chlorarachnion reptans* (AB009416), *Chlorarachnion sp.* CCMP238 (AB009396), *Cyphoderia amphoralis* isolate Rila 4 (HM187647), *Cyphoderia ampulla* isolate Vitosha 2 (HM187673), *Cyphoderia ampulla* isolate Cape Breton 2 (HM187663), *Cyphoderia cf. compressa* (ADK22128), *Cyphoderia compressa* isolate Rila 1 (HM187661), *Cyphoderia trochus palustris* (HM187669), *Euglypha rotunda* (AB009417), and *Pseudocorythion acutum* (HM187672) were obtained from the NCBI database. An outgroup composed of *Plasmodium falciparum* (AAC63390) and *Plasmodium fragile* (AAW52538) was created. The IQ-Tree version 1.6.7 (Nguyen et al., 2015) software was used for the tree construction. The sequences were aligned using MAFFT version 7.407 (Katoh and Standley, 2013). The selection of model was made with ModelFinder (Kalyaanamoorthy et al., 2017) and the Bayesian Information Criterion. The phylogeny used maximum likelihood and bootstrap with 100 replicates. For the creation and editing of tree picture, FigTree version 1.4.3 (http://tree.bio.ed.ac.uk/) and Inkscape version 0.92 (www.inkscape.org) were used.

2.6. Analysis of codon usage in nuclear and mitochondrial genomes

For each particular codon, the Relative Synonymous Codon Usage (RSCU) was measured in the nuclear genome on one hand and in the mitochondrial genome on the other hand, with the DAMBE software (Xia, 2018).

2.7. Metabolic pathway database

Pathways were predicted with MetaCyc database version 20.5 (Caspi et al., 2016). A GenBank file was created containing DNA sequences and functional annotations. This file was used as an input of the software Pathway-Tools version 21 (Karp et al., 2011). Using functional annotations (especially Enzyme Commission number), Pathway-Tools software ran the PathoLogic program that compared the *P. brassicae* annotations to the database MetaCyc version 20.5 (Caspi et al., 2016). The inferred reactions in the organism were then compared to the pathways of the database. Pathways were then deduced and used to produce a genome-scale metabolic network. At the end of the process by Pathway-Tools, a Pathway Genome DataBase (PGDB) was created for *P. brassicae*: ClubrootCyc. PGDB functionalities were validated using data from the transcriptomics analysis of lipid droplets by Bi et al. (2016).

2.8. Data Availability

The eH *P. brassicae* genome sequence has been deposited in the public domain in GenBank Data Librairies under accessions SRR7819149, SRR6395507, SRR6395506 and SRR6395505 for raw data sequences, and POCA00000000 for the genomic data. The genome annotation is available at https://doi.org/10.15454/1.51904983296096e12. The Pathway

Genome DataBase ClubrootCyc is available online at https://pathway-tools.toulouse.inra.fr/CLUBROOT

3. Results and discussion

3.1. Nuclear and mitochondrial genomes: general characteristics in structure and composition

Main features of the nuclear genome

For a better understanding of the pathogenesis and metabolic potential of *P. brassicae*, we generated a *de novo* assembled genome of the eH isolate (belonging to pathotype 1 according to Somé et al., 1996). A paired-end shotgun library consisting of 73,082,688 reads (read length of 101 bp and insert size of 430 bp) and approximately 225-fold genome sequence coverage was obtained. Two additional paired-end libraries were also generated. A 3 kb insert mate-pair Illumina library resulted in 97,629,456 reads (read length of 101 bp and insert size of 3200 bp) and a 5 kb insert mate-pair Illumina library resulted in 81,813,222 reads (read length of 101 bp and insert size of 5050 bp). Both libraries displayed also an approximately 225-fold genome sequence coverage. The final assembly consisted of 136 gap-filled scaffolds with a N50 of 741,293 bp and the total sequence length was 24,559,637 bp with a 59.29% GC content. Basic information about the assembled genome and predicted genes is shown in Table 1, in comparison with the assembly of e3 and ZJ-1 genomes. While the ZJ-1 resources corresponded to draft data, the assemblies of eH and e3 shared the same levels of quality. The main features of the eH genome were indeed similar to the reference e3 genome, particularly concerning GC content (Figure 1), but the eH genome displayed a less fragmented assembly (Table 1) and a higher number of predicted genes (12,811 in eH and

9,730 in e3). Figure 1 shows a visual representation of the eH genome’s first 43 scaffolds (which represent 95% of the total genome sequence) compared to the other available *P. brassicae* sequences (the sequences published by Rolfe et al. (2016) were not included because data are not available). The gene density, calculated as the number of genes every 20 kb, was nine on average. This gene density is relatively high as previously described (Schwelm et al., 2015; Rolfe et al., 2016), and contributes to the small genome size compared to the two known genomes of free-living Rhizaria, *Bigelowiella natans* (~100 Mb) (Curtis et al., 2012) and *Reticulomyxa filosa* (~320 Mb) (Glöckner et al., 2014).

Gene similarity between the eH and e3 isolates was studied by comparing the gene sequences of both genomes. Using a stringent query span of 80% (meaning that the percentage of alignment coverage of the query sequence should be at least 80%, making sure that we compared genes with common regions), 60% of the genes found in the eH genome showed similarity with the e3 genes. Among the 6,654 alignments displaying the query span of 80%, 4,702 were with 100% gene identity and only 52 were with <60% gene identity (value often used to define genes as highly divergent). For the ZJ-1 isolate’s genome (Bi et al., 2016), annotated regions are not available which made it unfeasible to carry out such accurate comparisons between genes and only contigs were compared which include non-coding regions.

The scaffolds #44 to #242 (corresponding to 5% of the total genome) and in particular scaffolds #88 to #242 were more difficult to assemble (Figure S1) and thus few genes were annotated in these regions.

Mitochondrial genome: structure and gene content

In the circular genome representation (Figure 1), scaffold #43 showed a low GC content and one match with a *P. brassicae* mitochondrial gene (Siemens 2002, Unpublished, NCBI ID: AF537102). In addition, as a high AT content was found in some Plasmodiophorid mitochondrial genomes, for example in *S. subterranea* (Gutiérrez et al., 2014), these observations suggested that the scaffold #43 could correspond to the *P. brassicae* mitochondrial genome. To ensure that no errors occurred in its assembly, a new assembly of this scaffold #43 was performed using software specifically designed for organelle assembly (NOVOPlasty). Both assemblies were similar suggesting consistent results.

The newly assembled sequence of eH scaffold #43 is 102,962 bp, with a 24.67% GC content (Figure 2) and was very similar (Max score = 91831, Total score = 1.728e+05, Query cover = 90%, E-value = 0.0, Identity = 99%) to scaffold #49 of the genome of the *P. brassicae* isolate ZJ-1 (Bi et al., 2016). The size of the mitochondrial *P. brassicae* sequence (102,962 bp) was larger than that of *S. subterranea* (37,699 bp) (Gutiérrez et al., 2014). Such diversity in genome size was previously described in Chlorarachniophytes, a small group of algae from the supergroup Rhizaria phylum Cercozoa, in which the mitochondrial genomes ranged from 34 to 180 kbp (Tanifuji et al., 2016). Genome sizes vary considerably because of variation in the intergenic regions and intron content, which both contribute to mitochondrial genome plasticity. In *P. brassicae*, numerous long introns were found in mitochondrial genes in contrast to the *Lotharella oceanica* (Rhizaria group) and *B. natans* mtDNAs (Tanifuji et al., 2016).

To assess the extent of mitochondrial genome rearrangements more closely, mitochondrial genomes of *P. brassicae* (eH and ZJ-1), *S. subterranea*, *B. natans*, *L. oceanica* and *Paracercomonas marina* were aligned using the SimpleSynteny alignment tool and the protein sequences. A genome synteny map is shown in Figure 3. Despite the significant difference in total size, mitochondrial genomes of *P. brassicae* eH and ZJ-1, and *S. subterranea* exhibited a similar architecture in which conservation of homologous genes and gene order were identified, but some genes are missing in ZJ-1. The organization of the three other mitochondrial sequences was also in part conserved in the content of genes but in a lesser extent concerning the gene order. The genome sizes varied considerably between these species, mainly because of the variations of the intergenic regions and the intron content, making impossible an accurate synteny's analysis using gene sequences (data not shown).

Both mitochondrial sequences were also quite similar between *P. brassicae* and *S. subterranea*: when the blast best-hit analyses of the *P. brassicae* mitochondrial genes were displayed for the "protein sequences" (32 protein-coding genes (Table S1)) and for the "complete gene sequences" (the 28 other types of genes (Table S2)), 60 genes were found among which 56 were also found among the 59 described in *S. subterranea* (Gutiérrez et al., 2014). Genes coding for proteins involved in respiration and oxidative phosphorylation (17 genes: *ATP 1-7*, *4L*, *9*, *COB*, *COX 1-3*, *ATP 1*, *6*, *8*, *9*), and for ribosomal subunit proteins (10 for small subunit: *RPS, 4, 7*, *8*, *10*, *11*, *12*, *13*, *14*, *19* and 4 for large subunit: *RPL 5*, *6*, *14*, *16*) were identified. As in *S. subterranea,* 24 genes coding for tRNA were found in the *P. brassicae*, but *tRNH* described in *S. subterranea* was not found in *P. brassicae* and *tRNQ* was found only in *P. brassicae* (and not in *S. subterranea*). Interestingly, the protein-coding genes in *P. brassicae* displayed a mosaic structure with small parts spread across the scaffold, as described for mitochondrial sequences of the chlorarachniophyte *B. natans* (belonging to the Rhizaria group) and the model cryptophyte protist *Guillardia theta*; both algal nuclear genomes are mosaics of genes derived from the host and endosymbionts (Curtis et al., 2012). No transposable elements were detected in the *P. brassicae* mitochondrial genome.

