University of Tuscia

Department for Innovation in Biological, Agro-food and Forest systems – DIBAF

Ph.D. Course**: Science, Technology and Biotechnology for Sustainability**

Curriculum**: Forest Ecology and Environmental Technologies**

Functional and comparative genome analysis of microorganisms suitable for the developing of microbe assisted phytotechnologies

(AGR/05)

Ph.D. candidate: Andrea Firrincieli

Tutor: Prof. Giuseppe Scarascia Mugnozza

Ph.D. Coordinator: Prof. Mauro Moresi

Academic Year 2016/2017

Summary

Shor Abstract **4**

Extended Abstract **4**

Aim of this Work **6**

State of the art **7**

Chapter 1: Comparative genome analysis of *Rhodotorula graminis* WP1

Introduction **10**

Material and methods **11**

Results and discussion **15**

Conclusion **38**

Chapter 2: Arsenic speciation and bioaccumulation in *Rhodococcus atherivorans* BCP1

Introduction **40**

Material and methods **41**

Results and discussion **44**

Conclusion **57**

Conclusions and future perspectives **57**

List of papers **57**

References **58**

Supplementary material **67**

Short Abstract

Microbial assisted phytoremediation is a promising remediation technique, which exploit the utilization of plant associated microorganism for the rehabilitation of polluted environments. The aim of this study was to dissect the biotechnological potential of *Rhodotorula graminis* WP1 and *Rhodococcus atherivorans* BCP1 as candidate microorganisms for the developing of a microbial consortia capable to couple important aspects of microbe assisted phytotechnologies (e.g. improved growth performances and increased plant tolerance to environmental stresses) to phytoremediaton. In this context, a comparative genome analysis was performed on WP1 to highlight specific genetic features associated with plant-microbe interactions, and characterize the presence of virulence traits typical of the *Pucciniomycotina* phytopathogens. Furthermore, from the perspective phytotechnologies, the genetic potential of WP1 involved in the cellular homeostasis of heavy metals, mainly arsenic, was analyzed. Finally, different aspects regarding the homeostasis and response of BCP1 to As(V) and As(III) were investigated using both physiologic and bioinformatic approaches.

Extended Abstract

Rhodotorula graminis WP1 and *Rhodococcus atherivorans* BCP1 were previously characterized for their applicative potential in the field of microbe assisted phytotechnologies and bioremediation respectively. In particular, the red yeast *R. graminis* WP1, isolated from healthy poplar trees, promotes plant growth in numerous plant hosts by producing plant hormones such as salicylic acid, abscisic acid, indole-3-aceticacid, jasmonic acid, gibberellins-3-acid and epibrassinolides, which are important in plant growth promotion, response to phytopathogens, and plant tolerance against abiotic stresses. In the context of bioremediation, strains of the *Rhodococcus* genus are ubiquitous in nature, especially in plants and soil, and are ideal microorganisms for the implementation of technologies aimed to the removal of organic pollutants. The remarkable capacity to co-metabolize a very wide range of xenobiotic compounds and their environmental adaptability granted by the synthesis of mycolic acids which confer the high tolerance of these bacteria to antibiotics and other forms of environmental stress, makes these bacteria suitable for *In Situ* bioremediation approaches.

***Rhodotorula graminis* WP1.** In-silico analysis of the draft genome and transcriptome of WP1 was performed to reveal genomic features important for the growth in association with plant. Using a comparative approach, we focused our analysis in the characterization of WP1 secretome (small secreted proteins), as molecular elicitors involved in the interaction with the host and the surrounding microbiome, and the genetic effectors involved in plant colonization such as genes involved in the biosynthesis of secondary metabolites acting as plant hormones was assessed. Furthermore, the

presence of genes with beneficial effect on improving plant tolerance to heavy metals in the context of microbe assisted phytoremediation was investigated. These results highlight for the first how important virulence traits and genetic effectors were differentially selected in WP1 in the context of its lifestyle as non-pathogenic plant associated microorganism.

***Rhodococcus atherivorans* BCP1.** Arsenic ranks among the priority metals that are of public health significance. In the environment, the metalloid arsenic mainly exists under two forms: the arsenite As(III) and arsenate As(V); the former being more toxic due to its high mobility and stability. Bacteria have developed multiple strategies for arsenic detoxification. *Rhodococcus aetherivorans* BCP1 can cometabolize chlorinated compounds, mineralize a wide range of hydrocarbons, resist to different stress conditions and convert tellurite and selenite into less toxic forms, making this strain an ideal candidate for microbial biotechnology applications. In this study, we show that BCP1 tolerate high concentrations of As(V) during its growth under aerobic conditions. Furthermore, different aspects regarding the arsenic homeostasis and the response of BCP1 to As(V) were investigated: (i) the different capability to convert As(V) into As(III) depending on the initial concentration of arsenate; (ii) the arsenic biosorption; (iii) the effect of arsenic on polyphosphate granule formation and (iv) the genetic/genomic aspects involved in arsenic detoxification. Finally, the detection of electrondense nanoparticles after the incubation with As(V) suggested the ability of BCP1 strain to generate As-based nanostructures.

Conclusions. The results presented in this study highlights how BCP1 and WP1 would be suitable candidate for the developing of microbe assisted phytotechnology aimed to increase the uptake and sequestration of As(III) and for the decontamination of polluted aqueous solution by organic pollutant and heavy metals. Also, the capability of WP1 to increase plant growth would also drastically improve the plant fitness in presence of toxic contaminants, and phytoremediation performances as well.

Keywords

Comparative genome analysis; Microbe-assisted phytoremediation; endophyte; phytopathogens; arsenic; *Pucciniomycotina, Rhodococcus, Rhodotorula*

Aim of the Work

The selection of a microorganism suitable for the implementation of microbe assisted phytotecnologies is achieved by assessing its potential for biological features commons in plant growth promoting microorganism. However, only through an in depth genetic analysis is possible to detect molecular effectors putatively involved in tolerance to abiotic and biotic stress, antibiotic production, plant growth promotion, plant colonization and adaptation to plant niches, providing an important source of information necessary to understand the genetic potential of a specific microorganism. In this context, genome sequences can be used for this purpose. The aim of this research project was to characterize from a genetic and functional perspective two different microorganisms suitable for the developing of microbe assisted phytotechnologies. The microorganism studied are the plant growth promoting yeast *Rhodotorula graminis* WP1 and the PCB and alkene degrader *Rhodococcus atherivorans* BCP1.

WP1 is an endophytic yeast isolated from *Populus trichocarpa*, and studied for its capability to synthetize a wide array of plant phytohormones: salicylic acid, abscisic acid, indole-3-aceticacid, jasmonic acid, gibberellins-3-acid and epibrassinolides. Because its phylogenetic affiliation to *Pucciniomycotina*, a taxonomic group characterized by important plant pathogens, wood-rotting fungi and parasites, WP1 is a good candidate for genome comparative studies aimed to define potential genetic traits positively selected for plant-host interactions and adaptation to plant niches. In this regard, the identification of major loss events of common virulent traits in *Pucciniomycotina* phytopathogens will be used to better define the evolutionary history of this WP1

BCP1 was characterized for its capability to cometabolize chlorinated compounds, mineralizes a wide range of alkanes (Cappelletti et al., 2013), making him a good candidate for the developing of microbe assisted phytotechnologies. Furthermore, from the analysis of BCP1 resistance to heavy metals by using phenotype microarray, a general tolerance to various heavy metals salts was observed (Orro et al., 2015).

Because of its high degree of inducing toxicity at low level of exposure, arsenic, along with cadmium, chromium, lead, and mercury, rank among the priority metal of public health significance (Tchounwou et al., 2012). In the environment, the metalloid arsenic mainly exists under two forms: the arsenite (As(III)) and arsenate (As(V)), the former being more toxic due to its high mobility and stability (Kumari and Jagadevan; 2017). Microorganisms play a crucial role in the environmental speciation of arsenic. The oxyanions enter in the biosphere from anthropogenic activities (i.e. pesticides, herbicides and fungicides) and biological leaching. The latter is the major reason why microorganism, mainly bacteria, have developed multiple strategies for arsenic detoxification (Lebrun et al., 2003). Indeed, arsenic is a common element in the Earth's crust and exists at an average

concentration of approximately 5 mg/kg (Garelick et al., 2008). Roughly, two major detoxification systems involved in the arsenic homeostasis exist in bacteria: i) arsenate detoxification occur through the conversion and consequent extrusion of arsenite outside the cell; ii) arsenite is either methylated into a volatile form or converted into arsenate by a cytochrome-c oxidase (arsenite oxidase), which is finally extruded outside the cell (Hudson-Edwards and Santini, 2013). The latter can use arsenite as electron donor coupled with the reduction of oxygen, nitrate or chlorate (Osborne and Santini, 2012).

In this work, we have investigated the arsenic homeostasis of BCP1. The *Rhodococcu*s genus possess exceptional capabilities to convert heavy metals to less toxic forms and increase plant performances exposed to the arsenic oxyanions. Shagol et al., (2014) reported that *Rhodococcus atherivorans JS2210* could increase plant tolerance to arsenic and enhance plant growth in presence of the heavy metal while Presentato et al., (2016) have reported that BCP1 is also able to produce intracellular nanorods of tellurite and nanoparticles of selenite. In this context, the response of BCP1 to arsenic and its capability to bioaccumulate the heavy metal has been investigated.

State of the Art

Soil contamination by heavy metals and organic pollutants is a major global environmental issue because of anthropogenic activities. In Europe, the number of potentially contaminated sites and the number of sites known to be contaminated and needing remediation has been estimated at 3.5 and 0.5 million, respectively (Majone et al. 2014). Bioremediation includes all those techniques that take advantage from the utilization of biological systems for the remediation of polluted soils and waters. In this context, phytoremediation depends on the use of plants for the extraction, degradation and immobilization and mobilization of organic and inorganic pollutants from environmental matrices. From an economic point of view, phytoremediation, compared to the traditional remediation approaches, offers the opportunity to take an advantage from contaminated agricultural lands, unsuited for food crops production (Jiang et al., 2015), since the biomass obtained from no-food cultivation may be used to produce biofuels in liquid or gas form (i.e. biodiesel and biomethane) and thus turning the whole phytoremediation process into a sustainable and cost-effective solution.

An understanding of the factors affecting phytoremediation is necessary to develop a well-performing system in term of pollutants removal and plant growth. In the case of the degradation of organic pollutants, the plant acts indirectly through the stimulation of microbial communities that populate the rhizosphere. The rhizosphere soil is a very selective environmental niche where biotic interactions between plant and soil microorganism are sustained through the production and secretion of root exudates (Bais et al., 2006). For this process, known as “phytostimulation”, the root exudate can

enhance the degradation of organic pollutants by promoting the growth of bacteria able to metabolize or co-metabolize organic pollutants (i.e. plant roots release phenolic compounds such as flavonoids, that stimulate the growth of PCB degrading bacteria) (Narasimhan et al., 2003). Yergeau et al. (2014) have studied the metatranscriptomic profile of microbial communities that populate the rhizosphere of willows growing in presence of polycyclic aromatic hydrocarbons. For these plants, genes involved in toluene and biphenyl degradation pathways, and in the degradation of chloroaromatic compounds were highly expressed by the microbial community populating the rhizosphere soils compared to bulk soil. Indeed, microbial consortia have been successfully used in association with plants growing under different regimes or in presence of various types of contaminants such as organic pollutant and heavy metals. Recently, in a study conducted on mesocosms contaminated with crude oil, polyaromatic hydrocarbons (PAHs), and heavy metals, through the utilization of microbial consortia in association with plant (*Vetiveria zizanioides*), Nanekar et al., (2015) achieved a 79 % removal of crude oil, coupling the almost complete mineralization of PAHs, with a marked decrease in heavy metals such as cadmium and lead. Similarly, by selecting specific microorganisms isolated from plants naturally growing on heavy metals contaminated soils, Langella et al., (2013) could identify optimal plant–microbe combinations to bioremove various heavy metals species such as Cd, Cu, Mn, Ni, Pb, Sr, Zn and U, showing how the preferential uptake of certain heavy metals species is deeply affected either by the presence of a microbial inoculum and by the plant species selected. This study suggests that a prior knowledge regarding the plant system and the genetic and biological potential of the microbial consortium is necessary to achieve an efficient decontamination process.

Recent advances in the microbe-assisted phytoremediation is represented using plant growth promoting microorganisms (PGPMs), coupling the bioremoval of organic and inorganic pollutants with an increased biomass production, improved tolerance toward biotic and abiotic stresses, and enhanced phytoremediation effectiveness. Indeed, a "common" but non-essential feature shared by microorganisms isolated from plant tissues, and, to some extent, by microorganisms isolated from plant environmental niches such as the rhizosphere and rhizoplane, is the capability to interfere/modulate primary plant metabolic pathways through the biosynthesis of elicitors, maintaining a delicate balance between the endophyte/rhizobacteria and plant. Regarding this aspect, phytohormones of microbial origin, the secretion of secondary metabolites with antimicrobial activity and microbial metabolisms specialized in nutrient uptake (i.e. phosphate and iron ions) have a profound impact on plant adaptation and response to environmental stresses and pathogens (Rungin et al., 2012; Mei and Flinn, 2010; Chen et al., 2016). By exploiting plant growth promoting bacteria belonging to *Burkholderia, Variovorax, Bacillus, Pseudomonas* and *Ralstonia* species, Thijs et al., (2014) were able to develop an efficient plant-microbe system capable to degrade the 2,4-DNT, and

improving also roots development and plant tolerance to the organic contaminant. Similarly, the nitrogen fixing bacteria *Burkholderia vietnamiensis* G4, isolated from contaminated wastewater (Fries et al., 1997), can grow in association with plants, (O'Sullivan and Mahenthiralingam, 2005) improving growth and tolerance to biotic stresses (biocontrol) (Chiarini et al., 2006). Finally, the poplar endophyte *Pseudomonas putida* PD1 improve plants tolerance to high concentration of phenanthrene, enhancing the removal of the organic pollutants by 25-40%, granting better growth performances compared to plants exposed to the toxicant without the presence of the endophyte (Khan et al., 2014).

Different strategies are used for the selection of PGPMs. In-depth studies are often conducted to dissect the genomic and physiological features of the microorganisms to better understand the biology behind the molecular mechanism involved in plant-microbe interaction regarding the capabilities to improve plant growth, and contaminants degradation. Indeed, is not uncommon that massive genome sequencing projects are aimed to define the metagenome (the genetic potential) of cultivable plant associated microbes (PRJNA365409). By dissecting the genetic information retrieved from such projects, an invaluable source of data can be exploited for the optimization and developing of microbes-assisted phytoremediation technologies.

Chapter 1

Comparative genome analysis of Rhodotorula graminis WP1

1. Introduction

Rhodotorula graminis WP1 was isolated in August 2002 from surface-sterilized shoot cuttings of poplar trees collected at the Three Forks Natural Area in King County, WA in the riparian zone of the Snoqualmie River (+47° 31' 14.30", -121° 46' 28.32") (Xin et al., 2009). Although WP1 was originally isolated from poplar, when inoculated onto corn, rice, turf grass, pepper, squash, and hybrid poplar, an enhanced plant growth and fruits yield was observed in nutrient-limited soils, suggesting a possible role of WP1 in the improvement of plant nutrient uptake and plant developing (Khan et al., 2012; Knoth et al., 2013; Knoth et al., 2014). Strain WP1 has also useful biochemical properties (Doty, 2014), including the ability to ferment both pentoses and hexoses and to degrade fermentation inhibitors (Xu et al., 2011).

From a phylogenetic point of view, WP1 belongs to *Pucciniomycotina*, a subdivision of Basidiomycota characterized by known plant pathogens such us rust fungi (*Cronartium quercuum f. sp. fusiforme, Melampsora allii-populina, Melampsora lini, Melampsora laricis-populina, Puccinia graminis, Puccinia stiriformis, Puccinia triticina*) (Tavares et al., 2014) and the anther smut fungus *Microbotryum violaceum* (Shykoff and Kaltz, 1997).

In general, depending on the nature of the interaction with the host, microorganism can be classified as: symbiont, endophyte and pathogens. Compared to endophytic and symbiotic microorganism, plant pathogens are usually classified as biotroph and necrotroph, depending on the type of "strategy/behavior" adopted during the acquisition of nutrients from the host. While biotroph extract nutrients from living tissues by establishing and maintaining long-term interactions, necrotroph acquire nutrients from tissues by actively killing the host cells through the secretion of a wide array hydrolytic enzymes. Either biotrophs and necrotrophic microorganisms possess the capability to modulate and suppress host immunity to maintain and complete their lifecycle within the host tissues (Okmen and Doehlemann, 2014).

On the other hand, the nature of the interactions between plant and endophytes, which are common plant inhabitants, is still poorly understood. To some extent, endophytes, just like the gut microbiome in humans, represent an extension of the host genetic potential, where biological function as the nutrients uptake, defense against pathogens and hormonal response to environmental stimuli can be only accomplished through a tight cross-talk between the plant and its microbiome. Endophytes

complete their life cycle within plant tissues without causing any symptoms of disease. However, under certain conditions are capable to switch from commensal or weak parasitism to pathogenic or mutualistic lifestyles (Comby et al., 2016). Such behavior strongly suggests how some endophytic fungi might have retained that genetic potential typical of the obligated biotroph counterpart, involved in the generation and maintenance of strong interaction with the host in response to biotic and abiotic stimuli. Because WP1 is capable to colonize and complete its lifecycle within the plant, this study will explore the predicted proteome and transcriptome, focusing on secretion factors (small secreted proteins) and hydrolytic enzymes with the aim of identifying genetic traits involved in its symbiosis mechanisms acting as host-specific effectors. Small secreted proteins represent then a means through which an organism interact with the surrounding environment. For this reason, functional aspects such as carbohydrate hydrolysis, proteases, protein-protein interaction domain, oxidoreductases, membrane transporter and pool of small secreted proteins in WP1 were investigated and compared against the predicted proteome of *Pucciniomycotina* species such as rust-fungi, parasites and non-endophytic fungi. Other than a functional aspect, the presence of multi-domain domain protein was investigated in order to link specific domain combination to biological processes such as host-pathogen interaction and elicitation of plant immune responses.

This *in silico* analysis combined with pathway assignment and protein annotation from multiple databases and from comparative genomic analysis with manual curation of select sequences

In this chapter will be discussed the genetic traits of WP1 putatively involved in plant colonization, adaptation to plant niches and interaction with the host, phytohormone production and resistance to heavy metals.

2. Materials and Methods

2.1. RNA extraction

To improve genes annotation, the total RNA was extracted to perform a transcriptome analysis. A single WP1 colony, harvested from solid nitrogen free media (NFMS), was used to inoculate 100ml of liquid +NF, grown with agitation (200-250 rpm) for 24 hours at RT. Cells were harvested and suspended in 100 ml of YPD broth and grown with agitation at RT. After two days, cells were spun down at 3400 rpm and then washed twice with NFMS media. Cell density was adjusted to 0.4 OD_{600} in fresh NFMS and grown with agitation (200-250 rpm) at RT until cell density reached an OD_{600} of 1.2. Cells were spun down at 3400 rpm, washed twice in NFMS media, frozen at -80 °C overnight, and lyophilized. Lyophilized biomass was ground in liquid nitrogen until a fine powder was achieved

and RNA extracted using standard protocols as suggested by the manufacturers of the reagent TRlzol (Life Technologies).

2.2. DNA Sequencing and assembly

The *R. graminis* strain WP1 genome was sequenced by the Joint Genome Institute (JGI) using the Sanger whole genome shotgun approach. The DNA extracted was sheared allowing to generate and sequencing three libraries with and insert size of 3, 6, and 33.8 kb. Raw sequenced reads were quality checked and filtered for vector sequence, mitochondria, unanchored rDNA and DNA contamination based on GC content. Filtered reads were finally assembled into contigs and supercontigs with the Arachne assembler (Jaffe et al., 2003).

2.3. Transcriptome sequencing and genome annotation

To improve gene detection and annotation the construction of an expression sequence tag (EST) library was performed by purifying the mRNA using PolyA. Purified mRNAs were subsequently retro-transcribed to construct cDNA libraries. The cDNA library was sequenced on a Roche-454 GS-FLX platform. The reads were filtered and screened for quality as described in 2.2 and finally assembled into contigs using Newbler (v2.3-PreRelease-6/30/2009) with default parameters. Prior to gene prediction, repeats in assembled scaffolds were masked using RepeatMasker (Smit et al., 2010), RepBase library (Jurka et al., 2005), and the most frequent (>150 times) repeats recognized by RepeatScout (Price et al., 2005). The following combination of gene predictors was run on the masked assembly: ab initio including Fgenesh (Salamov and Solovyev 2000a) and GeneMark (Ter-Hovhannisyan et al., 2008); homology-based including Fgenesh+ (Salamov and Solovyev 2000b) and Genewise (Birney and Durbin, 2000) seeded by BLASTx alignments against the NCBI NR database; and transcriptome-based using Fgenesh package. In addition, tRNAs were predicted using tRNAscan-SE (Lowe and Eddy, 1997).

The assembled genome was annotated in 2010 using the JGI Annotation Pipeline (Grigoriev et al., 2006), which combines several gene prediction and functional annotation methods, and integrates the annotated genome into JGI web-based resource for fungal comparative genomics, MycoCosm (http://genome.jgi.doe.gov/fungi) (Grigoriev et al., 2014). For each genomic locus, the best representative gene model was selected based on a combination of protein homology and transcriptome support. In this study, a manual curation of genome annotation was carried out through Blast2GO (Conesa et al., 2005). All predicted proteins were functionally annotated using SignalP (Nielsen et al., 1997) for signal sequences, TMHMM (Melén et al., 2003) for transmembrane domains, interProScan (Quevillon et al., 2005) for integrated collection of functional and structural

protein domains, and protein alignments to NCBI nr, SwissProt (UniProt Consortium, 2014), KEGG (Kanehisa et al., 2008) for metabolic pathways, and KOG (Koonin et al., 2004) for eukaryotic clusters of orthologs. InterPro and SwissProt hits were used to map Gene Ontology terms (Ashburner et al., 2000).

