



# Metagenomics approach for *Polymyxa betae* genome assembly enables comparative analysis towards deciphering the intracellular parasitic lifestyle of the plasmodiophorids

Alain Decroës<sup>a,\*</sup>, Jun-Min Li<sup>b</sup>, Lorna Richardson<sup>c</sup>, Euphemia Mutasa-Gottgens<sup>c,d</sup>, Gipsi Lima-Mendez<sup>e</sup>, Mathieu Mahillon<sup>a,1</sup>, Claude Bragard<sup>a</sup>, Robert D. Finn<sup>c</sup>, Anne Legrève<sup>a,\*</sup>

<sup>a</sup> Phytopathology-Applied Microbiology, Earth and Life Institute, UCLouvain, Louvain-la-Neuve, Belgium

<sup>b</sup> State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Key Laboratory of Biotechnology in Plant Protection of Ministry of Agriculture and Zhejiang Province, Institute of Plant Virology, Ningbo University, 315211 Ningbo, China

<sup>c</sup> European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, UK

<sup>d</sup> University of Hertfordshire, Hatfield, Herts AL10 9AB, UK

<sup>e</sup> Louvain Institute of Biomolecular Science and Technology, UCLouvain, Louvain-la-Neuve, Belgium

## ARTICLE INFO

### Keywords:

*Polymyxa betae*

Rhizaria

Plasmodiophora

*Spongospora*

Ankyrin repeats

CatSper

## ABSTRACT

Genomic knowledge of the tree of life is biased to specific groups of organisms. For example, only six full genomes are currently available in the rhizaria clade. Here, we have applied metagenomic techniques enabling the assembly of the genome of *Polymyxa betae* (Rhizaria, Plasmodiophorida) RES F41 isolate from unpurified zoospore holobiont and comparison with the A26–41 isolate. Furthermore, the first *P. betae* mitochondrial genome was assembled. The two *P. betae* nuclear genomes were highly similar, each with just ~10.2 k predicted protein coding genes, ~3% of which were unique to each isolate. Extending genomic comparisons revealed a greater overlap with *Spongospora subterranea* than with *Plasmodiophora brassicae*, including orthologs of the mammalian cation channel sperm-associated proteins, raising some intriguing questions about zoospore physiology. This work validates our metagenomics pipeline for eukaryote genome assembly from unpurified samples and enriches plasmodiophorid genomics; providing the first full annotation of the *P. betae* genome.

## 1. Introduction

The Plasmodiophorida (plasmodiophorids), is an eukaryotic order of mainly strict plant endoparasites, previously considered as fungi and now classified in the Phytomyxea embedded within the Cercozoa in the Rhizaria protist supergroup ([1]; [157]; [66]). They were discovered early in the 19th century with the description of *Plasmodiophora brassicae* in 1877 by Woronin [153] and, in spite of this early scientific interest, this group remains poorly studied. Research on their interactions with plants and on virus transmission under controlled conditions is complicated by their strict biotrophic endoparasitic habit and continues

mainly in the context of agricultural plant hosts.

Four of the known plasmodiophorids species are considered to be of economic significance in agriculture. *P. brassicae* is responsible for clubroot disease in Brassica crops (e.g. canola, cabbage). Infected plants develop root galls and are stunted, leading to significant economic losses [123]. *Spongospora subterranea* infects potatoes causing powdery scab, characterized by small circular lesions on the tuber filled with a powdery mass of spores. *S. subterranea* also causes galls on roots and stolons and transmits the *Potato mop-top virus* (PMTV), a damaging virus in potato crops [103,143]. *Polymyxa graminis* symptomlessly invades roots of numerous Poaceae species but becomes damaging as a virus vector,

**Abbreviations:** AAI, amino acid identity; ANI, average nucleotide identity; Anks, proteins containing ankyrin repeats; CatSper, cation channel sperm-associated proteins; COG, clusters of orthologous groups; EC, enzyme codes; EST, Expressed Sequence Tags; ER, Endoplasmic reticulum; GO, Gene ontology; GPCR, G protein-coupled receptor; GPI, Glycosylphosphatidylinositol; KO, KEGG Ortholog; LSU, large ribosomal subunit; MAG, metagenome-assembled genome; ncRNA, non-coding RNA; PR, pathogenesis-related protein; QS, quality score; SAR, Stramenopiles, Alveolata and Rhizaria; SEA, Singular Enrichment Analyses; SP, signal peptide; SSU, small ribosomal subunit; TF, transcription factors; TM, transmembrane domains.

\* Corresponding authors.

E-mail addresses: [alain.decroes@uclouvain.be](mailto:alain.decroes@uclouvain.be) (A. Decroës), [anne.legreve@uclouvain.be](mailto:anne.legreve@uclouvain.be) (A. Legrève).

<sup>1</sup> Present affiliation of MM: Plant Protection Department, Agroscope, 1260 Nyon, Switzerland.

<https://doi.org/10.1016/j.ygeno.2021.11.018>

Received 29 March 2021; Received in revised form 24 June 2021; Accepted 10 November 2021

Available online 16 November 2021

0888-7543/© 2021 The Authors.

Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

being capable of transmitting at least 14 plant viruses (Benyvirus, Bymovirus, Furovirus, Pecluvirus) to major crops worldwide such as rice, wheat, maize, barley, rye, oat, triticale, sugarcane and peanut [48,134]. *Polymyxa betae* has been reported in all sugar beet (*Beta vulgaris* subsp. *vulgaris*) production regions. Even though, like *P. graminis*, it is widely accepted that *P. betae* is not pathogenic to its host, in sugar beet crops it is responsible for the transmission of the *Beet necrotic yellow vein virus* (BNYVV). This benyvirus is the causal agent of rhizomania syndrome, a devastating disease in susceptible sugar beet. *P. betae* also transmit another benyvirus, the *Beet soil-borne mosaic virus* (BSBMV) and two pomoviruses, the *Beet soil-borne virus* (BSBV) and the *Beet virus Q* (BVQ) which are often associated with the rhizomania syndrome [58,104]. *P. brassicae* and *S. subterranea* are both associated with cell proliferation and the formation of hypertrophic cells in the host plants [62,100] whereas roots of plants infected by non-viruliferous *Polymyxa* spp. generally appear normal. *P. betae* and *S. subterranea* share the ability to transmit plant viruses to their hosts while no virus transmission has yet been reported for *P. brassicae* [122].

Until 2000, most research on plasmodiophorids consisted of ultrastructural and karyological observations as well as rDNA-based and immuno-based detection and co-occurrence with viruses [18,90,93]. In the early 2000s, Mutasa-Gottgens et al. [108] generated 11 *P. betae* Expressed Sequence Tags (EST) with phagemid library of cDNA clones from infected sugar beet roots. A few years later, Bulman and colleagues used a suppression subtractive hybridization strategy to identify 76 new *P. brassicae* genes [20,21]. The same strategy was applied for *P. graminis*, generating four EST [138] and for *P. betae*, generating 76 *P. betae* EST as well as 120 sugar beet ESTs up-regulated during the different stages of the infection process [44]. The rapid evolution of sequencing technologies brought new perspectives in plasmodiophorids research starting with the release of the *Arabidopsis thaliana* annotated genome, which, as a susceptible host plant, promoted *P. brassicae*-plant interactions studies [46,99,101,130,131]. For example, Siemens et al. [131] used a transcriptomics approach to study regulation of defense genes in *A. thaliana* during compatible interaction with *P. brassicae*. Cao et al., [31] studied reactive oxygen species metabolism in *Brassica napus* infected with *P. brassicae*. Their findings suggest a fine-grained molecular interplay between plasmodiophorids and their host, probably by means of direct effector secretion from intracellular plasmodia to their host's cytoplasm but also by means of membrane proteins such as receptors and transporters. Subsequently, Burki et al. [23] used 454 pyrosequencing technology to sequence cDNA from five rhizarian species including *P. brassicae* and *S. subterranea*. In the last five years, the release of genomes of *P. brassicae* in 2015 [123,124] *S. subterranea* in 2018 [37] and *P. betae* in 2019 [42] facilitated transcriptomic studies oriented towards plant-plasmodiophorids interactions, especially *P. brassicae* [10,11,38,120,128]. Several *P. brassicae* isolates have now been (re)sequenced [11,41,120,125,137]. Before 2015, the lack of enough plasmodiophorid genomic information naturally led scientists to focus on the plant side of the interaction [109].

The sequencing effort granted to the Rhizaria supergroup remains small by comparison to other eukaryotic groups [22,129]. In addition to plasmodiophorids only three genomes from other rhizaria protists are available: the Chlorarachnea *Bigelowiella natans* [39], the Foraminifera *Reticulomyxa filosa* [60] and the Imbricatea *Paulinella micropora* (accession no.: PRJDB3528). More protist genomes are needed for comparative analyses that fully encompass the eukaryotic diversity and the microbial biosphere [129]. This would be expected to provide new clues on evolution and adaptation of eukaryotes, and to improve knowledge of the functional repertoire of eukaryotic genomes. The great majority of sequenced eukaryotic genomes belong to animals, plants and fungi, representing only two of the eight main eukaryotic groups summarized by Burki et al. [24].

Although methods for purification of plasmodiophorid resting spores have been developed, the sequencing and assembly of plasmodiophorid genomes remain challenging because, whatever the technical approach

chosen, the purified samples are never free of host and other contaminating microorganisms [16,43,115,138,150,151]. This is especially true for *Polymyxa* spp. which, compared to *P. brassicae* and *S. subterranea*, do not produce structures similar to galls filled with numerous resting spores which help to bulk protist DNA for extraction. Moreover, isolation and purification of *Polymyxa* spp. sporosori have proven to be difficult due to the variation of their size and shape. Additional confounders include the possibility that some “contaminant” microorganisms might be able to develop within hatched individual spores in sporosori [34] or that DNA from the host could pass to the plasmodia [16]. Despite the delays caused by these challenges, a first draft genome of *P. betae* was recently recovered through a metagenomic approach using DNA obtained from sporosori of the monosporosorus *P. betae* A26–41 isolate grown in sugar beet in a soil-free hydroponic system [42].

The objectives of this new study were to validate the metagenomics approach for recovering a new *P. betae* genome from a geographically distinct isolate (RES F41; DNA sequence generated from zoospores contaminated with bacteria), and to carry out some comparative analysis with *P. betae* isolate A26–41 and exiting genomes from other plasmodiophorids (*P. brassicae* and *S. subterranea*). To achieve this, a bioinformatic pipeline including Metaspades, Metawrap and EukCC softwares was applied to assemble and isolate two *P. betae* genomes from the RES F41 unpurified zoospore holobiont and the A26–41 partially purified resting spores sample. The new genomes were used to carry out some comparative sequence analysis. This study is the first to provide a full nuclear genome annotation for *P. betae*, and to also present the first *P. betae* mitochondrial genome.

## 2. Materials and methods

### 2.1. Construction of two *Polymyxa betae* draft genomes

#### 2.1.1. *P. betae* isolates origin and DNA extraction

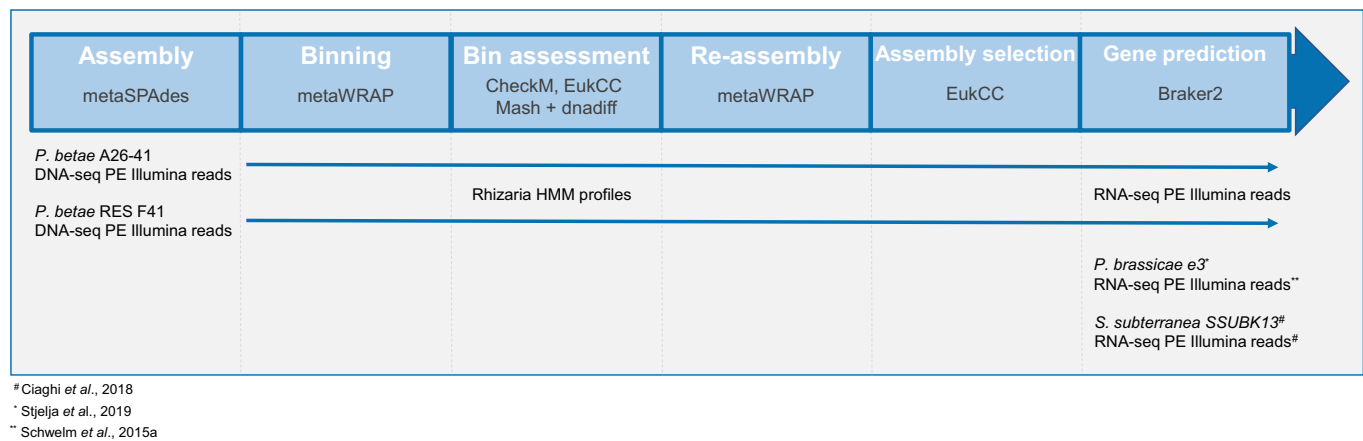
Two *P. betae* isolates were used as starting material: the Belgian isolate A26–41 [91] and the UK isolate RES F41 [107]. The methodology followed to obtain high-quality DNA extract from each isolate is as described in the respective papers.

#### 2.1.2. Libraries preparation and sequencing

The methodology used to prepare library and sequence A26–41 isolate is described in Decroës et al. [42]. For the RES F41 isolate, paired-end libraries were prepared from fragmented dA-tailed DNA. After addition of dT-tailed DNA adaptors, the ligated products were purified and whole-genome sequencing was performed using the Illumina paired-end HiSeq 2000 platform with a sequencing length of 2 × 150 bp (Novogene, Beijing, China).

#### 2.1.3. Genome assembly and cleaning

The obligate nature of *P. betae* excludes the option of a large-scale axenic culture [45]. Contaminating DNA from bacteria, protozoa, algae, fungi is unavoidable even after extensive purification. Both sample sequence datasets were independently processed using the same bioinformatic workflow (see Fig. 1) aiming to assemble reads and filter *P. betae* contigs from contaminating DNA sequences. Reads were assembled with the metagenome assembler metaSPAdes v3.12.0 [111] and assembled contigs were subsequently binned using the metaWRAP software [144]. MetaWRAP produces consensus bins based on three tools, namely MetaBAT 2.0, MaxBins 2.0 and CONCOCT, which are then assessed for completeness and contamination using CheckM [112]. However, as CheckM is optimized for bacterial genomes, the bins were assessed using a bespoke version of EukCC [121], a tool designed for microbial eukaryote genome quality estimation. Briefly, a bespoke single-copy marker gene reference set was constructed from five available Rhizaria genomes (GenBank accession no. GCA\_000320545.1, GCA\_001049375.1, GCA\_003833335.1, GCA\_000512085.1,



**Fig. 1.** Schematic representation of the workflow of metagenomics analysis. The softwares used in each step are visible in the boxes. Input data are given below the boxes.

GCA\_900404475.1), enabling the identification of 636 profile HMMs from PANTHER (v14.1), which were utilized by EukCC to estimate the completeness and contamination of the bins. Mitochondrial genome was re-assembled for both *P. betae* isolates with the assemblers MITObim 1.8 using MIRA 4.0.2 [65] and NOVOPlasty [47].

