Functional genomics in symbiotic interactions

From high-throughput sequencing to functional characterization

Dissertation

der Fakultät für Biologie

der Ludwig-Maximilians-Universität München

vorgelegt von

Anne Hoffrichter

München, 2018

Functional genomics in symbiotic interactions

From high-throughput sequencing to functional characterization

Dissertation

an der Fakultät für Biologie

der Ludwig-Maximilians-Universität München

vorgelegt von Anne Hoffrichter, München, 2018

1. Gutachter: **Prof. Dr. Martin Parniske**

2. Gutachterin: **Prof. Dr. Silke Werth**

Tag der Abgabe: **02. August 2018**

Tag der mündlichen Prüfung: **10. Dezember 2018**

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

München, den 02. August 2018 .....

(Anne Hoffrichter)

Erklärung

Hiermit erkläre ich,

- dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist.

- dass ich mich anderweitig einer Doktorprüfung ohne Erfolg **nicht** unterzogen habe.

München, den 02. August 2018 .....

(Anne Hoffrichter)

Acknowledgements

I would like to thank Prof. Dr. Martin Parniske for his support and the opportunity to conduct my thesis in his group.

My special thanks go to my supervisor Dr. Andreas Brachmann, who believed in me and passed his knowledge on to me. Thanks for always motivating me and showing me a world of science. I had a great time with you!

I also want to thank past and current members of the AG Brachmann: Dr. Kinga Sedzielewska Toro and Dr. Charissa de Bekker for scientific discussions and for being two very strong and special women in science!
Also special thanks to Gisela Brinkmann and Ute Bergmann for their help and company in the lab!

Further I want to show my gratitude to the students that I helped supervise:
Andrea Holzer, who was a great Master-Student. Thanks especially for your help on the yeast complementation work.
And thanks to Julia Puchan who helped on this project with cloning, but was also a participant in fruitful discussions in our group meetings.

Of course I also want to thank our collaborators for giving me the opportunity to work on very interesting projects:
Thanks to Dr. Macarena Marín, for giving me the chance to play with 3rd generation sequencing data!
Thanks to Dr. Nicolas Corradi for having me participate on this very exciting work on recombination in AM fungal genomes!
Thanks also to Dr. Arthur Schüßler, for starting the *Geosiphon* project and being always willing to answer questions about this special little fungus.

I am also thankful to all the members of the WeiGuBra-meetings (AG Weiberg, AG Gutjahr & AG Brachmann) for the scientific discussions in a very interactive forum. It was always nice to present fresh data and hear about your ideas in these meetings!

My warmest thanks go out to the friends I made along the way, Aline & Priya I hope we will keep in touch! Janet, my best office-pal! And the kicker connection Flo, Salar, Michi, Antoine & Aline!

I am also very grateful for everybody who helped me with proofreading this thesis, Andreas & Dagmar but also to two very important people in my life, Tina & Julius!

For always being supportive in high tides or in low tides I am very thankful to my family and the most special people in my life, Julius, Tina, Kathi, Lisi & Christina!

List of publications

Parts of this work have already been submitted for publication:

Juan Liang*, Anne Hoffrichter*, Andreas Brachmann, Macarena Marín (2018). Complete genome of *Rhizobium leguminosarum* Norway, an ineffective *Lotus* micro-symbiont. (submitted to "Standards in Genomic Sciences", 24.01.2018).

* Equal contribution

From the work presented in this thesis the following manuscripts are in preparation:

Anne Hoffrichter, Arthur Schüßler, Andreas Brachmann (2018). The *Geosiphon pyriformis* transcriptome reveals common evolutionary origin and special adaptations in glomeromycotan symbioses.

The results in this manuscript describe the analysis of the *Geosiphon pyriformis* symbiotic transcriptome. They offer insights into the evolutionary origin of arbuscular mycorrhiza symbiosis. These data can be found in section 2.2 (page 24-32).

Anne Hoffrichter, Andrea Holzer, Kinga Sędzielewska-Toro, Andreas Brachmann (2018). Day-Night rhythm has only minimal effect on arbuscules and trehalose biosynthesis in *R. irregularis*.

The manuscript shows the analysis of the *R. irregularis* genes responsible for trehalose biosynthesis. The expression of the highly conserved and functional genes shows a slight diurnal rhythm. These results can be found in section 2.3.5-2.4.2 (page 38-44).

From work not presented in this thesis the following manuscript has been submitted:

Eric C.H. Chen, Stephanie Mathieu, Anne Hoffrichter, Kinga Sędzielewska-Toro, Max Peart, Adrian Pelin, Steve Ndikumana, Jeanne Ropars, Steven Dreissig, Jörg Fuchs, Andreas Brachmann and Nicolas Corradi (2018). Single nucleus sequencing reveals evidence of inter-nucleus recombination in arbuscular mycorrhizal fungi. (submitted to "eLife", 25.07.2018)

Summary

Functional genomics provide information beyond simple genotype-phenotype relationships. Whole genome and transcriptome sequencing is used to compare existing results to sequencing data from newly sequenced species and to thereby infer functions. With this approach, also analysis of organisms that are difficult to culture, or for which no transformation methods are available, is feasible. This applies e.g. to the Glomeromycota fungi, which are obligate symbionts and their hyphae contain thousands of nuclei in a common cytoplasm.

The work presented here shows the different stages of functional genomics in three parts. In the first part, a complete and gap-less assembly of the bacterial genome of *Rhizobium leguminosarum* Norway was created by using different 2nd and 3rd generation sequencing techniques. This assembly shows the advantages of using short read assemblies in combination with long reads. This is particularly relevant to bridge for example long repeat regions and retain base accuracy, which is also applicable for larger genomes.

In the second part, the assembly and analysis of the transcriptome of the early diverging glomeromycotan fungus *Geosiphon pyriformis* was performed. The results of functional annotation of the non-redundant virtual transcripts show higher similarity to Asco-, Basidio-, and Zygomycota than to Glomeromycota. This indicates that the different symbiotic life-style of *G. pyriformis*, compared to other Glomeromycota, also demands a different setup of expressed genes. On the other hand, the absence of the "Missing Glomeromycota Core Genes" set in this transcriptome suggests that there was a common obligate symbiotic ancestor, depending on either cyanobacteria or algae. In the course of this analysis three transcripts were identified that show similarities to genes encoding amyloid proteins and pose promising candidates for functional characterization.

The third part regards the functional characterization of effector candidate genes that had been identified in *Rhizophagus irregularis* through a genome comparison approach. Using host induced gene silencing one effector candidate gene was downregulated and showed a strong phenotype in colonization and arbuscules. Further analyses indicate that the protein encoded by this gene is involved in trehalose biosynthesis and is not an effector. The phenotype of downregulation shows that trehalose biosynthesis plays an important role in the establishment or maintenance of a functional symbiosis.

In conclusion, this work demonstrates the possibilities of functional genomics: creating and analyzing complete gap-less genomes, using the acquired data to find candidate genes for specific functions, and functionally characterizing newly identified genes *in vitro*.

Zusammenfassung

Das Feld “Functional Genomics” ermöglicht Einblicke über einfache Genotyp-Phänotyp-Beziehungen hinaus. Die vollständige Genom- und Transkriptom-Sequenzierung wird verwendet, um bestehende Ergebnisse mit Sequenzierungsdaten von neu sequenzierten Arten zu vergleichen und damit auf Funktionen zu schließen. Mit diesem Ansatz ist auch eine Analyse von Organismen möglich, die schwierig zu kultivieren sind oder für die keine Transformationsmethoden verfügbar sind. Dies gilt z.B. für Glomeromycota-Pilze, welche obligate Symbionten sind und in deren Hyphen Tausende von Kernen in einem gemeinsamen Zytoplasma enthalten sind.

Die hier vorgestellte Arbeit zeigt die verschiedenen Stadien der funktionellen Genomik in drei Teilen. Im ersten Teil wurde ein vollständiges und lückenloses Bakteriengenom von *Rhizobium leguminosarum* Norway mit verschiedenen Sequenzierungstechniken der 2. und 3. Generation erstellt. Dieses Assembly zeigt welche Vorteile die Verwendung einer Kombination aus kurzen und langen Sequenzier-Reads mit sich bringt, wie zum Beispiel die Möglichkeit lange “repeat regions” zu überbrücken, so dass sich diese Kombination auch gut auf größere Genome anwenden lässt.

Im zweiten Teil wurde das Transkriptom des früh divergierenden Glomeromycota-Pilzes *Geosiphon pyriformis* assembliert und analysiert. Die Ergebnisse der funktionellen Annotation der “non-redundant virtual transcripts” zeigen eine höhere Ähnlichkeit zu Asco-, Basidio- und Zygomycota als zu Glomeromycota. Dies deutet darauf hin, dass der unterschiedliche symbiotische Lebensstil von *G. pyriformis* im Vergleich zu anderen Glomeromycota auch ein verändertes Genexpressionsprofil erfordert. Andererseits legt die Abwesenheit der "Missing Glomeromycota Core Genes" in diesem Transkriptom nahe, dass es einen gemeinsamen obligat symbiotischen Vorfahren gab, abhängig entweder von Cyanobakterien oder Algen. Im Zuge dieser Analyse wurden drei Transkripte identifiziert, die strukturell Ähnlichkeiten zu Genen aufweisen, welche für Amyloidproteinen kodieren, und vielversprechende Kandidaten für die funktionelle Charakterisierung darstellen.

Der dritte Teil befasst sich mit der funktionellen Charakterisierung von Effektorkandidaten, die in *Rhizophagus irregularis* durch einen Genomvergleich identifiziert wurden. Unter Verwendung von "Host Induced Gene Silencing" wurde ein Kandidatengen herunterreguliert und zeigte in der Folge einen starken Phänotyp in Kolonisierung und der Form und Größe der Arbuskeln. Weitere Analysen zeigen, dass das von diesem Gen kodierte Protein an der Trehalose-Biosynthese beteiligt ist und kein Effektor ist. Der Phänotyp der Herunterregulation zeigt, dass die Trehalose-Biosynthese eine wichtige Rolle bei der Bildung oder Aufrechterhaltung einer funktionellen Symbiose spielt.

Zusammenfassend zeigt diese Arbeit die umfangreichen Möglichkeiten von “functional genomics” auf: Erstellung und Analyse vollständiger lückenloser Genome, Verwendung der gewonnenen Daten, um Kandidatengene für spezifische Funktionen zu finden, und funktionelle Charakterisierung dieser so identifizierten Gene *in vitro*.

List of Abbreviations

1stGS	First generation sequencing
2ndGS	Second generation sequencing
3rdGS	Third generation sequencing
AA	Auxiliary activities
aa	Amino acid
ABC	ATP-binding cassette
af	After filtering
AM	Arbuscular mycorrhiza
AMF	Arbuscular mycorrhiza fungus/fungi
AMT	Ammonium transporter
ANI	Average nucleotide identity
bf	Before filtering
BLAST	Basic local alignment search tool
bp	Basepair
CAzyme	Carbohydrate-active enzyme
CBM	Carbohydrate-binding module
CDD	Conserved domain database
cDNA	Complementary DNA
CDS	Coding sequence
CE	Carbohydrate esterase
COG	Cluster of orthologous groups
dbCAN	Database for automated Carbohydrate-active enzyme annotation
dNTP	Deoxynucleotide tri-phosphate
dsRNA	Double stranded RNA
EV	Empty vector
Gb	Gigabases
GC	Guanine and Cytosine
GFP	Green fluorescent protein – transformation marker
GH	Glycoside hydrolase
gsA-gsE	Genospecies A-E
GT	Glycosyl transferase
HD	Homeo domain
HIGS	Host induced gene silencing
HMG	High mobility group
HR4	Hairy root experiment 4
IRM	Intraradical mycelium
Kb	Kilobases
KOG	Eukaryotic orthologous groups
LECA	Last eukaryotic common ancestor
MATA-HMG	Mating type high mobility group
MAT-like	Mating type locus like
Mb	Megabases
MGCGs	Missing Glomeromycota core genes
mRNA	Messenger RNA
MST	Monosaccharide transporter
mya	Million years ago
NCBI	National Center for Biotechnology Information
NHGRI	National Human Genome Research Institute
NLS	Nuclear localization signal
NPC	Nuclear pore complex
nr	Non-redundant protein sequence database
NRVT	Non-redundant virtual transcript
nt	Nucleotide
PC1-3	Principal components 1-3
PCA	Principal component analysis
PCW	Plant cell wall
PIGS	Pipetting induced gene silencing
PPA	Pre-penetration apparatus
PT	Phosphate transporter
qRT-PCR	Quantitative real-time PCR
RCP	Repeats containing proteins
RI	*Rhizobium leguminosarum*
RNAi	RNA interference
RNS	Root nodule symbiosis
ROC	Root organ culture
rRNA	Ribosomal RNA
SBS	Sequencing by synthesis
SCR	Small and cysteine rich
SL	Strigolactones
SMRT	Single molecule real time (sequencing)
SMS	Single molecule sequencing
SNP	Single nucleotide polymorphism
SP	Signal peptide/secretion peptide
T6P	Trehalose-6-phosphate
TAG	Triacylglycerol
Tb	Terabases
TCDB	Transporter classification database
TPS	Trehalose-6-phosphate synthase
TPP	Trehalose-6-phosphate phosphatase
tRNA	Transfer RNA
WGS	Whole genome sequencing
wpi	Weeks past inoculation

Table of contents

1 Introduction

1.1 Functional genomics

Life is complex, but in the time of genomics, scientists have new possibilities to decipher it.

There are millions of different species living on our planet. These range from as small as a fraction of a micrometer to as large as a 30-meter blue whale or even about 10 km^2 clonal colonies of fungi[1]. These species range from unicellular to multicellular organisms with highly differentiated cell types. Despite their immense differences all living organisms have in common that their blue-prints are coded in DNA.

Functional genomics is a field that emerged with the evolution of sequencing techniques and the generation of a big publicly available database of genome sequences and analysis tools. It is a collective approach to shed light on the connections between DNA, RNA, proteins and their biological function.

In the classical candidate gene based approaches in biology, scientists focus mostly on only one gene or phenotype of a whole organism or ecosystem. These simple research approaches offer valuable results and have created huge amounts of knowledge in the past and present. The advantage of focusing on one gene or process is the possibility to formulate a sound hypothesis and the experiments are usually straightforward and offer clear results. On the other hand by focusing on only one gene/phenotype of a complex organism, important connections with other genes/processes may be neglected.

Furthermore while it was formerly only feasible to sequence and analyze species from cultivation-based methods, now environmental samples can be sequenced from uncultivable or low abundant organisms. This enables us to identify new species and study these under natural conditions[2]. In functional genomics a lot of information can be deduced by comparison. Thereby it is possible to use the extensive knowledge base that was built up over decades of study on model organisms and specific genes, proteins and their functions using traditional methods. In studying microbes the term of a pan-genome was forged[3]. This includes the “core genome”, genes that are shared between all strains, the “dispensable genome”, shared by two or more strains, and “singleton genes”, unique to an individual strain[3]. For some bacterial species mathematical modeling showed that even after sequencing the genomes of hundreds or thousands of additional strains, novel unique genes will be discovered[4].

Similarly, with comparative genomics it is also possible to find core genes for a specific conserved function in different organisms with the same life-style. For example, Delaux *et al.* compared genomes and transcriptomes of arbuscular mycorrhiza (AM) host and non-host plant species in order to find presence or absence of known core symbiosis genes. Their

findings suggest that AM fungi shape the genome evolution of the host plants and the loss of AM symbiosis in different plant lineages leads to convergent modifications of their genomes[5].
Another part of functional genomics is phylogenomics, the study of evolutionary relationships based on genome-wide data. Resolving phylogenetic trees was for a long time based on morphology of species or multiple alignments of single genes (e.g. rRNA genes). Phylogenomics can make use of genome data that is nowadays vastly available; for example Spatafora and his collaborators thereby resolved the phylogenetic classification of zygomycete fungi. In this approach 46 fungal genomes were analyzed using profile hidden markov models of a pan-eukaryotic protein set. Instead of using alignments of single proteins, the single alignments were concatenated into a single super matrix alignment and analyzed[6]. The advantage of an approach like this is that single genes that may be outliers in their evolution are averaged by combining several conserved genes.
The invention of new sequencing techniques together with the creation of higher computing capacity and applications to handle this big amount of data made all this new research possible.

1.1.1 Evolution of sequencing techniques creates basis for functional genomics analyses

In the beginnings of genome sequencing only a few organisms with small genomes and important model organisms were completely sequenced. These included between 1977 and 2000 a lot of "firsts": in 1977 the first ever DNA genome (bacteriophage PhiX174; genome size: ~5,375 bp)[7], in 1995 the first prokaryote (*Haemophilus influenza*; genome size: 1.8 Mb)[8], in 1996 the first eukaryote (*Saccharomyces cerevisiae*; genome size: 12.1 Mb)[9], in 1998 the first animal (*Caenorhabditis elegans*; genome size: 97 Mb)[10] and in 2000 the first plant (*Arabidopsis thaliana*; genome size: 115 Mb)[11] were fully sequenced.
One of the first large scale attempts of a holistic research approach was the human genome project. It was launched in 1990 and was the first initiative to sequence the complete human genome and the aim to deduce the function of all genes from this genetic information. In 2003, more than 10 years after the start of the project, the compilation of genetic information of the human genome was finished at a cost of somewhere between $500 million and $1 billion[12]. This project was mainly conducted with first generation sequencing techniques (Sanger), but during this time, research was conducted in order to improve old and invent new techniques[13].
Hand in hand with the improvements in micro-fabrication and high-resolution imaging, a stepping-stone was set for sequence-by-synthesis (SBS) based sequencing techniques. The so-called second generation sequencing (2ndGS, sometimes also called next generation

sequencing) techniques are mainly characterized by massively parallelized sequencing setups, thereby reducing time and cost per base sequenced. The cost of sequencing with these techniques was at $0.012 per Mb by July 2017. For the sequencing of a human genome, this would equate to $1,121 per genome[14].

The era of 2ndGS is mainly shaped by the sequencing platforms that are developed by Illumina (formerly Solexa). The technique used by Illumina relies on massively parallelized sequencing on a flow cell that is covered with a lawn of single stranded oligonucleotides. Single stranded DNA fragments that have adapters on both ends are distributed over this lawn. Here they can bind to the complementary oligo nucleotides, and are subsequently amplified via a solid phase (bridge) PCR. This results in clusters of DNA fragments of the same sequence on the flow cell. During the SBS step, fluorescently labeled dNTPs are incorporated. These dNTPs contain a terminator, so that the elongation process by polymerase is stopped after each incorporation event until the terminator is cleaved enzymatically. The fluorescent signal is documented in each cycle of incorporation (see Fig. 1). Sequencing the DNA fragments from both sides allows paired end sequencing. Knowing the size of the fragments and the sequences on both ends is a great advantage for the assembly process. With the NovaSeq 6000 platform from Illumina, paired end read length of up to 2x150 bp are possible. One run takes 16-36 hours with the maximum output being 6,000 Gb per run[15] (see Table 1).

More recently a third generation of sequencing (3rdGS) techniques has emerged. These techniques rely on the sequencing of single molecules (single molecule sequencing – SMS), which further reduces the time of library preparation and sequencing and removes any PCR biases from pre-sequencing PCR amplification steps. The two major players of 3rdGS are the single-molecule real-time (SMRT) technique developed by Pacific Biosciences, and nanopore sequencing developed by Oxford Nanopore Technologies. In SMRT technology the incorporation of single phospholinked dNTPs by polymerase is directly detected in a zero-mode waveguide. In contrast nanopore sequencing relies on the change of electric currents over a membrane. A nanopore is a very small hole with an inner diameter of about 1 nm. This is at the same scale as the width of DNA strands. Protein nanopores are common in nature and can be genetically adjusted to fit the specific purposes for sequencing.

The protein nanopores are placed into an electrically resistant polymer bilayer to which an electric field is applied and the characteristic nucleotide specific disruptions in the electric current are measured[16,17] (see Fig. 1). Oxford Nanopore announced on their website that Martin Smith and collaborators from the Kinghorn centre for clinical genomics (Garvan institute for medical research) were able to sequence more than one million bases in one continuous read using this technology[18]. The average error rate of nanopore sequencing is

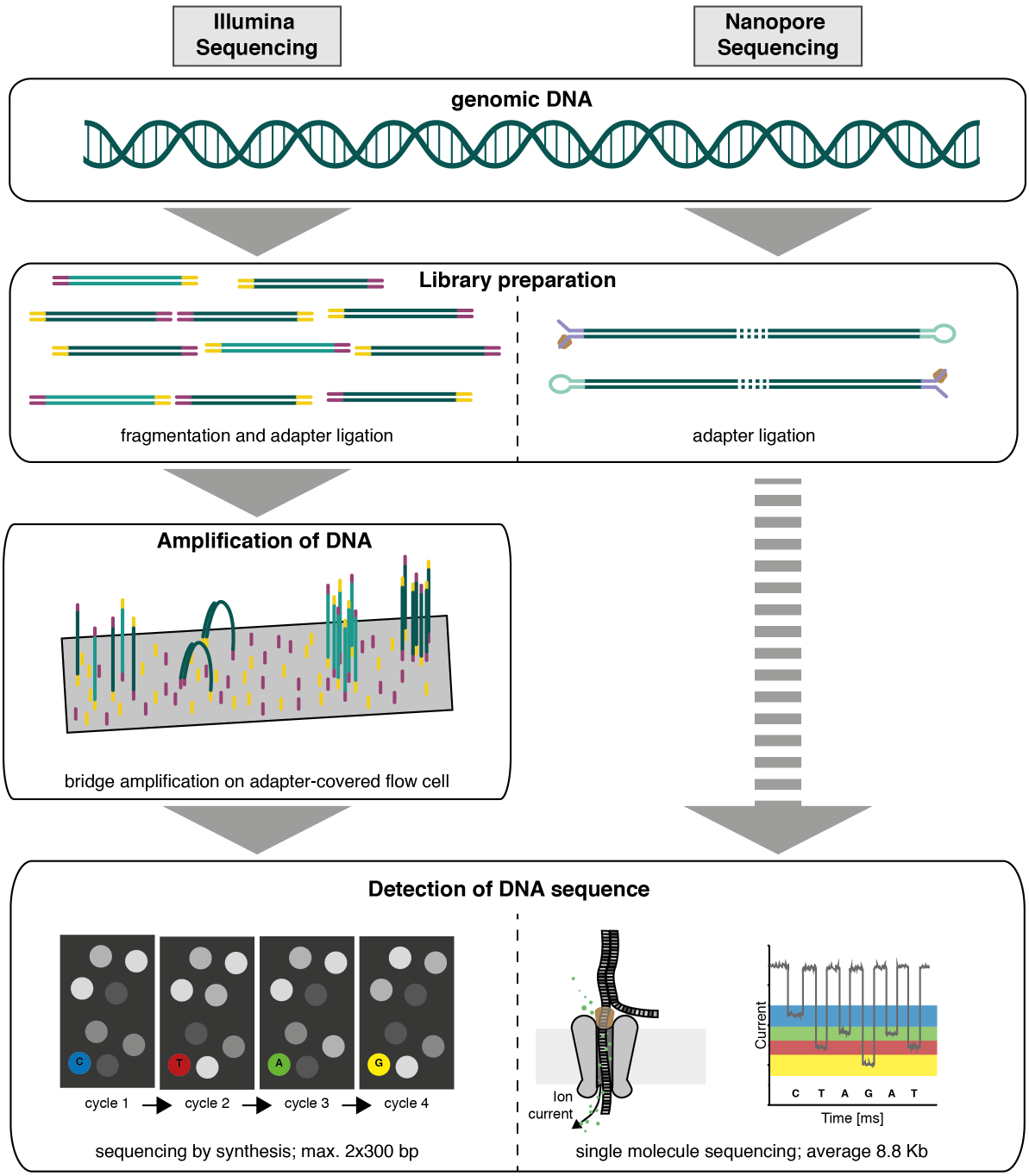

Fig. 1: Next generation sequencing techniques used in this work. Both Illumina and Nanopore sequencing can start from genomic DNA and proceed with library preparation. For Illumina sequencing the DNA is fragmented. Adapters are ligated to the DNA ends in both techniques. For Illumina sequencing an amplification step before sequencing is necessary. Detection of DNA sequence happens in a sequencing-by-synthesis manner in Illumina, and a single-molecule-sequencing by measuring changes in electric current (simplified representation) in Nanopore sequencing.

between 7-15%[19]. In this technique though, the base calling mistakes are unbiased and appear at random. This leads to a consensus accuracy of more than 99% by sequencing each molecule more than once (see Table 1).

General advantages of the 3rdGS are long read length that can span several genes and also long repetitive regions. These are difficult to assemble with short reads from 2ndGS. Furthermore, there is no requirement of PCR amplification steps before sequencing thereby reducing introduced PCR biases. An additional advantage is that epigenetic modifications

like methylation of single bases or abasic sites can be detected through different changes in the ionic current during nanopore sequencing[20,21]. At present the long reads of 3rdGS are mainly used to guide de novo assembly of 2ndGS reads as scaffold. By combining the two methods gap-less genomes can be generated easily and repetitive regions can be deciphered.

Table 1: Comparison of sequencing techniques. Sanger (1stGS), Illumina (2ndGS), SMRT and Nanopore (3rdGS) sequencing techniques[19,22-26].

	Read length [bp]	Cost	Run time	Accuracy	Other advantages/disadvantages
Sanger	700	$2.3/Kb	1-3h	99.9%	Impractical for whole genome sequencing
Illumina	2x50-600	$7-1,866 /Gb	4h-11d	75-80%>Q30 (>99.9%)	Paired reads enable scaffolding
PacBio (SMRT)	10,000-12,000	$180-300/Gb	<6h	85-89%	Accuracy of >99% when sequencing several passes of the same molecule; no PCR bias; reads spanning repetitive regions
Nanopore MinION	10,000	$170-2,300/Gb	1 min- 48 h (~10 Mb/h)	85-93%	Accuracy of up to 99.96% in consensus; no PCR bias; reads spanning repetitive regions

1.1.2 Computing capacities and bioinformatics applications

With the introduction of 2ndGS the capabilities of sequencing technology outcompeted the law of Moore regarding the price for sequencing. This law is usually addressed to computing capacity of microchips, which was predicted to double (measured by number of transistors per unit cost) approximately every two years[27]. The national human genome research institute (NHGRI) documented the costs of sequencing performed in their sequencing centers. With the introduction of 2ndGS in 2008 the cost of sequencing per Mb dropped drastically, from $102.13 in January 2008, to $0.52 two years later in January 2010 (see Fig. 2)[12]. Accordingly instead of doubling, the capacity of sequencing (measured by Mb per cost unit) grew by a magnitude of about 200 in these two years.

The advance in technology and reduction in cost per Mb (see Fig. 2, Table 1) resulted in a massive increase in sequencing data. Since 1982, the GenBank sequence database exists. By December 1982 only 606 (0.7 Mb) sequences were deposited in this database, but it has been growing ever since. In 1990 when the Basic Local Alignment Search Tool (BLAST)[28] was introduced, which can be used to search the database for similar sequences within seconds, there were already 41,057 (51.3 Mb) sequence entries. The number of bases in the database is doubling approximately every 18 months. And by April 2018 GenBank hosted

208.5 million sequences (260 Gb), and 2.8 Tb from whole genome sequencing (WGS) entries (see Fig. 2)[29,30]. By May 2018 a total of 79,448 genomes was sequenced and deposited at RefSeq, a non-redundant curated database. Of these publicly available genomes 50,406 are bacterial and 9,217 are fungal. This huge amount of data is publicly accessible allowing each researcher to search the database and compare the sequences they work on with already existing data.

For handling of such amounts of data and for creating whole genomes from short sequences complex algorithms were developed. CLC Genomics Workbench (Qiagen), a program for analyzing genomics sequencing data, uses de Bruijn graphs[31] to connect overlapping sequences by using sub-sequences of a certain length ("words"). With this algorithm it is difficult if not impossible to resolve repeat sequences that are longer than the used "words" or the sequence reads. This can be solved adding 3rdGS long sequencing reads as bridges on short read assembled repeat regions as for example with the command line program "Unicycler"[32].

After assembly a genome is nothing more than a succession of As, Cs, Gs and Ts. Only by running these sequences through an annotation process, the information about genes and their function can be deduced. For gene annotation the comparison of genomes and

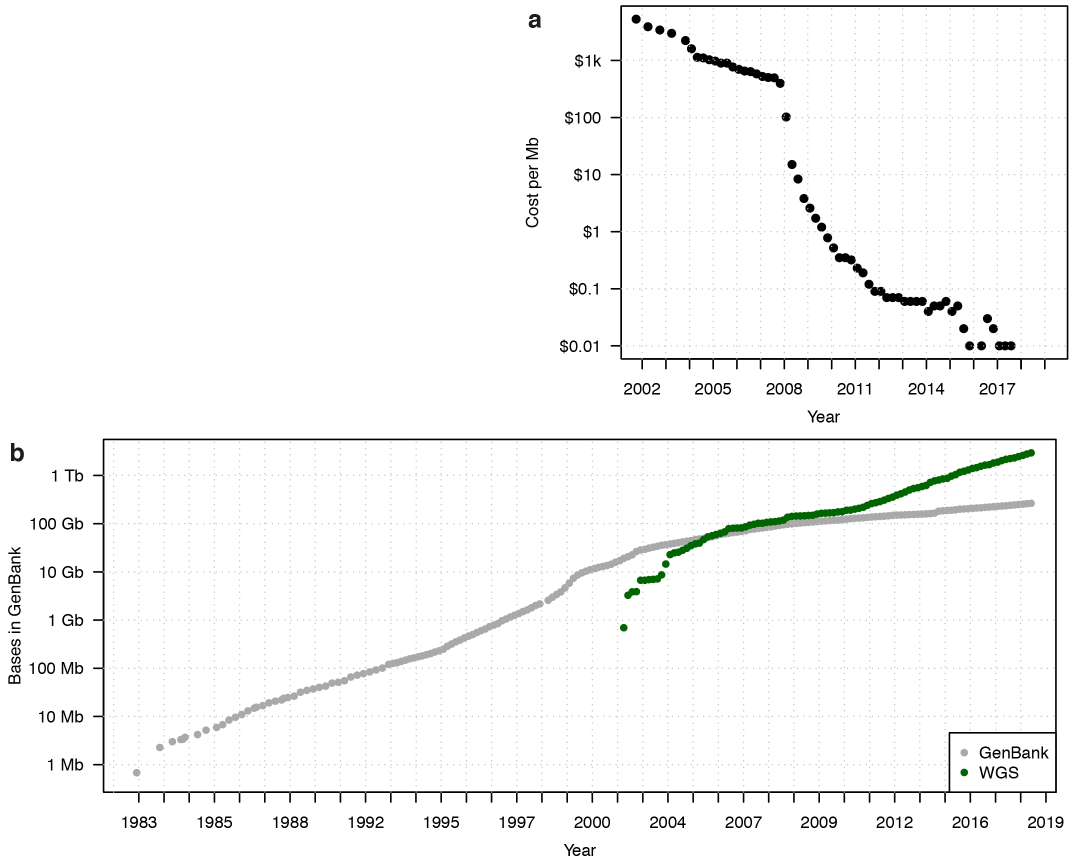

Fig. 2: Development of sequencing costs and sequence deposition in GenBank. (a) The cost of sequencing per Mb in $. (b) The amount of bases stored in GenBank from curated sequences (gray) and whole genome sequencing (WGS, green).

transcriptomes of the same organisms can be of great support. Robust annotation can also be obtained by using BLAST based programs such as Blast2GO[33] or online platforms such as MicroScope[34] by comparison of unknown sequences to already annotated genes from other organisms in the database. More insight into the functional organization of an organism can be acquired by comparing the sequences to existing data-sets for certain functions such as the transporter classification database (TCDB)[35] or the database for automated carbohydrate-active enzyme annotation (dbCAN)[36].