2.4 Phylogenetic classification of *Rhodotorula graminis* WP1 and synteny analysis

A Maximum likelihood tree with 1000 replicates was constructed concatenating the phylogenetic markers RNA polymerase subunit 1 (Rpb1), RNA polymerase subunit 2 (Rpb2) and the Translation elongation factor 1 alpha (TEF1) (Table 1). Protein sequences were downloaded from Mycocosm database and aligned against each other using Muscle (Edgar and Robert, 2004). Aligned sequence were finally concatenated and a phylogenetic tree was constructed in MEGA 6.0. Syntenic region were analyzed in SyMAP v4.2 (Soderlund et al., 2006). Orthologues clusters were detected using OrthoVenn (Wang et al., 2015)

Table 1. Molecular markers used for the phylogenetic classification of WP1: RNA polymerase subunit 1 (Rpb1), RNA polymerase subunit 2 (Rpb2) and Translation elongation factor 1 alpha (TEF1) protein IDs

Species	Protein IDs		
	Rpb1	Rpb2	TEF1
Agaricostilbum hyphaenes	Agahy1|287844	Agahy1|303066	Agahy1|263013
Atractiellales sp.	Atrsp2|1220916	Atrsp2|1277670	Atrsp2|1341582
Chionosphaera apobasidialis	Chiap1|575547	Chiap1|1148406	Chiap1|1170677
Chionosphaera cuniculicola	Chicu1|824675	Chicu1|826886	Chicu1|790770
Cronartium quercuum f. sp. fusiforme	Croqu1|651104	Croqu1|75937	Croqu1|53309
Erythrobasidium hasegawianum	Eryha1|1082531	Eryha1|1052852	Eryha1|1072163
Heterogastridium pycnidioideum	Hetpy1|286935	Hetpy1|248646	Hetpy1|289232
Hyalopycnis blepharistoma	Hyabl1|647022	Hyabl1|629774	Hyabl1|651275
Melampsora allii-populina	Melap1finSC_191|1455444	Melap1finSC_191|1696169	Melap1finSC_191|1312015
Melampsora lini	Melli1|206525	Melli1|202015	Melli1|201958
Melampsora laricis-populina	Mellp2_3|50989	Mellp2_3|47615	Mellp2_3|73378
Microbotryum violaceum	Micvi1|2552	Micvi1|862	Micvi1|2806
Mixia osmundae	Mixos1|95435	Mixos1|14551	Mixos1|89758
Puccinia graminis	Pucgr2|13322	Pucgr2|13448	Pucgr2|11829
Puccinia stiriformis	Pucst1|500184	Pucst1|499040	Pucst1|493585
Puccinia triticina	Puctr1|1101	Puctr1|7720	Puctr1|2073
Rhodotorula graminis WP1	Rhoba1_1|54819	Rhoba1_1|66967	Rhoba1_1|40458
Rodotorula minuta	Rhomi1|1964	Rhomi1|50033	Rhomi1|71099
Rhodosporidium turoloides NP11	Rhoto1|7984	Rhoto1|878	Rhoto1|5076
Sporobolomyces lindare	Spoli1|204427	Spoli1|186502	Spoli1|184981
Sporobolomyces roseus	Sporo1|14395	Sporo1|22879	Sporo1|9732
Lipomyces starkeyi	Lipst1_1|75854	Lipst1_1|73345	Lipst1_1|63951

2.5 Assessing the presence of horizontal gene transfer (HGT) events *in Rhodotorula graminis* WP1

The presence of genes acquired through HGT events was assessed using a phylogenetic based approach in combination with a BLAST database search for best hits. Part of the analysis was performed using the pipeline described in HGTector (Zhu et al., 2014). Because the number of fungal

genomes available in databases is too low, to fully define the phylogenetic history of WP1 proteome, the following phylogenetic scenario was adopted to detect candidate HGT genes from prokaryotes: (i) self-group, fungi; (ii) close-group, eukaryotes; (iii) distal-group, prokaryotes. Self and close groups contain all those species strongly involved in a "vertical" gene transfer with WP1 while the distal group represent all the species that are phylogenetically distant from WP1. The BLASTP (Altschul et al., 1990) parameter used in HGTector were: minimum identity, 45%; max number of returning hit per sequence, 100; e-value, $1e^{-20}$.

2.6 In-silico detection of small secreted proteins (SSPs) in *Rhodotorula graminis* WP1 and *Pucciniomycotina* species

SSPs proteins in WP1 and *Pucciniomycotina* species were analyzed as follow. For each species, protein sequences with a signal peptide were downloaded from Mycocosm database. An in-house script was used to filter out all those sequence longer than 300 amino acids. Only sequences with a length equal or lower than 300 amino acids were kept and analyzed for the presence of transmembrane (TM) motif using TMHMM. Those sequences without TM motifs where analyzed using TargetP (Emmanuelson et al., 2000) to remove putative mitochondrial protein. For the comparative analysis of SSPs in *Pucciniomycotina* the following species were used: *Agaricostilbum hyphaenes, Atractiellales sp., Chionosphaera apobasidialis, Chionosphaera cuniculicola, Cronartium quercuum f. sp. fusiforme, Erythrobasidium hasegawianum, Heterogastridium pycnidioideum, Hyalopycnis blepharistoma, Melampsora allii-populina, Melampsora lini, Melampsora laricis-populina, Microbotryum violaceum, Mixia osmundae, Puccinia graminis, Puccinia stiriformis, Puccinia triticina, Rhodotorula graminis* WP1, *Rodotorula minuta , Rhodosporidium turoloides* NP11, *Sporobolomyces lindare, Sporobolomyces roseus*

2.7 Cluster analysis and functional annotation of SSPs

Putative SSPs from species listed in 2.6 were aligned "all against all" using BLAST and clustered together using E-values. TribeMCL (Enright et al., 2002), an implementation of The Markov Cluster Algorithm, was used in combination with SPCPS (Nepusz et al., 2010) to convert E-values in similarity values. The latter was used for the clusterization of SSPs. Protein annotation from each cluster of SSPs was manually curated in Blast2GO, by associating SSFAM, PFAM and IPR terms to the biological functions defined by Gene Ontology (GO) terms. Clusters containing annotated WP1 SSPs were analyzed as follows: (i) all proteins from each cluster were retrieved and aligned against each other in MUSCLE; (ii) alignments were edited in order to remove sequences with a poorly aligned. Finally, aa sequences were analyzed in Gblock (Castresana 2000) to retrieve only conserved

aa blocks. Parameters used in GBlock for the selection of conserved regions were: “allow smaller final blocks”; “allow gap position within the final blocks”, and “allow less strict flanking position”. Aa blocks obtained from each cluster were then analyzed in PhyML (Guindon et al., 2010) using the Approximate Likelihood-Ratio Test (aLRT) algorithm (Anisimova and Gascuel, 2006). The number of duplications, loss, and eventually HGT events were computed in Notung 2.8 (Stolzer et al., 2012) by inferring the protein tree with the species tree.

2.8 Comparative genome analysis of *Rhodotorula graminis* WP1.

The Cluster analysis tool from Mycocosm database was used to data mine WP1 genome for genetic feature related to plant beneficial functions, adaptation to plant host, response to oxidative and environmental stresses and detoxification of heavy metals. The cluster used in this analysis was the *Pucciniomycotina* UM 2015.638.

3. Results and discussion

3.1 Genome features and phylogenetic analysis of *Rhodotorula graminis* WP1

WP1 genome features are reported in Table 2. Based on MCL clustering, 3552 predicted proteins cluster within 929 multigene families. The largest families include protein kinases, transporters, and transcription factors. With a 21.01 Mbp genome, represented by 26 scaffolds, WP1 show one of the highest GC content (67%) among the all publicly available fungal genomes. Interestingly, its intron content, which is also GC-rich, show a significant asymmetry in the C (42.7%) and G (23%) levels. Because the average asymmetry of all fungal genomes in Mycocosm database is 1.8 ± 3.1%, and the second highest asymmetry of 7.3% (29.5% C and 22.2% G) was observed in the related *Pucciniomycotina* yeast *Sporobolomyces roseus*, a deviation of 19.7% for WP1, is the largest asymmetry ever observed. Unlike eubacteria, where the G in the leading strand is enriched over C within the replication origin, the biases in terms of GC asymmetry of intronic sequences in eukaryotes are likely due to sequences motif recognized by the splicing machinery (Touchon et al., 2004). Indeed, compared to other eukaryotes, where the conserved consensus is present from position +1 to +6 of introns, (Rogozin and Milanesi, 1997; Bhasi et al., 2007), an extension rich in C from +7 to +10 of WP1 introns was observed, while +3 and +4 positions show stronger consensus relative to what is observed in other Basidiomycetes.

Table 2. Genome feature of *Rhodotorula graminis* WP1

Genome features	Annotation results	
Number of genes	7283	
Avg. gene length, bp	2126	
Avg. protein length, aa	501	
Avg. exon length, bp	254	
Avg. intron length, bp	105	
Avg. exons per gene	6.2	
PROTEINS WITH:		
Similarity to KEGG	6384	88%
Similarity to KOG	5910	81%
Similarity to Swissprot	5989	82%
Similarity to NCBI NR	5724	79%
Pfam domain	4545	62%
Complete CDS	6415	88%
Transmembrane helices	1283	18%
Signal peptide	1626	22%

Based on amino acid concatenated sequences of the nuclear protein-coding genes of the two subunits of DNA polymerase II (RPB1 and RPB2) and the translation elongation factor 1-α (TEF1), WP1 is positioned inside the *Sporidiobolaceae* family (Figure 1), which is in line with the recent phylogenetic reconstruction of the *Pucciniomycotina* taxonomic group (Wang et al., 2015). The phylogenetic classification of WP1 within the *Sporidiobolaceae* Family is also supported by the high degree of synteny shared with two representative members of this family: *Sporobolomyces roseus* and *Rhodosporidium turoloides* (Figure 2 A).

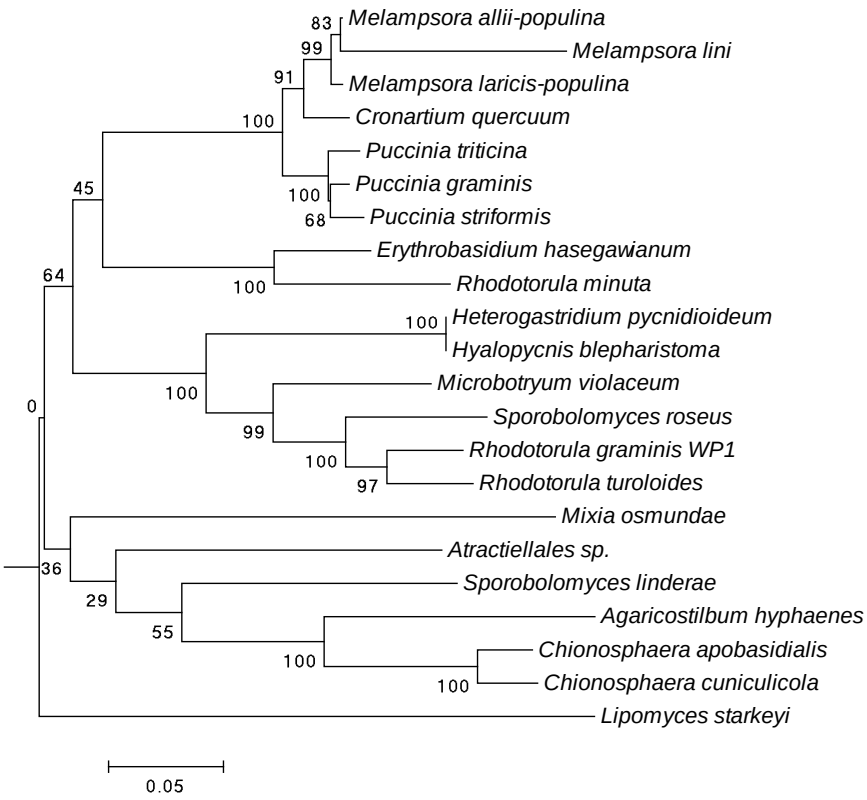
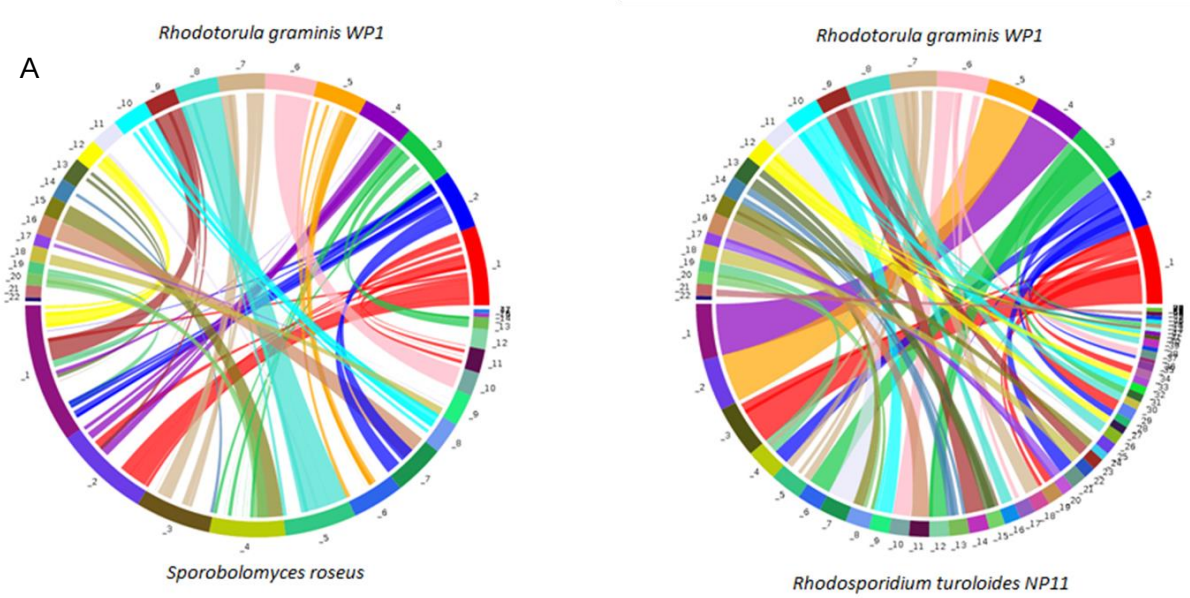


Figure 1. Molecular phylogeny of Pucciniomycetous species and related filamentous fungi.

At proteome level, the correlation of WP1 to *Rhodosporidium turoloides* NP11 is supported by 12159 shared orthologues, organized in 5920 clusters, while 222 unique genes in WP1 were organized in 64 clusters (Figure 2B). By using the Hypergeometric test with a *p-value* of 0.05 on GO terms of WP1 unique proteins, this fraction was enriched with proteins involved in the metabolism of nitrogen compounds such us xanthine metabolism (xanthine dehydrogenase; IDs: 10194, 46252 and 55432), asparagine catabolic process (L-asparaginase; IDs: 44992 and 4557) and serine metabolism (O-acetyltransferase; IDs: 1199 and 66239).

A



Species	Blocks	% Coverage	Inverted Blocks	< 100 Kb	100 – 1 Mb	1 – 10 Mb
Rhodosporidium turoloides NP11	85	88	44	39	43	3
Sporobolomyces roseus	90	57	48	50	39	1

B

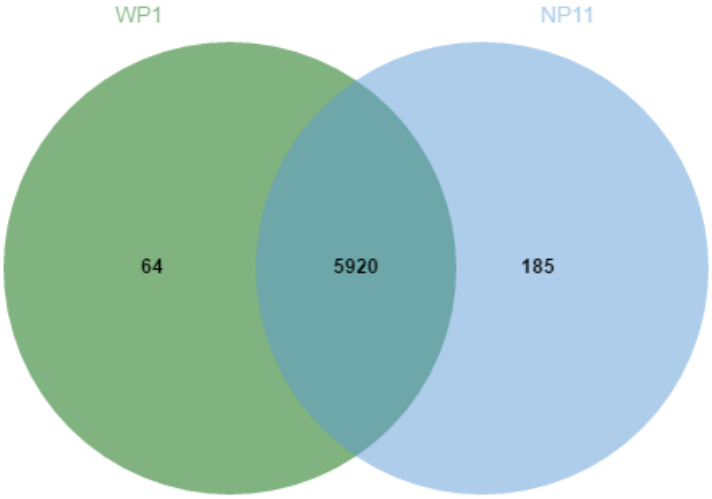


Figure 2. Comparative analysis of WP1 genome against *Rhodosporidium turoloides* NP11 and *Sporobolomyces roseus*. A) Synteny map and statistic of WP1 genome against *S. roseus* and *R. turoloides* NP11. B) Venn diagram of orthologues and unique genes between WP1 and NP11

3.1 Analysis of SSPs in *Pucciniomycotina* species and in *Rhodotorula graminis* WP1

In Figure 3 is reported a scheme of the pipeline used to determine and characterize the pool of SSPs.

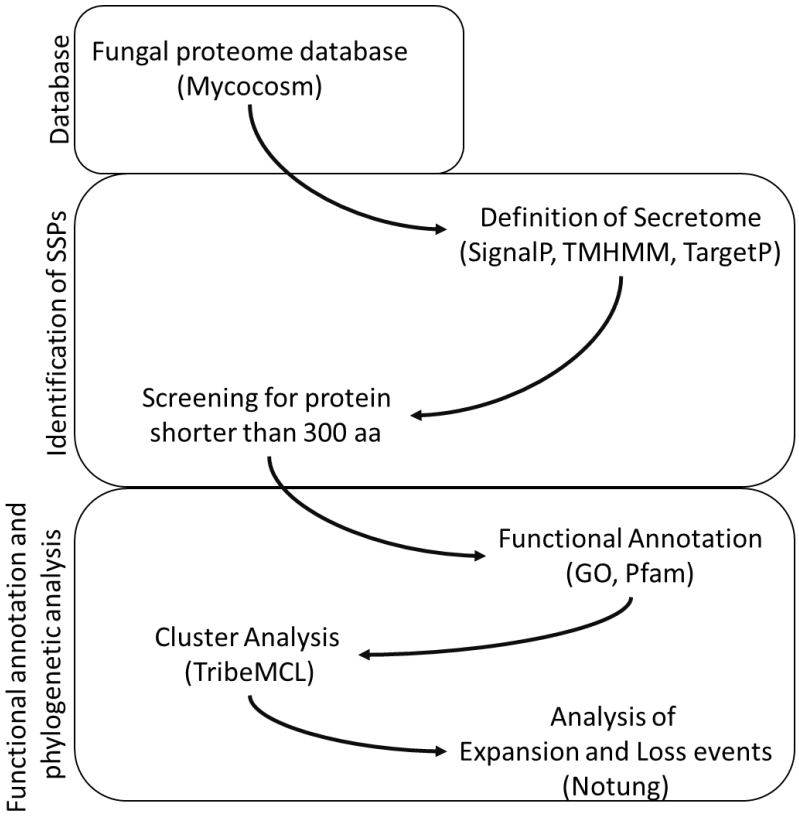


Figure 3. Pipeline used for the characterization, functional annotation and phylogenetic analysis of SSPs in *Pucciniomycotina* and *R. graminis* WP1

As previously reported (Kim et al., 2016), *Pucciniomycotina* represent the phylogenetic group with the highest number of small secreted proteins among Basidiomycota. With 91 SSPs and a proteome size of 7282, a neat difference between plant pathogens (rust lineage) and WP1, with the only exception for the plant pathogens *M. violaceum*, was observed (Figure 4). Also a correlation between the number of SSPs and the microorganism lifecyle was not observed in the *Pucciniomycotina* parasites *Chionosphaera apobasidialis* and *Hyalopycnis blepharistoma*, confirming that the size of secretome is more lineage specific rather than related to the organism lifestyle (Krijger et al., 2014). Indeed, by comparing the phylogenetic tree in Figure 1 against the cluster dendrogram represented in Figure 4, rust fungi, with an exception for *Atractiellales* sp., generate a separate cluster from the other *Pucciniomycotina* species

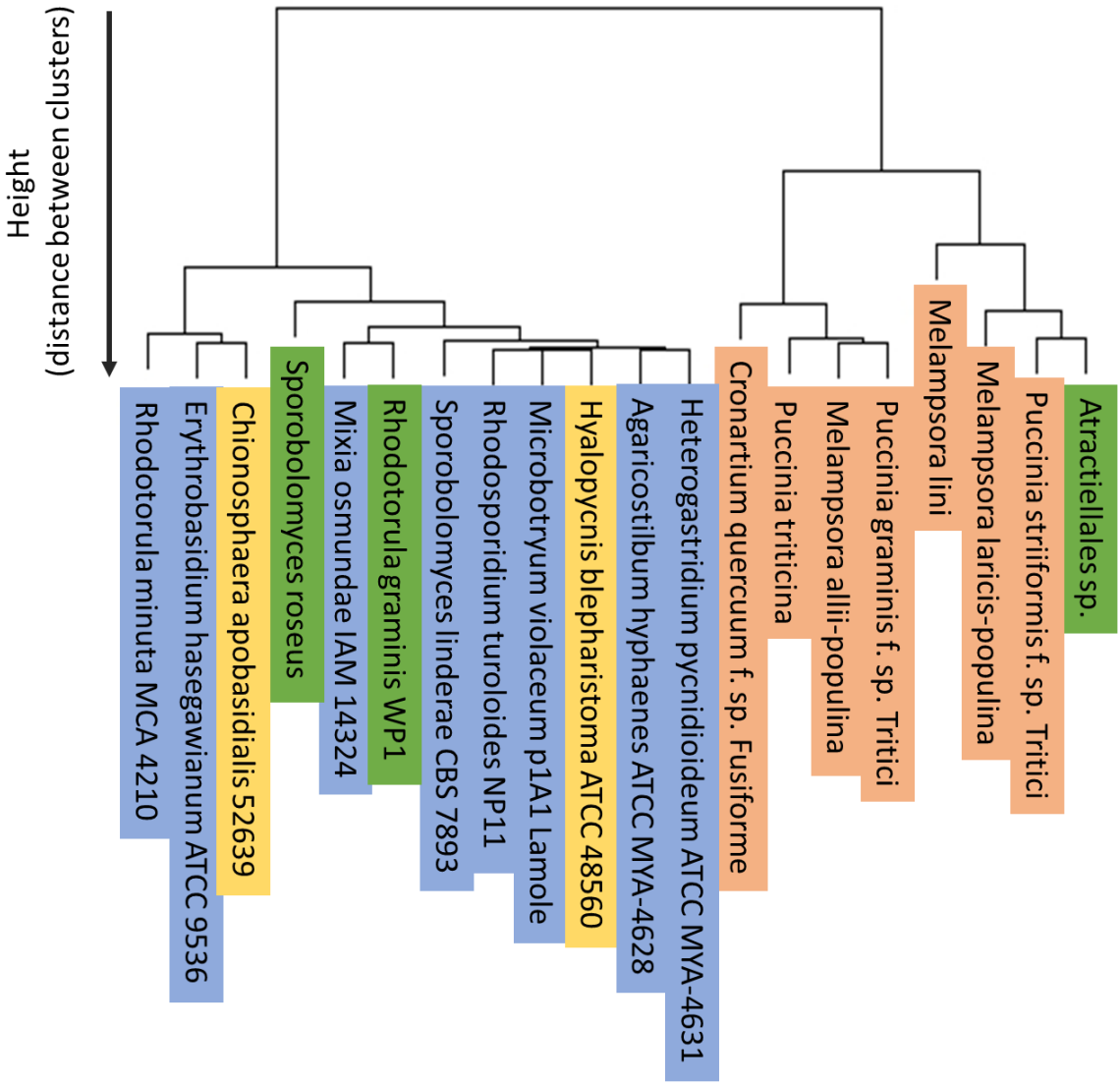


Figure 4. Hierarchical clustering of SSPs in relation to proteome size. Each colors define the life style: non-obligated biotrophs (blue), plant pathogens (pink), symbionts/endophytes (green) and parasites (yellow). Hierarchical clustering was performed in R statistical software.

In WP1, distinct SSPs annotated with protein-protein interaction domain (IDs: 46692, 54421, 65786, 49363, 10697, 5062) carbohydrates recognition domain (ID: 40477) and proteins involved in ROS scavenging (ID: 55820) were detected (Table 3). Whit the exception for the Rare Lipoprotein A (ID: 49363) these genes were also identified within syntenic regions of the WP1 non-endophytic counterpart NP11.