For comprehensiveness, bins were compared to a set of complete reference genomes of bacterial and unicellular eukaryotic genomes in RefSeq as of January 2018 and bespoke references from other Rhizaria species: *B. natans* (GenBank accession no. GCA\_000320545.1), *P. brassicae* (GenBank accession no. GCA\_003833335.1), *R. filosa* (GenBank accession no. GCA\_000512085.1), *S. subterranea* (GenBank accession no. GCA\_900404475.1). This was performed using a two-stage process, as described in [3]. First, Mash was used to find the best matching reference genome, followed by a whole-genome alignment between the query metagenome-assembled genome (MAG) and reference genome using dnadiff (part of the MUMmer package), to evaluate the alignment fraction and average nucleotide identity (ANI).

Bins from each isolate with the highest completeness/lowest contamination score based on the bespoke EukCC reference set were determined to represent *P. betae*. These were then reassembled using the *reassemble bins* module within metaWRAP, in order to produce the best quality assembly. Reassembled bins were compared against the original bins using the bespoke EukCC again, and the best assembly for each isolate (completeness, contamination, genome size and N50) was selected as the *P. betae* draft genome. Reads were mapped back to corresponding assemblies using Bowtie2 v2.3.4 [88] and per-contig average coverage was computed.

Overall ANI value between both draft genomes, fragmented with a window size of 1000 base pairs (bp) and a step size of 200 bp, were computed with the “ani.rb” ANI calculator [118,119]. Thresholds to consider alignments for ANI computation were a minimum length of 700 bp and identity of 0.7. Moreover, both draft genomes were compared using ProgressiveMauve v2.4.0 [40] with default parameters.

Low complexity sequences from both *P. betae* genomes were masked prior to any further analysis. *De novo* identification of novel repeat families was performed with RepeatModeler v-open-1.0.11 and both genomes were then softmasked using RepeatMasker with rmbblastn v2.2.27+. RepBase-20170127 and Dfam\_Consensus-20170127 databases were combined to the *de novo* identified repeats.

The remaining (non-*P. betae*) bins initially generated by metaWRAP for each isolate represented bacteria found in association with *P. betae*. These bins were assigned a quality score (QS) calculated as “completeness” - (5\* “contamination”). Bins with a QS > 50 are considered to be medium/high quality, and were assigned a taxonomic classification based on the Genome Taxonomy Database (GTDB) using GTDB-Tk [33].

## 2.2. *Polymyxa betae* RNA preparation, sequencing and data processing

Three-week-old sugar beet seedlings monogerm DH line KWS2320 were inoculated with a zoospore suspension of the isolate A26–41 and grown as previously described [91]. Heavily infected roots were harvested after four weeks, surface disinfected with 1 N NaOH and 0.1% SDS and rinsed three times with sterile water. Total RNA was then isolated using Trizol™ (Invitrogen, Carlsbad, CA) and DNase treated with RQ1 DNase (Promega, Madison, WI). Roots were first examined under bright field microscopy to confirm that all *P. betae* life stages, *i.e.* undifferentiated plasmodia, sporangia, sporogenous plasmodia, sporosori, were present. mRNA was purified from total RNA through Poly(A)-tail selection then fragmented, retro-transcribed with random primers and one library was prepared using the TruSeq stranded mRNA kit (Illumina, San Diego, CA). The library was paired-end sequenced using an Illumina NextSeq instrument and the “NextSeq High kit”, with a sequencing length of 2 × 150 bp. Raw reads were trimmed for adapters and low-quality sequences with bbdutk.sh [25] and quality checked with FastQC v.0.72 [4]. They were then mapped against the sugar beet draft genome KDHBv-1.0 [51] with Bowtie2 v2.3.4 [88] as a first filtering step to remove plant sequences. Unaligned read pairs were then mapped against the *P. betae* draft genomes with HISAT2 v2.1.0 [82].

## 2.3. Genome annotation

### 2.3.1. Gene predictions and annotation

The resulting HISAT2-created BAM-files and the softmasked *P. betae* draft genomes were used as input for gene prediction with the BRAKER2 v2.4.1 pipeline [15,68]. This takes advantage of GeneMark-ET v.4.48\_3.60 [97,140] and AUGUSTUS v3.3.3 [136] for RNA-seq-based genome annotation. The ProtHint pipeline v2.1.1 from Georgia Institute of Technology [14] was also used to generate genome-wide footprints of homologous proteins to assist gene prediction with BRAKER2. In our case, the reference protein set constituted 1,383,384 amino acid (aa) sequences originated from UniProtKB/TrEMBL (all rhizarian proteins, 20-01-2020) + UniProtKB/Swiss-Prot (all sequences, 20-01-2020) + Stramenopiles, Alveolata and Rhizaria (SAR) orthologs retrieved from the “protist” clusters node of OrthoDB protein database [84] if the corresponding clusters were represented in at least 2 SAR species. Gene prediction was performed for each *P. betae* draft genome, with and without using protein hints. Both sets of predicted genes (with and without protein hints), were compared with rhizarian proteins present in UniProtKB/TrEMBL using OrthoVenn2 [154]. This, along with visual inspection of gene models and aligned RNA-seq reads using Integrated Genomic Viewer v9.1.2 [117], allowed us to identify the gene set (generated with the protein hints) for further analysis. Protein

products of genes predicted by BRAKER2 without protein hints, which were absent from the dataset predicted with protein hints, were then compared to NCBI nr database through BLASTP searches [28]. Sequences showing >30% identity with deposited proteins from other SAR species were manually added to the gene set predicted with protein hints. Predicted proteins were scanned for matches against the InterPro v78.0 [105] protein signatures database using the InterProScan v5.42–78.0 tool [76]. EggNOG-mapper v2 [70] with the EggNOG v5.0 clusters and phylogenies [71] was also used to assign predicted proteins to clusters of orthologous groups (COG) categories. The dbCAN2 meta server [155] allowed the annotation of carbohydrate active enzymes (CAZymes) with DIAMOND (E-Value ( $10^{-102}$ ), HMMER (E-Value  $<10^{-15}$ , coverage  $>0.35$ ) and Hotpep (Frequency  $>2.6$ , Hits  $>6$ ) using CAZy, dbCAN and PPR database, respectively. The KEGG release 93.0 was searched for orthologies by HMMER/HMMSEARCH using Kofam-KOALA (E-Value ( $10^{-2}$ ) [6]. Gene prediction and annotation from *P. brassicae* and *S. subterranea* genomes [37,137] and transcriptomes [123] were also performed for the purpose of proteome comparisons.

The *P. betae* draft genomes were analysed for non-coding RNA (ncRNA) using cmsearch against the Rfam database v14.4 [78]. tRNA matches were further categorised using tRNAscan-SE [98].

Proteins predicted from both draft genomes, i.e. from isolates A26–41 and RES F41, were compared with each other using BLASTP and clustered with UniProtKB/TrEMBL rhizarian proteins using OrthoVenn2. Since both draft genomes were resolved from contaminated DNA datasets, predicted proteins not found on the other draft genome or in rhizarian proteins were then compared to the NCBI nr database with BLASTP in order to evaluate if they had originated from foreign DNA. Contigs containing only predictions found on one draft genome, and matching bacterial proteins or bacterial RNA models were manually removed as bacterial contamination.

Overall amino acid identity (AAI) between both *P. betae* predicted proteomes was computed with the “aai.rb” AAI calculator [118,119]. Thresholds to consider alignments for AAI computation were a minimum alignment of 50 aa and a minimum identity of 0.2.

The mitochondrial genome was annotated using AGORA [77], GeSeq [142], MITOS2 [9,52] and tRNAscan-SE softwares with the support of additional blast and Rfam searches. The synteny figure was produced with the SimpleSynteny web-based tool [146].

### 2.3.2. Secretome, membranome and putative effector selection

Secreted and membrane protein predictions were investigated following a computational pipeline developed for fungi, and adapted from Vivek-Ananth et al. [148]. First, all predicted proteins were searched for signal peptide (SP) with SignalP v5.0b [7], in “Eukarya” organism group and Phobius [80], for Glycosylphosphatidylinositol (GPI) anchor using big-PI with the “Protozoa” learning set [55] and for transmembrane domains (TM) using Phobius and TMHMM v2.0 [85]. Proteins with a SP and/or a GPI and/or a TM were selected through an inclusive rule and retained for the second step. Endoplasmic reticulum (ER) retrieval signal was then searched using PS SCAN [59] for the PROSITE [132] pattern PS00014. Proteins with an ER signal were discarded. Next, proteins were separated into two groups, one containing proteins with transmembrane domains and/or GPI anchors, and one containing proteins without clues of membrane localization. TMHMM tends to predict TM domain in the first 60 N-terminal amino acids (aa) when SP are present. However, the presence of a SP is a suggestion of extracellular localization rather than membrane localization. Therefore, TM predicted by TMHMM in the first 60 aa were not considered at this step. The last step was the subcellular localization of proteins classified in these two groups using three tools: TargetP v2.0 [56] with the “Non-plant” organism group, WoLF PSORT v0.0.1.1 [69] with the “Animal” organism type, and ProtComp v9.0 (Animal & Fungi) [83]. Proteins were considered as secreted or transmembrane if they are predicted as such by at least two tools. Candidate secreted proteins were eventually screened for effector using the EffectorP 2.0 machine learning classifier

[135].

### 2.4. Comparison of plasmodiophorid predicted proteomes

Predicted proteins from *P. betae* isolates A26–41 and RES F41, *P. brassicae* isolate e3, and *S. subterranea* isolate SSUBK13 were clustered and compared with OrthoVenn2 [154] with E-value and inflation value parameters set to  $1e^{-2}$  and 1.5, respectively. This clustering analysis was performed with *S. subterranea* data kindly provided by the *S. subterranea* genome paper authors [37] and the *P. brassicae* proteome UP000039324 retrieved from the UniProtKB/TrEMBL database.

COG categories were parsed from the EggNOG mapping results obtained for each isolate as well as for *B. natans* (Bigna1), *Phytophthora infestans* (ASM14294v1) and *Plasmodium falciparum* 3D7 (ASM276v2) retrieved from EnsemblProtists. Comparisons of COG categories abundances in full proteomes, secretomes and membranomes were then performed and visualized with R software.

Gene ontology (GO) terms were assigned to the predicted proteins according to the InterProScan analysis results and parsed to produce customized GO-annotation files for each isolate. For *B. natans*, *P. infestans* and *P. falciparum*, GO terms were retrieved from EnsemblProtists with Biomart.

Different custom protein subsets were then generated for each isolate: membrane proteins, secreted proteins, proteins containing ankyrin repeats according to Pfam searches (Anks), and proteins with a hit in PHI-base (Pathogen Host Interactions). Comparisons of GO terms abundances in secretomes and membranomes were then performed with WEGO v2.0 [158,159] and visualized with R software. Significant differences between protein frequencies were assessed using Chi-square tests. Singular Enrichment Analyses (SEA) were performed on those subsets using the agriGO v2.0 [141] analysis toolkit with customized GO-annotated reference sets for each isolate as background. A hypergeometric statistical test followed by the Hochberg adjustment method was used to compute adjusted P-values, only GO terms with an adjusted P-value  $<0.05$  were considered.

Predicted proteins were also scanned for Pfam domain signatures [57] with hmmscan v.3.2.1. Best non-overlapping Pfam matches were parsed from hmmscan results for each protein dataset with an E-value of  $10^{-4}$ . Pfam enrichment analysis was calculated as described by Chandran et al. [32]. Briefly, a hypergeometric distribution was used to compute P-values which were then adjusted for multiple hypothesis testing as explained by Benjamini and Hochberg [8] with an allowed false discovery rate of less than 10%. Enrichment analyses were performed for each species and species combination, with all domains identified from all compared species as background. Moreover, enrichments were also achieved from protein subsets corresponding to orthologs and co-orthologs identified from each species combination with the OrthoVenn2. Only matches with an adjusted P-value  $<0.05$  and mapped on at least three proteins were kept.

#### 2.4.1. Sequence manipulation and phylogeny

Alignment of sequences was achieved with MAFFT [81,87] and visualized either in Mview [13] or in MEGAX software [86]. A maximum-likelihood (ML) concatenation-based species tree was inferred from MAFFT alignments of 31 single copy genes in IQ-TREE [110] in combination with ModelFinder [79] and ultra-fast bootstrap with 1000 replicates [67]. The tree was rooted using *Plasmodium falciparum* and visualized using iTOL v.5 [94]. Dotplots from mitogenomes were obtained with LAST v1170.

## 3. Results

### 3.1. Assembly and annotation of two *P. betae* draft genomes

The sequencing run for the A26–41 and RES F41 isolates yielded 12.3 and 7.1 Gbps accounting for 24,584,570 and 23,751,328 paired reads

after adaptor and quality trimming, respectively. After metaSPAdes assembly, 235,649 and 1,115,426 contigs were generated, respectively. Following binning with metaWRAP, and evaluation of the resulting bins using the bespoke EukCC as described, one bin from each isolate was selected as most likely to be *P. betae*, based on a genome completeness of more than 92% and contamination of less than 3% (Supplementary Table S1). Some medium to high quality bacterial genomes were also recovered and assigned to a taxonomic classification. These are shown in Supplementary Table S2, with bins common between the two isolates highlighted. The contents of both predicted *P. betae* bins were reassembled to achieve the best quality assembly for each draft genome. The resulting draft genomes for A26–41 and RES F41 encompassed 1039 and 1476 contigs, respectively. The A26–41 average contig read coverage was 266-fold whereas the average RES-F41 contig coverage was 37-fold. The absence of significant *Beta vulgaris* (sugar beet host) DNA contamination was confirmed by mapping raw reads from the fully sequenced monogerm DH line KWS2320 (SRA accession no. **SRR869626** and **SRR869627**) to both *P. betae* draft genomes. The GC-content (45.0%) is comparable to *S. subterranea* (45.7%) but lower than *P. brassicae* (58.5–59.4%). Total assembly length (25.4 Mb for RES F41 and 26.5 Mb for A26–41) is comparable to the assemblies of other available full genomes in plasmodiophorids (general genome statistics are detailed in Supplementary Table S3). Similarity between both *P. betae* draft genomes was evaluated through ProgressiveMauve full length genome alignment (Supplementary Fig. S1). From Mauve alignment, 49 A26–41 contigs did not find correspondence in RES F41 whereas 234 RES F41 contigs were absent from A26–41. A total of 26,717 SNP were identified between the isolates as well as 3566 and 3695 gaps in A26–41 and RES F41 respectively. To get a better idea of full-length similarity between isolates, ANI value was computed. The “ani.rb” tool generated 127,606 and 120,483 fragments and discarded 936,345 bp (3.5%) and 1,320,853 bp (5.2%) from the A26–41 and the RES F41 isolates, respectively. From the aligned portions, the average nucleotide identity was 99.87% (Supplementary Fig. S2). These statistics demonstrate that the two draft genomes, despite being from different isolates, had a high degree of similarity. One contig, containing *P. betae* ribosomal DNA genes (18S ribosomal RNA gene, internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, and 28S ribosomal RNA gene) was not properly binned along with the other A26–41 contigs, possibly due to extremely high average read coverage (17,237×). It was manually added to the A26–41 draft genome.

