Analysis of Carbohydrate Metabolism Genes of *Spongospora subterranea* Using 454 Pyrosequencing

Análisis de Genes del Metabolismo de Carbohidratos de *Spongospora subterranea* Utilizando Pirosecuenciación 454

Pablo Andrés Gutiérrez Sánchez[1]; Juan Fernando Alzate[2] and Mauricio Marín Montoya[3]

Abstract. ***Spongospora subterranea***, *the causal agent of Potato powdery scab, is an important soil-borne obligate protozoan commonly found in Andean soils. This is a serious problem that causes cosmetic damage on the skin of tubers and induces root gall formation, diminishing the yield and commercial value of the potato. Genetic studies on **S. subterranea** are difficult due to its obligate parasitism, which explains the lack of available knowledge on its basic biology. **S. subterranea** is a member of the Plasmodiophorida order, a protist taxa that includes other important plant pathogens such as **Plasmodiophora brassicae** and **Spongospora nasturtii.** Little is known about the genomes of Plasmodiophorida; however, with the use of Next-Generation Sequencing technologies combined with appropriate bioinformatic techniques, it is possible to obtain genomic sequences from obligate pathogens such as **S. subterranea**. To gain a better understanding of the biology of this pathogen and Plasmodiophorida in general, DNA sequences from a cystosori-enriched sample of **S. subterranea** were obtained using 454 pyrosequencing technology. As a first step in understanding the nutritional requirements of **S. subterranea** as well as its infective and resistance structures, we present a bioinformatic analysis of 24 contigs related to genes involved in the glycolysis, starch, cellulose and chitin metabolism. Intron structure and codon usage is also discussed. The genes analyzed in this study are a good source of information for studies aimed at characterizing these enzymes **in vitro**, as well as the generation of new methods for the molecular detection of **S. subterranea** in either soils or infected plants.*

Key words: Genomics, *Plasmodiophora*, powdery scab, *Solanum tuberosum*.

Resumen. ***Spongospora subterranea***, *el agente causal de la sarna polvosa de la papa, es un protozoo y patógeno obligado presente en los suelos andinos. Esta enfermedad es un serio problema para el cultivo de papa, al causar lesiones cosméticas en la piel de los tubérculos, así como agallas en las raíces, limitando sus rendimientos y valor comercial. Debido al caracter de parásito obligado de **S. subterranea**, es muy poco lo que se conoce sobre su biología básica. Este patógeno es un miembro del orden Plasmodiophorida, un taxón que incluye otros microorganismos de importancia como **Plasmodiophora brassicae** y **Spongospora nasturtii**. Sin embargo, con el uso combinado de métodos de secuenciación de nueva generación y herramientas bioinformáticas apropiadas, es posible obtener información genómica sobre microorganismos no cultivables. Con el fin de profundizar en la comprensión de la biología de **S. subterranea** y del orden Plasmidiophora en general, se llevó a cabo la secuenciación mediante la metodología 454 de una muestra enriquecida con quistorosoros del patógeno. En este trabajo se identificaron 24 ensamblajes (contigs) de secuencias relacionadas con la glicólisis y el metabolismo del almidón, celulosa y quitina. Tambien se discute la estructura de los intrones y el uso de codones. Los genes analizados en este estudio son una buena fuente de información para estudios futuros enfocados en la caracterización de estas enzimas, así como para la generación de nuevos métodos de detección molecular de **S. subterranea** en suelos y plantas infectadas.*

Palabras clave: Genómica, *Plasmodiophora*, sarna polvosa, *Solanum tuberosum*.

Potato powdery scab is a disease caused by *Spongospora subterranea*, a soil-borne obligate parasitic protozoan commonly found in Andean soils. This is a serious problem that can cause cosmetic damage on the skin of potato tubers, limiting their use for commercial purposes, as well as the induction of root galls that can seriously compromise nutrient and water intake (Merz and Falloon, 2009). Moreover, tubers presenting symptoms of powdery scab are rejected as seeds because they can infest disease-free areas and spread the pathogen to new crops. *S. subterranea* is also responsible for the transmission of *Potato mop top virus* (PMTV), the causal agent of spraing disease in potato tubers, which produces significant reductions in plant growth and induces internal tuber necrosis (Latvala-Kilby *et al*., 2009). Both diseases are major problems in the potato industry worldwide as there are no adequate control measures or resistant cultivars available. The rising incidence of powdery scab is also of great concern due to increased international trading and commerce agreements (Osorio *et al*., 2012).

Powdery scab was discovered in Germany in 1841 and spread soon afterwards to the rest of Europe,

---

[1] Associate Professor. Universidad Nacional de Colombia - Sede Medellín - Facultad de Ciencias – Escuela de Biociencias. A.A. 3840, Medellín, Colombia
. <paguties@unal.edu.co>
[2] Associate Professor. Universidad de Antioquia - Escuela de Medicina. A.A. 1226, Medellín, Colombia. <juan.alzate@medicina.udea.edu.co>
[3] Associate Professor. Universidad Nacional de Colombia - Sede Medellín - Facultad de Ciencias – Escuela de Biociencias. A.A. 3840, Medellín, Colombia. AA. 3840, Medellín, Colombia. <mamarinm@unal.edu.co>

the American continent and potato-growing regions throughout the world (Osorio *et al.*, 2012). In Colombia, *S. subterranea* was first reported in 1965 and is currently widely distributed in the provinces of Boyacá, Cundinamarca, Nariño and Antioquia, which are responsible for 87% of the total potato production (Osorio *et al.*, 2012). The incidence of *S. subterranea* is on a steady increase due to the use of low quality seeds, the absence of *S. subterranea* tolerant cultivars and the lack of appropriate diagnostic tools to certify tuber seeds as *S. subterranea* and PMTV free (Gil *et al.*, 2011). The powdery scab is responsible for losses of about 50% in potato varieties such as Parda pastusa, Diacol capiro (*Solanum tuberosum* L.) and Criolla Colombia (*S. phureja* (Juz. et. Buk)) (SINAIPA, 2002), while damage associated with PMTV have been estimated at 25% (Guerrero, 1997, 2000).

S. subterranea can persist in the environment as a spongy aggregate of resting spores, or cystosori, which germinate to release biflagellate zoospores in the presence of a host (Merz and Falloon, 2009). The infection is characterized by a biphasic development featuring multinucleate plasmodia that eventually cleave to form zoosporangia. In the secondary phase, germinating zoosporangia release secondary zoospores, which infect roots, stolons, and tubers. Plasmodia can migrate through the root tissue by breaking through host cell walls (Osorio *et al.*, 2012). *S. subterranea* can cause infections in a wide range of host plants outside the Solanaceae family which in turn can act as reservoirs of infection (Webster and Weber, 2007). Due to its obligate parasitism, it is difficult to conduct genetic studies on *S. subterranea,* explaining the lack of available knowledge on its basic biology (Qu and Christ, 2006).

S. subterranea is a member of the Plasmodiophorida order, a protist taxa that includes other important plant pathogens such as *Plasmodiophora brassicae*, which causes clubroot of Brassicaceae plants, and *Spongospora nasturtii* responsible for crooked root of water-cress (Archibald and Keeling, 2004; Webster and Weber, 2007). Little is known about the genomes of Plasmodiophorida (Braselton, 1995; Siemens *et al.*, 2009; Bulman *et al.*, 2011); however, with the use of Next-Generation Sequencing (NGS) technologies combined with appropriate bioinformatic procedures, it is possible to obtain genomic sequences from obligate pathogens such as *S. subterranea*.