Table 3. WP1 SSPs with a protein-protein interaction domain and ROS scavenging

Family proteins and Protein domains Ids	WP1 Protein IDs (Mycocosm)	Functional relevance in plant-microbe Interaction and adaptation to plant niches
IPR008427 [DOMAIN] - Extracellular membrane protein, CFEM domain	46692, 54421	Host interaction (Zhang et al., 2015)
IPR023566 [FAMILY] - Peptidyl-prolyl cis-trans isomerase, FKBP-type	65786	Host Interaction (Pemberton, 2006)
IPR001179 [DOMAIN] - FKBP-type peptidyl-prolyl cis-trans isomerase domain		
IPR029000 [DOMAIN] - Cyclophilin-like domain	65265	
IPR020892 [CONSERVED_SITE] - Cyclophilin-type peptidyl-prolyl cis-trans isomerase, conserved site		
IPR002130 [DOMAIN] - Cyclophilin-type peptidyl-prolyl cis-trans isomerase domain		
IPR024936 [FAMILY] - Cyclophilin-type peptidyl-prolyl cis-trans isomerase		
IPR010987 [DOMAIN] - Glutathione S-transferase, C-terminal-like	55820	Response to oxidative stress
IPR012336 [DOMAIN] - Thioredoxin-like fold		
IPR009009 [DOMAIN] - RlpA-like protein, double-psi beta-barrel domain	49363, 10697, 5062	Host interaction (Mak et al., 2013)
IPR013320 [DOMAIN] - Concanavalin A-like lectin/glucanase domain IPR000757 [DOMAIN] - Glycoside hydrolase family 16	40477	Host Interactions (Guyon et al., 2014)

Within the GO category "Biological Process" a neat distinction between NP11 and WP1 SSPs was observed in the enrichment of GO terms with a biological relevance in the uptake of plant metabolites. Indeed the functional profile for proteins involved in the binding of cyclic and heterocyclic organic compounds and for protein with an hydrolase activity was more similar to the obligated biotroph *P. graminis* rather than to *R. turoloides* NP11 (Figure 5). The enrichment of such domains in WP1 secretome may confer a huge benefit in terms of competition for nutrients uptake against other microorganisms.

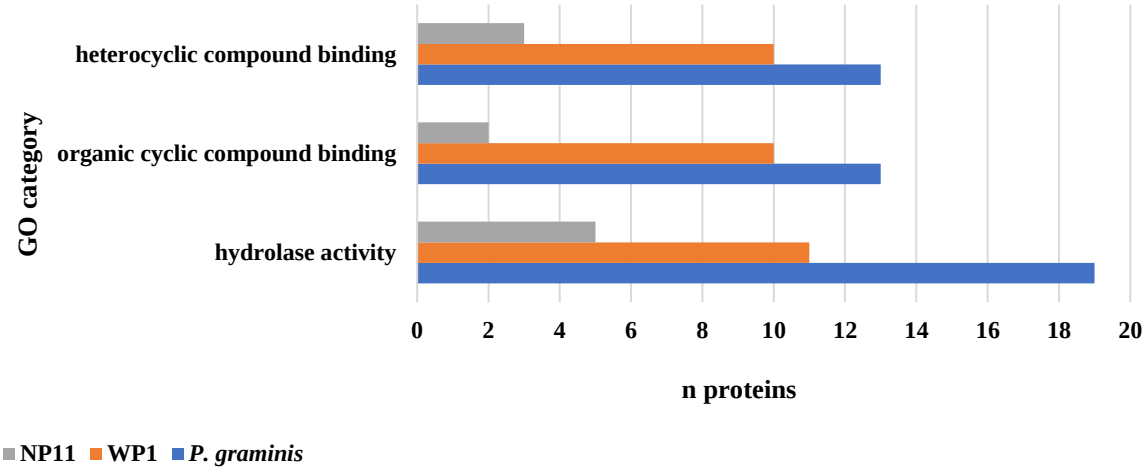


Figure 5. Analysis of GO terms within the category Molecular functions (GO: 0003674). Heterocyclic compound binding: enzymes that interact selectively and non-covalently with heterocyclic compound; Organic cyclic compound binding: enzymes that interact selectively and non-covalently with an organic cyclic compound, any molecular entity that contains carbon arranged in a cyclic molecular structure; Hydrolase activity: enzymes involved in the hydrolysis of various chemical bonds, e.g. C-O, C-N, C-C, phosphoric anhydride bonds. NP11, *R. turoloides* NP11; WP1, *R. graminis* WP1.

3.2 Cluster analysis of SSPs

Based on sequence similarities, from cluster analysis conducted on *Pucciniomycotina* SSPs 410 protein clusters were retrieved. An overview of the most common protein domains detected within the entire pool of SSPs is reported (Figure 6). Sixty-three out of 91 putative SSPs in WP1 were clustered. However, many of the SSPs in WP1 were annotated as hypothetical proteins and only few of them were successfully annotated with Pfam, InterPro and, by consequence, with GO terms. Here, will be discussed all those SPPs in WP1 with biological relevance in adaptation to plant niche and plant-microbe interactions.

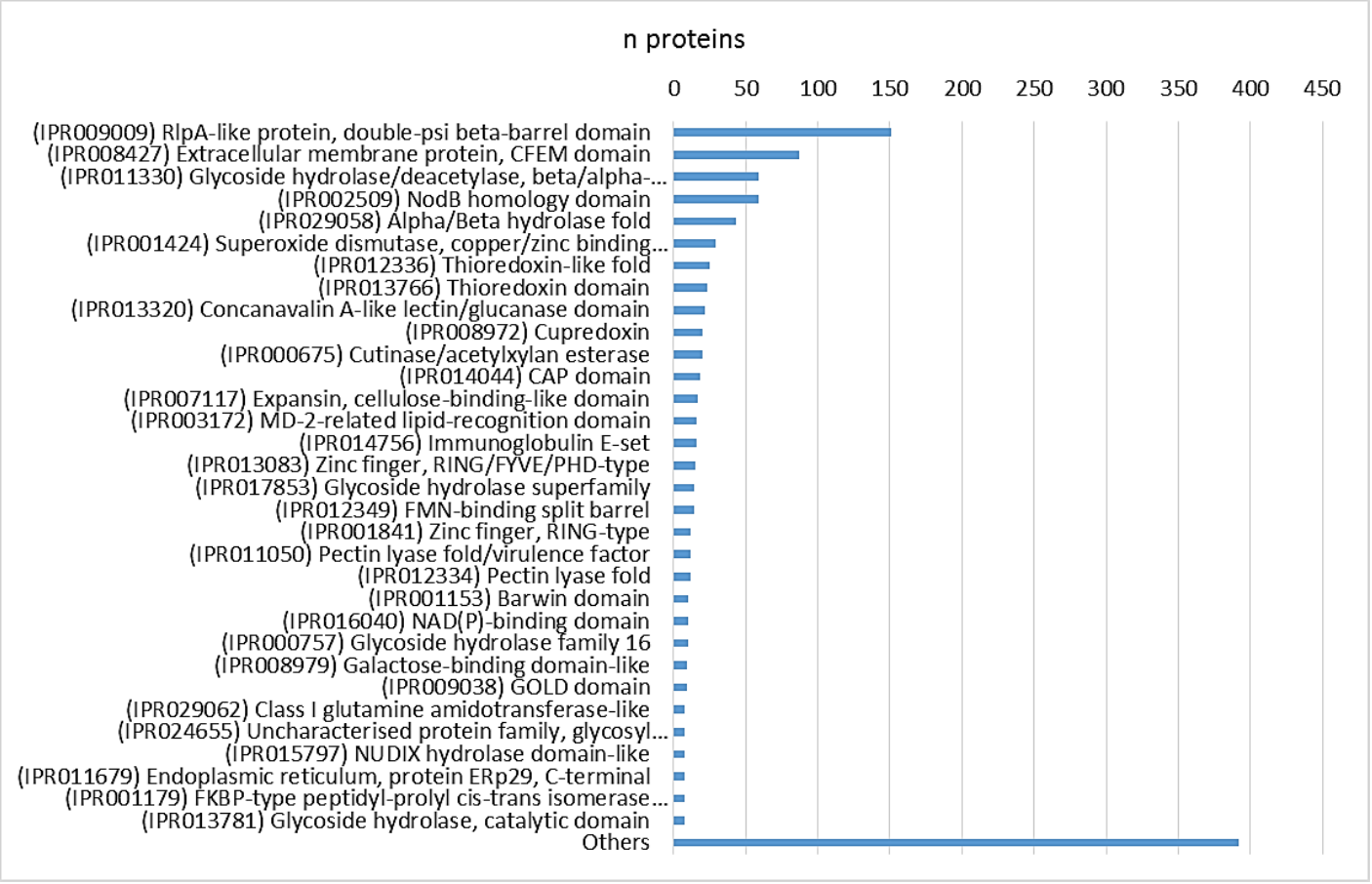


Figure 6. Functional annotation of *Pucciniomycotina* SSPs. InterPro accession numbers (IPR) were retrieved from the EMBL-EBI databases

3.2.1 The Rare lipoprotein A (RlpA) domain (IPR009009).

The first and most abundant protein cluster detected from the clustering analysis of *Pucciniomycotina* SSPs was represented by hypothetical proteins from the rust lineage. Proteins annotated with an RlpA domain were detected in the second biggest cluster of *Pucciniomycotina* SSPs (Figure 7A). Furthermore, additional domains such as the cellulose binding domain, LysM domain and the Carbohydrate binding module-5 were detected for some member of this cluster (figure 7B).

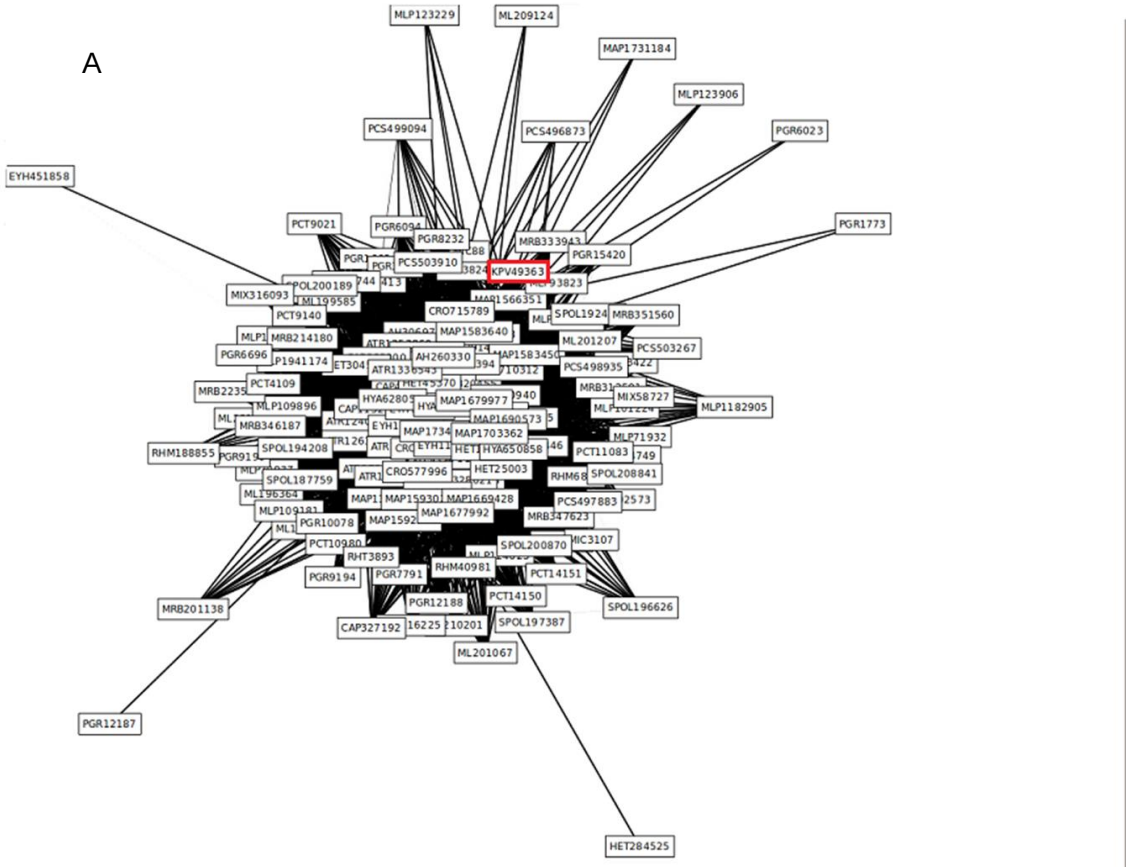


Figure 7. A) RlpA cluster of SSPs; in red the RlpA SSPs of WP1. B) Annotation results of functional domains for RlpA cluster. Names used in this chart follow the Pfam nomenclature system. Protein ID in Figure 7A are written in the general form Species-Mycocosm protein IDs. PGR (*Puccinia graminis*), KPV (*Rhodotorula graminis* WP1), RHT (*Rhodosporidium turoloides* NP11), MLP (*Melampsora laricis-populina*), ML (*Melampsora lini*), MAP (*Melampsora alli populina*), PCT (*Puccinia triticina*), AH (*Agaricostilbum hyphaenes*), CRO (*Cronartium quercuum*), ATR (*Atractiellales sp*.), CAP (*Chionosphaera apobasidialis*), MIX (*Mixia osmundae*), HYA (*Hyalopycnis blepharistoma*), HET (*Heterogastridium pycnidioideum*).

The RlpA domain (IPR009009) is a non-enzymatic domain commonly detected in the secretome of rust fungi (Saunders et al., 2012). We attempt to perform a phylogenetic reconstruction of the RlpA cluster in order to calculate the number of loss and duplication events occurring in the *Pucciniomycotina* subdivision but due to the lack of variable regions outside conserved blocks, and consequent low bootstrap values, we failed in the attempt to reconstruct a congruent phylogenetic tree for this class of proteins. However, as general trend the presence of duplication events was more frequent within the phylogenetic lineage of rust fungi, suggesting that the RlpA domain was positively selected and expanded in relation to the microorganism lifestyle (data not shown). Furthermore, the RlpA SSPs from the rust lineage were the only with a cellulose and carbohydrates biding domain. Proteins with such domains usually recognize components of plant cell wall by working in association with hydrolytic enzymes such as xylanases and cellulases (Tovar-Herrera et al., 2015). As reported in Table 3, WP1 carries three genes encoding for SSPs with an RlpA domain. The protein 49363, which was placed within the cluster discussed in this section, was detected in a chromosome region unique in WP1 and not in synteny with NP11 and others phylogenetically related species. The gene encoding for the protein 49363 was detected in a gene cluster encoding for proteins with relevant functions in plant-host interaction such as iron (ID: 51878) and ammonium (ID. 51873) uptake, polysaccharides hydrolysis (ID: 64610) and a member of the membrane peptidase S54 family proteins (IDs: 12722).

3.2.2 CFEM protein domain (IPR008427)

This major cluster is represented by members of SSPs annotated with a CFEM domain (Figure 8). Such protein domain is fungal-specific, characterized by eight cysteines and commonly detected in protein involved in fungal pathogenesis and plant-microbe interactions. Unlike the RlpA cluster, no additional domains were detected. In a comparative study between *Ascomycota* and *Basidiomycota* (Zhang et al., 2015), *Pucciniomycotina* SSPs with a CFEM domain from rust fungi were the most representative members of this cluster. Unfortunately, despite a fully conserved 8-cys domain was detected among almost all members of this cluster, it was not possible to construct a phylogenetic

tree due to very short internal branches supported by low bootstrap values. WP1 protein with a CFEM domain were detected in chromosome region highly conserved in phylogenetic related species suggesting that such genetic traits were not maintained or selected as consequence of the adaptation to plant niches but for a more general involvement in other biological functions Indeed, in *C. albicans* CFEM proteins in are involved in biofilm formation and iron acquisition from host heme, while CFEM proteins in *A. fumigatus* are exclusively involved in cell wall stabilization (Vaknin et al., 2014) and not acting as virulence factors, even for pathogenic fungi. Interestingly, even within the same phylogenetic line, the functional aspects of CFEM proteins are extremely diversified. Indeed, the role of the CFEM family in biofilm formation is not conserved in the human pathogen *Candida parapsilosis* (Ding et al., 2011), suggesting that the CFEM proteins conserved across WP1 and phylogenetically related species might not share the same functional aspects.

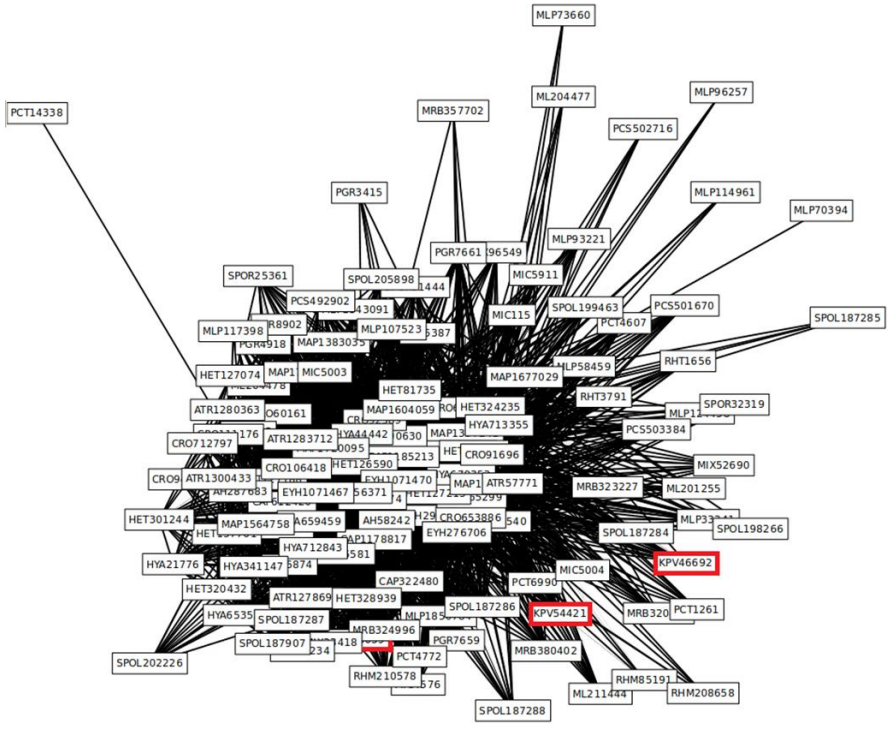


Figure 8. SSPs CFEM cluster. WP1 SSPs with a CFEM domain are marked in red. Protein ID are written in the general form Species-Mycocosm protein IDs. PGR (*Puccinia graminis*), KPV (*Rhodotorula graminis* WP1), RHT *(Rhodosporidium turoloides* NP11), MLP (*Melampsora laricis-populina*), ML (*Melampsora lini*), MAP (*Melampsora alli populina*), PCT *(Puccinia triticina*), AH (*Agaricostilbum hyphaenes*), CRO

(*Cronartium quercuum*), ATR (*Atractiellales sp.*), CAP (*Chionosphaera apobasidialis*), MIX (*Mixia osmundae*), HYA (*Hyalopycnis blepharistoma*), HET (*Heterogastridium pycnidioideum*).

3.2.3 MD-2-related lipid-recognition domain (IPR003172)

Lipids have been reported to act as messengers in plant response to hormones, with important implication in plant development, and response to pathogens and environmental stresses (Wang and Chapman, 2013). Secreted proteins with a lipid recognition domain may be able to interfere/respond to hormones signaling. For the MD-2 like proteins cluster, an expansion was observed within the rust lineage of *Melampsora* species, with four duplications and only one loss event (Figure 9). No expansion and loss events were detected within the phylogenetic lineage of WP1.

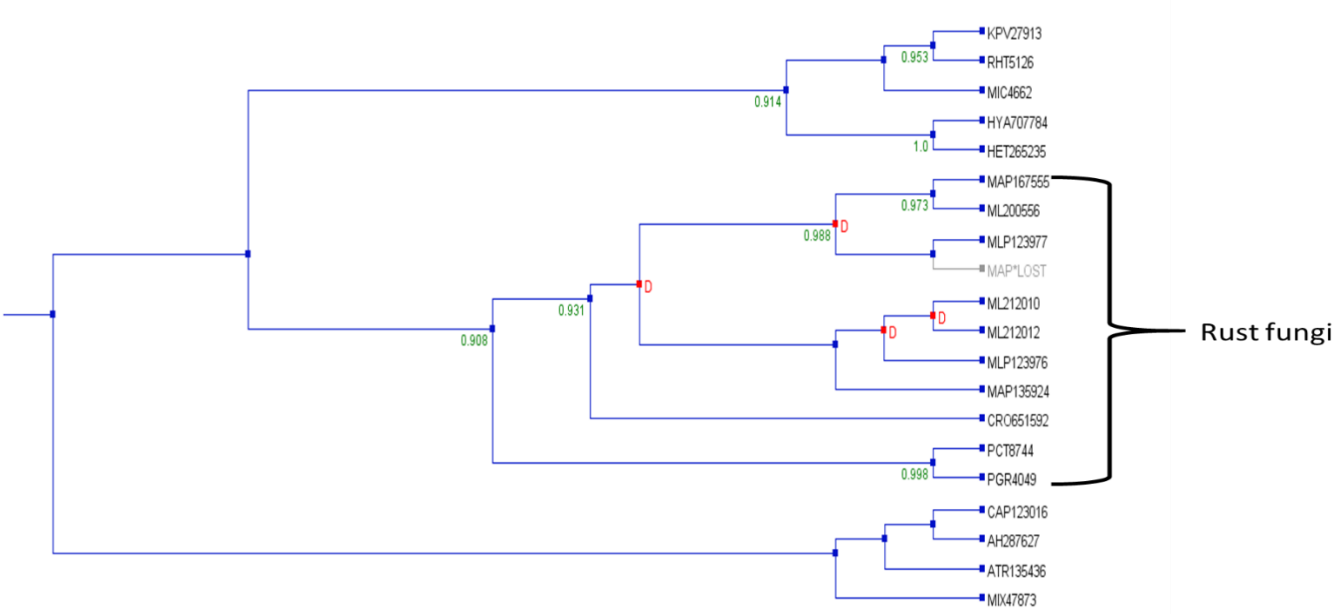


Figure 9. Expansion and loss events in *Pucciniomycotina* MD-2 like proteins. Protein ID are written in the general form Species-Mycocosm protein IDs. PGR (*Puccinia graminis*), KPV (*Rhodotorula graminis* WP1), RHT (*Rhodosporidium turoloides* NP11), MLP (*Melampsora laricis-populina*), ML (*Melampsora lini*), MAP (*Melampsora alli populina*), PCT (*Puccinia triticina*), AH (*Agaricostilbum hyphaenes*), CRO (*Cronartium quercuum*), ATR (*Atractiellales* sp.), CAP (*Chionosphaera apobasidialis*), MIX (*Mixia osmundae*), HYA (*Hyalopycnis blepharistoma*), HET (*Heterogastridium pycnidioideum*).

Furthermore, the MD-2 like protein coding gene in WP1 is located within a chromosomal region highly conserved across phylogenetically related species to WP1 (*R. turoloides NP11, Rhodotorula sp. JG-1b and Sporobolomyces roseus*) (data not shown).

3.3 HGT genes in WP1

In the acquisition of new function and adaptation to environmental niches, later gene transfer may play an important role. Twelve genes in WP1 were putatively acquired through HGT (Table 4) where some of the donor species belong to bacterial taxa commonly detected in rhizosphere or isolated as plant associated bacteria.

Table 4. Putative HGT genes in WP1. Protein top hit reported in this table are based on the highest e-value score

Predicted donor	Number of hits	Gene product
Bacillales	2	nitroreductase; acetyltransferase
Burkhoderiales	2	Glutathione S-transferase; short chain dehydrogenase
Enterobacteriales	2	2-hydroxyacid dehydrogenase; Orotate phospribosyl-transferase
Oscillatoriales	1	hypotethical protein
Pseudomonadales	2	maleylacetate reductase; glutathione peroxidase
Rhizobiales	2	peroxiredoxin; 3-oxoacyl ACP-reductase
Rhodocaterales	1	4a-hydroxytetrahydrobiopterin dehydratase

However, because these genes were also detected in other *Microbotryomycetes* (*Rhodosporidium* and *Sporobolomyces* species), and in limited number of genomes within the phylogenetic group of *Pucciniomycotina*, it is possible that a mixed pattern in the distribution of genes within closely related species is the result of loss events rather than HGT. Only the gene encoding for a peroxiredoxin protein show a strong evidence of lateral transfer. Indeed, in WP1 the genomic neighborhood of this gene show synteny with regions found in *R. turoloides* NP11 and *R. glutinis* where the gene encoding for the PRX protein is missing (Supplementary material, Fig. 1A).

