UPPSALA
UNIVERSITET

UPTEC X 21041
Examensarbete 30 hp
September 2021

The evolutionary history of phosphorus transporters in arbuscular mycorrhizal fungi

Lovisa Lundberg

UPPSALA
UNIVERSITET

Teknisk- naturvetenskaplig fakultet
UTH-enheten

Besöksadress:
Ångströmlaboratoriet
Lägerhyddsvägen 1
Hus 4, Plan 0

Postadress:
Box 536
751 21 Uppsala

Telefon:
018 – 471 30 03

Telefax:
018 – 471 30 00

Hemsida:
http://www.teknat.uu.se/student

Abstract

The evolutionary history of phosphorus transporters in arbuscular mycorrhizal fungi

Lovisa Lundberg

Arbuscular mycorrhizal (AM) fungi are obligate biotrophs that form symbiosis with plants by colonizing their roots. The fungus supplies the host plant with various nutrients, including phosphorus. Phosphorus is crucial for the development of plants and is hard to acquire in soil since it can be scarce and has a slow motility. The fungus utilizes its long hyphal threads to contact more soil to obtain phosphorus and transport it back to the plant. It does so with its use of different phosphorus transporters (PTs) located in its membranes. Here we have done a phylogenetic analysis of different PTs from a variety of fungi from different phyla together with plants and new sequence data from AM fungi. In total, 955 genomes were screened, 26 of which belong to AM fungi. This work resulted in a database of 1351 PT sequences, 907 from fungi (243 from AM) and 444 from plants, and two phylogenetic trees to visualize the data. One phylogeny was made of the branch of the PT Pho87 which was selected for building a Hidden Markov model, which can facilitate future searches of PTs.

Handledare: Anna Rosling, Maliheh Mehrshad
Ämnesgranskare: Martin Ryberg
Examinator: Pascal Milesi
ISSN: 1401-2138, UPTEC X 21041
Tryckt av: Uppsala

Evolutionär kartläggning: Svampar som gör växter mer livskraftiga genom transport av fosfor

Fosfor är ett ämne som är livsviktigt för tillväxt och överlevnad för alla organismer. Det kan dock vara svårt att komma åt för växter då det rör sig långsamt genom jord, vilket utgör ett problem. Genom evolutionen har växter anpassat sig tillsammans med en viss sorts jordlevande svampar för att lösa detta problem. Tillsammans samarbetar växten och svampen för att komma åt alla de näringsämnen som de behöver. Svampen sprider ut sina långa trådar under marken för att kunna täcka så mycket yta som möjligt. Med trådarna kan de sedan ta upp olika ämnen, däribland fosfor men även vatten, och transportera tillbaka det till växtens rötter. I utbyte får svampen kol, fetter och andra ämnen som växten bygger upp men som svampen inte kan eller har svårt att få tag i själv men fortfarande behöver. Växter som samarbetar med denna sorts svamp har högre tolerans mot extern stress, såsom torka, och är därför mer livskraftiga. Vartefter klimatet förändras och behovet av mat ökar i världen så är detta något som man vill kunna utnyttja genom att tillföra dessa svampar i jordbruk genom biogödsel.

Svampen använder sig av olika fosfortransportörer, som är ett sorts protein, för att förflytta fosfor över sina membran och sedan transportera det vidare till växten. Dessa transportörer är aktiva under olika förhållanden, bland annat beroende på hur mycket fosfor som finns tillgängligt. Genom att titta på arvsmassan som kodar dessa transportörer och jämföra den mellan olika svamparter och växter är det möjligt att se hur dessa hänger ihop evolutionärt. Detta visualiseras i form av träd där varje löv ytterst på trädet består av en gen. Grenarna mellan löven visar hur de olika generna är relaterade till varandra. Där en förgrening uppstår har en gen utvecklats och blivit två. Vi har i denna studie byggt sådana träd av gener för fosfortransportörer från olika växter och svampar. Genom att analysera dessa har vi fått en bättre uppfattning av hur den evolutionära historien ser ut. Träden kan delas in i fyra undergrupper, där en består av gener från växter och tre främst består av gener från svampar, även om några växtgener även ingår här. Den svampart som ingår i en viss grupp har gener som kodar för den fosfortransportör som gruppen utgör. En och samma svampart kan finnas i flera grupper, vilket innebär att den har gener för flera olika transportörer. Den kan också finnas med flera gånger i samma grupp, vilket innebär att arten har fler än en genkopia som kodar för samma transportör. De svamparter som är nära relaterade till varandra evolutionärt har också gener för fosfortransportörer som ligger intill varandra i trädet. Detta innebär att grupper av fosfortransportörer har utvecklats i takt med att arterna har gjort det.

Table of contents

Abbreviations

AM	arbuscular mycorrhiza
AQP	aquaporin
BAS	branched absorbing structures
HGT	horizontal gene transfer
HMM	Hidden Markov model
FACS	fluorescence-activated cell sorting
MAT-loci	mating-type loci
MDA	multiple displacement amplification
P	phosphorus
Pi	inorganic phosphorus
PCR	polymerase chain reaction
PT	phosphorus transporter
VTC	vacuolar transporter chaperon
WGA	whole genome amplification

1 Introduction

1.1 The phosphorus problem

Phosphorus (P) is a macronutrient needed to construct nucleic acids, phospholipids and more. All living organisms including plants need it for survival and growth. A limited supply of P harms plant development and slows down growth. Plants preferentially absorb P in the form of inorganic phosphate (Pi). This molecule has a low solubility and motility in soil, leaving the surroundings of the roots depleted since plants have a higher demand for Pi than the surrounding soil can supply. The depleted area around the roots is referred to as the depletion zone (Ferrol *et al.* 2018). P is especially scarce in ancient soils which are nutrient deprived due to age. Plants have over evolutionary time developed different strategies to cope with this problem. Apart from the concentration of P in soil, root characteristics are the main limiters of P uptake, including the area of the roots available for uptake and the distribution of the roots in the soil. This makes generating more root surface to contact more soil important for P acquisition. This strategy is referred to as scavenging and includes fast development of roots and root hairs, extensive branching, local extension of roots in P rich areas and symbiosis with mycorrhizal fungi. There are several mycorrhizal types including arbuscular, ericoid and ectomycorrhizal. They all form long hyphal branches that are used to transport P to the plant, thus covering more soil than the roots could (Lambers *et al.* 2008).

Another strategy, referred to as mining, releases P from other forms, which is useful since phosphate is quickly absorbed into or onto soil particles. Mining is performed by the release of exudates that solubilize or hydrolyse nutrients, increasing their mobility, and/or change their form into one that the roots are capable of absorbing. For instance, some plant species form cluster roots which produce carboxylates and phosphatases. Carboxylates release P in alkaline soils by locally reducing the pH or by replacing P bound to calcium. In acid soils they replace P bound to iron or aluminium. Phosphatases release P from organic sources. This is mainly used in ancient soils where there is little P left in solution, whereas younger soils have enough P to maintain symbiosis through only scavenging (Lambers *et al.* 2008).

Different mycorrhizal fungi can access different chemical forms of P. Arbuscular mycorrhizal (AM) fungi mainly use scavenging and takes up Pi, whereas ericoid and ectomycorrhizal also can utilize insoluble organic forms with the use of phosphatases, and are thus able to use both strategies (Lambers *et al.* 2008).

1.2 Arbuscular mycorrhizal fungi

AM fungi are soil-inhabiting obligate biotrophs belonging to the phylum Glomeromycota. They form symbiosis with around 80 percent of all terrestrial plant species by penetrating the inner root tissue, known as cortical tissue, and forming treelike structures called arbuscles inside the cells (Plassard *et al.* 2019). The arbuscules are connected inside the root by the intraradical mycelium, that also continues outside of the root forming the extraradical mycelium. The extraradical mycelium branches out from the root, colonizing the surrounding soil. The mycelia can have a range that is 10 to 40-fold more extensive than the roots, depending on the host plant (Ferrol *et al.* 2018). The extraradical mycelium consists of runner hyphae and branched absorbing structures (BAS). Runner hyphae are long, relatively thick, straight branches that are connected to the intraradical mycelium in the root. The runner hyphae are responsible for the size of the colony. Connected to the runner hyphae are the BAS, which consists of thinner hyphae that are morphologically similar to the arbuscles in their tree-like shape. Daughter spores of the fungus form and mature on the BAS (Kameoka *et al.* 2019).