In order to shed further light on the *P. brassicae* taxonomic classification, mitochondrial sequences of the *COX1* gene were used in phylogenetic analyses. For this, 20 sequences of the *COX1* gene from diverse taxa were retrieved from NCBI (not found for ZJ-1 isolate), aligned, and used to create a maximum likelihood phylogenetic tree (Figure 4), knowing that analyses with the Bayesian method gave the same results (data not shown).The

COX1 gene was chosen because it was retrieved from more species than the other mitochondrial genes. Both *P. brassicae* and *S. subterranea* strains were clustered in a branch with very high bootstrap and a posterior probability of 0.99, confirming that both were closely related and belonged to the order of Plasmodiophoridae. The different known orders inside the Rhizaria were also relatively well related since families of Cercozoa (Plasmodiophoridae, Chlorarachniophyceae, Cercomonadidae, Euglyphidae and Cyphoderiidae) were clustered in a clade supported by a posterior probability superior to 0.9.

3.2. In silico genome annotation's improvement and functional prediction: metabolic pathway database

Nuclear genome annotation

For annotation, the RNAseq of resting spores was used and resulted in 23,540,954 reads. Other transcriptomes obtained *in planta* from other studies did not improve the annotation (data not shown), showing that the transcriptional machinery was operational in resting spores. This allowed us to well compare our results to the e3 reference genome for which seven transcriptomes were used, including those of resting and maturing spores (Schwelm et al., 2015).

The genome annotation was improved compared to the e3 reference genome by interrogating several external databases as described in the materials and methods. The annotation data predicted 12,590 protein-coding genes in eH genome (Table 1). In total, 8,989 genes were associated with GO terms, 4,523 with EC numbers and 8,231 with InterPro signatures. In rare cases of conflicting annotation, the results from manual annotation and Blast2Go were favored.

Carbohydrate-active enzyme sequences of both the e3 and eH genomes were compared using dbCAN, a web server and DataBase for automated Carbohydrate-active enzyme ANnotation. For the eH isolate, 287 enzymes were predicted (Figure S2; Table S3): 89 Glycoside Hydrolases (GH), 111 Glycosyltransferases, 47 Carbohydrate Esterases (CE), 7 Polysaccharide Lyases, 23 Carbohydrate-Binding Modules (CBM), and 10 Auxiliary Activities. GH18, CE4 and CBM18 were the most abundant enzymes (respectively 19, 21, 7 were predicted) within some of these groups. Amongst these 287 eH genes with predicted carbohydrate-active enzyme activity, 223 matched with e3 genes. The list of CAZymes was similar in both the eH and e3 genomes.

Using SignalP and TMHMM to predict the secretome, 741 protein sequences were found to harbor a secretion signal prediction without transmembrane helix. Furthermore, 543 sequences corresponded to small proteins (less than 450 amino acids). Schwelm et al. (2015) analyzed the e3 genome and predicted 553 secreted proteins including 416 with a length smaller than 450 amino acids. Among the proteins with a signal prediction, 355 were common to e3 and eH genomes (Table S4). The higher number of predicted proteins found in the eH genome could be due to the annotation using Eugene which probably created more gene sequences compatible with secretome pattern detection. In addition, the use of updated software versions may have also upgraded the secretome.

An additional annotation was carried out by blasting eH data against clubroot published sequences. All available sequence data for concerning eH and other *P. brassicae* isolates are summarized in Table S5. Ninety-nine percent and 100% identity matches were observed for the *P. brassicae* methyltransferase gene (Ludwig-Müller et al., 2015) and the serine protease gene (Feng et al., 2010), respectively. Three hundred and eighty eight out of the 426 genes described by Bi et al. (2016) involved in lipid metabolism were found in the eH genome (Table S5B). From the 118 genes identified by Feng et al. (2013) in the Led09 isolate’s genome, 23 matched the eH genome (Table S5C). Genes identified as being involved in pathogenicity (*i.e.* with functions such as P-type ATPase activity, Mps1 binder-like protein) were mapped to the eH genome. To complete the annotation, out of the 9,730 genes in the e3 genome, 6,626 were matched to the eH genome (Table S5A). To go further in the comparison of eH and e3 genomes, a global comparison of orthologous gene clusters study was performed (Table S6). Usually, orthology is the most accurate way of describing differences and similarities in the composition of genomes from different species, but it can be also used to compare strains intra-species. The number of orthologous gene clusters was 8168, among which 7485 orthogroups were shared by e3 and eH isolates, and 683 clusters were specific to one of the two isolates, showing a certain relatedness between both stains.

The annotations were added to the corresponding eH genes using python scripts querying the Uniprot database. Using this updated annotation, the genes were classified into subsystems whose main terms are described in Figure 5. Most of the annotated genes were classed in the protein-binding (1,433) GO term category. This class, not accurately defined for their function, still remained high because of the lack of functional data in Plasmodiophorid genomics resources. Ninety-six genes encoding potential RNA-binding proteins having at least one RNA-binding domain were identified. This protein family could play a role in *P. brassicae* pathogenicity, as described in *Ustilago maydis*, the basidiomycete causative agent of corn smut disease (Becht et al., 2005). From the InterPro (IP) database, the ankyrin repeat (258 and 259) was found to be the major protein domain. The ankyrin repeat is the most common protein-protein interaction motif in nature. It is predominantly found in eukaryotic proteins, but also in many bacterial pathogens which employ various types of secretion systems to deliver ankyrin-containing proteins into eukaryotic cells where they mimic or manipulate various host functions (Al-Khodor et al., 2010). In eukaryotes, proteins with ankyrin repeats are involved in many cellular functions, such as transcriptional regulation, signal transduction, cell cycle but also in virulence in Amoebae, suggesting an adaptation to their intracellular localization (Molmeret et al., 2005). The high number of ankyrin repeats in the *P. brassicae* genome could suggest a role in its capacity to invade host plant cortical cells.

To summarize the long list of GO terms by removing redundancy, a semantic similarity-based graph was constructed with REViGO (Supek et al., 2011) (Figure S3). This allowed a representative subset of the terms to be found using a clustering algorithm that relies on semantic similarity measures. The Revigo Treemap showed a high representation of the GO terms linked to protein phosphorylation. Changes in the phosphorylation state have a broad impact on cells because they affect many processes such as the subcellular protein localization, protein-protein interactions or signaling cascades (Johnson, 2009). Moreover, interference of the host phosphorylation machinery is a common strategy used by pathogens to promote infection and development in host tissues (Albataineh and Kadosh, 2016), particularly in plant-pathogenic fungi (Turra et al., 2014). A better understanding of the regulation of protein phosphorylation could help in developing targets for regulation of *P. brassicae* infection processes. Another GO term found frequently in the *P. brassicae* genome annotation was related to lipid metabolism. This is consistent with Bi et al. (2016) who reported proteins associated with lipid droplets in resting spores of *P. brassicae* and showed that lipid conversion is high in this protist during cortex infection.

Finally, the codon usage, a pattern that describes the degeneracy of the genetic code and illustrates differential usage of synonymous codons, was deciphered in nuclear genome *versus* mitochondrial genome. Using the genetic invertebrate mitrochondrial code from NCBI, the codon usage of the *P. brassicae* mitochondrial ORFs was determined. The Figure S4 presents the RSCU in both nuclear and mitochondrial genomes. This shows that, as usually described in other species, the codon usage was not identical between nuclear and organelle in *P. brassicae* (Sablok et al., 2017). As the major pattern in codon usage was a result of differences among the genes in the frequencies of G+C, our results are coherent with the higher G+C content in mitochondrial sequence than in nuclear genome (Zhang et al., 2007).

Metabolic pathway database

Taking advantage of the improved annotation described here, a *P. brassicae* specific database was created using Pathway Tools Software. New genome annotation and available databases (Uniprot, Gene Ontology and ECdomainMiner) served to create a Genbank entry. Using this file as input, the PathoLogic component of Pathway Tools could be used to create a new Pathway Genome DataBase (PGDB) containing the predicted *P. brassicae* metabolic pathways. The database is called ClubrootCyc (Figures 6 and S5) and predicts 115 pathways, 1,938 enzymatic reactions, 71 transport reactions, 4,468 enzymes, 123 transporters, and 1,567 compounds (https://pathway-tools.toulouse.inra.fr/CLUBROOT). It links genomic data to additional information such as protein annotation, enzymatic reactions and pathways.

To validate the relevance and functionality of the ClubrootCyc database, the metabolic reconstruction was tested with available published transcriptomics data described by Bi et al. (2016). Visual analysis of omics gene-expression, highlighted in Figure S6, showed that the major pathway was involved in fatty acid and lipid biosynthesis and degradation, consistent with the results of Bi et al (2016) who showed that lipid metabolism is enriched in cortical infection stage of gall development.

The development of this *P. brassicae* - specific database is thus a useful tool for integrating transcriptomics data, gene and metabolic networks and to gain further information than just a list of over- or under-expressed genes between experimental conditions.

Prediction of new pathways

Moreover, ClubrootCyc also allowed us to access new pathways and to complete annotation data. For instance, a complete pathway for the spermidine biosynthesis was predicted in eH *P. brassicae* (Figure S7). Particularly, the three enzymes (EC 2.7.2.11, EC 1.2.1.41, and EC 2.6.1.13) required for conversion of glutamate to ornithine, were present in the eH genome although they were not found in the genomes described by Rolfe et al. (2016) or by Schwelm et al. (2015). Ornithine is converted into putrescine, a precursor of the polyamine spermidine. The roles of polyamines synthesised by plants are well documented. They are involved in fundamental cellular processes, such as chromatin organization, cell proliferation, differentiation, and programmed cell death (Thomas and Thomas, 2001; Bais and Ravishankar, 2002), but also in adaptive responses to abiotic (Bouchereau et al., 1999; Urano et al., 2003; Kuznetsov et al., 2006) and biotic stresses (Walters, 2003). Polyamines have been found being accumulated in the galls caused by *P. brassicae* (Walters and Shuttleton, 1985). In turnip (*Brassica rapa var. rapa),* the levels of putrescine, spermidine and spermine were reported to be higher in the roots infected with clubroot than in the non-infected normal roots (Walters 2003). In cabbage, broccoli and Komatsuna clubroot galls, the levels of putrescine, spermidine and spermine also increased after infection with *P. brassicae* (Hamana et al., 2015). The high level of polyamines measured in the roots may be contributed by the host plant, as reported by Cao et al. (2008) who showed the up-regulation of the *B. napus* polyamine biosynthetic enzyme spermidine synthase in response to *P. brassicae* infection. However, it could also be due to *P. brassicae* since it is difficult to distinguish between metabolites of plant or pathogen origin. Polyamine biosynthesis could be vital to the survival and pathogenesis of *P. brassicae* as described in other pathogens, such as bacteria (Shah et al., 2011), fungi (Valdes-Santiago et al., 2012) and protists (Rhee et al., 2007).