The gene product is a member of the Ohr-like proteins, a sub cluster of Ohr/OsmC proteins, which comprise members detected in Eukaryotes (Meireles et al., 2017) and Prokaryotes as well, mainly *Rhizobiales* (in this study, Supplementary material, Fig. 1B and Supplementary material, Table 1). Ohr-like proteins are characterized by the presence of two catalytic cysteines separated by approximately 60 amino acid residues in the primary sequence (Supplementary material, Fig. 1B). This feature is conserved in all members belonging to OsmC/Ohr family, while the only features that differentiate Ohr-like proteins form the OsmC/Ohr, is the lack of two additional amino acid residues, Glu and Arg, which are required, but not essential, for the peroxidase activity (Meireles et al., 2017). OsmC/Ohr and Ohr-like proteins has been linked to the capability of an organism to quench hydroperoxide, which are also a class of lethal stressor and signal molecules commonly produced by the basal host metabolism or as results of the response to environmental stresses (biotic and abiotic) (Tripathy and Oelmuller, 2012; Meireles et al., 2017). In WP1, the presence of this protein might

conferee an enhanced tolerance to oxidative stresses and the capability to interfere with the ROS signaling.

3.4 Plant colonization

Cutinases belong to the alpha/beta hydrolase superfamily, and catalyze the hydrolysis of carboxylic ester bonds. Such proteins are involved in leaves colonization and carbon acquisition, playing a central role in fungal pathogenesis. Two cutinase coding genes (CCGs) with a conserved GYSQ motif (Figure 10A) were detected in WP1. Such genes are organized within the same chromosomal region, suggesting that one copy was originated from a duplication event, or acquired through horizontal gene transfer. The CCGs from WP1 encodes for secretion proteins, and are also conserved in a syntenic region in NP11. Interestingly, in NP11 the only copy of the CCG encode for protein lacking the signal peptide (Figure 10B). Furthermore, CCGs from Wp1 and NP11 are the only cutinase genes detected within the phylogenetic lineage of WP1 (*Rhodotorula*, *Rhodosporidium* and *Sporobolomyces* species). In *Pucciniomycotina*, the presence of gene clusters carrying multiple copies of CCGs were detected only in obligated biotroph (Figure 10C). From an evolutionary point of view, gene duplication, as consequence of lateral gene transfer or a "self" duplication events, is believed to facilitate the adaptation of an organism through a functional and regulatory diversification. For example, the CCGs from the plant pathogens *Phytophtora* ssp, were acquired as result of lateral gene transfer from a prokaryotic lineage. In our case, CCGs from WP1 were not detected as HGT genes, meaning that one copy was acquired via duplication.

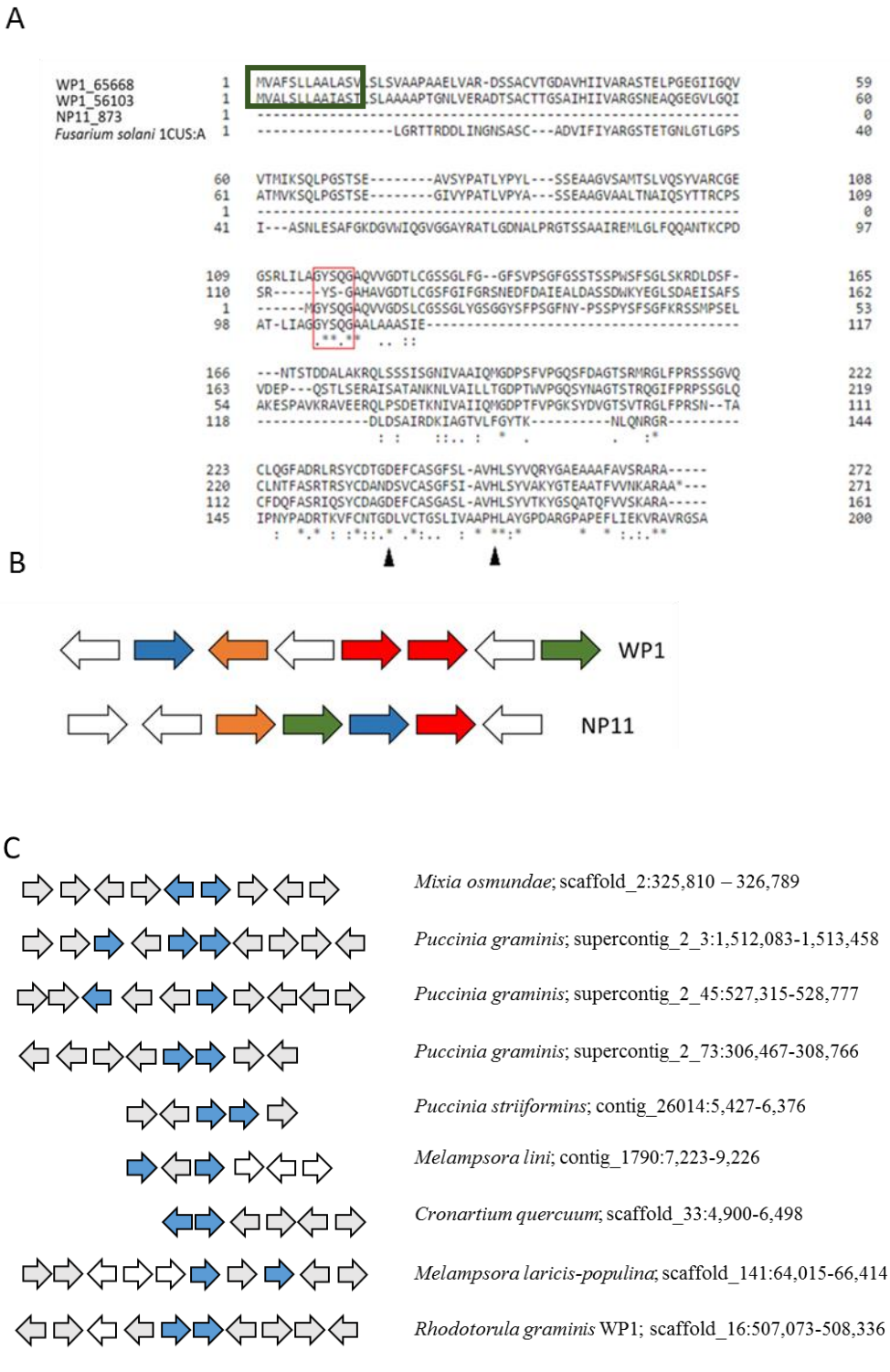


Figure 10. A) Alignment of NP11 (R. turoloides) and WP1 cutinase protein sequences against *Fusarium solani* cutinase. Red box marks the GYSQ motif conserved in all cutinase proteins. Black triangles mark the aspartate and histidine in the active site (Pietrokovski, 1998). Signal peptide in green. B) Alignment of the genome regions containing the CCGs in WP1 and NP11 (red arrows). White arrows, hypothetical proteins. C) Chromosomal organization of CCGs (pale blue arrows) in *Pucciniomycotina* obligated biotroph.

Another interesting feature of WP1 is the presence of a polysaccharidic capsule that surrounds the cellular body (Figure 11). A well-studied encapsulated yeast is represented by *Cryptococcus neoformans*, an opportunistic pathogen causative of meningoencephalitis in immunocompromised patients (Mitchell and Perfect, 1995).

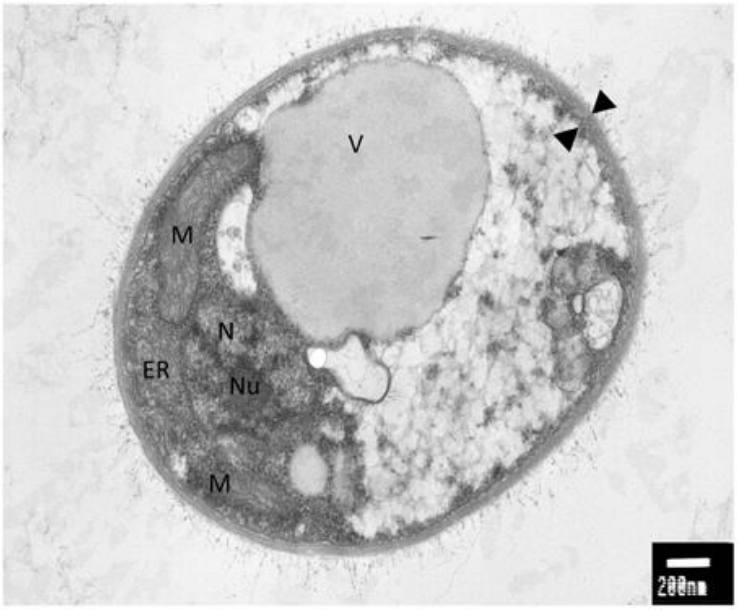


Figure 11. Electron microscopy analysis of Rhodotorula graminis strain WP1; M, mitochondria; ER, endoplasmic reticulum; N, nucleus; Nu, nucleulus; V, vacuole; black triangles, capsule. Photo credit: Prof. Jimmie Lara, Department of Microbiology, University of Washington.

C. neoformans is a ubiquitous yeast where the capability to form biofilms increase its survival capabilities under different environmental conditions. Within biofilms, microbial cells show an increased resistance and tolerance against a wide range of biotic and abiotic stress but also, biofilms represent a strategy colonize new ecological niches through cells dispersion (Ramage et al., 2009). The capability of *C. neoformans* to adhere and grow as biofilm is dependent on the presence of the capsule. Deletion mutants in *cap*59 and *cap*10, two genes involved in the capsule synthesis, are unable to form biofilms, implicating that this structure exerts an important role in the adhesion and subsequent formation of cells aggregates (Garcia-Rivera et al., 2004; Martinez and Casadevall, 2005). Since a putative CAP59 (Protein ID: 4796) and CAP10 (Protein ID: 7100) were detected in WP1, the genome sequence of a non-pathogenic encapsulated yeast would be interesting for a comparative analysis between the capsule synthesis in WP1 and *C. neoformans*.

3.5 Nutrient uptake from plant

One of the major differences between biotrophs and nonobligate biotrophs lie in a different set of transport system specialized in the uptake of specific class of molecules of plant metabolites to sustain a specific metabolism important for host-associated lifestyle. For example, as result of the activity of secreted aspartic proteases, the human pathogens *Candida albicans* can utilize oligopeptides as nitrogen source to sustain its growth (Mayer et al., 2013). Another example is represented by *Aspergillus fumigatus* where the oligopeptide (OPT)-transport systems and the extracellular proteolytic activities are co-regulated but not involved in virulence (Harttman et al., 2011).

In *Pucciniomycotina*, the OPT-transporter family went through 59 duplications and 130 loss events (Figure 12), where the highest number of duplication and loss events were observed within the rust lineage. Compared to *R. turoloides* and *S. roseus,* in WP1 the OPT-transporter family have undergone through an expansion, with two duplication events and zero loss. Similarly, in the poplar endophyte *Atractiellales sp.,* the overall number of duplication and loss events were 4 and 3 respectively. Such expansions may reflect an adaptive response to the acquisition of organic nitrogen within plant niches.

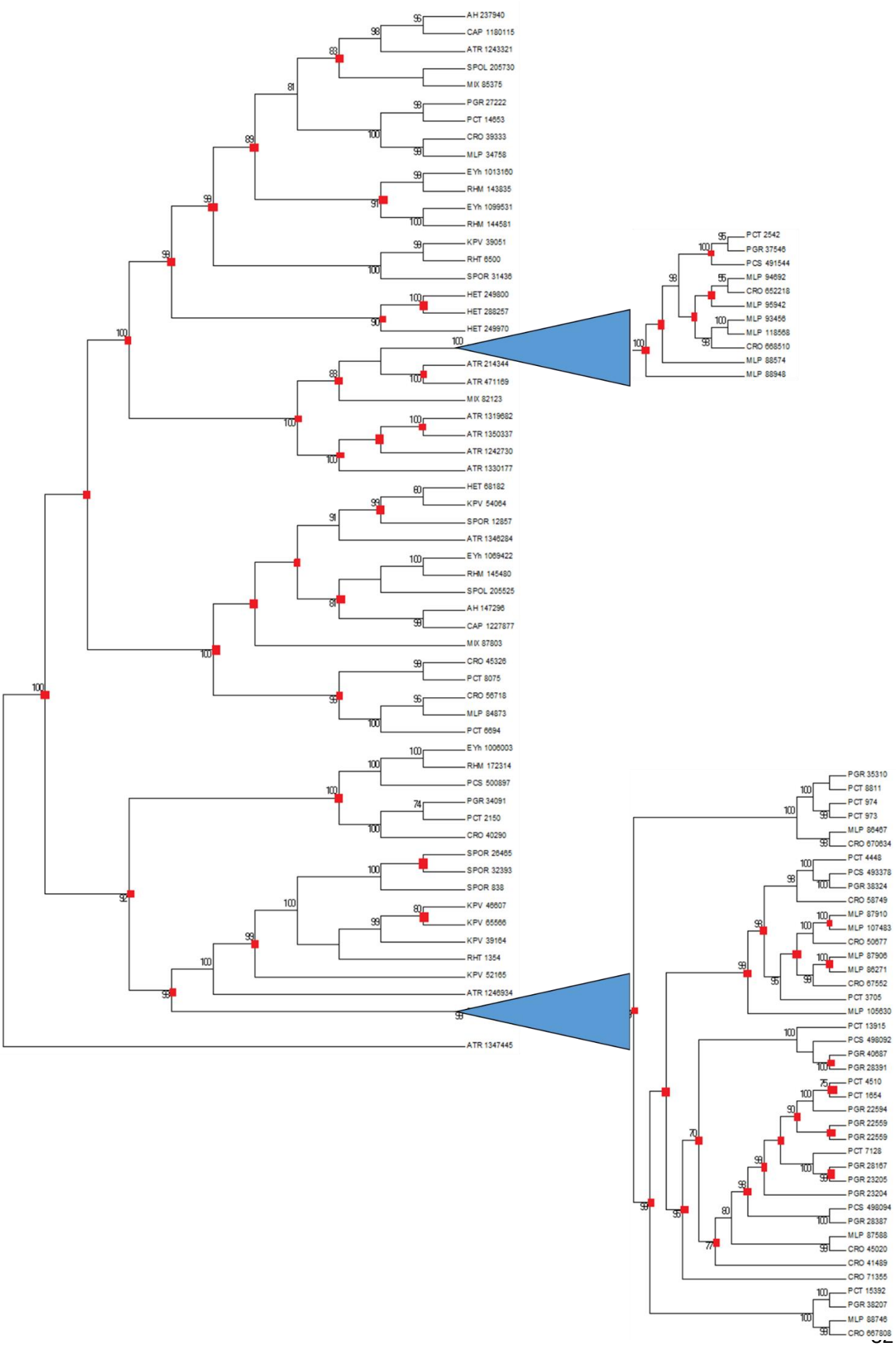

AH 237940
CAP 1180115
ATR 1243321
SPOL 205730
MIX 85375
PGR 27222
PCT 14653
CRO 39333
MLP 34758
EYh 1013160
RHM 143835
EYh 1099531
RHM 144581
KPV 39051
RHT 6500
SPOR 31436
HET 249800
HET 288257
HET 249970
ATR 214344
ATR 471169
MIX 82123
ATR 1319682
ATR 1350337
ATR 1242730
ATR 1330177
HET 68182
KPV 54064
SPOR 12857
ATR 1346284
EYh 1069422
RHM 145480
SPOL 205525
AH 147296
CAP 1227877
MIX 87803
CRO 45326
PCT 8075
CRO 56718
MLP 84873
PCT 6694
EYh 1006003
RHM 172314
PCS 500897
PGR 34091
PCT 2150
CRO 40290
SPOR 26465
SPOR 32393
SPOR 838
KPV 46607
KPV 65566
KPV 39164
RHT 1354
KPV 52165
ATR 1246934
ATR 1347445
PCT 2542
PGR 37546
PCS 491544
MLP 94692
CRO 652218
MLP 95942
MLP 93456
MLP 118568
CRO 668510
MLP 88574
MLP 88948
PGR 35310
PCT 8811
PCT 974
PCT 973
MLP 86467
CRO 670634
PCT 4448
PCS 493378
PGR 38324
CRO 58749
MLP 87910
MLP 107483
CRO 50677
MLP 87906
MLP 86271
CRO 67552
PCT 3705
MLP 105630
PCT 13915
PCS 498092
PGR 40687
PGR 28391
PCT 4510
PCT 1654
PGR 22594
PGR 22559
PGR 22559
PCT 7128
PGR 28167
PGR 23205
PGR 23204
PCS 498094
PGR 28387
MLP 87588
CRO 45020
CRO 41489
CRO 71355
PCT 15392
PGR 38207
MLP 88746
CRO 667808

Figure 12. aRLT tree of OPT-transporter in *Pucciniomycotina*. Nodes in red represent a duplication event. Loss events are not shown for a better visualization of tree topology. Nodes with a score lower than 70 are not shown. Protein ID are written in the general form Species-Mycocosm protein IDs. PGR (Puccinia graminis), KPV (*Rhodotorula graminis* WP1), RHT (*Rhodosporidium turoloides* NP11), MLP (*Melampsora laricis-populina*), ML (*Melampsora lini*), MAP (*Melampsora alli populina*), PCT (*Puccinia triticina*), AH (*Agaricostilbum hyphaenes*), CRO (*Cronartium quercuum*), ATR (*Atractiellales* sp.), CAP (*Chionosphaera apobasidialis*), MIX (*Mixia osmundae*), HYA (*Hyalopycnis blepharistoma*), HET (*Heterogastridium pycnidioideum*). aRLT tree was constructed using OPT-transporter amino acid sequences with a complete or partial conservation of the NPG and KIPPR motifs (Koh et a., 2002)

On the other hand, the number of sugar transporter detected in WP1 and related species in greatly higher when compared to phytopathogens (Figure 13). Because endophytic microorganism might colonize a wide range of environmental niches, such disparity may suggest that phytopathogens possess only classes of sugar transporters specialized in the uptake of specific plant metabolites.

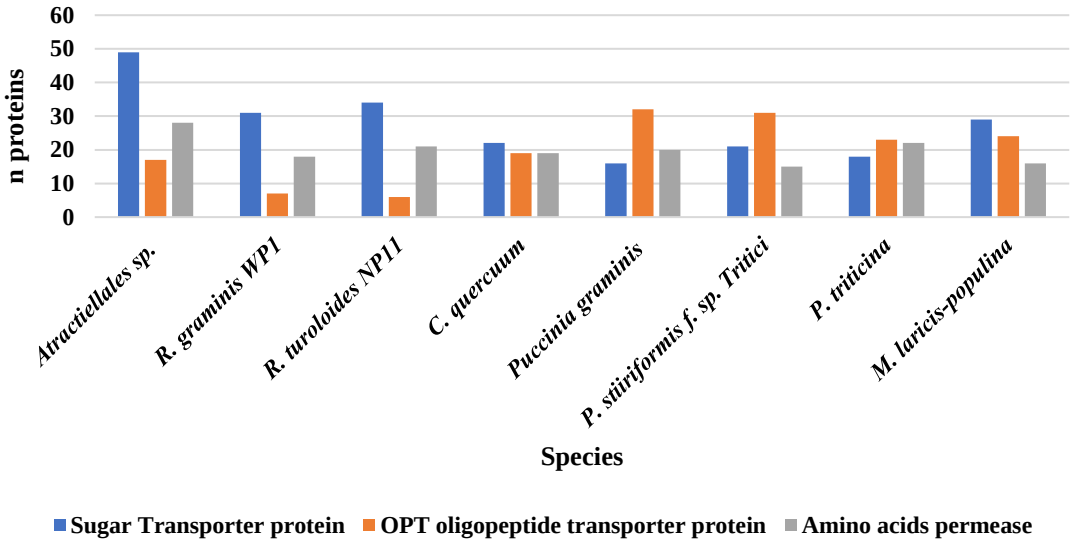


Figure 13. Enrichment of sugars and nitrogen compounds transporters in *Pucciniomycotina* species

3.5 Glycosyl Hydrolase coding genes

Plant cell wall play a crucial role in defense mechanism against pathogens. Many phytopathogens and mycorrhizae possess wide array of hydrolytic enzymes with a catalytic activity against protein and polysaccharide components of the plant cell wall. The number and typology of glycolytic enzymes possessed by WP1 is comparable to those observed in non-obligated biotrophs, while glycosyl hydrolases were more represented in phytopathogens (Figure 14). WP1 and non-obligated biotroph lack in glycolytic enzymes belonging to family 12, 7, and 10, which instead were exclusive

in *Puccinia* species, *M. larici populina*, *C. quercuum*. Similarly, genes encoding for the family 5 of glycosyl hydrolases were mainly represented in the phytopathogens *P. graminis* and *M. laricis-populina*. Families 12, 7, and 10 are mainly represented by xylanase and endoglucanase, which are known to exert hydrolytic activity against the plant cell wall (Kubicek et al., 2014). In WP1, two secreted protein with a glycoside hydrolase 5 domain were detected. However, none of these proteins presents a cellulose-binding module 1, 3 and 4, meaning that are not specialized in the recognition of plant cell wall. The lack of protein families capable to degrade the polysaccharide components of cell walls clearly suggest how massive loss/reduction events at the expense of these gene families were beneficial to WP1 adaptation to symbiotic lifestyle. Finally, a marked difference in the number of chitinases (glycosyl hydrolase family 18) genes was observed between obligated biotrophs and non-obligated biotrophs.

Chitin is one of the most abundant polysaccharide in fungal cell wall and chitinases play a central role in the cell wall remodeling and thus in hyphal developing (Hartl et al., 2011). In phytopathogens, chitinases are differentially regulated during the growth and infection mechanisms, playing an important role in the developing of biotrophic and necrotrophic hyphae. The latter one, have a more invasive growth in host cells. Thus, the expansion of chitinase genes in phytopathogens may somehow reflect their important role in the remodeling of fungal cell wall of biotrophs lifestyle (Oliveira-Garcia and Deising et al., 2013).