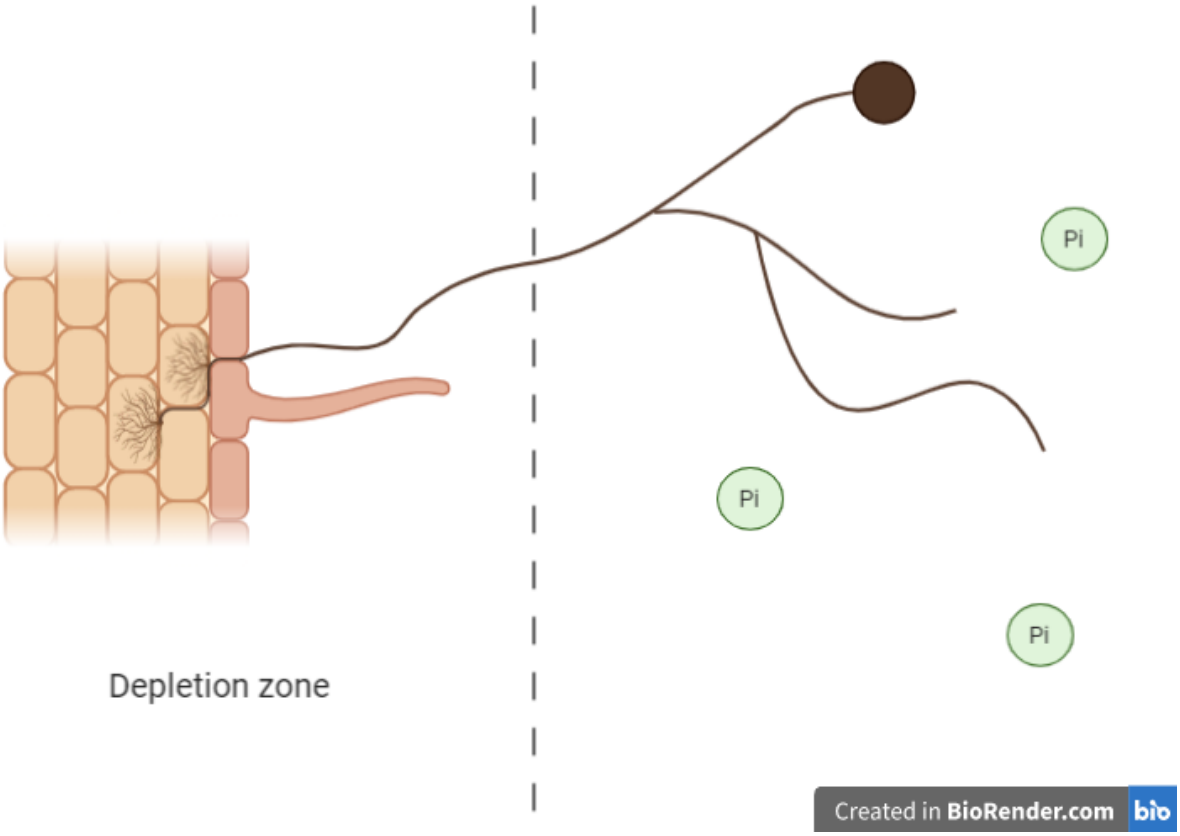


Figure 1. Schematic figure of the plant root and the hyphae of AM fungi. The soil surrounding the roots that are accessible for the root hairs is depleted of Pi. The fungus penetrated the cortical root cells of the plant to form arbuscles connected with intraradical hyphae. The extraradical hyphae covers a lot more soil with its long thin threads where it can access Pi and other nutrients. At the end of the hyphae the spores mature to later bud off. Created with BioRender.com.

1.3 Nuclear dynamics and spores

AM fungi have aseptate hyphae which allow nuclei to flow freely through the continuous mycelium. AM fungi are thus multinucleate, with nuclei moving bidirectionally across the entire fungus (Kokkoris *et al.* 2021). The nuclei move in an erratic manner where they change direction, pass each other and remain immobile, sometimes for several minutes (Jany and Pawlowska 2010). Most studied AM fungi have nuclei with a uniform genotype, so called homokaryons. But a few instances of AM fungi with nuclei from two different genotypes have been reported. The two different nucleotypes cohabit one fungus which is referred to as dikaryon. Recently there has been evidence that AM fungi have mating-type loci (MAT-loci) which are required for sexual propagation in other fungi. The MAT-loci determine sexual identity and are necessary for sexual interaction to occur between compatible strains (Kokkoris *et al.* 2021). However, AM fungi are known only to propagate through asexual spores containing hundreds of nuclei. Glomeromycota is one of the most long lived multicellular asexual linages. One issue of asexual propagation is the irreversible accumulation of deleterious mutations, according to Muller's ratchet. How AM fungi have been able to avoid this for so long is unclear. One suggestion is the AM fungus ability to harbor multiple coexisting gene copies which can act as a buffer for deleterious mutations. Another suggestion is the creation of dikaryons through fusion of hyphae (Jany and Pawlowska 2010). Kokkoris *et al.* (2021) showed that stable dikaryons are unlikely over time since differences in fitness between the two genotypes rapidly leads to allele fixation in the nuclei population which in turn leads to the creation of a homokaryon instead. Even if the two nucleotypes have the same fitness and competitive ability bottlenecks during spore formation can occur, where one nucleotype has a higher abundance of nuclei in the initial spore, limiting coexistence (Kokkoris *et al.* 2021).

Jany and Pawloska (2010) studied the movement of nuclei during spore formation in *Glomus etunicatum.* The sporogenesis was initiated by swelling of hyphae, either at the terminal or intercalary. Nuclei moved bidirectionally in and out of the spore from the start of formation until the pigmentation of the spore wall was finished. This indicates that nuclei can move in and out of the spore through cytoplasmic streaming until the spore is fully mature and closed off from the hyphae. The study showed that the influx of nuclei from the mycelium was larger than the efflux generating a net accumulation of nuclei in the spore (Jany and Pawlowska 2010). Marleau *et al.* (2011) did not find any correlation between the number of nuclei in a spore and the ability to germinate. Spores that failed to germinate had the same number of nuclei as other spores that germinated successfully. They did however find a correlation of the number of spores and the spore size for four different AM fungi taxa, where larger spores contain more nuclei (Marleau *et al.* 2011). The spores germinate by growing thin hyphal structures called germ tubes that reach for the host. AM fungi spores can reinitiate germination if it fails due to damage of the germ tube or if the host is unreachable. This is due to their multinucleate nature since only the nuclei that have migrated through the germ tube

are lost while the rest remain in the spore. The spore can keep trying to germinate until it depletes all of it's nuclei (Jany and Pawlowska 2010).

By introducing the DNA polymerase inhibitor aphidicolin to the growth medium of *Glomus irregulare* mitosis is blocked in the early S phase giving the ability to study nuclear division *in vitro*. While aphidicolin did not affect the spore size it did reduce the number of nuclei in the new spores with 54.86% in addition to reducing the number of spores 14-fold compared to the control. This indicates that there is at least one mitotic cycle of the nuclei that have migrated from the hyphae into the spore. The hyphal growth was not affected by aphidicolin (Marleau *et al.* 2011).

1.4 The symbiosis

The origin of the AM symbiosis is dated back to around 450 million years ago and is considered to have an essential role in the adaptation of vascular plants on land. Because of the coevolution with AM fungi the plants could manage the limited Pi supply and other potential difficulties that hindered them from colonizing land (Ferrol *et al.* 2018).

1.4.1 The benefits

There are several host benefits that have been connected to the symbiosis with AM fungi. The host has a greater resistance to different stresses, both biotic and abiotic, together with an overall improved mineral nutrition. In addition to Pi the host receives other molecular macronutrients improving nutrition such as nitrogen, zinc, copper and water, from beyond the depletion zone due to the extraradical fungal mycelia (Venice *et al.* 2019). It also works as a protector against minerals that are toxic in high concentrations, e.g. zinc and copper. AM fungi can alleviate this issue with the use of their homeostatic control. Due to this AM fungi have also been suggested to have a key role in cleaning up polluted soil, phytoremediation (Ferrol *et al.* 2016).

In exchange the fungus receives fixed carbon, lipids and sugar from the plant (Ferrol *et al.* 2018). AM fungi can manipulate fatty acids through elongation and desaturation (Kameoka *et al.* 2019) but they lack the genes to synthesize them which confirms their obligate biotrophic nature (Venice *et al.* 2019). This suggests that all fatty acids in AM fungi are obtained from their host plants (Kameoka *et al.* 2019).

Because of the benefits of the host plant AM fungi have been discussed every now and then as a biofertilizer. The idea is to inoculate crop fields with AM fungi to get a greater, less stress affected yield with less input. However, the outcomes vary. There have been reports of positive, negative and no results. Positive or no results are the majority, especially for studies which include multiple taxa. One reason for the varying results is that field and greenhouse studies are grouped together. There are a lot of different factors playing a part in the outcome of inoculation. Because of this it is hard to detect meaningful results (Hart *et al.* 2017). Some

studies show a positive outcome with local inoculants but other factors apart from the actual inoculant and plant affect the result, such as soil fertility and timing. One reason for inflated positive results is the lack of proper controls. In general, in nature, plants without AM symbiosis are unlikely and since the controls are grown in nurseries before they are transplanted onto the field they are without the symbiosis during crucial developmental stages (Hart *et al.* 2017).

1.4.2 Root colonization

Even though AM fungi cannot complete their life cycle without a host the spores can still germinate and grow but with limited hyphal growth (Akiyama *et al.* 2005). Both symbionts secret signalling molecules which initiates the symbiosis (Ferrol *et al.* 2018). Plants secrete strigolactones which are a group of sesquiterpene lactones. In response to strigolactones the germinated spore starts to branch extensively in order to reach the root (Akiyama *et al.* 2005) and produce "Myc factors", which are a mixture of chito- and lipooligosaccharides (Ferrol *et al.* 2018). Once the plant recognizes the "Myc factors" a series of cellular events and expression patterns start which leads to the beginning of formation of arbuscules in the cortical cells of the root (Ferrol *et al.* 2018). The exchange of molecules can then occur at the symbiotic interface, referred to as the periarbuscular space. The interface includes the fungal cell wall and plasma membrane together with the periarbuscular membrane which is a plant-derived plasma membrane that encapsules the arbuscules separating them from the cytoplasm of the host cell (Ferrol *et al.* 2018). In addition to creating the new membrane the host cell undergoes other structural changes to accommodate the fungus, such as, relocation of the nucleus from the periphery to the centre, fragmentation of the vacuole and morphology changes of the plastids to avoid the build-up of starch (Bonfante and Genre 2008).

However, it should be mentioned that AM fungi does not necessarily colonize roots. AM fungi can develop a symbiosis with ferns and other plants that lack root hairs using their rhizoids instead, which have a corresponding function. They can also colonize and grow in adventitious roots, which comes from shoot cells. For example, in the genus *Psilotum*, commonly known as whisk ferns, AM fungi colonize rhizomes. Studies have shown that also vascular plants can have colonized adventitious roots, this is the case for onion and leek, among others. AM fungi preferably colonize the cortical tissue, with that said, not all cell types of a host organ can be colonized, this applies for differentiating cells, the endodermis and vascular tissue for example (Bonfante and Genre 2008).

1.4.3 Uptake of phosphorus

Mycorrhizal plants have two different ways to obtain P, the direct and the mycorrhizal pathway. The direct pathway is the plant's own strategy for acquiring P and is located on the epidermal cells of the roots and root hairs. Studies using radioactively labelled Pi have traced which pathways are used for Pi acquisition (Ferrol *et al.* 2018). It has been shown that a plant

in symbiosis with AM fungi changes its strategy for Pi uptake, favouring the mycorrhizal pathway while downregulating the uptake from the direct pathway. The fungus can then be the main contributor of P for the plant. These studies have also shown that the amount of Pi received by the plant correlates to the amount of extraradical hyphae and not with the amount of colonized cortical cells. The ratio of Pi being obtained from the fungal pathway and the direct pathway varies depending on the host and P availability. The direct pathway is in most cases used in some sense since not all root cells are colonized, leaving the direct pathway open. Studies using plants grown in split-root systems, where the root system is split in half and isolated from each other where one half is in symbiosis and the other is not, have shown that Pi acquisition from the direct pathway is down regulated locally but not in the parts of the root without fungal colonization. However, if the soil surrounding the roots are rich in Pi so that the direct pathway through root hairs and epidermal cells can be used it might have a negative impact on the development of symbiosis with AM fungi. This is most likely dependent on the host plant since it regulates the Pi uptake. Although, expression of fungal Pi transporters correlates to the levels of external Pi in the soil, being up regulated at low Pi and down regulated at high Pi, which also might play a part in the favouring of the direct pathway at high Pi levels (Ferrol *et al.* 2018).