These tools have been used with success in sequencing intracellular protozoa such as *Enterocytozoon bieneusi*

(Keeling et al., 2010) and in the identification of lower eukaryotes and bacteria present in complex environments, such as the biofilms formed in water distribution systems (Liu et al., 2012). Using NGS, the genome of *Blumeria graminis*, an ascomycete biotroph pathogen causing barley powdery mildew revealed a genome size expansion caused by transposon proliferation concomitant with a dramatic reduction in gene content, such as genes encoding sugar-cleaving enzymes, transporters and assimilatory enzymes for inorganic nitrate and sulfur (Spanu *et al.*, 2010). More recently, the genomes of *Melampsora laricipopulina* (101 Mb) and *Puccinia graminis* f. sp. *tritici* (89 Mb) were completed; these pathogens are responsible for some of the most devastating diseases of crop plants: poplar leaf rust wheat and barley stem rust, respectively. The comparison of the 16,399 predicted proteins of *M. larici-populina* with the 17,773 predicted proteins of *P. graminis* f.sp. *tritici* revealed genomic features related to their obligate biotrophic lifecycle, including expanded lineage-specific gene families, a large repertoire of effector-like small secreted proteins, impaired nitrogen and sulfur assimilation pathways, and expanded families of amino acids and oligopeptide membrane transporters. The dramatic up-regulation of transcript coding for small secreted proteins, secreted hydrolytic enzymes, and transporters in planta suggests that they play a role in host infection and nutrient acquisition (Duplessis *et al.*, 2011).

To gain a better understanding of the biology of *S. subterranea* and Plasmodiophorida in general, DNA sequences from a cystosori-enriched sample were obtained in this study using 454 pyrosequencing technology. As a first step in understanding the nutritional requirements of *S. subterranea* as well as its infective and resistance structures, we present a bioinformatic analysis of genes involved in glycolysis, starch, celullose and chitin metabolism.

MATERIALS AND METHODS

454 pyrosequencing. Cystosori were purified from a 500 g soil sample collected at Chuscalito (05°58'38"N, 75°21'54"W) in the municipality of La Unión (Antioquia), an area with a prolonged historical record of powdery scab disease. After drying at 25°C for 48 h, cystosori were obtained by using 250 and 90 μm sieves following the method of Jaramillo and Botero (2007). The presence of *S. subterranea* was confirmed with PCR, using primers Spo8 (5' CTG GGT GCG ATT

GTC TGT TG 3') and Spo9 (5' CAC GCC AAT GGT TAG AGA CG 3') (Bulman and Marshall, 1998). The ITS sequence of ribosomal DNA corresponding to this pathogen was also identified in the 454 data. DNA extraction was performed with the PowerSoil DNA Isolation kit (Mo Bio Laboratories, EEUU), following the manufacturer's recommendations. The DNA was fragmented by nebulization at 30 psi for 1 min and further purified with a MinElute PCR purification kit (Qiagen, Germany). The DNA library was sequenced in a FLX 454 Life Sciences system (Roche) at Centro Nacional de Secuenciación Genómica (Universidad de Antioquia) using TITANIUM reagents (Roche). Sequencing reads were assembled with Newbler v2.5.3 (Roche).

Bioinformatics. Genes involved in carbohydrate metabolism were identified using a local protein database in combination with a BLASTX search their putative function was confirmed manually. Exons/intron junctions were identified with the intron consensus reported for *P. brassicae* (Bulman *et al*., 2006, 2007) together with protein translations and reading frames obtained from BLASTX. The proposed protein sequence product was used in a new BLASTP search to identify the closest homolog available and to verify that the size of the putative protein was correct. The evolutionary relationship of sequences was inferred using the Neighbor-Joining method (Saitou and Nei, 1987) and distances computed using the JTT matrix-based method (Schwarz and Dayhoff, 1979) and 1,000 bootstrap replicates. The rate of variation among the sites was modeled with gamma distribution (shape parameter = 1). All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA5 (Tamura *et al*., 2011).

RESULTS AND DISCUSSION

The assembly from 454 pyrosequencing data resulted in a total of 54,049 contigs from 717,646 reads for a total of 33,812,768 bp. A global GC composition analysis exhibited a bimodal distribution with local maxima at 44% and 62%, suggesting the presence of a mixture of sequences from different organisms, most likely potato, *S. subterranea* and soil microorganisms (Figure 1A). A database of all DNA sequences available from the Solanaceae family was used in combination with BLASTN to remove plant DNA. 6,745 unique sequences (contigs and singletons) assembled from 111,065 reads accounting for 3,507,101 bp (10.3% of the total) were identified as potato DNA. Further

removal of contaminating sequences from the soil microorganisms resulted in a total of 30,976 unique sequences assembled from a total of 505,168 reads and corresponding to 21,586,028 bp (63.8%) with a high probability of belonging to *S. subterranea*. Several lines of evidence support this assumption: 1) the only ITS sequence remaining was identified as associated to *S. subterranea*; 2) The average GC content of *S. subterranea* was estimated to be 44%, which correlates well with the GC content sequences attributed to this pathogen; 3) the curated database included all the sequence data previously obtained for *S. subterranea* on rDNAs and protein coding genes (Bulman *et al*., 2011); 4) The presence of introns, which suggests DNA of an eukaryotic origin with an intron/structure very similar to that determined in the closely related *P. brassicae* (Bulman *et al*., 2006, 2007).

A total of 24 contigs were annotated as having functions associated with carbohydrate metabolism. BLASTX hits revealed a very moderate identity at the amino acid (a.a.) levels with currently known protein, suggesting an organism with distant phylogenetic relationship with those available at GenBank (Table 1). In a previous study, Bulman *et al*. (2011), using an *in vitro* dual culture system of *P. brassicae* and *S. subterranea* constructed a pilot-scale DNA library and determined some important characteristics of the *S. subterranea* genome, such as a high content of retrotransposable elements and the presence of introns in protein-coding genes. This intron structure was confirmed in our study and was very useful in determining the protein-coding regions in the absence of a reference transcriptome. A Intron splice site at the 5´ border revealed a preference for the AGGTA sequence. At the 3´ border, the (C/T)AG sequence was preferred and a run of T/A residues was observed 10-13 nt upstream from the AG splice site (Figure 1B). Overall, as shown in the codon usage table, there was no strong bias in codon preference. However, some rarely used codons included CTA (leucine), ATA (isoleucine), GGG (glycine) and GTA (valine) (Figure 1C).

Glycolysis genes. Due to the importance potato tubers play in the lifecycle of *S. subterranea*, the pathogen should be specialized in metabolizing sugar compounds within this rich food source. As glycolysis is the central pathway in the utilization of carbohydrates, *S. subterranea* is expected to have all the enzymes in this important metabolic pathway.

Table 1. Preliminary identification of the *Spongospora subterranea* sequences involved in the carbohydrate metabolism using BLASTX.

Contig	nt	Best BLASTX hit	Organism	Putative domains	QC*	Ev	Id
00323	3107	Hypothetical protein (EGZ09330)	*Phytophthora sojae*	pfam00224: Pyruvate kinase	60%	3e-44	66%
00693	2954	Hypothetical protein (EEC80343.1)	*Oryza sativa Indica*	pfam00342:Phosphoglucose isomerase	46%	9e-77	67%
00720	2575	Hypothetical protein (EEC79748)	*Oryza sativa Indica*	pfam00083: Sugar (and other) transporter	36%	2e-36	44%
01250	2203	GADPH (YP_007373664)	*Clostridium stercorarium*	pfam02800: GADPH	32%	2e-74	74%
01599	2057	Probable · -D-xylosidase 2-like (XP_003383636)	*Amphimedon queenslandica*	pfam01915: Glycosyl hydrolase family 3	53%	4e-54	35%
01982	1919	Hypothetical protein (EFX90443)	*Daphnia pulex*	pfam00162: Phosphoglycerate kinase	33%	9e-27	74%
02507	1781	Predicted · -glucosidase (XP_003384091)	*Amphimedon queenslandica*	pfam01055: Glycosyl hydrolases family 31	40%	4e-36	67%
02634	1743	Phosphoglycerate kinase (XP_637130)	*Dictyostelium discoideum*	pfam00162: Phosphoglycerate kinase	50%	2e-62	77%
03218	1638	Cytosolic phosphoglucose isomerase (CAC85647)	*Clarkia similis*	pfam00342: Phosphoglucose isomerase	19%	3e-38	63%
03335	1617	Chitin synthase (ACA49151)	*Aphanomyces euteiches*	pfam01644: Chitin synthase	24%	1e-20	40%
03461	1595	Phosphofructokinase, putative (XP_002515946)	*Ricinus communis*	cl00204: PFK Superfamily	63%	4e-51	65%
03607	1576	Pyruvate kinase muscle isozyme (ADY42759)	*Ascaris suum*	pfam00224: Pyruvate kinase	49%	5e-43	65%
04184	1492	Glucokinase (YP_002481374)	*Cyanothece sp.*	pfam02685: Glucokinase	29%	4e-31	47%
04261	1469	Phosphoglycerate mutase (EJU03001)	*Dacryopinax sp.*	pfam01676: Metalloenzyme superfamily	60%	2e-54	70%
04468	1456	Sugar hydrolase (ACN18098)	Uncultured bacterium	cd06543: GH18_PF-ChiA-like[	57%	5e-31	49%
05356	1355	Predicted protein (XP_002157859)	*Hydra magnipapillata*	pfam06415: BPG-independent PGAM	37%	2e-39	78%
05404	1356	Hypothetical protein (EGZ27195)	*Phytophthora sojae*	pfam03142: Chitin synthase	99%	2e-106	42%
06526	1251	Hypothetical protein (EFA85346)	*Polysphondylium pallidum*	pfam01055: Glycosyl hydrolases family 31	62%	6e-43	44%
07316	1192	Chitin synthase (AEZ51041)	*Pythium aphanidermatum*	cd04190: C-terminal domain of Chitin Synthase	40%	2e-31	49%
09275	1063	Triosephosphate isomerase (YP_003505684)	*Denitrovibrio acetiphilus*	pfam00121: Triosephosphate isomerase	22%	3e-17	71%
09967	1018	Probable · -D-xylosidase 2-like (XP_003383636)	*Amphimedon queenslandica*	pfam01915: Glycosyl hydrolase family 3	56%	1e-47	58%
12920	874	Fructose-bisphosphate aldolase B (NP_001007978)	*Gallus gallus*	pfam00274: Fructose-bisphosphate aldolase class-I	81%	3e-39	68%
13579	844	Chitinase (YP_005059432)	*Granulicella mallensis*	cd06543: GH18_PF-ChiA-like	70%	2e-24	53%
15341	772	· -glucosidase (EFA84959)	*Polysphondylium pallidum*	pfam00933: Glycosyl hydrolase family 3	50%	1e-24	43%