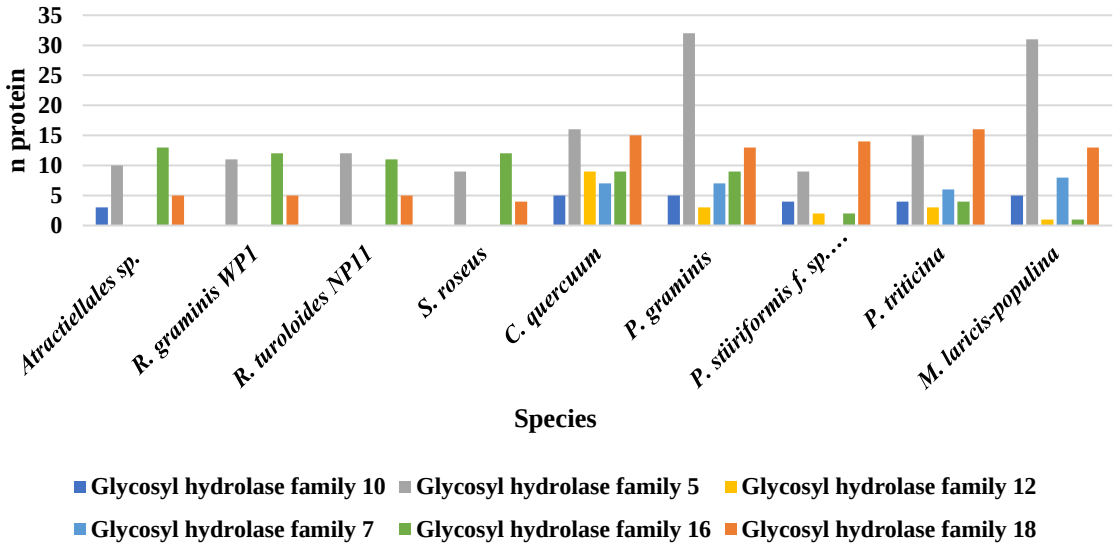


Figure 14. Comparative analysis of plant cell wall degrading enzymes in WP1 and *Pucciniomycotina* species. Glycosyl hydrolase family 5: Chitosanase (EC 3.2.1.132); b-mannosidase (EC 3.2.1.25); endo-b-1,4-glucanase / cellulase (EC 3.2.1.4); glucan b-1,3-glucosidase (EC 3.2.1.58); licheninase (EC 3.2.1.73); glucan endo-1,6-b-glucosidase (EC 3.2.1.75). Glycosyl hydrolase family 7: Endo-b-1,4-glucanase (EC 3.2.1.4); reducing end-acting cellobiohydrolase (EC 3.2.1.176); chitosanase (EC 3.2.1.132); endo-b-1,3-1,4-glucanase (EC 3.2.1.73). Glycosyl hydrolase family 10: Endo-1,4-b-xylanase (EC 3.2.1.8); endo-1,3-b-xylanase (EC 3.2.1.32). Glycosyl hydrolase family 12: Endoglucanase (EC 3.2.1.4); xyloglucan hydrolase (EC 3.2.1.151); b-1,3-1,4-

glucanase (EC 3.2.1.73); xyloglucan endotransglycosylase (EC 2.4.1.207). Glycosyl hydrolase family 16: Xyloglucan: xyloglucosyltransferase (EC 2.4.1.207); keratan-sulfate endo-1,4-b-galactosidase (EC 3.2.1.103); endo-1,3-b-glucanase (EC 3.2.1.39); endo-1,3(4)-b-glucanase (EC 3.2.1.6); licheninase (EC 3.2.1.73).

3.6 Heavy metal resistance genes

During the past few years, the raising interest in the study of resistance and bioaccumulation capabilities toward heavy metals in *Rhodotorula* species revealed them as good candidate for the developing of new strategies aimed to remove heavy metals from contaminated matrix (Garza et al., 2016). Recently, a proteomic study aimed to assess the physiological response of *Rhodotorula mucillaginosa* to the exposure of cadmium, arsenic, lead and copper revealed the presence of an efficient detoxification system and bioremoval capabilities (Ilyas et al., 2016).

In WP1, genes involved in heavy metal homeostasis and arsenic resistance were detected. Two contiguous genes encoding for a hypothetical protein (Protein IDs: 27572) and a trans-membrane protein member of the ACR3 family of arsenite permease (Protein ID: 53976). The hypothetical protein and the arsenite permease share a sequence similarity of 37 and 57 % with the arsenate reductase (ARR2) and the arsenite permease ARR3 of *S. cerevisiae* respectively (Bobrowicz et al., 1997). The arsenate reductase from WP1 possess the typical catalytic amino acid motif HCX_5R (Landrieu et al., 2004) (Figure 15A). Interestingly, downstream to the ACR3 and ARR2 coding sequences, genes involved in the arrest of cell cycle were detected, suggesting how *ars* genes might be part of a cluster under the control of ROS signaling in response to heavy metal exposure (Figure 15B). On a different chromosomal region, three additional genes encoding for the arsenical resistance protein ArsH, and the genes encoding for a protein-tyrosine phosphatase and an N-6 adenine-specific DNA methylase, respectively involved in arsenate reduction and bio-methylation of arsenite, were detected. However, in addition to the canonical catalytic site HCX_5R of ACR2 arsenate reductases, this second phosphatase possess multiple glycine-rich phosphate binding motif (GXGXXG) indicating that it might have retained a phosphotyrosine phosphatase catalytic activity (Mukhopadhyay et al., 2003).

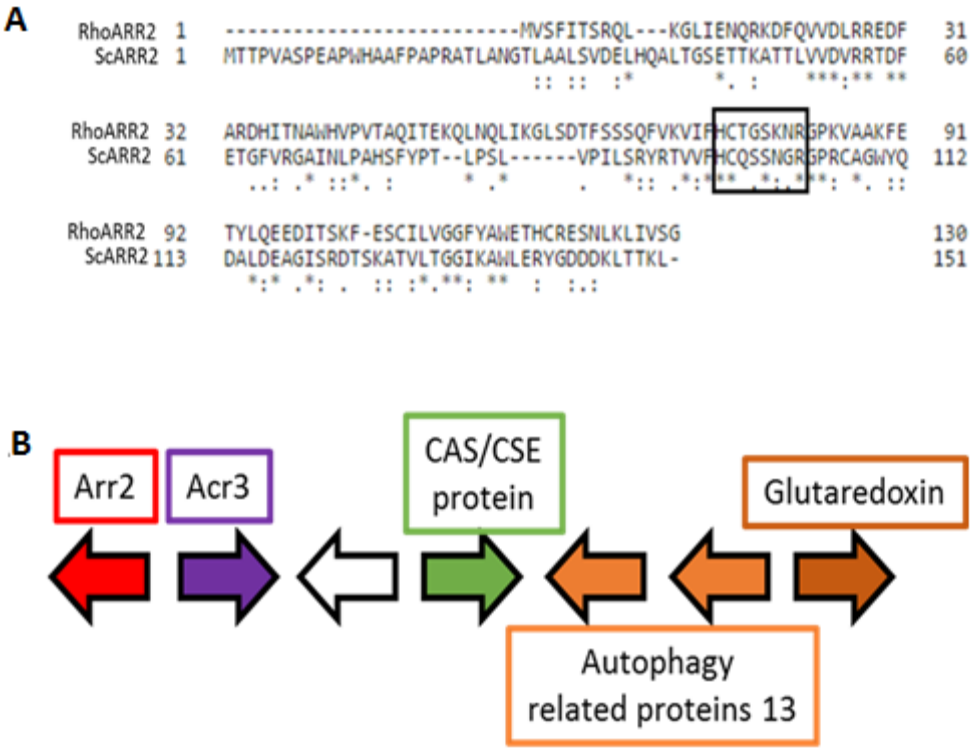


Figure 15. A) Sequence alignment of WP1 ACR2 proteins against S. cerevisiae arsenate reductase (Q06597). Black box underline the conserved the active site motif HCX_5R conserved in all ACR2 arsenate reductases. B) Genes map of *orfs* from position scaffold_7:1380515-1366031 involved in arsenic resistance (Arr2 and Acr3), response to oxidative stresses (glutaredoxine) and regulation of cell cycle/apotosis (CAS/CSE and Autophagy related protein).

Another interesting feature of WP1 regarding the resistance to heavy metals is the presence of a phytochelatine synthase coding gene (Protein IDs: 55075) and a putative cysteine-rich protein of 75 amino acids of the metallothionein (MT) superfamily (Protein IDs: 43376). The WP1 metallothionein possess two almost identical 7-cysteine motif [CXC]-X_5-[CX_2C]-X_4-[CXC]-X_2-C and [CXC]-X_5-[CXC]-X_3-[CXC]-X_2-C, which are probably the result of a domain duplication event. From position 45 to position 71, the cysteine motif presents a perfect match with the cys residues of the motif X_2-[CXC]-X_5-[CXC]-X_3-[CXC]-X_2-C-X_3 of *Neurospora crassa* metallothionein. While the amino acidic sequence upstream to position 45 show a partial conservation with in the cys motif of the 7-cys building block-[CXC]-X_3-[CSC]-PPGX-[CXC]-AXCP from *Tremella mesenterica strain PYCC 5472* metallothionein (Figure 16).

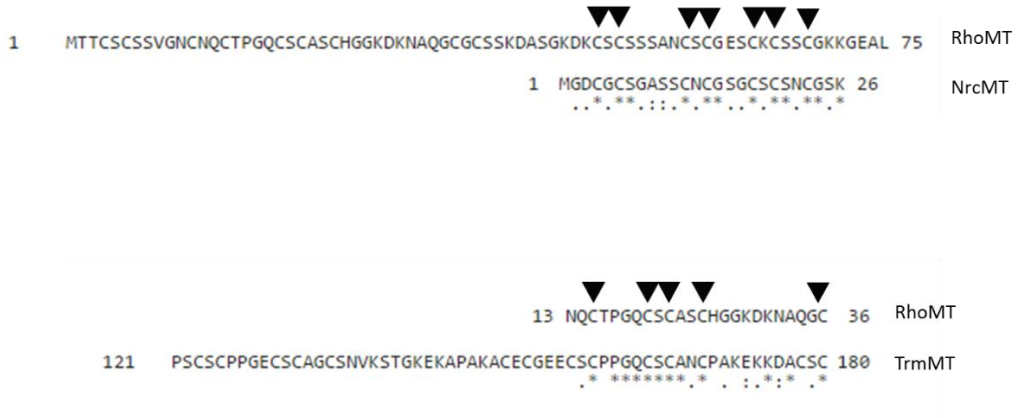


Figure 16. Alignment of of R. graminis WP1 metallothinein (RhoMT) metallothionein cys residues of *N. crassa* metallotionein (NrcMT; UniProt ID: P02807) and *T. mesenterica* (TrmMT) (Iturbe-Espinoza et al., 2016)

3.7 Phytohormone production

Under in vitro condition WP1 was able to produce the phytohormones giberellins (GA3), Jasmonic acids (JA), abscisic acid (ABA) and Brassinosteroids (Khan et al., 2016). Among endophytic bacteria isolated from wild poplar, WP1 was the best producer of Indole-3-acetic acid (IAA) and Brassinosteroids (Annex 1). The WP1 genome lacks the standard genes encoding for proteins involved in the biosynthesis of IAA via indole-3-pyruvate (KEGG Entry: R00677; R00684) and indole-3-acetamide (KEGG Entry: R00679). However, three putative proteins, an aromatic-L-amino-acid decarboxylase (Protein ID: 35429), a monoamine oxidase (Protein ID: 54216) and an indol-3-acetaldehyde dehydrogenase (Protein ID: 14581) can participate in the conversion of L-tryptophan (trp) to IAA via tryptamine. In this pathway, trp is first decarboxylated to tryptamine (KEGG Entry: R00685) which is subsequently oxidized into indole-3-acetaldehyde and then to IAA through two consecutive oxidation steps carried out by an amine oxidase (KEGG Entry: R02173) and an aldehyde dehydrogenase, respectively (KEGG Entry: R02678). Furthermore, two putative hydrolases (Protein ID: 34153; 66162) belonging to the nitrilase superfamily, which are important for the microbial colonization of plants due to their role in nitriles detoxification and utilization of plant nitriles as carbon and nitrogen source (Howden and Preston, 2009; Howden et al., 2009), can also convert the indole-3-acetonitrile (IAN) into IAA (KEGG Entry: R03093). However, how IAN synthesis occur in microbes is still unclear (Fu and Wang, 2011). Annotation of functional domains in WP1 proteome, 15 proteins were annotated as containing a 2OG-Fe(II) oxygenase domain (IPR005123) which are primarily involved in the biosynthesis of gibberellins and other plant hormones (Prescott and Lloyd, 2000; Zhao et al., 2013; Farrow and Facchini, 2014). Finally, R. graminis WP1 has a set of genes involved in the synthesis of (R)-acetoin and (R,R)-2,3-butandiol, two well-known VOCs that increase

resistance to plant pathogens and also act as growth promoting factors (Johnston-Monje and Raizada, 2011; D'Alessandro et al., 2014). In WP1, a putative metabolic pathway which leads to the synthesis of (R)-acetoin and (R,R)-2,3-butandiol starts with the decarboxylation of pyruvate into 2-acetolactate (KEGG Entry: R00006) by a putative acetolactate synthase (Protein ID: 32290; 35922). Under aerobic conditions, the synthesis of (R,R)-2,3-butandiol (KEGG Entry: R02946) from (R)-acetoin by two putative NADH-dependent dehydrogenase (Protein ID: 39181; 46342) occurs through the spontaneous decarboxylation of 2-acetolactate into (R)-acetoin (Atsumi et al., 2009).

Conclusion

There is a growing interest in the characterization plant microbiome due to its impacts on plant health and growth. With the sequence of the poplar genome (Tuskan et al., 2006) and multiple studies of the poplar microbiome (Hacquard and Schadt, 2015), poplar can become a model system for studying tree-microbiome interactions. The Laccaria-Populus interaction is a well-studied mycorrhizal mutualism at the molecular level (Podila et al., 2009). However, nature of the interactions between mycorrhizal and plant might be quite different when compared to endophytes.

From a comparative genome study, we show that WP1 possess unique genetic features, not detected in phylogenetically related species, for genes with a biological relevance in the uptake and metabolism of plant secondary metabolites and plant colonization. Indeed, similarly to obligated biotroph, in WP1 two copies of secreted CCGs were detected in a chromosomal region partially in synteny with the non-endophytic counterpart NP11. Furthermore, compared to obligated biotrophs, we found that, in line with its phylogenetic affiliation, the pool of SSPs was involved in a massive loss of genes encoding for proteins with a protein-protein interaction domain (elicitors) and for hydrolytic enzymes (proteases and glycosyl hydrolases), suggesting that WP1 does not possess genetic potential to generate that strong interaction with the host. On the other hand, WP1 possess a unique set of SSPs involved in the metabolism of nitrogen compounds and in the recognition of cyclic and heterocyclic organic compounds. These findings suggest that WP1 is a microorganism well adapted to colonize and survive within plant niches without creating a strong interaction with the host. The utilization of WP1 as plant growth promoting microorganism is also supported by its capability to synthetize a wide spectrum of plant hormones. Despite the effectors involved in the synthesis of IAA and gibberellins were detected, the lack of jasmonate and brassinosteroids biosynthetic pathways suggest that the synthesis of such phytohormones is accomplished through undiscovered pathways. The synthesis of phytohormones is highly relevant in the context of microbe-host interaction, since these secondary metabolites exert a relevant role in the host response to pathogens, abiotic stresses and in the improvement growth performances.

Studies conducted on well-known plant symbiont or plant pathogens such as *Laccaria bicolor* and rust-fungi respectively, have been extremely precious in the identification of the molecular systems exploited by endophytes to colonize and complete its lifecycle in association with the host. However, such dual system excludes the contribution of the microbiome in the context of such mechanism.

Thus, the characterization of WP1 genome would allow us to focalize our next studies in the understanding how such genetic traits are involved in a more complex interaction system: plant-(microbiome)-microbe of interest. A possible strategy to accomplish this, would be the utilization of of brand new molecular techniques such as Tag-Seq (Rozemberg et al., 2016) which would allow us to focus and monitor the expression level of hundreds of genes within the plant microbiome for the microorganism of interest. Without an in-depth genomic characterization, this would not be possible.

Chapter 2

Arsenic speciation and bioaccumulation in *Rhodococcus atherivorans* BCP1

1. Introduction

The environmental forms of inorganic arsenic are represented by arsenate (As(V)) and arsenite (As(III), where the latter is the most cytotoxic due to its high reactivity against biomolecules such as DNA and proteins (Rodriguez-Gabriel and Russel et al., 2005). Microorganism have developed different strategies used in arsenic detoxification. Under aerobic condition, the most common strategy involves the utilization of specific class of redox enzymes, classified as arsenate reductase, that convert As(V) into As(III) using glutathione (GSH). In Actinobacteria, the reduction of As(V) into As(III) is carried out by the enzyme arsenate-mycothiol transferase, which utilize the mycothiol instead of GSH (Ordonez et al., 2009). At this stage, the trivalent form of the oxyanion is usually extruded outside the cell through the ABC transporter ArsB with the hydrolysis of an ATP molecule. However, in plant and fungi As(III) can be chelated and stored inside vacuoles or, methylated and converted into a volatile form. These, are the most common detoxification mechanism that occur in bacteria, plant and fungi. While these processes usually drive the cellular redox potential through a progressive oxidative state, other microorganisms, mainly anaerobic bacteria, can couple the arsenic detoxification with the production of chemical energy (Páez-Espino, 2009). From an applicative point of view, it has been stated that a microorganism suitable for the decontamination of arsenic from environmental matrix should possess the following characteristics: (i) capability to tolerate and bioaccumulate high level of the oxyanion; and (ii) reduced capability to convert As(V) in the most toxic form As(III) (Mateos et al., 2006). However, regarding the capability of a strain to limit the conversion of As(V) into As(III), this aspect is more dependent by the chemical-physical properties of the soil. The equilibrium between As(III) and As(V) in soil is strictly dependent by the pH and the redox potential. Indeed, at negative soil redox potential most of the arsenic is present as As(III) while oxidizing conditions both As(III) and As(V) are present (Marin et al., 1993). Thus, the capability of strain to limit the conversion of As(V) into As(III) is more dependent by the microenvironment than by the microorganism itself.

In the context of arsenic phytoremediation, BCP1 show some intriguingly characteristic in terms of bioaccumulation potential, tolerance toward high concentration of As(V) and As(III) and capability to convert As(V) into As(III). Indeed, the latter could be exploited in order to enhance the plant uptake of the most toxic form As(III).

2. Materials and Methods

2.1 Strain isolation, growth media composition and reagents.

Rhodococcus atherivorans BCP1 it is a Gram negative, aerobic non-sporulating bacterium belonging to the Mycolata group of *Actinomycetes* and isolated from a butane-utilizing microcosms for its capability to grow on gaseous n-alkanes (Presentato et al., 2017; Frascari et al., 2005). Liquid cultures of BCP1 were grown in agitation at 150 rpm at the 30 °C. The compositions of the media used in this work are reported. Luria Bertani growth media: NaCl 10 g L^{-1}, Yeast Extract 5 g L^{-1}, Tryptone 10 g L^{-1}; M9 minimal growth media: Na_2HPO_4 (7H2O) 12.8 g L^{-1}, KH_2PO_4 6.8 g L^{-1}, NaCl 0.5 g L^{-1}, NH_4Cl 1 g L^{-1}, $MgSO_4$ 1 mM, $CaCl_2$ 100 µM. All ICP-OES analysis were performed using ultrapure reagents for trace element analysis.

2.2 BCP1 growth in presence of Arsenate and Arsenite

Evaluation of Minimum Inhibitory Concentration (MIC) was assessed in M9 minimal growth media in presence of glucose 0.2% (w/v). Metal exposure was carried out on a bacterial suspension obtained from a pre-inoculum stopped in mid log phase. Before exposure to arsenic (As(III) and As(V)), cells were washed two times in PBS and suspended in M9 to match a final OD_{600} of 0.04 – 0.05. The exposure to arsenate and arsenite was carried out for 24 h at 30 °C under 150 rpm.

2.3 Minimum biocide concentration to As(V) and As(III)

BCP1 biofilm growth was performed on the MBEC-P&G system, commercially known as Calgary Biofilm Device (CBD), as described by Ceri et al. (1999). The CBD consist of a plastic lid, with 96 conical protruding pegs, that can be adapted on a 96 well microtiter plate loaded with the bacterial suspension. At optimal temperature and under shacking condition, biofilm growth occurs on peg surface. In order to obtain a BCP1 microbial biofilm, from the cryogenic stock, bacterial biomass was streaked out in LB agar plates and incubated at 30 °C for 48 hours. Second subculture was made streaking out a single colony from the first subculture and incubated for 72 hours. After that, the bacterial biomass from the second subculture was gathered with a sterile cotton swab and suspended in LB liquid media to match an optical density at 600 nm of 0.76-0.8 (OD_{600}). A dilution 1:30 was prepared to obtain standardized inoculum, which was used to inoculate the CBD with 180 µl of bacterial suspension. Biofilm growth was performed at 30 °C under shacking condition. After 48 h of incubation in LB, biofilms were exposed to increasing concentrations of As(V) and As(III) in the range of 75 – 450 mM and 2 – 25 mM respectively. Metal exposure of 48 h-old biofilms was carried

out in M9 media in presence of glucose 0.2 % (w/v). Prior to metal exposure, the plastic lid with pegs colonized by 48 h-old biofilms were rinsed in saline solution to remove loosely adherent cells and transferred into Challenge plate. The Challenge plate is a 96 microtiter plate where each well was loaded with 180 µl of M9 containing the heavy metal. Before metal exposure, few pegs were collected to assess biofilm vitality at 48 h by viable cell count. 48-h-old biofilm where then exposed to increasing concentration of As(III) and As(V). Each concentration was tested in quadruplicate. Exposure to the heavy metal was carried out for 24 h at 150 rpm. After 24 h, biofilms were rinsed twice in saline solution, to remove any trace of the metal.

In order to assess cells vitality within biofilms, pegs were removed from the lid with a pair of sterile pincers, rinsed in saline solution (0.9% w/v NaCl), to remove loosely adherent cells from peg surface, and transferred in 300 µl of saline solution with 0.1 % (v/v) of Tween 20. After that, pegs were sonicated in ice for 10 minutes to detach cells from pegs surface. The bacterial suspensions obtained from sonication was serial diluted to assess biofilm viability.

2.4 Analysis of arsenic speciation

After exposure to the metals, cells were collected via centrifugation and washed two time in PBS and freeze-dried. For analysis of arsenic level in growth media, 10 mL of the growth media were filtered using 0.22 µm PVDF filters and stored at – 80 °C until arsenic speciation analysis.

Freeze-dried biomass was accurately weighed on an analytical balance and added to 25 mL of a mixture of 1% of HNO_3 and left to react overnight. The samples were then subjected to a cycle microwave assisted extraction according to the following program:

Table 1. Microwave settings

Steps	Time (min)	Temperature (° C)
1	3	increased up to 55 ° C
2	10	55 ° C
3	2	increased up to 75 ° C
4	10	75 ° C
5	2	increased up to 95 ° C
6	30	95 ° C

Samples were collected into falcon tubes and subjected to centrifugation for 15 minutes at 10000 rpm at a temperature of 4 ° C. The supernatant was filtered using 0.22 µm PVDF filter and subjected to speciation analysis. As(V) and As(III) concentrations were assessed using the Hydride generation technique coupled with a PerkinElmer™ Optima 7300 DV Inductively coupled plasma – Optical Emission Spectrometry (ICP-OES) (Edmundson and Horsfall, 2015). The arsenic in the samples was

determined as follows. For quantifying total arsenic (As(III)) and As(V)), samples were mixed with equal volumes of a reducing solution composed by potassium iodide (KI) 5% and ascorbic acid 5 %, and incubated at room temperature for 1 h. Doing this, all the As(V) is converted into As(III). Finally, As(III) is converted into arsine (AsH_3) by adding an equal volume of 0.2% $NaBH_4$. Arsine quantification was performed using the ICP-OES. As(III) was quantified as stated before but without adding the reducing agents to the samples. Doing this, As(V) was quantified by subtracting As(III) level to total arsenic. A calibration curve was made using arsenic trioxide (As2(O3)) at the following concentration: 0, 0.5, 1, 3, 5, 10, 20, 40 ppb. All the samples were opportunely diluted in order to fit within the As(III) concentration range used for the calibration curve. The Hydride generation technique has been coupled with a PerkinElmerTM Optima 7300 DV Inductively coupled plasma – Optical Emission Spectrometry (ICP-OES) for the determination of arsenic species. Arsenic speciation experiments were performed in triplicates.

2.5 Ultrastructural analysis of BCP1 cells exposed to arsenic

Ultrastructural analysis of BCP1 cells exposed to arsenate were carried out as described in Pesciaroli et al., (2012). Samples were fixed and dehydrated for 3 days with decreasing ethanol-LR White resin (SPI Supplies, West Chester, PA) ratios. After three days of dehydration, samples were embedded in LR White resin and thin sections ranging from 60 to 80 nm were obtained using a Reichert Ultracut ultramicrotome (Leica Microsystems Srl, Milan, Italy) equipped with a diamond knife. Thin sections were transferred on copper grids, stained with uranyl acetate and lead citrate and observed with a JEOL 1200 EX II electron microscope (JEOL). Images were acquired with the Olympus SIS VELETA charge-coupled device camera equipped with iTEM software (Olympus Soft Imaging Solutions GmbH, Muenster, Germany). Negatively stained BCP1 cells grown in the presence of arsenic were obtained using a 1% phosphotungstic acid solution (pH 7.3). TEM observations of BCP1 were carried out by mounting 5 µL of each sample on carbon-coated copper grids (CF300-CU, Electron Microscopy Sciences). Air-dried samples were observed with a Hitachi H7650 TEM.