1.2 Plant-Microbe symbioses – the art of thinking independently together

A mutualistic symbiosis lifestyle describes the situation in which two organisms live together and each partner benefits from the other. When taking a closer look at the roots of a plant, we find a diverse community of organisms living in, on and around them (in the rhizosphere). Two key symbiotic players in the rhizosphere are mycorrhizal fungi and nitrogen-fixing bacteria.

1.2.1 Root Nodule symbiosis

During root nodule symbiosis (RNS) a group of diazotrophic bacteria known as Rhizobia colonize the roots of legume plants. Plants sense the presence of beneficial bacteria by detection of specific chemical compounds known as NOD factors, while the bacteria detect plant-secreted flavonoids. This specific cross talk is required for intracellular accommodation of the bacteria into plant roots. For the accommodation specialized organs, called nodules, are formed. Here, bidirectional nutrient exchange takes place between plants and bacteria[37]. The beneficial Rhizobia are able to fix atmospheric nitrogen and provide the plant with ammonia. In return the plant delivers carbon in the form of dicarboxylic acids to the bacteria[38]. However, this interaction is not always beneficial for the plant. So-called cheaters such as *Rhizobium leguminosarum* (*Rl*) Norway also exist[39]. These are bacteria that are able to initiate nodule formation and colonize them, but do not fix nitrogen. The genome organization of *Rl* is of one circular chromosome and several large plasmids. The number and size of the larger plasmids is varying between isolates[40,41].

1.2.2 Arbuscular mycorrhiza symbiosis

1.2.2.1 Evolutionary history

Arbuscular mycorrhizal fungi (AMF) of the phylum Glomeromycota[6] engage in an endosymbiosis with more than 80% of all land plant species[42], in total more than 200,000 plant species out of 280,000 species[43]. Up to date, there are about 300 species known belonging to the phylum of Glomeromycota indicating a very wide host range for these fungi[44]. It is hypothesized that mycorrhizal fungi played a key role in the terrestrialization and true root development of plants[45]. When plants first came to land, the habitats were of poor soils and lacking water. Possibly AM fungi enabled plants to adapt to these harsh conditions from the former habitats in sea or freshwater. The emergence of land plants happened earlier than the oldest fossil evidence of fungal colonization available, but there is evidence of fossilized glomales-like fungal spores and hyphae from the Ordovician of Wisconsin (460 million years ago (mya)). This supports the presence of mycorrhiza-like fungi at the time of land plant emergence[46]. Fossil evidence of fungi, which form arbuscule-like structures, exists

in land plants from the early Devonian period about 407 mya[47] These early land colonizing plants were rootless, but it was shown that AMF are also capable of colonizing a great number of extant rootless hornworts[48]. Furthermore it was shown that key symbiotic genes (*DMI1*, *DMI3*, *IPD3*) that are required for mycorrhiza formation, are present in basal liverworts and hornworts and are able to rescue the respective mutants in *Medicago truncatula*[49].

The modern AMF are obligate symbionts and can't complete their life cycle to form secondary fertile spores without a plant host[50]. AM symbiosis plays an important ecological role[51]. Many crop plants, like rice, maize, or soy can be colonized by AMF. The symbiosis increases crop yield and protects the plant from drought stress and some pathogens[52,53]. The plant alone is only able to take up nutrients from an area less than 1 cm around the root system[54] (see Fig. 3). This area is called the root depletion zone as it is quickly deprived of nutrients. In symbiosis with a fungus though the nutrient uptake, mainly phosphate and nitrate, can take place via the fungal hyphae network that expands further into smaller soil pores. In turn the fungus receives photosynthetically fixed carbon in the form of sugars and lipids from the plant.

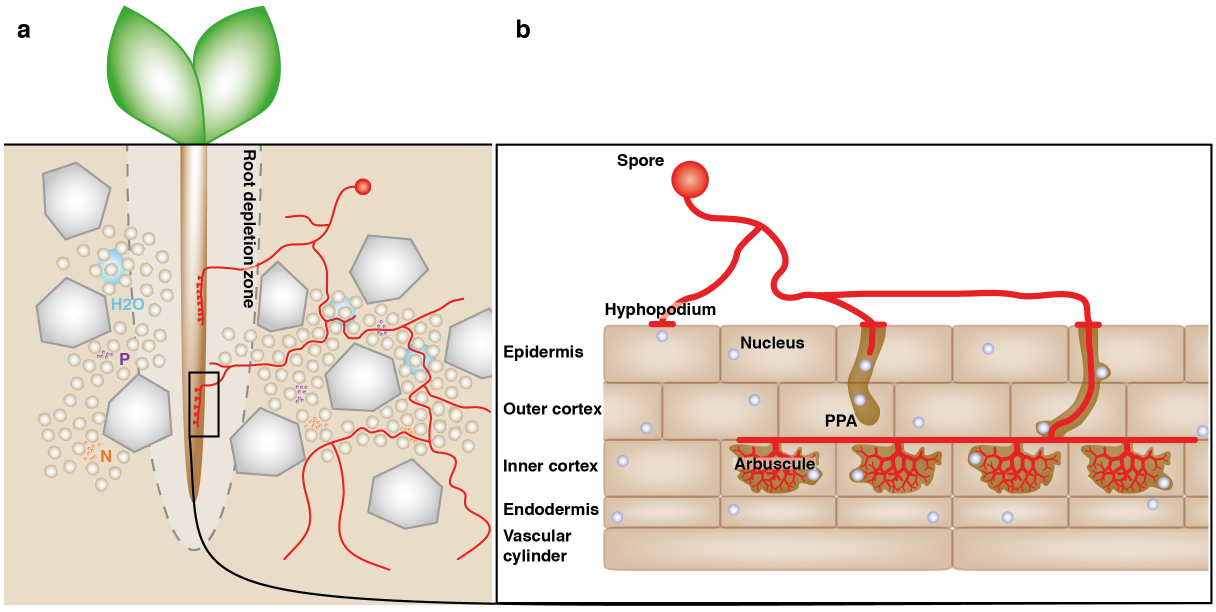

Fig. 3: Underground network and establishment of AM symbiosis. (a) The plant can take up nutrients in a small zone around the root (root depletion zone). For nutrient uptake deeper and in smaller pores of the soil, indirect uptake is facilitated via an AM fungus. (b) When fungal hyphae reach the root, hyphopodia structures are formed. The plant is forming a pre penetration apparatus (PPA) to allow fungal entry. When the fungus reaches the inner cortex of the root, arbuscules are formed.

1.2.2.2 Establishment of the symbiosis

For the establishment of symbiosis, a crosstalk between the plant and the fungus takes place. Plant root exudates, containing strigolactones (SL) can trigger the germination of fungal spores and hyphal branching[55]. Also AM fungi send signals to the plant, collectively called Myc factors. A diffusible molecule was found that activates transcription of symbiosis

related genes in *M. truncatula*, when AM fungi are grown in vicinity of the plant but separated by a fungus-impenetrable membrane[56].

Plants detect invading species by the recognition of non-self molecules, which leads to the induction of a local and systemic immune response[57]. For the establishment of a successful symbiosis the symbiont must either evade the recognition by the immune system or must suppress the immune responses. Many fungi secrete so-called effector proteins, for immune suppression[58]. The presence of effector candidate genes in AM fungi suggests that effectors might be similarly employed by AM fungi[59-61]. The conservation of effector candidates between different fungal species in the Glomeromycota further indicates an important role for the establishment of symbiosis[61].

Based on structural similarity, it was postulated that AM effectors should be secreted and have one or more of the following characteristics: (1) nuclear localization signal (NLS), (2) be small and cysteine rich (SCR; less than 150 bp, cysteine content >3%), (3) be repeats containing (RCP)[61]. Recent genome and transcriptome analyses allowed the prediction of the effector repertoire of several AMF. In the closely related species *Rhizophagus irregularis*[59,62] and *Rhizophagus clarus*[61], as well as in *Gigaspora rosea*[60], a repertoire of secreted proteins was identified via *in silico* pipelines.

So far only two effector proteins in AMF have been characterized in more detail. SP7 is a small, secreted protein that was identified in *R. irregularis* with an improved version of the yeast secretion sequence trap method[63]. Kloppholz *et al*. showed that SP7 is expressed in fungal tissue that is in contact with the plant and is localized to the nucleus when expressed *in planta* without the secretion signal. Furthermore SP7 suppressed defense responses associated with the plant protein ERF19, which is highly induced in the presence of the legume pathogen *Colletotrichum trifolii*[63]. The SL-induced putative secreted protein 1 (SIS1) in *R. irregularis* was identified in a screen of RNA-seq data of genes induced upon SL treatment or symbiosis. In a host induced gene silencing (HIGS; see Section 1.2.3, Fig. 6) approach, Tsuzuki *et al*. showed that downregulation of SIS1 expression leads to suppression of colonization and formation of stunted arbuscules[64].

After this presymbiotic stage, the fungal hyphae reach the plant root and form attachment structures, the so-called hyphopodia[55]. Four to five hours after hyphopodia formation, the plant cell actively forms a structure called prepenetration apparatus (PPA)[50]. During PPA formation the plant cell nucleus first moves to the expected point of fungal entry, and then moves ahead of the thick cytoplasmic bridge that is developing across the plant cell vacuole (see Fig. 3)[65]. Currently it is unknown if PPA formation is induced by any signal exchange. However, moving of the nucleus is also inducible by applying a needle to the epidermis, leading to nuclear movement towards the point of pressure application[66]. Nevertheless, the

disruption of gene expression of several plant symbiosis genes such as *SYMRK* or *CCamK* in *L. japonicus* (*DMI2*, *DMI3* in *M. truncatula* respectively) leads to the abolishment of PPA formation[65]. Upon completion of PPA formation the fungal hyphae can enter the host cell and grow towards the inner cortex of the plant root[65].

Once the fungal hyphae reach the inner cortex, they start growing intercellularly in the apoplastic space along the longitudinal axis of the root[67]. Additionally, the fungus grows into the host cells of the inner cortex and starts forming hyper-branched structures, so-called arbuscules (from Latin 'arbusculum', meaning bush or little tree; see Fig. 3)[68]. The fungal hyphae branch repeatedly in a dichotomous manner, creating a very large symbiotic interface. On the plant side, the arbuscule is surrounded by the plant-derived periarbuscular membrane. Between plant and fungal plasma membrane is the periarbuscular space, formed mainly of plant cell wall components (e.g. cellulose, expansins, xyloglucans)[69,70]. The thick and rigid fungal cell wall surrounding the spores becomes thinner and thinner in the intracellular structures[71]. In the smallest arbuscular branches it is only a thin amorphous structure[71]. In this expanded symbiotic interface a number of membrane transporters can be found on the plant and the fungal side to facilitate the nutrient exchange[72-75].

1.2.2.3 Nutrient transport and transfer

It is estimated that between 4% and 20% of the photoassimilates are transferred from the host to the fungus[76,77]. Within the plant, sucrose is transported via the phloem from the source tissues to the roots. Here, in the sink tissues, sucrose is converted to fructose and glucose by plant invertases[78]. Additionally, SWEET sugar exporters were characterized in potato and it was shown that they are differentially expressed upon AM colonization, suggesting a role in sugar transport towards the fungal partner[79]. SWEETs can transport sucrose to the periarbuscular space, where it is cleaved by cell wall bound invertases (see Fig. 4)[78,80]. On the other side of the periarbuscular space, two fungal monosaccharide transporters were so far identified. These are GpMST1, which transports glucose, mannose, galactose, and fructose with decreasing affinity[81], and RiMST2 which is expressed in arbuscules and internal hyphae and able to transport glucose, xylose, mannose, and fructose[75]. In the fungus the imported hexoses are rapidly turned over to glycogen and trehalose, presumably to buffer the intracellular concentrations of glucose[82,83].

The gained carbohydrates are transported to the extraradical mycelium and growing hyphal tips[84,85]. In the extraradical hyphae glycogen can also be transformed to trehalose, a non-reducing disaccharide consisting of two glucose moieties that are linked with an alpha,alpha-1,1 bond. This sugar is present in AM fungal spores and was found to be the main storage sugar of AMF[86]. Additionally to sugars also lipids that are produced in the plant are transferred directly to the fungus as was shown in a recent study[87]. This is necessary, as

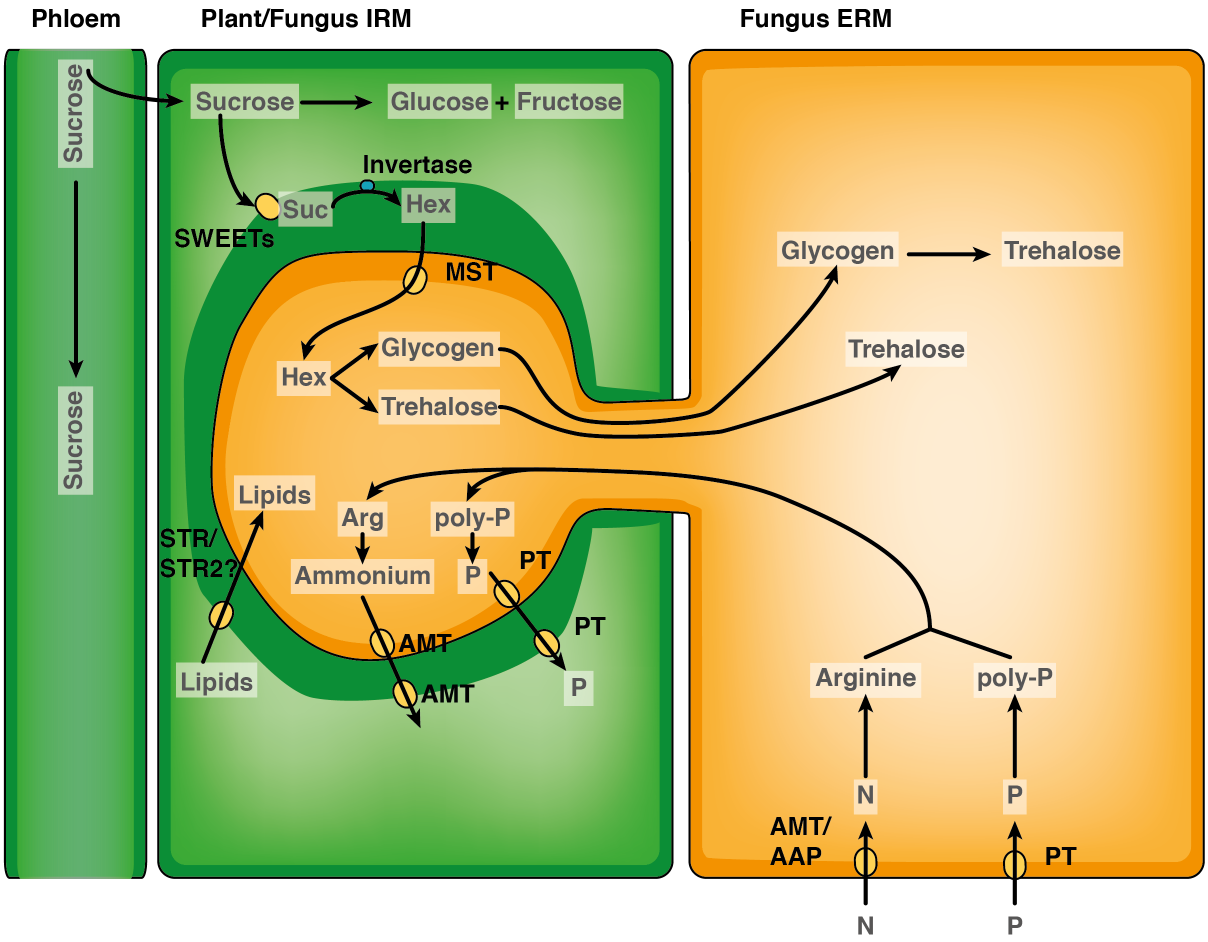

Fig. 4: Nutrient transfer and transport between the symbiotic partners. Different transport routes for carbohydrate from the plant and phosphate and nitrogen from the fungus exist. N: Nitrogen, AMT: ammonium transporter, AAP: amino acid permease, Arg: arginine, P: inorganic phosphate, PT: phosphate transporter, poly-P: poly phosphate, Suc: sucrose, Hex: hexoses, MST: monosaccharide transporter.

genomic mining of Glomeromycota genomes revealed that there is no de novo fatty acid synthase present in these fungi[60,88,89]. It was proposed that the transport of lipids towards the fungus is performed with the STR/STR2 complex in the plant membrane[90]. However so far this has not been experimentally verified (see Fig. 4).

In the large underground network of AM symbiosis, which is connecting plants with the hyphal network, there is also nutrient transfer towards the plant. In rice plants that are colonized by AMF it was shown that about 70% of overall acquired phosphorus is coming over the fungal route[91]. There are a few characterized fungal phosphate transporters e.g. GvPT from *Glomus versiforme*[74], GiPT from *R. irregularis*[92], GmosPT from *Glomus mossae*[93] and GigmPT from *Gigaspora margarita*[94]. These transporters are expressed either in the extraradical mycelium and/or in arbuscules. Within the fungus poly-phosphate is transported from the extraradical hyphae to the arbuscules where inorganic phosphate is released and transported to the periarbuscular space. Also in the plant symbiosis-specific phosphate transporters exist, such as MtPT4 from *Medicago truncatula*[73], LjPT4 from *L. japonicus*[95] or

OsPT11 from *Oryza sativa*[96]. Through these transporters the plant can take up phosphate from the periarbuscular space.

For the assimilation of the second major nutrient, nitrogen, the fungus can import nitrogen via amino acid permeases[97] or ammonium transporters[72,98,99]. Most likely nitrogen is then translocated as arginine to the intraradical mycelium (IRM), where it is converted to ammonium via the urea cycle[100,101]. Ammonium transporters on the plant side of the symbiotic interface were shown to localize to the branch domain of the periarbuscular membrane. This indicates that the active transfer of ammonium is taking place in this area[102,103].

1.2.2.4 Sexuality of arbuscular mycorrhiza fungi

All hyphae of AM fungi contain a common coenocytic cytoplasm that harbors thousands of nuclei and the propagation units, the spores, still contain hundreds of nuclei[104]. Sexual reproduction (i.e. karyogamy, meiosis) has never been observed, but the exchange of nuclei is possible through anastomosis (fusion of hyphae) between closely related strains[105]. The absence of sexual reproduction together with the long history of obligate biotrophy in which the genome of AM fungi stayed stable in the absence of selection imposes key questions[106]. One main question deals with genome organization of AM fungi. Are AM fungi heterokaryotic, meaning that they harbor genetically divergent nuclei in their cytoplasm thereby counteracting deleterious mutations in genes with copies of intact alleles present in alternative nuclei? Or are AM fungi homokaryotic with each hyphae containing genetically identical nuclei. A recent study showed that the size of single nuclei from different isolates suggests that their nuclei are haploid[107]. Single nucleotide polymorphism (SNP) analysis from genomic sequencing further suggests that some isolates contain two dominant divergent haploid nuclei (dikaryon), whilst others contain only one population of haploid nuclei (homokaryon)[108].

In several other genomic approaches a set of meiosis genes[109], a vast expansion of high-mobility group (HMG) coding genes[110] and MAT-like mating type loci[108] could be found in Glomeromycota species. HMG genes encode transcription factors known to be involved in mating of other fungi. These genes are usually found in the so-called mating type locus, together with e.g. homeodomain (HD), pheromones and pheromone receptor encoding genes[111]. Citing Riley *et al.* (2014): " […] a fungal MAT locus is a genomic region found in most fungi that serves the common purpose of determining sexual compatibility between two individuals of outcrossing species, or a region required for sexual differentiation in self-fertile species"[110].

More recently a putative AMF mating-type locus was identified in *R. irregularis*[108]. This locus has two HD containing genes that are transcribed in opposite direction. Analysis of single

nuclei of different isolates showed that there are on the one hand dikaryon-like isolates containing two sets of nuclei with different alleles in this locus and on the other hand homokaryons, in which all nuclei contain the same alleles in this locus. Between the six isolates, five different sets of HD alleles were identified. Together, these findings suggest that these long thought ancient asexual organisms might well have some sort of cryptic sex or parasexual processes, which enables the exchange of genetic material between individuals.

1.2.3 Two model systems, one symbiotic interface

In this study two model systems were used for studying the symbiosis between a photoautotroph partner and a glomeromycotan fungus (see Fig. 5). These two model systems have certain similarities, but also differences that provide advantages for certain experimental setups. The first system is a symbiosis between *Geosiphon pyriformis* as fungal macro-symbiont with *Nostoc punctiforme* as cyanobacterial micro-symbiont. The second is

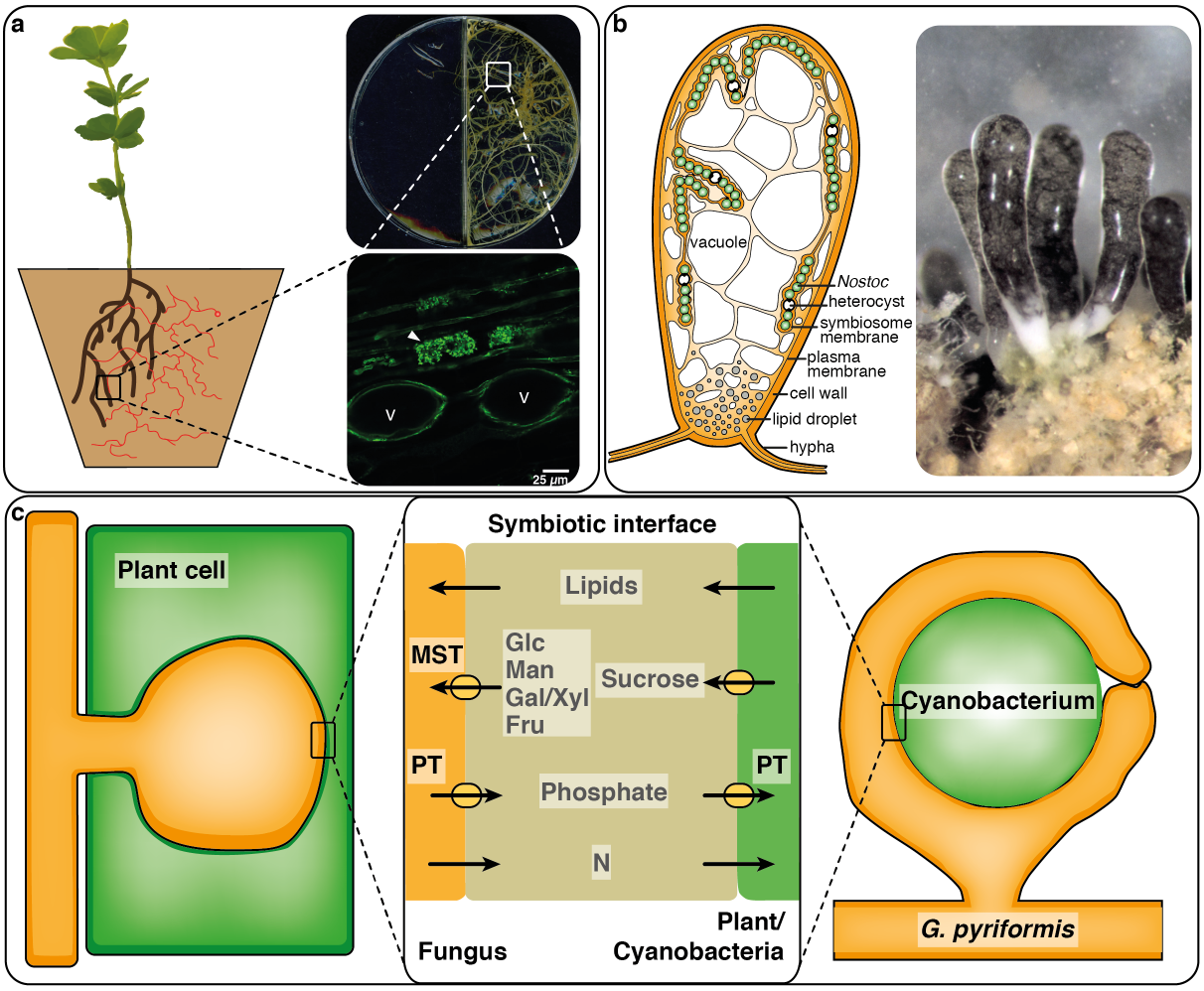

Fig. 5: Two model systems for Glomeromycota interaction with a photoautotroph symbiotic partner. (a) Pot culture with *L. japonicus* and *R. irregularis* or root organ culture on plates with carrot roots. In both cases intraradical structures as arbuscules (arrowhead) and vesicles (v) are formed (WGA-stained colonized *L. japonicus* root). (b) Schematic representation of a *G. pyriformis* symbiotic bladder with *N. punctiforme* filaments and photograph of the same (Figure adapted with permission A. Schüßler pers. communication). (c) Symbiotic interface in AM and *Geosiphon-Nostoc* symbiosis has the same structure and function. The symbiotic interface contains fungal and plant/cyanobacterial cell wall components.

Rhizophagus irregularis as fungal micro-symbiont with *Lotus japonicus* as plant macro-symbiont.

G. pyriformis is a early diverging member of the Glomeromycota[112]. So far it has not been observed that this fungus is able to perform arbuscular mycorrhizal symbiosis with plant roots. However, a photosynthetic prokaryote is taken up by fungal hyphae via endocytosis, which subsequently swell and form so called “bladders”. These structures can grow up to two millimeters out of the soil. The engulfed cyanobacterial *Nostoc* filaments can grow and multiply in the bladders and can fix carbon photosynthetically[113]. In the classical AM symbiosis, which is performed by *R. irregularis*, the photosynthetic partner is the model legume *L. japonicus*. Here the micro- and macro-symbiont-roles are reversed and both partners are eukaryotes. Despite these differences, the symbiosis itself appears to be similar. The symbiotic interface in the *Geosiphon-Nostoc* symbiosis is built by the fungal plasma membrane, a thin and amorphous fungal cell wall that contains chitin, the *Nostoc* cell wall and the *Nostoc* plasma membrane. Also in the AM symbiosis the symbiotic interface is made of the fungal membrane, a thin fungal cell wall, a space containing plant cell wall components and the plant plasma membrane. In both types of symbiosis, this interface is structurally the same and the main point of nutrient exchange. Here the cyanobacteria, or the plant, trade the produced sugar for nutrients as phosphate and water. The striking similarity of this symbiotic interface suggests that many molecular aspects, e.g. transcription of symbiosis specific genes, must be the same in the early diverging glomeromycotan fungus *Geosiphon* and its relative *Rhizophagus*. For transcriptomic experiments however, it is difficult to retain pure fungal RNA from the symbiotic stage within the plant root in the case of *R. irregularis*. In the case of *G. pyriformis,* the fungal and bacterial transcripts can easily be separated, by taking advantage of the poly-adenilation of eukaryotic transcripts. A disadvantage of studying *G. pyriformis* is a current lack of cultured material and it also cannot be found any longer in its last known habitat (pers. communication Arthur Schüßler).

1.2.3.1 Current cultivation and experimental methods and available data

Due to the obligate biotrophic nature of AM fungi, cultivation always has to occur with a host plant in pots or with so-called root organ cultures (ROC) on plates. The latter system uses *Agrobacterium rhizogenes* transformed carrot (*Daucus carota*) or chicory (*Cichorium intybus*) roots and enables access to pure fungal material. As AM fungi have a coenocytic cytoplasm, containing thousands of nuclei, it is so far impossible to stably transform or create mutants of them.

Lotus japonicus is a model legume plant that is used as a host to study the in vivo interactions between AM fungi and plants. In this system *Agrobacterium rhizogenes* mediated methods are available to transform host plant roots. Thereby it is possible to e.g.

overexpress putative fungal effector genes to study their function *in planta* on colonization levels or phenotypes. Also a method called host induced gene silencing (HIGS) has been described in two cases for studying the function of *R. irregularis* genes[64,75]. This method can be used to down-regulate fungal genes. The host roots are transformed with an RNAi construct that targets the fungal genes. This construct is transcribed and processed into small double stranded RNAs (dsRNAs) *in planta* and then translocated via an unknown process to the fungus at the interaction sites. In the fungus the targets are then silenced by the translocated dsRNAs (see Fig. 6).

In order to functionally analyze fungal genes the use of heterologous systems is necessary. For cellular localization studies tobacco (*Nicotiana benthamiana*) leaves can be transiently transformed with the genes of interest fused to a fluorophore and analyzed by microscopy. Yeast mutants and expression strains offer a suitable system for functional characterization.