The chorismate/Shikimate pathway, which was predicted by Schwelm et al. (2015) but lacking the chorismate mutase (EC 5.4.99.5), was found to be complete in the eH *P. brassicae* genome suggesting that this pathway is operational in *P. brassicae***.** This pathway is known in plants, bacteria, fungi, and apicomplexan parasites and is central for the biosynthesis of carbocyclic aromatic compounds. Chorismate is the common branchpoint for the production of these metabolites, and some intermediates in these pathways also have important physiological relevance. For example, salicylate, which is derived from chorismate, is involved in plant defense against *P. brassicae* (Lemarié et al., 2015). Anthranilate, another derivative of chorismate, is the precursor of the *Pseudomonas* quinolone signal, which regulates numerous virulence factors in *Pseudomonas aeruginosa* (Palmer et al., 2013). Furthermore, the chorismate mutase enzyme of this pathway, newly identified in this study in the eH genome, was described as being involved in gall formation in *Meloidogyne javanica* (Doyle and Lambert, 2003) and in *Ustilago maydis* (Djamei et al., 2011). The potential role of this enzyme in *P. brassicae* pathogenicity would require functional studies.

ClubrootCyc also predicted the L-Dopachrome biosynthesis pathway, which is involved in melanin synthesis. Melanin is a broad term for a group of natural pigments found in most organisms and localized in cell walls of different structures of fungi (Toledo et al., 2017). They represent well-known virulence factors for several pathogenic fungi (Jacobson, 2000) and for bacteria (Nosanchuk and Casadevall, 2003).

So ClubrootCyc, a freely accessible tool that can be updated with newly published results, provides a way to easily visualize omics data and represent metabolic pathways potentially involved in particular experimental conditions of interest. Metabolic pathway prediction in *P. brassicae* is still, however, incomplete and poorly annotated because there is still not enough information in the reference databases. To make further progress in metabolic network reconstruction and modelling, appropriate gap-filling methods are required. For this, software is being developed to identify and add lacking pathways, such as Meneco, a tool dedicated to the topological gap-filling of genome-scale draft metabolic networks (Prigent et al., 2017).

4. Conclusion

In this study, we investigated the genome sequence of an additional *P. brassicae* isolate, providing new genomics resources for further diversity studies. Additional genome sequencing of *P. brassicae* isolates may allow more precise diagnostic molecular tools than the current pathotyping systems to be developed. A major result was also the description of *P. brassicae* mitochondrial sequence, which can now be used for phylogenetic analyses in the supergroup Rhizaria. We also improved the annotation of *P. brassicae* genome allowing a better description of the metabolic pathway for this protist. This allowed the biosynthesis pathways of some specific compounds that could play a role in infection to be clarified. By combining genomic analyses with metabolic network reconstruction in *P. brassicae,* the present study provides useful post-genomics tools devoted to this protist. This will help improve our understanding of the infection processes involved in the diseases caused the pathogens of this family by targeting experiments to assess the function of, as yet, unknown enzymes or transporters.

Acknowledgments

We thank the BRC *BrACySol* (INRA Rennes, France) for providing the seeds. We thank Leigh Gebbie for revising the English written style of the manuscript.

References

Albataineh MT, Kadosh D (2016) Regulatory roles of phosphorylation in model and pathogenic fungi. Medical Mycology 54, 333-352.

Alborzi SZ, Devignes MD, Ritchie DW (2017) ECDomainMiner: discovering hidden associations between enzyme commission numbers and Pfam domains. BMC Bioinformatics 18, 107.

Al-Khodor S, Price CT, Kalia A, Abu Kwaik Y (2010) Ankyrin-repeat containing proteins of microbes: a conserved structure with functional diversity. Trends in Microbiology 18, 132-139.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25, 3389-3402.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium Nature Genetics 25, 25-29.

Aurrecoechea C, Barreto A, Basenko E Y, Brestelli J, Brunk BP, Cade S, Crouch K, Doherty R, Falke D, Fischer S, Gajria B, Harb OS, Heiges M, Hertz-Fowler C, Hu S, Iodice J, Kissinger JC, Lawrence C, Li W, Pinney DF, Pulman JA, Roos DS, Shanmugasundram A, Silva-Franco F, Steinbiss S, Stoeckert CJ, Spruill D, Wang H, Warrenfeltz S, Zheng J (2017) EuPathDB: the eukaryotic pathogen genomics database resource. Nucleic Acids Research 45, D581-D591.

Bais HP, Ravishankar GA (2002) Role of polyamines in the ontogeny of plants and their biotechnological applications. Plant Cell, Tissue and Organ Culture 69, 1-34.

Baxter L, Tripathy S, Ishaque N, Boot N, Cabral A, Kemen E, Thines M, Ah-Fong A, Anderson R, Badejoko W, Bittner-Eddy P, Boore JL, Chibucos MC, Coates M, Dehal P, Delehaunty K, Dong S, Downton P, Dumas B, Fabro G, Fronick C, Fuerstenberg SI, Fulton L, Gaulin E, Govers F, Hughes L, Humphray S, Jiang RHY, Judelson H, Kamoun S, Kyung K, Meijer H, Minx P, Morris P, Nelson J, Phuntumart V, Qutob D, Rehmany A, Rougon-Cardoso A, Ryden P, Torto-Alalibo T, Studholme D, Wang Y, Win J, Wood J, Clifton SW, Rogers J, Van den Ackerveken G, Jones JDG, McDowell JM, Beynon J, Tyler BM (2010) Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. Science 330, 1549-1551.

Becht P, Vollmeister E, Feldbrügge M (2005) Role for RNA-binding proteins implicated in pathogenic development of *Ustilago maydis*. Eukaryotic Cell 4, 121-133.

Bi K, He Z, Gao Z, Zhao Y, Fu Y, Cheng J, Xie J, Jiang D, Chen T (2016) Integrated omics study of lipid droplets from *Plasmodiophora brassicae*. Scientific Reports 6, 36965.

Bouchereau A, Aziz A, Larher F, Martin-Tanguy J (1999) Polyamines and environmental challenges: recent development. Plant Science 140, 103-125.

Brodmann D, Schuller A, Ludwig-Muller J, Aeschbacher RA, Wiemken A, Boller T, Wingler A (2002) Induction of trehalase in *Arabidopsis* plants infected with the trehalose-producing pathogen *Plasmodiophora brassicae.* Molecular Plant-Microbe Interactions 15, 693-700.

Bulman SR, Braselton JP (2014) Rhizaria: Phytomyxea. In: Mc Laughlin DJ, Spatafora JW (eds), The Mycota VII, 2nd ed. Systematics and Evolution, Part A. Springer-Verlag. Berlin-Heidelberg, Germany. pp. 99-112.

Burger G, Gray MW, Forget L, Lang BF (2013) Strikingly bacteria-like and gene-rich mitochondrial genomes throughout Jakobid protists. Genome Biology and Evolution 5, 418-438.

Burki F, Kudryavtsev A, Matz MV, Aglyamova GV, Bulman S, Fiers M, Keeling PJ, Pawlowski J (2010) Evolution of Rhizaria: new insights from phylogenomic analysis of uncultivated protists. BMC Evolutionary Biology 10, 377.

Cao T, Srivastava S, Rahman MH, Kav NNV, Hotte N, Deyholos MK, Strelkov SE (2008) Proteome-level changes in the roots of *Brassica napus* as a result of *Plasmodiophora brassicae* infection. Plant Science 174, 97-115.

Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Subhraveti P, Weaver DS, Karp PD (2016) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Research 44, D471-D480.

Conesa A, Götz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674-3676.

Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, Arias MC, Ball SG, Gile GH, Hirakawa Y, Hopkins JF, Kuo A, Rensing SA, Schmutz J, Symeonidi A, Elias M, Eveleigh RJ, Herman EK, Klute MJ, Nakayama T, Oborník M, Reyes-Prieto A, Armbrust EV, Aves SJ, Beiko RG, Coutinho P, Dacks JB, Durnford DG, Fast NM, Green BR, Grisdale CJ, Hempel F, Henrissat B, Höppner MP, Ishida K, Kim E, Kořený L, Kroth PG, Liu Y, Malik SB, Maier UG, McRose D, Mock T, Neilson JA, Onodera NT, Poole AM, Pritham EJ, Richards TA, Rocap G, Roy SW, Sarai C, Schaack S, Shirato S, Slamovits CH, Spencer DF, Suzuki S, Worden AZ, Zauner S, Barry K, Bell C, Bharti AK, Crow JA, Grimwood J, Kramer R, Lindquist E, Lucas S, Salamov A, McFadden GI, Lane CE, Keeling PJ, Gray MW, Grigoriev IV, Archibald JM (2012) Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature 492, 59-65.

Devos S, Laukens K, Deckers P, Van Der Straeten D, Beeckman T, Inzé D, Van Onckelen H, Witters E, Prinsen E (2006) A hormone and proteome approach to picturing the initial metabolic events during *Plasmodiophora brassicae* infection on *Arabidopsis*. Molecular Plant-Microbe Interactions 19, 1431-1443.