We analysed both draft genomes (including the extra contig identified for A26–41) with cmsearch as described to determine the non-coding RNA (ncRNA) content (Supplementary Table S12). For both draft genomes we were able to recover all components of the spliceosome (U1–U6) and the ribosome (both the large ribosomal subunit (LSU) and the small ribosomal subunit (SSU)), as well as signal recognition particles, ribonucleoproteins, a self-splicing ribozyme, and small nucleolar RNAs common to both genomes. We used tRNAscan-SE to further categorize the tRNA genes present (Supplementary Table S12). The results show that both draft genomes include tRNA genes for all 20 amino acids, as well as an unusual stop codon (TCA) found in the mitochondria of some green algae. This coverage of both rRNA and tRNA genes is consistent with the MIMAG criteria for a “high-quality” genome.

MITObim reassembly resulted in the first sequence of *P. betae* mitochondrial genome from isolate A26–41. Reassembly was also performed with the NOVOPlasty software with highly similar results, confirming the accuracy of reassembly. The total length of the MITObim assembly is 33,862 bp, shorter than the mitogenomes of *S. subterranea* (37,699 bp; [63]) and, from *P. brassicae* (102,962 - 114,663; [41] - [137]). Its GC content is 29.0%, slightly higher than the percentage reported for *S. subterranea* (26.8%) and *P. brassicae* (26.2%). The mitogenome annotation using AGORA, GeSeq, MITOS2 and tRNAscan-SE softwares as well as additional blast and Rfam searches identified 30 protein-coding genes and 27 structural RNAs on the *P. betae* mitochondrion. Sixteen proteins involved in the mitochondrial respiratory chain were

found among protein-coding genes, namely ATP1,6,9, COX1–3, COB, NAD1–7, NAD4I, NAD9. Twelve ribosomal proteins were also identified: RPL6,14,16, RPS3,4,7,8,11,12,13,14,19. Two ORF encoding hypothetical proteins were also detected. The structural RNAs that were found are rRNS, rRNL, rRN5 and 24 tRNA (Supplementary Table S12). A comparison with *P. brassicae* and *S. subterranea* mitochondrial genomes revealed that the three plasmodiophorid mitogenomes show a near perfect synteny (Fig. 2). However, *P. betae* and *S. subterranea* mitochondrial genomes are closer in length and in gene content. This difference is explained by the high intron content, containing additional ORFs, and larger intergenic spaces found within the *P. brassicae* mitochondrial genome.

Next, we investigated the proteome encoded by each *P. betae* nuclear genome, aided by gene annotations predicted using sequence data (31,798,532 paired reads) from poly(A)-RNA purified from sugar beet roots infected with high levels of *P. betae*. After filtering steps, ~16% of the RNAs-eq reads mapped against both *P. betae* draft genomes (this low rate was expected due to the presence of sugar beet RNA). The RNA-seq alignments on both draft genomes were used along with protein hints to assist gene prediction using the BRAKER2 pipeline. In total 10,224 protein-coding genes accounting for 10,799 protein-coding mRNAs were predicted for the A26–41 isolate while 10,222 protein-coding genes accounting for 10,808 protein-coding mRNAs were predicted for the RES-F41 isolate. These two mRNA datasets were then compared to each other using BLASTP and showed that 3.04% (328 out of 10,799) of predicted proteins in isolate A26–41 found no match in the predicted protein set of isolate RES F41. Similarly, 3.22% (348 out of 10,808) of isolate RES F41 predicted proteins found no match in the predicted protein set of isolate A26–41. Overall amino acid identity between the aligned portion of both *P. betae* predicted proteomes computed with the “aai.rb” AAI calculator was again very high, reaching 99.18%.

The InterProScan analysis of the A26–41 predicted proteins found at least one annotation in 9379 predicted proteins (86.8%) from which 4749 (44.0%) could be associated with at least one GO term. For the RES F41 predicted proteins, annotations were found in 9363 predicted proteins (86.6%) from which 4736 (43.8%) could be associated with at least one GO term. EggNOG mapping found seed orthologs for a total of 5633 predictions (52.2%) from A26–41 among which 4247 (39.3%) were assigned to a known Cluster of Orthologous Group (COG) category (other than “S”). Similarly, 5591 predictions (51.7%) from RES F41 were mapped to a seed ortholog from which 4220 (39.0%) were assigned to a known COG category. Functional profiles generated from EggNOG COG categories are presented in Fig. 3. Interestingly, these functional profiles are remarkably similar across the compared plasmodiophorids, but distinct from those of *B. natans*, *Plasmodium falciparum* and *P. infestans*, the three other SAR species included in the comparison. The major difference being the distribution of poorly characterized proteins. A large proportion of predicted rhizaria proteins (including *B. natans*) were unmapped, indicating that they are unrepresented in EggNOG, whereas more poorly characterized proteins from *P. falciparum* and *P. infestans* were matched but unassigned to a category or assigned to the “unknown function” category, probably due to the inclusion of these two species in EggNOG v5.0. Functional profiles have also been generated from selected GO terms derived from the InterProScan analysis (Supplementary Fig. S3). A comparison of genome metrics for A26–41, RES F41 and other plasmodiophorids is presented at Table 1.

CAZyme annotation (see methods) allowed the identification of 177 and 175 carbohydrate-utilizing protein candidates from the A26–41 and RES F41 isolates respectively. To be stringent, we restricted the results to CAZymes predicted with at least two methods on dbCAN2 meta server which represent 64 and 62 proteins, respectively. Identification of isolate A26–41 enzymes involved in known biosynthetic pathways with KofamKOALA allowed us to assign 3001 predicted proteins to a KEGG Ortholog (KO). From these, 1204 were associated with enzyme codes (EC). Similarly, 3010 RES F41 proteins were assigned to a KO from which 1204 were also linked to an EC.

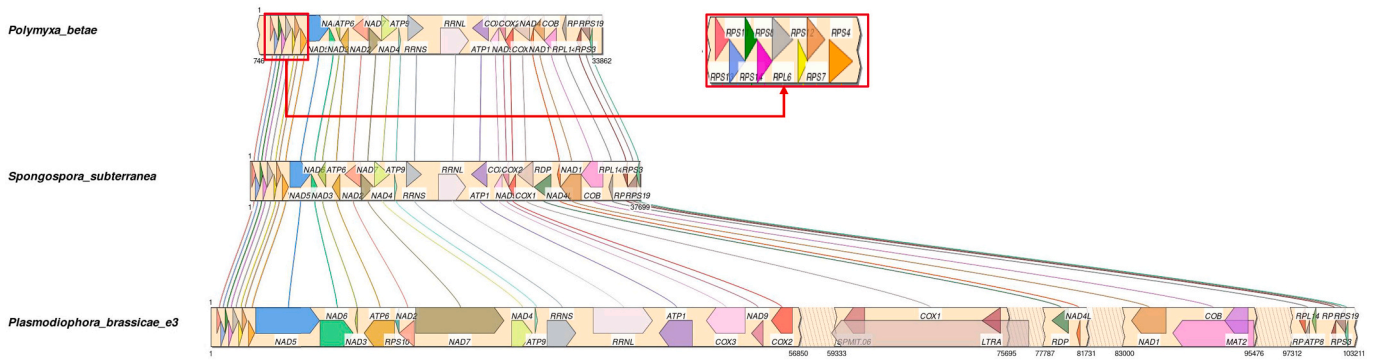


Fig. 2. Syntenic comparison between the mitochondrial genomes of *P. betae*, *P. brassicae* and *S. subterranea*. Homologous genes share the same colors and are connected by curved lines. Gene names are indicated. To make the comparison easier, the *P. brassicae* genome (LS992577) was opened at position 43,000, then both fragments were flipped and merged.

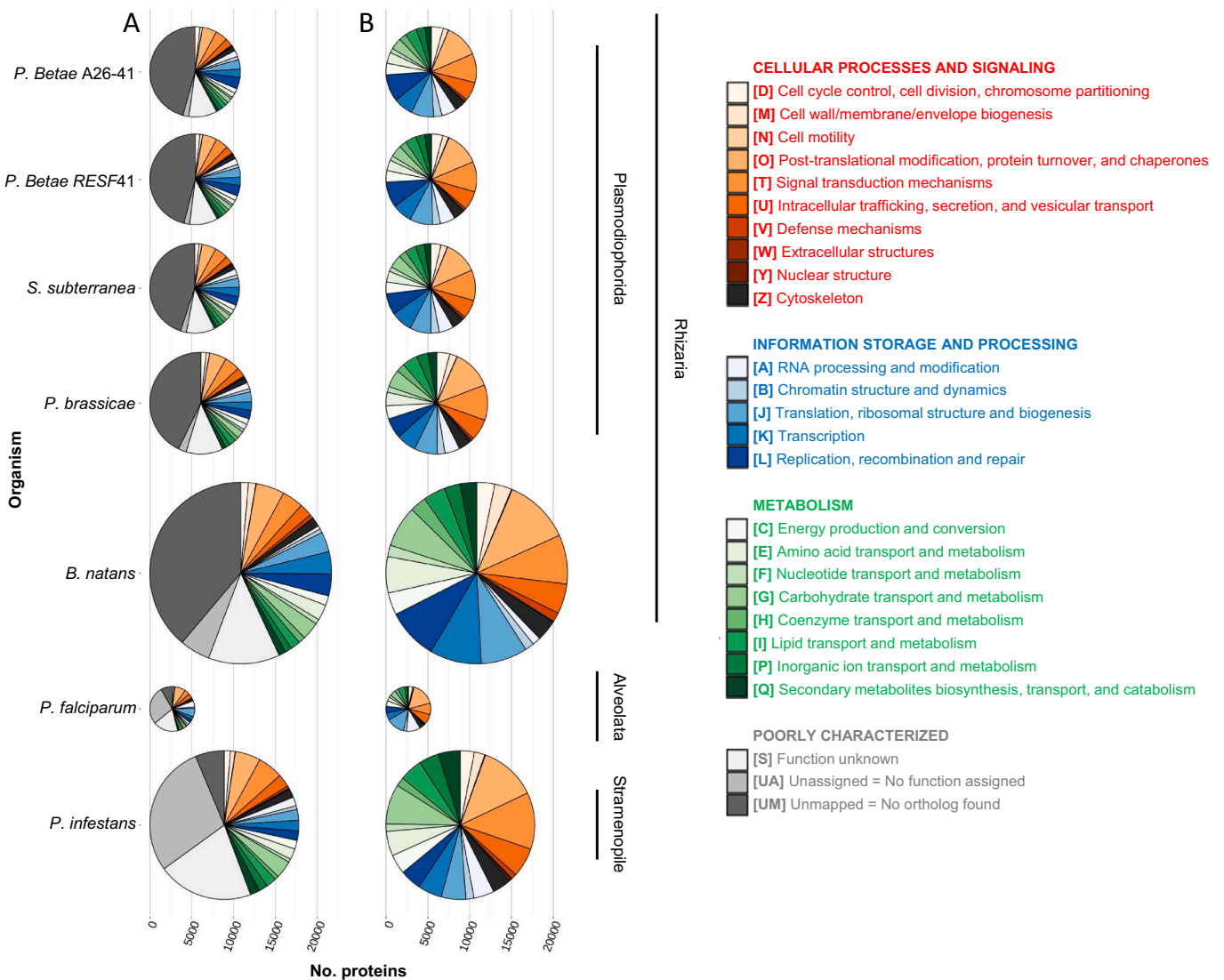


Fig. 3. Comparison of COG categories predicted for *P. betae* (Pbet) draft genomes A26–41 and RES F41, *P. brassicae* (Pbra) e3 and *S. subterranea* (Ssub) SSUBK13 with COG from *B. natans* (Bnat), *Plasmodium falciparum* (Pfal) et *P. infestans* (Pinf). A: COG category profiles with poorly characterized proteins. B: COG category profiles without poorly characterized proteins. The diameter of pie charts is proportional to the number of proteins in full proteomes.

### 3.2. Protein clustering with *S. subterranea* and *P. brassicae*

A protein clustering analysis was performed on proteins predicted

from both *P. betae* nuclear genomes as well as from *P. brassicae* (GenBank accession no. GCA\_900303365.2) and *S. subterranea* (GenBank accession no. GCA\_900404475.1). For a true comparison, the protein sets for

**Table 1**

Comparison of *P. betae* (Pbet) draft genomes A26–41 and RES F41 with other fully sequenced plasmodiophorids: *P. brassicae* (Pbra) e3, eH, ZJ-1 and *S. subterranea* (Ssub) SSUBK13.

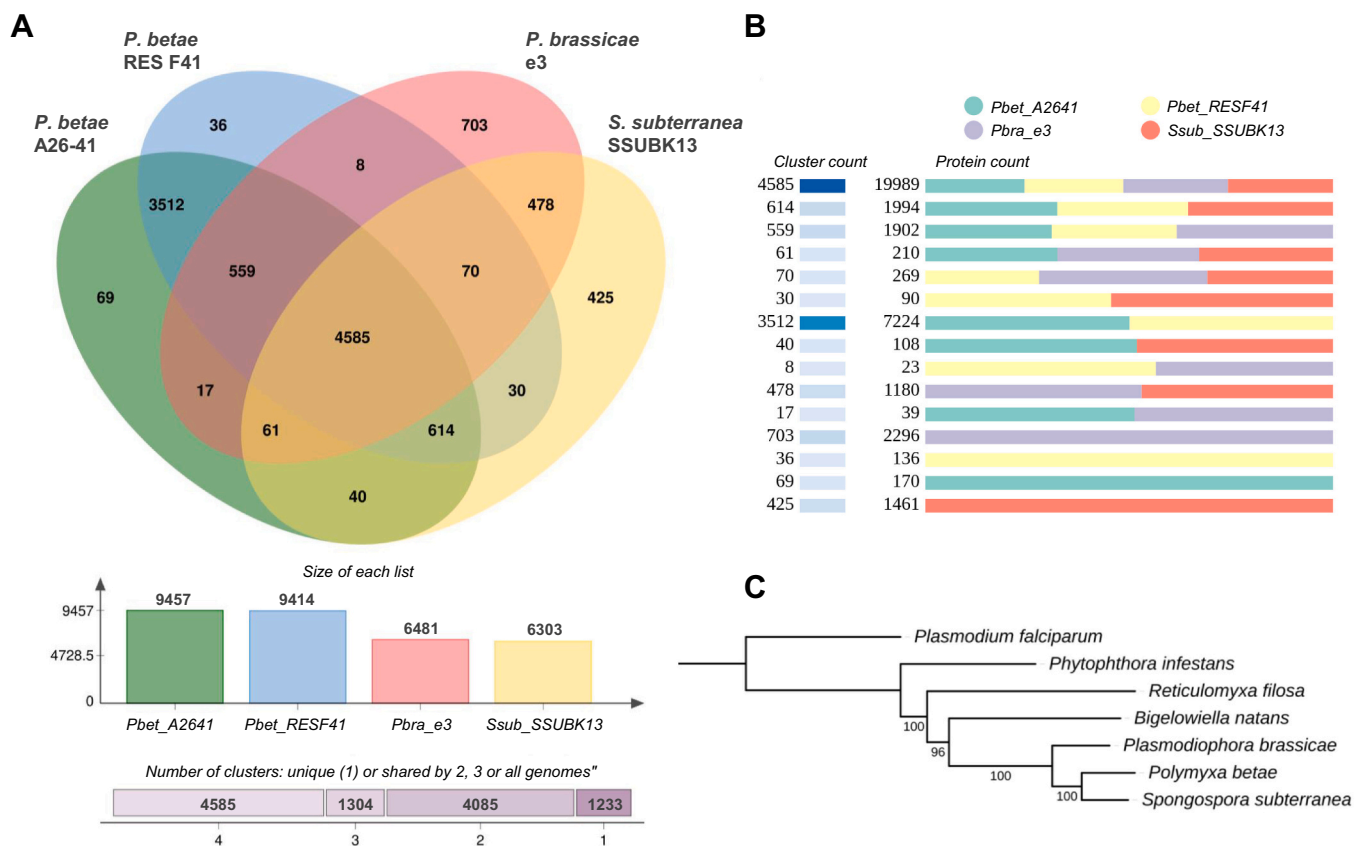
Features	Pbet A26–41 (this study)	Pbet RES F41 (this study)	Ssub SSUBK13 [37]	Pbra Pb3 [120]	Pbra e3 [124]	Pbra e3 [137]	Pbra eH [41]	Pbra ZJ-1 [10]
N° of contigs/scaffolds	1040	1476	2340	107	165	20	136	667
N50 (Kb)	49	57	28	723	473	1300	741	510
Assembly length (Mb)	26.5	25.4	28.1	24.2	24.0	25.1	24.6	24.1
GC content (%)	45.0	45.0	45.7	59.0	58.5	59.0	59.3	59.4
Protein-coding Genes	10,224	10,222	9742 (this study)	10,851	9730	9231	12,590	10,951
CDSs percentage	52.4	54.3	47.9 (this study)	ND	56.8	ND	ND	60.2
Repeats percentage	7.9	5.7	10.7 (this study)	<2%	5.4	11.5	ND	5.9