The first glomeromycotan genome was published in 2013[59]. This genome of the model fungus *R. irregularis* DAOM 197198 is 101 Mb big and contains 30,282 predicted genes. For the same fungal isolate also a sequenced transcriptome is available which is 20.3 Mb and contains 25,906 non redundant virtual transcripts (NRVTs)[114]. Furthermore two transcriptomes of Gigasporacea are sequenced and available, namely *G. rosea* (55.5 Mb,

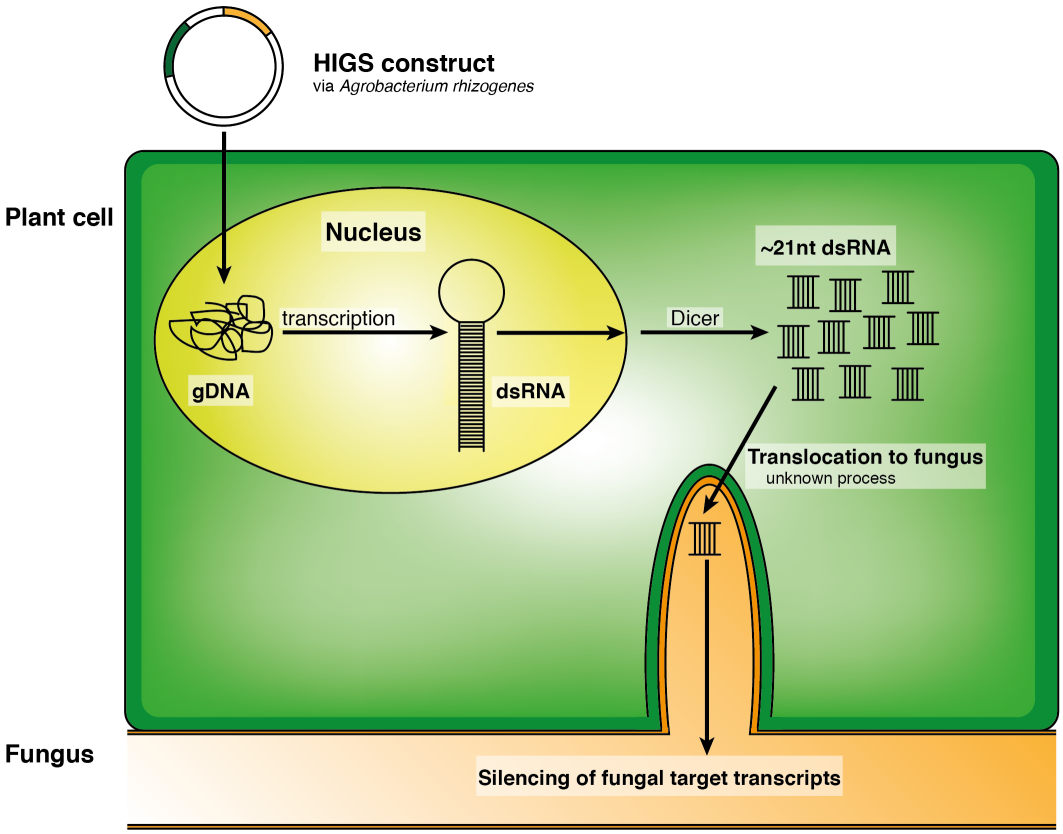

Fig. 6: Schematic overview of the host induced gene silencing (HIGS) method. The plant is transformed with a HIGS construct by *A. rhizogenes* mediated transformation. dsRNA is transcribed in the plant nucleus and processed by Dicer to small dsRNA (~21 nt). Translocation to the fungus at interaction sites is happening via an unknown process. In the fungus the target transcripts are silenced.

86,332 NRVTs)[60] and *G. margarita* (51.1 Mb, 86,108 NRVTs)[89]. For *Rhizophagus clarus* an effector candidate repertoire was predicted from a draft genome[61]. All of these available genome and transcriptome data can be used for comparison and function prediction in newly sequenced genomes and further for functional characterization of promising repertoires like the effector candidates.

1.3 Goals of the thesis

In this thesis I present one example for each of the three stages of functional genomics:

(1) Sequencing and assembly of a complete genome

For the first part I used 2ndGS (Illumina) and 3rdGS (Nanopore) in order to sequence the genome of the bacterium *Rhizobium leguminosarum* Norway. Using a hybrid assembly algorithm, the sequencing reads of both techniques could be combined and resulted in a gap-less complete genome

(2) Genome or transcriptome annotation and analysis

In the second part the symbiotic transcriptome of *Geosiphon pyriformis* was analyzed and compared to already sequenced glomeromycotan transcriptomes. With these analyses a set of genes could be found that are promising candidates for future functional characterization.

(3) Functional characterization of candidate genes *in vivo* and in heterologous systems

The work of Sedzielewska Toro *et al*. was used as a basis for the third part. Here the genomes of *Rhizophagus clarus* and *Rhizophagus irregularis* have been used to identify conserved fungal effectors. Several of these were further functionally characterized, resulting in a deeper insight into fungal trehalose metabolism.

2 Results

2.1 Assembly and annotation of the *Rhizobium leguminosarum* Norway genome

Rhizobium leguminosarum (*Rl*) Norway was identified as a so-called "cheater" strain, inducing ineffective nodules on a wide range of hosts, i.e. nodules in which little or no nitrogen fixation occurs[39]. To investigate the basis of this phenotype, a whole genome Illumina sequencing approach had been conducted (about 15 million paired 150 bp reads), resulting in an assembly with a total length of 7,765,030 bp in 83 contigs, with an average coverage of 309x (pers. communication, Macarena Marín). This amount of contigs is quite common for bacterial genomes assembled from the comparably short Illumina sequencing reads, due to unresolved regions likely from repetitive sections including repeats, pseudogenes, transposable elements, or rRNAs.

In order to obtain a complete gap-less genome of this strain, additional long-read 3rdGS was performed with Oxford Nanopore MinION. This resulted in an additional 180,000 long Nanopore reads with an average length of about 6,000 bp. The command line program Unicycler was used to perform a hybrid assembly of these reads together with the above-mentioned Illumina reads. Unicycler first uses the SPAdes algorithm to create a short read assembly of the provided Illumina reads. A range of k-mers ("words") is used to identify the best assembly with the lowest number of contigs (here: 194) and dead ends (here: 22) (see Table 2, Fig. 7a).

Table 2: SPAdes assembly of short reads with different K-mer length. The score function is 1/(c*(d+2)), where c is the contig count and d is the dead end count.

K-mer	Contigs	Dead ends	Score
27			Too complex
47	938	39	2.60×10^{-5}
63	446	22	9.34×10^{-5}
77	326	22	1.28×10^{-4}
89	269	22	1.55×10^{-4}
99	227	22	1.84×10^{-4}
107	215	22	1.94×10^{-4}
115	203	22	2.05×10^{-4}
121	197	22	2.12×10^{-4}
127	194	22	2.15×10^{-4} ← best

With this primary assembly Unicycler performed various cleaning procedures to remove overlaps and simplify the structure (see Fig. 7b). The long sequencing reads (Nanopore) where subsequently used for bridging to resolve repeat regions (see Fig. 7c). For the final polishing of the genome, all short reads were mapped against the bridged assembly (see

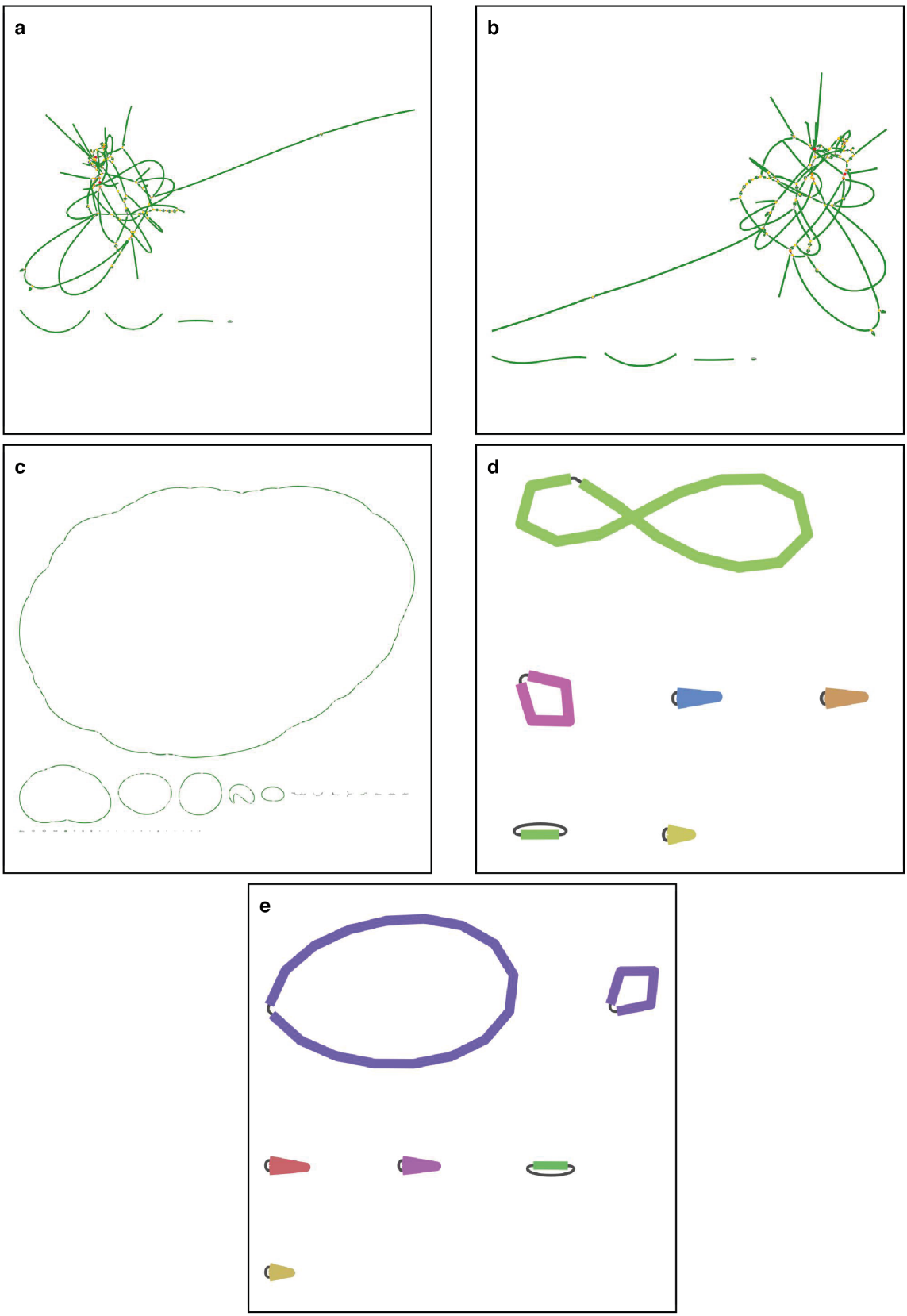

Fig. 7: Assembly graphs of the different steps in the Unicycler assembly of the *R. leguminosarum* Norway genome. (a) SPAdes short read assembly. Unresolved regions are apparent at yellow, orange and red bridges.(b) After removal of overlaps and simplification of the graph there are still many unresolved regions and dead ends present. (c) Bridging with long reads (Nanopore) already results in six circular sequences. (d) Mapping of all short reads against bridged assembly results in final circular sequences. (e) Definition of start position of circular contigs by rotating starting gene to zero position on forward strand.

Fig. 7d). This removes errors that may have been introduced by less accurate long reads. Using Unicycler the start position of the circular contigs were defined by searching for a starting gene (*dnaA* or *repA*) and rotating the contig to start with that gene on the forward strand (see Fig. 7e).

This hybrid assembly approach resulted in the complete gap-less genome of *Rl* Norway of 7,788,085 bp (average base coverage of 380x), distributed on a circular chromosome containing 63% of the genomic information and five large circular plasmids ranging from 280-1,098 Kb (see Fig. S1). The annotation of the complete genome was performed using the online tool MicroScope[34].

Table 3: Genome statistics of *Rl* Norway.

Replicon	Size [bp]	GC content	Protein encoding genes	Proportion coding sequences	Mean protein length [aa]	rRNA operons	tRNA genes
Chromosome	4,906,123	61.0%	5,045	87.6%	284	3	54
pRLN1	1,098,158	60.5%	1,079	90.8%	308		
pRLN2	592,529	60.9%	595	88.9%	295		
pRLN3	557,386	57.4%	570	83.5%	272		
pRLN4	354,350	60.7%	312	89.0%	337		
pRLN5	279,539	61.3%	265	90.2%	317		
Total	7,788,085	60.3%	7,866	88.3%	302	3	54

The chromosome contains three identical rRNA operons and 54 tRNA genes, none of which is found on any of the five plasmids (see Table 3 and Fig. S1). In total 7,866 protein-encoding genes were identified. BUSCO analysis[115] confirmed complete presence of the core bacteria dataset, which consists of 148 genes. All genes from the BUSCO core bacteria dataset are located to the chromosome, with only few additional gene duplications on the plasmid replicons. 6,016 of the protein coding genes have a functional class description (COG). Most are of the class “general function prediction only” (11.5%), followed by “amino acid transport and metabolism” (10.6%) and “carbohydrate transport and metabolism” (9.0%). The six replicons have a comparable mix of functional classes (see Fig. 8a).

For genome comparison *R. leguminosarum* bv. *viciae* 3841 (*Rlv* 3841) was used, a well characterized nitrogen fixing *Rl* strain. The genomes of *Rl* Norway and *Rlv* 3841 have a very similar relative occurrence of genes encoding functional proteins (see Fig. 8b). *Rl* Norway contains more protein encoding genes than *Rlv* 3841 (7,866 vs. 7,263 genes), but the number of genes for which a functional annotation could be retrieved is almost identical (6,106 vs. 6,105 genes). The major difference is the number of functionally not classifiable genes (1,760 vs. 1,158 genes).

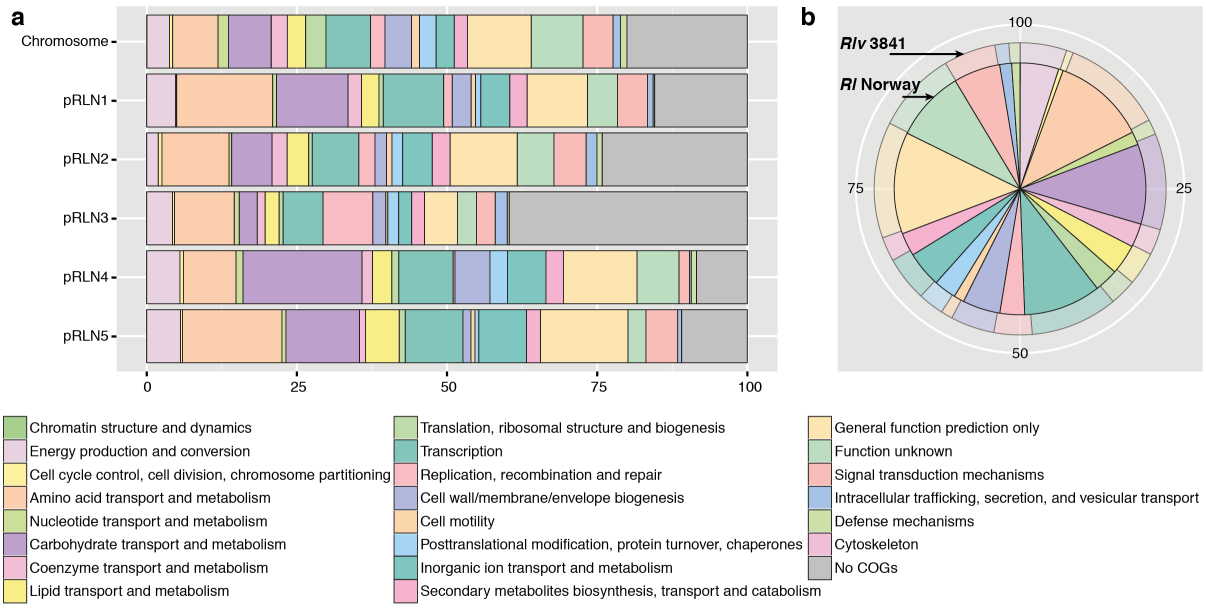

Fig. 8: Distribution of functional classes of protein encoding genes within the *RI* Norway genome. (a) Functional class distribution across the six *RI* Norway replicons. (b) Comparison of the relative occurrence of functionally classified protein encoding genes between the *RI* Norway and *RIv* 3841 genomes. Functional annotation (COG) was performed using COGNiTOR.

Comparison of the replicons of *RI* Norway and *RIv* 3841 showed high similarity between the chromosomes of both strains, also large parts of pRLN1 and pRL12, pRLN2 and pRL11, pRLN4 and pRL9, and pRLN5 and pRL10 show similarities. pRLN3 does not show large segments similar to any of the replicons of *RIv* 3841 (see Fig. 9).

For genospecies classification, the *RI* Norway genome was compared to representatives of the five proposed genospecies (gsA-gsE)[116]. The two highest average nucleotide identity (ANI) scores (*RI* CC278f: 96.34%; *RI* SM51: 95.59%) were found with members of the genospecies gsD. All other comparisons resulted in ANI scores below 95% (see Table 4). The ANI score between *RI* Norway and *RIv* 3841, which belongs to gsB, is only 93.26%.

The nucleotide sequences of all replicons have been deposited at GenBank under the accession numbers CP025012-CP025017.

Table4: Genome comparison of *RI* Norway with members of the five genospecies and respective ANI scores.

	Norway vs	One-way ANI 1	One-way ANI 2	Two-way ANI
(gsA)	WSM1325	93.45%	93.52%	93.70%
gsB	3841	93.01%	93.06%	93.26%
gsC	TA1	93.75%	93.80%	93.94%
gsD	SM51	95.40%	95.40%	95.59%
(gsD)	CC278f	96.11%	96.19%	96.34%
gsE	128C53	94.66%	94.75%	94.84%

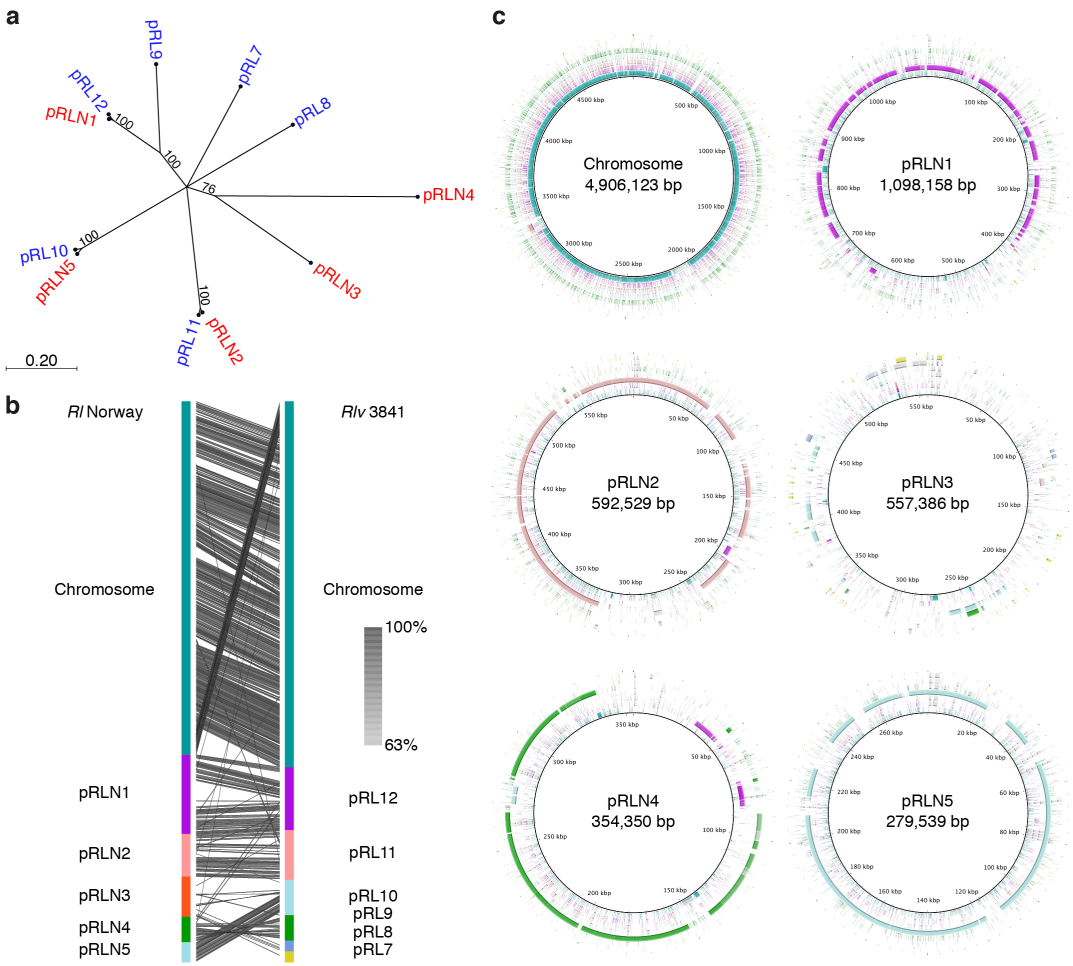

Fig. 9: Genome comparison between *RI* Norway and *Rlv* 3841. (a) Neighbor-joining tree of rep proteins from both strains. Protein sequences for repA, repB, and repC from the individual plasmids were aligned and the resulting alignments concatenated for analysis. *RI* Norway proteins are depicted in red, *Rlv* 3841 proteins in blue. Bootstrap values indicated on the nodes strongly support the relations between pRLN2 - pRL11, pRLN5 - pRL10, and pRLN1 - pRL12. Only bootstrap values >70% are depicted. Branch lengths are given in terms of expected numbers of substitutions per nucleotide site. (b) For whole genome comparison the sequences of chromosome and plasmids were concatenated for *RI* Norway and *Rlv* 3841 and compared with BlastN in Easyfig 2.2.2. Levels of sequence identity are indicated by different shades of grey. (c) Gene contents comparison between the two strains. Depicted are the *RI* Norway replicons and their respective homologous regions from the *Rlv* 3841 replicons. Plasmid maps were generated using BRIG. Colors in the rings are the same as for the *Rlv* 3841 replicons in (b).

2.2 Analysis of the *Geosiphon pyriformis* symbiotic transcriptome

2.2.1 Assembly of the *G. pyriformis* transcriptome

To obtain a *G. pyriformis* transcriptome, Illumina sequencing was performed on extracted cDNA inserts from a *G. pyriformis* expression library[81], resulting in almost 100 million primary reads (see Fig. S2a). After quality trimming and removal of known contaminants the remaining ca. 62 million reads were assembled into 47,645 NRVTs (non-redundant virtual transcripts) with a minimum length of 200 bp. This initial assembly ("*G. pyriformis* before filtering"; average length: 985 nt; N50: 5,818 nt; GC content: 52%; total length: 46.9 Mb; average coverage: 137.5x) encompassed a total of 58,927,166 primary reads (see Fig. S2a). Blobplot analysis[117] of the initial assembly revealed high contamination with sequences from other species, especially bacteria (see Fig. S2b). Subsequent filtering of NRVTs with best blastn[28] hits under non-stringent conditions (e-value < 10) against archeabacteria, bacteria, plants, and viruses led to the removal of 23, 5,011, 610, and 62 NRVTs, respectively. The known Glomeromycota transcriptomes have a very low mean GC content of around 32% (see Table 5). Accordingly, 16,103 NRVTs with a GC content above 55% were also removed (see Methods). Later analyses revealed that 3,899 NRVTs with a GC content of about 40% belonged to three different repetitive primary transcripts from *G. pyriformis* (see below) and the associated primary reads were thus merged into three NRVTs. For the submission of the transcriptome data to NCBI all NRVTs shorter than 200 bp were removed and further NRVTs

Table 5: Comparison of available Glomeromycota genome and transcriptome datasets. *Ri*: *R. irregularis* (Genome[59], predicted transcriptome[59], transcriptome[114]), *Gr*: *G. rosea*[60], *Gm*: *G. margarita*[89]

Organism	Dataset	Number of reads	Number of initial NRVTs	Number of final NRVTs after filtering	Average length of NRVTs [bp]	N50 [bp]	GC [%]	Total length	Origin of sample
Ri	Genome	30,067,076	28,371	28,371	NA	15,160	28	101	Pure sterile spores
Ri	Predicted transcriptome	NA	NA	30,282	890	1,234	33	27	NA
Ri	Transcriptome	437,411	25,906	25,906	773	984	31	20.3	Mix of germinated spores, extraradical mycelium (with *Dc*, *Mt*), and laser-microdissected arbuscule-containing cells
Gr	Transcriptome	230,084,705	97,463	86,332	643	948	33	55.5	Mix of non-symbiotic pure fungal reads and cleaned symbiotic reads(with *Mt*)
Gm	Transcriptome	NS	86,183	86,108	1,186	1,684	32	51.1	Mix of quiescent spores, germinating spores, spores treated with SL, and extraradical mycelium
Gp	Transcriptome	24,450,538	47,645	17,604	705	1,110	41	12.4	Symbiotic bladders and attached mycelium

* NA, not applicable; NS, not stated; *Ri*, *R. irregularis*; *Gr*, *G. rosea*; *Gm*, *G. margarita*; *Gp*, *G. pyriformis*; *Dc*, *Daucus carota*; *Mt*, *M. truncatula*

identified by the automatic screening of NCBI upon upload had to be removed. Examination of the discarded NRVTs supported the elimination of mainly bacterial genomic sequences and artificially multiplied NRVTs from the *G. pyriformis* transcriptome (see Fig. S2c).

After these filtering and removal steps a total of 17,604 NRVTs remained (average length: 705 nt; N50: 1,110 nt; GC content: 41%; total length: 12.4 Mb; average coverage: 49x), encompassing around 33.5% of the primary cDNA library insert reads (see Fig. 10a).

This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GGEK00000000. The version described in this work is the first version, GGEK01000000.

In blobplot analysis, more than 90% of these NRVTs did not show similarity to known DNA sequences (blastn, e-value < 10) outside of the Glomeromycota (see Fig. S2d). Comparisons on the amino acid level using blastx[28] revealed 64% NRVTs with matches against database sequences (e-value $1e^{-5}$), of which 8% were best hits to the *R. irregularis* genome[59]. All best matches are either to eukaryotic sequences or to sequences of unknown origin, indicating efficient removal of contaminating sequences of bacterial origin.

In comparison to the other available glomeromycotan transcriptomes (see Table 5) a lower number of obtained NRVTs and accordingly a smaller total length is apparent. Furthermore *G. pyriformis* has a higher GC content (41% vs. 31-33%; Table 5), supporting the early divergence of other Glomeromycota and subsequent separate evolution.

2.2.2 *G. pyriformis* symbiotic transcriptome is valid for comparative analysis

The completeness of the initial and the final *G. pyriformis* transcriptome assemblies were controlled by comparison against a set of 290 fungal BUSCOs[115] in relation to the three published glomeromycotan transcriptomes (see Fig. 10b).

In the original assembly before filtering (bf, 47,645 NRVTs) 181 complete BUSCOs could be identified; after filtering (af, 17,604 NRVTs) the number of complete BUSCOs was reduced to

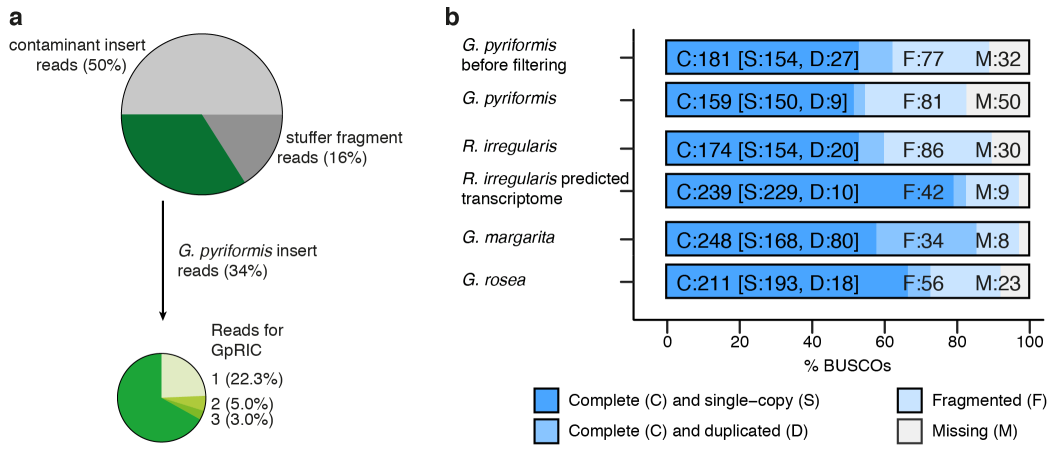

Fig. 10: *G. pyriformis* read distribution and comparison of BUSCO results. (a) Distribution of all sequenced insert reads of cDNA library. (b) BUSCO results for the datasets used in this analysis.

159. Of these complete BUSCOs, the duplicated BUSCOs were reduced by 18 (bf: 27, af: 9), and there were only four less single-copy BUSCOs (bf: 154, af: 150). The number of missing BUSCOs increased from 32 to 50.

In the *R. irregularis* transcriptome the number of complete and single-copy BUSCOs is comparable to that of *G. pyriformis* (154 vs. 150), but there are more fragmented and duplicated BUSCOs in *R. irregularis* (106 vs. 90), resulting in a lower number of missing BUSCOs (30 vs. 50). Both transcriptomes are not complete, when compared to the predicted complete *R. irregularis* transcriptome, although even here 9 BUSCOs are missing. The transcriptomes of *G. rosea* and *G. margarita* both appear to have a higher coverage of BUSCOs, but they also contain a much higher number of NRVTs.

2.2.3 Allele frequency distribution of single nucleotide polymorphisms (SNPs) suggests homokaryosis

To further establish the accuracy of the transcriptome and rule out increased numbers of sequence errors an analysis of SNP distribution was performed. These errors can originate from the amplification in *E. coli*, from PCR steps or by the sequencing itself.

In the *G. pyriformis* transcriptome a frequency of 5.84 SNPs/Kb (72,468 SNPs in 6,656 NRVTs) was identified. 3,246 NRVTs were highly polymorphic (> 10 SNPs; see Fig. S3a). The SNP allele frequency distribution with few NRVTs showing a frequency around 0.5 indicates a homokaryotic genome organization (see Fig. S3b). *G. pyriformis* has a higher SNP frequency in comparison to other Glomeromycota, although it is still lower than the expected frequency for true heterokaryotic fungi (> 10 SNPs/Kb)[118].