Diederichsen E, Frauen M, Linders EG, Hatakeyama K, Hirai M (2009) Status and perspectives of clubroot resistance breeding in crucifer crops. Journal of Plant Growth Regulation 28, 265-281.

Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. Nucleic Acids Research 45, e18.

Dixon GR (2009) The occurrence and economic impact of *Plasmodiophora brassicae* and clubroot disease Journal of Plant Growth Regulation 28,194-202.

Djamei A, Schipper K, Rabe F, Ghosh A, Vincon V, Kahnt J, Osorio S, Tohge T, Fernie AR, Feussner I, Feussner K, Meinicke P, Stierhof YD, Schwarz H, Macek B, Mann M, Kahmann R (2011) Metabolic priming by a secreted fungal effector. Nature 478, 395-398.

Doyle EA, Lambert KN (2003) *Meloidogyne javanica* chorismate mutase 1 alters plant cell development. Molecular Plant-Microbe Interactions 16, 123-131.

Emms DM, Kelly S (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biology 16, 157.

Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, Jassal B, Jupe S, Korninger F, McKay S, Matthews L, May B, Milacic M, Rothfels K, Shamovsky V, Webber M, Weiser J, Williams M, Wu G, Stein L, Hermjakob H, D'Eustachio P (2016) The Reactome pathway knowledgebase. Nucleic Acids Research 44, D481-D487.

Fähling M, Graf H, Siemens J (2003) Pathotype separation of *Plasmodiophora brassicae* by the host plant. Journal of Phytopathology – Phytopathologische Zeitschrift 151, 425-430.

Feng J, Hwang R, Hwang SF, Strelkov SE, Gossen BD, Zhou QX, Peng G (2010) Molecular characterization of a serine protease Pro1 from *Plasmodiophora brassicae* that stimulates resting spore germination. Molecular Plant Pathology 11, 503-512.

Feng J, Hwang SF, Strelkov SE (2013) Assessment of gene expression profiles in primary and secondary zoospores of *Plasmodiophora brassicae* by dot blot and real-time PCR. Microbiological Research 168, 518-524.

Foissac S, Gouzy J, Rombauts S, Mathé C, Amselem J, Sterck L, Van de Peer Y, Rouzé P, Schiex T (2008) Genome Annotation in Plants and Fungi: EuGène as a Model Platform. Current Bioinformatics 3, 87-97.

Glöckner G, Hulsmann N, Schleicher M, Noegel AA, Eichinger L, Gallinger C, Pawlowski J, Sierra R, Euteneuer U, Pillet L, Moustafa A, Platzer M, Groth M, Szafranski K, Schliwa M (2014). The genome of the Foraminiferan Reticulomyxa filosa. Current Biology 24, 11-18.

Gravot A, Grillet L, Wagner G, Jubault M, Lariagon C, Baron C, Deleu C, Delourme R, Bouchereau A, Manzanares-Dauleux MJ (2011) Genetic and physiological analysis of the relationship between partial resistance to clubroot and tolerance to trehalose in *Arabidopsis thaliana*. New Phytologist 191, 1083-1094.

Gravot A, Deleu C, Wagner G, Lariagon C, Lugan R, Todd C, Wendehenne D, Delourme R, Bouchereau A, Manzanares-Dauleux MJ (2012) Arginase induction represses gall development during clubroot infection in Arabidopsis. Plant and Cell Physiology 53, 901-911.

Gutiérrez P, Bulman S, Alzate J, Ortíz MC, Marín M (2014) Mitochondrial genome sequence of the potato powdery scab pathogen. Mitochondrial DNA 27, 58-59.

Hahn C, Bachmann L, Chevreux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads - a baiting and iterative mapping approach. Nucleic Acids Research 41, e129.

Hamana K, Hayashi H, Niitsu M (2015) Polyamines in different organs of Brassica crop plants with or without clubroot disease. Plant Production Science 18, 476-480.

Ingram DS, Tommerup IC (1972) The life history of *Plasmodiophora brassicae* Woron. Proceedings of the Royal Society B: Biological Sciences 180,103-112.

Jacobson ES (2000) Pathogenic roles for fungal melanins. Clinical Microbiology Reviews 13, 708-717.

Johnson LN (2009) The regulation of protein phosphorylation. Biochemical Society Transactions 37, 627-641.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: Fast model selection for accurate phylogenetic estimates. Nature Methods 14, 587–589.

Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research 45, D353-D361.

Karp PD, Latendresse M, Caspi R (2011) The pathway tools pathway prediction algorithm. Standards in Genomic Sciences 5, 424-429.

Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution 30, 772-780.

Kemen E, Gardiner A, Schultz-Larsen T, Kemen AC, Balmuth AL, Robert-Seilaniantz A, Bailey K, Holub E, Studholme DJ, MacLean D, Jones JDG (2011) Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis thaliana.* PLoS Biology 9, e1001094.

Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. Journal of Molecular Biology 305, 567-580.

Krzywinski M, Schein J, Birol İ, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. Genome Research. 19, 1639-1645.

Kuznetsov VV, Radyukina NL, Shevyakova NI (2006) Polyamines and stress: biological role, metabolism, and regulation. Russian Journal of Plant Physiology 53, 583-604.

Lemarie S, Robert-Seilaniantz A, Lariagon C, Lemoine J, Marnet N, Jubault M, Manzanares-Dauleux MJ, Gravot A (2015) Both the jasmonic acid and the salicylic acid pathways contribute to resistance to the biotrophic clubroot agent *Plasmodiophora brassicae* in *Arabidopsis*. Plant and cell Physiology 56, 2158-2168.

Ludwig-Müller J (2009a) Glucosinolates and the clubroot disease: defense compounds or auxin precursors? Phytochemistry Reviews 8, 135-148.

Ludwig-Müller J (2009b) Plant defence - what can we learn from clubroots? Australasian Plant Pathology 38, 318-324.

Ludwig-Müller J, Jülke S, Geiss K, Richter F, Mithöfer A, Šola I, Rusak G, Keenan S, Bulman S (2015) A novel methyltransferase from the intracellular pathogen *Plasmodiophora brassicae* methylates salicylic acid. Molecular Plant Pathology 16, 349-364.

Malinowski R, Novak O, Borhan MH, Spichal L, Strnad M, Stephen A. Rolfe SA (2016) The role of cytokinins in clubroot disease. European Journal of Plant Pathology 145, 543-557.

Manzanares-Dauleux MJ, Barret P, Thomas G (2000) Development of a pathotype specific SCAR marker in *Plasmodiophora brassicae.* European Journal of Plant Pathology 106, 781-787.

Molmeret M, Horn M, Wagner M, Santic M, Abu Kwaik A (2005) Amoebae as training grounds for intracellular bacterial pathogens. Applied Environmental Microbiology 71, 20-28.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Molecular Biology and Evolution 32, 268-274.

Nosanchuk JD, Casadevall A (2003) The contribution of melanin to microbial pathogenesis. Cellular Microbiology 5, 203-223.

Palmer GC, Jorth PA, Whiteley M (2013) The role of two *Pseudomonas aeruginosa* anthranilate synthases in tryptophan and quorum signal production. Microbiology 159, 959-969.

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods 8, 785-786.

Prigent S, Frioux C, Dittami SM, Thiele S, Larhlimi A, Collet G, Gutknecht F, Got J, Eveillard D, Bourdon J, Plewniak F, Tonon T, Siegel A (2017) Meneco, a topology-based gap-filling tool applicable to degraded genome-wide metabolic networks. PLoS Computational Biology 13, e1005276.

Rhee HJ, Kim E-J, Lee JK (2007) Physiological polyamines: simple primordial stress molecules. Journal of Cellular and Molecular Medicine 11, 685-703.

Rizk G, Gouin A, Chikhi R, Lemaitre C (2014) MindTheGap: integrated detection and assembly of short and long insertions. Bioinformatics 30, 3451-3457.

Rolfe SA, Strelkov SE, Links MG, Clarke WE, Robinson SJ, Djavaheri M, Malinowski R, Haddadi P, Kagale S, Parkin IAP, Taheri A, Borhan MH (2016) The compact genome of the plant pathogen *Plasmodiophora brassicae* is adapted to intracellular interactions with host *Brassica spp.* BMC Genomics, 17, 272.

Sablok G, Chen TW, Lee CC, Yang C, Gan RC, Wegrzyn JL, Porta NL, Nayak KC, Huang PJ, Varotto C, Tang P (2017) ChloroMitoCU: Codon patterns across organelle genomes for functional genomics and evolutionary applications. DNA Research 24, 327-332.

Schuller A, Kehr J, Ludwig-Müller J (2014) Laser microdissection coupled to transcriptional profiling of *Arabidopsis* roots inoculated by *Plasmodiophora brassicae* indicates a role for brassinosteroids in clubroot formation. Plant and Cell Physiology 55, 392-411.

Schwelm A, Fogelqvist J, Knaust A, Jülke S, Lilja T, Bonilla-Rosso G, Karlsson M, Shevchenko A, Dhandapani V, Choi SR, Kim HG, Park JY, Lim YP, Ludwig-Müller J, Dixelius C (2015) The *Plasmodiophora brassicae genome* reveals insights in its life cycle and ancestry of chitin synthases. Scientific Reports 5, 11153.

Schwelm A, Dixelius C, Ludwig-Müller J (2016) New kid on the block – the clubroot pathogen genome moves the plasmodiophorids into the genomic era. European Journal of Plant Pathology 145, 531-542.

Shah P, Nanduri B, Swiatlo E, Ma Y, Pendarvis K (2011) Polyamine biosynthesis and transport mechanisms are crucial for fitness and pathogenesis of *Streptococcus pneumoniae*. Microbiology 157, 504-515.

Siemens J, Keller I, Sarx J, Kunz S, Schuller A, Nagel W, Schmülling T, Parniske M, Ludwig-Müller J (2006) Transcriptome analysis of *Arabidopsis* clubroots Indicate a key role for cytokinins in disease development. Molecular Plant-Microbe Interactions 19, 480-494.