*P. brassicae* and *S. subterranea* were generated with the same bioinformatic pipeline (BRAKER2 v2.1.4) as used for *P. betae* (further analysis using “published” gene models for *P. brassicae* and *S. subterranea* are presented in Supplementary notes 3.1, but show near identical results). Fig. 4 shows there was a high similarity between both *P. betae* isolates, with 9270 shared clusters containing 9718 and 9673 proteins from A26–41 and RES F41 respectively. Moreover, a closer relationship between both *P. betae* and *S. subterranea* (5400 shared clusters) than between both *P. betae* and *P. brassicae* (5300 shared clusters) is indicated. Results from separate analyses for each *P. betae* isolate are available in Supplementary Fig. S4. Fig. 4 also shows a closer relationship between *P. brassicae* and *S. subterranea* than between *P. betae* and *P. brassicae*, further illustrated by the maximum likelihood phylogenetic tree (Fig. 4C). It should be noted that 767, 870, 3178 and 2564 proteins have been classified as singletons (only clusters are shown in Fig. 4) in

*P. betae* A26–41, RES F41, *P. brassicae* and *S. subterranea*, respectively. Taking singletons and species-specific clusters into account, 42.5%, 45%, and 37.6% of predicted proteins are specific to *P. betae*, *P. brassicae* and *S. subterranea*, respectively. Lastly, 4585 protein clusters are shared by the four organisms, and therefore taken to represent the plasmodiophorids core proteome.

### 3.3. A pipeline for secretome and membrane protein predictions

Proteins predicted from both *P. betae* isolate draft genomes as well as from *P. brassicae* and *S. subterranea* genomes were screened for secreted and membrane proteins using a four-step pipeline adapted from Vivek-Ananth et al. [148]. As presented in Table 2, numbers of remaining proteins at each step are very similar in both *P. betae* draft genomes and closer to *S. subterranea* than to *P. brassicae*. The effectorP 2.0 machine



**Fig. 4.** Protein clustering from *P. betae* (Pbet) draft genomes A26–41 and RES F41, *P. brassicae* (Pbra) e3 and *S. subterranea* (Ssub) SSUBK13 with OrthoVenn2. A: Venn diagram of ortholog clusters. B: cluster and protein counts. C: Maximum likelihood phylogenetic tree of the fully sequenced rhizarias computed with the A26–41 *P. betae* isolate with 31 single copy gene encoded aa sequences. Bootstrap support (in %) is shown on branches.

**Table 2**

Secretome and membranome prediction pipeline results for *P. betae* (Pbet) draft genomes A26–41 and RES F41, *P. brassicae* (Pbra) e3 and *S. subterranea* (Ssub) SSUBK13.

	Pbet A26–41		Pbet RES F41		Pbra e3		Ssub SSUBK13	
Proteome	10,799		10,808		12,163		10,700	
Step 1	3423		3420		4332		3098	
Step 2	3413		3409		4319		3092	
	Secreted	Membr.	Secreted	Membr.	Secreted	Membrane	Secreted	Membr.
Step 3	991	2422	1010	2399	1283	3036	796	2296
Step 4	655	1525	670	1514	834	1942	526	1489

Step 1: proteins containing a signal peptide and/or a GPI anchor and/or a transmembrane domain. Step 2: step 1 proteins containing an ER retrieval signal were discarded. Step 3: Membr. = step 2 proteins containing a GPI anchor and/or a transmembrane domain, Secreted = step 2 proteins lacking GPI anchor and transmembrane domain. Step 4: step3 proteins filtered according to their predicted localization. For details, see methods.

learning classifier was used to predict effector candidates among the secretomes determined from the four genomes. Respectively 104 (15.9%), 104 (15.5%), 94 (17.9%) and 112 (13.4%) of these proteins from *P. betae* A26–41, *P. betae* RES F41, *S. subterranea* and *P. brassicae* are predicted to be effectors. No RXLR effector-like proteins were detected.

Level 2 GO annotation of secretomes and membranomes predicted from the four data sets is presented in Fig. 5. Globally, the same tendencies are visible in each plasmodiophorid. Prevailing molecular functions predicted in the plasmodiophorid secreted proteins are linked to binding and catalytic activities and intervention in metabolic and cellular processes. Other GO terms are poorly represented in all predicted secretomes with fewer than ten proteins mapped even for GO related to cellular components, and of those that fall in this category, some are annotated as part of the cell, organelles or protein complexes. Prevailing molecular functions predicted in the membranomes are linked to binding, catalytic and transporter activities and over-represented biological processes include cellular and metabolic processes as well as localization. Notably, membrane proteins involved in response to stimulus, signaling and regulation of biological process have been revealed.

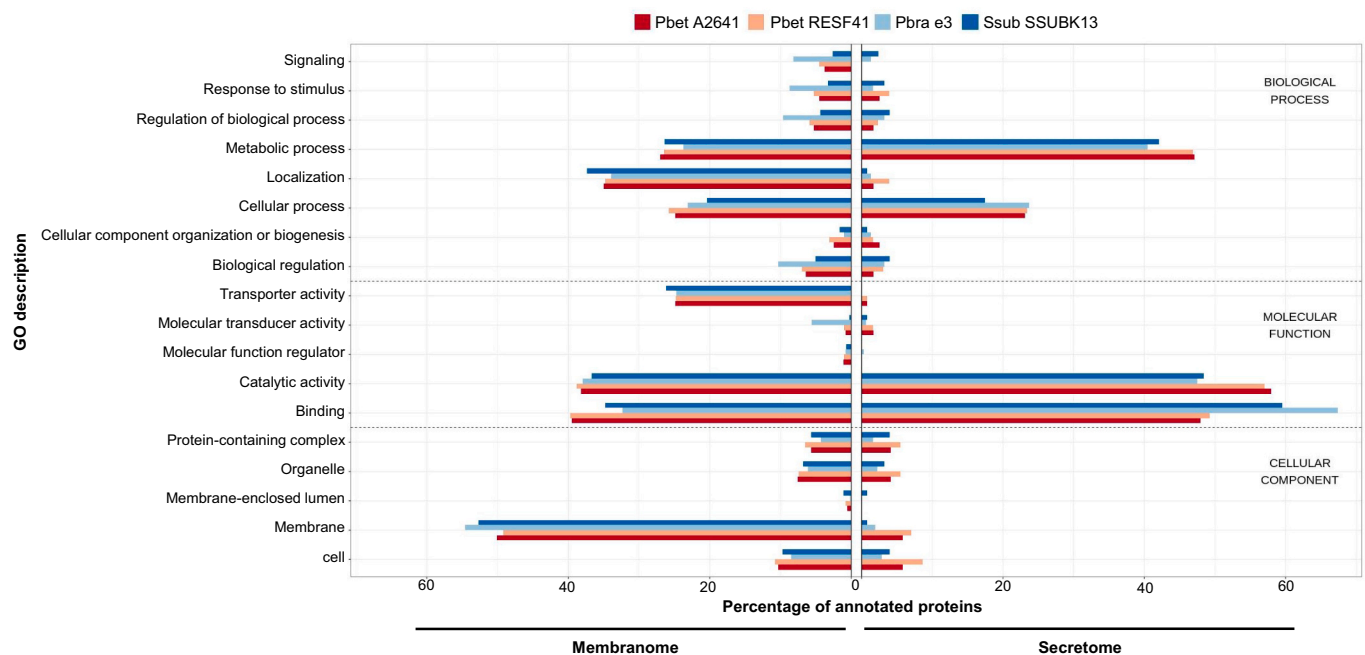
A clustering analysis was performed with OrthoVenn2 and revealed

that most secreted proteins are not shared in plasmodiophorids (Supplementary Fig. S5). Functional profiles generated from EggNOG COG categories showed that the majority of plasmodiophorid secreted proteins remain poorly characterized (Supplementary Fig. S6).

#### 3.4. Comparison of *P. betae* with gall-forming plasmodiophorids

GO terms and Pfam domains overrepresented in protein clusters absent from *P. betae* but shared in *P. brassicae* and *S. subterranea* were examined to detect potential molecular functions, biological processes or domains associated with the gall-forming phenotype (Supplementary Tables S5 and S10). Among GO biological processes enriched in overlapping clusters shared only by *P. brassicae* and *S. subterranea*, we noted “cell differentiation” and “histidyl-tRNA aminoacylation”, “histone acetylation”, “methylation”, as well as “dicarboxylic acid metabolic process”. Moreover, the GO molecular functions “metalloprotease activity”, “kinase regulator activity” and “FMN binding” were also found remarkably enriched. Lastly, GO terms linked to endosome cellular trafficking are also overrepresented in *P. brassicae* and *S. subterranea* overlapping clusters.

Each predicted proteome and secretome was also searched for enriched Pfam domains (Supplementary Tables S6 and S7).



**Fig. 5.** Assignments of proteins predicted as secreted (right) and as membrane proteins (left) to selected general GO terms (level 2). Results for *P. betae* (Pbet) drafts genomes A26–41 and RES F41, *P. brassicae* (Pbra) e3 and *S. subterranea* (Ssub) SSUBK13 are presented as percentages of annotated proteins in membranome or secretome.



Substantially more proteins bearing ankyrin repeats (Anks) were discovered in the gall-makers *S. subterranea* (245) and *P. brassicae* (243) than in *P. betae* (54 in A26–41 and 56 in RES F41), even though based on our clustering and phylogenetic analysis (Fig. 32), *S. subterranea* seems globally closer to *P. betae*. To better characterize proteins containing ankyrin repeats for potential involvement in interactions with the plant host, GO terms were investigated from secreted Anks. Only the ubiquitous “protein binding activity” term was associated with these proteins in both *P. betae* and *S. subterranea*. Enrichment tests were also carried out for subsets representing Anks from each plasmodiophorid proteome (not only secretomes). The sole enriched GO term found was again “protein binding activity” despite the presence of a few other domains. Anks were then searched for pathogenicity relevant characteristics including for the presence of a high confidence blastp hits in PHI database, for the classification or not as a CAZyme and for their predicted secreted and effector nature. Despite the lower number of Anks in *P. betae*, proportionally less *P. betae* Anks (<10%) are considered as secreted than *S. subterranea* Anks (11.8%) and *P. brassicae* Anks (30.5%). Only one *P. betae* Ank from the A26–41 isolate is an effector candidate compared with two in *P. brassicae* and six in *S. subterranea*. However, proportionally more correspondence in the PHI database was found among *P. betae* Anks (16.1%–18.5%) than in *S. subterranea* (13.9%) and *P. brassicae* Anks (11.1%). Altogether, this suggests that the spectrum of Anks is different between the compared plasmodiophorids and from references available in databases.

The respective secretomes and membranomes of all the plasmodiophorids were also searched for enriched GO terms (Supplementary Tables S8 and S9). A weaker representation of carbohydrate metabolism related proteins, in particular “carbohydrate binding” was found in both *P. betae* computational secretomes than in both gall-makers secretomes. In the computational membranome, proportionally more proteins involved in and transferase activities and carbohydrate binding are found in *P. betae* than in *P. brassicae* and *S. subterranea* membranomes.

### 3.5. Plant defense related genes

We also searched the plasmodiophorid genomes for genes encoding plant defense related proteins. Three *P. betae* secreted proteins containing a Barwin domain were recovered in both *P. betae* isolates and have no orthologs in *P. brassicae* and *S. subterranea* secretomes. The proteins share the same general structure: a signal peptide followed by a Barwin domain (barley wound-induced, RlpA-like domain superfamily) or PR-4 in the pathogenesis-related (PR) protein nomenclature. The C-terminal end is characterized by the presence of a long intrinsic disorder region. As shown in the Supplementary Fig. S7, the N-terminal region of those three proteins, after signal peptide cleavage, is conserved over a ~120 bp region corresponding to the PR-4-like domain. Conversely, they are greatly dissimilar through the C-terminal region. Another PR-like protein from the Thaumatin family (PR5-like) was also recovered in both *P. betae* and *P. brassicae* predicted secretomes. Proteins belonging to the CAP superfamily with BLASTP hits to plant PR-1 proteins were also found in the four secretomes.

Lastly, we identified other proteins which could interfere with plant defense. One SOBER1-like immunity suppressor protein was found in *P. betae*, as well as in both of the other compared plasmodiophorids. Three isochorismatases were also found in both *P. betae* proteomes against four in *P. brassicae* and one in *S. subterranea*. These also lack signal peptide and are thus not predicted as secreted through the “classical” secretion pathway. Nevertheless, these three proteins are homologs of experimentally characterized fungal isochorismatases which are known to be secreted through unconventional pathways.