2.2.4 KOG functional classification indicates life style differences to other Glomeromycota

Comparison against the Refseq database revealed KOG functional classifications for only 3,487 (19.9%) of the translated NRVTs (see Fig. S4a). Not surprisingly, when compared against sets of predicted fungal proteomes, conserved proteins with homologs present within Glomeromycota, Ascomycota, Basidiomycota, and Zygomycota show a much higher proportion of ortholog groups with KOG classification than Glomeromycota-specific proteins or proteins without known fungal homologs (49.3% vs. 5.4% vs. 2.0%). The KOG functional class distribution of *G. pyriformis* is quite distinct from other Glomeromycota (see Fig. 11a; Fig S3b) and much more similar to other fungi, independent of their life style, nicely illustrated by the results of a principal component analysis (see Fig. 11b; PC1 explaining 44.2% of variability, PC2 explaining 13.9% of variability, PC3 explaining 11.6% of variability). With 418 NRVTs (12%) the most abundant KOG class is of “general function prediction only”, followed by “posttranslational modification, protein turnover, chaperones” (335 NRVTs, 9.6%) and

a

KOG protein functional class	Number of NRVTs	*G. pyriformis* ratio	Glomeromycota transcriptomes	Ascomycota	Basidiomycota	Zygomycota
General function prediction only	418	12.0%	12.3%	12.2%	11.7%	12.1%
Posttranslational modification, protein turnover, chaperones	335	9.6%	13.7%	9.0%	8.8%	8.8%
Signal transduction mechanisms	271	7.8%	26.3%	7.2%	8.6%	9.9%
Translation, ribosomal structure and biogenesis	260	7.5%	3.9%	7.0%	5.8%	6.3%
Lipid transport and metabolism	239	6.8%	3.8%	4.3%	4.6%	4.2%
Function unknown	234	6.7%	4.8%	5.9%	5.7%	5.6%
Transcription	227	6.5%	4.3%	6.5%	6.7%	8.6%
RNA processing and modification	201	5.7%	4.9%	4.7%	4.7%	4.3%
Intracellular trafficking, secretion, and vesicular transport	195	5.6%	4.2%	7.0%	6.9%	6.4%
Replication, recombination and repair	145	4.2%	3.5%	3.7%	4.2%	2.8%
Amino acid transport and metabolism	125	3.6%	2.4%	4.5%	3.6%	3.6%
Energy production and conversion	123	3.5%	2.8%	4.7%	4.2%	4.7%
Cytoskeleton	108	3.1%	2.2%	3.2%	3.8%	3.5%
Cell cycle control, cell division, chromosome partitioning	94	2.7%	3.1%	3.6%	3.8%	3.4%
Carbohydrate transport and metabolism	89	2.5%	3.0%	3.8%	3.8%	3.4%
Inorganic ion transport and metabolism	70	2.0%	1.8%	2.4%	2.2%	2.4%
Secondary metabolites biosynthesis, transport and catabolism	69	2.0%	3.2%	2.5%	3.2%	1.8%
Nucleotide transport and metabolism	65	1.9%	0.7%	1.5%	1.3%	1.0%
Chromatin structure and dynamics	63	1.8%	1.6%	2.0%	2.1%	2.6%
Cell wall/membrane/envelope biogenesis	55	1.6%	7.2%	1.2%	1.4%	1.9%
Coenzyme transport and metabolism	47	1.3%	1.0%	1.7%	1.5%	1.0%
Defense mechanisms	31	0.9%	0.8%	0.6%	0.7%	0.6%
Nuclear structure	13	0.4%	2.0%	0.6%	0.5%	0.7%
Extracellular structures	9	0.3%	0.1%	0.2%	0.2%	0.2%
Cell motility	4	0.1%	0.1%	0.1%	0.1%	0.1%

b

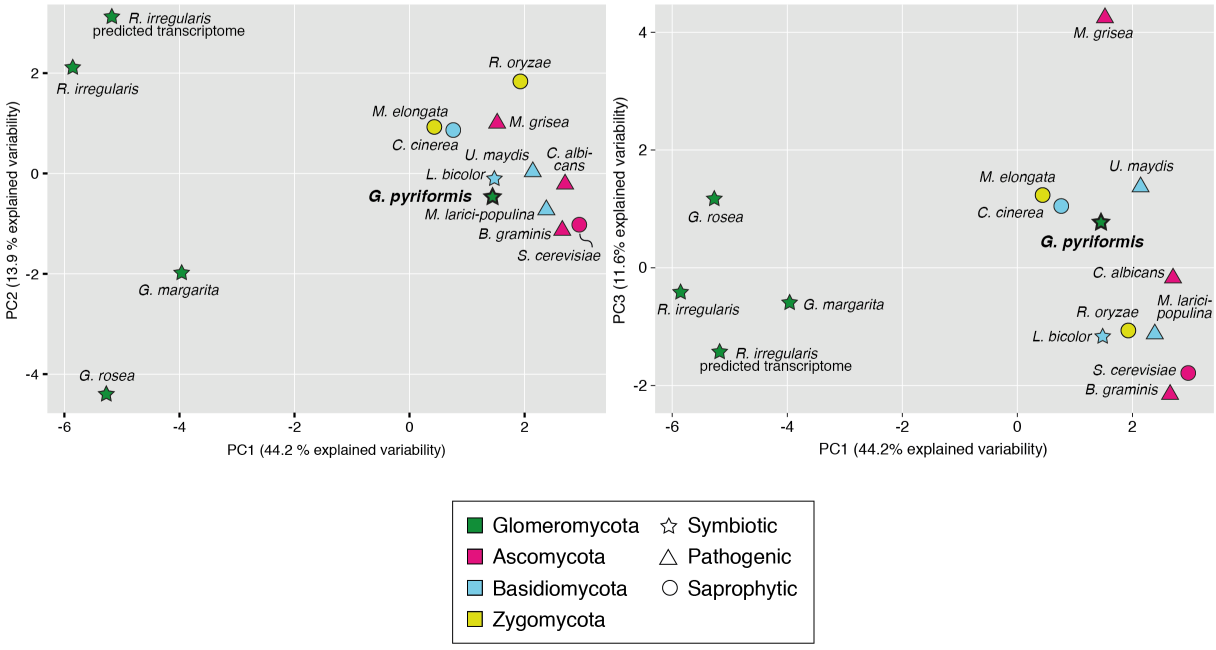

Fig. 11: Distribution of KOG protein functional classes. (a) Comparison of G. pyriformis KOG hits with other fungi. (b) PCA analysis of KOG hit distribution in several fungi from different phyla and life-styles.

"signal transduction mechanisms" (271 NRVTs, 7.8%) which is the most represented class in other glomeromycotan transcriptomes (26.3%). Another highly represented class in other glomeromycotan transcriptomes is "cell wall/membrane/envelope biogenesis" (7.2%), which is in *G. pyriformis* only represented by 1.6% of the classified NRVTs similarly to Asco-, Basidio- and Zygomycota (1.2%, 1.4%, 1.9% respectively).

2.2.5 The presence of meiosis-related genes is conserved among Glomeromycota

Using a set of 86 fungal meiosis-related genes[119], seven of which are meiosis-specific, a reciprocal blastx/tblastn search against the *G. pyriformis* transcriptome was performed. 57% of these genes are present (see Fig. 12, Table ES1), which is similar to the numbers identified in the transcriptomes of *G. rosea*[60] and the genome of *R. irregularis*[120]. In comparison, in the *R. irregularis* transcriptome only 26% of the meiosis genes were found[114]. In the more complete *G. margarita* transcriptome 70% of the meiosis gene set could be identified.

Of the seven meiosis-specific genes in this gene set, five (*SPO11*, *DMC1*, *MSH5*, *HOP2*, *MND1*) are present in the *G. pyriformis* transcriptome, and among the investigated transcriptomes only *G. margarita* contains all seven genes (additionally *MSH4* and *REC8*). All seven are also present in the *Rhizophagus* genome but only two in the *R. irregularis* transcriptome.

The *G. pyriformis* transcriptome was also searched for *MATA-HMG* genes and was able to identify 36 NRVTs with similarity (blastn, e-value < 1) to a set of 76 *R. irregularis MATA-HMG*

Process involved (number of genes)	*G. pyriformis* transcriptome	*G. margarita* transcriptome	*G. rosea* transcriptome	*R. irregularis* transcriptome	*Rhizophagus spp.* genome
DSB (double-strand break) generation (10)	40%	40%	30%	10%	20%
Removal of Spo11 (4)	50%	75%	50%	50%	50%
Strand invasion (10)	60%	70%	50%	20%	60%
DNA damage checkpoint (5)	40%	80%	60%	0%	40%
Regulation of crossover frequency (10)	50%	70%	50%	30%	70%
Synaptonemal complex (7)	43%	63%	43%	29%	43%
DNA Repair (8)	88%	88%	75%	25%	88%
Mismatch repair (5)	60%	100%	80%	60%	60%
Resolution of recombination intermediates (9)	56%	67%	67%	0%	56%
Nonhomologous end joining (4)	75%	75%	75%	0%	75%
Others (14)		64%	57%	50%	79%
Meiosis-specific genes (7)	71%	100%	57%	29%	100%
Total (86)	57%	70%	56%	26%	59%

Fig. 12: Presence of meiosis genes in glomeromycotan transcriptomes and genomes. Presence of 86 meiosis genes was checked for in *G. pyriformis*, and compared to the data from *G. margarita*, *G. rosea*, *R. irregularis* transcriptomes and *R. irregularis* genomes.

genes (see Table S1)[110]. Although, compared to *R. irregularis*, *G. pyriformis* contains only

half as many of these genes, there is still an expansion of this group in comparison to other fungi, which contain a lot less (usually < 10) *MATA-HMG* genes[110].

2.2.6 Carbohydrate-active enzyme (CAZyme) profile suggests adaptation to symbiotic lifestyle

A total of 72 NRVTs encoding for 42 CAZyme families were identified. These NRVTs included nine auxiliary activities (AA), seven carbohydrate-binding module (CBM), 17 carbohydrate esterases (CE), 26 glycoside hydrolases (GH) and 16 glycosyl transferases (GT) (see Fig. 13, Table S2). As in the *G. rosea*[60] and *R. irregularis*[114] transcriptomes, no polysaccharide lyases (PL), were found. Of these NRVTs seven belong to families that are plant cell wall degrading (CE1, GH5, GH27, GH31), and 21 belong to chitin metabolism involved families (AA11, CBM18, CBM19, CBM50, CE4, GH16, GH18, GH23, GT2). Compared to *G. rosea*[60], *R. irregularis*[114] and *G. margarita* (CAZyme analysis in this work) the numbers of CAZymes relative to the number of NRVTs are comparable. In total though, there are much less CAZymes in *G. pyriformis* (72) compared to the other analyzed transcriptomes (*G. margarita*: 632, *G. rosea*: 293, *R. irregularis*: 94). The percentage of plant cell wall degrading (PCW) families is similarly low in all four fungi, whereas chitin metabolism is overrepresented in the *G. pyriformis* CAZyme families. A clear difference between *G. pyriformis* and the other Glomeromycota is visible in the GT family. About one third of the CAZymes that are not involved in chitin metabolism or PCW degradation in the other Glomeromycota belong to GT families. *G. pyriformis* has only about 12.5%.

CAZyme groups	*G. pyriformis* transcriptome	*G. margarita* transcriptome	*G. rosea* transcriptome	*R. irregularis* transcriptome
Plant cell wall degrading	7	68	33	11
Chitin metabolism	21	74	46	26
Other "Auxiliary Activities Family"	6	35	34	0
Other "Carbohydrate-binding Module Family"	1	86	6	3.5
Other "Carbohydrate Esterase Family"	10	79	35	1
Other "Glycoside Hydrolase Family"	18	75	38	17.5
Other "Glycosyl Transferase Family"	9	210	101	35
Other "Polysaccharide Lyase Family"	0	5	0	0
Total	72	632	293	94
% of total NRVTs	0.40%	0.73%	0.30%	0.36%

Fig. 13: *G. pyriformis* NRVTs classified as CAZymes. Presence and distribution of CAZymes in *G. pyriformis* compared to *G. margarita*, *G. rosea* and *R. irregularis* transcriptomes.

2.2.7 Missing Glomeromycota Core Genes (MGCGs)

A set of 39 missing glomeromycotan core genes (MGCGs)[60] was analyzed for their presence in the *G. pyriformis* transcriptome. The loss of this MGCG set is conserved among glomeromycotan fungi and can be connected to the obligatory biotrophic life-style of these fungi. In *G. pyriformis* most of these genes were missing. In a first screen five NRVTs coding for genes in the MGCG set were identified. However, subsequent blastx of the respective NRVTs against the NCBI database revealed that four of these NRVTs are most probably bacterial contaminations. One NRVT had similarities to *URE2*, a gene coding for a nitrogen catabolism repression transcriptional regulator that was also found in the *R. irregularis* genome (see Table S3)[60].

2.2.8 The *G. pyriformis* transporter repertoire

In this analysis a total of 1,726 NRVTs that are putative transporters were identified with a reciprocal BLAST approach. Blastx of these putative transporters revealed that 287 of the NRVTs might be of bacterial/protiste/plant origin and had to be removed from the analysis. This leaves 1439 NRVTs coding for putative *G. pyriformis* transporters in 219 transporter families (see Table ES2). The most abundant transporter families are the nuclear pore complex (NPC) family, the domain of unknown function 3339 (DUF3339) family and the ATP-binding cassette (ABC) superfamily. Performing this analysis with the same parameters also for the transcriptomes of *G. margarita*, *G. rosea* and *R. irregularis* resulted in 5194 NRVTs (229 families), 3307 NRVTs (259 families) and 2182 NRVTs (215 families) respectively.
Also in *G. margarita* and *R. irregularis* the NPC family is the most abundant transporter family. Although the number of putative transporter NRVTs in *G. pyriformis* is lower than in the other glomeromycotan transcriptomes, compared to the total number of NRVTs in the transcriptomes, they have similar numbers. Also in regard to superfamilies, class and subclass of the transporters, *G. pyriformis* shows similar numbers as *R. irregularis* and *G. margarita*. The already published ammonium transporters (AMT1, AMT2, AMT3)[121] as well as the monosaccharide transporter MST1[81] of *G. pyriformis* could be identified as well as an additional NRVT showing similarities to *R. irregularis* MST4[75] and NRVTs similar to *Rhizophagus sp*. phosphate transporters[74,92,122] and aquaporins[123].

2.2.9 The *G. pyriformis* repertoire of secreted proteins

The complete transcriptome was searched for putative secreted proteins. Therefore the predicted protein sequences were analyzed for the presence of a secretion peptide in combination with the absence of a transmembrane helix. Only proteins strongly predicted to be located in the secretory pathway (parameter LOC = S and Reliability class = 1) were

included. Thereby 104 secreted candidate proteins were obtained. The respective NRVTs were analyzed by blastx against the non-redundant protein sequences (nr) database and 15 sequences of possible bacterial origin were removed. In the remaining 89 sequences there are 17 with a KOG functional classification including predicted lipase, palmitoyl protein thioesterase, 5'-3' exonuclease HKE1/RAT1 and carbonic anhydrase. None of the CAZymes are secreted (see Table ES3).

2.2.10 Identification of three repeat containing proteins - GpRICs

In three of the secreted proteins 34% of the *G. pyriformis* insert reads were assembled. These proteins contain repeats and were therefore called "**R**epeat containing proteins **I**n symbiosis with **C**yanobacteria" (GpRIC1, GpRIC2, GpRIC3). 22.3% of *G. pyriformis* reads mapped to GpRIC1, 5.0% to GpRIC2 and 3.0% to GpRIC3. As the assembly of the transcriptome was performed with short reads, the assembly algorithm was challenged by these repeat-containing sequences. This resulted in several misassembled contigs (GpRIC1: 3,587; GpRIC2: 171; GpRIC3: 141), containing stretches of GpRICs sequences.

In order to determine the correct sequences of these sequences colony hybridization was performed on the original cDNA expression library. Clones of all three GpRICs could be identified through this approach and were sequenced using Sanger sequencing.

GpRIC1 is 435 aa long and contains 6 repeats of 68 aa. All of the repeats are separated by KEX2 recognition sites. There are three sites in the repeats with only 50% conservation, three sites with 67% conservation and seven sites with 83% conservation. With 32.6%

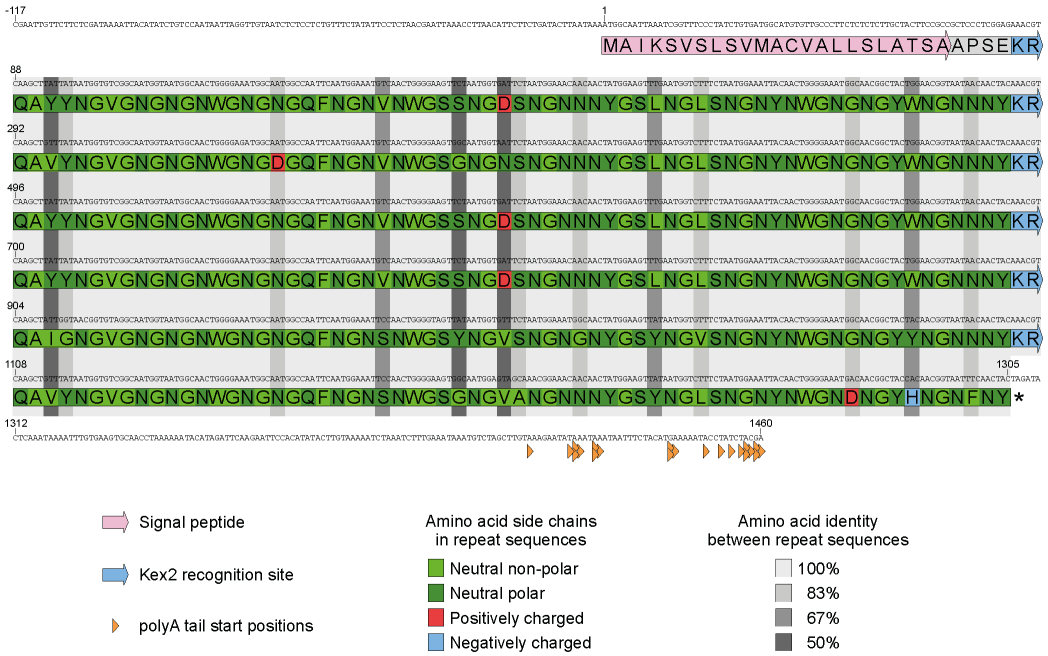

Fig. 14: mRNA and amino acid sequence of GpRIC1. Amino acid sequence contains a signal peptide (pink), several Kex2 recognition sites (blue) and six almost identical repeats containing mainly un-charged amino acids. Several polyA tail start positions were identified in different clones (orange triangles).

Asparagine and 25.5% Glycine the protein is mostly neutral without any charges (see Fig. 14). GpRIC2 is 456 aa long and contains three perfect repeats, separated by two KEX2 recognition sites. The repeats are 145 aa (last repeat without KEX2 cleavage site 143 aa) long. Also GpRIC2 contains mainly neutral amino acids with 38.9% Asparagine and 23.3% Glycine. GpRIC3 is with 470 aa the longest GpRIC. This one has less regular repeats compared to the other two. In total there are 20 KEX2 recognition sites, dividing the protein into four different classes of peptides: (a) three repeats of 55-58 aa length, (b) six repeats of 32-36 aa length, (c) eight peptides of 7 aa length that start with an aspartic acid (D) and a proline (P) and (d) three peptides of 3-5 aa length without similarity. The sequence of these peptides is “a-c-c-b-b-b-c-c-c-c-b-b-a-c-c-b-a-d-d-d”. GpRIC3 has a similar amino acid composition as the other two GpRICs with 33.5% asparagine and 21.9% glycine.

The cDNA as well as protein sequences are deposited at GenBank under the accession numbers MH580277 (GpRIC1), MH580278 (GpRIC2) and MH580279 (GpRIC3).

2.3 Host induced gene silencing of *R. irregularis* candidate effector genes

Up-to-date no *in vivo* methods are available for the transformation of glomeromycotan fungi. This makes the functional characterization of glomeromycotan genes difficult. The establishment of a method called host induced gene silencing (HIGS) could therefore offer a valuable tool for these species. In HIGS, roots of the host plant are transformed with an RNAi construct targeting fungal genes. Upon colonization, the processed dsRNA is translocated via an unknown process to the fungus, where the target genes are silenced (see Fig. 6).
In order to establish this method in our lab, a set of putative fungal effector genes was chosen. These genes are conserved between *R. irregularis* and *R. clarus*, two closely related species of the Glomeromycota, which makes it likely that they might play an important role in symbiosis establishment and an effect could be observed upon downregulation.

2.3.1 Expression of selected *R. irregularis* effector genes in roots

To demonstrate that HIGS is working in this context, the native amount of mRNA of the target gene must be detectable. Therefore the expression of 19 putative *R. irregularis* effector genes (pers. comm. K. Sedzielewska Toro) was analyzed via qRT-PCR in colonized carrot roots (continuous root organ culture; *EX_25359*, *EX_27056*, *EX_54675*, *NLS_30765*, *NLS_334409*, *NLS_343100*, *NLS_7749*/*RiTSL1*, *NLS_98735*, *RCP_230436*, *RCP_335225*, *RCP_340423*, *RCP_84949*) or in colonized *L. japonicus* roots (5 weeks post inoculation; *NLS_26232*, *NLS_320155*, *NLS_32853*, *NLS_349288*, *RCP+NLS_349824*, *SCR_339199*, *SCR_343180*). Three of them were later excluded from the set of putative effectors (EX_). The other effectors included nine candidates with a nuclear localization signal (NLS_), four genes coding for repeat-containing proteins (RCP_), two genes coding for small and cysteine rich effectors (SCR) and one with a NLS and coding for a RCP.
The mean relative expression (normalized to *Ri_β-Tubulin*) of three biological replicates ranged between 0.0001 and 0.36 (see Fig. 15). The lowest expression was detected for *EX_54675*, which was later excluded from the effector candidates (pers. comm. K. Sedzielewska Toro). The effector candidate with the highest expression was *RCP+NLS_349824*, which contains repeats as well as an NLS. For some of the effector candidates the expression of the different biological replicates varied up to 28-fold (*RCP_335225*). Others also showed substantial variation between the biological replicates, e.g. *EX_54675*: 6.8-fold, *NLS_334409*: 7.2-fold, *NLS_26232*: 8.6-fold, *SCR_339199*: 14.4-fold. On the other hand there are also effector candidates that are more similarly expressed in the different biological replicates, e.g. *NLS_320155*: 1.07-fold, *RCP_230436*: 1.14-fold, *NLS_343100*: 1.18-fold. The effector candidates for HIGS targets were chosen considering a well detectable expression level and variation between biological replicates being not higher

than 6-fold. These parameters should ascertain that a significant downregulation is measureable.

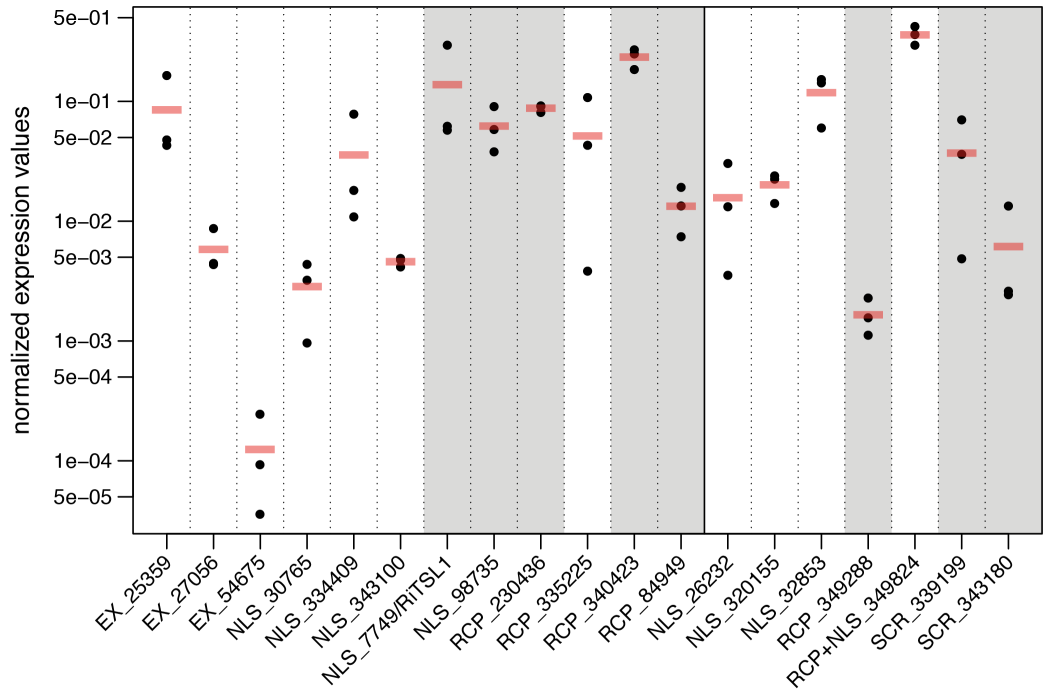

Fig. 15: Expression of *R. irregularis* effector candidates in colonized carrot and *L. japonicus* roots. qRT-PCR for three biological replicates per analyzed gene was performed. The values were normalized to *Ri_β-Tubulin*. The samples in the first box were obtained from carrot root organ culture. The samples in the second box were obtained from *L. japonicus* colonized roots. Grey background shows the effector candidates that were subsequently targeted in HIGS experiments. Red bars indicate the mean value of the biological replicates.

2.3.2 Downregulation of one effector candidate out of eight effectors tested

Eight of the aforementioned effector candidates were chosen as HIGS targets. For three (*NLS_98735*, *RCP_340423*, *NLS_7749*/*RiTSL1*) two target sites were selected. For all others only one target site was used. Usually in hairy root transformation experiments not the complete root system is transformed. Therefore GFP was used as a transformation marker in all experiments. In a first experiment (*NLS_98735*, *RCP_340423*, *NLS_7749*/*RiTSL1*), the GFP-positive (+) roots and GFP-negative (-) *L. japonicus* roots were separated upon harvesting and subsequently analyzed separately. Only for *NLS_7749*/*RiTSL1* a significant downregulation (two-sided t-test: EV+/Target1+ p-value: 0.02, EV+/Target2+ p-value: 0.03) could be observed (see Fig. 16). All of the other eight targets on seven effector candidate genes that were tested showed no significant difference between the expression in HIGS and EV control roots. The lower expression of *RiTSL1* in HIGS roots appeared not only in the GFP-positive roots, but also in the GFP-negative roots.

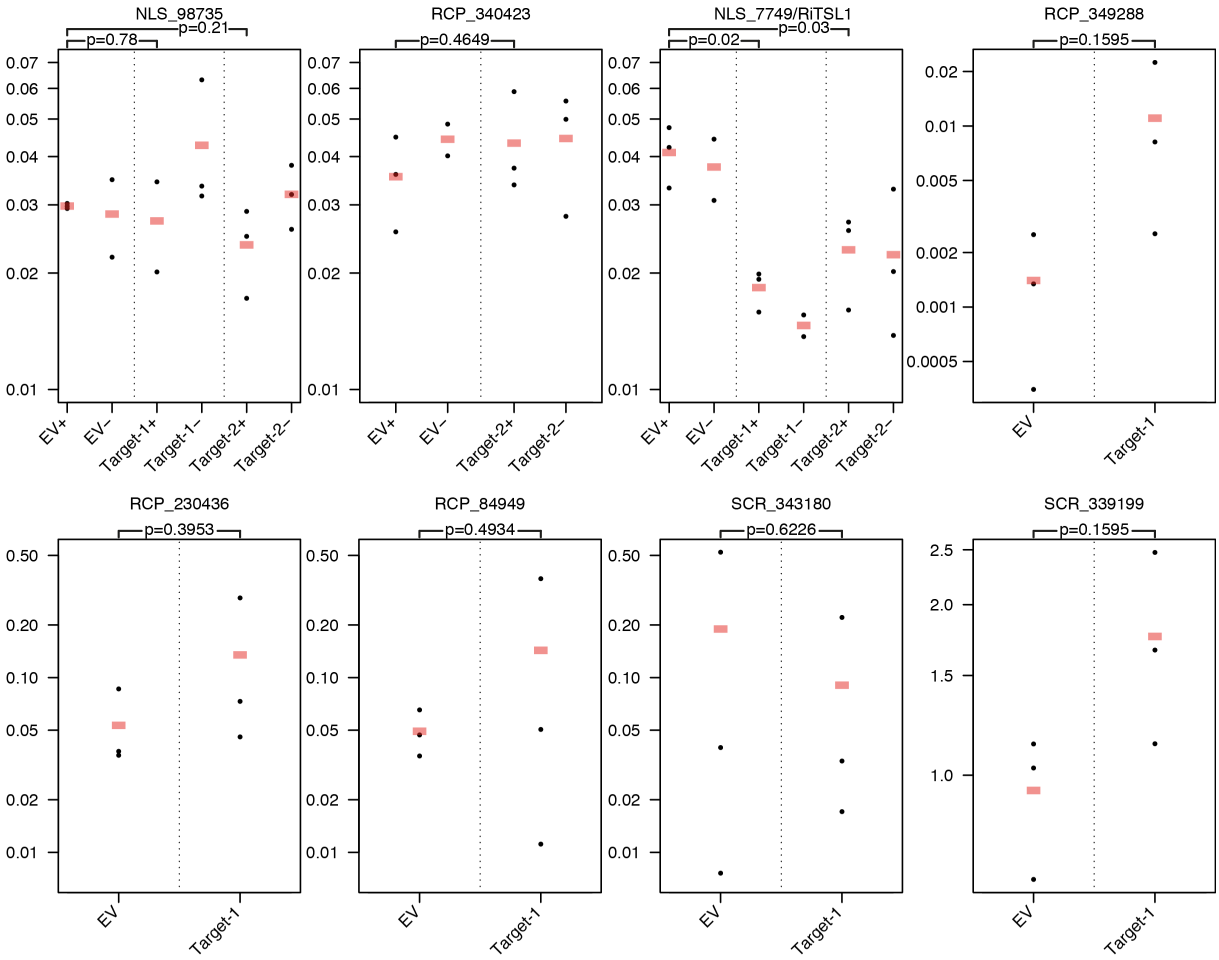

Fig. 16: qRT-PCR results of HIGS experiments with different fungal effector candidate targets. Expression values are normalized to the arithmetic mean of the expression values for *Ri_Ef1-alpha* and *Ri_β-Tubulin*. Empty vector (EV) control roots were compared to roots expressing dsRNA constructs for HIGS target genes (Target-1/Target-2: different target regions on same gene). For HIGS targeting *NLS_98735*, *RCP_340423* and *NLS_7749/RiTSL1*, transformed (+) and untransformed (-) roots of the same root systems were used. Two-sided t-test was performed and shows only for *NLS_7749/RiTSL1* significant lower expression in HIGS roots compared to EV control roots. Red bars indicate the mean value of the biological replicates.