Somé A, Manzanares MJ, Laurens F, Baron F, Thomas G, Rouxel F (1996) Variation for virulence on *Brassica napus L* amongst *Plasmodiophora brassicae* collections from France and derived single-spore isolates. *Plant Pathology* 45, 432-439.

Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, Ver Loren van Themaat E, Brown JK, Butcher SA, Gurr SJ, Lebrun MH, Ridout CJ, Schulze-Lefert P, Talbot NJ, Ahmadinejad N, Ametz C, Barton GR, Benjdia M, Bidzinski P, Bindschedler LV, Both M, Brewer MT, Cadle-Davidson L, Cadle-Davidson MM, Collemare J, Cramer R, Frenkel O, Godfrey D, Harriman J, Hoede C, King BC, Klages S, Kleemann J, Knoll D, Koti

PS, Kreplak J, López-Ruiz FJ, Lu X, Maekawa T, Mahanil S, Micali C, Milgroom MG, Montana G, Noir S, O'Connell RJ, Oberhaensli S, Parlange F, Pedersen C, Quesneville H, Reinhardt R, Rott M, Sacristán S, Schmidt SM, Schön M, Skamnioti P, Sommer H, Stephens A, Takahara H, Thordal-Christensen H, Vigouroux M, Wessling R, Wicker T, Panstruga R (2010) Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science 330, 1543-1546.

Supek F, Bosnjak M, Skunca N, Smuc T (2011) REVIGO summarizes and visualizes long lists of Gene Ontology terms. PLoS One 6, e21800.

Tanifuji G, Archibald JM, Hashimoto T (2016) Comparative genomics of mitochondria in chlorarachniophyte algae: endosymbiotic gene transfer and organellar genome dynamics. Scientific Reports 6, 21016.

The Gene Ontology Consortium (2017) Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Research 45, D331-D338.

The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Research 45, D158-D169.

Thomas T, Thomas TJ (2001) Polyamines in cell growth and cell death: molecular mechanisms and therapeutic applications. Cellular and Molecular Life Sciences 58, 244-258.

Toledo AV, Franco MEE, Lopez SMY, Troncozo MI, Saparrat MCN, Balatti PA (2017) Melanins in fungi: types, localization and putative biological roles. Physiological and Molecular Plant Pathology 99, 2-6.

Turra D, Segorbe D, Di Pietro A (2014) Protein kinases in plant-pathogenic fungi: conserved regulators of infection. Annual Review of Phytopathology 52, 267-288.

Urano K, Yoshiba Y, Nanjo T, Igarashi Y, Seki M, Sekiguchi F, Yamaguchi-Shinozaki K, Shinozaki K (2003) Characterization of Arabidopsis genes involved in biosynthesis of polyamines in abiotic stress responses and developmental stages. Plant, Cell & Environment 26, 1917-1926.

Valdes-Santiago L, Cervantes-Chavez JA, Leon-Ramırez CG, Ruiz-Herrera J (2012) Polyamine metabolism in fungi with emphasis on phytopathogenic species. Journal of Amino Acids 2012, 13 pages.

Veltri D, Malapi-Wight MM, Crouch JA (2016) SimpleSynteny: a web-based tool for visualization of microsynteny across multiple species. Nucleic Acids Research 44 (Web Server issue), W41-W45.

Wagner G, Charton S, Lariagon C, Laperche A, Lugan R, Hopkins J, Frendo P, Bouchereau A, Delourme R, Gravot A, Manzanares-Dauleux MJ (2012) Metabotyping: A new approach to investigate rapeseed (Brassica napus L.) genetic diversity in the metabolic response to clubroot infection. Molecular Plant-Microbe Interactions 25, 1478-1491.

Walters DR (2003) Polyamines and plant disease. Phytochemistry 64, 97–107.

Walters DR, Shuttleton MA (1985) Polyamines in the roots of turnip infected with *Plasmodiophora brassicae* Wor. New Phytologist 100, 209-214.

Xia X (2018) DAMBE7: New and improved tools for data analysis in molecular biology and evolution. Molecular Biology and Evolution 35, 1550-1552.

Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Research 40, W445-W451.

Zhang W-J, Zhou J, Li Z-F, Wang L, Gu X, Zhong Y (2007) Comparative analysis of Codon Usage patterns among mitochondrion, chloroplast and nuclear genes in *Triticum aestivum* L. Journal of Integrative Plant Biology 49, 246-254.

Table 1. Genome statistics of *P. brassicae* assemblies.

BUSCO analyses were done on genome sequences with the EnsemblProtists database.

Isolates	eH	e3 (Schwelm et al., 2015)	ZJ-1 (Bi et al., 2016)
Assembly accession	POCA00000000.1	GCA_001049375.1	GCA_002093825.1
Assembly assessment			
Genome assembly size	24,559,637 nt	24,050,826 nt	24,831,078 nt
Number of assembled scaffolds	136	165	667
Mean scaffold size	180,585 nt	145,762 nt	37,228 nt
Median scaffold size	11,351 nt	36,592 nt	2,588 nt
Minimal scaffold size	509 nt	1,868 nt	2,001 nt
Maximal scaffold size	1,480,576 nt	1,189,627 nt	1,479,440 nt
N50	741,293 nt	472,887 nt	510,270 nt
Genomic G + C content	59.29 %	59.37%	59.38%
Busco 3.0.2			

Complete and single-copy	192	191	184
Complete and duplicated	3	3	10
Fragmented	3	3	3
Missing	17	18	18
Annotation statistics			
Number of protein coding genes	12,590		
Average gene size	1,398 nt		
Average exon size	335 nt		
Average exon number per gene	3.65		
Average intron size	66 nt		
Average intron number per gene	2.65		
Genes with 5' UTR	25%		
Genes with 3' UTR	35%		
Gene with transcript evidence	6736		
Gene without transcript evidence	6065		

Figure 1. Visual representation of the eH genome first 43 scaffolds.

In the first outer circle (A) are represented the first 43 scaffolds of the assembled eH genome that correspond to 95% of the total genome sequence. In the second and third outer circles (B and C), the percentage in GC and the gene density (calculated as the number of genes every 20kb) are indicated, respectively. In the two following inner circles, is represented the gene similarity between the isolates eH and e3 (Schwelm et al., 2015): D circle pictures the relative rate every 10kb of the eH genes number to the eH genes number that have similarity to e3 genes (80% threshold); E circle displays this similarity as lines when genes from eH and e3 have 80% similarity. The inner circle shows the gene similarity between the isolates eH and ZJ-1 (E, purple lines; Bi et al., 2016), and eH and Led09 (E, green lines; Feng et al., 2013). The grey scaffold 43 represented the mitochondrial genome.

Figure 2. Circular DNA map of the mitochondrial genome of *Plasmodiophora brassicae*.

The outer circle shows the position of protein coding genes with exons represented in green and introns in grey, and the position of ribosomal potein coding genes in orange. The inner rings depicts the position of structural RNAs: rRNA and one ncRNA (rnpB) in blue, and tRNA in red.

Figure 3. Syntenic comparison between mitochondrial genomes.

The names of the genes are indicated. Lines connecting regions of sequence similarity are drawn. Dotted lines show missing sequences.

Figure 4. Maximum likelihood phylogenetic tree of the *COX1* mitochondrial sequences.

The Alveolates cluster was used to root the tree. Bootstrap values are shown for each node. The scale bar corresponds to 0.3 substitutions per site.

Figure 5. Assignment of genes in different annotation categories.

GO, Gene Ontology term (Blast2GO); EC, Enzyme Commission number (Blast2GO and ECDomainMiner); IP, Interpro term (Blast2GO and Gene Ontology Database). The terms in which at least 80 genes were classified are represented.

Figure 6. A view of ClubrootCyc homepage.

Its mains features and functionalities, such as the structure of each scaffold , the gene local context, the pathway genome database overview, and the toolbar for several searchs are displayed in the Figure S4.

The complete pathway genome database is available at http://pathway-tools.toulouse.inra.fr/organism-summary?object=CLUBROOT

- Sequence report of an additional genome of a *P. brassicae* isolate
- Identification of the mitochondrial genome sequence
- Refinement of the annotation
- Prediction of new pathways
- Creation of a *P. brassicae* pathway genome database