Electron dense nanoparticles produced by BCP1 exposed to arsenic were extracted as described in Presentato et al., (2016). Roughly, BCP1 biomass were collected by centrifugation (3700 rpm) for 20 min after 5 days exposure to the heavy metal in M9. The pellets were washed twice with saline solution (NaCl 0.9% w/v). Cells were lysed in Tris-HCl (1.5 mM) buffer pH 7.4 by ultrasonication (MICROSON™ Ultrasonic Cell Disruptor XL, Qsonica Misonix Inc.) on ice at 22W for 10 min pausing 30 second 30 second of sonication. Cellular debris were separated via centrifugation at 3700 rpm for 20 minutes. The supernatant was then recovered and incubated overnight (16h) at 4°C with

1-Octanol (Sigma-Aldrich®) in a ratio 4:1 (v/v). Metal nanoparticles were recovered by centrifugation (16,000 rpm) for 15 minutes and finally suspended in deionized water.

2.6 Genome analysis of arsenic resistance genes

The genetic determinants and analysis of metabolic pathways involved in arsenic detoxification were computed using Pathways Tools v18.2 (Caspi et al., 2014) in combination with MetaCyc and the Integrated Microbial Genome database as reference database (Markowitz et al., 2012), for metabolic pathway and genome data respectively. To predict metabolic pathways outside the BCP1 taxonomic group, the pruning mode was turned off. Finally, promoter analysis was performed using the Neural Network Promoter Prediction to detect putative pseudogenes within the gene clusters of interest (Reese, 2001).

2.7 Statistical methods

BCP1 growth under planktonic and biofilm conditions was carried out with five and four biological replicates. Arsenic (As(III) and As(V)) speciation was tested using five biological replicates. Data analysis was carried out in R v3.2.3, performing a Tukey's HSD (Honest Significant Difference) as post-hoc test in conjunction with a One-Way ANOVA.

Results and Discussion

3.1 As(V) and As(III) MIC

Prior to arsenic speciation experiments, the MIC was assessed. Resistance to As(V) and As(III) were assayed in M9 minimal media in presence of glucose. With a starting number of Log (CFU mL^{-1}) ranging from 5 and 5.5, the concentration of As(V) where BCP1 growth was inhibited was > 240 mM (Figure 1).

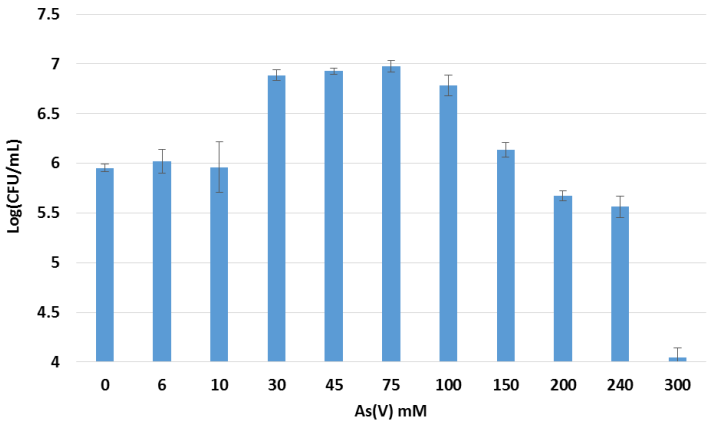


Figure 1. BCP1 resistance to As(V). The values and standard deviations are the means of five biological replicates. The number of colony forming unit (CFU) per mL (planktonic) are reported in log_{10}.

Similar levels of resistance were observed in other *Rhodococcus* species isolated from arsenic contaminated soils (Shagol et al., 2014; Ghosh and Sar, 2013). Surprisingly, the growth of BCP1 was greatly enhanced at arsenic concentration higher than 30 mM, appearing to gain metabolic energy from arsenate reduction. However, such hypothesis can be excluded since only anaerobic bacteria are able to use arsenate as electron acceptor in anaerobic respiration (Saltikov and Newman, 2003). A similar behaviour was observed in the aerobic bacteria *Exiguobacterium* sp. WK6 and *Aeromonas* sp. CA1 where, as consequence of arsenate reduction to arsenite, the alkalinisation of the growth media allowed these bacteria to grow for an extended period (Anderson and Cook, 2004). In our case, BCP1 growth was exceptionally higher just after 24 h of incubation with As(V) and no change in pH values where observed between cells with and without arsenic (data not shown). Thus, the increased growth performance of BCP1 in presence of As(V) was not determined by a lower degree of acidification of the growth media during the exposure to the metal. Finally, As(III) resistance in BCP1 is reported (Figure 2). With a starting number of 5.79 Log(CFU mL^{-1}), the MIC values are in line with the inhibitory concentration observed in *Rhodococcus* sp. NAU-1 (> 8 mM) (Jain et al., 2012).

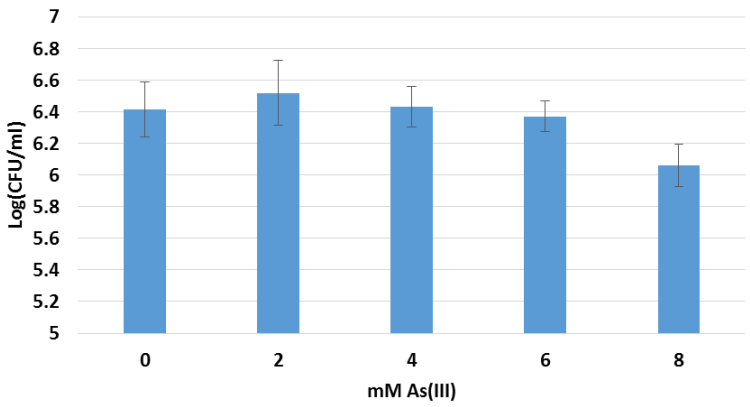


Figure 2. BCP1 resistance to As(III). The values and standard deviations are the means of five biological replicates. The number of colony forming unit (CFU) per mL (planktonic) are reported in log10.

3.2 Arsenic MIC and Minimum Biocide Concentration (MBC) in BCP1 biofilms

High tolerance to antimicrobial is a common feature of microbial biofilms. Within biofilms, heavy metals are highly reactive against the extracellular matrix secreted by the cells during the biofilm growth. Polysaccharides, nucleic acids and proteins are the principal components of such matrix, creating a poly-ionic surface highly reactive against a wide range of metals and metalloids (Harrison

et al., 2007). For this reason, the analysis of MIC and MBC_{90} will give to us a better understanding of BCP1 tolerance (MBC) and resistance (MIC) to As(III) and As(V) (Table 2).

Table 2. BCP1 biofilms resistance (MIC) and tolerance (MBC_{90}) to As(V) and As(III).

	As(V)			As(III)		
Strain	MIC	MBC_{90}	% Kill[1]	MIC	MBC_{90}	% Kill[1]
BCP1	>270 mM	420 mM	91 ± 2.4	10 mM	18 mM	91 ± 3

[1]Kill % was calculated as follow: % = [1 – (remaining cfu/PEG) / (initial cfu (before metal exposure)/PEG)] × 100. The values and standard deviations are the means of four biological replicates.

As expected, As(V) MIC in BCP1 biofilms were higher when compared to the planktonic counterpart. Intriguingly, when BCP1 is exposed to As(III), no substantial differences in MIC values were observed between planktonic and biofilms cells. A possible explanation is that as result of As(III) exposure. This might be the results of a dispersal mechanism triggered after the exposure to As(III). In presence of an oxidative stress a "specialized" subpopulation of microbial cells within biofilm switch from a sessile mode of growth to a planktonic state (Gambino and Cappitelli, 2015). Due to polysaccharide production and high metabolic heterogeneity, microbial cells within biofilm usually show different degrees of metal susceptibility, where only a subpopulation of cells is exposed or susceptible to the toxicant. Because in the environment, many bacteria tend to growth as complex microbial community in association to a biotic or abiotic surface, the MBC better represent the actual tolerance of BCP1 to arsenic (Teitzel and Parsek, 2003).

3.3 As (V) speciation in BCP1.

Arsenic speciation in BCP1 was assessed at sub inhibitory concentration of As(III) and As(V). Because BCP1 show a better growth performance at high concentrations of As (V), the speciation of the oxyanion was assayed at 6 and 34 mM (Figure 3).

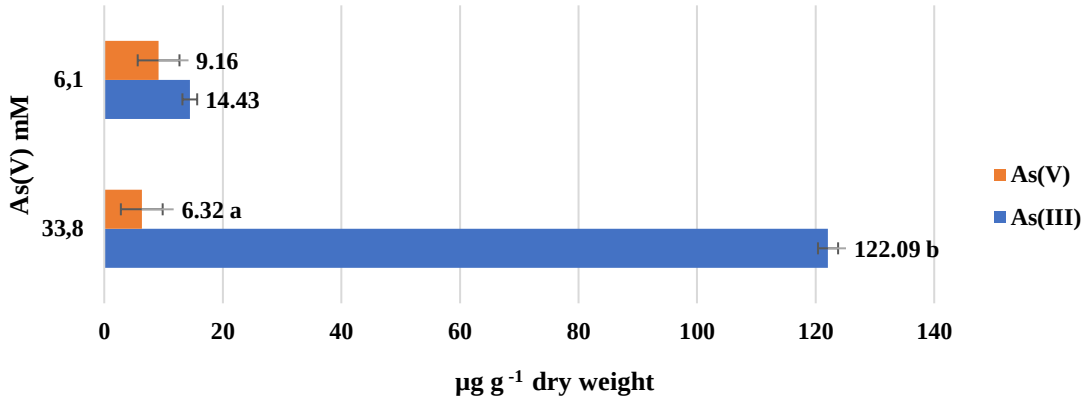


Figure 3. As(V) speciation in BCP1. No change in arsenate concentration were observed for the abiotic control. As(V) and As(III) concentrations were reported as µg of metal per g of BCP1 biomass (dry weight). Each concentration was tested in triplicate. Letters represent groups with statistically significant differences between means (alpha = 0.05). Statistically different groups were determined using a Tukey's HSD post-hoc test.

As expected, the amount of biomass collected after 24 h exposure to As(V) was higher when BCP1 was exposed to 34 mM of As(V). Indeed, with a starting amount of biomass of 7 mg (dry weight) the biomass recovered at 6 and 34 mM after 24 h growth was 30 ± 1.7 mg and 85 ± 7 mg respectively. In presence of 33.8 mM of As(V) the As(III) detected in association with the biomass was 19.3 fold higher (122.09 ± 1.69 µg/g) compared to intracellular levels of As(V) (6.32 ± 3.52) µg/g. Such disparity in the intracellular ratio As(V)/As(III) in presence of 34 and 6 mM of As(V) are possibly due to a strong upregulation of the arsenate reductase genes. Indeed, at 6.1 mM of As(V) the concentration of total arsenic (As) in the growth media was 30 percent less than the initial concentrations, while at 34 mM of As(V) only 10 percent less of the initial concentration was detected. A decrease in arsenic removal at higher concentration can be attribute to an increased activity of the arsenite transporter ArsB, that is involved in the extrusion of intracellular As(III).

3.4 Arsenic Bioaccumulation in BCP1

The capability of BCP1 to reduce As(V) was discussed in paragraph 3.3. Because the experiment was carried out exposing to the oxyanion As(V) a mid-log phase culture, the changes observed in As(III) and As(V) are representative of the genetic potential of BCP1 to detoxify the arsenic. Dobrowlosky et al., (2016) reported that the extracellular matrix recovered from 72 h-old planktonic culture of *Rhodococcus opacus* and *Rhodococcus rhodochrous* have a high capability to adsorb Cd(II), Pb(II), Ni(II), Co(II) and Cr(VI). That said, a 96 h-old planktonic culture of BCP1 was exposed to 34 mM of As(V) for 24 h in M9 in presence of 0.2% of glucose. The analysis of the intracellular and

extracellular arsenic concentration shows a significant reduction of the arsenic in the growth media from 4.69 g L^{-1} to 2.5 ± 0.36 g L^{-1} and an intracellular concentration of arsenic equal to 35.3 ± 13 mg g^{-1}.

3.5 Arsenite oxidation in BCP1

The biological oxidation from As(III) to As(V) is carried out by a class of molybdopterin containing proteins known as arsenite oxidase. The first crystal structure of an Arsenite oxidase purified from *Alcaligenes faecalis* was characterized by Ellis et al., (2001). The enzyme is composed by a small subunit with a Riesky-type cluster [2Fe-2S], and a large subunit containing a Mo site and a [3Fe-4S] cluster. Such enzymes have been detected in various *Rhodococcus* species but many of these do not possess the [3Fe-4S] cluster (Sultana et al., 2012). To date, experimental evidences of a *Rhodococcus* strain (NAU-1) able to convert As(III) into As(V) has been reported in Jain et al., (2012), showing an intracellular and extracellular concentration of 0.051 ± 0.007 mM and 2.56 ± 0.09 mM respectively. Similarly, when BCP1 was exposed for 24 h to 2 mM of As(III), the intracellular levels of As(V) were 5.66 ± 0.34 μg g^{-1}. No As(V) were detected in the growth media suggesting that the capability of BCP1 to oxidize As(III) to As(V) is most likely the result of an enzymatic action non-specific toward the As(III).

3.6 Ultrastructural analysis of BCP1 exposed to Arsenic

The capability of BCP1 to produce intracellular nanostructure of As was investigated at 24 and 96 h . At 24 h we observed the presence of electron dense Polyphosphate (polyp) granules (Tobin et al., 2007) in correspondence (Figure 4A) or inside (Figure 4B) vacuole-like structure which are potentially rich in poly-β-hydroxybutyrate and similar to those observed in GFAJ-1 cells when exposed to As in absence of phosphate (Wolfe-Simon et al., 2012).

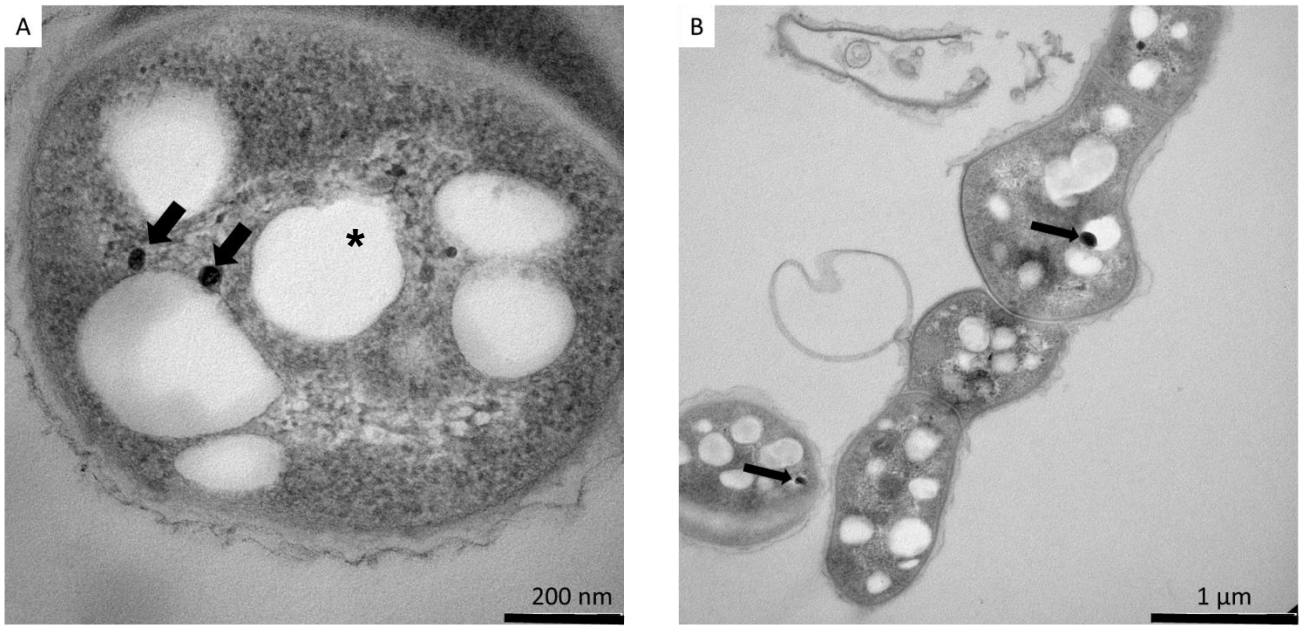


Figure 4. Thin sections of BCP1 cells exposed to 34 mM of As(V) show the presence of polyP granules (black arrows) in correspondence (A) or inside (B) vacuole-like structures (asterisk). No polyP granules where detected in absence of As(V) (data not shown)

In *Herminiimonas arsenicoxydans*, the expression of genes encoding for the phosphate transport systems is strictly connected with the accumulation of polyP granules in vacuole-like structures in arsenic-rich aquatic environments (Muller et al., 2007). Furthermore, has been shown that the genes involved in arsenic detoxification and phosphate uptake were located near to each other, suggesting that both process are coregulated. The capability of BCP1 to synthetize polyP granules in presence of arsenic arise some intriguingly utilization of this strain in presence of multiple heavy metals. For example, in *Chlamydomonas acidophyla* polyP granules are capable to chelate cadmium and lead (Nishikawa and Yamakoshi, 2003). However, the most interesting cellular structures were detected in the cellular extract of BCP1 cells exposed to 34 mM of As(V) for 96 hours. Indeed, in presence of 34 mM of As(V) hexagonal electron dense nanoparticle were detected (Figure 5A). Electron dense material in the cellular extract was also detected when BCP1 was treated with 6 mM of As(V) (Figure 5 B), but the dimension and the shape of such nanoparticles were not comparable to those observed at 34 mM. The utilization of Energy-dispersive X-ray (EDX) spectroscopy for the elemental analysis and chemical characterization these nanoparticles is ongoing, allowing us to determine if the Arsenic is part of this structures or mainly accumulated at the level of the extracellular matrix, as observed in the cyanobacterium *Synechocystis* PCC 6803 (Sure et al., 2016).

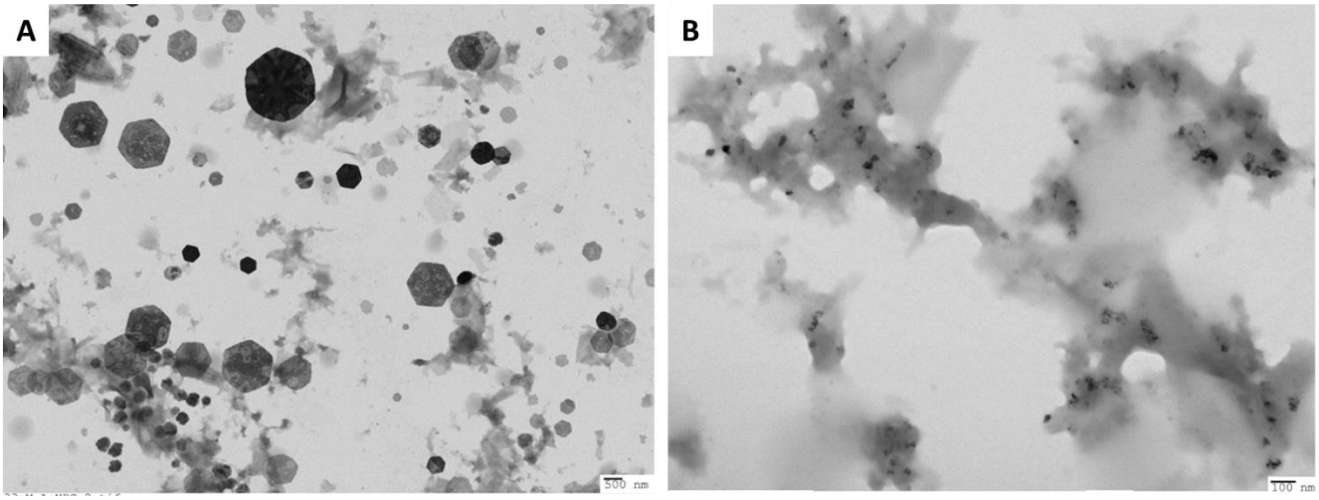


Figure 5. Electron dense nanoparticles detected in the cellular extract of BCP1 treated with 34 mM (A) and 6 mM (B) of As(V) after 96 h of treatment.

3.7 In-silico analysis of genes involved in arsenic resistance and detoxification

BCP1 *ars* operon is shown in Figure 6. The *ars* operon in BCP1 and phylogenetically related species to BCP1 carry three different arsenate reductases represented by an arsenate-mycothiol transferase (*ars*C1), a low molecular weight (LMW) Protein thyrosine-phosphatase (*ars*C) and a glutathione reductase (*ars*C'). A second set of protein coding genes *ars*B, *ars*A and *ars*D, which are involved in the recognition and extrusion of As(III) through the ABC transporter ArsB were also detected. BCP1 possess the complete set of genes necessary to perform the arsenic detoxification. As mentioned before, phosphate uptake and arsenic detoxification are tightly connected (Huertas and Michan, 2013). This finding is enforced by the presence of the *pst*SCAB operon downstream the genes involved in arsenate and arsenite detoxification in other *Rhodococcoccus* species such as *R. erythropolis* TUHH-12, *R. erythropolis* SK121 and *R. eclensis* NIO-1009.

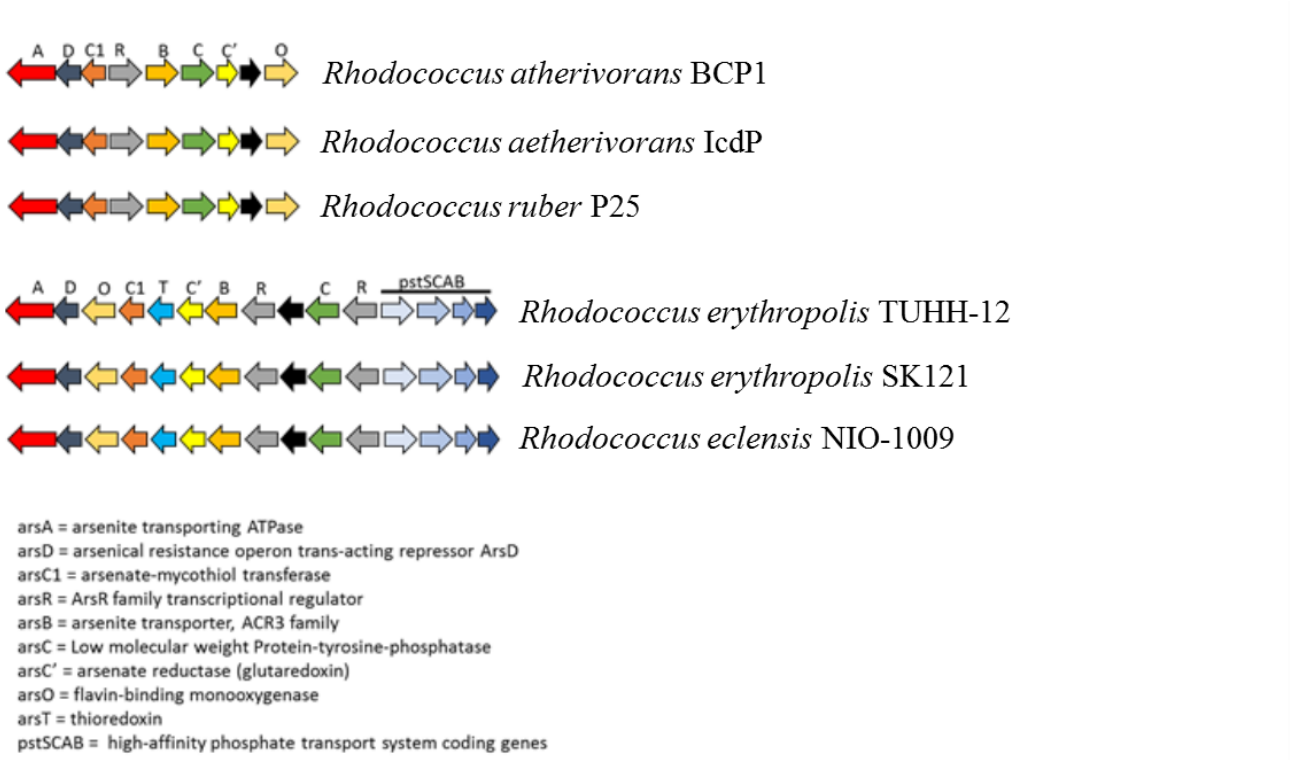


Figure 6. *ars* operon in BCP1 and phylogenetically related species *Rhodococcus aetherivorans* IcdP, *Rhodococcus ruber* P25, *Rhodococcus erythropolis* TUHH-12, *Rhodococcus erythropolis* SK121 and *Rhodococcus eclensis* NIO-1009. Black arrow: hypothetical protein.