1.4.4 Phosphorus transporters

The proteins that mediate the transport of Pi over the membranes are symporters, meaning that they cotransport Pi against its concentration gradient together with another ion that is transported along its gradient. The most studied symporters are the ones that are coupled together with H^+, e. g. PT1, PT2. These are high-affinity H^+/Pi symporters fuelled by a H^+ -ATPase and are active when the concentration of Pi is low (Plassard *et al.* 2019). H+-ATPases are needed for transport of Pi across the membrane because metabolic energy is required since Pi is a negatively charged ion that is transported across the plasma membrane into the cell which has a higher negative electric potential than the soil (Ferrol *et al.* 2018).

However, if the soil is alkaline there is a lack of protons to drive the H^+ coupled symporters. To secure the acquisition of Pi the fungus uses low-affinity Na^+/Pi symporters which are active at higher concentrations of Pi. There are two known symporters of this kind in yeast, Pho87 and its paralog Pho90. In addition to these there is another low-affinity transport protein, Pho91, in the vacuolar membrane that transports Pi from the vacuole to the cytosol in the intraradical mycelium. All of these have an SPX-domain that interacts with cytosolic 5-inositol-P7 which is an indicator of the internal P levels in *Saccharomyces cerevisiae* (Plassard *et al.* 2019).

One high-affinity phosphorus transporter (PT) that is a homolog to the *S. cerevisiae* transporter Pho84 has been found in several AM fungi, such as *Glomus versifone, Rhizophagus irregularis* and *Gigaspora margarita* among others. These transporters were found to be expressed not only in the extraradical mycelium as expected but also in the

intraradical mycelium towards the periarbuscular interface. This is surprising since the primary function of these transporters is the uptake of Pi from soil. In the intraradical mycelium there is expected to be an efflux of Pi and not an influx. This suggests that there might be a secondary function to these transporters where they control the flux of Pi into the periarbuscular interface, therefore controlling the amount of Pi that is delivered to the plant. It has been shown in *G. margarita* that the homolog of Pho84 has a crucial part of the maintenance of symbiosis since silencing of the gene causes harm to the development and growth of the fungus and its arbuscules (Ferrol *et al.* 2018).

As previously mentioned, Pho91 is a part of the transport from the intraradical mycelium to the fungal cytosol where it can be transported into the periarbuscular space. This process starts once the Pi is in the cytosol of the extraradical mycelium where it is converted into ATP in the mitochondrion. After that the Pi residues are joined together with high-energy bonds in the vacuole of the fungal branch to create polyphosphate which are linear polymers that can be three to thousands of residues in length. It has been proposed that this occurs in the vacuolar transporter chaperon (VTC) complex since orthologs of this gene are up-regulated when the extraradical mycelium is exposed to Pi. To accommodate the negative charge of polyphosphate there is an equal accumulation of the inorganic cations Na^{+}, K^{+}, $Ca2^{+}$ and MG^{2+} in the extraradical mycelium (Ferrol *et al.* 2018). The polyphosphate is then translocated through the fungal branch via the vacuole. There are two possible ways for the fungus to do this, either through cytoplasmic streaming or along a motile tubular vacuole system. Both pathways can be used simultaneously. It is probable that the cytoplasmic streaming occurs through water flow through aquaporins (AQPs) in the fungal plasma membrane. Once the polyphosphate is translocated to the intraradical mycelium it is hydrolysed by fungal polyphosphatases which leads to a high concentration of Pi which in turn leads to efflux of Pi into the fungal cytosol, likely via Pho91. After that the Pi needs to pass the fungal membrane into the periarbuscular space. There are two ways the fungus can achieve this, either by a yet unidentified efflux protein that transports Pi over the membrane or by a VTC complex that polymerizes Pi back into polyphosphate which will later be broken down into Pi by a plant acid phosphatase. This leaves free Pi in the periarbuscular space which the plant can access through a mycorrhiza-inducible Pi transporter (Ferrol *et al.* 2018), see figure 2.

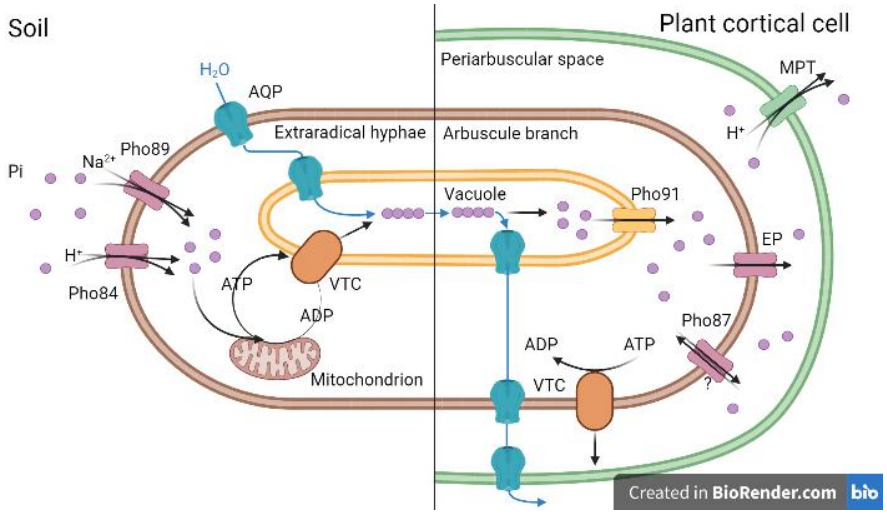


Figure 2. Simplified schematic figure of the transport of Pi from soil to the plant cortical cell. Pi is transported into the extraradical hyphae through PTs, Pho84 and Pho89 are shown here. After that it is probable that the Pi is turned into ATP in the mitochondrion to be polymerized into polyphosphate in the vacuole transport chaperon (VTC) complex. The polyphosphate is then transported through the hyphae likely by a water flow created by aquaporins (AQP). The polyphosphate is then hydrolyzed and transported into the arbuscle branch likely via Pho91. From the arbuscle the Pi is transported into the periarbuscular space probably through an unknown efflux protein (EP), another VTC complex that polymerizes the Pi or the PT Pho87. The direction of the Pi flow of Pho87 is unknown. It could act as both in- and efflux. In the periarbusclar space it is likely that the Pi is transported into the plant cell by a mycorrhiza-inducible transporter (MPT). Created with BioRender.com.

1.5 Aim of this study

The aim of this study is to characterize the diversity of phosphorus transporters across different phyla of fungi with focus on different subgroups, especially the Pho87 complex. This is done in order to get a better understanding of phosphorus pathways in fungi. A dataset containing different phosphorus transporters is collected to give an overall picture of the divergence. The set can be used in future studies to construct models for annotation purposes.

2 Methods

2.1 Data collection

2.1.1 AM fungi genome data

Due to the cryptic life cycle of AM fungi it cannot be grown in axenic cultures. Single cell techniques are not optimal for multicellular organisms even though they yield a better result than metagenomic approaches (Montoliu-Nerin *et al.* 2020). Montoliu-Nerin *et al.* (2020) developed and compared three workflows to assemble fungal genomes to accomodate for these difficulties. All three workflows are based on the isolation and sequencing of spores

from trap culture. For this project, the assemblies from workflow three are used since they yielded the highest BUSCO derived completeness result with 89% completeness (Montoliu-Nerin *et al.* 2020).

In brief to generate these sequences the genomic material was obtained from nuclei that were collected from crushed spores of the fungal strain *Claroideoglomus claroideum/C. luteum* from a trap culture from the INVAM pot culture collection. The presence of fungal DNA was confirmed using DNA stain, fluorescence-activated cell sorting (FACS) and polymerase chain reaction (PCR) with primers specific to both fungi and bacteria. The genomes of each separate nucleus were then amplified using whole genome amplification (WGA) and multiple displacement amplification (MDA). After screening the amplified DNA using specific barcodes to ensure that there were only fungal DNA it was sequences with Illumina HiSeqX. The reads from all nuclei were combined and then normalized using SPADES. The assembly was scaffolded using Satsuma based on Nanopore data of pooled nuclei assembled with Canu (Montoliu-Nerin *et al.* 2020). In the same way 26 AM genomes were sequenced and assembled (Montoliu-Nerin *et al.* 2021), see supplementary table S1 for full species names.

2.1.2 Reference sequences used for analysis

We also collected the fungal PT reference set consisting of a total of 323 sequences from Plassard *et al.* (2019) and Venice *et al.* (2019) publications. The data from Plassard *et al.* (2019) were retrieved from Mycocosm database based on their functional annotation. In total the set contains 231 amino acid PT sequences from different fungi of the phyla Ascomycota and Basidiomycota, together with one sequence of the AM fungi *Rhizophagus irregularis* (Plassard *et al.* 2019). The data from Venice *et al.* (2019) contains 92 amino acid PT sequences from the phyla Ascomycota, Basidiomycota, Glomeromycota, Murcoromycota and Morteirellomycota. The sequences were obtained from JGI genome portal and NCBI, except the sequence for *G. margarita* which they obtained from their own study (Venice *et al.* 2019).

The set of PT plant sequences used as reference consists of 11 amino acid sequences from Rausch *et al.* (2001) and eight sequences from Yadav *et al.* (2010) from six different plant species.

2.1.3 Mycocosm fungal genome data

929 fungal genomes were downloaded (2020-10-15) from the Mycocosm database (Grigoriev *et al.* 2014). We did not include information about the taxonomy for all genomes due to time constraint but for the ones that have the distribution is the following 491 Ascomycota, 309 Basidiomycota, 56 Mucoromycota, 26 Chytridiomycota and 20 Zoopagomycota.

2.2 Database and blast

An initial small blast database was created with the ten protein sequences for Pho87 from Plassard *et al.* (2019) together with the sequences for Pho87/90/91 from *Saccharomyces cerevisiae* as reference. This database was then used to blast (v. 2.9.0+) against 929 fungal genomes collected from the Mycocosm database (Grigoriev *et al.* 2014). The e-value threshold was set to 10^{-5}.