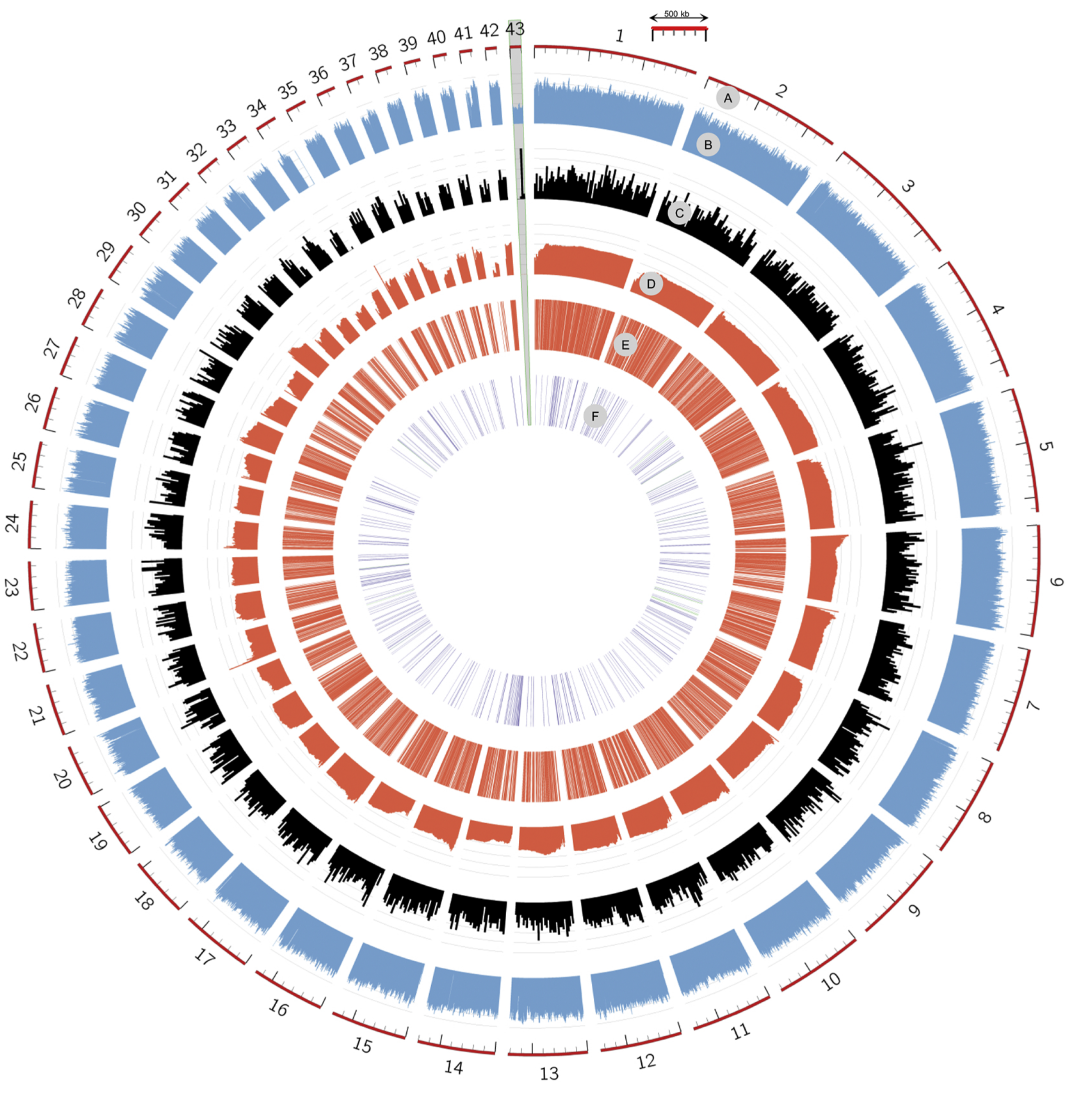

500 kb
A
B
C
D
E
F
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43

Figure 1

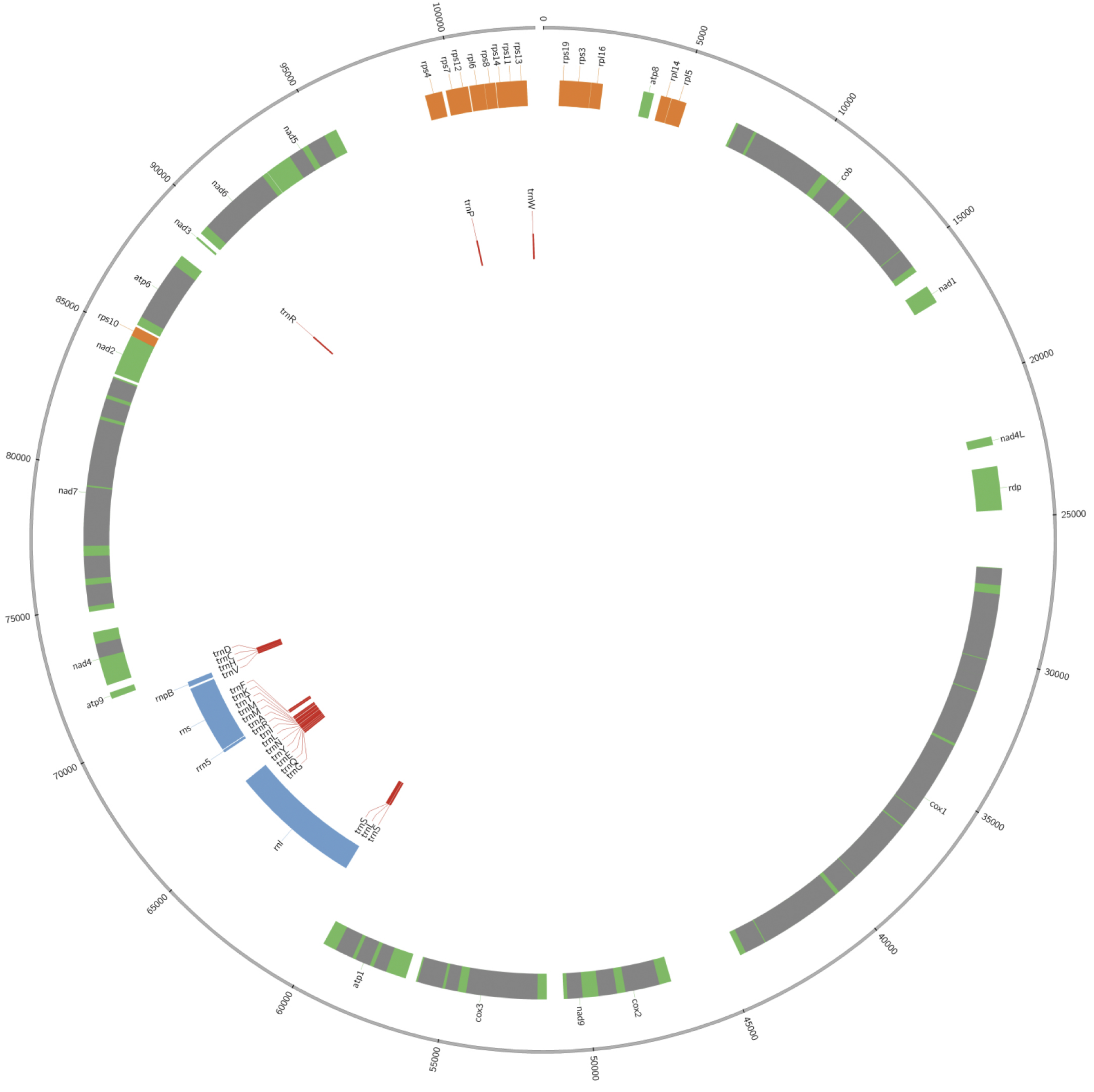

0
5000
10000
15000
20000
25000
30000
35000
40000
45000
50000
55000
60000
65000
70000
75000
80000
85000
90000
95000
100000
rps19
rps3
rpl16
atp8
rpl14
rpl5
cob
nad1
nad4L
rdp
cox1
cox2
nad9
cox3
atp1
rnl
trnS
trnL
trnS
rrn5
rns
rnpB
trnF
trnK
trnT
trnM
trnM
trnA
trnR
trnI
trnL
trnN
trnY
trnE
trnQ
trnG
trnD
trnC
trnH
trnV
atp9
nad4
nad7
nad2
rps10
atp6
nad3
nad6
nad5
trnR
trnP
trnW
rps4
rps7
rps12
rpl6
rps8
rps14
rps11
rps13