### 3.6. Sperm-specific voltage-gated calcium channel (CatSper)

The clustering and comparative analysis revealed the presence of six orthologs of the mammalian cation channel sperm-associated proteins

(CatSper complex) in *P. betae* and *S. subterranea* proteomes and not in *P. brassicae* proteome (Supplementary Table S5). These include proteins annotated as CatSper1,2,4, CatSper $\beta$ , CatSper $\delta/\epsilon$  and CatSper $\gamma$  by InterProScan. Since there is no InterPro entry for CatSper3, we identified CatSper3 homologs in *P. betae* and *S. subterranea* proteomes through a BLASTP search. The CatSper  $\zeta$  subunit, specific to mammals, raised no BLASTP hits in any plasmodiophorid. The recently identified CatSper associated protein EF-hand calcium-binding domain-containing protein 9 (EFCAB9) also found homologous sequences in both *P. betae* and *S. subterranea* but not in *P. brassicae*. Furthermore, we searched the InterPro database for CatSper protein families and used identified sequences as BLASTP queries against NCBI nr. and as tBLASTn queries against the *P. brassicae* genome sequence [137]. This confirmed the absence of the CatSper complex from *P. brassicae* but also from sequences of any other SAR species except the stramenopile *Hondaea fermentalgiana*, which also contains the CatSper $\alpha$ 1,2,3,4 and CatSper $\beta$ , $\delta/\epsilon$ , $\gamma$  InterPro families. CatSper protein sequences are 100% conserved between both *P. betae* isolates except CatSper $\gamma$  where one conserved aa substitution is present and EFCAB9 in which a 2 aa deletion is present in the gene model for isolate *P. betae* A26–41. *P. betae* and *S. subterranea* CatSper1, 2, 3 and 4 share 51.73%, 63.11%, 57.35% and 50.65% identity, respectively. Identities for CatSper $\beta$ ,  $\delta$ ,  $\gamma$  are a bit lower at 37.08%, 39.84% and 38.29% (38.43%), respectively. The *S. subterranea* EFCAB9 homolog was also found in both *P. betae* isolates (76.19% identity). In *P. betae* the EFCAB9 protein is fused in the N-terminal portion with a sequence stretching across a 394–396 aa region containing a mitogen-activated protein kinase (MAPK) domain. The CatSper genes are located on different contigs in both the *P. betae* and the *S. subterranea* draft genomes. More generally, the “calcium activated cation channel activity” molecular function was found overrepresented in *P. betae* and *S. subterranea* membranomes but not in *P. brassicae*.

### 3.7. GPCR signal transduction pathway

Comparative analyses revealed a much lower level of annotations associated with signal transduction, especially with the “G protein-coupled receptor signaling pathway” (GPCR), in *P. betae* and *S. subterranea* proteomes (with 17 and 14 proteins, respectively) than in *P. brassicae* proteome (82 proteins) (Supplementary Table S4). Significantly enriched Pfam domains were searched from each plasmodiophorid proteome against the entire plasmodiophorids (Supplementary Tables S7 and S8). Several enriched Pfam domains potentially linked to signaling process were identified: “GPCR proteolysis site”, “7 transmembrane sweet-taste receptor of 3 GPCR”, “Leucine rich repeat”, “NHL repeat”, “receptor family ligand binding region”, “Regulator of G protein signaling domain”. Membranome GO enrichments also confirmed the overrepresentation of signal transduction related terms. In particular, *P. brassicae* computational membranome contains 50 predicted proteins belonging to the “G protein-coupled receptor signaling pathway” against six in *P. betae* and none in the *S. subterranea* membranome. This is partly due to the high number of proteins (26) annotated as “G-protein coupled GABA receptor activity” against none in *S. subterranea* and *P. betae* membranomes.

## 4. Discussion

### 4.1. An efficient method to sequence and compare obligate endoparasites eukaryotic genomes

The results presented in this study were generated with DNA extracted from two distinct *P. betae* life stages, sporosori and zoospores, and from two geographically distinct isolates (Belgium and UK). Despite these differences, and the fact that both draft genomes were assembled from samples including distinct levels of contaminating plant and microbial DNA sequences, the alignment of the “cleaned” draft genomes showed a high degree of coverage and identity. Moreover, predicted

protein comparisons from both also showed outstanding similarity. These results endorse the use of the bioinformatic pipeline (Fig. 1) to extract accurate eukaryotic assembled genomes from datasets highly contaminated with prokaryotic sequences. Interestingly, this strategy showed effectiveness with both datasets despite distinct *P. betae* mean coverages and the fact that the RES F41 isolate dataset was substantially more contaminated than the A26–41 isolate dataset (1,115,426 vs. 235,649 putative prokaryotic contigs, respectively). While it would have been advantageous to generate long reads, this was unfortunately not an option in this study because we were not able to produce enough good quality DNA for long-read sequencing. Using such long reads to perform hybrid assemblies is possible with metaSPAdes and would have been useful for reducing the numbers of contigs and gaps in the final assemblies. However, the fact that two different samples provided comparable assemblies with short reads assures the reliability of our results.

Pfam domain signatures were searched against the *P. betae* A26–41 and RES F41 predicted proteins as well as *P. brassicae* and *S. subterranea* *de novo* and published sequence sets (Supplementary Table S11). It is noteworthy that *de novo* gene prediction from published *S. subterranea* and *P. brassicae* genomes noticeably increased the number of Pfam domains returned by hmmscan. This could indicate either more sensitivity, or more redundancy through gene prediction with the BRAKER2 pipeline. Similarly, *P. betae* protein clustering was carried out with *de novo* and published sequence sets. Again, the number of *P. betae* proteins clustered with proteins from *P. brassicae* and *S. subterranea* was higher if all compared gene models had been predicted with the BRAKER2 pipeline. These results indicate a potential bias in comparative analyses depending on the gene prediction method used. However, our analyses show that use of *de novo* gene predictions in place of published gene data is suitable for such comparative analyses. Indeed, it may be a better approach in that it helps to normalise the gene predictions across the different data sets, and gave us the confidence that we were comparing like with like.

Successful plant pathogens have the ability to manipulate the defense response of their host through the secretion of effector proteins [50]. Plasmodiophorids have the particular character to develop plasmodia inside plant root cells, in close contact with their host cytoplasm. Thus, we might speculate that membrane and secreted proteins will constitute, through a long co-evolution, a significant proportion of the proteome co-ordinating the interaction between *P. betae* and its host. Previously, various methods have been used to predict secreted proteins in plasmodiophorid species including, using only SignalP to detect signal peptide [37], following a two-step method relying on SignalP and TMHMM to eliminate proteins with a TM domain [124], or using a four-step pipeline involving SignalP, TMHMM in combination with PredGPI to keep only proteins lacking GPI, and defined as secreted by TargetP [120]. In this study, a four-step pipeline (adapted from Vivek-Ananth et al. [148]) was used to concomitantly predict secreted and membrane proteins. The results of the secretome and membranome GO annotation classification support the predictions generated from this pipeline, as applied to plasmodiophorid species. In fact, the number of *P. brassicae* secreted proteins that were predicted with this pipeline (834) is greater than in previous studies (553–741) [41,120,124] while the methodology is at least as stringent. More recently, two studies focused on *P. brassicae* effectors prediction, using RNA-seq data and reducing the scope to small secreted proteins highly expressed in primary infection of *Brassica napus* [35] and secondary infection of *A. thaliana* [114], resulting in the prediction of 33 (<300aa) and 32 (<400aa) effectors, respectively. The comparison of the expression of predicted *P. betae* secreted proteins in zoospores, sporosori and at different stages of the infection cycle would be useful to generate a short list for experimental functional analysis.

More shared protein clusters were found between *P. betae* isolates than had been previously reported for *P. brassicae* [41]. While the protein clustering results also suggest a similarity in proteomes across plasmodiophorids, there were a significant number of differences which

should be further investigated to better understand their biology. Globally, the clustering results and phylogenomics support the idea that *P. betae* is more phylogenetically related to *S. subterranea* than to *P. brassicae* and that *S. subterranea* should be classified between *P. betae* and *P. brassicae*. This contrasts with some previous studies based only on ribosomal DNA sequence information which grouped *P. brassicae* and *S. subterranea* in a poorly supported clade distinct from *P. betae* and *P. graminis* [17,53,152] but agrees with a recent study revising the taxonomy of phytomyxea [66]. The analysis of the mitochondrial genomes further supports this hypothesis, even if the three mitogenomes showed a nearly perfect synteny. In comparison with the *P. brassicae* mitochondrial genome, *P. betae* and *S. subterranea* mitogenomes are both approximately three times smaller with a reduced number of introns and intergenic sequences.

#### 4.2. Host-plasmodiophorids interactions

In this comparative study, we noticed a lower enrichment in proteins bearing ankyrin-repeat domains (Anks) in the predicted proteomes and secretomes of the asymptomatic *P. betae* than in both gall-makers, *P. brassicae* and *S. subterranea*. We also found that functional prediction of Anks through homology researches raised poor results, especially for secreted Anks. Ankyrin repeats consist of 30–34 aa repeats widely spread in eukaryote proteins [12] and involved in various interactions, primarily with proteins [95] but also with lipids and sugars [73]. The versatile interaction potential of Anks suggests that they are involved in a diverse range of functions enumerated in Voronin and Kiseleva [149] and Islam et al. [73]. Interestingly, some Anks are considered as virulence genes in bacteria, being used to hijack host cellular mechanisms [2]. In particular, Anks genes could be key in establishing the intracellular parasitic lifestyle of some bacteria such as *Wolbachia* [61,64,133] as well as the symbiotic interaction between *Shewanella* sp. and sponges [5]. More generally, Anks abundance seems to be correlated with the microbe symbiotic lifestyle, being particularly enriched in the reduced genomes of some obligate intracellular bacteria in comparison to free living bacteria [74,75]. Plasmodiophorids also have compact genomes [120]. Moreover, Schwelm et al. [124] observed that *P. brassicae* effectors highly expressed in plasmodia contained ankyrin repeats. The expression of Anks putative effectors was further confirmed during primary and secondary infection with *P. brassicae* ([35] [114]) and Voronin and Kiseleva [149] suggested that Anks genes acquisition could be a pathway used for Protozoa evolution and adaptation. Collectively, current evidence indicates that it is possible that Anks could have played an important role during the co-evolution of plasmodiophorids with their hosts.

Genes harboring similarities with plant defense genes were also uncovered. Three PR-4-like proteins containing a signal peptide were detected in *P. betae* computational secretomes but not in other plasmodiophorids. Schwelm et al. (2015) already identified a PR-1-like secretory protein in *P. brassicae* with no homologs in the non-plant pathogenic Rhizarians as well as PR-5-like proteins in *P. brassicae* secretome. Likewise, PR-1-like and PR-5-like proteins were also uncovered in the *P. betae* secretome. Interestingly, a *P. betae* PR-1-like (Pbef2) was previously identified by suppressive subtractive hybridization in the *P. betae*-*B. vulgaris* pathosystem [44]. In plants, PR genes are generally induced in response to an infection or an attack and are part of the plant defense response [127,145]. However, PR protein recognition was also shown to regulate plant defense response. For instance, PR4-LRR1 interaction in pepper resulted in suppression of cell death and defense responses [72]. Therefore, the possible implication of secreted PR-like effectors in plasmodiophorids pathogenicity is worth further investigation. Three isochorismatases-like proteins were also predicted in *P. betae* as well as in other plasmodiophorids. Secreted isochorismatases effectors encoded by the root fungal and oomycete pathogens *Verticillium dahliae* and *Phytophthora sojae* have previously been experimentally shown to suppress salicylate pathway through hydrolysis of its

isochorismate precursor, notably in potato [96,156]. As for plasmodiophorid isochorismatases, these are lacking a signal peptide, suggesting they would be secreted through an unconventional pathway. Moreover, it was previously reported that the manipulation of the salicylic acid pathway using a secreted effector is involved for *P. brassicae* virulence [19,49,99]. Further research is needed to determine if secreted isochorismatases are involved in the molecular interaction between the plasmodiophorids and their respective hosts.

On the other hand, the *P. brassicae* genes encoding PbBSMT, a methyltransferase acting on plant salicylic acid [99] and the auxin-responsive Gretchen Hagen 3 (PbGH3), a protein involved in plant defense hormone homeostasis (Schwelm et al., 2015), were not identified in either *P. betae* or *S. subterranea*, suggesting that plasmodiophorids have adopted different strategies to facilitate the infection and their development within host tissues. In a similar vein, we report that GPCR pathway genes are by far less represented in both *P. betae* and *S. subterranea* than in *P. brassicae* e3. Bi et al. [10] highlighted the importance of GPCR signaling pathway for *P. brassicae* development, pathogenicity and induction of clubroot symptoms. They hypothesized that it could be involved in the recognition of extracellular environment signals (e.g. root exudates) and intracellular signals in the host cells [10,113].

Our results showed that homologs of CatSper complex principal ( $\alpha 1-4$ ) and auxiliary subunits ( $\beta\delta\gamma$ ) were detected in *P. betae* and *S. subterranea* proteomes but not in *P. brassicae*, and to the best of our knowledge, this is peculiar within the protist world [27]. In mammals, CatSper are sperm-specific  $Ca^{++}$  ion channel complex proteins initially known for their role in sperm hyperactivation, motility, fertility and chemotaxis to the egg [36,116,139]. This complex requires the presence of the four subunits to work, controlling the cytoplasmic  $Ca^{++}$  concentrations in a pH-sensitive manner. Globally, an intracellular alkalization and high level of  $Ca^{++}$  inside the cell favors sperm activation [139]. The remarkable conservation of the full CatSper complex as well as the presence of other key elements of  $Ca^{++}$  signaling machinery lead us to hypothesize that the  $Ca^{++}$  signaling could have a critical role in *P. betae* and *S. subterranea* life cycles, putatively through zoospore motility regulation and chemotaxis [27,102] where the CatSper complex would conceivably play an important role. These results could also explain a previous field observation that the infection potential of *P. betae* is positively correlated with  $Ca^{++}$  content and pH in soils [92] and it should be noted that liming to neutralise soil acidity was historically used to tackle clubroot of crucifers caused by *P. brassicae* [30,106] and is still widely practiced as a soil conditioner in agriculture, including for sugar beet crops [54]. Our results are in accordance with the hypothesis that  $Ca^{++}$  signaling machineries appeared from early eukaryotic lineages, before initial divergence, rather than from a subsequent horizontal gene transfer event as previously suggested [27]. It is interesting to note that no CatSper subunit signature was found in *P. brassicae*, *B. natans* or *R. filosa*. If evolutionary origin is common, it would mean that all three organisms have lost the four  $\alpha$  as well as the three auxiliary subunits during evolution as has been demonstrated for numerous metazoa [26]. More generally, we should keep in mind that calcium channels underwent rapid evolution leading to an impressive diversity [147]. Further research should clearly be conducted to better characterize these proteins.

## 5. Conclusions

In the last decade, most plasmodiophorid research focused on *P. brassicae*. Taken together, our results suggest that though if *P. betae*, *P. brassicae* and *S. subterranea* are classified in the *Plasmodiophorida* and share a similar underground intracellular lifestyle, they probably have different strategies, relying on distinct key molecular mechanisms, to invade hosts root cells and multiply. For example, we pointed out striking differences such as a lower enrichment of proteins with ankyrin-repeat domains, the detection of CatSper complex proteins, different

enrichment of GPCR genes for signaling pathways that could be relevant for development and pathogenicity. This emphasizes the need to enlarge plasmodiophorids research focus to more species. Extending the comparison to new plasmodiophorid genome sequences such as *Polymyxa graminis* could be of great help. In any case, the methods and analytical strategy we have developed for eukaryote genome assembly from unpurified samples will no doubt make some contribution towards current efforts to sequence representatives of all living organisms as in the newly established Earth BioGenome Project (<https://www.eartbhigenome.org/>).