2.3.3 Lower root length colonization and small arbuscules in *RiTSL1* HIGS roots

In the above mentioned hairy root experiment, in which a lower expression of *RiTSL1* in HIGS roots as compared to the EV control roots was observed, also a lower root length colonization was detected (see Fig. 17a). The total root length colonization, as well as internal hyphae and vesicles, were significantly less in roots with a lower expression of *RiTSL1*. The root length colonization is correlated with the expression of RiTSL1 (see Fig. 17b; correlation coefficient: 0.64; adjusted R-squared value: 0.33). Also the form of the arbuscules in this experiment differed. In HIGS roots the arbuscules appeared to be smaller and not cell filling (see Fig. 17c, representative picture from different experiment). Furthermore, known symbiosis induced plant genes (*AMT2.2*, *Bcp1*, *PT4*, *SbtM1*), were expressed at lower levels in HIGS roots, confirming the lower root length colonization (see Fig. 17d).

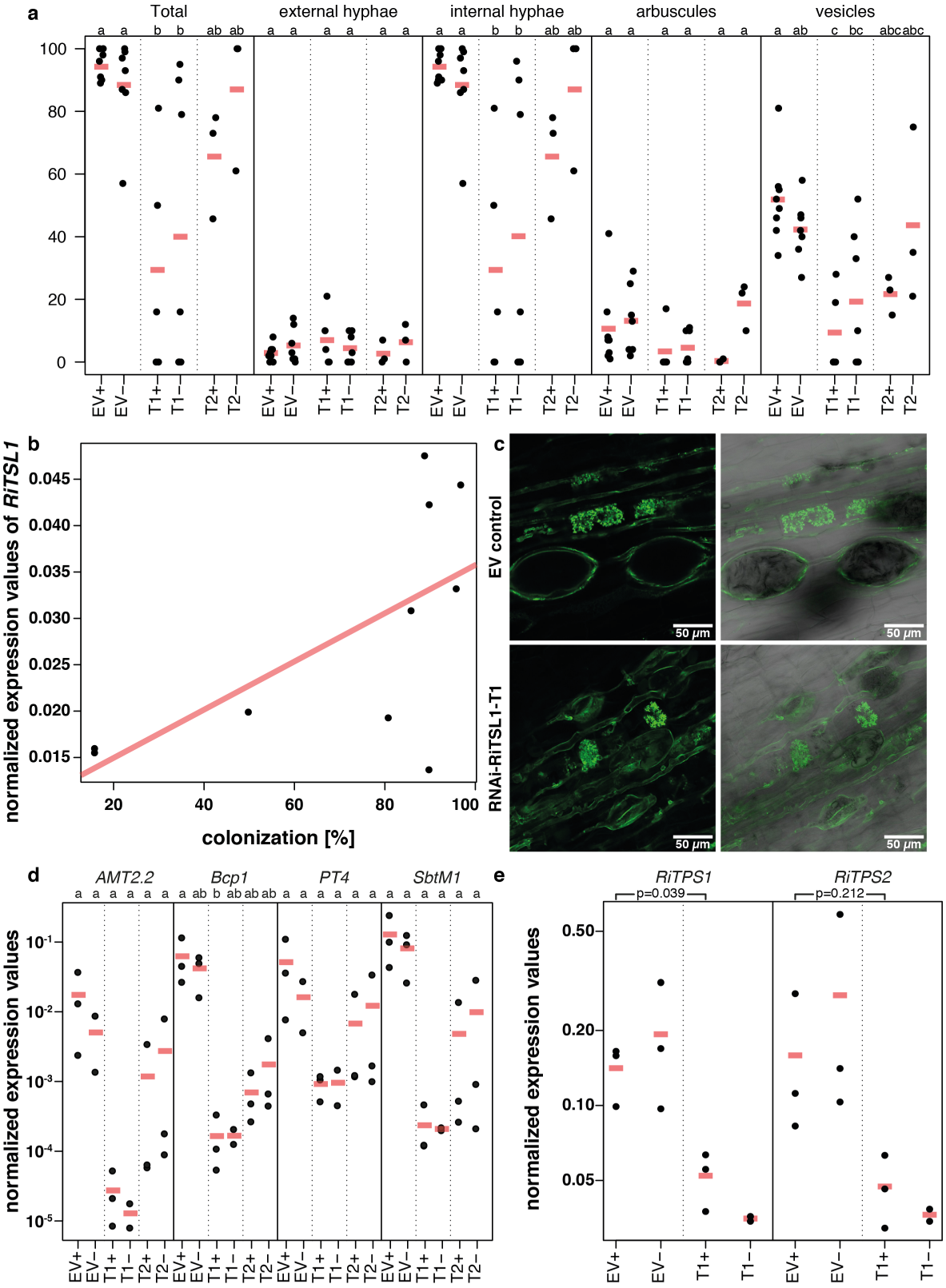

Fig. 17: Downregulation of *RiTSL1* in *L. japonicus* roots and the respective phenotypes. T1 = HIGS-RiTSL1 Target1, T2 = HIGS-RiTSL1 Target2, EV = empty vector control. All results except (d) from same experiment. Red bars indicate the mean value of the biological replicates. (a) Root length colonization levels in RiTSL1 HIGS (T1 & T2), and EV control roots. (b) The expression of *RiTSL1* is correlated with root length colonization in HIGS-RiTSL1 and EV control roots (corr. eff.: 0.64; R-squared: 0.33). (c) Example of observed arbuscule phenotype, small and not cell-filling arbuscules. (d) qRT-PCR results show expression of AM symbiosis induced *L. japonicus* genes *AMT2.2*, *Bcp1*, *SbtM1* and *PT4* (normalized to *LjUbi*). (e) qRT-PCR results show expression of *RiTPS1* and *RiTPS2* (normalized to arithmetic mean of *Ri_Ef1-alpha* and *Ri_β-Tubulin* in hairy root experiment with downregulation of *RiTSL1*

The homologues of RiTSL1 in yeast (ScTSL1, ScTPS3) are involved in trehalose biosynthesis, together with two other proteins (ScTPS1, ScTPS2), where these proteins form a complex[124-126]. Also in *R. irregularis* two additional genes were identified that code for homologues of the yeast proteins (see Section 2.4.1). Therefore the expression of these other two *R. irregularis* genes involved (*RiTPS1* & *RiTPS2*) was analyzed as well. Also *RiTPS1* was significantly lower expressed in GFP-positive RiTSL1-HIGS (transformed) roots compared to GFP-positive EV-control roots (untransformed). *RiTPS2* showed more variation in expression between the biological replicates and expression was not significantly lower in HIGS roots (see Fig. 17e).

2.3.4 Downregulation of RiTSL1 via HIGS not reproducible

The results of downregulation of *RiTSL1* and the respective phenotype in a first experiment (see above, Table 6 HR4) looked very promising. To quantify the reduced size arbuscules and for confirmation of the results, further experiments with the same experimental setup were performed. In the following six experiments no downregulation of *RiTSL1* in HIGS roots as compared to EV control roots was observed. Also decrease in root length colonization and the stunted arbuscules were only observed in one further experiment (HR5; see Table 6). The detection of hairpin cDNA to confirm that the RNAi construct was expressed was successful in HR4 and two further experiments in which no downregulation was detected (see Table 6). One attempt to directly apply a dsRNA substrate to the fungus grown on plate with root organ culture (PIGS1) also didn't result in downregulation of *RiTSL1*.

Table 6: HIGS-RiTSL1 hairy root transformation experiments (HR) and one PIGS experiment with occurrence of downregulation of RiTSL1 and respective phenotypes.

Experiment name	Stunted arbuscules	Decrease in root length colonization	Downregulation of RiTSL1	Detection of hairpin dsRNA
HR4	Yes	Yes	Yes	Yes
HR5	Yes	Yes	No	Yes
HR6	NA	No	NA	NA
HR7	Yes	No	NA	NA
HR8	No	No	No	NA
HR11	NA	NA	No	Yes
HR14*	NA	No	NA	NA
PIGS1	NA	NA	No	NA

*Instead of 35S promoter, UBQ10 promoter for transformation marker (GFP); NA: not available

2.3.5 Arbuscule size measurements show slight reduction of size in the morning

The fact that the HIGS results were not reproducible can have two possible reasons. Either in the first experiment HIGS indeed worked and is not reproducible to some unknown factors, or HIGS didn't work and the downregulation in the first experiment was due to some other circumstances. One possible reason for the different size and shape of arbuscules could be the different time of harvest of EV control and HIGS-RiTSL1 roots. Arbuscules are short-lived structures. Therefore a time course experiment was performed to identify changes in the arbuscules. *R. irregularis* was grown in pots with *L. japonicus*. 5.5 weeks past inoculation (wpi) the roots were harvested each 3 h within a period of 24 h (T1-T8; see Fig. 18). For each time point four biological replicates (pooled roots from one pot containing five plants and *R. irregularis*) were harvested. Overall root colonization varied between the biological replicates so that for some replicates only few arbuscules could be measured. For each biological replicate 10-29 arbuscules were measured, which equals to a total of 61-103 arbuscules per time point. The average arbuscule size compared to the arbuscule containing plant cell varied between 46 and 66% amongst all biological replicates. The average size of all arbuscules in one time point was smallest in T2 with 52% and largest in T4 and T8 with 58% and 59%, respectively. Multicomparison analysis showed that the size of arbuscules differs significantly between these time points (see Fig. 18).

For the above-mentioned HIGS-experiment all roots were harvested during the light phase. In this phase the arbuscules have overall the same size, meaning the observed stunted arbuscules are not caused by different times of harvest. This offers strong support that the observed arbuscule phenotype in HIGS-RiTSL1 roots compared to EV control roots was indeed the effect of a successful downregulation of *RiTSL1* expression.

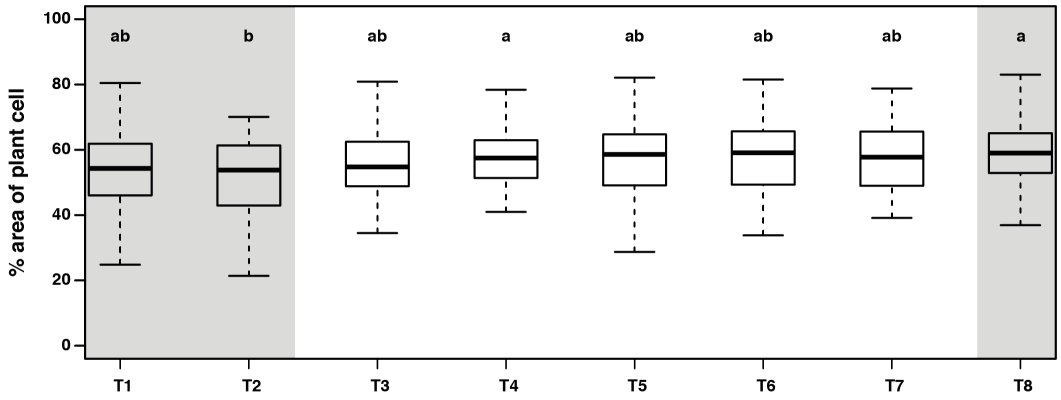

Fig. 18: Arbuscule size in %-age of plant cell over the course of one day. Timepoints (T1-T8) of sampling are 3 hours apart from each other. Dark shaded timepoints are nighttime. For each timepoint the roots of four biological replicates (one biological replicate = five plants grown in one pot) were harvested and for each root 10-29 arbuscules were measured. Multicomparison was performed with ANOVA and Tukey Honest Significant Differences (confidence level = 0.95) analysis.

2.3.6 Expression of *RiTPS1*, *RiTPS2* and *RiTsl1* shows a slight diurnal rhythm in *L. japonicus* roots

From yeast it is known that *TSL1* is involved in trehalose metabolism together with *TPS1* and *TPS2*. Trehalose is a disaccharide that is used as a storage sugar in AM fungi. As sugar from the plant may not be available at the same concentrations over the course of the day, it is possible that these fungal genes are differentially expressed depending on the time of the day.

A time course experiment was performed to show the expression profiles of trehalose biosynthesis genes of *R. irregularis* in order to determine if the observed downregulation of

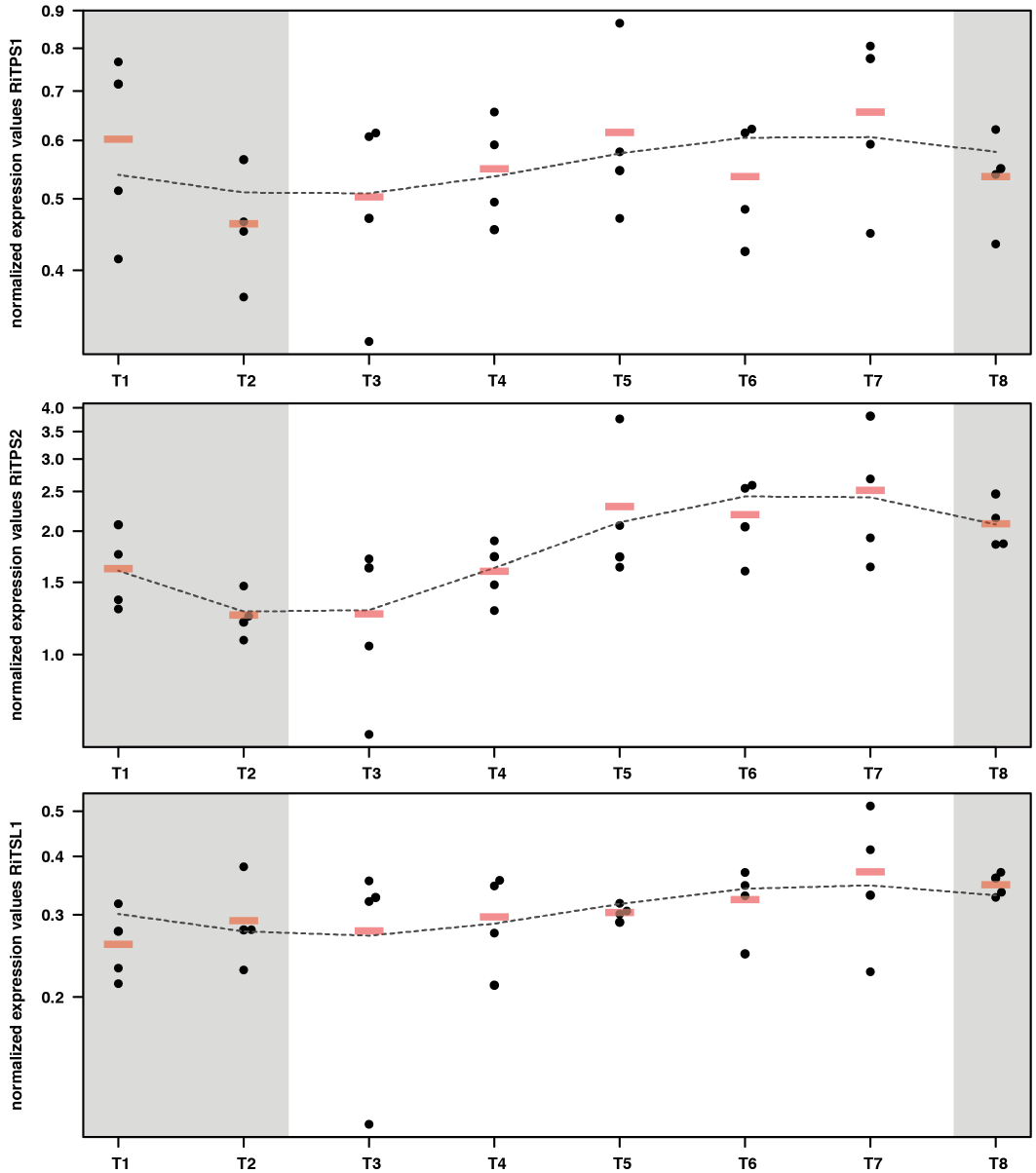

Fig. 19: qRT-PCR results of trehalose biosynthesis gene expression in *R. irregularis* colonized *L. japonicus* roots over the course of one day. Colonized root samples were harvested every three hours over a time course of 24 hours in a light/dark cycle of 16/8 h. For each time point four biological replicates were analyzed. Red bars indicate the mean value of the biological replicates.The absolute expression levels were normalized to the geometric mean of three reference genes. ANOVA and Tukey Honest Significant Differences (confidence level = 0.95) analysis revealed no significant differences between timepoints. Data points were fit to a sinus-model and the resulting curve plotted on top of the results.

RiTSL1 in HIGS-RiTSL1 roots was due to different harvesting times of EV control and HIGS-RiTSL1 roots.

The experiment was performed *in vivo* in colonized roots of *L. japonicus*. For each time point four biological replicates were analyzed. There is no significant difference between the different time points in the expression of *RiTPS1*, *RiTPS2* and *RiTSL1* (see Fig. 19). Nevertheless a slight trend of rhythmicity is observable, especially for the higher expressed *RiTPS2*. *RiTPS2* mean relative expression values ranged between 1.25 in the morning (T2 & T3) to 2.52 in the evening (T7). The average expression of *RiTPS2* over all time points is 3.3x and 6x higher than that of *RiTPS1* and *RiTSL1*, respectively. For these two genes the relative expression values range from 0.46 (T2) to 0.66 (T7) for *RiTPS1* and from 0.26 (T1) to 0.37 (T7) for *RiTSL1*.

This slight rhythmicity of expression of trehalose biosynthesis genes coincides with the size of arbuscules. T2 was the time point with smallest arbuscule size and lowest expression of RiTPS1, RiTPS2 and RiTSL1. On the other hand the largest arbuscules were observed just after the highest expression of trehalose biosynthesis genes (T8 and T7 respectively).

The changes of expression are not significant indicating that the significantly lower expression of RiTSL1 in HIGS-RiTSL1 roots is indeed the result of successful downregulation.

2.4 Trehalose metabolism in *R. irregularis*

Even though the downregulation of *RiTSL1* was not reproducible, the respective phenotype indicates an important role for *RiTSL1* in the establishment or maintenance of AM symbiosis. Further experiments were conducted to investigate the function of *RiTSL1* in *R. irregularis*. The sequence of *RiTSL1* shows homology to trehalose biosynthesis genes in yeast. Therefore knowledge of this process in yeast was used to elucidate the function of *RiTSL1* and other trehalose biosynthesis genes in *R. irregularis*.

2.4.1 Trehalose-biosynthesis proteins in *R. irregularis* fall into three clades of OtsA/OtsB pathway that are found in most fungi

A common biosynthesis pathway for trehalose is the OtsA/OtsB pathway (see Fig. 20a), known from archaea, bacteria, plants, arthropods and fungi. This pathway is composed of two consecutive enzymatic reactions, which are catalyzed by trehalose-6-phosphate-synthase (TPS) and trehalose-6-phosphate-phosphatase (TPP). In the first step a glucose moiety is transferred from UDP-glucose to glucose-6-phosphate by TPS. In the subsequent step trehalose-6-phosphate (T6P) is dephosphorylated by TPP to trehalose. Most fungi possess three enzymes that contain either a TPS, a TPP or both of the domains.

Also in the genome of *R. irregularis* three sequences containing both of the conserved domains were identified (see Fig. 20b). All three genes were isolated from cDNA of *L. japonicus* roots colonized with *R. irregularis*.

The gene encoding RiTPS1 has 7 exons and a coding sequence (CDS) of 2331 nt resulting in a 777 aa protein. The length of the CDS of RiTPS2 is also 2331 nt, but the gene contains 12 exons. The CDS of RiTSL1 is the longest with 2568 nt with a resulting 856 aa protein. All of the three proteins contain both TPS and TPP domains. In the conserved domain database (CDD) of the National Center for Biotechnology Information (NCBI) the TPS domain is described to have 13 active site residues whereas the TPP domain has 25 active site residues. RiTPS1 and RiTPS2 contain all 13 active site residues of TPS whereas RiTSL1 only contains 11. All 25 active site residues of TPP are present in RiTPS2 and RiTSL1, whereas RiTPS1 only contains 23.

To find further indications, which of the three proteins are taking over the TPS or TPP function, a phylogenetic analysis of 156 proteins containing only the TPS, or both TPS and TPP domains was performed. The TPS domains of these proteins from 41 different fungi, three oomycetes and one chlorophyte were aligned to each other and a maximum-likelihood phylogeny was calculated. The phylogeny resulted in a tree with three distinct clades (see Fig. 20c) that were also described by Avonce *et al.*[127] as Fungi A, Fungi B and Fungi C for enzymes with TPS function, only regulatory function, and TPP function respectively. The

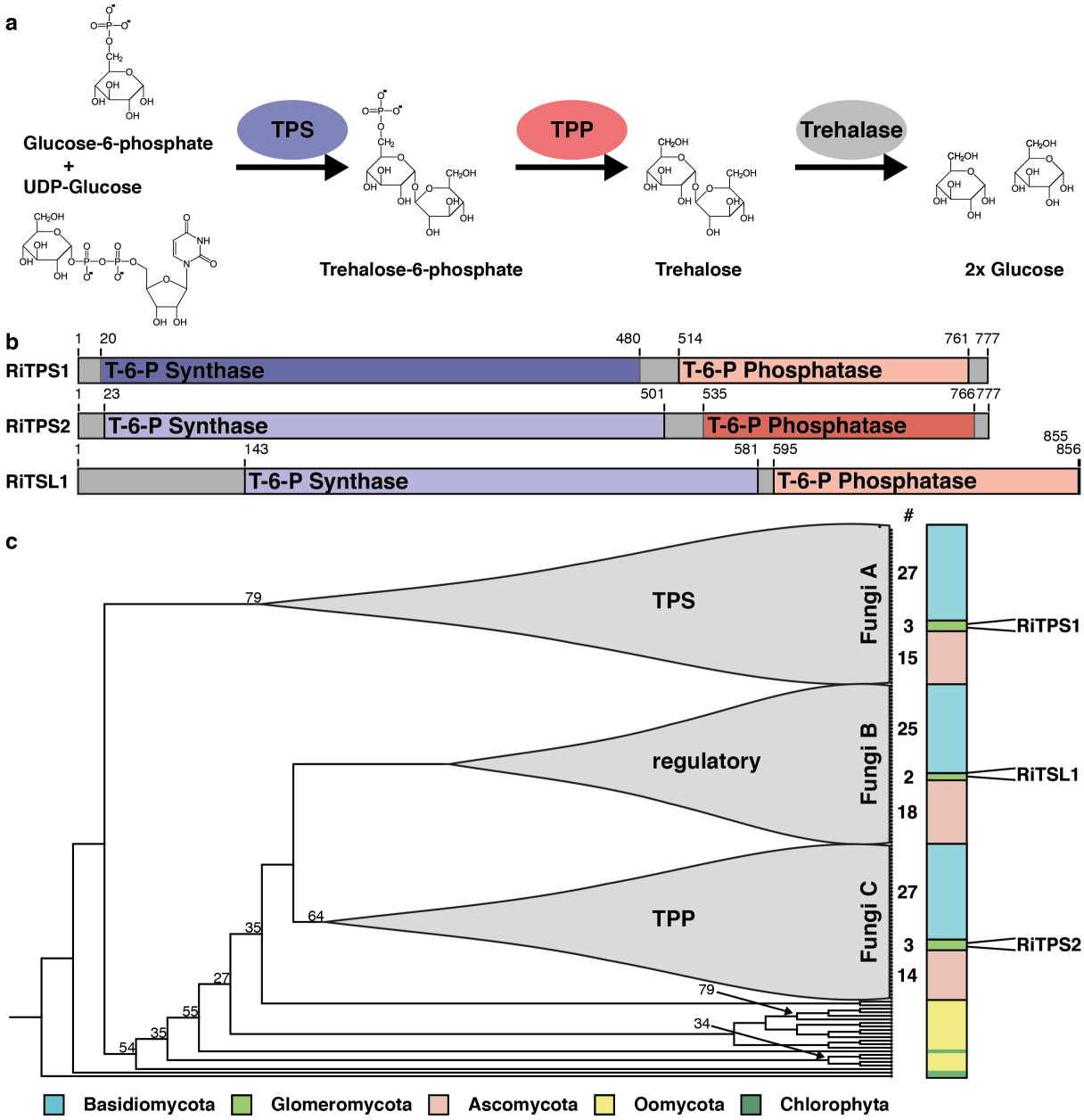

Fig. 20: Trehalose metabolism and the involved proteins in trehalose biosynthesis in *R. irregularis*. (a) Trehalose OtsA/OtsB biosynthesis pathway with trehalose-6-phosphate-synthase (TPS) and trehalose-6-phosphate-phosphatase (TPP) and release of trehalose to glucose by trehalase. (b) The three proteins in *R. irregularis* that show sequence similarity to proteins involved in OtsA/OtsB pathway. Darker shade shows functional domains. (c) Maximum likelihood phylogenetic tree of TPS domain of 156 proteins from 41 fungi, three Oomycetes and one Chlorophyte. Numbers of proteins in the different clades are shown in #-column. Only bootstrap values (1000 replicates) below 90 are shown.

clades were named in my analysis regarding the presence of the yeast proteins ScTPS1 (TPS function), ScTPS2 (TPP function) and ScTPS3 & ScTSL1 (regulatory function). RiTPS1 falls into one clade with ScTPS1, RiTPS2 with ScTPS2 and RiTSL1 with ScTPS3 & ScTSL1. This indicates that RiTPS1 might have TPS function, RiTPS2 might have TPP function and RiTSL1 might play a putative regulatory role in the process of trehalose biosynthesis.

In these analyses a conserved N-terminal region was identified in the proteins of the Fungi B clade (see Fig. 21a). In 35 of the 45 analyzed regulatory proteins the first 17-23 aa (Ascomycota), or the first 28-29 aa (Basidiomycota, Glomeromycota) are highly conserved.

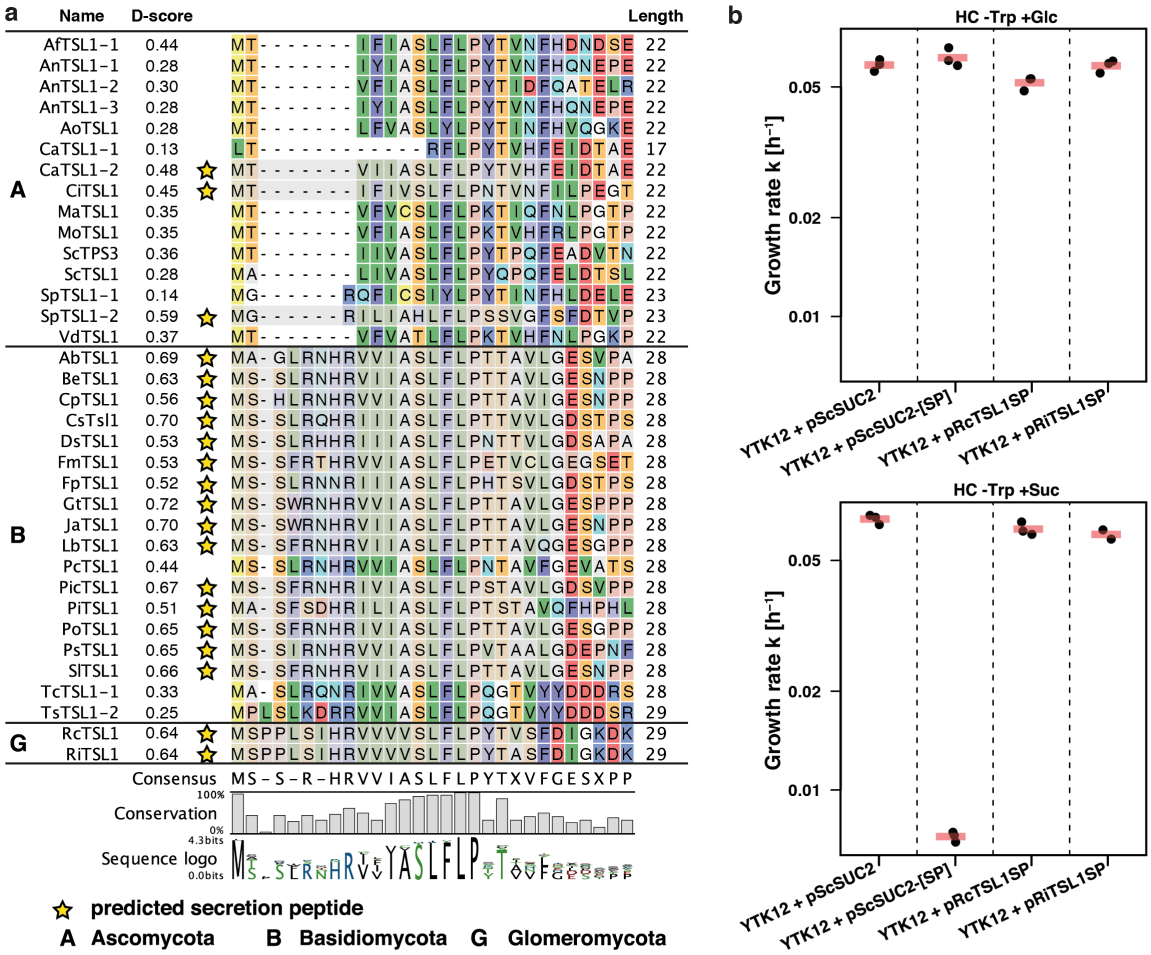

Fig. 21: N-termini of TSL1 proteins in Ascomycota, Basidiomycota and Glomeromycota are conserved and are in some cases predicted as secretion peptides (SP). (a) Alignment of N-termini of TSL1 of 10 Ascomycota (A), 18 Basidiomycota (B) and 2 Glomeromycota (G) shows high conservation. Most of the N-termini of TSL1 of Basidiomycota and Glomeromycota are predicted as SP by SignalP4.1 with a D-value cut-off of 0.45 (star; lighter shade indicates length of SP). (b) In a yeast-secretion assay the predicted SP of *R. clarus* and *R. irregularis* are functional. Red bars indicate the mean value of the biological replicates.

In most of the proteins of Basidiomycota (15 of 18) and Glomeromycota (2 of 2) and some of the Ascomycota (3 of 15) this region is furthermore predicted to be a signal peptide (SP). To test the functionality of the SP a yeast secretion trap assay was performed.