QC: Query coverage; Ev: BLAST expectation value; Id: Percent identity.

A

B

C
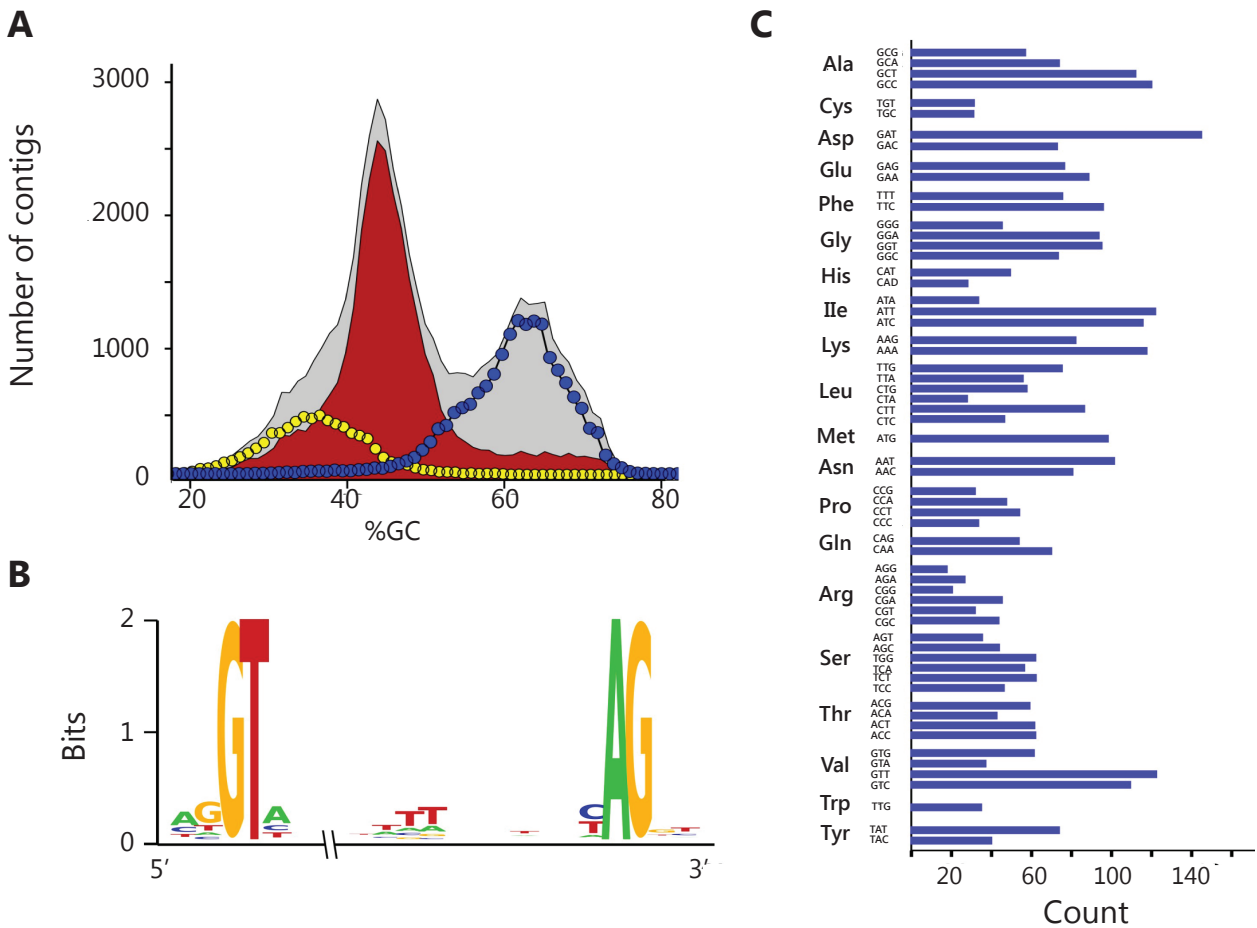
Figure 1. Overall genomic features of *Spongospora subterranea*. **A)** Percent GC composition of contigs identified as *S. tuberosum* (yellow circles) and microbial origin (Blue circles). The remaining sequences (red) are attributed to *S. subterranea*. **B)** Sequence consensus at intron splice sites at the 5´and 3´ border. **C)** Codon usage in the coding regions discussed in this paper.

Glycolysis consists of ten steps that sequentially convert glucose to pyruvate with the production of ATP and NADH. Thirteen contigs were annotated as glycolytic enzymes; only enolase, responsible for the conversion of 2-phospohglycerate to phosphoenolpyruvate, was not found. No evidence of either fructose or galactose utilization was found (Figure 2).

Glucokinase is the point of entry of glucose into the glycolitic pathway and phosphorylate glucose using ATP as a donor to give glucose-6-phosphate and ADP. Contig 04184 was identified as corresponding to the C-terminal portion of this enzyme, including a portion of the nucleotide-binding domain. Phosphoglucose isomerase (PGI) was identified in contigs 00693 and 03218. The third step of glycolysis, which consists of the phosphorylation of F6P to fructose-1,6-biphosphate (FBP), requires Phosphofructokinase (PFK). Contig 03461 corresponds to the N-terminal region of the PFK gene (326 a.a). Contig 12920 corresponds to a central portion of the aldolase gene, which catalyzes the conversion of FBP to glyceraldehyde-3-phosphate (GAP) and dihydroxyacetone phosphate (DHAP). Contig 09275 encodes for the C-terminal region of triose phosphate isomerase (TIM) and is split by two introns. The following steps in the glycolytic pathway are catalyzed by dehydrogenase (GAPDH), Phosphoglycerate kinase (PGK) and Phosphoglycerate mutase (PGM). Contig 01250 encodes the C-terminal end of GAPDH with the corresponding coding sequence split by three introns. PGK is encoded by contigs 01982 and 02634, corresponding to the N- and C-terminal portions, respectively. The PGM enzyme from *S. subterranea* was identified as a 2,3-bisphosphoglycerate-

independent phosphoglycerate, encoded at the N- and C-terminal ends by contigs 05356 and 04261, respectively. Finally, the Pyruvate kinase gene was identified in contigs 03607 and 00323. Contig 03607 encodes for residues 1-264 and is divided into at least six exons; contig 00323

contains two introns and corresponds to positions 294-507 in the homologous protein (Figure 2).