## Availability of data and materials

Illumina raw reads are deposited at European Nucleotide Archive (ENA) under accession no. ERR5922443 for *P. betae* isolate RES-F41 DNA-seq data and ERR5905993 for *P. betae* isolate A26–41 RNA-seq data in the Bioproject PRJEB43927. *P. betae* isolate A26–41 DNA-seq data are available under accession no. SRR8105394. Both Genome assemblies and the mitochondrial genome of *P. betae* isolate A26–41 are available in the PRJEB43927.

## Funding

This work was supported by UCLouvain, Belgium and China Agriculture Research System from the Ministry of Agriculture of the P. R. China (CARS-03). RF is supported by EMBL core funds, UK.

## Declaration of Competing Interest

The authors declare no conflict of interest.

## Acknowledgments

*Spongopora subterranea* predicted proteins set used in the *S. subterranea* genome announcement [37] was kindly provided by Stephan Ciaghi and Arne Schwelm. We express our gratitude to Britta Schulz (KWS SAAT SE, Einbeck, Germany) for kindly providing seeds from the KWS2320 sugar beet. Computational resources have been partially provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Walloon Region.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ygeno.2021.11.018>.

## References

- [1] S.M. Adl, Alastair G. Simpson, C.E. Lane, J. Lukeš, D. Bass, S. Bowser, et al., The revised classification of eukaryotes, *J. Eukaryot. Microbiol.* 59 (2012) 429–493, <https://doi.org/10.1111/j.1550-7408.2012.00644.x>.
- [2] S. Al-Khodori, C.T. Price, A. Kalia, Y. Abu Kwaik, Functional diversity of ankyrin repeats in microbial proteins, *Trends Microbiol.* 18 (2010) 132–139, <https://doi.org/10.1016/j.tim.2009.11.004>.
- [3] A. Almeida, A.L. Mitchell, M. Boland, S.C. Forster, G.B. Gloor, A. Tarkowska, et al., A new genomic blueprint of the human gut microbiota, *Nature* 568 (2019) 499–504, <https://doi.org/10.1038/s41586-019-0965-1>.
- [4] S. Andrews, FastQC: A Quality Control Tool for High Throughput Sequence Data, 2010.
- [5] A. Anoop, A. Agostinho, Whole-genome comparisons among the genus *Shewanella* reveal the enrichment of genes encoding Ankyrin-repeats containing proteins in sponge-associated bacteria, *Front. Microbiol.* 10 (2019), <https://doi.org/10.3389/fmicb.2019.00005>.
- [6] T. Aramaki, R. Blanc-Mathieu, H. Endo, K. Ohkubo, M. Kanehisa, S. Goto, et al., KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold, *Bioinformatics* 36 (2020) 2251–2252, <https://doi.org/10.1093/bioinformatics/btz859>.
- [7] J.J.A. Armenteros, K.D. Tsirigos, C.K. Sønderby, T.N. Petersen, O. Winther, S. Brunak, et al., SignalP 5.0 improves signal peptide predictions using deep

- neural networks, *Nat. Biotechnol.* 37 (2019) 420, <https://doi.org/10.1038/s41587-019-0036-z>.
- [8] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Stat. Soc. Ser. B Methodol.* 57 (1995) 289–300.
- [9] M. Bernt, A. Donath, F. Jühling, F. Externbrink, C. Florentz, G. Fritzsche, et al., MITOS: Improved de novo metazoan mitochondrial genome annotation, *Mol. Phylogenet. Evol.* 69 (2013) 313–319, <https://doi.org/10.1016/j.ympev.2012.08.023>.
- [10] K. Bi, T. Chen, Z. He, Z. Gao, Y. Zhao, H. Liu, et al., Comparative genomics reveals the unique evolutionary status of *Plasmodiophora brassicae* and the essential role of GPCR signaling pathways, *Phytopathol. Res.* 1 (2019) 12, <https://doi.org/10.1186/s42483-019-0018-6>.
- [11] K. Bi, Z. He, Z. Gao, Y. Zhao, Y. Fu, J. Cheng, et al., Integrated omics study of lipid droplets from *Plasmodiophora brassicae*, *Sci. Rep.* 6 (2016), <https://doi.org/10.1038/srep36965>.
- [12] P. Bork, Hundreds of ankyrin-like repeats in functionally diverse proteins: Mobile modules that cross phyla horizontally? *Proteins Struct. Funct. Bioinforma.* 17 (1993) 363–374, <https://doi.org/10.1002/prot.340170405>.
- [13] N.P. Brown, C. Leroy, C. Sander, MView: a web-compatible database search or multiple alignment viewer, *Bioinforma. Oxf. Engl.* 14 (1998) 380–381, <https://doi.org/10.1093/bioinformatics/14.4.380>.
- [14] T. Brůna, A. Lomsadze, M. Borodovsky, GeneMark-EP and -EP+: automatic eukaryotic gene prediction supported by spliced aligned proteins, *bioRxiv* (2020), <https://doi.org/10.1101/2019.12.31.891218>, 2019.12.31.891218.
- [15] T. Brůna, K.J. Hoff, A. Lomsadze, M. Stanke, M. Borodovsky, BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database, *NAR Genomics Bioinforma.* 3 (2021), <https://doi.org/10.1093/nargab/lqaa108>.
- [16] H. Buhariwalla, S. Greaves, R. Magrath, R. Mithen, Development of specific PCR primers for the amplification of polymorphic DNA from the obligate root pathogen *Plasmodiophora brassicae*, *Physiol. Mol. Plant Pathol.* 47 (1995) 83–94, <https://doi.org/10.1006/pmpp.1995.1044>.
- [17] S.R. Bulman, S.F. Kühn, J.W. Marshall, E. Schnepf, A phylogenetic analysis of the SSU rRNA from members of the Plasmodiophorida and Phagomyxida, *Protist* 152 (2001) 43–51, <https://doi.org/10.1078/1434-4610-00042>.
- [18] S. Bulman, S. Neuhauser, Phytomyxa, in: J.M. Archibald, A.G.B. Simpson, C. H. Slamovits (Eds.), *Handbook of the Protists*, Springer International Publishing, Cham, 2017, pp. 783–803, [https://doi.org/10.1007/978-3-319-28149-0\\_24](https://doi.org/10.1007/978-3-319-28149-0_24).
- [19] S. Bulman, F. Richter, S. Marschollek, F. Benade, S. Jülke, J. Ludwig-Müller, *Arabidopsis thaliana* expressing PbbSMT, a gene encoding a SABATH-type methyltransferase from the plant pathogenic protist *Plasmodiophora brassicae*, show leaf chlorosis and altered host susceptibility, *Plant Biol. (Stuttg.)* (2018), <https://doi.org/10.1111/plb.12728>.
- [20] S. Bulman, H.J. Ridgway, C. Eady, A.J. Conner, Intron-rich gene structure in the intracellular plant parasite *Plasmodiophora brassicae*, *Protist* 158 (2007) 423–433, <https://doi.org/10.1016/j.protis.2007.04.005>.
- [21] S. Bulman, J. Siemens, H.J. Ridgway, C. Eady, A.J. Conner, Identification of genes from the obligate intracellular plant pathogen, *Plasmodiophora brassicae*, *FEMS Microbiol. Lett.* 264 (2006) 198–204, <https://doi.org/10.1111/j.1574-6968.2006.00466.x>.
- [22] F. Burki, P.J. Keeling, Rhizaria, *Curr. Biol.* 24 (2014) R103–R107, <https://doi.org/10.1016/j.cub.2013.12.025>.
- [23] F. Burki, A. Kudryavtsev, M.V. Matz, G.V. Aglyamova, S. Bulman, M. Fiers, et al., Evolution of Rhizaria: new insights from phylogenomic analysis of uncultivated protists, *BMC Evol. Biol.* 10 (2010) 377, <https://doi.org/10.1186/1471-2148-10-377>.
- [24] F. Burki, A.J. Roger, M.W. Brown, A.G.B. Simpson, The new tree of eukaryotes, *Trends Ecol. Evol.* 35 (2020) 43–55, <https://doi.org/10.1016/j.tree.2019.08.008>.
- [25] B. Bushnell, BBTools Software Package, Available at: <http://sourceforge.net/projects/bbmap/>, 2014.
- [26] X. Cai, D.E. Clapham, Evolutionary genomics reveals lineage-specific gene loss and rapid evolution of a sperm-specific ion channel complex: CatSper and CatSperβ, *PLoS One* 3 (2008), e3569 <https://doi.org/10.1371/journal.pone.0003569>.
- [27] X. Cai, X. Wang, D.E. Clapham, Early evolution of the eukaryotic Ca<sup>2+</sup> signaling machinery: conservation of the CatSper channel complex, *Mol. Biol. Evol.* 31 (2014) 2735–2740, <https://doi.org/10.1093/molbev/msu218>.
- [28] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, et al., BLAST+: architecture and applications, *BMC Bioinformatics* 10 (2009) 421, <https://doi.org/10.1186/1471-2105-10-421>.
- [30] R.N. Campbell, A.S. Greathead, Control of clubroot of crucifers by liming, in: *Soilborne Plant Pathogens: Management of Diseases with Macro- and Microelements*, APS Press, St. Paul, Minnesota, 1989, pp. 90–101. Available at: <https://agris.fao.org/agris-search/search.do?recordID=US9037509>. (Accessed 9 September 2020).
- [31] T. Cao, S. Srivastava, M.H. Rahman, N.N.V. Kav, N. Hotte, M.K. Deyholos, et al., Proteome-level changes in the roots of *Brassica napus* as a result of *Plasmodiophora brassicae* infection, *Plant Sci.* 174 (2008) 97–115, <https://doi.org/10.1016/j.plantsci.2007.10.002>.
- [32] D. Chandran, Y.C. Tai, G. Hather, J. Dewdney, C. Denoux, D.G. Burgess, et al., Temporal global expression data reveal known and novel salicylate-impacted processes and regulators mediating powdery mildew growth and reproduction on *Arabidopsis*, *Plant Physiol.* 149 (2009) 1435–1451, <https://doi.org/10.1104/pp.108.132985>.
- [33] P.-A. Chaumeil, A.J. Mussig, P. Hugenholtz, D.H. Parks, GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database, *Bioinformatics.* 36 (2020) 1925–1927, <https://doi.org/10.1093/bioinformatics/btz848>.
- [34] J. Chen, Z. Wang, J. Hong, C.R. Collier, M.J. Adams, Ultrastructural studies of resting spore development in *Polymyxa graminis*, *Mycol. Res.* 102 (1998) 687–691, <https://doi.org/10.1017/S0953756297005406>.
- [35] W. Chen, Y. Li, R. Yan, L. Xu, L. Ren, F. Liu, et al., Identification and characterization of plasmodiophora brassicae primary infection effector candidates that suppress or induce cell death in host and nonhost plants, *Phytopathology*® 109 (2019) 1689–1697, <https://doi.org/10.1094/PHYTO-02-19-0039-R>.
- [36] J.-J. Chung, K. Miki, D. Kim, S.-H. Shim, H.F. Shi, J.Y. Hwang, et al., CatSperγ regulates the structural continuity of sperm Ca<sup>2+</sup> signaling domains and is required for normal fertility, *eLife* 6 (2017), <https://doi.org/10.7554/eLife.23082>.
- [37] S. Ciaghi, S. Neuhauser, A. Schwelm, Draft genome resource for the potato powdery scab pathogen *Spongospora subterranea*, *Mol. Plant-Microbe Interact.* (2018), <https://doi.org/10.1094/MPMI-06-18-0163-A>.
- [38] S. Ciaghi, A. Schwelm, S. Neuhauser, Transcriptomic response in symptomless roots of clubroot infected kohlrabi (*Brassica oleracea* var. gongyolodes) mirrors resistant plants, *BMC Plant Biol.* 19 (2019) 288, <https://doi.org/10.1186/s12870-019-1902-z>.
- [39] B.A. Curtis, G. Tanifuji, F. Burki, A. Gruber, M. Irimia, S. Maruyama, et al., Algal genomes reveal evolutionary mosaicism and the fate of neulomorphs, *Nature* 492 (2012) 59–65, <https://doi.org/10.1038/nature11681>.
- [40] A.E. Darling, B. Mau, N.T. Perna, progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement, *PLoS One* 5 (2010), e11147, <https://doi.org/10.1371/journal.pone.0011147>.
- [41] S. Daval, A. Belcour, K. Gazengel, L. Legrand, J. Gouzy, L. Cottret, et al., Computational analysis of the *Plasmodiophora brassicae* genome: mitochondrial sequence description and metabolic pathway database design, *Genomics* 111 (2019) 1629–1640, <https://doi.org/10.1016/j.ygeno.2018.11.013>.
- [42] A. Decroës, M. Calusinska, P. Delfosse, C. Bragard, A. Legrève, First draft genome sequence of a *Polymyxa* genus member, *Polymyxa betae*, the protist vector of Rhizomania, *Microbiol. Resour. Announc.* 8 (2019), <https://doi.org/10.1128/MRA.105109-18>.
- [43] P. Delfosse, A.S. Reddy, A. Legrève, K.T. Devi, M.D. Abdurahman, H. Maraite, et al., Serological methods for detection of *Polymyxa graminis*, an obligate root parasite and vector of plant viruses, *Phytopathology* 90 (2000) 537–545, <https://doi.org/10.1094/PHYTO.2000.90.5.537>.
- [44] N. Desoignies, J. Carbonell, J.-S. Moreau, A. Conesa, J. Dopazo, A. Legrève, Molecular interactions between sugar beet and *Polymyxa betae* during its life cycle, *Ann. Appl. Biol.* 164 (2014) 244–256, <https://doi.org/10.1111/aab.12095>.
- [45] N. Desoignies, A. Legrève, In vitro dual culture of *Polymyxa betae* in agrobacterium rhizogenes transformed sugar beet hairy roots in liquid media, *J. Eukaryot. Microbiol.* 58 (2011) 424–425, <https://doi.org/10.1111/j.1550-7408.2011.00563.x>.
- [46] S. Devos, K. Laukens, P. Deckers, D. Van Der Straeten, T. Beeckman, D. Inzé, et al., A hormone and proteome approach to picturing the initial metabolic events during plasmodiophora brassicae infection on *Arabidopsis*, *Mol. Plant-Microbe Interactions*® 19 (2006) 1431–1443, <https://doi.org/10.1094/MPMI-19-1431>.
- [47] N. Dierckxens, P. Mardulyn, G. Smits, NOVOPlasty: de novo assembly of organelle genomes from whole genome data, *Nucleic Acids Res.* 45 (2017), e18, <https://doi.org/10.1093/nar/gkw955>.
- [48] B. Dieryck, J. Weyns, D. Doucet, C. Bragard, A. Legrève, Acquisition and transmission of Peanut clump virus by *Polymyxa graminis* on cereal species, *Phytopathology* 101 (2011) 1149–1158, <https://doi.org/10.1094/PHYTO-12-10-0335>.
- [49] M. Djavaheri, L. Ma, D.F. Klessig, A. Mithöfer, G. Gropp, H. Borhan, Mimicking the host regulation of salicylic acid: a virulence strategy by the clubroot pathogen *Plasmodiophora brassicae*, *MPMI* 32 (2018) 296–305, <https://doi.org/10.1094/MPMI-07-18-0192-R>.
- [50] P.N. Dodds, J.P. Rathjen, Plant immunity: towards an integrated view of plant–pathogen interactions, *Nat. Rev. Genet.* 11 (2010) 539–548, <https://doi.org/10.1038/nrg2812>.
- [51] J.C. Dohm, A.E. Minoche, D. Holtgräwe, S. Capella-Gutiérrez, F. Zakrzewski, H. Tafer, et al., The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*), *Nature* 505 (2014) 546–549, <https://doi.org/10.1038/nature12817>.
- [52] A. Donath, F. Jühling, M. Al-Arab, S.H. Bernhart, F. Reinhardt, P.F. Stadler, et al., Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes, *Nucleic Acids Res.* 47 (2019) 10543–10552, <https://doi.org/10.1093/nar/gkz833>.
- [53] G.J. Down, L.J. Grenville, J.M. Clarkson, Phylogenetic analysis of *Spongospora* and implications for the taxonomic status of the plasmodiophorids, *Mycol. Res.* 106 (2002) 1060–1065, <https://doi.org/10.1017/S0953756202006391>.
- [54] A.P. Draycot, Nutrition, in: *The Sugar Beet Crop: Science Intro Practice*, Cooke DA & Scott RK, London, 1993, pp. 265–270.
- [55] B. Eisenhaber, P. Bork, F. Eisenhaber, Prediction of potential GPI-modification sites in proprotein sequences, *J. Mol. Biol.* 292 (1999) 741–758, <https://doi.org/10.1006/jmbi.1999.3069>.
- [56] O. Emanuelsson, H. Nielsen, S. Brunak, G. von Heijne, Predicting subcellular localization of proteins based on their N-terminal amino acid sequence, *J. Mol. Biol.* 300 (2000) 1005–1016, <https://doi.org/10.1006/jmbi.2000.3903>.