In BCP1 the connection between the phosphate metabolism and the arsenic detoxification process might lie on a different level. Using PathwayTools, the detoxification of As(V) in BCP1 is achieved by the utilization of two different reducing agents: glutathione and mycothiol (Figure 7). Both pathways might be active or differentially regulated during a specific metabolic status determined by the intracellular and extracellular levels of phosphate.

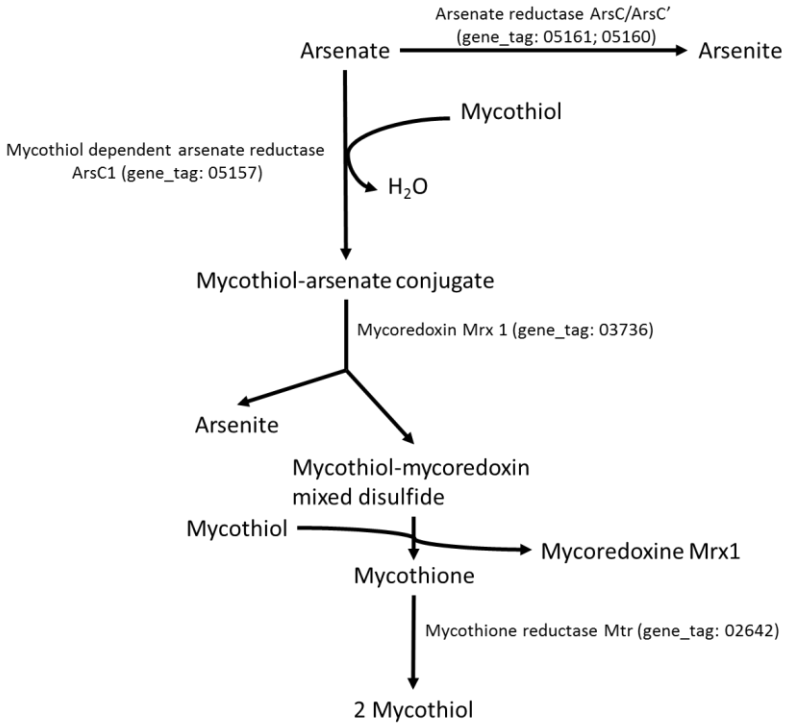


Figure 7. Mycothiol and glutathione dependent pathways involved in arsenic detoxification in BCP1.

The genes encoding for the mycothiol synthase (Locus tag: Ga0035244_01311), which is involved in the final step of the mycothiol synthesis (Buchmeier et al., 2006), was detected upstream the *pst*SCAB operon (Locus tag: Ga0035244_01312-01315), suggesting that under phosphate starvation, and consequent induction of the *pstSCAB* operon, positive effect in the expression of the mycothiol synthase genes would increase the intracellular levels of the reducing agent used in the first step of the conversion of As(V) into As(III). The effect of As(V) on the phosphate metabolism was observed when the cells were exposed to to high concentration (34 mM) of the oxyanion. Indeed, under 34 mM of As(V), BCP1 accumulate the phosphate ions under the form of polyP granules. At the same concentration, the intracellular levels of As(III) were roughly 10 fold higher compared to BCP1 treated with 6 mM of As(V). Would be interesting to test the detoxification performances of BCP1 under different phosphate concentrations without changing the levels of As(V) in the growth media. In *E. coli*, the over-accumulation of phosphate as polyp granules was achieved by overexpressing the genes encoding for polyP kinase (ppk) and the Pi-specific transport system (*pst*SCAB) (Morohoshi et al., 2002). Compared to the phosphate transport system PitA, the Pst system is 100-fold more selective toward phosphate than arsenate (Sousa et al., 2015). That said, the resistance to arsenic in BCP1 might be achieved through two different but highly interconnected mechanism: by increasing

the synthesis of mycothiol, and at the same time limiting the uptake of As(V) through the expression of the *pstSCAB* operon.

Conclusion

Under planktonic condition BCP1 was capable to accumulate arsenic up to 35.3 ± 13 mg g^{-1}, causing significant reduction of the oxyanion in the growth media from 4.69 g L^{-1} to 2.5 ± 0.36 g L^{-1}. The results of As(V) speciation suggest that BCP1 might actually increase the extracellular level of As(III), and thus increasing the toxicity of the surrounding media. However, this behaviour should not be considered a collateral effect of the exceptional detoxification capabilities of BCP1. In association with plants, there would be a positive effect in terms of As(III) removal from the surrounding media. As reported in Briat (2010), the detoxification of arsenic in plant mainly occur through two mechanisms: volatilization and vacuole sequestration of As(III) where the pool of glutathione is involved either in the reaction that convert the As(V) into As(III) and in the synthesis of phytochelatines (PC). The latter chelates the As(III) granting the transport of the PC-As(III) complex into the vacuole through the ABC transporter ABCC1-2. Hyperaccumulator plants like *Pteris vittata* L have a great detoxification potential due to high intracellular levels of free glutathione (Singh et al., 2006). That said, by exploiting the association with an hyperaccumulator , BCP1 would be a suitable candidate for the developing of microbe assisted phytotechnology aimed to increase the uptake and sequestration of As(III) and for the decontamination of organic pollutants and heavy metals. Finally, due to its high resistance to As(V) and also to the capability to grow better in presence of the oxyanion, outside the phytoremediation context, BCP1 is a good candidate for the production of engineered strains capable to convert arsenic into an insoluble form. Recently, an *E. coli* strain transformed with the phytochelatin analogue EC20, was able to convert arsenic into the stable and insoluble metallic form As(0) (Edmundson et al., 2016).

Conclusions and Future Perspectives

In the current study we employed different approach to characterize the suitability of WP1 and BCP1 for the developing of microbes assisted phytotechnologyes. Despite their unique features in the as plant growth promoting microorganisms and degradation of organic pollutants, interesting aspects in the context of arsenic, and more in general heavy metal homeostasis, were detected in both microorganisms. WP1 possess a complete set of genes (e.g. phytochelatine synthase and metallothioneines) which are known to be involved in the chelation of a very wide range of heavy metals, while BCP1 possess the exceptional capability to perform the reduction of arsenate into

arsenite in presence of high concentration of the pentavalent form of the oxyanion. In the context of arsenic removal, BCP1 would be deleterious if associated with non-hyperaccumulator plants. However, we might speculate that by associating WP1 to BCP1, would have a positive effect in the reduction of As(III) bioavailability, by significantly reducing the uptake of the oxyanion. More specifically, the capability of BCP1 to reduce high levels of As(V) into As(III), combined with the catalytic activity of phytochelatine and metallothionine from WP1, would create a scenario where the most cytotoxic form of the arsenic is converted into an insoluble form, thereby reducing the bioavailability of the arsenic for the plant. To date an effective system capable to convert As(III) into an insoluble form has been accomplished by the utilization of bacterial strains engineered with phytochelatine analogues, where the utilization of the latter for bioremediation processes will require extensive risks assessment studies.

List of Papers

- Doty SL, Freeman JL, Cohu CM, Burken JG**, Firrincieli A***, Simon A, Khan Z, Isebrands JG, Lukas J, Blaylock MJ (2017) Enhanced Degradation of TCE on a Superfund Site Using Endophyte-Assisted Poplar Tree Phytoremediation. Environ Sci Technol.;51(17):10050-10058. (*Soil Chloride in the Rhizosphere of CP and EP; Genome Analysis for Genes Related to Bioremediation; Verification of the Colonization Ability of Endophyte Strain PDN3; Statistical Analysis)

- Evangelistella C, Valentini A, Ludovisi R, **Firrincieli A***, Fabbrini F, Scalabrin S, Cattonaro F, Morgante M, Mugnozza GS, Keurentjes JJB, Harfouche A1 (2017) De novo assembly, functional annotation, and analysis of the giant reed (Arundo donax L.) leaf transcriptome provide tools for the development of a biofuel feedstock. Biotechnol Biofuels.;30;10:138 (*Field experiments; Bioinformatics analyses during the peer review process)

- Kandel SL, **Firrincieli A***, Joubert PM, Okubara PA, Leston ND, McGeorge KM, Mugnozza GS, Harfouche A, Kim SH, Doty SL (2017) An In vitro Study of Bio-Control and Plant Growth Promotion Potential of Salicaceae Endophytes. Front Microbiol.; 8:386. (*Comparative analysis of genomic regions related to the synthesis of anti-Rhizoctonia metabolites in *Burkholderia* species; Assessing the plant growth promoting properties in Burkholderia spp. (WP40 and WP42) genomes)

- **Firrincieli A**, Otillar R, Salamov A, Schmutz J, Khan Z, Redman RS, Fleck ND, Lindquist E, Grigoriev IV, Doty SL (2015) Genome sequence of the plant growth promoting endophytic yeast Rhodotorula graminis WP1. Front Microbiol.;6:978.

- Khan Z, Rho H, **Firrincieli A***, Han Hung S, Luna V, Oscar Masciarelli O, Kima Doty SL (2016) Growth enhancement and drought tolerance of hybrid poplar upon inoculation with endophyte consortia. Curr. Plant Biolo.;6,38-47 (Comparative genome analysis)

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol.;215(3):403-10.

- Anderson CR, Cook GM (2004) Isolation and characterization of arsenate-reducing bacteria from arsenic-contaminated sites in New Zealand. Curr Microbiol.;48(5):341-7

- Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol. 55(4):539-52

- Ashburner M, Ball C. A, Blake J. A, Botstein D, Butler H, Cherry J. M, et al (2000) Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat. Genet. 25:25–29.

- Atsumi S, Li Z, Liao JC (2009) Acetolactate synthase from Bacillus subtilis serves as a 2-ketoisovalerate decarboxylase for isobutanol biosynthesis in Escherichia coli. Appl. Environ. Microbiol. 75:6306–6311.

- Bais HP, Weir TL, Perry LG, Gilroy S, Vivanco JM (2006) The role of root exudates in rhizosphere interactions with plants and other organisms. Annu Rev Plant Biol. ;57:233-66.

- Bhasi A, Pandey RV, Utharasamy SP, Senapathy P (2007) EuSplice: a unified resource for the analysis of splice signals and alternative splicing in eukaryotic genes. Bioinformatics 23:1815–1823

- Birney E, Durbin R (2000) Using GeneWise in the Drosophila annotation experiment. Genome. Res. 10, 547–548.

- Bobrowicz P, Wysocki R, Owsianik G, Goffeau A, Ułaszewski S (1997) Isolation of three contiguous genes, ACR1, ACR2 and ACR3, involved in resistance to arsenic compounds in the yeast Saccharomyces cerevisiae. Yeast.13(9):819-28.

- Buchmeier N, Fahey RC (2006) The mshA gene encoding the glycosyltransferase of mycothiol biosynthesis is essential in Mycobacterium tuberculosis Erdman. FEMS Microbiol Lett.;264(1):74-9.

- Cappelletti M, Di Gennaro P, D'Ursi P, Orro A, Mezzelani A, Landini M, Fedi S, Frascari D, Presentato A, Zannoni D, Milanesi L (2013) Genome Sequence of Rhodococcus sp. Strain BCP1, a Biodegrader of Alkanes and Chlorinated Compounds. Genome Announc.;1(5).

- Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res. 42, D459–D471.

- Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution 17, 540-552.

- Ceri H, Olson ME, Stremick C, Read RR, Morck D, Buret A (1999) The Calgary Biofilm Device: new technology for rapid determination of antibiotic susceptibilities of bacterial biofilms. J Clin Microbiol.;37(6):1771-6.

- Chen F, Ren CG, Zhou T, Wei YJ, Dai CC (2016) A novel exopolysaccharide elicitor from endophytic fungus Gilmaniella sp. AL12 on volatile oils accumulation in Atractylodes lancea. Sci Rep. 5(6):34735.

- Chiarini L, Bevivino A, Dalmastri C, Tabacchioni S, Visca P (2006) Burkholderia cepacia complex species: health hazards and biotechnological potential. Trends Microbiol. 14(6):277-86

- Chijioke U, Winney L (2017) Optimal Removal of Heavy Metals From Leachate Contaminated Soil Using Bioaugmentation. CLEAN–Soil, Air. Wiley Online Library

- Comby M, Lacoste S, Baillieul F, Profizi C, Dupont J (2016) Spatial and Temporal Variation of Cultivable Communities of Co-occurring Endophytes and Pathogens in Wheat. Front Microbiol. 31;7:403.

- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 21(18):3674-6.

- D'Alessandro M, Erb M, Ton J, Brandenburg A, Karlen D, Zopfi J (2014) Volatiles produced by soil-borne endophytic bacteria increase plant pathogen resistance and affect tritrophic interactions. Plant Cell Environ. 37, 813–826.

- Ding C, Vidanes GM, Maguire SL, Guida A, Synnott JM, Andes DR, Butler G (2011) Conserved and divergent roles of Bcr1 and CFEM proteins in *Candida parapsilosis* and *Candida albicans*. PLoS One.;6(12):e28151.

- Dobrowolski R, Szcześ A, Czemierska M, Jarosz-Wikołazka A (2016) Studies of cadmium(II), lead(II), nickel(II), cobalt(II) and chromium(VI) sorption on extracellular polymeric substances produced by Rhodococcus opacus and Rhodococcus rhodochrous. Bioresour Technol.;225:113-120.

- Doty SL (2014) Endophytic Yeast Strains, Methods for Ethanol and Xylitol Production, Methods for Biological Nitrogen Fixation, and a Genetic Source for Improvement of Industrial Strains. Patent No. 8,728,781.

- Edgar, Robert C (2004), MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Research 32(5), 1792-97.

- Edmundson MC, Horsfall L (2015). Construction of a Modular Arsenic-Resistance Operon in E. coli and the Production of Arsenic Nanoparticles. Front Bioeng Biotechnol.; 20(3):160.

- Ellis PJ, Conrads T, Hille R, Kuhn P. Crystal structure of the 100 kDa arsenite oxidase from Alcaligenes faecalis in two crystal forms at 1.64 A and 2.03 A (2001) Structure;9(2):125-32.

- Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence J Mol Biol. ;300(4):1005-16.

- Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res.;30(7):1575-84.

- Farrow SC, Facchini PJ (2014) Functional diversity of 2-oxoglutarate/Fe(II)-dependent dioxygenases in plant metabolism. Front. Plant Sci. 5:524

- Frascari D, Zannoni A, Fedi S, Pii Y, Zannoni D, Pinell D (2005) Aerobic cometabolism of chloroform by butane-grown microorganisms: long-term monitoring of depletion rates and isolation of a high-performing strain. Biodegradation 16:147–158.

- Fries MR, Forney LJ, Tiedje JM (1997) Phenol- and toluene-degrading microbial populations from an aquifer in which successful trichloroethene cometabolism occurred. Appl Environ Microbiol. ;63(4):1523-30

- Fu J, Wang S (2011) Insights into auxin signaling in plant-pathogen interactions. Front. Plant Sci. 2:74
- Gambino M, Marzano V, Villa F, Vitali A, Vannini C, Landini P, Cappitelli F (2015) Effects of sublethal doses of silver nanoparticles on Bacillus subtilis planktonic and sessile cells. J Appl Microbiol.;118(5):1103-15.

- García-Guzmán G, Heil M (2014) Life histories of hosts and pathogens predict patterns in tropical fungal plant diseases. New Phytologist, - Wiley Online Library

- García-Rivera J, Chang YC., Kwon-Chung KJ, Casadevall A (2004) Cryptococcus neoformans CAP59 (or Cap59p) is involved in the extracellular trafficking of capsular glucuronoxylomannan. Eukaryot. Cell 3:385–392.

- Garelick H, Jones H, Dybowska A, Valsami-Jones E (2008) Arsenic pollution sources. Rev Environ Contam Toxicol;197:17-60.

- Garza-Gonzalez MT, Barboza Perez D, Vazquez Rodriguez A, Garcia-Gutierrez DI, Zarate X, Cantú Cardenas ME, Urraca-Botello LI, Lopez-Chuken UJ, Trevino-Torres AL, Cerino-Córdoba Fde J, Medina-Ruiz P, Villarreal-Chiu JF, Morones-Ramirez JR (2016) Metal-Induced Production of a Novel Bioadsorbent Exopolysaccharide in a Native Rhodotorula mucilaginosa from the Mexican Northeastern Region. PLoS One;11(2):e0148430.

- Ghosh S, Sar P (2013) Identification and characterization of metabolic properties of bacterial populations recovered from arsenic contaminated ground water of North East India (Assam) (2013). Water Res;47(19):6992-7005.

- Grigoriev IV, Martinez DA, Salamov A (2006) Fungal genomic annotation. Appl. Mycol. Biotechnol. 6, 123–142.

- Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, et al (2014) MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res. 42, D699–D704.

- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol;59(3):307-21.

- Guyon K, Balagué C, Roby D, Raffaele S (2014) Secretome analysis reveals effector candidates associated with broad host range necrotrophy in the fungal plant pathogen Sclerotinia sclerotiorum. BMC Genomics. 4;15:336.

- Harrison JJ, Ceri H, Turner RJ (2007) Multimetal resistance and tolerance in microbial biofilms. Nat Rev Microbiol. 5(12):928-38.

- Hartl L, Zach S, Seidl-Seiboth V. Fungal chitinases: diversity, mechanistic properties and biotechnological potential (2012) Appl Microbiol Biotechnol ;93(2):533-43.

- Hartmann T, Cairns TC, Olbermann P, Morschhauser J, Bignell EM, Krappmann S (2011) Oligopeptide transport and regulation of extracellular proteolysis are required for growth of Aspergillus fumigatus on complex substrates but not for virulence. Mol. Microbiol. 82, 917–935

- Huang J. Horizontal gene transfer in eukaryotes: the weak-link model (2013) Bioessays. ;35(10):868-75.

- Huertas MJ, Michán, C. (2013) Indispensable or toxic? The phosphate versus arsenate debate. Microb Biotechnol 6: 209–211.

- Ilyas S, Rehman A, Coelho AV, Sheehan D (2016) Proteomic analysis of an environmental isolate of Rhodotorula mucilaginosa after arsenic and cadmium challenge: Identification of a protein expression signature for heavy metal exposure. J Proteomics;141:47-56.

- Iturbe-Espinoza P, Gil-Moreno S, Lin W, Calatayud S, Palacios Ò, Capdevila M, Atrian S (2016) The Fungus Tremella mesenterica Encodes the Longest Metallothionein Currently Known: Gene, Protein and Metal Binding Characterization PLoS One;11(2):e0148651.

- Jaffe D. B, Butler J, Gnerre S, Mauceli E, Lindblad-Toh K, Mesirov J, et al. (2003) Whole-genome sequence assembly for mammalian genomes: arachne 2. Genome. Res. 13, 91–96.

- Jain R, Adhikary H, Jha S, Jha A, Kumar GN (2012) Remodulation of central carbon metabolic pathway in response to arsenite exposure in Rhodococcus sp. strain NAU-1. Microb Biotechnol.;5(6):764-72.

- Jiang Y, Leib M, Duanc L, Longhursta P (2015) Integrating phytoremediation with biomass valorisation and critical element recovery: A UK contaminated land perspective. Biomass and Bioenergy 83:28–339

- Johnston-Monje D, Raizada MN (2011) Conservation and diversity of seed associated endophytes in Zea across boundaries of evolution, ethnography and ecology. PLoS ONE 6:e20396. 10.1371

- Jurka J, Kapitonov V. V, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome. Res. 110, 462–467
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. Nucleic Acids Res. 32, D277–D280.

- Khan Z, Guelich G, Phan H, Redman R. S, Doty S. L (2012) Bacterial and yeast endophytes from poplar and willow promote growth in crop plants and grasses. ISRN Agron.

- Khan Z, Rho H, Firrincieli A, Han HS, Luna V, Masciarelli O, Kim SH, Doty SL. Growth enhancement and drought tolerance of hybrid poplar upon inoculation with endophyte consortia (2016) Curr. Plan. Biol. http://dx.doi.org/10.1016/j.cpb.2016.08.001

- Khan Z, Roman D, Kintz T, delas Alas M, Yap R, Doty S (2014) Degradation, phytoprotection and phytoremediation of phenanthrene by endophyte Pseudomonas putida, PD1. Environ Sci Technol. ;48(20):12221-8.

- Kim KT, Jeon J, Choi J, Cheong K, Song H, Choi G, Kang S, Lee YH (2016) Kingdom-Wide Analysis of Fungal Small Secreted Proteins (SSPs) Reveals their Potential Role in Host Association. Front Plant Sci. 19;7:186.

- Knoth J, Kim SH, Ettl G, Doty SL (2013) Effects of cross host species inoculation of nitrogen-fixing endophytes on growth and leaf physiology of maize. GCB Bioenery 5, 408–418.

- Knoth JL, Kim SH, Ettl GJ, Doty SL (2014) Biological nitrogen fixation and biomass accumulation within poplar clones as a result of inoculations with diazotrophic endophyte consortia. New Phytol. 201, 599–609

- Koh S, Wiles AM, Sharp JS, Naider FR, Becker JM, Stacey G (2002) An oligopeptide transporter gene family in Arabidopsis. Plant Physiol;128(1):21-9.

- Krijger JJ, Thon MR, Deising HB, Wirsel SG (2014) Compositions of fungal secretomes indicate a greater impact of phylogenetic history than lifestyle adaptation. BMC Genomics.;15:722.
- Kubicek CP, Starr TL, Glass NL (2014) Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. Annu Rev Phytopathol.;52:427-51

- Kumari N, Jagadevan S (2016) Genetic identification of arsenate reductase and arsenite oxidase in redox transformations carried out by arsenic metabolising prokaryotes - A comprehensive review Chemosphere 163:400–412.

- Landrieu I, da Costa M, De Veylder L, Dewitte F, Vandepoele K, Hassan S, Wieruszeski JM, Corellou F, Faure JD, Van Montagu M, Inzé D, Lippens G (2004) A small CDC25 dual-specificity tyrosine-phosphatase isoform in Arabidopsis thaliana. Proc Natl Acad Sci U S A. ;101(36):13380-5

- Langella F, Grawunder A, Stark R, Weist A, Merten D, Haferburg G, Büchel G, Kothe E (2014) Microbially assisted phytoremediation approaches for two multi-element contaminated sites. Environ Sci Pollut Res Int. 21(11):6845-58.