The blast hits were then sorted based on e-value, bit score and alignment length. All genomes that did not have a hit with alignment length above 500 were removed. After the filtering 480 genomes remained. From these 480 genomes the sequences that matched were extracted using BEDTools (v. 2.29.2).

A new enriched database was created using the extracted sequences, the reference sequences and the nucleotide sequences for the protein that was used in the previous database. Since the previous database was built on protein sequences the nucleotide sequences needed to be obtained. Out of the ten species used for Pho87 from Plassard *et al.* (2019) the genomes for eight of them were present among the genomes from Mycocosm (Grigoriev *et al.* 2014). However, only one of them had a good enough match to already be present among the extracted sequences. The best hit for the remaining seven was extracted manually since they were filtered out in a previous step.

The new database was then used to blast against the 26 AM genomes sequence by Montoliu-Nerin *et al.* (2020). The blast results only yielded short matches in the AM genomes which motivates the need to use a Hidden Markov model. The extracted sequences from the Mycocosm genomes, the nucleotide sequences for the species in the original database and the sequences from the blastn of the AM genomes were aligned using MUSCLE (v. 3.8.31, Edgar 2004). FastTree (v. 2.1.10, Price *et al.* 2010) was used to create a first draft phylogeny, see supplementary figure S2. The sequences did not cluster as expected from which the conclusion was drawn that we need to enrich the dataset with more sequences from different PTs. In order to keep working with the sequences as proteins instead of nucleotides the 480 Mycocosm scaffolds that generated a hit in the original blastx were gene predicted and annotated using a Snakemake pipeline (Montoliu-Nerin *et al.* 2020) so that amino acid sequences for the reference PTs could be used. Blastp (v. 2.9.0+, Altschul *et al.* 1990) was then performed on the gene prediction using a database containing amino acid sequences of PTs from a spread of fungi from Venice *et al.* (2019) and Plassard *et al.* (2019). The AM genomes were gene predicted and blasted against in the same way.

2.3 Further enrichment

We chose to include sequences of plant PTs together with the fungal sequences from Venice *et al.* (2019) and Plassard *et al.* (2019). The sequences for the plant PTs were retrieved from Rausch *et al.* (2001) and Yadav *et al.* (2010). These sequences were then used to do a blastp (Altschul *et al.* 1990) on NCBI only targeting plants (taxid:3193, Madden *et al.* 1996). All sequences with a percentage identity of 60 or above were kept. Ten fungal sequences from

different PTs with a good spread across the phyla Ascomycota, Basidiomycota, Glomeromycota and Mucoromycota from Venice *et al.* (2019) were selected. Again, a blastp (v. 2.9.0+, Altschul *et al.* 1990) was done on NCBI only targeting plants, this time keeping all sequences with a percentage identity of 30 or above. In total this generated 75780 plant PT sequences. These were reduced using CD-hit (Li and Godzik 2006, Fu *et al.* 2012) clustering sequences with 99 percent identity and with a short sequence coverage of 90 percent, resulting in 3397 sequences, see supplementary table S3.

2.4 Functional annotation and filtering

All sequences generated from blast were functionally annotated using pHMMer (v. 3.2.2, hmmer.org) with default settings. A search list was used to filter out sequences which did not have the required function, see supplementary list S4. This resulted in 408 AM sequences, 852 plant sequences and 1138 fungal sequences from Mycocosm (Grigoriev *et al.* 2014). All 2398 sequences generated from blast were added together with the sequences used for the database from Venice *et al.* (2019) and Plassard *et al.* (2019) resulting in a set of 2722 sequences.

2.5 Gene alignment and phylogeny reconstruction

The set of PTs were aligned using MUSCLE (v. 3.8.31, Edgar 2004). The alignment lacked conserved domains and is therefore not suitable to use for phylogenies. When studying the phylogeny generated with FastTree (v. 2.1.10, Price *et al.* 2010) the plant sequences cluster together and have a lot of redundancy. Once again CD-hit (Li and Godzik 2006, Fu *et al.* 2012) was used to reduce the amount of sequences with a sequence identity threshold of 90%. Resulting in a reduction from 852 sequences to 510 leading to a full set of 2380 sequences. To reduce the risk of short sequences spreading out over the alignment all sequences shorter than 200 amino acids were removed resulting in a set of 2198 sequences. Again the set was aligned using MUSCLE (v. 3.8.31, Edgar 2004) but when inspecting the alignment with MEGA-X (v. 10.2.6) no conserved domains were found. The whole set was then functionally annotated again but this time with NCBI CD-hit batch search (Shennan Lu *et al.* 2020) with default settings to again remove protein sequences that do not have the required function. The list generated was manually curated to meet the requirements. The set was then filtered based on the search list which was taken out from the full list, see table supplementary S5. After filtering on the new functions the set was reduced to 2029 sequences. Once again these were aligned using MUSCLE (v. 3.8.31, Edgar 2004) and once again no conserved domains were found. The full set was then divided into two, also based on the protein function. One of the sets contained sequences that have an SPX domain in addition to the PT sequence and the other one had pure PT sequences. The set with only PT sequences generated a better alignment, containing 1254 sequences. This set was then realigned with kalign (v. 2.04, Lassman 2019) together with alignments (accession cd01115, cl21473 and cl34877,

downloaded 2021-08-25) from NCBI, resulting in an alignment of 1879 sequences. Kalign (v. 2.04, Lassman 2019) was used instead of MUSCLE (v. 3.8.31, Edgar 2004) to reduce the alignment time. FastTree (v. 2.1.10, Price *et al.* 2010) was then used to construct a phylogeny of this set, see supplementary figure S6. Once examining the tree we decided to remove all the sequences containing MFS domains since they all grouped together and their function as PTs are uncertain. In addition to this all outliers were removed together with the references that were duplicates. This resulted in a final set of PTs containing 1351 sequences. This set was realigned with kalign (v. 2.04, Lassman 2019) and a final tree was constructed using FastTree (v. 2.1.10, Price *et al.* 2010). For a full schematic overview of the workflow, see figure 3.

2.6 Tree visualization

All trees were visualized and coloured with FigTree (v. 1.4.4, http://tree.bio.ed.ac.uk/software/figtree/). The colouring was based either on the phylum of the taxa or which group of interest (AM fungi, references, Mycocosm and plants) the taxon belongs to. By locating the reference sequences in the tree we could define three subgroups for fungi, Pho84, Pho87 and Pho89, and one for plants.

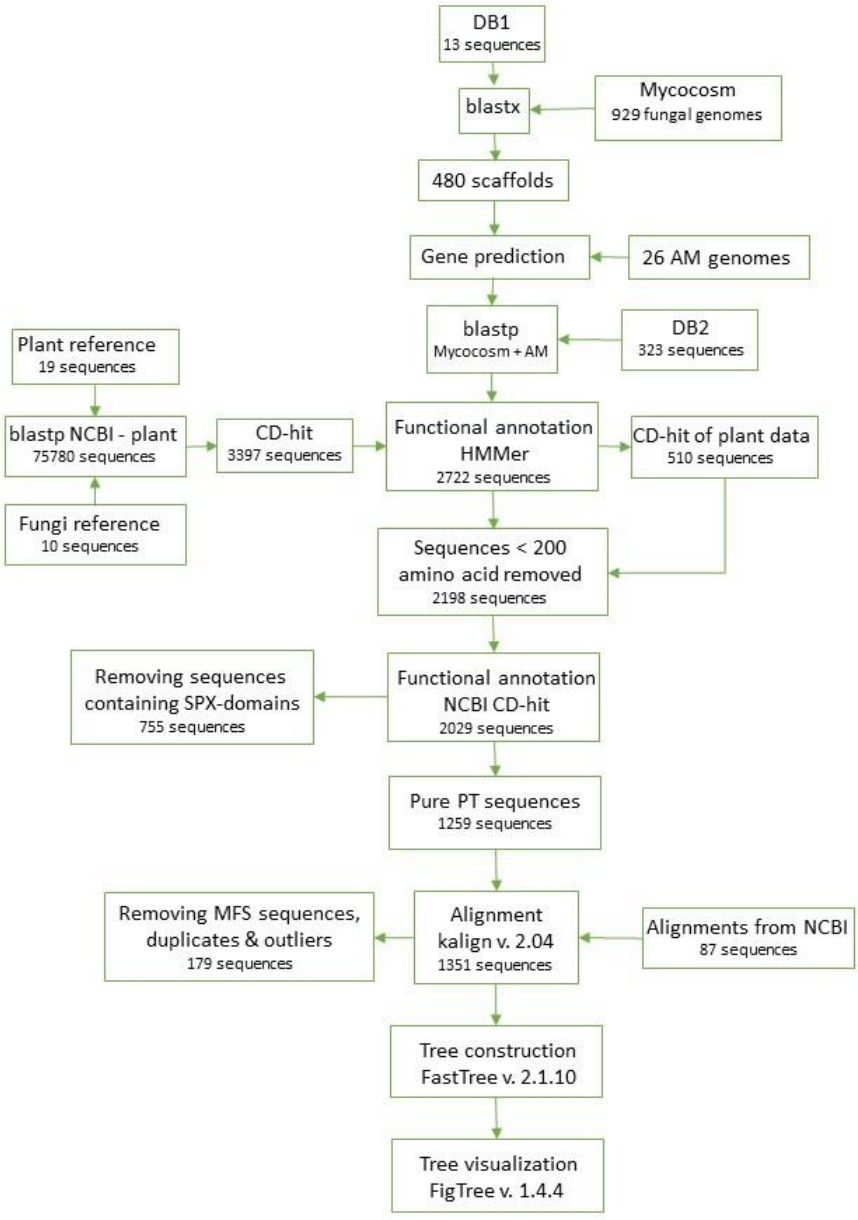


Figure 3. Methodical overview of the data acquiring and tree construction of the full trees. Starting with the initial database (DB1) containing 10 amino acid sequences of Pho87 together with sequences for Pho87/90/91 from *S. cerevisiae*. DB1 was then used to perform a blastx on the genome data from Mycocosm. The data was then gene predicted and a blastp was done with the reference sequences from Plassard *et al.* (2019) and Venice *et al.* (2019) as database (DB2). The dataset was then enriched with plant data generated with NCBI blast with both plant and fungal sequences as reference. Redundant sequences were removed with CD-hit before being functionally annotated with pHMMer together with the fungal sequences. Additional sequences of the plant data were removed again using CD-hit before removing all sequences shorter than 200 amino acids from the set. After another functional annotation this time using NCBI CD-hit all sequences containing an SPX-domain were removed leaving only pure PT sequences. These sequences were enriched with alignments from the relevant superfamilies, all duplicates and MFS sequences were removed, before being aligned with kalign (v. 2.04). The alignment was then used to construct the tree.