In this assay an invertase-deficient yeast strain is unable to grow on medium containing sucrose as a sole carbon source. An expression vector containing a gene for the N-terminal fusion of the SP of interest to invertase is created. By transforming a vector containing a functional SP to the invertase-deficient yeast strain, growth on sucrose is reestablished, showing that the predicted SPs of the two glomeromycotan TSL1 proteins (*R. irregularis*: RiTSL1, *R. clarus*: RcTSL1) are functional in yeast (see Fig. 21b).

2.4.2 Yeast complementation confirms TPS function of RiTPS1 and TPP function of RiTPS2

To confirm that RiTPS1 has TPS function and RiTPS2 has TPP function, while RiTSL1 doesn't have either of these, yeast complementation assays were performed. The test

condition for haploid W303 *tps1Δ* mutants was growth on glucose. Yeast cells that are able to produce T6P can grow on medium in which the sole carbon source is glucose. Knockout mutants that are not able to produce T6P show no growth on this medium (see Fig. S5). Transforming *tps1*Δ mutants with an expression vector for RiTPS1 resulted in the same growth as *tps1*Δ mutants transformed with an expression vector for ScTPS1. Both show the same growth rates as the wild type cells (see Fig. 22a). Expression of RiTPS2 and RiTSL1 couldn't complement the function of ScTPS1.

Yeast W303 *tps2Δ* mutants accumulate the intermediate compound T6P and are unable to react to certain abiotic stresses. These mutants were unable to grow at 37°C (see Fig. S5). Expression of RiTPS2 in *tps2Δ* rescued the growth at 37°C the same as expression of ScTPS2 in these mutants. The growth rate of complemented yeast was below the growth rate of wild type cells. *tps2Δ* expressing RiTPS1 or RiTSL1 could not grow at 37°C (see Fig. 22b).

In both assays RiTSL1 was expressed either as full-length protein or as –SP protein, missing the first 22 aa, which were predicted to be a signal peptide. This predicted SP of RiTSL1 is functional when fused to an invertase protein in yeast (see Fig. 21b, Section 2.4.1). Removing the predicted SP ensures that the protein stays inside the yeast cell, on the other hand a possible important function of the N-terminal could only be observed in the full length protein.

The results confirm the function of RiTPS1 as TPS and of RiTPS2 as TPP. No function was observed for RiTSL1.

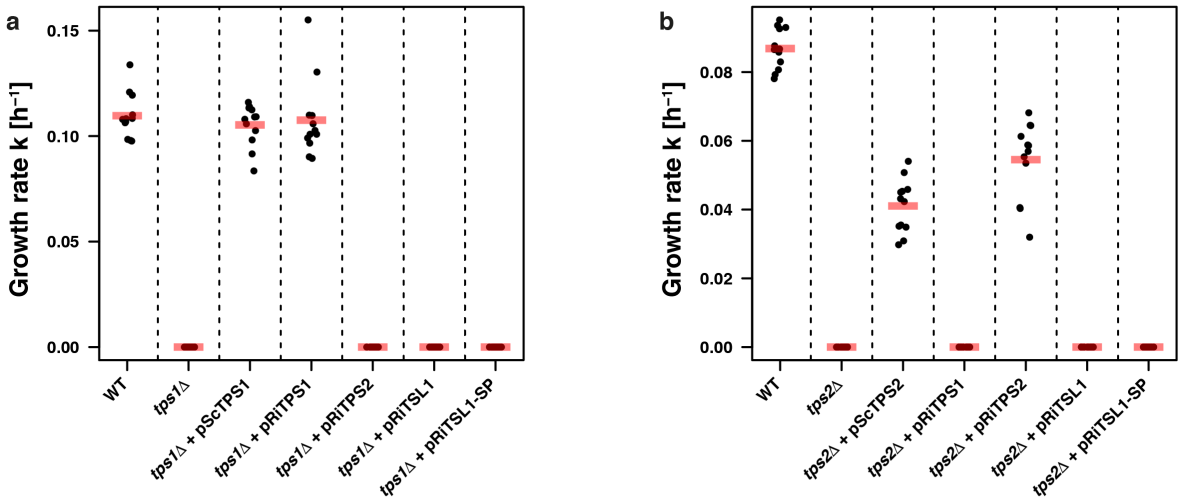

Fig. 22: Yeast complementation assay shows that RiTPS1 and RiTPS2 are able to complement the respective yeast phenotypes. Red bars indicate the mean value of the biological replicates.(a) *S. cerevisiae* strains were grown in liquid medium containing glucose. Wildtype and *tps1*Δ mutant were transformed with an empty vector (WT, *tps1*Δ) and with the indicated constructs. *RiTSL1* was transformed as full length CDS and the truncated version without the predicted secretion peptide (-SP). (b) *S. cerevisiae* strains were grown at 37°C.

3 Discussion

Functional genomics consists of three key stages: whole genome/transcriptome sequencing and assembly; annotation and comparative analysis of a whole genome/transcriptome; functional characterization of candidate genes found in whole genome/transcriptome analysis.

With the rise of new sequencing technologies, the possibility emerged to obtain great amounts of accurate sequencing data. These data can help to understand organisms and ecosystems in a greater context and allow the comparison and deduction of information on a large scale. By finding commonalities or differences between closely related species (and also more distantly related species), conclusions about the evolutionary context and function can be drawn.

A good example for this is the arbuscular mycorrhiza symbiosis. Even though important symbiosis genes have been identified, still relatively little is known about how these fungi interact with their host plants. By comparing large data of symbiotic plants, a symbiotic "toolkit" of plant genes was identified. Presence or absence of these genes in host and non-host species led to the conclusion that the loss of AM symbiosis was followed in several cases to the loss of parts or the entire set of genes required for symbiosis[5]. In the same study they also compared which other genes are absent in non-hosts and present in hosts. Thereby sets of 175 *M. truncatula* and 167 *Oryza sativa* genes (65 shared between the two) were identified. These sets were enriched in already known symbiosis related genes and contain promising candidates for further functional characterization.

In another example the analysis of large sequencing datasets of glomeromycotan fungi showed that there is a conserved meiosis machinery in all of these species[109]. With this knowledge new light is shone on the hypothesis of "ancient asexuality" in these fungi, and specific experiments can be planned to identify the distinct mechanisms of their sexuality.

The long-term interaction between plants and AM fungi appears to follow similar rules in a large group of host plants[42]. This points to a more conserved effectome for the interaction with their hosts, in contrast to pathogenic microbes, where effectors are under high selective pressure as plant immunity is adapting[128].

By comparing the genomes of *R. irregularis* and *R. clarus* a set of conserved effector candidates was identified[61]. With this availability of a distinct set of genes, identified through genome-wide analyses, arises the opportunity to characterize these genes in more detail and find how plants and fungi communicate.

In this study, the possibility of assembling a complete gap-less genome with the help of 2ndGS short and 3rdGS long sequencing reads was tested on the example of a bacterial strain, *R. leguminosarum* Norway. Further the symbiotic transcriptome of *G. pyriformis*, as an

early diverging member of the Glomeromycota with a different symbiotic partner and life-style, was analyzed in order to compare the expressed genes to other glomeromycotan fungi. Lastly a set of *R. irregularis* effector candidates, identified through a genome-wide screen, was used for functional characterization, leading to insights on fungal trehalose metabolism.

3.1 Assembly of the complete and gap-less genome of *R. leguminosarum* Norway

With 2ndGS short sequencing reads a lot of data were produced over the years. However, assemblies of the relatively short sequencing reads are challenging and the resulting genomes are often fragmented. This was also the case for the assembly of the *Rl* Norway genome performed with CLC Genomics Workbench (Qiagen) from about 15 million Illumina reads. The resulting genome had a length of 7,765,030 bp scattered on 83 contigs. The high amount of contigs is a problem of resolving sequence conflicts, which especially occurs in regions containing repeat sequences, pseudogenes, transposable elements, rRNA, LTR retrotransposons, or telomeres.

Here it is shown that with the combination of 2ndGS (Illumina) and 3rdGS (MinION) techniques the assembly of a complete gap-less genome of *Rl* Norway was possible. Within the Unicycler program, which was used for performing the complete assembly, several sub-processes give an insight on the advantage of using reads from both of these sequencing technologies. In the first sub-process only short reads from Illumina sequencing were used to create an assembly. Fig. 7a shows that this assembly with 194 contigs and 22 dead ends still has many unresolved regions, even more than the assembly done in CLC Genomics Workbench. Just by taking one step further and using the long reads from Nanopore sequencing, all of these regions could be resolved and resulted in six circular replicons, one chromosome and five large plasmids (see Fig. 7e). In comparison with the original CLC-assembly this hybrid assembly is only about 23,000 bp larger. However, instead of a distribution on 83 contigs, now the assembly is arranged on the six complete circular replicons, giving a complete picture of the genome.

That a combination of long and short reads increases the assembly quality greatly, even though the single long reads have high error rates was also demonstrated by Goodwin and colleagues. They therefore implemented a hybrid error correction for Nanopore long reads, similar to error correction pipelines for SMRT sequencing reads[129]. With this technique they showed for the *S. cerevisiae* genome that the assembly with Nanopore reads leads to greatly increased N50 values and is more contiguous across the entire size spectrum and across problematic regions than an Illumina-only assembly[130]. The possibility of having high quality, (close to) gap-less genomes opens many doors for research especially in non-coding and repetitive regions. Also for the research on AM fungi this technique is promising as here many assumptions are made on the basis of absence genes. With an estimated size of ~150 Mb the genome of the model AM fungus *R. irregularis* is almost 20 times larger than that of *Rl* Norway. Complete and gap-less genome assemblies are more difficult with that size, but

improvements especially in repetitive regions, or regions with duplicated genes are to be expected.

3.2 Data-mining on the *G. pyriformis* symbiotic transcriptome reveals common evolutionary origin and special adaptations in glomeromycotan symbioses

G. pyriformis poses a model system for an early diverging glomeromycotan fungus that lives in symbiosis with the cyanobacterium *N. punctiforme*. Compared to other Glomeromycota it has a symbiotic partner from a different kingdom (bacteria vs. plant) and it has a quite different life-style. In this symbiosis the fungus seems to be in control over the bacterial symbiotic partner, whereas in other Glomeromycota the plant seems to be in control over the fungus inside its roots[131]. The exchange of nutrients though appears to happen in the same way via a symbiotic interface harboring membrane transporters on each side. The comparison of *G. pyriformis* and other glomeromycotan fungi on a transcriptome level offers the opportunity to identify evolutionary changes for the adaptation to different hosts. Due to the nature of this symbiosis it was possible to construct a cDNA library of symbiotic fungal transcripts, removing bacterial transcripts by taking advantage of the lack of polyadenylation[81].

The analysis of a whole transcriptome opens up possibilities to find new important symbiosis genes and make large-scale comparisons between similar, but also more divergent organisms. In this work it was possible to sequence the transcriptome of *G. pyriformis* bladders, the symbiotic organs. Although the preparation of the cDNA library contained a step using poly-T primers, in order to only select eukaryotic sequences[81], the original assembly contained many bacterial sequences. In the work of Holger Martin, who prepared the original cDNA library, 56 clones of the cDNA library were sequenced. Out of these 56 clones, six didn't show homology to any eukaryotic genes using BLAST (best hit)[132]. About the same amount of bacterial sequences (~10.5%) in the present assembly of Illumina sequencing reads was found. The high contamination rate is most likely due to non-sterile culturing conditions of the original *G. pyriformis* culture used for cDNA library construction. Together with non-specific binding of poly-T primers in the cDNA library construction this can be the origin of contaminations. Nevertheless, after filtering and removing contaminant sequences, a transcriptome comprising 17,604 NRVTs (average length: 705 nt; N50: 1,110 nt; GC content: 41%; total length: 12.4 Mb; average coverage: 49x) could be generated, encompassing around 33.5% of the primary cDNA library insert reads (see Fig. 10a). In blobplot analysis, more than 90% of these NRTVs did not show similarity to known DNA sequences (blastn, e-value < 10; see Fig. S2d). Comparisons on amino acid level using blastx[28] revealed 64% NRVTs with matches against database sequences (e-value $1e^{-5}$), of which 8% were best hits to the genome of *R. irregularis*. All best matches are either to

eukaryotic sequences or to sequences of unknown origin, indicating efficient removal of contaminating sequences of bacterial origin.

In comparison to other available glomeromycotan transcriptomes (Table 5) a lower number of obtained NRVTs and accordingly a smaller total length is apparent. This could be a consequence of lower starting material complexity: the *G. pyriformis* cDNA library was constructed from symbiotic bladders with some attached hyphae. In contrast the *R. irregularis*[114], *G. rosea*[60], and *G. margarita*[89] transcriptomes originate from a mix of different morphological structures, resulting in a larger variety of expressed genes. Interestingly, the *G. pyriformis* transcriptome has a higher GC content than the other Glomeromycota (41% vs. 31-33%; Table 5). At present, without access to more glomeromycotan genomes and transcriptomes, it is unclear whether this is a sign of relative phylogenetic distance or a consequence of differences in symbiotic life style.

A good measure for the completeness of a newly sequenced genome or transcriptome is the BUSCO[115] assessment. In this analysis, a set of 290 core fungal genes was checked for presence in the assembled data. This analysis does not only provide an idea about completeness, but also about how much duplication or fragmentation is present in the assembly. In order to validate the filtering steps to remove contaminant NRVTs, the BUSCO results of the final transcriptome and the original assembly were compared. There was a slight increase in missing BUSCOs and no change in the fragmentation. Although there are less complete BUSCOs after filtering, most of them were duplicated in the original assembly. This indicates that the stringent filtering conditions did not result in a significant loss of *G. pyriformis* transcriptome sequences.

In the *R. irregularis* transcriptome the number of complete and single-copy BUSCOs is comparable to that of *G. pyriformis* (154 vs. 150). However, there are more fragmented and duplicated BUSCOs in *R. irregularis* (106 vs. 90), resulting in a lower number of missing BUSCOs (30 vs. 50). The comparable completeness between the *G. pyriformis* and *R. irregularis* transcriptomes is surprising given the fact that the former contains 31% less NRVTs and has a 38% shorter total length (see Table 5). This further supports the hypothesis that the differences may concern differentially rather than constitutively transcribed genes. Both transcriptomes are not complete, as can be seen in comparison to the predicted complete *R. irregularis* transcriptome, although even here 9 BUSCOs are missing. The transcriptomes of *G. rosea* and *G. margarita* both appear to be more complete. However they also contain a much larger number of NRVTs. This is probably due to incomplete assembling of redundant transcripts, gene duplication, or expansion of gene

families[60]. The high number of duplicated BUSCOs in the *G. margarita* transcriptome would support the hypothesis at least for this species (see Fig. 10b).

The quality of the *G. pyriformis* transcriptome is remarkable considering the nature of the starting material and confirms the validity of the dataset for further analysis. However taking into account the low sampling complexity, absence of genes in the *G. pyriformis* transcriptome that are differentially expressed in other species does not necessarily indicate absence in the genome itself. This should not be the case for genes encoding enzymes involved in general biosynthetic pathways. Construction of the cDNA library involved removal of transcripts smaller than 600 bp and a PCR amplification step[81,132], which might have introduced single nucleotide changes by polymerization errors. Also, quantitative data may include possible amplification biases in the PCR as well as during amplification in *E. coli*. The final assembly of 17,604 NRVTs can be considered to be a conservative estimation of the *G. pyriformis* transcriptome present in the cDNA expression library.

The genome organization of Glomeromycota has been a thoroughly debated subject[133-135]. Early studies showed a high degree of polymorphism in coding regions of certain genes in clonally grown isolates and it was proposed that the multinucleate organization in these fungi helps to cope with mutations in individual nuclei. However, more recent studies applying whole genome/transcriptome and single nucleus sequencing revealed low SNP frequencies with a range of 0.1-1 SNP/Kb[106]. In addition, one study analyzing single nuclei genomes from different *R. irregularis* isolates showed that some of them have SNP allele frequencies of 0.5, indicating a dikaryotic organization[108].
In the *G. pyriformis* transcriptome a frequency of 5.84 SNPs/Kb (72,468 SNPs in 6,656 NRVTs) was identified (see Fig. S3a). 3,246 NRVTs were highly polymorphic (> 10 SNPs). The SNP allele frequency distribution indicates a homokaryotic genome organization (see Fig. S3b). *G. pyriformis* has a higher SNP frequency in comparison to other Glomeromycota, although it still is below the value of what would be expected for true heterokaryotic fungi (> 10 SNPs/Kb) [118]. The amplification and cloning steps during preparation of the cDNA library and the sequencing library before the final sequencing might have contributed to the larger SNPs number. For other Glomeromycota the possibility was proposed that the symbiosis with the plant host may have a regulatory effect on which nuclei to propagate and thereby on genome stability[106,131]. Such regulation would not be expected to happen in interaction with the cyanobacterial microsymbiont.

Glomeromycota fungi are believed to be ancient asexuals[136], because no sexual processes have ever been observed. The identification of meiosis-related genes[120], a vast expansion of

mating-type high-mobility group (*MATA-HMG*) genes[110], and the presence of different mating type loci encoding for homeodomain proteins[137] in the genome of Glomeromycota suggest a form of cryptic sex for this phylum. An aim of this study was to investigate whether this potential is also present in *G. pyriformis*. Using a set of 86 fungal meiosis-related genes[119], seven of which are meiosis-specific, a reciprocal blastx/tblastn search against the *G. pyriformis* transcriptome was performed. 57% of these genes are present (see Fig. 12; Table ES1), which is similar to the numbers identified in the transcriptomes of *G. rosea*[60] and the genome of *R. irregularis*[120]. In comparison, in the *R. irregularis* transcriptome only 26% of the meiosis genes were found[114]. In the more complete *G. margarita* transcriptome 70% of the meiosis gene set could be identified.

Of the seven meiosis-specific genes in this gene set, five (*SPO11*, *DMC1*, *MSH5*, *HOP2*, *MND1*) are present in the *G. pyriformis* transcriptome, and among the investigated transcriptomes only *G. margarita* contains all seven genes (additionally *MSH4* and *REC8*). All seven are also present in the *Rhizophagus* genome but only two in the *R. irregularis* transcriptome. This suggests that the genes could also be present in the other genomes but are probably not sufficiently expressed in the tissues used for sampling.

The *G. pyriformis* transcriptome was searched for *MATA-HMG* genes and it was possible to identify 36 NRVTs with similarity (blastn, e-value < 1) to a set of 76 *R. irregularis MATA-HMG* genes[110]. Although, compared to *R. irregularis*, *G. pyriformis* contains only half as many of these genes, there is still an expansion of this group in comparison to other fungi, which contain a lot less (usually < 10) *MATA-HMG* genes[110]. Together with the presence of meiosis-related genes this indicates that *G. pyriformis* might be able to perform sexual processes and that this ability is conserved among the Glomeromycota.

A good starting point for transcriptome analyses is to predict biochemical functions to the proteins encoded by NRVTs. This is done by comparing the predicted protein sequences to sequences from other organisms with already known function. The database of eukaryotic orthologous groups (KOG) is a collection of functionally characterized and categorized protein sequences that can be used for this purpose.

Comparison with the KOG database[138] revealed KOG functional classifications for only 19.9% of the translated NRVTs (see Fig. S4a). Not surprisingly, when compared to sets of predicted fungal proteomes, conserved proteins with homologs present within Glomeromycota, Ascomycota, Basidiomycota, and Zygomycota show a much higher proportion of ortholog groups with KOG classification than Glomeromycota-specific proteins or proteins without known fungal homologs (49.3% vs. 5.4% vs. 2.0%). Interestingly, the KOG functional class distribution of *G. pyriformis* is quite distinct from the other

Glomeromycota (see Fig 12a; Fig S3b). The main difference between other glomeromycotan fungi and *G. pyriformis* is found in "signal transduction mechanisms" and "cell wall/membrane/envelope biogenesis". These two functional classes are highly represented in *R. irregularis*[114], *G. rosea*[60] and *G. margarita* possibly due to the adaptation to mycorrhizal symbioses, which involves intensive signaling processes[60]. As *G. pyriformis* is engulfing its bacterial microsymbiont, which subsequently has no other choice than to work in this new environment, it might not be necessary to exchange many signals. The overrepresentation of "cell wall/membrane/envelope biogenesis" in arbuscular mycorrhizal fungi is on par with arbuscule development. In this process a large membrane surface is built. In *G. pyriformis* on the other hand it is not necessary, as the symbiotic bladders do not comprise an equally vast enlargement of membranes. The overall distribution of NRVTs in the functional classes is much more similar to other fungi outside the Glomeromycota, independently of their life style (symbiotic, pathogenic, saprophytic). This is nicely illustrated by the results of a principal component analysis (see Fig. 11b).

This difference indicates that the evolution of the symbiosis between Glomeromycota and plants went hand in hand with a major change in the expression profiles of distinct functional classes. Supporting this hypothesis, *G. pyriformis*, as an early diverging member of Glomeromycota and not engaging in arbuscular mycorrhiza symbiosis, did not go through the necessary adaptations and still has a similar functional class expression set-up as fungi from other clades.

The set of missing glomeromycotan core genes, as proposed by Tang and colleagues[60] is another example of the adaptations of glomeromycotan fungi to symbiosis. Many biosynthetic processes are no longer performed by the fungus, but by the plant. Therefore these fungi lost the ability to produce for example thiamine, or to do fatty acid de novo synthesis.

Even in the early diverging member of this group, *G. pyriformis*, most of the genes of this set are missing. In a first screen five NRVTs coding for genes in the MGCG set were identified. Subsequent blastx against the NCBI database though revealed that four of these NRVTs are most probably bacterial contaminations that were not identified in the previous filtering steps based on blastn results. One NRVT had similarities to *URE2*, a nitrogen catabolism repression transcriptional regulator that was also found in *R. irregularis*[60].

These results suggest that the cyanobacterial symbiont took over all the necessary biosynthetic pathways known from plant - glomeromycotan fungi interactions. This indicates that dependency on a symbiotic partner must have developed before the split of *G. pyriformis* from the other Glomeromycota. Thus at some point a partner switch either from

cyanobacteria to green algae/plant, or the other way around probably has happened. Further indications for a possible partner switch are the close relationships of AM fungi in bryophytes with cyanobacteria. Here it was shown that when cyanobacteria are more abundant in the soil, AM fungi are less likely to associate with the hornwort as a symbiotic partner, possibly caused by a functional relationship between glomeromycotan fungi and cyanobacteria[48].

In several obligate biotrophic pathogenic and in mutualistic ectomycorrhizal and AM fungi plant-microbe interactions the repertoire of carbohydrate-active enzymes (CAZymes) is reduced[60,114,139-142] compared to non-obligate phyto-pathogens[143]. These CAZymes are usually involved in degrading plant cell wall (PCW) polysaccharides and the reduced number of them likely prevents immune responses of the plant[60,114,144]. In the symbiotic bladders of *G. pyriformis* that were used in this study, the symbiosis does not involve a plant-partner but a cyanobacterial one. Therefore an even lower number would have been expected, considering that there are no PCW polysaccharides to degrade in the symbiotic bladders.
A total of 72 NRVTs encoding for 42 CAZyme families were identified. As in the *G. rosea*[60] and *R. irregularis*[114] transcriptomes, no polysaccharide lyases (PL), were found. Of these 72 NRVTs seven belong to families that are plant cell wall degrading, and 21 belong to chitin metabolism involved families. Compared to *G. rosea*[60], *R. irregularis*[114] and *G. margarita* (CAZyme analysis in this work) the numbers of CAZymes relative to the number of NRVTs are comparable. In total though, there are much less CAZymes in *G. pyriformis* (72) compared to the other analyzed transcriptomes (*G. margarita*: 632, *G. rosea*: 293, *R. irregularis*: 94). The percentage of plant cell wall degrading families is similarly low in all four fungi, whereas chitin metabolism is overrepresented in the *G. pyriformis* CAZyme families. Plants are able to detect the fungal chitin oligosaccharides and subsequently activate their defense responses[145,146]. Since *G. pyriformis* does not interact with a plant host, chitin metabolism might be less restricted than in the other Glomeromycota.
A clear difference between *G. pyriformis* and the other Glomeromycota is visible in the glycosyl transferase (GT) family. About one third of the CAZymes that are not involved in chitin metabolism or PCW degradation in the other Glomeromycota belong to GT families. *G. pyriformis* has only about 12.5%. GTs are enzymes that catalyze the glycosylation of proteins, which can be important for their secretion, stabilization or localization[147]. The low abundance of GTs in the *G. pyriformis* symbiosome could be due to its specific symbiotic life-style. The control over the cyanobacterial partner may be the reason for less communication/secreted proteins from *G. pyriformis*, thereby less need for GTs.
Furthermore also saprophytic fungi use CAZymes in order to degrade polysaccharides from the soil[148]. The low number of CAZymes in *G. pyriformis* therefore might also be an indication

for the adaptation to a symbiotic lifestyle in which additional carbohydrate nutrition from the soil is dispensable.

The nutrient exchange between mutual symbiotic partners plays a key role in the establishment and maintenance of a successful relationship. In arbuscular mycorrhiza symbiosis the identification of transporters on the fungal (e.g. GiMST2[75], GigmPT[94], GintAMT1, 2, 3[72,98,99]) as well as on the plant side (e.g. SWEETs[79], OsPT2, 6, 11[96], LjAMT2;2[149]) is an indication of nutrient exchange at the interface between the two organisms.

In *G. pyriformis* a symbiotic monosaccharide transporter (GpMST1[81]) was identified. Also three ammonium transporters (GpAMT1, 2, 3[121]) were found, of which two (GpAMT1, 2) were localized to the plasma membrane in yeast and GpAMT3 was localized to the vacuolar membrane. Only GpAMT1 was able to complement an ammonium uptake deficient yeast mutant. Thus, nutrient exchange during the symbiosis likely occurs by similar means as between AM fungi and plants[150].

In this analysis using a reciprocal BLAST approach a total of 1726 NRVTs that are putative transporters could be identified. Blastx of these putative transporters revealed that 287 of the NRVTs might be of bacterial/protiste/plant origin and had to be removed from the analysis. This leaves 1439 NRVTs coding putative *G. pyriformis* transporters in 219 transporter families (see Table ES2). The most abundant transporter families are the nuclear pore complex (NPC) family, the domain of unknown function 3339 (DUF3339) family and the ATP-binding cassette (ABC) superfamily. Performing this analysis using the same parameters for the transcriptomes of *G. margarita*, *G. rosea* and *R. irregularis* resulted in 5194 NRVTs (229 families), 3307 NRVTs (259 families) and 2182 NRVTs (215 families) respectively.

Also in *G. margarita* and *R. irregularis* the NPC family is the most abundant transporter family. The NPC family contains proteins that form the nuclear pores localized in the nuclear envelope and facilitate macromolecular transport in both directions[151]. It also plays an important role in many nuclear processes, for example gene activation and cell cycle regulation[151]. As the NPC is an integral part of a functioning nucleus, the presence of NRVTs coding for NPC proteins was expected. However the high amount of 135 NRVTs is surprising. It might be that this expansion of proteins with similarity to NPC proteins is actually due to a high similarity to protein kinases, which is the superfamily of NPCs.

Although the number of putative transporter NRVTs in *G. pyriformis* is lower than in the other glomeromycotan transcriptomes, they have similar numbers when compared to the total number of NRVTs in the transcriptomes. Also in regard to superfamilies, class and subclass of the transporters, *G. pyriformis* shows similar numbers as *R. irregularis* and *G. margarita*.

The already published ammonium transporters (AMT1, AMT2, AMT3)[121] as well as the monosaccharide transporter MST1[81] of *G. pyriformis* could be identified as well as an additional NRVT showing similarities to *R. irregularis* MST4[75] and NRVTs similar to *Rhizophagus sp*. phosphate transporters[74,92,122] and aquaporins[123]. These results indicate that indeed the symbiotic interface between fungus and bacterium or fungus and plant requires a similar composition of the transporter repertoire as already proposed by Schüßler and colleagues[81].

The secretion of certain proteins is used as a tool by many pathogens, but also symbiotic fungi, to evade host immunity. The effector SP7 from *R. irregularis* is one of the few glomeromycotan secreted proteins already characterized. When expressed *in planta*, this protein is localized to the nucleus, where it can interact with ERF19, a protein that is usually highly expressed when the host is challenged by a pathogen. The interaction with ERF19 counteracts the immune response of the plant and thereby facilitates fungal entry[63]. On the other hand fungal effectors that are secreted to their symbiotic partner may also activate certain cellular pathways in order to alter plant cell structure or function in a way beneficial for the fungus.

In the *Geosiphon-Nostoc* symbiosis the fungus is the macrosymbiont and there is no need to counteract any plant immune response, as the photosynthetic microsymbiont is simply engulfed into the fungus. On the other hand effectors that activate biosynthesis pathways are expected to be present. Therefore a much lower number of secreted fungal proteins was expected, than in other fungi of the Glomeromycota. In the *G. pyriformis* transcriptome, 89 putative secreted proteins were identified. In absolute numbers this is much less than in the *G. rosea* transcriptome and the *R. irregularis* genome (441 and 376 respectively)[59,60]. If compared to the transcriptome size, the relative numbers are similar to those in *G. rosea* (about 0.5% of all NRVTs). In the genome of the lichen-forming fungus *Endocarpon pusillum* 135 secreted proteins were found, which is less than in the three in that study analyzed pathogenic fungi[152]. It is also less when compared to the other Glomeromycota analyzed in the present study. As in *Geosiphon* in this symbiosis the microsymbiont is a photosynthetic cyanobacterium (or alga). Other than in lichen, in the *Geosiphon-Nostoc* symbiosis the microsymbiont is engulfed in the fungus. Therefore the symbiosis with a macrosymbiont, a plant, could require more “crosstalk” between the two partners than the symbiosis with a unicellular, or even engulfed, photobiont.

As of yet very little is known about fungal effectors. Unsurprisingly only for 17 of the 89 NRVTs, KOG functional classes could be predicted (e.g. predicted lipase, palmitoyl protein thioesterase, 5’-3’ exonuclease HKE1/RAT1, carbonic anhydrase). Overall it is rather

unexpected that *G. pyriformis* in its symbiotic state is expressing so many genes coding for secreted proteins. This finding suggests that there might indeed be some signaling with the cyanobacteria or degradation of bacterial components in the perisymbiotic space. To elucidate this further, the predicted proteins would have to be characterized in detail.

In the set of predicted secreted proteins three sequences were found, which have a very high sequencing coverage compared to the other NRVTs. Of all *G. pyriformis* reads 22.3%, 5.0% and 3.0% mapped to these three genes. BLAST analysis didn't result in any known sequences on nucleotide level and only a few hits on protein level with hypothetical proteins and a maximum identity score of 52%.