- Lebrun E, Brugna M, Baymann F, Muller D, Lièvremont D, Lett MC, Nitschke W (2003). Arsenite oxidase, an ancient bioenergetic enzyme. Mol Biol Evol. 20(5):686-93.

- Lowe T. M, Eddy S. R (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964. 10.1093/nar/25.5.0955

- Majone, Verdini R, Aulenta F, Rossetti S, Tandoi V, Kalogerakis N, Agathos S, Puig S, Zanaroli G, Fava F (2014). In situ groundwater and sediment bioremediation: barriers and perspectives at European contaminated sites. N. Biotechnol. 25;32(1):133-46

- Mak TN, Schmid M, Brzuszkiewicz E, Zeng G, Meyer R, Sfanos KS, Brinkmann V, Meyer TF, Brüggemann H (2013) Comparative genomics reveals distinct host-interacting traits of three major human-associated propionibacteria. BMC Genomics. 22;14:640.

- Marin AR, Masscheleyn PH, Patrick Jr WH (1993). Soil redox-pH stability of arsenic species and its influence on arsenic uptake by rice. Plant and Soil.;152(2):245-253

- Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC (2012) IMG: the Integrated Microbial Genomes database and comparative analysis system. Nucleic Acids Res.;40(Database issue):D115-22.

- Martinez LR., Casadevall A (2005) Specific antibody can prevent fungal biofilm formation and this effect correlates with protective efficacy. Infect. Immun. 73, 6350–6362.

- Mateos LM, Ordóñez E, Letek M, Gil JA (2006) Corynebacterium glutamicum as a model bacterium for the bioremediation of arsenic. Int Microbiol.;9(3):207-15

- Mayer L, Wilson D, Hube B (2013) Candida albicans pathogenicity mechanisms Virulence; 4(2): 119–128.

- Mei C, Flinn BS (2010) The use of beneficial microbial endophytes for plant biomass and stress tolerance improvement. Recent Pat Biotechnol.;4(1):81-95

- Meireles DA, Domingos RM, Gaiarsa JW, Ragnoni EG, Bannitz-Fernandes R, da Silva Neto JF, de Souza RF, Netto LES (2017) Functional and evolutionary characterization of Ohr proteins in eukaryotes reveals many active homologs among pathogenic fungi. Redox Biol. 12:600-609.

- Melén K, Krogh A, von H. G (2003) Reliability measures for membrane protein topology prediction algorithms. J. Mol. Biol. 327, 735–744.
- Mitchell TG, Perfect JR (1995) Cryptococcosis in the era of AIDS–100 years after the discovery of Cryptococcus neoformans. Clin. Microbiol. Rev. 8, 515–548

- Morohoshi T, Maruo T, Shirai Y et al. (2002) Accumulation of inorganic polyphosphate in phoU mutants of Escherichia coli and Synechocystis sp. strain PCC6803. Appl Environ Microb 68: 4107–4110.

- Mukhopadhyay R, Zhou Y, Rosen BP (2003) Directed evolution of a yeast arsenate reductase into a protein-tyrosine phosphatase. J Biol Chem;278(27):24476-80.

- Muller D, Médigue C, Koechler S, Barbe V, Barakat M, Talla E, Bonnefoy V, Krin E, Arsène-Ploetze F, Carapito C, Chandler M, Cournoyer B, Cruveiller S, Dossat C, Duval S, Heymann M, Leize E, Lieutaud A, Lièvremont D, Makita Y, Mangenot S, Nitschke W, Ortet P, Perdrial N, Schoepp B, Siguier P, Simeonova DD, Rouy Z, Segurens B, Turlin E, Vallenet D, Van Dorsselaer A, Weiss S, Weissenbach J, Lett MC, Danchin A, Bertin PN. (2007) A tale of two oxidation states: bacterial colonization of arsenic-rich environments. PLoS Genet.;3(4):e53.

- Nanekar S, Dhote M, Kashyap S, Singh A, Juwarkar A (2015). Microbe assisted phytoremediation of oil sludge and role of amendments: a mesocosm study. Int. J. Environ. Sci. Technol. 12:193–202

- Narasimhan K, Basheer C, Bajic VB, Swarup S (2003) Enhancement of plant-microbe interactions using a rhizosphere metabolomics-driven approach and its application in the removal of polychlorinated biphenyls. Plant Physiol. 132(1):146-53.

- Nepusz T, Sasidharan R, Paccanaro A (2010) SCPS: a fast implementation of a spectral method for detecting protein families on a genome-wide scale. BMC Bioinformatics. 9;11:120.

- Nielsen H, Engelbrecht J, Brunak S, von H. G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. 10:1–6.

- Okmen B, Doehlemann G (2014) Inside plant: biotrophic strategies to modulate host immunity and metabolism. Curr Opin Plant Biol. 20:19-25.

- Oliveira-Garcia E, Deising HB. Infection structure-specific expression of β-1,3-glucan synthase is essential for pathogenicity of Colletotrichum graminicola and evasion of β-glucan-triggered immunity in maize (2013) Plant Cell.;25(6):2356-78

- Ordonez E., Van Belle K., Roos G., De Galan S., Letek M., Gil J. A., et al. . (2009). Arsenate reductase, mycothiol, and mycoredoxin concert thiol/disulfide exchange. J. Biol. Chem. 284, 15107–15116.

- Orro A, Cappelletti M, D'Ursi P, Milanesi L, Di Canito A, Zampolli J, Collina E, Decorosi F, Viti C, Fedi S, Presentato A, Zannoni D, Di Gennaro P (2015) Genome and Phenotype Microarray Analyses of Rhodococcus sp. BCP1 and Rhodococcus opacus R7: Genetic Determinants and Metabolic Abilities with Environmental Relevance. PLoS One;10(10):e0139467

- Osborne, T.H.; Santini, J.M. Prokaryotic Aerobic Oxidation of Arsenite (2012). In The Metabolism of Arsenite; Arsenic in the Environment 5; Santini, J.M., Ward, S.A., Eds.; CRC Press: London, UK; pp. 61–72

- O'Sullivan LA, Mahenthiralingam E (2005) Biotechnological potential within the genus Burkholderia. Lett Appl Microbiol;41(1):8-11.

- Páez-Espino D. Microbial responses to environmental arsenic (2009) Biometals 22:117–130

- Pemberton TJ (2006) Identification and comparative analysis of sixteen fungal peptidyl-prolyl cis/trans isomerase repertoires. BMC Genomics. 22;7:244.
- Pesciaroli L, Petruccioli M, Fedi S, Firrincieli A, Federici F, D'Annibale A (2012). Characterization of Pleurotus ostreatus biofilms by using the calgary biofilm device. Appl Environ Microbiol.; 79(19):6083-92

- Pietrokovski S (1998) Modular organization of inteins and C-terminal autocatalytic domains. Protein Sci.;7(1):64-71.

- Pitkin J. W, Panaccione D. G, Walton J. D (1996) A putative cyclic peptide efflux pump encoded by the TOXA gene of the plant-pathogenic fungus Cochliobolus carbonum. Microbiology 142(Pt 6), 1557–1565.

- Podila G. K., Sreedasyam A., Muratet M (2009) Populus rhizosphere and the ectomycorrhizal interactome. Crit. Rev. Plant Sci. 28, 359–367.

- Prescott AG, Lloyd MD (2000) The iron(II) and 2-oxoacid-dependent dioxygenases and their role in metabolism. Nat. Prod. Rep. 17, 367–383.

- Presentato A, Piacenza E, Anikovsky M, Cappelletti M, Zannoni D, Turner RJ (2016) Rhodococcus aetherivorans BCP1 as cell factory for the production of intracellular tellurium nanorods under aerobic conditions. Microbial Cell Factories; 15(1):204.

- Price A. L, Jones N. C, Pevzner P. A (2005) De novo identification of repeat families in large genomes. Bioinformatics 21(Suppl. 1), i351–i358.

- Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. (2005) InterProScan: protein domains identifier. Nucleic Acids Res. 33, W116–W120.

- Ramage G, Mowat E, Jones B, Williams C, Lopez-Ribot J (2009) Our current understanding of fungal biofilms. Crit. Rev. Microbiol. 35, 340–355.

- Reese MG, (2001) Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome", Comput Chem 26(1),51-6.

- Rodriguez-Gabriel MA, Russell P (2005) Distinct signaling pathways respond to arsenite and reactive oxygen species in Schizosaccharomyces pombe. Eukaryot. Cell.;4:1396–1402

- Rogozin IB, Milanesi L (1997) Analysis of donor splice sites in different eukaryotic organisms. J. Mol. Evol. 45(1):50-9.

- Rozenberg A, Leese F, Weiss LC, Tollrian R (2016) Digital gene expression analysis with sample multiplexing and PCR duplicate detection: A straightforward protocol. Biotechniques 61(1): 26–32

- Rungin S, Indananda C, Suttiviriya P, Kruasuwan W, Jaemsaeng R, Thamchaipenet A (2012) Plant growth enhancing effects by a siderophore-producing endophytic streptomycete isolated from a Thai jasmine rice plant (Oryza sativa L. cv. KDML105). • Antonie Van Leeuwenhoek;102(3):463-72.

- Salamov AA, Solovyev VV (2000) Ab initio gene finding in Drosophila genomic DNA. Genome. Res. 10, 516–522.

- Saltikov CW, Newman DK. Genetic identification of a respiratory arsenate reductase (2003) Proc Natl Acad Sci U S A.;100(19):10983-8.

- Santini JM1, vanden Hoven RN (2004) Molybdenum-containing arsenite oxidase of the chemolithoautotrophic arsenite oxidizer NT-26. J Bacteriol.;186(6):1614-9.
- Saunders DG, Win J, Cano LM, Szabo LJ, Kamoun S, Raffaele S (2012) Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. PLoS One. ;7(1):e29847.

- Shagol CC, Krishnamoorthy R, Kim K, Sundaram S, Sa T (2014) Arsenic-tolerant plant-growth-promoting bacteria isolated from arsenic-polluted soils in South Korea. Environ Sci Pollut Res Int. 21(15):9356-65

- Shykoff J and Kaltz O (1997) Effects of the Anther Smut Fungus Microbotryum violaceum on Host Life-History Patterns in Silene latifolia (Caryophyllaceae). International Journal of Plant Science Vol. 158, pp. 164-171

- Singh N, Ma LQ, Srivastava M, Rathinasabapathi B (2006). Metabolic adaptations to arsenic-induced oxidative stress in Pterisvittata L and Pterisensiformis L. Plant Science, 2: 274-282

- Smit A. F. A, Hubley R, Green P (2010) Repeat Masker Open-3.0. Available online at: http://www.repeatmasker.org

- Soderlund C, Nelson W, Shoemaker A, Paterson A (2006) SyMAP: A system for discovering and viewing syntenic regions of FPC maps. Genome Res.;16(9):1159-68.

- Sousa T, Branco R, Piedade AP, Morais PV. Hyper Accumulation of Arsenic in Mutants of Ochrobactrum tritici Silenced for Arsenite Efflux Pumps (2015). PLoS One. 1;10(7):e0131317.

- Stolzer M, Lai H, Xu M, Sathaye D, Vernot B, Durand D (2012) Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. Bioinformatics. 15;28(18):i409-i415.

- Sultana M, Vogler S, Zargar K, Schmidt AC, Saltikov C, Seifert J, Schlömann M. (2012) New clusters of arsenite oxidase and unusual bacterial groups in enrichments from arsenic-contaminated soil. Arch Microbiol.;194(7):623-35.

- Sure S, Ackland ML, Gaur A, Gupta P, Adholeya A, Kochar M. (2016) Probing Synechocystis-Arsenic Interactions through Extracellular Nanowires. Front Microbiol. 19;7:1134.

- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0 (2013) Mol Biol Evol.;30(12):2725-9.

- Tavares S, Ramos AP, Pires AS, Azinheira HG, Caldeirinha P, Link T, Abranches R, Silva Mdo C, Voegele RT, Loureiro J, Talhinhas P (2014) Genome size analyses of Pucciniales reveal the largest fungal genomes. Front Plant Sci. 26;5:422.

- Tchounwou PB, Yedjou CG, Patlolla AK, Sutton DJ (2012) Heavy Metals Toxicity and the Environment. EXS. 101: 133–164.

- Teitzel GM, Parsek MR. (2003) Heavy metal resistance of biofilm and planktonic Pseudomonas aeruginosa. Appl Environ Microbiol.;69(4):2313-20.

- Thijs S, Weyens N, Sillen W, Gkorezis P, Carleer R, Vangronsveld J (2014) Potential for plant growth promotion by a consortium of stress-tolerant 2,4-dinitrotoluene-degrading bacteria: isolation and characterization of a military soil. Microb Biotechnol.;7(4):294-306.

- Tobin KM, McGrath JW, Mullan A, Quinn JP, O'Connor KE (2007) Polyphosphate accumulation by Pseudomonas putida CA-3 and other medium-chain-length polyhydroxyalkanoate-accumulating bacteria under aerobic growth conditions. Appl Environ Microbiol. 73(4):1383-7.

- Touchon M, Arneodo A, d'Aubenton-Carafa Y, Thermes C (2004) Transcription-coupled and splicing-coupled strand asymmetries in eukaryotic genomes Nucleic Acids Res.; ;32(17):4969-78

- Tovar-Herrera OE, Batista-García RA, Sánchez-Carbente Mdel R, Iracheta-Cárdenas MM, Arévalo-Niño K, Folch-Mallol JL (2015) A novel expansin protein from the white-rot fungus Schizophyllum commune. PLoS One. 10(3):e0122296 .

- Tripathy BC, Oelmüller R (2012) Reactive oxygen species generation and signaling in plants. Plant Signal Behav. 7(12):1621-33.

- Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, et al (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray) Science 313, 1596–1604.

- Vaknin Y, Shadkchan Y, Levdansky E, Morozov M, Romano J, Osherov N (2012) The three Aspergillus fumigatus CFEM-domain GPI-anchored proteins (CfmA-C) affect cell-wall stability but do not play a role in fungal virulence. Fungal Genet Biol. 63:55-64.

- Wang QM, Yurkov AM, Göker M, Lumbsch HT, Leavitt SD, Groenewald M, Theelen B, Liu XZ, Boekhout T, Bai FY (2015) Phylogenetic classification of yeasts and related taxa within Pucciniomycotina. Stud Mycol.;81:149-89.

- Wang X, Chapman KD (2013) Lipid signaling in plants. Front Plant Sci. 27;4:216.

- Wang Y, Coleman-Derr D, Chen G, Gu YQ. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 1;43(W1):W78-84.

- Wolfe-Simon F, Switzer Blum J, Kulp TR, Gordon GW, Hoeft SE, Pett-Ridge J, Stolz JF, Webb SM, Weber PK, Davies PC, Anbar AD, Oremland RS (2012) A bacterium that can grow by using arsenic instead of phosphorus. Science.;332(6034):1163-6.

- Xin G, Glawe D, Doty SL (2009) Characterization of three endophytic, indole-3-acetic acid-producing yeasts occurring in Populus trees. Mycol Res. 113(Pt 9):973-80.

- Xu P, Bura R, Doty SL (2011) Genetic analysis of D-xylose metabolism by endophytic yeast strains of Rhodotorula graminis and Rhodotorula mucilaginosa. Genet. Mol. Biol. 34, 471–478.

- Yergeau Etienne, Sanschagrin Sylvie, Maynard Christine, St-Arnaud Marc and W Greer Charles (2014). Microbial expression profiles in the rhizosphere of willows depend on soil contamination. ISME Journal 8: 344–358.

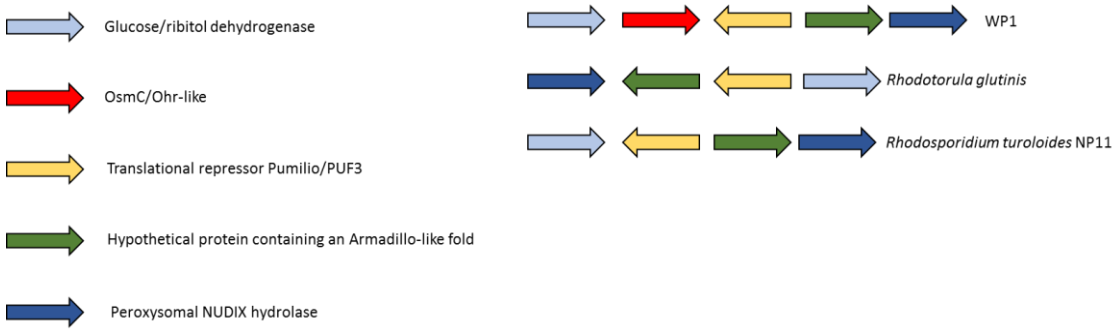
- Zhang ZN, Wu QY, Zhang GZ, Zhu YY, Murphy RW, Liu Z, Zou CG (2015) Systematic analyses reveal uniqueness and origin of the CFEM domain in fungi. Sci Rep.; 10;5:13032.

- Zhao Z. Zhang Y, Liu X, Zhang X, Liu S, Yu X (2013) A role for a dioxygenase in auxin metabolism and reproductive development in rice. Dev. Cell 27, 113–122.

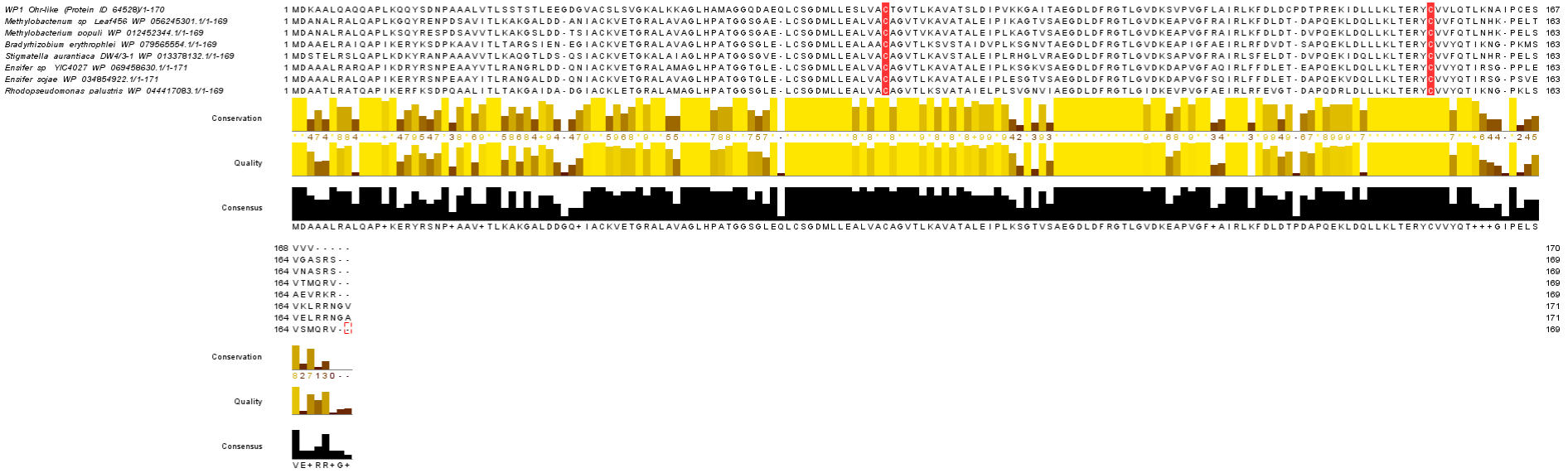
- Zhu Q, Kosoy M, Dittmar K. (2014) HGTector: an automated method facilitating genome-wide discovery of putative horizontal gene transfers. BMC Genomics. 26;15:717.

Supplementary Material, Figure 1.

A



B



Supplementary material, Figure 1. A) WP1 Chromosome region containing the Ohr-like coding sequence (red arrow) compared against *R. glutinis* (scaffold_63:56188-75827) and *R. turoloides* NP11 (scaffold_4:899345-910798). B) Multiple sequence alignment of WP1 Ohr-like against selected species from *Rhizobiales*

Supplementary material, Table 1. List of Ohr-like proteins retrieved from prokaryotes

Sequences ID	Protein Hit[1]	Genus	Taxonomic affiliation
WP_056245301.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_012452344.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_079565554.1	peroxiredoxin	Bradyrhizobium	Rhizobiales
WP_017482753.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_013378132.1	peroxiredoxin	Stigmatella	Myxococcales
WP_069458630.1	peroxiredoxin	Ensifer	Rhizobiales
WP_048451243.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_071015752.1	peroxiredoxin	Ensifer	Rhizobiales
WP_066920200.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_048446006.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_037019867.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_036356424.1	peroxiredoxin	Microvirga	Rhizobiales
WP_056202198.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_056114401.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_012332059.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_060846081.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_068289095.1	peroxiredoxin	Labrys	Rhizobiales
WP_075011427.1	peroxiredoxin	Stigmatella	Myxococcales
WP_048437211.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_027315826.1	peroxiredoxin	Microvirga	Rhizobiales
WP_034854922.1	peroxiredoxin	Ensifer	Rhizobiales
OJW63718.1	peroxiredoxin	Afipia	Rhizobiales

WP_009763614.1	peroxiredoxin	Microvirga	Rhizobiales
WP_068884422.1	peroxiredoxin	Paramesorhizobium	Rhizobiales
ANY79716.1	peroxiredoxin	Microvirga	Rhizobiales
WP_044417083.1	peroxiredoxin	Rhodopseudomonas	Rhizobiales
WP_047187307.1	peroxiredoxin	Microvirga	Rhizobiales
WP_046829653.1	peroxiredoxin	Afipia	Rhizobiales
WP_011580036.1	peroxiredoxin	Chelativorans	Rhizobiales
WP_048464417.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_015927852.1	peroxiredoxin	Methylobacterium	Rhizobiales
OJY00197.1	peroxiredoxin	Rhizobiales	Rhizobiales
WP_008599137.1	peroxiredoxin	Nitratireductor	Rhizobiales
WP_056300093.1	peroxiredoxin	Afipia	Rhizobiales
WP_056215633.1	peroxiredoxin	Aeromicrobium	Actinomycetales
WP_056913952.1	peroxiredoxin	Pseudolabrys	Rhizobiales
WP_046862089.1	peroxiredoxin	Microvirga	Rhizobiales
WP_086087147.1	peroxiredoxin	Pseudorhodoplanes	Rhizobiales
WP_028347731.1	peroxiredoxin	Bradyrhizobium	Rhizobiales
WP_027232068.1	peroxiredoxin	Phyllobacterium	Rhizobiales
WP_039894376.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_055730754.1	peroxiredoxin	Bosea	Rhizobiales
WP_038357770.1	peroxiredoxin	Bosea	Rhizobiales
WP_056280750.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_040984606.1	peroxiredoxin	Mesorhizobium	Rhizobiales
WP_056090806.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_071924337.1	peroxiredoxin	Chelatococcus	Rhizobiales
WP_069689885.1	peroxiredoxin	Bosea	Rhizobiales
OJV06493.1	peroxiredoxin	Bosea	Rhizobiales
WP_007559990.1	peroxiredoxin	Methylobacterium	Rhizobiales

WP_050732267.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_047572555.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_027171501.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_009734532.1	peroxiredoxin	Bradyrhizobiaceae	Rhizobiales
WP_018045809.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_056163792.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_056143139.1	peroxiredoxin	Methylobacterium	Rhizobiales
WP_056708382.1	peroxiredoxin	Bosea	Rhizobiales
WP_062314438.1	peroxiredoxin	Bradyrhizobium	Rhizobiales
WP_024516499.1	peroxiredoxin	Bradyrhizobium	Rhizobiales
WP_074820566.1	peroxiredoxin	Bradyrhizobium	Rhizobiales
WP_066618856.1	peroxiredoxin	Bosea	Rhizobiales

1. *Each proteins reported in this table share a sequence similarity of 58-62 % and a blast score of 180 – 240 against the Ohr-like protein of WP1*