Glycolysis must be essential for supplying the reducing power required for electron transport and
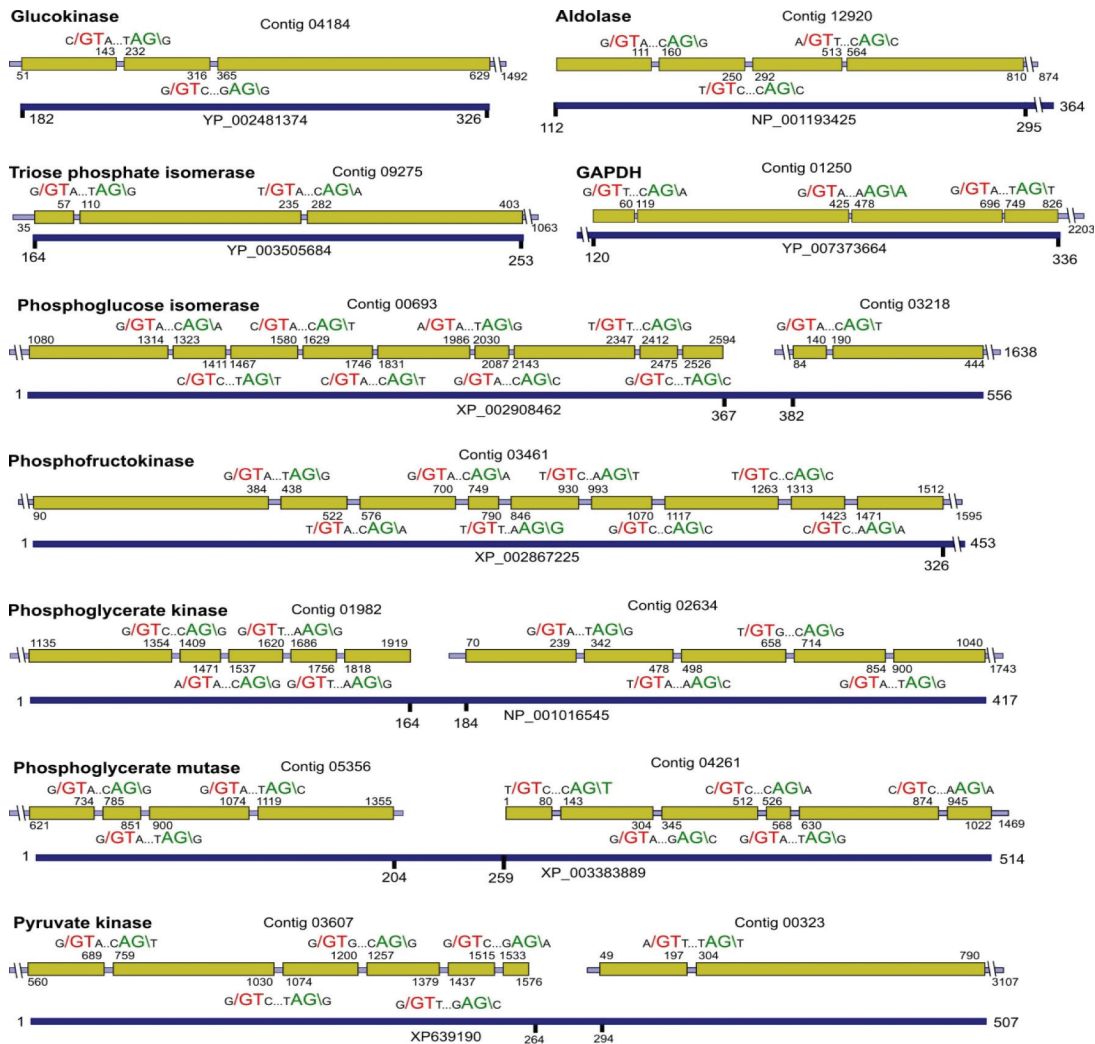
Figure 2. Gene structure of *Spongospora subterranea* genes involved in glycolysis. Yellow bars correspond to predicted exons for each contig. The sequence of each splice is shown. Blue bars correspond to the closest protein BLASTP hit using the putative translation product; GenBank accession codes are shown below. Numbers correspond to nucleotide position for contigs and amino acids for the reference protein.

oxidative phosphorylation; the main energy source of aerobic organisms. It is known that mithochondria are abundant in plasmodiophorids (Clay and Walsh, 1997), which was confirmed by an overrepresentation of mtDNA sequences in our data, where nine NADH dehydrogenase subunits, a cytochrome b gene, three cytochrome oxidase subunits and three

ATP synthase genes were found (Gutiérrez *et al.*, 2014). Glycolysis is also thought to be important for eukaryotic microorganism spore survival and germination. Consistent with this hypothesis, Heinz *et al.* (2012), in a proteomic study of *Trachipleistophora hominis* microsporidian found that most of the enzymes needed for glycolysis were expressed. An

interesting result was the enrichment of PGK in the cytosol of spores, suggesting a role for glycolysis in *T. hominis* spore ATP generation. Something similar was also reported for *Antonospora locustae*, which suggests that glycolysis is mainly used to make ATP in spores rather than vegetative cells (Dolgikh *et al.*, 2011). An understanding of the mechanisms regulating the formation of zoospore and cystosori in *S. subterranea* is necessary for designing appropriate control measures; the data presented here could be used in further studies investigating the expression of glycolytic enzymes at different stages of the *S. subterranea* lifecycle.

α-glucosidases. Potato tubers are characterized by the storage of carbohydrates in the form of starch granules. The breakdown of starch normally requires the action of α-amylases and β-amylases that cleave α-1,4-glucosidic bonds and debranching enzymes that cleave the α-1,6-glucosidic bonds of amylopectin