The most abundant of these proteins contains six almost identical amino acid repeats and was therefore named GpRIC1 (**R**epeat containing proteins **I**n symbiosis with **C**yanobacteria). As all three of them share certain characteristics, the other two were subsequently named GpRIC2 and GpRIC3. All three of these proteins harbor a secretion peptide, Kex2 recognition sites and a poly-A-tail. All of this indicates that they are not only secreted, but also cleaved into smaller peptides and that they are of eukaryotic origin and not from any contamination. Further, the three proteins show a very characteristic amino acid composition. With 22-26% glycine and 33-39% asparagine these proteins are mostly non-charged (neutral). This low complexity indicates that the proteins are disordered, and could be amyloid forming. Several online tools for prediction of protein disorder and aggregation support this (Globplot2[153], pRANK[154], Waltz[155]).

Amyloids are proteins that can aggregate and form stable fibers with a cross β-sheet structure in which the β-strands run perpendicular to the fiber axis. These amyloids are insoluble and can be assembled and disassembled[156]. In humans they are often associated with neurodegenerative diseases, as Huntington's, Alzheimer's or Parkinson's disease.

Prions are a special kind of amyloids that are able to stimulate other prions in nonprion conformation to change into the prion conformation. In yeast this conformational change goes hand in hand with a change in phenotype, which can be inherited from mother to daughter cell[157].

In microorganisms there are also different classes of functional amyloids that can be cytotoxic but can also serve as support in biofilms or adhesins[156]. In fungi a form of amyloid proteins is known that enables aerial hyphae formation and is important for the transition of growth in a hydrophilic environment to growth in a hydrophobic environment (air)[158]. In the case of *G. pyriformis* there are several possible scenarios.

First the localization of these secreted peptides is undetermined. As they have a secretion signal it is possible that they are secreted to the outside of the bladders, thereby building

either a structural support or a water-repellent surface to enable these bladders to evade the soil and grow into the air.

The second possibility is that they are secreted to the perisymbiotic space to build a matrix surrounding the cyanobacteria. These stable assemblies of peptides possibly create a firm structural support for the bladders, to keep cyanobacteria in the upper part of the bladders, or to prevent extensive growth and replication of the cyanobacteria. A less likely reason would be *G. pyriformis* providing these peptides as nutrition to the cyanobacteria. The cyanobacteria are photoautotroph and are not dependent on *G. pyriformis* for providing a carbon source. Also the cyanobacteria grow heterocysts inside the *G. pyriformis* bladders. Usually nitrogen is fixed in heterocysts, making the cyanobacteria thereby also independent regarding the nitrogen supply.

It is further known from cyanobacteria that they are able to produce many different kinds of toxic compounds[159]. A less likely reason for the secretion of GpRICs by *G. pyriformis* is the counteraction of cyanotoxins as a filter for these toxins would have to be very specific and able to let other metabolites and nutrients pass.

The identification of these promising genes demonstrates the power of functional genomics approaches. Further experiments to functionally characterize these genes could include heterologous expression in yeast in order determine if prions can be formed, which can be inherited or transmissible. Also the expression in *E. coli* could give insight to the possible amyloid fibril assembly and disassembly. Experiments including the combined expression of the three proteins or the shorter peptides can offer information, if all of them can act independently or if the formation of structures is dependent on a mix of two or all three of the GpRICs. For localization studies antibodies can be used *in situ*, although this option is only feasible if *G. pyriformis* can be isolated and grown in culture again. Elsewise the heterologous expression of these genes in pathogenic fungi could further give insights into a possible role in infection or control over the host.

3.3 Downregulation of *RiTSL1*, a gene involved in trehalose biosynthesis

In pathogenic as well as symbiotic microbes so called effector proteins play a crucial role in the successful infection of the host plant. Effectors are defined as “molecules that manipulate host cell structure and function, thereby facilitating infection (virulence factors or toxins) and/or triggering defense responses (avirulence factors or elicitors)”[128]. In plant pathogenic microbes already several effectors have been identified and characterized. Here there is a great selective pressure on the evolution and thereby diversification of effectors. The so-called zig-zag-model describes this ongoing evolution, in which immunity mechanisms of the plant induce the evolution of pathogen effectors to the point that new plant immunity mechanisms have to evolve[160]. So-called microbial- or pathogen-associated molecular patterns (MAMPs or PAMPs) trigger the plant immune system, which in turn evolves PAMP-triggered immunity (PTI) mechanisms to stop the pathogen from infection. To circumvent these mechanisms, the pathogen evolves effector molecules. In turn the plant evolves again mechanisms against the effect of these molecules. These stages alternate in an ongoing evolutionary process[160].

On the other hand in AM fungi the wide host range combined with similar infection mechanisms and structure suggest that effectors playing a fundamental role in the infection process are conserved between glomeromycotan species. This idea in mind Sedzielewska and Brachmann recently published the predicted effectome of *R. clarus.* They found a set of 64 putative effector genes that is conserved between the *R. clarus* and *R. irregularis* genomes[61].

Thanks to recent whole genome/transcriptome sequencing projects for different glomeromycotan fungi there is a large amount of data available on putative effector proteins[59,60]. In the genome of *R. irregularis* 376 genes with signal peptide were identified, 79 of which are induced *in planta*[59] and in *G. rosea* 441 potential secreted proteins were found[60]. Despite the availability of candidate effectors, only two effector proteins were so far characterized in more detail in AM fungi (*SP7*[63] & *SIS1*[64], both in *R. irregularis*).

As *R. irregularis* is a coenocytic fungus, containing thousands of nuclei in one common cytoplasm, there are as of yet no methods to transform this fungus or create mutants of it. Therefore for the functional characterization of effector candidates a “host induced gene silencing” (HIGS) approach was chosen. HIGS is a method in which the host plant is, in this case transiently, transformed with an RNAi construct targeting genes in the symbiotic partner[75]. In theory the primary, about 200 bp long, double stranded (ds) RNA is processed by Dicer in the plant host. This results in small dsRNAs that are about 21 bp long. By an unknown process these small dsRNAs are translocated to the fungal cytoplasm, where they are loaded into the AGO complex and the target mRNA transcripts are degraded (see Fig. 6).

In pathogens as the powdery mildew *Blumeria graminis*[161] or the necrotrophic fungus *Fusarium graminearum*[162] this method was shown to be very effective to identify proteins important for infection. In *Fusarium* it was also shown that direct application of a dsRNA to the pathogen on leaves targeting the three fungal cytochrome P450 *lanosterol C14α-demethylases* showed inhibition of fungal growth[163].

In the present study a significant downregulation with an HIGS approach could be observed for the effector candidate *NLS_7749*/*RiTSL1*. The lower expression of *RiTSL1* in HIGS roots appeared to be not only in the GFP-positive roots, but also in the GFP-negative roots of the same plants. The downregulation in GFP-positive roots as well as GFP-negative roots suggests that the RNAi effect is systemic, either via the plant, or via the fungus. Further experiments as for example split root setups or transformed and non transformed plants with one connecting fungal system in one pot would be necessary to confirm which systemic way is possible for the RNAi effect.

In this successful experiment, in which a lower expression of *RiTSL1* in HIGS roots as compared to the EV control roots was detected, also lower root length colonization was observed (see Fig. 18a). Furthermore the shape of arbuscules was abnormal and stunted (see Fig. 18c, representative picture from different experiment). Also known symbiosis induced plant genes (*AMT2.2*, *Bcp1*, *PT4*, *SbtM1*), were expressed at lower levels in HIGS roots, confirming the lower root length colonization (see Fig. 18d). These results indicate that *RiTSL1* plays an important role in colonization.

TSL1 is a gene involved in trehalose biosynthesis. Trehalose is a stable, nonreducing disaccharide composed of two glucose-moieties linked by an α,α-1,1-glycosidic bond. It can be found in most organisms except vertebrates and usually functions as an osmoprotectant, for protein and lipid bilayer stabilization, or as a storage sugar[164,165]. It was also shown that in many pathogenic fungi an intact trehalose biosynthesis pathway is important for virulence. For example in the leaf and glume blotch fungus *Stagonospora nodorum* the absence of *TPS1* leads to smaller lesion size on leafs and reduced spore formation[166].

The other two fungal genes involved in this biosynthetic process (*RiTPS1* & *RiTPS2*) were identified in *R. irregularis* and also analyzed in the HIGS experiment. The expression of *RiTPS1* was significantly lower in GFP-positive RiTSL1-HIGS roots compared to GFP-positive EV-control roots. *RiTPS2* showed more variation in expression between the biological replicates and there was no significant lower expression in HIGS roots, although a trend is visible (see Fig. 18e). This implies that the three genes involved in trehalose biosynthesis are co-regulated. These results also suggest that, as in yeast, the proteins involved in trehalose biosynthesis might act closely together, or even form a complex. In this

case though it is quite unlikely that RiTSL1 is secreted towards the plant, whereas the other two players stay in the fungal cytoplasm (see below for further discussion).

The results of downregulation of *RiTSL1* in a first experiment (see above, Table 6, HR4) looked very promising. To quantify the reduced arbuscule size and for confirmation of the results, further experiments were performed. In the following six experiments though, no downregulation of *RiTSL1* in HIGS roots as compared to EV control roots was observed. Also decrease in root length colonization and the stunted arbuscules were only observed in one further experiment (HR5; see Table 6). The experiments were set up in the same way with one difference: In the successful experiment HR4, EV roots and HIGS roots were harvested on subsequent days at different times and EV plants were watered the night before harvesting.

In 1998 Schellenbaum *et al*. showed that mycorrhized maize plants that are exposed to a 12-day drought treatment do not show a significant difference in root length colonization, and mycorrhizal structures were unchanged. The trehalose content though was much higher in drought-exposed plants compared to well-watered plants[167]. Shachar-Hill and colleagues showed that trehalose in mycorrhized roots is mainly produced by the fungus[82]. This indicates that the results of HR4 are not due to harvesting during drought stress.

The question is: why can the results not be reproduced? Many unknown factors play a role in the setup of this experiment, and it is difficult to pin the reason to one of them. The hairpin dsRNA that is expressed *in planta* could be detected via PCR on the cDNA of HR4 and two experiments, where no downregulation could be shown (see Table 6). Possibilities are that the dsRNA was not processed into small dsRNAs in the plants, that the translocation process did not work efficiently, or that small dsRNAs could not be brought into vicinity to the target transcripts. Some of these possibilities would still be feasible to test. Small RNA extraction can be performed in order to show whether dsRNA is processed into small dsRNAs. The translocation of small RNAs into the fungus and whether the small dsRNAs get in contact with their mRNA target however seems difficult to prove.

For eight further targets on seven additional effector candidates HIGS experiments were performed and none of them showed downregulation (see Fig. 16). Also in literature there are only two reports for a successful downregulation of *R. irregularis* genes, namely *MST2* in 2011[75] and *SIS1* in 2016[64]. For HIGS on the fungal *MST2* the main differences are a slightly longer target region (300 bp vs. 200 bp), a different promoter (p35S vs. pLjUbi), a different plant host (*Medicago truncatula* vs. *L. japonicus*), and no off-target prediction[75]. Also for HIGS targeting *SIS1* a slightly longer target region was used (330 bp vs. 200 bp); the same vector and plant host as for *MST2* were used[64].

As the expression of the hairpin construct was also detected in experiments without downregulation, the difference in promoter is unlikely to be the crucial point. A longer target region would lead to more small RNAs and a higher likelihood of the small RNAs to bind to the target mRNA. Also the plant host may play a role in transferring the small RNAs to the fungus, as it is unknown how this process works and it should be considered for future experiments.

Even though the silencing effect mediated by HIGS could not be reproduced, the obvious phenotype of small and stunted arbuscules and reduced colonization that was observed in the first experiment led to the idea that *RiTSL1*, which plays a role in the fungal trehalose biosynthesis, may play an important role in the establishment and maintenance of the symbiosis.

It was shown earlier that arbuscules are relatively short lived structures and can go from maturation to degeneration in as short as 24 h[168]. In the case of diurnal rhythmicity of maturation/degeneration cycles, the time of root harvesting for observation of arbuscule phenotype would need to be taken into consideration.

In the present study the time of harvest for control and HIGS roots was not the same, therefore the changes in arbuscule size in wild type roots over the course of one day were analyzed. It could be shown that there is only slight but significant variation in the area of a plant cell that is filled by the arbuscule at different time points in one day. It was shown that arbuscules are slightly smaller in the morning time (end of dark phase/beginning of light phase, see Fig. 18).

These results propose that for disruption of proteins with a very distinct phenotype such as above mentioned RiTSL1, but also plant proteins as MtSucS1[169], MtRAM1[170] or the fungal monosaccharide transporter MST2[75] the time of harvest will not change the outcome of the experiment. If on the other hand the observed phenotypes are rather subtle, the time of harvest should be taken into consideration carefully when planning the experiment.

Most genes that were found to have an effect on arbuscule size or shape are involved in nutrient provision or exchange[75,91,169,171-173]. The observed smaller size of arbuscules in the morning might also be linked to nutrient exchange. Sugar production in plants is dependent on sunlight. During the day CO_2 is photosynthetically fixed and excess sugar is stored in the leaves as starch. In the nighttime the starch is degraded and sucrose is released. In *A. thaliana* it was shown that in the late night phase there is a drop in sugar availability in the roots[174]. In *L. japonicus* colonized roots, this would mean that there are less monosaccharides from the plant available to the fungus. Therefore less glucose would be taken up into the fungal cytoplasm. Usually monosaccharides are quickly converted to

triacylglycerols (TAG), glycerol and trehalose to increase the sink strength for monosaccharides. For trehalose production, in the most common biosynthetic pathway two catalytic enzymes are important. TPS is producing the intermediate T6P and TPP subsequently dephosphorylates T6P to gain trehalose.

In order to identify if the slight differences in arbuscule size are connected to the expression patterns of these trehalose biosynthesis genes in *R. irregularis*, they were analyzed over the course of one day. These analyses were also performed to elucidate whether possible fluctuations in sugar availability in *L. japonicus* roots, as in *A. thaliana*[174], lead to a change in production of trehalose in *R. irregularis*. The expression of *RiTPS2* is much higher than the expression of *RiTPS1* and *RiTSL1* during the whole day (see Fig. 19). This might lead to a shift in the T6P/trehalose ratio towards trehalose and remove T6P as fast as possible from the arbuscule. In plants it was shown that lower T6P levels are closely correlated with levels of sucrose. A decrease in T6P leads to an increase in sucrose and less conversion into storage products[175]. Since this process takes place in the fungus it is not known if it has a regulatory effect on the sucrose level in the plant. It is possible though that the immediate shift towards trehalose and no accumulation of T6P also increases the sink strength to make more sucrose available. Although there is no significant difference between the expression at the different time points, a clear trend of day-night rhythm is visible for *RiTPS2* (see Fig. 19). The lowest expression was determined to be in the early morning. This low expression of *RiTPS2* coincided with the smaller arbuscule size in the morning. These findings indicate that there is indeed a fluctuation of sugar availability in *L. japonicus* roots at different times of a day. These fluctuations would mean that *R. irregularis* also retrieves different amounts of carbon at different times of the day and thereby adapts the production of trehalose in the intraradical mycelium. As it was shown in RNAi plants of a symbiotic sucrose synthase[169] and an experiment in which the expression of a fungal monosaccharide transporter was downregulated[75], inhibition of sugar flow to the fungus results in a higher number of not fully developed or prematurely senescent arbuscules. In the normal case of day-night sugar fluctuations the arbuscules are probably not completely broken down, due to the fact that the differences in sugar availability are only slight and sugar flow towards the fungus is not completely at a halt.

Avonce *et al.*[127] showed that in bikonts (e.g. Rhizaria, Alveolates, Plants) and unikonts (Amoebozoa, Opisthokonts) there are three classes of TPS-TPP genes. The fusion of TPS and TPP domains into one gene probably already occurred in some bacterial and archaea groups. Here both domains in one fusion gene were still functional. Gene duplication events were proposed to have happened in the last eukaryotic common ancestor (LECA). Some

genes in unikonts (fungi) have lost the TPP domain[127]. In yeast the function of some of these proteins is known. ScTPS1 is performing the function of TPS and a deletion of this gene results in a growth defect on rapidly fermentable carbon sources like glucose due to an uncontrolled influx of glucose into glycolysis[176]. ScTPS2 is acting as TPP. The deletion of this gene results in the accumulation of T6P and subsequently in a defect in response to heat-shock[124]. Normally trehalose plays an important role during abiotic stresses like heat, or salt stress due to its properties as a non-reducing sugar and osmoprotectant[177]. No enzymatic function was found for ScTSL1 and ScTPS3 and a regulatory function is proposed[178].

In *R. irregularis* three genes were identified, all of which contain both domains, TPS and TPP. The results of comparing the TPS domains of 156 proteins showed a clear clustering of the fungal proteins into three clades (see Fig. 20c). In Avonce *et al.*[127] these three groups were described as Fungi A, Fungi B and Fungi C, containing proteins with TPS activity, regulatory function, and TPP activity, respectively. From the known functions of the yeast proteins the conclusions were drawn that RiTPS1 has TPS function, RiTPS2 has TPP function and RiTSL1 has neither TPS nor TPP function.

In RiTSL1 a N-terminal region was identified that is conserved in many of the analyzed proteins in the Fungi B clade. This region is predicted to be a signal peptide in *R. irregularis* and most of the Basidiomycota, including saprophytes. It could be shown that the predicted signal peptides of RiTSL1 and RcTSL1 are indeed functional in a yeast secretion assay (see Fig. 21b). However, this alone does not yet prove that in the natural environment, *i.e. R. irregularis,* it is secreted and as well. The occurrence of the predicted SP also in saprophytes, which feed on dead plant material and have no need for secreted effectors targeting plant immunity or metabolism, speaks strongly against a putative role as an effector protein.

The conservation of this N-terminal region in 35 of 45 proteins, and the fact that not all of them are predicted to be secreted, leads to the conclusion that this region may play some other important role. This could be in complex formation or regulation of trehalose biosynthesis.

From the phylogenetic analysis the function of RiTPS1 (TPS), RiTPS2 (TPP) and RiTSL1 (regulatory) can be hypothesized. Also the fact that in RiTPS1 and RiTPS2 all 13 active site residues of the TPS domain could be found and in RiTPS2 all 25 active site residues for the TPP domain could be found, gives a directive to their function. However also in RiTSL1 all 25 active site residues for the TPP domain could be found, which might indicate a TPP function for this protein. Therefore the results of the phylogenetic analysis were confirmed by heterologous expression in yeast knockout mutants and complementation of the respective phenotypes.

The successful complementation of the growth defect on glucose of the yeast *tps1*Δ mutant when expressing RiTPS1 confirmed the expectations of this enzyme functioning as TPS (see Fig. 22a). Also the complementation of the heat intolerant phenotype in the yeast *tps2*Δ mutant with RiTPS2 confirmed the expected function as TPP (see Fig. 22b). The fact that RiTSL1 could neither complement the yeast *tps1*Δ nor the *tps2*Δ mutant leads to the conclusion that this protein in fact has no enzymatic function in trehalose biosynthesis and might play a regulatory role in *R. irregularis* such as ScTSL1 and ScTPS3 in yeast.

In yeast it was shown that the trehalose biosynthesis proteins (ScTPS1, ScTPS2, ScTSL1, ScTPS3) form a complex. Also the deletion of all of the proteins in single or double, triple or quadruple mutants leads to a lower TPS activity in the exponential phase[178]. The co-expression of *RiTPS1*, *RiTPS2* and *RiTsl1* in HIGS-RiTSL1 roots and in the time course experiment already give a hint that also in *R. irregularis* a complex is formed. To elucidate this possible complex-formation of the *R. irregularis* proteins, yeast-two-hybrid experiments can be performed.

3.4 Concluding remarks

In this work I combined the work on three different projects using functional genomics on different microbe-photobiont-interactions. First I was able to show that in order to create a complete gap-less genome, the use of more than one next generation sequencing technology improves the assembly largely. By going away from only using the short reads from 2ndGS (Illumina) and additionally using long reads from 3rdGS as bridges a genome without any unresolved regions and without dead ends was created. This approach offers the basis for further sequencing efforts also on larger genomes, for example of AM fungi.

Also the analysis of the *G. pyriformis* transcriptome provides valuable information. The results indicate that although *G. pyriformis* has this different life-style compared to other Glomeromycota there must have been a common obligate symbiotic ancestor. A set of three secreted proteins was identified, that show characteristics of amyloid forming proteins. Functional analysis of these proteins, as detection of localization, prion or amyloid fibril formation, or effect of heterologous expression in other symbiotic or pathogenic fungi could give insights into a special adaptation to the symbiosis with cyanobacteria.

In the last part of my work I used a set of previously identified effector candidate genes in *R. irregularis* for functional characterization. Even though it was not possible to reproduce the results obtained by using HIGS, the effect of downregulation of *RiTSL1* on the symbiosis were evident. This phenotype as a positive control can be helpful in order to establish this method and functionally characterize fungal effectors.

In further steps of characterization it became apparent that RiTSL1 most probably is not an effector, but plays a role in trehalose biosynthesis, which happens in a slight diurnal rhythm in *R. irregularis* colonizing *L. japonicus* roots. Transcriptome time-course experiments together with monitoring metabolites in symbiotic roots could give further insights into possible diurnal rhythms of nutrient exchange between the two symbiotic partners.

This work demonstrates step-by-step the results and the possibilities of functional genomics that arise by researching biological systems, organisms and molecular mechanisms on an “–omics” scale.

4 Material and Methods

4.1 Material

4.1.1 Chemicals

The chemicals used in this work were all of analytical grade and commercially available from AppliChem, Becton Dickinson, Fermentas, Gerbu, Merck, Roth and Sigma-Aldrich.

4.1.2 Primers

Primers were purchased from Sigma-Aldrich or Integrated DNA Technologies (HIGSP290-301). A list of primers used in this work can be found in Table S4.

4.1.3 Media, buffers and solutions

All media were autoclaved, sterile filtered or mixed from autoclaved/sterile filtered components. Recipes for media, buffers and solutions can be found in the respective sections of methods they were used in.

4.1.4 Plasmids

Plasmid vectors used in this study are listed in Table S5. Intermediate plasmids for sequencing in TOPO vectors (Invitrogen) or cloning in golden gate[179] level I, or II vectors are not mentioned.

4.2 Methods

4.2.1 Organisms and Cultivations

4.2.1.1 Bacteria

For cloning and multiplication of plasmids *Escherichia coli* Top10 was used. Its genotype is: *F-*; *mcrA*; Δ(*mrr-hsdRMS-mcrBC*); ϕ*80lacZ*Δ*M15*; Δ*lacX74*; *recA1*; *araD139*; Δ(*araleu*)*7697*; *galU*; *galK*; *rpsL*(*StrR*); *endA1*; *nupG*; λ-. For selection for transformed organisms they were plated on LB medium (10 g/l Tryptone, 5 g/l yeast extract, 10 g/l NaCl) complemented with respective antibiotics (ampicillin 100 µg/ml, kanamycin 50 µg/ml).

For transient hairy root transformation *Agrobacterium rhizogenes* 1193 was used. For selection of transformed organisms they were grown on LB medium complemented with three antibiotics (rifampicin 50 µg/ml, kanamycin 50 µg/ml, carbomycin 50 µg/ml).

4.2.1.2 *Rhizophagus irregularis*

The AM fungus *R. irregularis* DAOM 197198, Biosystematic Research Center, Ottawa, Canada[180] was cultivated in *in vitro* root organ culture with *A. rhizogenes* transformed *Cichorium intybus* or *Daucus carota* roots. To continue the culture, the colonized roots were cut into pieces and transferred to the root compartments of new MSR plates[181].

4.2.1.3 *Saccharomyces cerevisiae*

The following yeast strains were used in this work:

AH109 (Clontech): *MAT**a** trp1-901 leu2-3*, *112 ura3-52*, *his3-200*, *gal4Δ*, *gal80Δ*, *LYS2::GAL1$_{UAS}$-GAL1$_{TATA}$-HIS3*, *GAL2$_{UAS}$-GAL2$_{TATA}$-ADE2*, *URA3::MEL1$_{UAS}$- MEL1$_{TATA}$-lacZ*

BY4741 *tps3::kanMX*: *MAT**a** his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 tps3::KanMX*

W303 (Reed Wickner, pers. comm.): *MAT**a**/MATα leu2-3*, *112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15 [phi$^+$]*

1686 (Reed Wickner, pers. comm.): *MATα thr-*

18 (Reed Wickner, pers. comm.): *MAT**a** lys-*

YTK12 (Joe Win pers. comm.): *suc2Δ9*, *ura3-52*, *ade2-101*, *trp::hisG*[182]

Yeast was grown, if not otherwise stated, at 30°C on YPAD (10 g/l yeast extract, 20 g/l peptone, 20 g/l glucose, 0.08 g/l adenine sulfate) or selection media (SC medium (-LW/-LWA/-LWAH): 6.7 g/l yeast nitrogen base without amino acids, 20 g/l glucose, appropriate dropout powder (-Leu, -Trp/-Leu, -Trp, -Ade/-Leu, -Trp, -Ade, -His), pH 5.8-6.0).

4.2.1.4 *Lotus japonicus*

Lotus japonicus ecotype MG20 was used for experiments with colonized roots.

The plants were grown in sterile and semi-sterile conditions. For germination and transformation, seeds were scarified with sandpaper and sterilized in bleach solution (4% (v/v) hypochlorite, 0.1% (w/v) SDS) for 10 minutes at room temperature with agitation. Seeds

were washed five times with sterile $_{dd}H_2O$ and after being imbibed they were put on bactoagar (8 g/l bactoagar) plates. For germination, the seeds were covered to keep them in dark and incubated for three days (Panasonic MLR-352H-PE growth chamber, 24°C, 16h light, 8h dark). After germination the plates were kept without the cover in the growth chamber for 10 days. One week before transferring the plants to pots, seedlings were put on 1/2 Hoagland medium (0.025% (w/v) KNO_3, 0.0059% (w/v) $Ca(NO_3)_2$ x 2 H_2O, 0.0247% (w/v) $MgSO_4$ x 7 H_2O, 3.4 x 10^{-5}% (w/v) KH_2PO_4, 0.0186% (w/v) KCl, 2.5 x 10^{-4}% (w/v) Fe-citrate, 8.65 x 10^{-6}% (w/v) $MnSO_4$ x 4 H_2O, 1 x 10^{-6}% (w/v) $ZnSO_4$ x 7 H_2O, 4 x 10^{-7}% (w/v) $CuSO_4$ x 5 H_2O, 1 x 10^{-5} (w/v) $Na_2B_4O_7$ x 10 H_2O, 5 x 10^{-7}% (w/v) $(NH_4)_6Mo_7O_2$, 0.05% (w/v) MES, 0.8% (w/v) Bactoagar, adjust to pH 6.1 with 1 M KOH). Plantlets were transferred to pots with sand/vermiculite (2/1 vol.) substrate. The plants were grown in a phytotron chamber (Imtech, 22.2°C, 16 h light, 8 h dark) for five to six weeks. Twice a week they were watered with an autoclaved mix of H_2O:dH_2O (1:1), once a week with 1/2 B&D solution[183].

4.2.2 Cell and molecular biological methods

4.2.2.1 Standard molecular biology methods

4.2.2.1.1 Sanger sequencing

Sanger sequencing was performed at the Genomics Service Unit (LMU Biocenter, Munich).

4.2.2.1.2 Cloning

(a) Gateway cloning

For yeast complementation vectors, genes were cloned into 426-GPD-ccdB-HA according to Alberti *et al.*[184]. Therefore Gateway® LR Clonase® II Enzyme mix (Thermo Fisher Scientific) was used according to manufacturer`s instruction.

(b) Golden gate cloning

The target genes/regions were cloned with the golden gate based gene silencing kit[179] into a binary expression vector.

(c) TOPO cloning

PCR-amplified genes/regions were cloned using the Zero Blunt TOPO PCR Cloning Kit (Incitrogen) according to manufacturer's instruction into the pCR™Blunt II-TOPO vector

4.2.2.1.3 PCR

Different kinds of PCR were performed depending on template and/or product of the PCR. The standard PCR program is:

T_D/2 min; [T_D/30 s; T_m/30s; T_{ex}/t_{ex}]x25-35; T_{ex}/5 min

For PCR with primers containing overhangs, a two-step PCR program was used:

T_D/2 min; [T_D/30 s; T_{m1}/30s; T_{ex}/t_{ex}]x10; [T_D/30 s; T_{m2}/30s; T_{ex}/t_{ex}]x25-35; T_{ex}/5 min

For PCR with very complex template a touch-down PCR program was used:

T_D/2 min; [T_D/30 s; T_{m1}(-0.5°C/cycle)/30s; T_{ex}/t_{ex}]x10; [T_D/30 s; T_{m2}/30s; T_{ex}/t_{ex}]x25-35; T_{ex}/5 min

The denaturation temperature T_D was chosen depending on the DNA polymerase (Taq: 95°C; Polymerase: 98°C). T_m is the melting temperature of the primers (see Table S4). The extension temperature T_{ex} was chosen depending on the DNA polymerase (Taq: 68°C; Phusion: 72°C), the extension time t_{ex} depends on the length of the PCR product and was chosen depending on the DNA polymerase (Taq: 1 min/kb; Phusion: 0.25 min/kb).

(a) PCR with Taq DNA polymerase

Colony PCRs or PCRs to check the length of a product were performed with Taq DNA polymerase. Reactions performed with Taq DNA Polymerase contain 1x Standard Taq Buffer (New England Biolabs (NEB)) and 0.125 u/μl Taq DNA Polymerase (NEB). The reactions were performed with 0.5 μM forward and reverse primers (see Table S4) and 0.25 mM dNTPs (NEB).

(b) PCR with Phusion® HF DNA polymerase

PCRs for cloning or sequencing were performed with Phusion® HF DNA polymerase (NEB). Reactions performed with Phusion® HF Polymerase contain 1x Phusion HF Buffer (NEB) and 0.02 u/μl Phusion® HF Polymerase (NEB). The reactions were performed with 0.5 μM forward and reverse primers (see Table S4) and 0.2 mM dNTPs (NEB).