2.7 Tree of Pho87 subgroup

One of the branches of the first node in the Pho87 subgroup was selected for construction of a subtree. The subset consists of 312 sequences. The sequences were aligned using kalign (v. 2.04, Lassman 2019) and FastTree (v. 2.1.10, Price *et al.* 2010) was used to construct the phylogeny. FigTree (v. 1.4.4, http://tree.bio.ed.ac.uk/software/figtree/) was used for visualization and colouring.

2.8 Profile Hidden Markov model

As concluded earlier the short hits from the blastn (v.2.9.0+, Altschul *et al.* 1990) of the AM genomes confirms that a model is useful in order to capture the genetic variety of the PTs as they have diverged. The model used for this is a profile Hidden Markov model (HMM) which uses a multiple sequence alignment to generate a scoring system that is position specific. A profile HMM takes into account what residue has the highest likelihood to occur at a specific position based on biases. This is also the case of insertions and deletions, which are more likely to occur at certain positions. A profile HMM is therefore better at catching homologues than the blast scoring system that does not take any specific positions into account and penalizes changes of residues, insertions and deletions equally (Zvelebil and Baum 2008). The branch containing the reference sequences for Pho87 and Pho91 from *S. cerevisiae* was selected for construction of an HMM. This subset consists of 312 sequences. The model was then tested by searching the full set of 1351 sequences.

3 Results

3.1 Statistics of collected PT sequences

The final set of PTs from all four datasets (AM fungi, Mycocosm, references and plants) consists of 1351 sequences. Among these sequences 243 belong to AM fungi, 444 to plants and 363 to Mycocosm, the different phyla can be seen in table 1.

Table 1. Statistics of the collected PT sequences in the final phylogeny. Table showing the spread of sequences from three of the datasets. The fungal sequences are divided according to phylum.

Dataset	Phylum	Number of sequences in the final phylogeny
AM fungi	Glomeromycota	243
Mycocosm	Ascomycota	190
	Basidiomycota	74
	Mucoromycota	73
	Zoopagomycota	13
	Chytridiomycota	6
	Blastocladiomycota	2
	Not specified	5
Plant	Plant	444

3.2 Visualizing phylogenetic relations of the collected PT sequences

The first tree generated gives an overview of the spread of the data, see figure 4. The tree is unrooted and coloured based on the different groups of data, AM fungi, plant, Mycocosm and reference sequences. In addition to these groups the tree is divided into three subgroups of fungi based on where the reference sequences from *S. cerevisiae* for Pho84, Pho87 and Pho89 are present. The groups also contain other reference sequences. For instance Pho84 contains PT2, PT3 and PT 5, Pho87 contains PT5-PT9 and Pho89 contains PT2, PT5 and PT9. The Pho87 group also contains Pho91.

In the tree there are two groups consisting of only plant sequences which represent PTs that are not present in fungi. One of them is the group closest to the centre of the tree which could suggest that those plant PTs are the origin of other PTs in both plants and fungi. The other large group of plants are in the Pho89 subgroup where it looks like the origin of the fungal sequences are from plants. A few plant sequences are present in all subgroups. This could be due to events of horizontal gene transfer (HGT).

In general the AM fungi sequences are grouped together with only a few exceptions. AM fungi is present in all fungal subgroups of the tree but mainly in the Pho84 subgroup where there is clear clade. Even though AM fungi form clades in the Pho87 subgroup, data from Mycocosm is dominating.

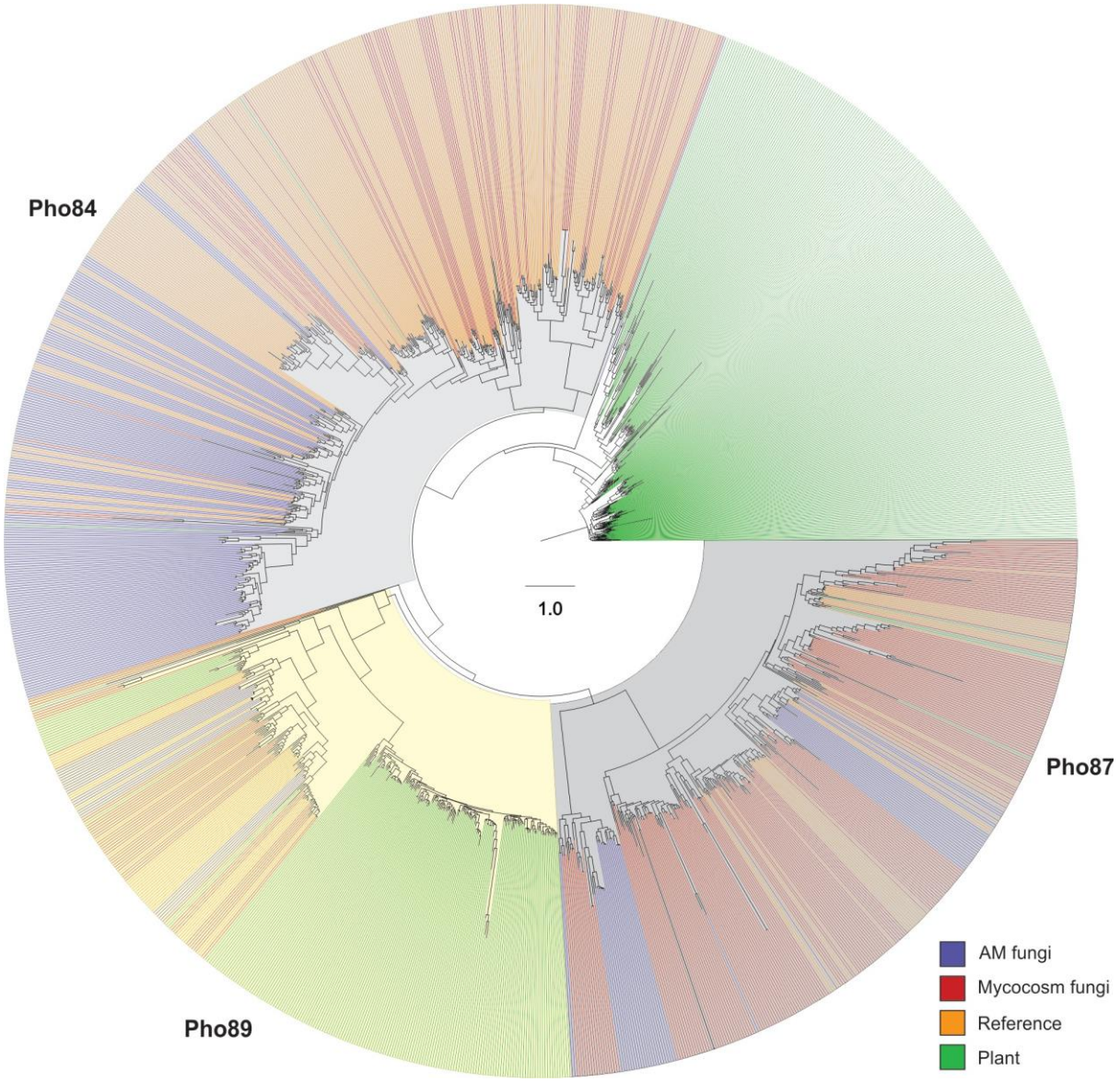
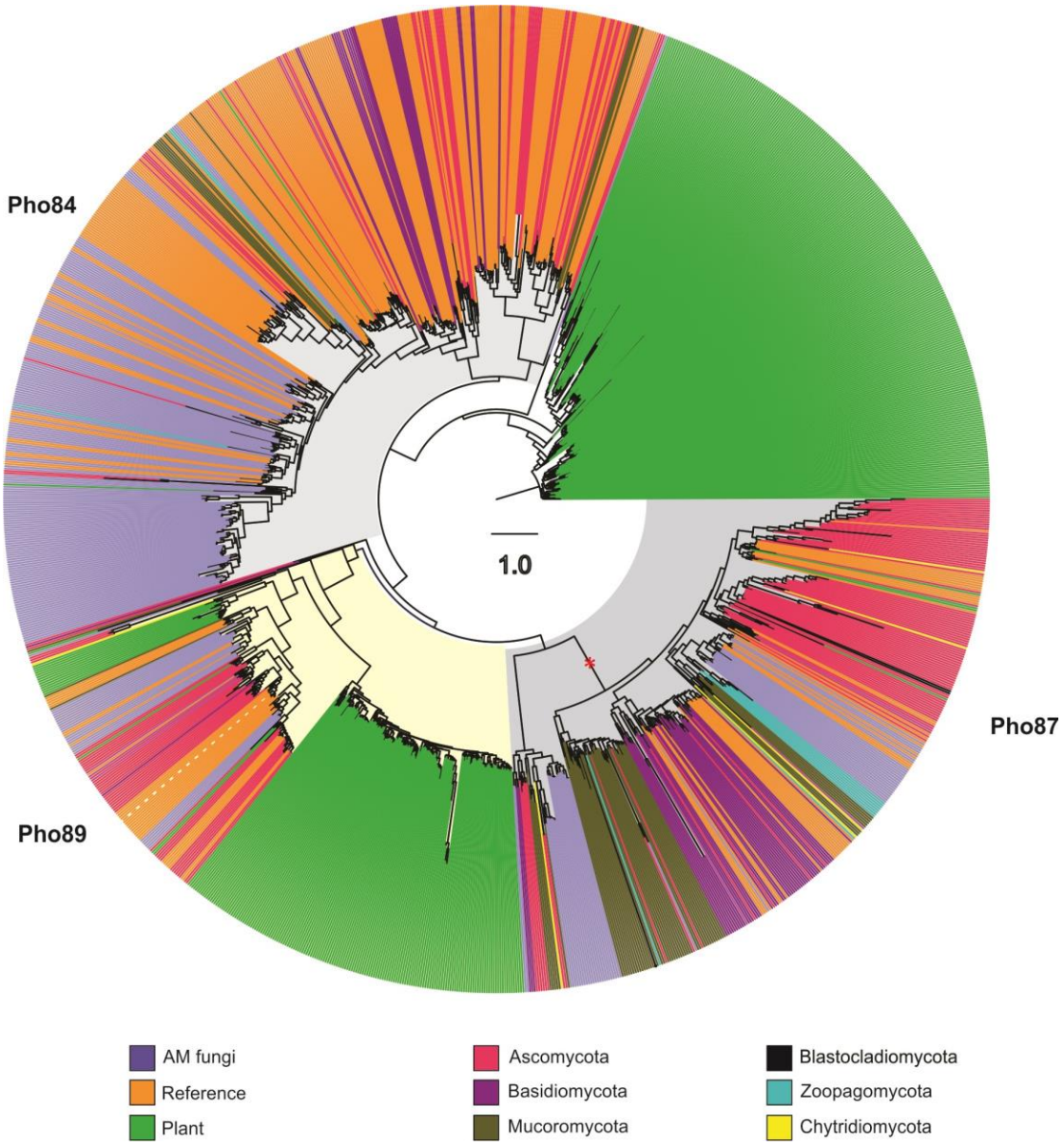