(Smith *et al.*, 2005). Two contigs, 02507 and 06526, were identified as α-glucosidases of the glycosyl hydrolase family 31, a group of enzymes that includes maltase-glucoamylase, sucrase-isomaltase and lysosomal acid alpha-glucosidase (Henrissat *et al.*, 1995). Contig 02507 encodes a partial protein product truncated at the C-terminus and split into five exons (Figure 3A). The corresponding protein product aligned in a region corresponding to 143-436 a.a of protein DDB_G0269790 from *D. discoideum* (867 a.a, XP_646281) with 48% identity. The N-terminus is characterized by the presence of a galactose mutarotase-like domain, which is found in proteins with isomerase activity, carbohydrate binding and aldose 1-epimerase activity (You *et al.,* 2010). BLASTP hits included α-xylosidases, α-glycosidases and lysosomal α-glycosidases. Phylogenetic analysis suggests that this contig is more closely related to enzymes of the latter group (Figure 3B). Lysosomal α-glucosidases are acid resistant enzymes essential for the degradation
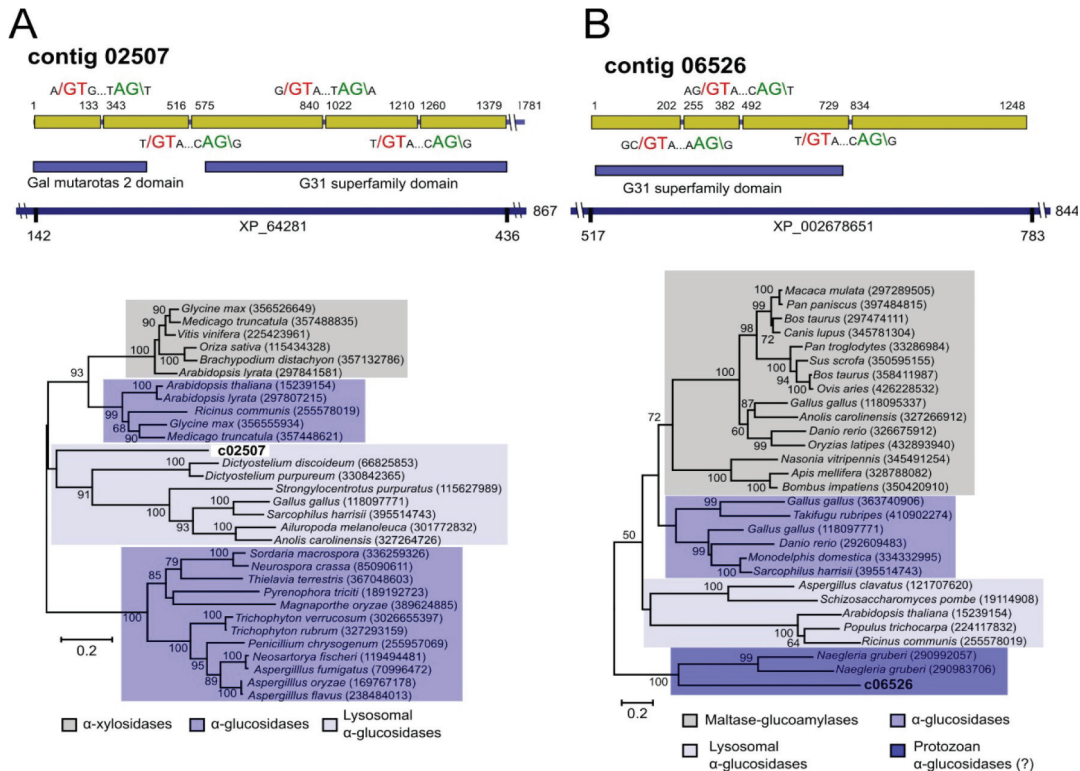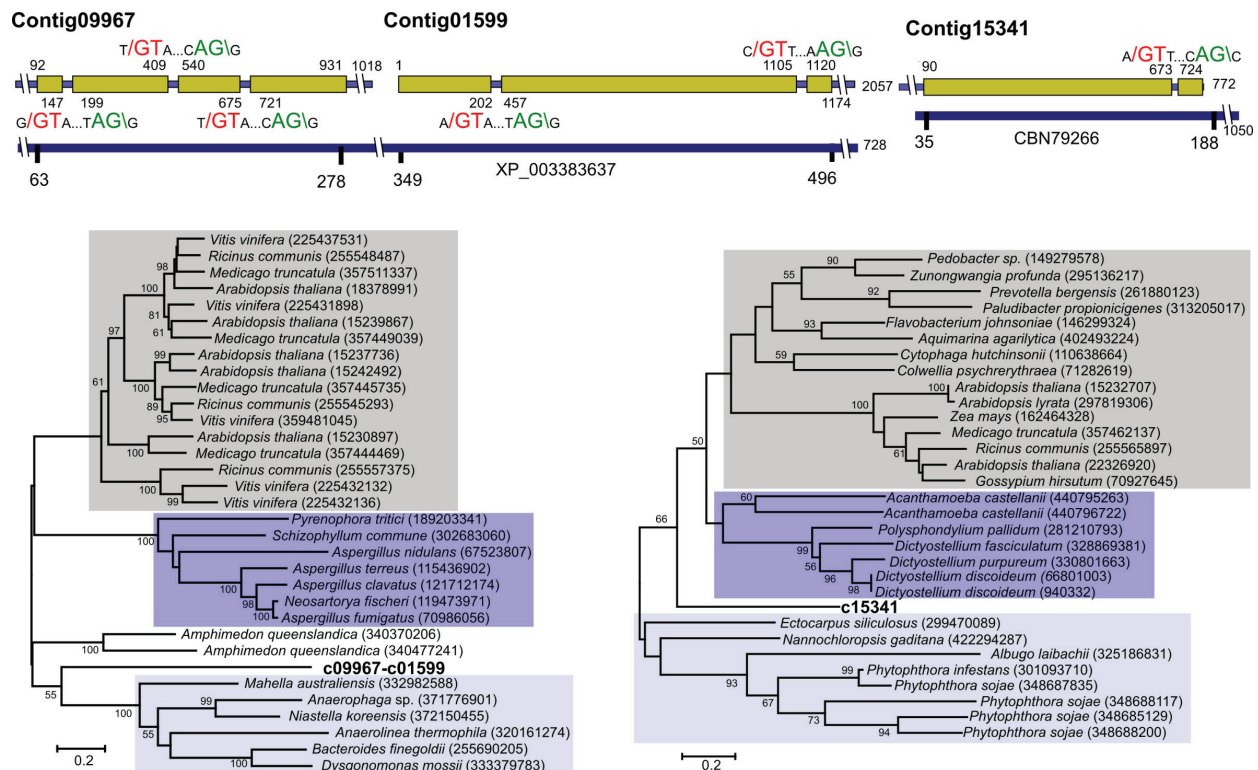

Figure 3. Structure of genes associated with α-glucosidase activity in *Spongospora subterranea*. **A.** Putative domains and the relative position with respect to the closest protein BLAST hit are shown below. **B.** NJ trees illustrate the phylogenetic affinity of the corresponding translated product with members of the G31 family of proteins. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test is shown next to the branches. Trees are drawn to scale, with branch lengths in units of the number of amino acid substitutions *per site*.

of starch to glucose in lysosomes (Jamil *et al.*, 2011). Contig 06526 clustered with a small clade comprising α-glucosidases more closely related to a predicted protein from the protist *Naegleria gruberi* (XP_002678651) (Figure 3B). Alignment comprises residues 517-783 from this 844-residue protein with 39% identity; the coding regions are split into four exons (Figure 3A). As the regions encoded by contigs 02507 and 06526 do not overlap, the possibility that both contigs encode the same protein product cannot be discarded. No obvious homologs of secreted α-amylases and β-amylases were found.

Cellulases. Nonstarch polysaccharides in the potato comprise the cell wall and the cementing substances between the cells, such as cellulose, crude fiber, pectic substances and hemicellulose (Smith, 1975). Xylan, the major component of hemicelluloses, is very heterogeneous and requires the action of a wide variety of enzymes, such as β-1,4-endoxylanase, β-D-xylosidase, α-L-arabinofuranosidase, α-D-glucuronidase, acetylxylan esterase, and phenolic acid esterase (Sunna and Antranikian, 1997). These enzymes hydrolize

xylan into its constituting sugars and have been found to be widely present in microorganisms, fungi and animals (Minic *et al.*, 2004). These enzymes are also characterized by broad substrate specificity and are important for the assimilation of glycosides, recycling of cell wall components, and modification of free glycosides (Faure, 2002). It is very likely that part of the infection mechanism of *S. subterranea* involves the action of enzymes involved in the degradation of cellulose.

Contigs 09967, 01599 and 15341 were identified as encoding family 3 glycosyl hydrolases, a group that includes enzymes involved in binding β-glucan such as β-D-glucosidases, α-L-arabinofuranosidases, β-D-xylopyranosidases and N-acetyl-β-D-glucosaminidases (Harvey *et al.*, 2000). In many cases, these enzymes have dual or broad substrate specificities with respect to monosaccharide residues, linkage position and chain length of the substrate. Contig 09967 is related to a β-D-xylosidase from the sponge *Amphimedon queenslandica* (XP_003383637). The *S. subterranea* protein aligns to residues 63-278 of
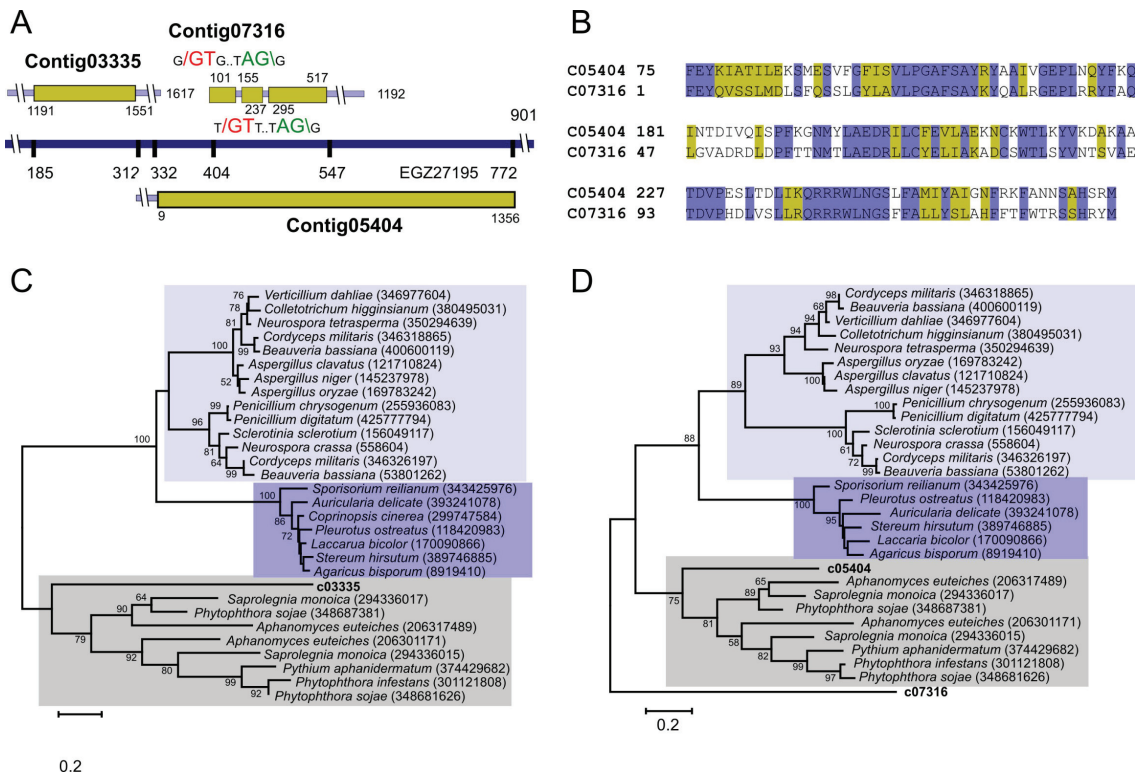