Figure 2

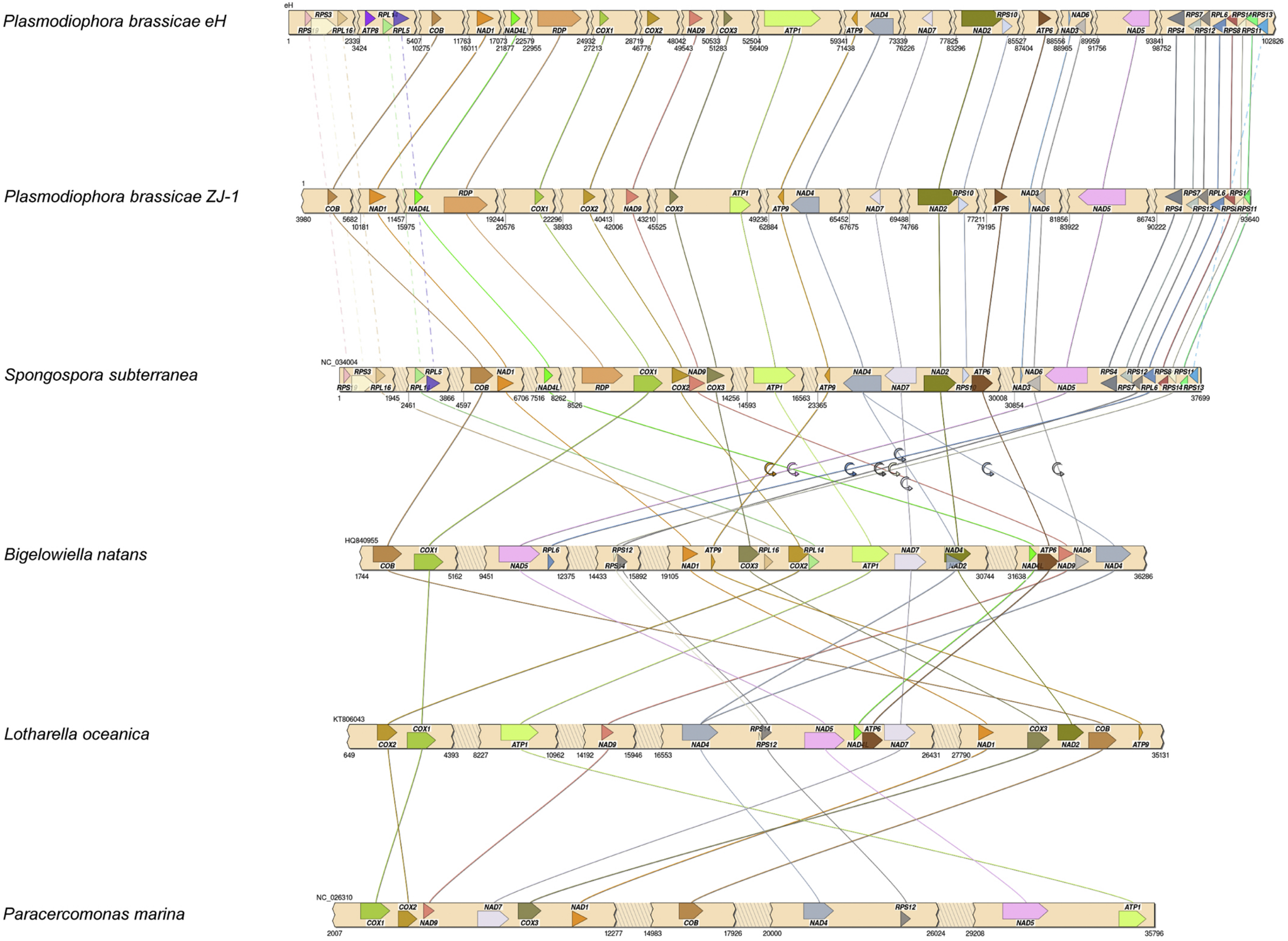

Plasmodiophora brassicae eH
eH
RPS3 RPS19 RPL16 ATP8 RPL14 RPL5 COB NAD1 NAD4L RDP COX1 COX2 NAD9 COX3 ATP1 ATP9 NAD4 NAD7 NAD2 RPS10 ATP6 NAD3 NAD6 NAD5 RPS4 RPS7 RPS12 RPL6 RPS8 RPS14 RPS11 RPS13
1 2339 3424 5407 10275 11763 16011 17073 21877 22579 22955 24932 27213 28719 46776 48042 49543 50533 51283 52504 56409 59341 71438 73339 76226 77825 83296 85527 87404 88556 88965 89959 91756 93841 98752 102826
Plasmodiophora brassicae ZJ-1
1
COB NAD1 NAD4L RDP COX1 COX2 NAD9 COX3 ATP1 ATP9 NAD4 NAD7 NAD2 RPS10 ATP6 NAD3 NAD6 NAD5 RPS4 RPS7 RPS12 RPL6 RPS8 RPS14 RPS11
3980 5682 10181 11457 15975 19244 20576 22296 38933 40413 42006 43210 45525 14256 49236 62884 65452 67675 69488 74766 77211 79195 81856 83922 86743 90222 93640
Spongospora subterranea
NC_034004
RPS19 RPS3 RPL16 RPL14 RPL5 COB NAD1 NAD4L RDP COX1 COX2 NAD9 COX3 ATP1 ATP9 NAD4 NAD7 NAD2 RPS10 ATP6 NAD3 NAD6 NAD5 RPS4 RPS7 RPS12 RPL6 RPS8 RPS14 RPS11 RPS13
1 1945 2461 3866 4597 6706 7516 8262 8526 14256 14593 16563 23365 30008 30854 37699
Bigelowiella natans
HQ840955
COB COX1 NAD5 RPL6 RPS12 RPS14 NAD1 ATP9 COX3 RPL16 COX2 RPL14 ATP1 NAD7 NAD4 NAD2 NAD4L ATP6 NAD9 NAD6 NAD4
1744 5162 9451 12375 14433 15892 19105 30744 31638 36286
Lotharella oceanica
KT806043
COX2 COX1 ATP1 NAD9 NAD4 RPS14 RPS12 NAD5 NAD4L ATP6 NAD7 NAD1 COX3 NAD2 COB ATP9
649 4393 8227 10962 14192 15946 16553 26431 27790 35131
Paracercomonas marina
NC_026310
COX1 COX2 NAD9 NAD7 COX3 NAD1 COB NAD4 RPS12 NAD5 ATP1
2007 12277 14983 17926 20000 26024 29208 35796

Figure 3

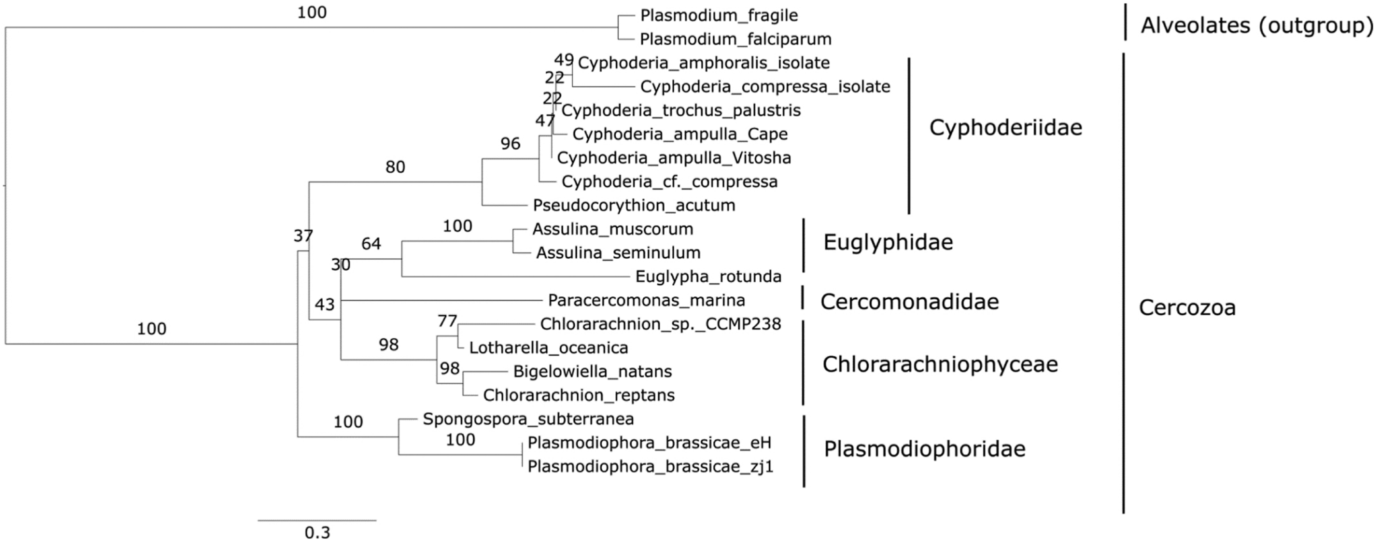

100
Plasmodium_fragile
Plasmodium_falciparum
Alveolates (outgroup)
49
Cyphoderia_amphoralis_isolate
22
Cyphoderia_compressa_isolate
22
Cyphoderia_trochus_palustris
47
Cyphoderia_ampulla_Cape
96
Cyphoderia_ampulla_Vitosha
80
Cyphoderia_cf._compressa
Pseudocorythion_acutum
Cyphoderiidae
100
Assulina_muscorum
37
64
Assulina_seminulum
Euglyphidae
30
Euglypha_rotunda
43
Paracercomonas_marina
Cercomonadidae
Cercozoa
77
Chlorarachnion_sp._CCMP238
100
98
Lotharella_oceanica
Chlorarachniophyceae
98
Bigelowiella_natans
Chlorarachnion_reptans
100
Spongospora_subterranea
100
Plasmodiophora_brassicae_eH
Plasmodiophora_brassicae_zj1
Plasmodiophoridae
0.3