- [57] R.D. Finn, A. Bateman, J. Clements, P. Coggill, R.Y. Eberhardt, S.R. Eddy, et al., Pfam: the protein families database, *Nucleic Acids Res.* 42 (2014) D222–D230, <https://doi.org/10.1093/nar/gkt1223>.
- [58] Y. Galein, A. Legrève, C. Bragard, Long term management of Rhizomania Disease—Insight into the changes of the beet necrotic yellow vein virus RNA-3 observed under resistant and non-resistant sugar beet fields, *Front. Plant Sci.* 9 (2018), <https://doi.org/10.3389/fpls.2018.00795>.
- [59] A. Gattiker, E. Gasteiger, A. Bairoch, ScanProsite: a reference implementation of a PROSITE scanning tool, *Appl. Bioinform.* 1 (2002) 107–108.
- [60] G. Glöckner, N. Hülsmann, M. Schleicher, A.A. Noegel, L. Eichinger, C. Gallinger, et al., The genome of the foraminiferan reticulomyxa filosa, *Curr. Biol.* 24 (2014) 11–18, <https://doi.org/10.1016/j.cub.2013.11.027>.
- [61] L. Gomez-Valero, C. Rusniok, C. Cazalet, C. Buchrieser, Comparative and functional genomics of legionella identified eukaryotic like proteins as key players in host-pathogen interactions, *Front. Microbiol.* 2 (2011) 208, <https://doi.org/10.3389/fmicb.2011.00208>.
- [62] P.V.D. Graaf, S.J. Wale, A.K. Lees, Factors affecting the incidence and severity of Spongospora subterranea infection and galling in potato roots, *Plant Pathol.* 56 (2007) 1005–1013, <https://doi.org/10.1111/j.1365-3059.2007.01686.x>.
- [63] P. Gutiérrez, S. Bulman, J. Alzate, M.C. Ortiz, M. Marín, Mitochondrial genome sequence of the potato powdery scab pathogen *Spongospora subterranea*, *Mitochondrial DNA Part A* 27 (2016) 58–59, <https://doi.org/10.3109/19401736.2013.873898>.
- [64] F. Habyarimana, S. Al-khodori, A. Kalia, J.E. Graham, C.T. Price, M.T. Garcia, et al., Role for the Ankyrin eukaryotic-like genes of *Legionella pneumophila* in parasitism of protozoan hosts and human macrophages, *Environ. Microbiol.* 10 (2008) 1460–1474, <https://doi.org/10.1111/j.1462-2920.2007.01560.x>.
- [65] C. Hahn, L. Bachmann, B. Chevreux, Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach, *Nucleic Acids Res.* 41 (2013), e129, <https://doi.org/10.1093/nar/gkt371>.
- [66] M. Hittorf, S. Letsch-Praxmarer, A. Windegger, D. Bass, M. Kirchmair, S. Neuhauser, Revised taxonomy and expanded biodiversity of the Phytomyxea (Rhizaria, Endomyxa), *J. Eukaryot. Microbiol.* 67 (2020) 648–659, <https://doi.org/10.1111/jeu.12817>.
- [67] D.T. Hoang, O. Chernomor, A. von Haeseler, B.Q. Minh, L.S. Vinh, UFBoot2: improving the ultrafast bootstrap approximation, *Mol. Biol. Evol.* 35 (2018) 518–522, <https://doi.org/10.1093/molbev/msx281>.
- [68] K.J. Hoff, A. Lomsadze, M. Borodovsky, M. Stanke, Whole-genome annotation with BRAKER, *Methods Mol. Biol. Clifton NJ* 1962 (2019) 65–95, [https://doi.org/10.1007/978-1-4939-9173-0\\_5](https://doi.org/10.1007/978-1-4939-9173-0_5).
- [69] P. Horton, K.-J. Park, T. Obayashi, N. Fujita, H. Harada, C.J. Adams-Collier, et al., WoLF PSORT: protein localization predictor, *Nucleic Acids Res.* 35 (2007) W585–W587, <https://doi.org/10.1093/nar/gkm259>.
- [70] J. Huerta-Cepas, K. Forslund, L.P. Coelho, D. Szklarczyk, L.J. Jensen, C. von Mering, et al., Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper, *Mol. Biol. Evol.* 34 (2017) 2115–2122, <https://doi.org/10.1093/molbev/msx148>.
- [71] J. Huerta-Cepas, D. Szklarczyk, D. Heller, A. Hernández-Plaza, S.K. Forslund, H. Cook, et al., eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses, *Nucleic Acids Res.* 47 (2019) D309–D314, <https://doi.org/10.1093/nar/gky1085>.
- [72] I.S. Hwang, D.S. Choi, N.H. Kim, D.S. Kim, B.K. Hwang, Pathogenesis-related protein 4b interacts with leucine-rich repeat protein 1 to suppress PR4b-triggered cell death and defense response in pepper, *Plant J.* 77 (2014) 521–533, <https://doi.org/10.1111/tpj.12400>.
- [73] Z. Islam, R.S.K. Nagampalli, M.T. Fatima, G.M. Ashraf, New paradigm in ankyrin repeats: beyond protein-protein interaction module, *Int. J. Biol. Macromol.* 109 (2018) 1164–1173, <https://doi.org/10.1016/j.ijbiomac.2017.11.101>.
- [74] K.K. Jernigan, S.R. Bordenstein, Ankyrin domains across the tree of life, *PeerJ* 2 (2014), e264, <https://doi.org/10.7717/peerj.264>.
- [75] K.K. Jernigan, S.R. Bordenstein, Tandem-repeat protein domains across the tree of life, *PeerJ* 3 (2015), e732, <https://doi.org/10.7717/peerj.732>.
- [76] P. Jones, D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, et al., InterProScan 5: genome-scale protein function classification, *Bioinformatics* 30 (2014) 1236–1240, <https://doi.org/10.1093/bioinformatics/btu031>.
- [77] J. Jung, J.I. Kim, Y.-S. Jeong, G. Yi, AGORA: organellar genome annotation from the amino acid and nucleotide references, *Bioinformatics* 34 (2018) 2661–2663, <https://doi.org/10.1093/bioinformatics/bty196>.
- [78] I. Kalvari, E.P. Nawrocki, J. Argasinska, N. Quinones-Olvera, R.D. Finn, A. Bateman, et al., Non-coding RNA analysis using the Rfam database, *Curr. Protoc. Bioinformatics* 62 (2018), e51, <https://doi.org/10.1002/cpbi.51>.
- [79] S. Kalyaanamoorthy, B.Q. Minh, T.K.F. Wong, A. von Haeseler, L.S. Jermin, ModelFinder: fast model selection for accurate phylogenetic estimates, *Nat. Methods* 14 (2017) 587–589, <https://doi.org/10.1038/nmeth.4285>.
- [80] L. Käll, A. Krogh, E.L.L. Sonnhammer, Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server, *Nucleic Acids Res.* 35 (2007) W429–W432, <https://doi.org/10.1093/nar/gkm256>.
- [81] K. Katoh, J. Rozewicki, K.D. Yamada, MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization, *Brief. Bioinform.* 20 (2019) 1160–1166, <https://doi.org/10.1093/bib/bbx108>.
- [82] D. Kim, B. Langmead, S.L. Salzberg, HISAT: a fast spliced aligner with low memory requirements, *Nat. Methods* 12 (2015) 357–360, <https://doi.org/10.1038/nmeth.3317>.
- [83] E.W. Klee, L.B. Ellis, Evaluating eukaryotic secreted protein prediction, *BMC Bioinformatics* 6 (2005) 256, <https://doi.org/10.1186/1471-2105-6-256>.
- [84] E.V. Kriventseva, D. Kuznetsov, F. Tegenfeldt, M. Manni, R. Dias, F.A. Simão, et al., OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs, *Nucleic Acids Res.* 47 (2019) D807–D811, <https://doi.org/10.1093/nar/gky1053>.
- [85] A. Krogh, B. Larsson, G. von Heijne, E.L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, *J. Mol. Biol.* 305 (2001) 567–580, <https://doi.org/10.1006/jmbi.2000.4315>.
- [86] S. Kumar, G. Stecher, M. Li, C. Nkayaz, K. Tamura, MEGA X: molecular evolutionary genetics analysis across computing platforms, *Mol. Biol. Evol.* 35 (2018) 1547–1549, <https://doi.org/10.1093/molbev/msy096>.
- [87] S. Kuraku, C.M. Zmasek, O. Nishimura, K. Katoh, aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity, *Nucleic Acids Res.* 41 (2013) W22–W28, <https://doi.org/10.1093/nar/gkt389>.
- [88] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with bowtie 2, *Nat. Methods* 9 (2012) 357–359, <https://doi.org/10.1038/nmeth.1923>.
- [89] A. Legrève, P. Delfosse, H. Maraite, Phylogenetic analysis of *Polymyxa* species based on nuclear 5.8S and internal transcribed spacers ribosomal DNA sequences, *Mycol. Res.* 106 (2002) 138–147, <https://doi.org/10.1017/S0953756201005391>.
- [90] A. Legrève, P. Delfosse, B. Vanpee, A. Goffin, H. Maraite, Differences in temperature requirements between *Polymyxa* sp. of Indian origin and *Polymyxa graminis* and *Polymyxa betae* from temperate areas, *Eur. J. Plant Pathol.* 104 (1998) 195–205, <https://doi.org/10.1023/A:1008612903927>.
- [91] A. Legrève, H. Maraite, Influence du calcium et du pH sur le potentiel infectieux du sol en *Polymyxa betae* Keskin et son expression en champ, in: 2nd International Conference of the Virologists - Proceedings. Colmar, France, 1991.
- [92] A. Legrève, B. Vanpee, P. Delfosse, H. Maraite, Host range of tropical and subtropical isolates of *Polymyxa graminis*, *Eur. J. Plant Pathol.* 106 (2000) 379–389, <https://doi.org/10.1023/A:1008784823899>.
- [93] I. Letunic, P. Bork, Interactive tree of life (iTOL) v4: recent updates and new developments, *Nucleic Acids Res.* 47 (2019) W256–W259, <https://doi.org/10.1093/nar/gkz239>.
- [94] J. Li, A. Mahajan, M.-D. Tsai, Ankyrin repeat: a unique motif mediating protein-protein interactions, *Biochemistry* 45 (2006) 15168–15178, <https://doi.org/10.1021/bi062188q>.
- [95] T. Liu, T. Song, X. Zhang, H. Yuan, L. Su, W. Li, et al., Unconventionally secreted effectors of two filamentous pathogens target plant salicylate biosynthesis, *Nat. Commun.* 5 (2014) 4686, <https://doi.org/10.1038/ncomms5686>.
- [96] A. Lomsadze, P.D. Burns, M. Borodovsky, Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm, *Nucleic Acids Res.* 42 (2014), <https://doi.org/10.1093/nar/gku557> e119–e119.
- [97] T.M. Lowe, S.R. Eddy, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, *Nucleic Acids Res.* 25 (1997) 955–964.
- [98] J. Ludwig-Müller, S. Jülke, K. Geiß, F. Richter, A. Mithöfer, I. Sola, et al., A novel methyltransferase from the intracellular pathogen *Plasmiodiophora brassicae* methylates salicylic acid: salicylic acid methyltransferase from *P. brassicae*, *Mol. Plant Pathol.* 16 (2015) 349–364, <https://doi.org/10.1111/mpp.12185>.
- [99] J. Ludwig-Müller, E. Prinsen, S.A. Rolfe, J.D. Scholes, Metabolism and plant hormone action during clubroot disease, *J. Plant Growth Regul.* 28 (2009) 229–244, <https://doi.org/10.1007/s00344-009-9089-4>.
- [100] R. Malinowski, J.A. Smith, A.J. Fleming, J.D. Scholes, S.A. Rolfe, Gall formation in clubroot-infected Arabidopsis results from an increase in existing meristematic activities of the host but is not essential for the completion of the pathogen life cycle, *Plant J.* 71 (2012) 226–238, <https://doi.org/10.1111/j.1365-3113.2012.04983.x>.
- [101] E.M. Medina, N.E. Buchler, Chytrid fungi, *Curr. Biol.* 30 (2020) R516–R520.
- [102] U. Merz, R.E. Falloon, Review: powdery scab of potato—increased knowledge of pathogen biology and disease epidemiology for effective disease management, *Potato Res.* 52 (2009) 17–37, <https://doi.org/10.1007/s11540-008-9105-2>.
- [103] A. Meunier, J.-F. Schmit, A. Stas, N. Kutluk, C. Bragard, Multiplex reverse transcription-PCR for simultaneous detection of beet necrotic yellow vein virus, beet soilborne virus, and beet virus Q and their vector *Polymyxa betae* KESKIN on sugar beet, *Appl. Environ. Microbiol.* 69 (2003) 2356–2360, <https://doi.org/10.1128/AEM.69.4.2356-2360.2003>.
- [104] A.L. Mitchell, T.K. Attwood, P.C. Babbitt, M. Blum, P. Bork, A. Bridge, et al., InterPro in 2019: improving coverage, classification and access to protein sequence annotations, *Nucleic Acids Res.* 47 (2019) D351–D360, <https://doi.org/10.1093/nar/gky1100>.
- [105] H. Murakami, S. Tsushima, Y. Kuroyanagi, Y. Shishido, Reduction of resting spore density of *Plasmiodiophora brassicae* and clubroot disease severity by liming, *Soil Sci. Plant Nutr.* 48 (2002) 685–691, <https://doi.org/10.1080/00380768.2002.10409258>.
- [106] E.S. Mutasa, E. Ward, M.J. Adams, C.R. Collier, D.M. Chwarszczynska, M.J. C. Asher, A sensitive DNA probe for the detection of *Polymyxa betae* in sugar beet roots, *Physiol. Mol. Plant Pathol.* 43 (1993) 379–390, <https://doi.org/10.1006/pmpp.1993.1066>.
- [107] E.S. Mutasa-Gottgens, D.M. Chwarszczynska, K. Halsey, M.J.C. Asher, Specific polyclonal antibodies for the obligate plant parasite *Polymyxa*—a targeted recombinant DNA approach, *Plant Pathol.* 49 (2000) 276–287, <https://doi.org/10.1046/j.1365-3059.2000.00446.x>.
- [108] S. Neuhauser, S. Bulman, M. Kirchmair, Plasmiodiophorids: the challenge to understand soil-borne, obligate biotrophs with a multiphasic life cycle, in: *Molecular Identification of Fungi*, Springer, Berlin, Heidelberg, 2010, pp. 51–78, [https://doi.org/10.1007/978-3-642-05042-8\\_3](https://doi.org/10.1007/978-3-642-05042-8_3).