Figure 4. Structure of genes associated with cellulase activity in *Spongospora subterranea*. Contigs 09967 and 01599 probably encode different segments of the same protein. NJ trees illustrate the phylogenetic affinity of the corresponding translated product with members of the family of 3-glycosyl hydrolases.

this 728-residue protein with 52% identity (Figure 4). Contig 01599 seems to correspond to an internal region of the same protein aligning with residues 349-496 with 28% identity. A phylogenetic analysis using the combined protein products of contigs 00967 and 01599 showed that this protein forms an independent cluster within β-D-xylosidases (Figure 4), these enzymes cleave the xylan backbone and hydrolyze xylo-oligosaccharides and xylobiose from their nonreducing ends to liberate D-xylose (Sunna and Antranikian, 1997). However, some enzymes from this group may act as bifunctional α-L-arabinofuranosidase/β-D-xylosidases (Faure, 2002). Contig 15341 is related to the N-terminal end of a β-glucosidase from the filamentous brown algae *Ectocarpus siliculosus* (CBN79266), a 1,050-residue protein sharing 45% identity at the amino acid level and aligning in the 35-177 region.

Chitin metabolism. The formation of cystosori is one of the most important biological features of Plasmodiophorids. These structures are irregular aggregates of resting spores of about 19-85 μm in diameter and have been shown to contain chitin, a polymer of N-acetylglucosamine, which serves as a structural component in the extracellular matrices of cell walls of fungi and the exoskeletons of arthropods. Composition studies on closely related *P. brassicae* concluded that the spore wall dry weight is composed of 25.1% chitin, ≥ 2,5% other carbohydrates, 33.6% protein and ≥ 17.5% lipid (Moxham and Buczacki, 1983). The biosynthetic machinery of chitin is highly conserved and involves several enzymes of which chitin synthase is the most important. This protein is an integral glycosyltransferase membrane that catalyzes the incorporation of GlcNAc from UDP-GlcNAc into chitin, a linear homopolymer of GlcNAc residues formed by covalent β-1,4 linkages (Merzendorfer, 2011). Studies with fungi have revealed that most of them contain more than one chitin synthase gene. Contigs 03335, 05404 and 07316 were identified as homologs of chitin synthase (Figure 5A, B). Contig 03335 contains the C-terminal domain involved in the incorporation of GlcNAc into chitin common
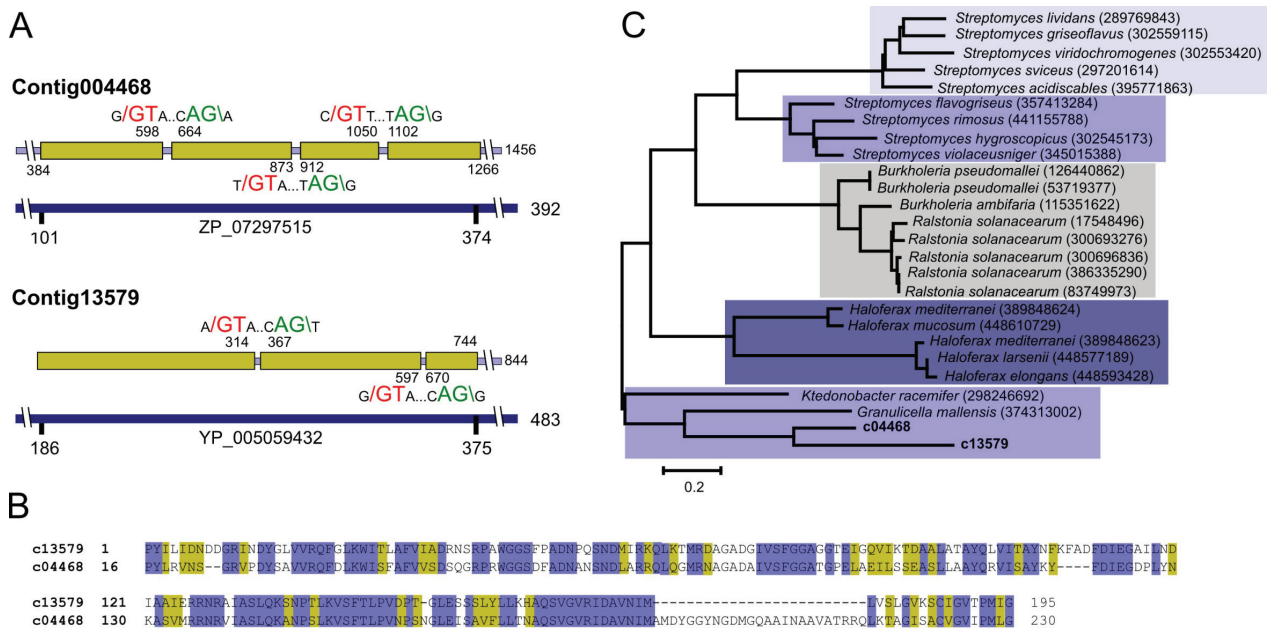

Figure 5. Structure of genes associated with chitin synthase activity in *Spongospora subterranea*. **A)** The relative position of each contig with respect to the closest protein BLAST hit is shown. **B)** Sequence alignment of the corresponding protein products encoded by contigs 05404 and 07316. **C)** and **D)** show phylogenetic trees of chitin synthase protein for contigs 03335, 05404 and 07316.

in chitin synthases classes I, II and III (Bowen *et al.,* 1992). The closest BLAST hit corresponded to chitin synthase 2 from the oomycete *Aphanomyces euteiches* (ACA49151), a plant pathogen responsible for root rot in several plants (Heyman, 2008). C03335 contains no introns and aligns in the region comprising residues 116 to 251 of this 802-residue protein with 40% identity. The closest hit to contig 05404 was the hypothetical protein PHYSODRAFT_470234 from *P. sojae* (EGZ27195), sharing 42% identity between residues 332-775. Contig 07316 is related to the C-terminal domain of chitin synthase from *Pythium aphanidermatum* (AEZ51041), a soil-borne plant pathogen Oomycete and causal agent of root rot in crops of economic importance, such as beets, peppers, chrysanthemum, cucurbits, cotton and turf grasses (El-Tarabily *et al.,* 2009). This 901-residue protein aligns at positions 413-454 with 56% identity (Figure 5C, D).

Another important enzyme in chitin metabolism is chitinase, which hydrolyzes the β-1,4-linkages in chitin. Chitinases are members of the glycoside hydrolase 18 family of proteins, an ancient gene family widely expressed in archaea, prokaryotes and eukaryotes (Henrissat, 1999). Chitinases, along with chitin synthetases, are essential for remodeling chitin-containing structures during growth and development (Funkhouser and Aronson, 2007). Two contigs were identified as chitinases: contig 4468 and 13579 (Figure 6A). Both sequences exhibit a PF-ChiA motif that has been shown to exhibit hydrolytic activity toward both colloidal and crystalline chitins (Nakamura *et al.,* 2007). Apparently, these proteins are paralogous, sharing 52% identity and 68% similarity (Figure 6B ,C).