Figure 4. Unrooted maximum-likelihood tree of the phylogenetic relationship of the collected PTs. The alignment consists of 1351 sequences from four different datasets. The taxons in the tree are coloured based on which dataset they come from. Blue for AM fungi, red for Mycocosm fungi, orange for references and green for plants. The tree has four subgroups, one for plants and three for fungi. The fungal subgroups have been divided based on where the reference sequence of Pho84, Pho87 and Pho89 from *S. cerevisiae* are.

3.3 Visualizing phylogenetic relations based on taxonomy

In the second tree the data from Mycocosm was coloured based on phylum, see figure 5. The phyla present are Ascomycota, Basidiomycota, Blastocladiomycota, Chytridiomycota, Mucoromycota and Zoopagomycota from Mycocosm and Glomeromycota from AM fungi. The taxonomy for the reference sequences varies and are not present in the tree. These are therefore still marked as reference. In general the different phyla group together. The three most abundant phyla in the dataset used are Ascomycota, Basidiomycota and Glomeromycota

(AM fungi). Ascomycetes and Glomeromycetes are well represented in all three Pho groups, keeping in mind that most of the reference sequences belong to Ascomycota. Basidiomycetes are present in the Pho84 and Pho87 groups but very few of them are in the Pho89 group. This could be due to fewer gene copies or that they have very few genes from that gene family. AM fungi is mostly present in the Pho84 group which could indicate that they use PTs belonging to this group more than PTs in the other groups, however it should be mentioned that this group is the one that is most inclusive and contains a lot of different PTs compared to the other two groups.



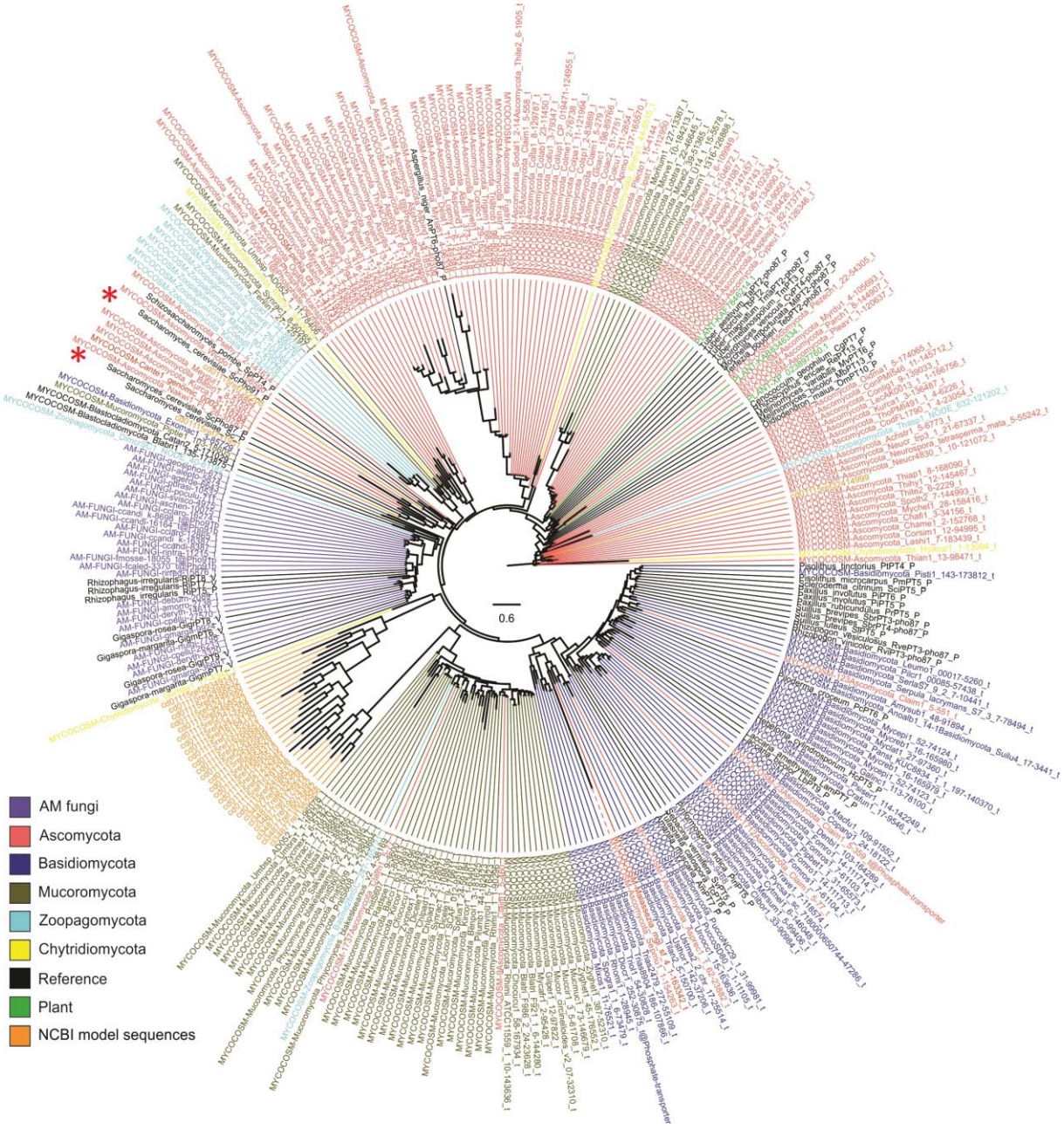
Figur 5. Unrooted maximum-likelihood phylogenetic tree coloured based on the taxonomy of the fungal sequences. The AM fungi belongs to the phylum Glomeromycota. Observe that there is no taxonomy for the reference sequences. The branch in the Pho87 subgroup marked with a red star was used for further analysis.

3.4 Tree of the Pho87 branch

From the subgroup Pho87 one branch was selected for further analysis, marked with a red star in figure 5. This branch contains 312 sequences. The second smaller branch, without a star, did not have a reference and thus could be a different gene family and was not included. When examining the taxons of the large Pho87 branch they have the same superfamily, cl34877, from the alignments from NCBI. Proteins belonging to this superfamily have an SPX-domain and are a part of the accumulation of vacuolar polyphosphate (https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=cl34877&spf=1).

In the Pho87 tree, see figure 6, three plant sequences are present and the rest are fungi. The sequences group according to phylum, except for the three Chytridiomycota sequences that are spread out over the tree and three sequences from Zoopagomycota. Two of the four Chytridiomycota sequences belong to the same species. The NCBI model sequences group together except a few of them, most likely the few that are spread out among the fungi are fungal sequences while the rest are bacteria and archaea. The three plant sequences all belong to trees from different families. The reference sequences from *S. cerevisiae* for Pho97 and Pho90, marked with red stars in figure 6, group together.

In the smaller branch of the Pho87 group that was not used to construct the tree there are 17 AM sequences from 15 unique species. 13 of these are present in both branches of the tree.



Figur 6. Unrooted maximum-likelihood phylogenetic tree of the subset from the Pho87 group. The alignment consists of 312 sequences. The tree is coloured based on phylum. The reference sequences from *S. cerevisiae* for Pho87 and Pho91 are marked with red stars.

3.5 Profile Hidden Markov model

The HMM consists of 312 sequences and has a length of 2022. Three of the sequences in the model are plants and the rest are fungi, as can be seen in figure 6. The model was tested with the default threshold on the whole dataset of 1351 sequences. The model captured all 312 sequences in the branch, as expected, together with 43 additional sequences. These 43

sequences also belonged to the Pho87 group but in the smaller branch that was left out from the model due to functional uncertainty.

4 Discussion

4.1 Bias in the dataset

The data from Mycocosm (Grigoriev *et al.* 2014) and the reference sequences both from Plassard *et al.* (2019) and Venice *et al.* (2019) used in this study is weighted heavily towards the phylum Ascomycota followed by Basidiomycota. Therefore, in the tree coloured after phylum it is reasonable that Ascomycota is the leading phylum, apart from Glomeromycota (AM fungi). Ascomycota is the leading phylum in the two reference sets as well which were used to create the initial database which makes it probable that after that database was used to search the Mycocosm genomes mostly Ascomycetes were detected. Since the matches then was made into a new database that was used to search the AM fungi genomes it is likely that the sequences captured are more closely related to Ascomycetes than any of the other phyla.

When using for instance blast or CD-hit the chosen thresholds and cut-off values are arbitrary which also affects the data in the sense of what is captured. In this study we also chose to remove other transporters from the set because we were not sure that they were PTs, for instance all sequences that were functionally annotated as MFS-transporters were removed.

All groups in the tree, see figure 4, contain at least one plant sequence. This could be due to HGT events but it could also be due to sampling error. If a plant sample is taken from the root of the plant it might not belong to the plant at all but instead be a fungal sequence.