- [110] L.-T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, *Mol. Biol. Evol.* 32 (2015) 268–274, <https://doi.org/10.1093/molbev/msu300>.
- [111] S. Nurk, D. Meleshko, A. Korobeynikov, P.A. Pevzner, metaSPAdes: a new versatile metagenomic assembler, *Genome Res.* 27 (2017) 824–834, <https://doi.org/10.1101/gr.213959.116>.
- [112] D.H. Parks, M. Imelfort, C.T. Skennerton, P. Hugenholtz, G.W. Tyson, CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes, *Genome Res.* 25 (2015) 1043–1055, <https://doi.org/10.1101/gr.186072.114>.
- [113] E. Pérez-López, M. Waldner, M. Hossain, A.J. Kusalik, Y. Wei, P.C. Bonham-Smith, et al., Identification of plasmodiophora brassicae effectors - a challenging goal, *Virulence* 9 (2018) 1344–1353, <https://doi.org/10.1080/21505594.2018.1504560>.
- [114] E. Pérez-López, M.M. Hossain, J. Tu, M. Waldner, C.D. Todd, A.J. Kusalik, et al., Transcriptome analysis identifies plasmodiophora brassicae secondary infection effector candidates, *J. Eukaryot. Microbiol.* 67 (2020) 337–351, <https://doi.org/10.1111/jeu.12784>.
- [115] X. Qu, B.J. Christ, Single cystosorus isolate production and restriction fragment length polymorphism characterization of the obligate Biotroph *Spongopora subterranea* f. sp. *subterranea*, *Phytopathology* 96 (2006) 1157–1163, <https://doi.org/10.1094/PHYTO-96-1157>.
- [116] D. Ren, B. Navarro, G. Perez, A.C. Jackson, S. Hsu, Q. Shi, et al., A sperm ion channel required for sperm motility and male fertility, *Nature* 413 (2001) 603–609, <https://doi.org/10.1038/35098027>.
- [117] J.T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, et al., Integrative genomics viewer, *Nat. Biotechnol.* 29 (2011) 24–26, <https://doi.org/10.1038/nbt.1754>.
- [118] L.M. Rodriguez-R, K.T. Konstantinidis, Bypassing cultivation to identify bacterial species: culture-independent genomic approaches identify credibly distinct clusters, avoid cultivation bias, and provide true insights into microbial species, *Microbe Mag.* 9 (2014) 111–118, <https://doi.org/10.1128/microbe.9.111.1>.
- [119] L.M. Rodriguez-R, K.T. Konstantinidis, The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes, *PeerJ Inc.* (2016), <https://doi.org/10.7287/peerj.preprints.1900v1>.
- [120] S.A. Rolfe, S.E. Strelkov, M.G. Links, W.E. Clarke, S.J. Robinson, M. Djavaheri, et al., The compact genome of the plant pathogen *Plasmodiophora brassicae* is adapted to intracellular interactions with host *Brassica* spp, *BMC Genomics* 17 (2016) 272, <https://doi.org/10.1186/s12864-016-2597-2>.
- [121] P. Saary, A.L. Mitchell, R.D. Finn, Estimating the quality of eukaryotic genomes recovered from metagenomic analysis, *bioRxiv* (2019), <https://doi.org/10.1101/2019.12.19.882753>, 2019.12.19.882753.
- [122] A. Schwelm, J. Badstöber, S. Bulman, N. Desoignies, M. Etemadi, R.E. Falloon, et al., Not in your usual top 10: protists that infect plants and algae, *Mol. Plant Pathol.* 19 (2018) 1029–1044, <https://doi.org/10.1111/mpp.12580>.
- [123] A. Schwelm, C. Dixelius, J. Ludwig-Müller, New kid on the block – the clubroot pathogen genome moves the plasmodiophorids into the genomic era, *Eur. J. Plant Pathol.* 145 (2015) 531–542, <https://doi.org/10.1007/s10658-015-0839-9>.
- [124] A. Schwelm, J. Fogelqvist, A. Knaust, S. Jülke, T. Lilja, G. Bonilla-Rosso, et al., The *Plasmodiophora brassicae* genome reveals insights in its life cycle and ancestry of chitin synthases, *Sci. Rep.* 5 (2015) 11153, <https://doi.org/10.1038/srep11153>.
- [125] A. Sedaghatkish, B.D. Gossen, F. Yu, D. Torkamaneh, M.R. McDonald, Whole-genome DNA similarity and population structure of *Plasmodiophora brassicae* strains from Canada, *BMC Genomics* 20 (2019) 744, <https://doi.org/10.1186/s12864-019-6118-y>.
- [127] J. Sels, J. Mathys, B.M.A. De Coninck, B.P.A. Cammue, M.F.C. De Bolle, Plant pathogenesis-related (PR) proteins: a focus on PR peptides, *Plant Physiol. Biochem.* 46 (2008) 941–950, <https://doi.org/10.1016/j.plaphy.2008.06.011>.
- [128] N. Shah, Q. Li, Q. Xu, J. Liu, F. Huang, Z. Zhan, et al., CRB and PbBa8.1 Synergically increases resistant genes expression upon infection of *plasmodiophora brassicae* in *Brassica napus*, *Genes* 11 (2020) 202, <https://doi.org/10.3390/genes11020202>.
- [129] S.J. Sibbald, J.M. Archibald, More protist genomes needed, *Nat. Ecol. Evol.* 1 (2017) 1–3, <https://doi.org/10.1038/s41559-017-0145>.
- [130] J. Siemens, H. Graf, S. Bulman, O. In, J. Ludwig-Müller, Monitoring expression of selected *Plasmodiophora brassicae* genes during clubroot development in *Arabidopsis thaliana*, *Plant Pathol.* 58 (2009) 130–136, <https://doi.org/10.1111/j.1365-3059.2008.01943.x>.
- [131] J. Siemens, I. Keller, J. Sarx, S. Kunz, A. Schuller, W. Nagel, et al., Transcriptome analysis of *Arabidopsis* Clubroots indicate a key role for Cytokinins in disease development, *Mol. Plant-Microbe Interactions®* 19 (2006) 480–494, <https://doi.org/10.1094/MPMI-19-0480>.
- [132] C.J.A. Sigrist, E. de Castro, L. Cerutti, B.A. Cuhe, N. Hulo, A. Bridge, et al., New and continuing developments at PROSITE, *Nucleic Acids Res.* 41 (2013) D344–D347, <https://doi.org/10.1093/nar/gks1067>.
- [133] K. Singhal, S. Mohanty, Genome organisation and comparative genomics of four novel *Wolbachia* genome assemblies from Indian *Drosophila* host, *Funct. Integr. Genomics* 19 (2019) 617–632, <https://doi.org/10.1007/s10142-019-00664-5>.
- [134] M. Smith, C. Rush, CHAPTER 25: *Polymyxa* species and the viruses they vector? With emphasis on benyviruses and furoviruses, in: Madeleine Smith, Charles Rush (Eds.), *Vector-Mediated Transmission of Plant Pathogens* General Plant Pathology, The American Phytopathological Society, 2016, pp. 353–364, <https://doi.org/10.1094/9780890545355.025>.
- [135] J. Spersneider, P.N. Dodds, D.M. Gardiner, K.B. Singh, J.M. Taylor, Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0, *Mol. Plant Pathol.* 19 (2018) 2094–2110, <https://doi.org/10.1111/mpp.12682>.
- [136] M. Stanke, M. Diekhans, R. Baertsch, D. Haussler, Using native and syntenically mapped cDNA alignments to improve de novo gene finding, *Bioinformatics* 24 (2008) 637–644, <https://doi.org/10.1093/bioinformatics/btn013>.
- [137] S. Stjelja, J. Fogelqvist, C. Tellgren-Roth, C. Dixelius, The architecture of the plasmodiophora brassicae nuclear and mitochondrial genomes, *Sci. Rep.* 9 (2019) 15753, <https://doi.org/10.1038/s41598-019-52274-7>.
- [138] Z.W. Subr, U. Kastirr, T. Kühne, Subtractive cloning of DNA from *Polymyxa graminis*—an obligate parasitic plasmodiophorid, *J. Phytopathol.* 150 (2002) 564–568, <https://doi.org/10.1046/j.1439-0434.2002.00795.x>.
- [139] X. Sun, Y. Zhu, L. Wang, H. Liu, Y. Ling, Z. Li, et al., The Catsper channel and its roles in male fertility: a systematic review, *Reprod. Biol. Endocrinol.* 15 (2017), <https://doi.org/10.1186/s12958-017-0281-2>.
- [140] V. Ter-Hovhannisyann, A. Lomsadze, Y.O. Chernoff, M. Borodovsky, Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training, *Genome Res.* 18 (2008) 1979–1990, <https://doi.org/10.1101/gr.081612.108>.
- [141] T. Tian, Y. Liu, H. Yan, Q. You, X. Yi, Z. Du, et al., agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update, *Nucleic Acids Res.* 45 (2017) W122–W129, <https://doi.org/10.1093/nar/gkx382>.
- [142] M. Tillich, P. Lehwark, T. Pellizzer, E.S. Ulbricht-Jones, A. Fischer, R. Bock, et al., GeSeq – versatile and accurate annotation of organelle genomes, *Nucleic Acids Res.* 45 (2017) W6–W11, <https://doi.org/10.1093/nar/gkx391>.
- [143] L. Torrance, K. Wright, F. Crutzen, G. Cowan, N. Lukhovitskaya, C. Bragard, et al., Unusual features of pomoviral RNA movement, *Front. Microbiol.* 2 (2011), <https://doi.org/10.3389/fmicb.2011.00259>.
- [144] G.V. Uritskiy, J. DiRuggiero, J. Taylor, MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis, *Microbiome* 6 (2018) 158, <https://doi.org/10.1186/s40168-018-0541-1>.
- [145] L.C. van Loon, M. Rep, C.M.J. Pieterse, Significance of inducible defense-related proteins in infected plants, *Annu. Rev. Phytopathol.* 44 (2006) 135–162, <https://doi.org/10.1146/annurev.phyto.44.070505.143425>.
- [146] D. Veltri, M.M. Wight, J.A. Crouch, SimpleSynteny: a web-based tool for visualization of microsynteny across multiple species, *Nucleic Acids Res.* 44 (2016) W41–W45, <https://doi.org/10.1093/nar/gkw330>.
- [147] A. Verkhatsky, V. Parpura, Calcium signalling and calcium channels: evolution and general principles, *Eur. J. Pharmacol.* 739 (2014) 1–3, <https://doi.org/10.1016/j.ejphar.2013.11.013>.
- [148] R.P. Vivek-Ananth, K. Mohanraj, M. Vandanashree, A. Jhingran, J.P. Craig, A. Samal, Comparative systems analysis of the secretome of the opportunistic pathogen *Aspergillus fumigatus* and other *Aspergillus* species, *Sci. Rep.* 8 (2018), <https://doi.org/10.1038/s41598-018-25016-4>.
- [149] D.A. Voronin, E.V. Kiseleva, Functional role of proteins containing ankyrin repeats, *Cell Tissue Biol.* 2 (2008) 1–12, <https://doi.org/10.1134/S1990519X0801001X>.
- [150] A.J. Wakeham, J.G. White, Serological detection in soil of *Plasmodiophora brassicae* resting spores, *Physiol. Mol. Plant Pathol.* 48 (1996) 289–303, <https://doi.org/10.1006/pmpp.1996.0024>.
- [151] J.A. Walsh, U. Merz, G. Harrison, Serological detection of spore balls of *Spongopora subterranea* and quantification in soil, *Plant Pathol.* 45 (1996) 884–895, <https://doi.org/10.1111/j.1365-3059.1996.tb02899.x>.
- [152] E. Ward, M.J. Adams, Analysis of ribosomal DNA sequences of *Polymyxa* species and related fungi and the development of genus- and species-specific PCR primers, *Mycol. Res.* 102 (1998) 965–974, <https://doi.org/10.1017/S0953756297005881>.
- [153] Woronin, *Plasmodiophora brassicae*, *Urheber der Kohlpflanzen-Hernie*, in: *Jahrbücher für Wissenschaftliche Botanik*, 1877, pp. 548–574.
- [154] L. Xu, Z. Dong, L. Fang, Y. Luo, Z. Wei, H. Guo, et al., OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species, *Nucleic Acids Res.* 47 (2019) W52–W58, <https://doi.org/10.1093/nar/gkz333>.
- [155] H. Zhang, T. Yohe, L. Huang, S. Entwistle, P. Wu, Z. Yang, et al., dbCAN2: a meta server for automated carbohydrate-active enzyme don't think that we should trye annotation, *Nucleic Acids Res.* 46 (2018) W95–W101, <https://doi.org/10.1093/nar/gky418>.
- [156] X. Zhu, A. Soliman, M.R. Islam, L.R. Adam, F. Daayf, Verticillium dahliae's isochorismatase hydrolase is a virulence factor that contributes to interference with potato's salicylate and jasmonate defense signaling, *Front. Plant Sci.* 8 (2017), <https://doi.org/10.3389/fpls.2017.00399>.
- [157] S. Neuhauser, M. Kirchmair, S. Bulman, D. Bass, Cross-kingdom host shifts of phytoxyd parasites, *BMC Evol. Biol.* 14 (2014) 33, <https://doi.org/10.1186/1471-2148-14-33>.
- [158] J. Ye, L. Fang, H. Zheng, Y. Zhang, J. Chen, Z. Zhang, et al., WEGO: a web tool for plotting GO annotations, *Nucleic Acids Res.* 34 (2006) W293–W297, <https://doi.org/10.1093/nar/gkl031>.
- [159] J. Ye, Y. Zhang, H. Cui, J. Liu, Y. Wu, Y. Cheng, et al., WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update, *Nucleic Acids Res.* 46 (2018) W71–W75, <https://doi.org/10.1093/nar/gky400>.