Sugar transport. A BLASTX search of sugar transporters using a local database identified contig 00720 as a
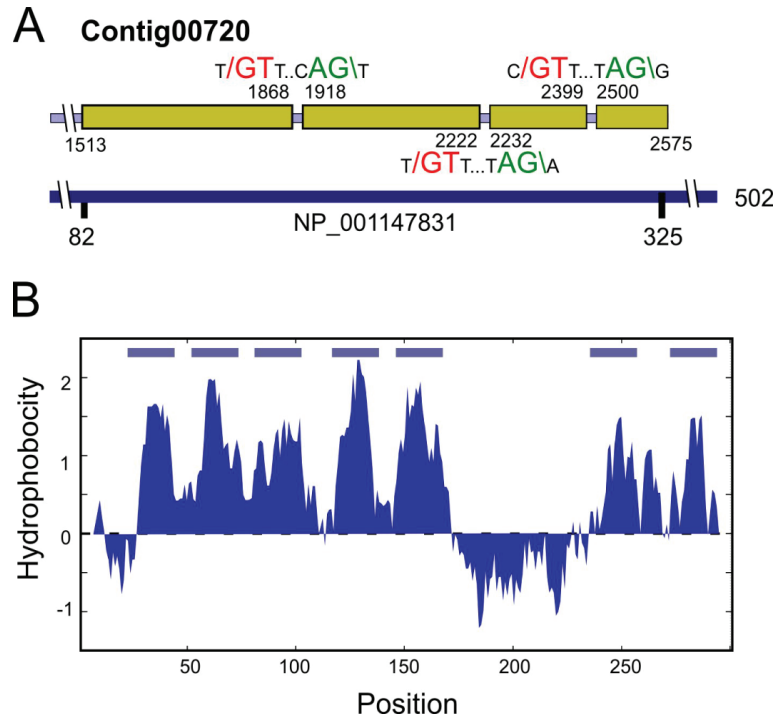

Figure 6. Structure of genes associated with chitinase activity in *Spongospora subterranea*. **A)** The relative position with respect to the closest protein blast hit is shown below. **B)** Sequence alignment of the corresponding protein products encoded by contigs 04468 and 13579 **C)** NJ tree illustrates the phylogenetic affinity of the corresponding translated product with members of the glycoside hydrolase 18 family.

protein belonging to the Major Facilitator Superfamily (MFS). This group comprises an extensive set of secondary transporters that facilitate the transport of substrates across membranes, such as ions, sugar phosphates, nucleosides, drugs and amino acids, among others (Saier, 2000). The closest BLASTP hit corresponded to sugar transporter ERD6-like

6-like isoform 1 from *Brachypodium distachyon* (XP_003567863), a protein of 502 residues with an identity of 36%. The predicted coding region of contig 00720 contains three introns and corresponds to residues 82-325 of *B. distachyon* protein, about 48% of the complete protein (Figure 7A). The N-terminal portion of the contig contains a motif characteristic

of the sugar porter (SP) subfamily, which is the largest group in MFS (Saier, 2000). Sugar porters are found in bacteria, archaea and eukarya and function by uniport or proton-coupled symport modes of transport (Pao *et al.*, 1998). MFS proteins normally contain 12 transmembrane regions, seven of which were identified in contig c00720 (Figure 7B).

Figure 7. Structure of contig coding for a sugar transporter in *Spongospora subterranea*. **A)** The relative position of each contig with respect to the closest protein blast hit is shown. **B)** Prediction of transmembrane helical domain for this predicted membrane protein.

CONCLUSIONS

The only way *S. subterranea* can be obtained outside the host plant is in the form of cystosori, which usually contain host plant tissues and are contaminated with soil microorganisms. For this reason, the 454 pyrosequencing technology in combination with bioinformatic applications is a good alternative for basic studies of this important plant pathogen. This analysis of partial genomic sequences provides insights into the structure of genes involved in the central metabolism of *S. subterranea*. A total of 18 genes (24 contigs) were identified as part of glycolysis, polysaccharide hydrolysis, chitin metabolism and sugar transport. Most of these genes were characterized by a complex intron/exon structure that should be confirmed using mRNA analysis of the corresponding transcripts.

The *S. subterranea* sequences reported here will allow for the primer design and development of molecular probes aimed at detecting and characterizing these enzymes *in vitro*. This information can also be used in detecting *S. subterranea* infected plants using methods such as qPCR or appropriate serological methods aimed at unique sequences of *S. subterranea*, especially as most of the genes presented in this study are expected to have high levels of expression. Future work should aim to complete the gene sequences and study their regulatory mechanisms.

BIBLIOGRAPHY

Archibald, J.M. and P.J. Keeling. 2004. Actin and ubiquitin protein sequences support a cercozoan/foraminiferan ancestry for the plasmodiophorid plant pathogens. Journal of Eukaryotic Microbiology 51(1): 113-118.

Bowen, A.R., J.L. Chen-Wu, M. Momany, R. Young, P.J. Szaniszlo and P.W. Robbins. 1992. Classification of fungal chitin synthases. Proceedings of the National Academy of Sciences of the United States of America 89(2): 519-523.

Braselton, J.P. 1995. Current status of the plasmodiophorids. Critical Reviews in Microbiology 21(4): 263-275.

Bulman, S.R. and J.W. Marshall. 1998. Detection of *Spongospora subterranea* in potato tuber lesions using the polymerase chain reaction (PCR). Plant Pathology 47: 759-766.

Bulman, S., J. Siemens, H.J. Ridgway, C. Eady and A.J. Conner. 2006. Identification of genes from the obligate intracellular plant pathogen *Plasmodiophora brassicae*. FEMS Microbiology Letters 264(2): 198-204.

Bulman, S., H.J. Ridgway, C. Eady and A.J. Conner. 2007. Intron-rich gene structure in the intracellular plant parasite *Plasmodiophora brassicae*. Protist 158(4):423-433.

Bulman, S., J.M. Candy, M. Fiers, R. Lister, A.J. Conner and C.C. Eady. 2011. Genomics of biotrophic, plant-infecting plasmodiophorids using *in vitro* dual cultures. Protist 162(3): 449-461.

Clay, C.M. and J.A. Walsh. 1997. *Spongospora subterranea* f.sp. *nasturtii*, ultrastructure of the plasmodial-host interface, food vacuoles, flagellar apparatus and exit pores. Mycological Research 101(6): 737-744.

Dolgikh, V.V., I.V. Senderskiy, O.A. Pavlova, A.M. Naumov, G.V. Beznoussenko. 2011. Immunolocalization of an alternative respiratory chain in *Antonospora* (*Paranosema*) *locustae* spores: Mitosomes retain their role in microsporidial energy metabolism. Eukaryotic Cell 10:588-593.

Duplessis, S., C. Cuomo, Y. Lin, A. Aerts, E. Tisserant, C. Veneault-Fourrey, D. Joly, S.Hacquard, J.Amselem,

B.L. Cantarel, R.Chiu, P.M. Coutinho, N. Feau, M. Field, P. Frey, E. Gelhaye, J. Goldberg, M.G. Grabherr, Ch.D. Kodira, A. Kohler, U. Kües, E.A. Lindquist, S.M. Lucas, R. Mago, E. Mauceli, E. Morin, C. Murat, J.L. Pangilinan, R. Park, M.Pearson, H. Quesneville, N. Rouhier, S. Sakthikumar, A.A. Salamov, J. Schmutz, B. Selles, H. Shapiro, P. Tanguay, G.A. Tuskan, B. Henrissat, Y. Van de Peer, P. Rouzé, J.G. Ellis, P.N. Dodds, J.E. Schein, S.Zhong, R.C. Hamelin, I.V. Grigoriev, L.J. Szabo and F. Martin. 2011. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proceedings of the National Academy of Sciences 108(22): 9166-9171.