Figure 4

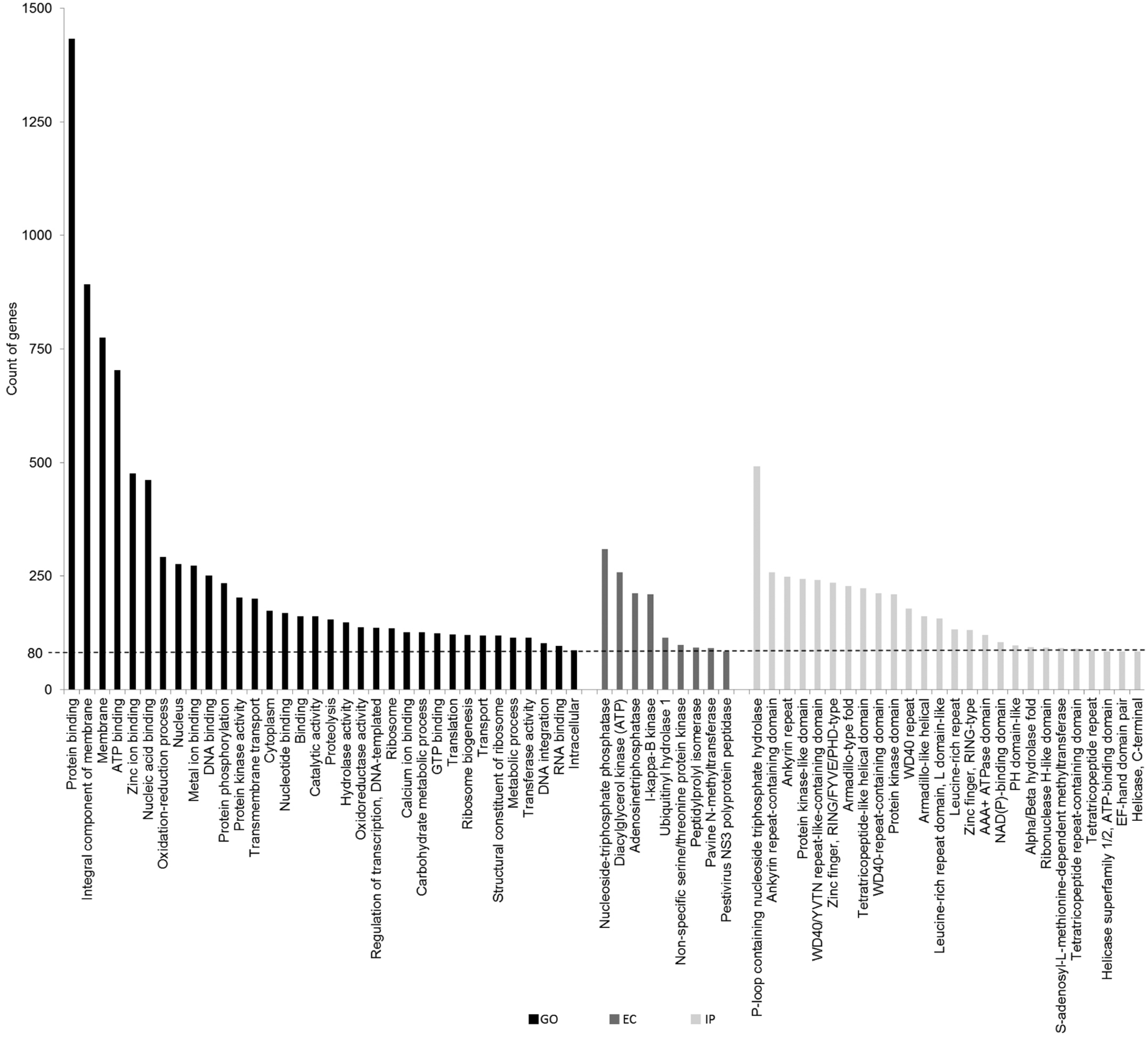
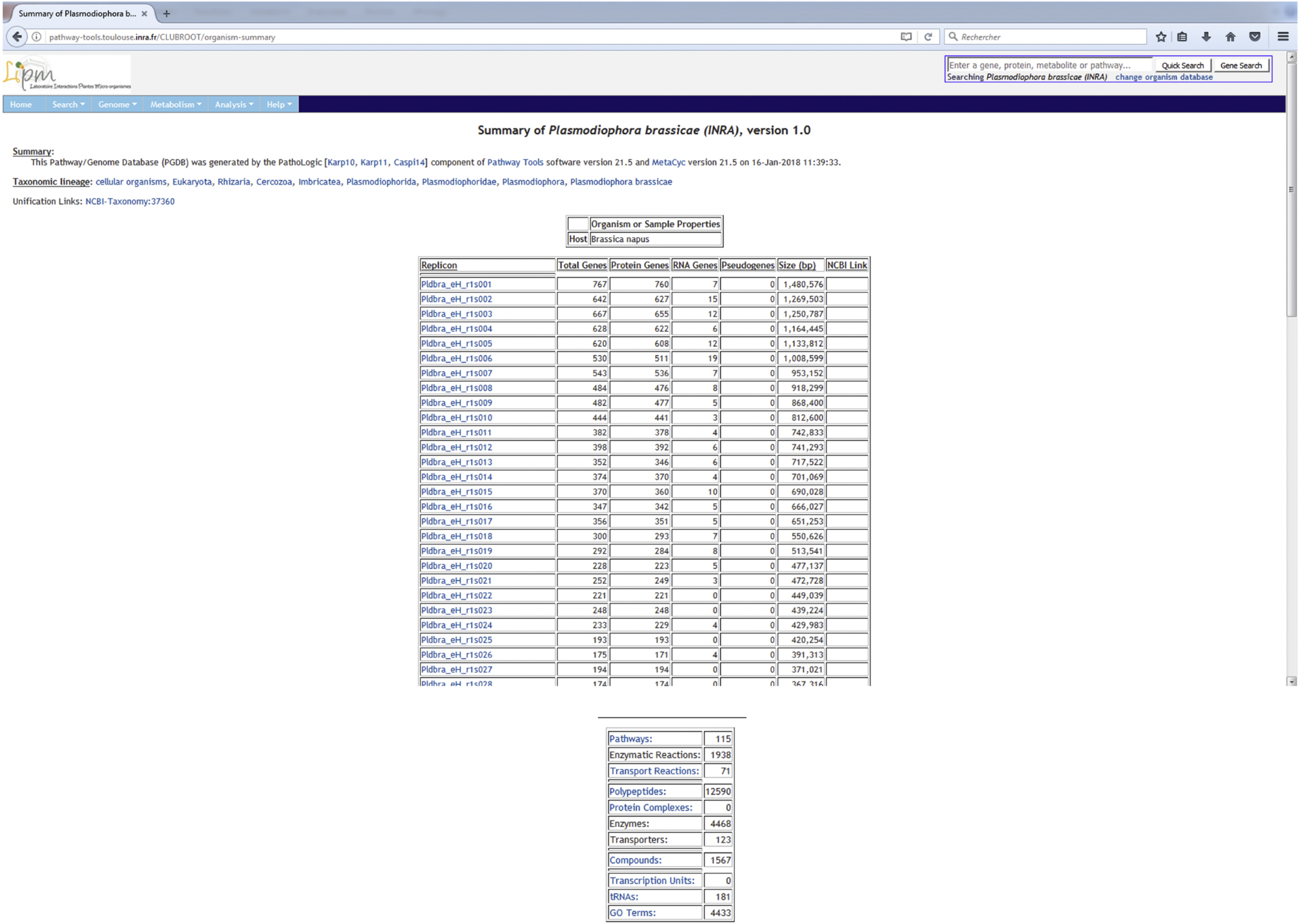

1500
1250
1000
750
500
250
80
0
Count of genes
Protein binding
Integral component of membrane
Membrane
ATP binding
Zinc ion binding
Nucleic acid binding
Oxidation-reduction process
Nucleus
Metal ion binding
DNA binding
Protein phosphorylation
Protein kinase activity
Transmembrane transport
Cytoplasm
Nucleotide binding
Binding
Catalytic activity
Proteolysis
Hydrolase activity
Oxidoreductase activity
Regulation of transcription, DNA-templated
Ribosome
Calcium ion binding
Carbohydrate metabolic process
GTP binding
Translation
Ribosome biogenesis
Transport
Structural constituent of ribosome
Metabolic process
Transferase activity
DNA integration
RNA binding
Intracellular
Nucleoside-triphosphate phosphatase
Diacylglycerol kinase (ATP)
Adenosinetriphosphatase
I-kappa-B kinase
Ubiquitinyl hydrolase 1
Non-specific serine/threonine protein kinase
Peptidylprolyl isomerase
Pavine N-methyltransferase
Pestivirus NS3 polyprotein peptidase
P-loop containing nucleoside triphosphate hydrolase
Ankyrin repeat-containing domain
Ankyrin repeat
Protein kinase-like domain
WD40/YVTN repeat-like-containing domain
Zinc finger, RING/FYVE/PHD-type
Armadillo-type fold
Tetratricopeptide-like helical domain
WD40-repeat-containing domain
Protein kinase domain
WD40 repeat
Armadillo-like helical
Leucine-rich repeat domain, L domain-like
Leucine-rich repeat
Zinc finger, RING-type
AAA+ ATPase domain
NAD(P)-binding domain
PH domain-like
Alpha/Beta hydrolase fold
Ribonuclease H-like domain
S-adenosyl-L-methionine-dependent methyltransferase
Tetratricopeptide repeat-containing domain
Tetratricopeptide repeat
Helicase superfamily 1/2, ATP-binding domain
EF-hand domain pair
Helicase, C-terminal
GO
EC
IP

Figure 5


Summary of Plasmodiophora b...
pathway-tools.toulouse.inra.fr/CLUBROOT/organism-summary
Rechercher
Lipm Laboratoire Interactions Plantes Micro-organismes
Enter a gene, protein, metabolite or pathway...
Quick Search
Gene Search
Searching Plasmodiophora brassicae (INRA) change organism database
Home
Search
Genome
Metabolism
Analysis
Help

Summary of *Plasmodiophora brassicae (INRA)*, version 1.0

Summary:
This Pathway/Genome Database (PGDB) was generated by the PathoLogic [Karp10, Karp11, Caspi14] component of Pathway Tools software version 21.5 and MetaCyc version 21.5 on 16-Jan-2018 11:39:33.

Taxonomic lineage: cellular organisms, Eukaryota, Rhizaria, Cercozoa, Imbricatea, Plasmodiophorida, Plasmodiophoridae, Plasmodiophora, Plasmodiophora brassicae

Unification Links: NCBI-Taxonomy:37360

	Organism or Sample Properties
Host	Brassica napus

Replicon	Total Genes	Protein Genes	RNA Genes	Pseudogenes	Size (bp)	NCBI Link
Pldbra_eH_r1s001	767	760	7	0	1,480,576	
Pldbra_eH_r1s002	642	627	15	0	1,269,503	
Pldbra_eH_r1s003	667	655	12	0	1,250,787	
Pldbra_eH_r1s004	628	622	6	0	1,164,445	
Pldbra_eH_r1s005	620	608	12	0	1,133,812	
Pldbra_eH_r1s006	530	511	19	0	1,008,599	
Pldbra_eH_r1s007	543	536	7	0	953,152	
Pldbra_eH_r1s008	484	476	8	0	918,299	
Pldbra_eH_r1s009	482	477	5	0	868,400	
Pldbra_eH_r1s010	444	441	3	0	812,600	
Pldbra_eH_r1s011	382	378	4	0	742,833	
Pldbra_eH_r1s012	398	392	6	0	741,293	
Pldbra_eH_r1s013	352	346	6	0	717,522	
Pldbra_eH_r1s014	374	370	4	0	701,069	
Pldbra_eH_r1s015	370	360	10	0	690,028	
Pldbra_eH_r1s016	347	342	5	0	666,027	
Pldbra_eH_r1s017	356	351	5	0	651,253	
Pldbra_eH_r1s018	300	293	7	0	550,626	
Pldbra_eH_r1s019	292	284	8	0	513,541	
Pldbra_eH_r1s020	228	223	5	0	477,137	
Pldbra_eH_r1s021	252	249	3	0	472,728	
Pldbra_eH_r1s022	221	221	0	0	449,039	
Pldbra_eH_r1s023	248	248	0	0	439,224	
Pldbra_eH_r1s024	233	229	4	0	429,983	
Pldbra_eH_r1s025	193	193	0	0	420,254	
Pldbra_eH_r1s026	175	171	4	0	391,313	
Pldbra_eH_r1s027	194	194	0	0	371,021	
Pldbra_eH_r1s028	174	174	0	0	367,316	

Pathways:	115
Enzymatic Reactions:	1938
Transport Reactions:	71
Polypeptides:	12590
Protein Complexes:	0
Enzymes:	4468
Transporters:	123
Compounds:	1567
Transcription Units:	0
tRNAs:	181
GO Terms:	4433

Figure 6