4.2 Phylogenetic separation within Dikarya

According to the study conducted by Plassard *et al.* (2019) fungi belonging to the phylum Basidiomycota lack a gene copy for Pho87. In the phylum tree, see figure 5, there are several clades dominated by Basidiomycetes in the Pho87 group. The Pho87 group also contains sequences for Pho91 which according to Plassard *et al.* (2019) Basidiomycetes have. When examining the references and the clades more closely in the Pho87 group the majority of Basidiomycetes are not in the same clade as the Pho91 reference but instead they group together with Pho87. The only Ascomycete they studied that had a gene copy for Pho91 was *S. cerevisiae*, which is used as a reference in this study. In contrast to their findings our clade with the Pho91 reference also contains several Ascomycetes. However, when comparing the results of the Pho89 group with the results of Plassard *et al.* (2019) they correspond. According to their study only a few of the Basidiomycetes species they used had a gene copy

for Pho89. In the phylum tree, see figure 5, there are visibly very few Basidiomycetes present.

One reason for this deviant result could be that there is no real standard for naming protein sequences, where two different proteins can have the same name and vice versa. This might also be seen when studying the different subgroups. For instance the Pho89 group contains the reference for PT2 and PT9 which are also present in the Pho84 and Pho87 groups, respectively and PT5 is present in all groups. This makes it difficult to determine which of the other PTs are related to the different Pho sequences. Another reason could be the amount of data studied. This study uses more sequences and species which makes the taxonomy a bit messier but also has the potential to capture greater variety.

4.3 Comparison of other PT phylogenies

Venice *et al.* (2019) did a similar study as this one with 82 fungal sequences of 58 amino acids in length. When comparing how the different PTs are spread out over the tree the results correspond even though this study contains more species, longer sequences and also includes plants. All three Pho subgroups are visible and contain the same reference sequences. For instance all groups in the tree contain PT5 and the group containing Pho87 and Pho91 also contain PT5-PT9. In the tree in the publication of Venice *et al.* (2019) Pho87 and Pho91 belong to the same group but even when taking out that branch, see figure 5, they group together supporting their sequence similarity.

4.4 Gene copies and duplication events

In this study we did not look further into the amount of gene copies but the trees show indications of duplication events. There are only 26 AM genomes in total and a lot of AM sequences in the Pho84 and the Pho87 group, see figure 4, this indicates duplication events which makes it likely that AM fungi have multiple gene copies for this group. The smaller branch in the Pho87 group could also be due to duplication events across the studied taxa since taxa belonging to the same phylum are present in both branches. When comparing the species for AM in the two branches 13 out of 15 species present in the small branch are also present in the large branch, supporting a duplication event. In figure 6 two of the four spread out sequences for Chytridiomycota belong to the same species indicating a duplication event, as well as at least two gene copies.

4.5 Future studies and the use of models

In this study only one HMM model for one of the branches in the Pho87 subgroup was constructed. This model was not tested properly for specificity and accuracy due to time constraint but on default settings it captured the 312 sequences in the large Pho87 branch, as expected. It also captured 43 sequences in the smaller branch of Pho87 indicating that these

might be Pho87 as well. Further testing of the model is needed to find a threshold that will only capture the 312 sequences the model is built on.

For future studies a model could be made for each of the fungal groups. Starting with a bigger dataset, maybe even of the whole group, and testing it on the dataset. If testing shows that higher accuracy is needed the dataset could be reduced by going further into the branchings. These models could then be used to search different fungal datasets to capture PTs. The need for models that properly capture PTs are motivated by the fact that blasting only gives limited hits. In this study we walked around this by gene prediction which enabled us to take out the full protein sequence for each hit, this step could be skipped with the use of models. A model could also help with answering questions such as how many gene copies there are and how diverse the genes are.

5 Conclusion

The main result of this study is a good collection of PT sequences across different fungal phyla together with plant PTs. The trees constructed show a great variety of PTs for different phyla of fungi. The trees also show indications of duplication events and potentially multiple gene copies of PTs for several fungal species. Even though only one HMM was constructed in this study it shows the potential use of the generated dataset for further research.

6 Ethical aspects

There are many benefits of the AM symbiosis which people want to utilize. One way to do so is to use AM fungi as a biofertilizer. But besides the varying outcomes, what are the potential risks? Introducing an exotic species into an ecosystem always comes with a risk. If the AM fungi inoculum can spread outside of the intended area it may become invasive which can change the ecosystem, especially the fungal and plant biodiversity (Hart *et al.* 2017). It is known that AM fungi can disperse through sediment, the atmosphere and through migrating birds. However, there is still a lot more to learn about the dispersal which makes the risk even greater since we cannot evaluate it properly. An inoculum can only spread if it is viable and can colonize the intended area, this is also a criterion for a successful inoculation. The successful establishment of AM fungi varies on a lot of different factors based on the host requirements, such as the nutritional needs, which change over time. Hence, the inoculum needs to be host specific enough to establish.

One issue with commercial inoculants is that they might contain other taxa than specified (Hart *et al.* 2017) which, in addition to being unsuccessful, can lead to an ever greater risk of creating an invasive species since the user cannot be certain of which species is being applied.

Many biofertilizers also contain other taxa than AM fungi, such as soil bacteria (Hart *et al.* 2017), which may increase the risk further. There is also a risk that AM fungi will replace, fully or partially, the resident species, since they are present almost everywhere (Hart *et al.* 2017).

AM inoculants reduce the establishment of native plants, thus reducing the plant biodiversity. However, it has been shown that introduction of native AM fungi may have a positive outcome on the ecosystem since it may help plants withstand invasions from pathogens for example, while non-native inoculants may favor exotic plants over native (Hart *et al.* 2017). The idea of using AM fungi as a biofertilizer seems like a good one, but are we really there yet? Taken together that much is still unknown about AM fungi taxa and that it seems to be hard to guarantee the biota in the inoculum, is it worth the risk when the outcomes vary so much?

7 Acknowledgements

I would like to thank my supervisors Anna Rosling and Maliheh Mehrshad for all the help and support during this project. I would also like to thank everyone in the Rosling lab group for their input and comments.

8 References

Akiyama K, Matsuzaki K-i, Hayashi H. 2005. Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. Nature 435: 824-827.

Altschul S F, Gish W, Miller W, Myers E W, Lipman D J. 1990. Basic local alignment search tool. Journal of Molecular Biology 215:403-410.

Bonfante P, Genre, A. 2008. Plants and arbuscular mycorrhizal fungi: an evolutionary-developmental perspective. Cell press, doi:10.1016/j.tplants.2008.07.001.

Edgar R C. Muscle. Nucleic Acids Res 32: 1792-97.

Ferrol N, Azcón-Aguilar C, Pérez-Tienda J. 2018. Review: Arbuscular mycorrhizas as key players in sustainable plant phosphorus acquisition: An overview on the mechanisms involved. Plant Science 280: 441-447.

Ferrol N, Tamayo E, Vargas P. 2016. The heavy metal paradox in arbuscular mycorrhizas: from mechanisms to biotechnological applications. Journal of Experimental Botany 67: 6253-6265.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next generation sequencing data, Limin Fu, Beifang Niu, Zhengwei Zhu, Sitao Wu & Weizhong Li. Bioinformatics 28:3150-3152.

Grigoriev I V, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I. 2014. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Research 42: 699-704.

Hart M M, Antunes P M, Chaudhary V B, Abbott L K. 2017. Fungal inoculants in the field: Is the reward greater than the risk? Functional Ecology, doi: 10.1111/1365-2435.12976.

Jany J-L, Pawlowska T E. 2010. Multinucleate Spores Contribute to Evolutionary Longevity of Asexual Glomeromycota. The American Naturalist 175: 424-435.

Kameoka H, Maeda T, Okuma N, Kawaguchi M. 2019. Structure-Specific Regulation of Nutrient Transport and Metabolism in Arbuscular Mycorrhizal Fungi. Plant & Cell Physiology 60: 2272-2281.

Kokkoris V, Chagnon P-L, Yildirir G, Dettman J, Stefani F, Corradi N. 2021. Host identity influences nuclear dynamics in arbuscular mycorrhizal fungi. Current Biology 31: 1-8.

Lambers H, Raven J A, Shaver G R, Smith S E. 2008. Plant nutrient-acquisition strategies change with soil age. Cell press, doi:10.1016/j.tree.2007.10.008.

Lassmann, T. 2019. Kalign 3: multiple sequence alignment of large data sets. Bioinformatics.

Lu S *et al*. 2020. NCBI CD-hit: CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res.

Marleau J, Dalpé Y, St-Arnaud M, Hijri M. 2011. Spore development and nuclear inheritance in arbuscular mycorrhizal fungi. BMC Evolutionary Biology 11: 51.

Montoliu-Nerin M, Sánchez-García M, Bergin C, Grabherr M, Ellis B, Kutschera V E, Kierczak M, Johannesson H, Rosling A. 2020. Building de novo reference genome assemblies of complex eukaryotic microorganisms from single nuclei. Scientific Reports 10: 1303.

Montoliu-Nerin M, Sánchez-García M, Bergin C, Kutschera V E, Johannesson H, Bever J D, Rosling A. 2021. In-depth Phylogenomic Analysis of Arbuscular Mycorrhizal Fungi Based on a Comprehensive Set of *de novo* Genome Assemblies. Frontiers, doi: 10.3389//ffunbb.2021.716385.

Plassard C, Becquer A, Garcia K. 2019. Phosphorus Transport in Mycorrhiza: How Far Are We? Trends in Plant Science 24: 794-801.

Price M N, Dehal P S, Arkin AP. 2010. FastTree 2 -- Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE, doi:10.1371/journal.pone.0009490.

Rausch C, Daram P, Brunner S, Jansa J, laloi M, Leggewie G, Amrhein N, Bucher M. 2001. A phosphate transporter expressed in arbuscule-containing cells in potato. Nature 414: 462-466.

Venice F, Ghignone S, Salvioli di Fossalunga A, Amselem J, Novero M, Xianan X, Sędzielewska Toro K, Morin E, Lipzen A, Grigoriev I V, Henrissat B, Martin F M, Bonfante P. 2019. At the nexus of three kingdoms: the genome of the mycorrhizal fungus Gigaspora margarita provides insights into plant, endobacterial and fungal interactions. Environmental Microbiology 22: 122-141.

Weizhong L, Godzik A. 2006. CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658-1659.

Yadav V, Kumar M, Deep D K, Kumar H, Sharma R, Tripathi T, Tuteja N, Saxena A K, Johri A K. 2010. A Phosphate Transporter from the Root Endophytic Fungus Piriformospora indica Plays a Role in Phosphate Transport to the Host Plant. The Journal of Biological Chemistry 285: 26532-26544.