El-Tarabily, K.A., A.H. Nassar, G.E. Hardy and K. Sivasithamparam. 2009. Plant growth promotion and biological control of *Pythium aphanidermatum*, a pathogen of cucumber, by endophytic actinomycetes. Journal of Applied Microbiology 106: 13-26.

Faure, D. 2002. The family-3 glycoside hydrolases: from housekeeping functions to host-microbe interactions. Applied and Environmental Microbiology 68: 1485-1490.

Funkhouser, J.D. and N.N. Aronson. 2007. Chitinase family GH18: evolutionary insights from the genomic history of a diverse protein family. BMC Evolutionary Biology 7: 96.

Gil, J.F, P.A. Gutiérrez, J.M. Cotes, E.P. González and M. Marín. 2011. Caracterización genotípica de aislamientos colombianos del *Potato mop-top virus* (PMTV, Pomovirus). Actualidades Biológicas 33(94): 69-84.

Guerrero, O. 1997. Reconocimiento del hongo *Spongospora subterranea* causante de la roña de la papa en el Departamento de Nariño. En: Fedepapa (ed.). Curso de Manejo Sanitario del Cultivo de la Papa. Pasto, Colombia. 25 p.

Guerrero, O. 2000. La roña o sarna polvosa en el Departamento de Nariño. pp. 127-129. En: Fedepapa (ed). Papas colombianas con el mejor entorno ambiental. Segunda edición. Bogotá, Colombia.

Gutiérrez P, Bulman S, Alzate J, Ortíz MC, Marín M. Mitochondrial genome sequence of the potato powdery scab pathogen Spongospora subterranea. 2014. Mitochondrial DNA 25: 1-2.

Harvey, A. J., M. Hrmova, R. de Gori, J.N. Varghese, and G.B. Fincher. 2000. Comparative modeling of the

three-dimensional structures of family 3 glycoside hydrolases. Proteins 41: 257-269.

Heinz E, T.A. Williams, S. Nakjang, C.J. Noël, D.C. Swan, A.V. Goldberg, S.R. Harris, T. Weinmaier, S. Markert, D. Becher, J. Bernhardt, T. Dagan, C. Hacker and T.M. Embley. 2012. The Genome of the obligate intracellular parasite *Trachipleistophora hominis*: new insights into microsporidian genome dynamics and reductive evolution. PLoS Pathogens 8(10): 1-23.

Henrissat, B., I. Callebaut, S. Fabrega, P. Lehn, J.P. Mornon and G. Davies. 1995. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. Proceedings of the National Academy of Sciences of the United States of America 92(15): 7090-7094.

Henrissat, B. 1999. Classification of chitinases modules. EXS 87: 137-156.

Heyman, F. 2008. Root Rot of Pea Caused by *Aphanomyces euteiches*. Doctoral thesis. Faculty of Natural Resources and Agricultural Sciences. Swedish University of Agricultural Sciences, Uppsala. 38 p.

Jamil, S., S. Ahmed and M. Tariq. 2011. Acid maltase deficiency-Pompe's disease. Journal of Pakistan Medical Association 61(8): 821-823.

Jaramillo, S. y J.M. Botero. 2007. Respuesta de diferentes poblaciones de *Spongospora subterranea* f. sp. *subterranea* a la rotación entre dos variedades de papa (*Solanum tuberosum* ssp. *andigena*). Revista Facultad Nacional de Agronomía Medellín 60(2): 3859-3876.

Keeling, P.J., N. Corradi, H.G. Morrison, K.L. Haag, D. Ebert, L.M. Weiss, D.E. Akiyoshi and S. Tzipori. 2010. The reduced genome of the parasitic Microsporidian *Enterocytozoon bieneusi* lacks genes for core carbon metabolism. Genome Biology and Evolution 2(1): 304-309.

Latvala-Kilby S., J.M. Aura, N. Pupola, A. Hannukkala and J.P. Valkonen. 2009. Detection of *Potato mop-top virus* in potato tubers and sprouts: combinations of RNA2 and RNA3 variants and incidence of symptomless infections. Phytopathology 99(5): 519-531.

Liu, R., Z. Yu, H. Guo, M. Liu, H. Zhang and M. Yang. 2012. Pyrosequencing analysis of eukaryotic and bacterial communities in faucet biofilms. Science of the Total Environment 435:124-131.

Merz, U. and R.E. Falloon. 2009. Review: Powdery scab of potato-increased knowledge of pathogen biology and disease epidemiology for effective disease management. Potato Research 52: 17-37.

Merzendorfer, H. 2011. The cellular basis of chitin synthesis in fungi and insects: common principles and differences. European Journal of Cell Biology 90(9): 759-769.

Minic, Z., C. Rihouey, C.T. Do, P. Lerouge and L. Jouanin. 2004. Purification and characterization of enzymes exhibiting beta-D-xylosidase activities in stem tissues of *Arabidopsis*. Plant Physiology 135(2): 867-878.

Moxham, S.E. and S.T. Buczacki. 1983. Chemical composition of the resting spore wall of *Plasmodiophora brassicae*. Transactions of the British Mycological Society 80(2): 297-304.

Nakamura, T., S. Mine, Y. Hagihara, K. Ishikawa and K. Uegaki. 2007. Structure of the catalytic domain of the hyperthermophilic chitinase from *Pyrococcus furiosus*. Acta Crystallographica Section F: Structural Biology and Crystallization Communications 63(1): 7-11.

Osorio, I., P.A. Gutiérrez y M. Marín. 2012. Revisión: *Spongospora subterranea* f.sp. *subterranea* y su virus asociado *Potato mop-top virus* (PMTV), dos patógenos reemergentes en los cultivos de papa de Colombia. Revista Facultad Nacional de Agronomía Medellín 65(1): 6361-6378.

Pao, S.S., I.T. Paulsen and M.H. Saier. 1998. Major facilitator superfamily. Microbiology and Molecular Biology Reviews 62(1): 1-34.

Qu X. and B.J. Christ. 2007. *In vitro* culture of the obligate parasite *Spongospora subterranea* (Cercozoa; Plasmodiophorida) associated with root-inducing transferred-DNA transformed potato hairy roots Journal of Eukaryotic Microbiology 54(6): 465-467.

Saier, M.H. 2000. Families of transmembrane sugar transport proteins. Molecular Microbiology 35(4): 699-710.

Saitou, N. and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing

phylogenetic trees. Molecular Biology and Evolution 4: 406-425.

Schwarz, R. and M. Dayhoff. 1979. Matrices for detecting distant relationships. pp. 353-358. In: Dayhoff, M (ed.) Atlas of protein sequences. National Biomedical Research Foundation, Washington, USA.

Siemens, J., S. Bulman, F. Rehn, and T. Sundelin. 2009. Molecular biology of *Plasmidiophora brassicae*. Journal of Plant Growth Regulation 28: 245-251.

Sistema Nacional de Información de Papa (SINAIPA). 2002. La roña de la papa. En: El Correo de la Papa. Boletín mensual No. 7, http://www.cevipapa.org.co/publicaciones/publicaciones.php.; consulta: noviembre 2011.

Smith, O. 1975. Potatoes: production, storing, processing. 3rd ed. Avi Publishing Co Inc., Conneticut, USA. 715 p.

Smith, A.M., S.C. Zeeman and S.M. Smith. 2005. Starch degradation. Annual Review of Plant Biology 56: 73-98.

Spanu, PD., J.C. Abbott, J. Amselem, T.A. Burgis, D.M. Soanes and K. Stüber. 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science 330: 1543-1546.

Sunna, A. and G. Antranikian. 1997. Xylanolytic enzymes from fungi and bacteria. Critical Reviews in Biotechnology 17: 39-67.

Tamura K., D. Peterson, N. Peterson, G. Stecher, M. Nei M and S. Kumar. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28(10): 2731-2739.

Webster, J. and R.W.S. Weber. 2007. Introduction to fungi. 3rd ed. Cambridge University Press, UK. 846 p.

You, W., X. Qiu, Y. Zhang, J. Ma, Y. Gao, X. Zhang, L. Niu and M. Teng. 2010. Crystallization and preliminary X-ray diffraction analysis of the putative aldose 1-epimerase YeaD from *Escherichia coli*. Acta Crystallographica Section F: Structural Biology and Crystallization Communications 66(8): 951-953.