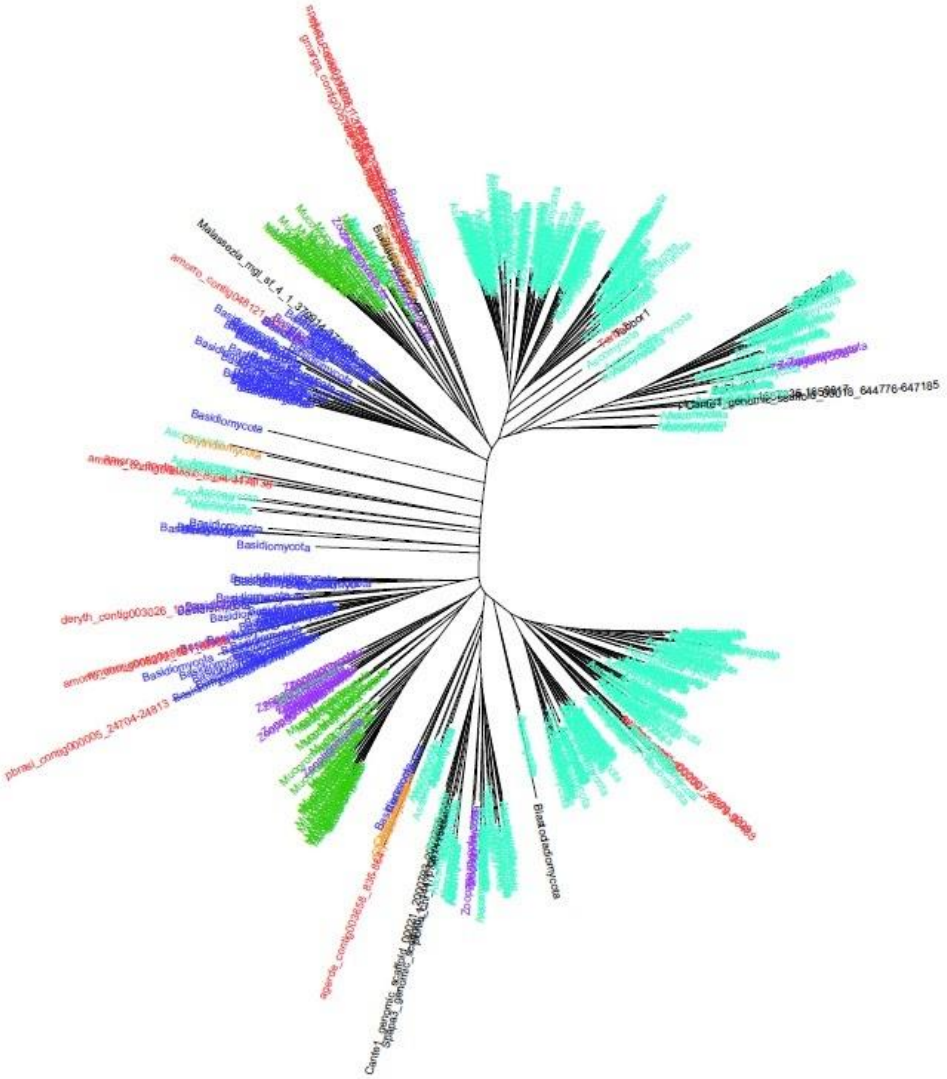
Zhang J, Madden, T L. 1997. PowerBLAST: A new network BLAST application for interactive or automated sequence analysis and annotation. Genome Res.

Zvelebil M, Baum J O. 2008. Patterns, profiles, and multiple alignments: Harbor J (pub.). Understanding Bioinformatics, s. 165-222. Garland Science, Taylor & Francis Group, LLC, New York City.

9 Supplementary

S1. Table of AM fungi species present in the data set with the short name and the species name (Montoliu-Nerin *et al.* 2021).

Short name	Species name
Acolom	*Acaulospora colombiana*
Agerde	*Ambispora gerdemannii*
Alepto	*Ambispora leptotricha*
Amorro	*Acaulospora morrowiae*
Aschen	*Archaeospora schenckii*
Ccandi	*Claroideoglomus candidum*
Ccandi_k	*Claroideoglomus candidum Kansas*
Cclaro	*Claroideoglomus claroideum*
Cpellu	*Cetraspora pellucida*
Deburn	*Diverspora eburnean*
Depige	*Diversispora Epigaea*
Deryth	*Dentiscutata erthropa*
Dheter	*Dentiscutata heterogama*
Fcaled	*Funneliformis caledonius*
Fmosse	*Funneliformis mosseae*
Gmarga	*Gigaspora margarita*
Grosea	*Gigaspora rosea*
Pbrasi	*Paraglomus brasilianum*
Poculu	*Paraglomus ocultum*
Rfulgi	*Racocetra fulgida*
Rintra	*Rhizophagus intraradices*
Rirreg	*Rhizophagus irregularis*
Rpersi	*Racocetra persica*
Scalos	*Scutellospora calospora*
Speluc	*Scutellospora pellucida*
Svisco	*Septoglomus viscosum*



S2. Unrooted maximum-likelihood draft tree of the first dataset coloured according to phylum. It was generated using the initial database containing 10 fungal sequences for Pho87 and reference sequences from *S. cerevisiae* for Pho87/90/91 which was blasted again the Mycocosm genomes to create a larger database which was used to blast against the AM genomes. The tree does not contain full protein sequences, only the blast hits.

S3. Table containing the blast results from NCBI. One for searches using plant reference and one for fungal references. The total number of sequences before reduction was 75780, after reduction using CD-hit 3397 sequences remained. Sequences retrieved 2021-03-05.

Plant to plant						
Species	Common name	Gene	Acc. No	Paper	Notes	Blast 60-100%
Lycopersicon esculentum/	Tomato	LePT1	24029 (SP)	Rausch		2397
Solanum lycopersicum		LePt2	22549 (SP)	Rausch		2866
		LePT4	AY885651 (GB)	Yadav	DNA	1107
Nicotiana tabacum	Tobacco	NtPT1	AAF74025 (GB)	Rausch		2896
		NtPT2	BAA86070 (GB)	Rausch		2881
		NtPT3	AB042951 (GB)	Yadav	mRNA	2797
		NtPT4	AB042956 (GB)	Yadav	mRNA	2855
Solanum tuberosum	Potato	StPT1	Q43650 (SP)	Rausch		2921
		StPT2	Q41479 (SP)	Rausch		2657
		StPT3	AJ318822 (GB)	Rausch	DNA	2819
		StPT4	AY793559 (GB)	Yadav	DNA	1184
		StPT5	AY885654 (GB)	Yadav	mRNA	2300
Medicago truncatula	Barrelclover	MtPT1	022301 (SP)	Rausch		2745
		MtPT2	022302 (SP)	Rausch		2824
		MtPT4	AY116210 (GB)	Yadav	mRNA	1930
Arabidopsis thaliana	Mouse-ear cress	AtPT1	Q96302 (SP)	Rausch		2933
		AtPT2	Q96303 (SP)	Rausch		2686
Sesbania rostrata		SrPT1	AJ286743 (GB)	Yadav	mRNA	2978
		SrPT2	AJ286744 (GB)	Yadav	mRNA	2933
Fungi to plant						
Group	Division	Species	Gene	Blast 30-100%		
Pi-repressible	Ascomycota	*Saccharomyces cerevisae*	ScPHO89	415	55.81%	
SPX domain	Glomeromycota	*Gigaspora margarita*	GigmPT7	317	63.32%	
Bas. PT2	Basidiomycota	*Laccaria bicolor*	LbPT2	3431	58.97%	
Muc. PT2	Mucoromycota	*Rhizopus delemar*	RdPT3	3548	60.34%	
Mor. PT2	Mucoromycota	*Mortierella elongata*	MePT4	3606	65.52%	
Muc. PHO84-like	Glomeromycota	*Rhizophagus irregularis*	RiPT3	3571	73.08%	
Glo. PT	Glomeromycota	*Funneliformis mossea*	FmPT	3446	72.00%	
Asc. PHO84-type	Ascomycota	*Aspergillus fumigatus*	AfPT	3614	71.61%	
Muc. PHO84-type	Mucoromycota	*Mucor circinelloides*	McPT2	3530	64.71%	
Bas. PHO84-type	Basidiomycota	*Hebeloma cylindrosporum*	HcPT1	3544	67.24%	

S4. List of search words used to filter the dataset after functional annotation with pHMMer.

- Inorganic phosphate transporter
- Inorganic phosphate transporter PHO84
- Low-affinity phosphate transporter
- Low-affinity phosphate transporter PHO91
- major facilitator superfamily domain-containing protein
- major facilitator superfamily transporter
- MFS domain-containing protein
- MFS domain-containing protein (Fragment)
- MFS general substrate transporter
- MFS transporter
- phosphate permease
- phosphate transport
- phosphate transport (Eurofung)
- phosphate transporter
- Plasma membrane phosphate transporter Pho87
- Putative phosphate transport (Eurofung)
- SPX domain-containing protein
- SPX domain-domain-containing protein
- SPX-domain-containing protein

S5. Table for filtering the NCBI CD-hit results. Only the sequences annotated as pure PTs were kept for further analysis.

Pure PTs				**SPX domains**		
Accession	Short name	Superfamily		Accession	Short name	Superfamily
cl31035	2A0109 superfamily	-		pfam03105	SPX	cl28943
TIGR00887	2A0109	cl31035		cl21499	SPX superfamily	-
pfam01384	PHO4	cl27575		cl28943	SPX superfamily	-
cl27575	PHO4 superfamily	-		cl28942	COG5408 superfamily	-
cd17364	MFS_PhT	cl28910		cd14478	SPX_PHO87_PHO90_like	cl21499
cd01115	SLC13_permease	cl21473		COG5408	COG5408	cl28942
cl21473	ArsB_NhaD_permease superfamily	-		cd14481	SPX_AtSPX1_like	cl21499
cl34877	COG5036 superfamily	-		cd14447	SPX	cl21499
pfam00153	Mito_carr	cl37969		cd14483	SPX_PHO81_NUC-2_like	cl21499
cl37969	Mito_carr superfamily	-				
cl28910	MFS superfamily	-				
cd17316	MFS_SV2_like	cl28910				
cd17380	MFS_SLC17A9_like	cl28910				

S6. Unrooted maximum-likelihood tree of the 1879 sequences where the sequences containing MFS domains are marked in grey and were removed before further analysis.