4.2.2.1.4 Agarose gel electrophoresis

The PCR products were mixed with 1 x Orange G loading dye (5 ml glycerin 99.5%, 5 ml TE buffer (pH8.0), 20 mg Orange G) and loaded on a 1% or 1.5% (w/v) agarose (in 1xTAE buffer (40mM Tris, 20mM acetic acid, 1mM EDTA)) gel for analysis. 2-log DNA Ladder (NEB) was used as a length standard. The gels were run for 30-40 minutes with 8 V/cm in 1x TAE buffer. Afterwards they were stained for 5 to 10 minutes in an ethidium bromide bath (0.5 g/l), rinsed in water and analyzed with UV light (302 nm).

4.2.2.1.5 DNA extraction from agarose gels

The respective band of DNA was cut with a razor blade from the gel. Subsequently the GeneJET gel extraction kit (Thermo Fisher Scientific) was used according to manufacturer's instructions.

4.2.2.1.6 Plasmid extraction

For plasmid extraction the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) was used according to manufacturer's instructions.

4.2.2.2 Ink staining of roots

Roots of *R. irregularis* inoculated plants were placed in 2 ml reaction tubes. The samples were covered with 1 ml 10% KOH and incubated for 15 min at 95°C. Afterwards the KOH solution was removed and the roots were rinsed three times with tap water and once with 10% acetic acid. Subsequently the roots were covered with ink solution (5% (v/v) black ink (Pelikan), 5% (v/v) acetic acid) and heated for 5 min at 95°C. The ink solution was removed and the roots rinsed with tap water three to five times. Finally the roots were destained using 5% acetic acid.

4.2.2.3 Microscopy

The successful transformation of hairy roots was validated by fluorescence microscopy using a stereo microscope (Leica M165 FC, magnification 7.3) with GFP filter (excitation light from the argon laser at 488 nm and detected at 500 - 550 nm).

The colonization level of ink stained roots was quantified using a light microscope (Leica): The roots of each sample were cut into 1 cm long fragments, of which ten were analysed. On each fragment ten loci (defined by a pin in the ocular at magnification 20 x) were analyzed for presence of external hyphae, internal hyphae, arbuscules and vesicles. The percentage of colonization incidents per 100 loci is given as root length colonization.

Pictures of arbuscules in ink-stained roots were taken with an inverted microscope (Leica DMI6000 B) in 10x and 20x magnification. Pictures were analyzed using FIJI[185] by free-hand-outlining the area of the arbuscule and the respective arbuscule-containing plant cell and taking the respective measurements.

4.2.2.4 Harvesting of *R. irregularis* spores

Spores of *R. irregularis* are cultured on MSR medium[181] plates. The agar was cut in small equal squares (approx. 1 cm^2 in size, around 50 spores). Each square was transferred to 300 µl citrate buffer (8.2% (v/v) 0.1 M sodium citrate, 1.8% (v/v) 0.1 M citric acid). The spores were placed on a shaker (1 h; 1000 rpm; room temperature) to dissolve the agar.

4.2.2.5 Mycorrhization

During potting the roots of *L. japonicus* were inoculated with *R. irregularis* by pipetting the citrate buffer containing the spores directly onto the roots while holding them over the holes in the potting substrate (vermiculite-sand (1:2) mix). Afterwards the plants were placed into those holes and stabilized by filling the holes with substrate.

4.2.2.6 Harvest of *G. pyriformis* cDNA expression library

A cDNA library of symbiotic *G. pyriformis* was generated by extracting polyA-RNA from a lysate of a non-sterile Geosiphon-culture in a symbiotic state with *Nostoc punctiforme*[132]. Inserts from the cDNA expression library of *G. pyriformis*, with 6 x 10^5 primary-insert-containing clones[81], were harvested by plasmid extraction.

The *E. coli* primary cDNA-library was grown on solid dYT medium (16 g/l tryptone, 10 g/l yeast extract, 5 g/l NaCl; 1.3% (w/v) agarose; 200 µg/ml ampicilline) over night at 37°C. Before plating, the bacteria were diluted 1:20,000 in liquid LB medium (10 g/l Tryptone, 5 g/l yeast extract, 10 g/l NaCl). On each plate (50 plates; 14.5 cm diameter) 200 µl of the bacteria suspension were distributed, which resulted in about 20,000 colonies per plate.

To harvest the bacteria from the plates, 2 ml LB-medium were added to each plate. The colonies were scraped from the plate with a spatula. This cell suspension was collected with a pipette and kept on ice. Another 2 ml LB-medium were added, to harvest remaining cells; this cell suspension was added to the next plate. These steps were repeated until all plates were harvested. The bacterial cells were harvested by centrifugation (12,000xg; 5 min).

4.2.2.7 Plasmid extraction of *G. pyriformis* cDNA library clones

Plasmid extraction was performed using the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) according to manufacturer's instructions. The extraction was performed with pellets from 1.5 ml cell suspension. The final elution volume was 600 *µ*l with a concentration of 291.6 ng/*µ*l measured with NanoDrop (Thermo Scientific).

4.2.2.8 Extraction of cDNA inserts from *G. pyriformis* cDNA library clones

100 *µ*g of plasmid DNA was digested with SfiI (62.5 U/ml, 50°C, over night). DNA was precipitated with 1/10 vol. 3 M NaAc and 2½ vol. 100% ethanol (-20°C, 1 h). The sample was centrifuged (4°C, 45000xg, 15 min). The supernatant was discarded and the pellet was washed with 70% (v/v) ethanol. Subsequently another centrifugation-step (20°C, 14000xg, 10 min) was performed. The supernatant was discarded and the pellet was air-dried. The dried pellet was dissolved in 100 µl 1x TE-buffer (10 mM Tris-HCl, pH8.0; 1 mM EDTA). Inserts were size-selected to obtain cDNA fragments within a size range from 600-6,000 bp on an agarose gel (0.8% (w/v)). The gel was run for 5 hours with 2.5-3.5 V/cm. The extraction of the DNA-fragments from the gel was performed using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific) according to manufacturer's instruction.

4.2.2.9 Illumina paired-end sequencing library preparation and sequencing of *G. pyriformis* cDNA library

Shearing of the cDNA-inserts was performed by using a M220 focused ultrasonicator (Covaris). 110 µl (2 µg) of cDNA insert plus 20 µl of 1xTE-buffer (10 mM Tris-HCl, pH8.0; 1

mM EDTA) were mixed in a snap-cap micro tube (Peak Incident Power (w): 50; Duty Factor: 20%; Cycles per Burst: 200; Treatment Time (s): 75; Temperature (°C): 20).

NEBNext DNA Library Prep Master Mix Set (NEB) was used according to manufacturer's instructions with NEBNext Index 12. PCR cycling conditions for PCR enrichment of adapter ligated DNA were as follows: [98°C/30'', (98°C/10'', 65°C/30'', 72°C/30'')x8, 72°C/5', 4°C/hold].

Two paired-end sequencing runs were performed on a MiSeq sequencer (Illumina) platform (run #1: 2x 150 bp, v2 chemistry; run #2: 2x 300 bp, v3 chemistry). All sequencing was performed at the Genomics Service Unit (LMU Biocenter, Munich).

4.2.2.10 Colony hybridization for extracting clones of GpRIC1, GpRIC2 and GpRIC3

For retrieving the correct sequence of these repeat-containing genes, colony hybridization was performed on clones from the original cDNA-expression library using the DIG-High Prime DNA Labeling and Detection Starter Kit I (Roche) according to manufacturer's instructions.

4.2.2.11 Transformation

4.2.2.11.1 *E. coli*

For plasmid multiplication, competent *E. coli* TOP10 cells were transformed with plasmid DNA using a standard heat shock method[186]. Occasionally, LB medium (10 g/l Tryptone, 5 g/l yeast extract, 10 g/l NaCl) was used instead of SOC medium. The cells were selected on according selective plates.

4.2.2.11.2 *A. rhizogenes*

Electro-competent *A. rhizogenes* 1193 cells were transformed with binary expression vectors using the Bio-RAD MicroPulser, delivering an electric pulse of 1.8 kV.

4.2.2.11.3 *S. cerevisiae*

The transformation of yeast was performed according to the protocol "High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method"[187]. The plates were incubated at 30°C for 4 days. Afterwards yeast colonies were streaked on fresh plates to obtain single colonies.

(a) with deletion cassettes

The yeast strain W303 was transformed with 3 µg to 4 µg PCR product of deletion cassettes. Transformants were selected on YPA Gal (10 g/l yeast extract, 20 g/l peptone, 20 g/l galactose, 0.04 g/l adenine sulfate) +200 mg/l geneticin plates. For the double mutant the resulting strain W303 tps3Δ::kanMX was used and transformed with the appropriate LEU2

cassette. Transformants were selected on HC -L plates[188]. This way the diploid yeast gene deletion strains W303 tps1Δ::kanMX, W303 tps2Δ::kanMX and W303 tps3Δ::kanMX tsl1Δ::LEU2 were generated.

(b) with complementation vectors

For the transformation mix 100 ng of the respective vector was used. The transformants were selected on HC -U Gal (tps1Δ::kanMX strain) and HC -U Glc (tps2Δ::kanMX strain) plates[188]. This way the following strains were generated:

Table 7: List of strains created by introducing complementation vectors.

Resulting strain	Original strain	Vector
W303 *tps1*Δ RiTSL1+SP	W303 *tps1*Δ::kanMX	HIGSV095
W303 *tps1*Δ RiTSL1-SP	W303 *tps1*Δ::kanMX	HIGSV096
W303 *tps1*Δ RiTPS1	W303 *tps1*Δ::kanMX	HIGSV097
W303 *tps1*Δ RiTPS2	W303 *tps1*Δ::kanMX	HIGSV098
W303 *tps1*Δ ScTPS1	W303 *tps1*Δ::kanMX	HIGSV099
W303 *tps1*Δ EV	W303 *tps1*Δ::kanMX	HIGSV101
W303 *tps2*Δ RiTSL1+SP	W303 *tps2*Δ::kanMX	HIGSV095
W303 *tps2*Δ RiTSL1-SP	W303 *tps2*Δ::kanMX	HIGSV096
W303 *tps2*Δ RiTPS1	W303 *tps2*Δ::kanMX	HIGSV097
W303 *tps2*Δ RiTPS2	W303 *tps2*Δ::kanMX	HIGSV098
W303 *tps2*Δ ScTPS2	W303 *tps2*Δ::kanMX	HIGSV100
W303 *tps2*Δ EV	W303 *tps2*Δ::kanMX	HIGSV101
W303 WT EV	W303	HIGSV101

4.2.2.11.4 Hairy root transformation of *L. japonicus*

A. rhizogenes transformed with binary expression vectors were cultured at 28°C for two days. Working in sterile conditions, the bacteria were scraped with a razor blade from the plates and each culture was resuspended in 700µl $_{dd}H_2O$. The suspension was transferred to a sterile filter paper. *L. japonicus* seedlings were placed with the hypocotyl in the bacteria solution on the filter paper. They were cut at the hypocotyl above the roots with a scalpel, dipped into the suspension and transferred to B5 medium (0.2% (w/v) sucrose, 0.8% (w/v) bactoagar, 0.033% (w/v) Gamborg`s B5 salt, 1 x B5 vitamin mix; adjust pH to 5.5 with 1M NaOH) plates. Covered in aluminum foil or a dark box, they were placed into a growth chamber (Panasonic MLR-352H-PE growth chamber, 24°C, 16h light, 8h dark). After two days the foil was removed and the plates placed back into the growth chamber. After three days the seedlings were transferred daily to B5 medium with cefotaxim (300 µg/ml) for four subsequent days. Transformed *L japonicus* MG20 plantlets were transferred to 1/2 Hoagland medium (see Section 4.2.1.4). After one week plantlets were transferred to pots.

4.2.2.12 Pipetting induced gene silencing (PIGS)

For PIGS 2 µg of a PCR product (RiTSL1: HIGSP252 & HIGSP253, GFP: HIGSP254 & HIGSP255) was used as a template for in vitro transcription with the T7 RNA Polymerase (NEB) according to manufacturer's instructions. 11 µg of the resulting dsRNA in 20µl of sterile $_{dd}H_2O$ was directly supplied to *R. irregularis* in root organ cultures with chicory roots (~2-3 weeks old) in a 24-well cell culture plate containing 500 µl MSR medium[181] (with adjustments: 0.15% Gelrite, 3 g/l sucrose, 3µM phosphate). Roots were harvested 2-3 weeks after application of dsRNA.

4.2.2.13 RNA extraction from roots

RNA was isolated from shock frozen root tissue using the Spectrum Plant Total RNA Kit (Sigma-Aldrich) according to manufacturer's instructions. The RNA was treated with DNase I amplification grade (Sigma-Aldrich) and tested for purity by PCR with primers HIGSP194&HIGSP195 and for quality using the 2100 Bioanalyzer (Agilent Genomics) with the Agilent RNA 6000 Nano kit and with the Plant RNA Nano program.

4.2.2.14 cDNA synthesis

Reverse transcription was performed with 300 ng RNA (concentration measured with 2100 Bioanalyzer (Agilent Genomics)) using the Superscript III kit (Invitrogen) and cDNA was subsequently treated with RNase H (NEB). Successful cDNA synthesis was tested by PCR.

4.2.2.15 Quantitative real-time PCR (qRT-PCR)

qRT-PCR was performed with Fast SYBR Green Master Mix (Thermo Fisher Scientific) on a QuantStudio 5 – 384-Well Block (Thermo Fisher Scientific) or on the CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad). Primers were designed using Primer3plus[189] with special settings "qPCR". Cycling conditions were: 95°C/2 min; (95°C/5 sec; 60°C/30 sec)x45 followed by dissociation curve analysis. The absolute expression values of *R. irregularis* genes were normalized to the geometric mean of three reference genes (KI300041.1/ERZ96911.1 – putative malate synthase; JEMT01026558.1/EXX58978.1 – Grx4; beta-tubulin), the arithmetic mean of two reference genes (beta-tubulin, elongation factor 1-alpha (EF1-alpha)), or only to one reference gene (beta-tubulin) according to the ΔΔCt method:

$$2^{(Cq_{reference\ gene} - Cq_{candidate\ gene})}$$

Each biological replicate is represented by two to three technical replicates.

4.2.2.16 Yeast growth assays

4.2.2.16.1 Sporulation and tetrad dissection

The diploid deletion mutant strains were sporulated and the tetrads dissected to generate haploid mutants of both mating types. A diploid mutant colony was streaked in a 16 cm^2 batch on a presporulation (10 g/l yeast extract, 100 g/l glucose, 20 g/l potassium acetate) plate and incubated for 4 days at 30°C. After incubation, from this plate cells were taken up with a flat toothpick, streaked on a sporulation (10 g/l potassium acetate, 1 g/l yeast extract, 0.5 g/l glucose/galactose, 0.8 g/l leucine, 0.2 g/l histidine, 0.8 g/l tryptophan, 0.35 g/l uracil, 0.7 g/l adenine) plate with glucose (for W303 *tps1*Δ::kanMX with galactose instead of glucose) and incubated for 4 days in the dark at RT. The sporulated cultures were examined with a light microscope (400x magnification) to confirm the formation of tetrads. For tetrad dissection the flat end of a sterile toothpick was used to transfer a dab of cells to 50 µl Zymolyase T100 (Zymo Research) (0.05 mg/ml in 1 M sorbitol) thawed on ice. The cells were resuspended and incubated at RT for 10 min. The digestion was stopped by placing the reaction tube on ice. A line of cells was drawn on the site of a flat and even YPA Gal (10 g/l yeast extract, 20 g/l peptone, 20 g/l galactose, 0.04 g/l adenine sulfate) plate using an inoculation loop. The plate was placed on the microscope stage of the Dissection Microscope MSM 400 (Singer Instruments). With a dissection needle (Singer Instruments) a cluster of four ascospores was picked and placed one below the other at a distance of ca. 1 cm using 600x magnification. This was repeated for 7 more tetrads, which were placed next to each other. The Plate was incubated for 4 days at 30°C.

4.2.2.16.2 Confirmation of deletion strains and determination of mating type

After tetrad dissection (see Section 4.2.2.16.1) the haploid cells were first analyzed for insertion of the KanMX or LEU2 Cassette and for the mating type using a replicator stamp. For deletion strains, plates with 8 tetrads in a row were replica plated on YPA Gal (10 g/l yeast extract, 20 g/l peptone, 20 g/l galactose, 0.04 g/l adenine sulfate) +200 mg/l geneticin plates. Plates with the double knockout strain were additionally replica plated on a HC-L plate[188]. Only haploid cells with knockouts should be able to grow on the corresponding selection plate(s). For determination of the mating type the yeast strains 1686 (Mat**α**) and 18 (Mat**a**) were used. With the flat end of sterile toothpicks a dab of both cell lines was taken up, each resuspended in 200 µl $_{dd}H_2O$ and plated on YPA Gal plates. Afterwards the plates with dissected tetrads were replica plated on both a plate with mating type **α** and on a plate with mating type **a** cells and incubated at 30°C over night. On the next day those plates were replica plated on SD minimal medium (6.7 g/l yeast nitrogen base without amino acids, 20 g/l glucose) plates which were then incubated at 30°C for another 2 days. On the minimal

medium plates, only those cells were able to grow which were in contact with cells of the opposite mating type. This way also haploid wild type cells could be detected. The identified cell lines were streaked onto new plates. The so found haploid knockout strains of each mating type and the diploid knockout strains were additionally confirmed by colony PCR.

4.2.2.16.3 Determination of mutant complementation growth rate

2 ml of an overnight preculture of transformed TPS1 Knockout strains were transferred into 2 ml reaction tubes and spun down at 12,000 x g. The supernatant was removed and the cells washed once with $_{dd}H_20$ (sterile). The centrifugation step was repeated and the cell pellet was resuspended in 2 ml HC-U Glc[188]. Afterwards the OD600 of both the cell suspensions and the overnight cultures of the transformed TPS2 knockout strains were measured and diluted with HC-U Glc medium to receive 100 µl suspension with an OD600 of 0.02 (2 x 105 cells/ml) in a Microtest Plate 96 Well,F (Sarstedt). Additionally 6 wells were filled only with medium to blank the values. The growth rates were analyzed using the Infinite 200 PRO multimode reader (Tecan) with the following parameters: OD620 was measured every 15 min for 72 hours. In the meantime the plate was shaken orbital with amplitude of 3.5 mm to prevent sedimentation of the cells. For TPS1 knockout cells and wild type cells as a control the temperature was set to 30°C; for TPS2 Knockout cells and wild type cells for comparison the temperature was set to 37°C. Afterwards the data were analyzed using Excel 2011 (Microsoft). The doubling time T_d was calculated using the following formula:

$$T_d = 4h \times \frac{\log(2)}{\log(\frac{q2}{q1})}$$

The optical densities of the wells with pure medium were averaged and subtracted from the OD620 of the wells with cell suspensions. The corrected OD620 values which were ≥ 0.15 were taken as q1. The OD620 values 4 h later were used as q2. The growth rate k was calculated using the following formula:

$$k = \frac{\log(2)}{T_d}$$

4.2.2.16.4 Yeast secretion assay

To test the functionality of the predicted secretion peptides (SP) of TSL1 from *R. irregularis* and *R. clarus* a yeast secretion assay[182] was used. In brief, the predicted SPs of RiTSL1, RcTSL1 and the original SP of the yeast invertase SUC2 were cloned into the yeast signal

sequence trap vector pSUC2T7M13ORI. The yeast SUC2-minus strain YTK12 was transformed[187] with the respective plasmids and the original vector without SP (pSUC2-[SP]). Transformed yeast was grown on HC dropout medium[188] with glucose as sole carbon source for control and sucrose as sole carbon source for testing. Growth rates were calculated as described above.

4.2.3 Bioinformatic methods

4.2.3.1 Hybrid assembly of *RI Norway* genome, genome annotation and comparisons

Hybrid genome assembly of *RI* Norway was performed with Unicycler v0.4.0[190] using default settings. Genome annotation was performed with RAST 2.0[191,192] and MicroScope[193]. BUSCO analysis of completeness was performed with BUSCO[115] version 3.0.2 and the lineage dataset "bacteria_odb9" (creation date: 2016-11-01; number of species: 3663; number of BUSCOs: 148) on the predicted protein sequences.

Clusters of orthologous groups (COGs) of proteins were predicted using the COGNiTOR software. Genome comparisons were performed with Easyfig 2.2.2[194] with concatenated sequences of whole chromosome and plasmids using blastn and default parameters. Comparison of plasmid contents was performed with BRIG[195], lower identity threshold = 70%, upper identity threshold = 90%.

Average nucleotide identity (ANI) comparison was performed with the genomes of members from the five proposed genospecies (gsA: WSM1325; gsB: 3841; gsC: TA1; gsD: CC278f & SM51; gsE: 128C53[116] (pers. comm. Peter Young). ANI scores were computed using the ANI calculator at http://enve-omics.ce.gatech.edu/ani/ with default parameters.

4.2.3.2 Basic Local Alignment Search Tool (BLAST) analyses

BLAST analyses were performed either using the online tools at https://blast.ncbi.nlm.nih.gov/Blast.cgi or using a locally installed version (blast 2.5.0, build Sep 9 2016 13:36:03) of the BLAST+ executables on a MacBook Pro (Retina, 15-inch, Late 2013, OS X El Capitan, Version 10.11.6)[28].

4.2.3.3 Trimming of *G. pyriformis* reads

Analysis of sequencing reads was performed using the commercial software CLC Genomics Workbench v9.5 (Qiagen). Reads were trimmed based on quality scores (limit: 0.05), end ambiguity (maximum allowed number of ambiguities: 2), adapter sequences and a short artifact sequence that was found during analysis (ATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGA).

4.2.3.4 Filtering *G. pyriformis* reads for known contaminants

Reads of known contaminants (vector sequence pDR196sfi of cDNA expression library including stuffer fragment, *E. coli* DH10B and DH5a, *Nostoc punctiforme* (GCF_000020025.1), bacteria like organisms[196]) were discarded by mapping all reads against these sequences (match score =1, mismatch cost=2, linear gap cost, insertion cost=3, deletion cost=3, length fraction=0.5, similarity fraction=0.8, global alignment=no, auto-detect paired distances=yes, non-specific match handling=map randomly, stand-alone read mapping) and proceeding with unmapped paired reads.

4.2.3.5 Assembly of *G. pyriformis* reads

The assembly of these reads from both sequencing runs was also performed using CLC software (Qiagen) with default settings.

4.2.3.6 Filtering of NRVTs from *G. pyriformis* assembly for unknown contaminants

In order to further remove contaminant sequences, the assembled NRVTs were compared to the NCBI nt DNA database (Nov 30th, 2016) using blastn (NCBI-BLAST 2.2.29+, e-value < 10). All contigs with taxon IDs (best hit) from archaea, bacteria, plant, or virus were discarded. Transcriptomes of *R. irregularis*[114] and *G. rosea*[60] have a mean GC content of around 32%. Therefore, NRVTs with GC content higher than 55%, corresponding to the minimum between two apparent peaks at 40% and 66%, were also discarded. All remaining contigs longer than 7000 bp were split into smaller parts and compared again to the NCBI nt database using blastn to remove bacterial sequences that were not recognized as best hit before. All NRVTs smaller than 200bp were discarded. During subsequent analyses 3,587, 171, and 141 NRVTs were identified to be distinct assemblies of the repetitive GpRIC1, GpRIC2, and GpRIC3 transcripts, respectively, and each were pooled into one NRVT.

4.2.3.7 Blobplot analysis of *G. pyriformis* assembly and final transcriptome

Blobplots were created using blobtools[117] with default parameters. Blobplots were created for all initial NRVTs, for all NRVTs removed due to filtering and for all NRVTs in the final assembly.

4.2.3.8 BUSCO assessment of completeness of the *G. pyriformis* transcriptome

BUSCO analysis was performed with BUSCO v1.22[115]. The dataset used for assessment was fungi_odb9 (21.10.2016) with default parameters for transcriptome analysis.

4.2.3.9 Comparison of *G. pyriformis* transcriptome to other fungal proteomes

NRVTs were compared to proteomes of other fungi (see Table ES4) using blastx (e-value < 1e-5).

4.2.3.10 Reciprocal BLAST for the identification of different sets of genes in the *G. pyriformis* transcriptome

Reciprocal BLAST (e-value < $1e^{-5}$) was performed with NRVTs and a set of 86 meiosis-related genes (Malik et al., 2007) from Saccharomyces cerevisiae and, if present, from Sordaria macrospora, Neurospora crassa, Rhizophagus spp. Only one best hit from tblastn of meiosis genes against NRVTs was checked against 20 best hits from blastx of NRVTs against meiosis genes. Pairs that were found in both directions were considered as true homologs. For the identification of HMG-domain containing genes reciprocal BLAST (e-value < 1) was performed with a set of 25 fungal HMG-domain containing proteins (see Table 8). The best 20 hits of each BLAST direction were compared to each other to find HMG-domain containing proteins.

Table 8: List of HMG domain containing proteins used for reciprocal BLAST analysis.

Accession	Species
AAF00498.1	*Davidsoniella eucalypti*
AAG42810.1	*Fusarium graminearum*
AAG42812.1	*Fusarium graminearum*
AAK15315.1	*Candida albicans*
AAK83343.1	*Cryphonectria parasitica*
AAK83344.1	*Cryphonectria parasitica*
AAL30836.1	*Zymoseptoria tritici*
AAP13349.1	*Pneumocystis carinii*
AET35419.1	*Syzygites megalocarpus*
AET35422.1	*Syzygites megalocarpus*
BAA33018.1	*Coprinopsis cinerea*
BAC66503.1	*Cordyceps tenuipes*
CAA06843.1	*Pyrenopeziza brassicae*
CAA06846.1	*Pyrenopeziza brassicae*
CAD21099.1	*Neurospora crassa*
C7U331.1	*Schizosaccharomyces pombe*
P25042.1	*Saccharomyces cerevisiae*
P35693.2	*Podospora anserina*
P25042.1	*Schizosaccharomyces pombe*
P36981.2	*Neurospora crassa*
Q02991.1	*Cochliobolus heterostrophus*
Q08143.1	*Podospora anserina*
Q10116.1	*Neurospora crassa*
Q99101	*Ustilago maydis*

Missing Glomeromycota core gene (MGCGs)[60] sequences were retrieved from theyeastgenome.org. Reciprocal BLAST (e-value < 1e-5) was performed to compare NRVTs to this set of MGCGs. The best 20 hits of each BLAST direction were compared to each

other to find homologues. For the prediction of transporters, reciprocal BLAST (e-value < $1e^{-5}$) was performed with NRVTs and the transporter classification database (TCDB)[35]. 20 best hits of blastx (NRVTs against TCDB) were compared to 200 best hits of tblastn (TCDB against NRVTs). With these parameters, already published Glomeromycota transporters could be found.

4.2.3.11 Prediction of amino acid sequences of the *G. pyriformis* transcriptome

Amino acid sequences were predicted using Blast2GO v4.1.9[33] with default parameters.

4.2.3.12 Prediction and analysis of euKaryotic Orthologous Groups (KOG) for predicted amino acid sequences of the *G. pyriformis* transcriptome

Functional annotation of KOG classes was performed using the WebMGA[138] server with default parameters. The predicted amino acid sequences of the *G. pyriformis* transcriptome were used as input sequences. Principal component analysis (PCA) was performed in R[197].

4.2.3.13 Prediction of CAZyme classes of *G. pyriformis* transcriptome

The predicted amino acid sequences of the *G. pyriformis* transcriptome were used to predict CAZyme classes on the dbCAN server[36] using default parameters.

4.2.3.14 Detection and analysis of SNPs in the *G. pyriformis* transcriptome

For SNP detection trimmed reads were mapped to NRVTs and the basic variant detection tool was used with following parameters (Ploidy =2, Ignore positions with coverage above = 100,000, Restrict calling to target regions = Not set, Ignore broken pairs = Yes, Ignore non-specific matches = Reads, Minimum coverage = 10, Minimum count = 2, Minimum frequency (%) = 2.0, Base quality filter = No, Read direction filter = No, Relative read direction filter = Yes, Significance (%) = 1.0, Read position filter = No, Remove pyro-error variants = No; CLC Genomics Workbench 9.5.3, Qiagen).

4.2.3.15 Prediction of secreted proteins in the *G. pyriformis* transcriptome

The predicted amino acid sequences were used for the prediction of secreted proteins. This was performed as described in[60]. Proteins were not filtered for length or starting amino acid.

4.2.3.16 Detection of redundant sequences in the *G. pyriformis* transcriptome

In order to detect redundant transcripts, all nucleotide sequences of the *G. pyriformis* assembly after filtering were blasted against each other (e-value < $1e^{-5}$). Edges were created by leaving only non-self BLAST-hit pairs in the list. Using the igraph package[198] in RStudio (Version 1.1.442) an igraph data frame was created (graph.data.frame(edges,directed=TRUE)) and entries with less than 11 connections were deleted (delete.vertices(igraph_df, V(igraph_df)[degree(igraph_df)<11])). The resulting .gml-

file was analyzed with Cytoscape (Version 3.6.0). Sequences that had common BLAST hits were clustered and assembled using CLC Genomics Workbench (Minimum aligned read length = 50, Alignment stringency = Medium, Conflicts = Vote (A, C, G, T)). The consensuses of the assembled sequences were analyzed regarding the previously obviously redundant sequences of GpRIC1. Meaning the presence of repeats, a high amino acid content of asparagine and glycine and the presence of Kex2 recognition sites. The *in silico* detected hypothetical sequences were subsequently found *in vitro* by colony hybridization.

4.2.3.17 Statistical analyses of qPCR data

Statistical analyses of qPCR data were performed using RStudio (Version 1.0.136). An "Analysis of Variance Model" (ANOVA) was fit to the data calling the "aov()"-function, subsequently "Tukey Honest Significant Differences" were computed calling the "TukeyHSD()"-function. With the "multicompView" package[199] a letter summary of similarities and differences was computed based on ANOVA and Tukey output.

For comparison of only two values a two-sided t-test was performed.

4.2.3.18 Phylogenetic analysis of trehalose biosynthesis proteins

Homologues of TPS1, TPS2 and TSL1 in *R. irregularis* and other fungi, oomycetes and one chlorophyte (11 Ascomycota, 27 Basidiomycota, three oomycetes) were found by blasting ScTPS1, ScTPS2, ScTPS3 or ScTSL1 against NCBI database or other fungal genome databases[200]. Homologues of *Rhizophagus clarus* and *Geosiphon pyriformis* were found by blasting the proteins from *R. irregularis* against the unpublished genome (*R. clarus*), or transcriptome (*G. pyriformis*) sequenced at the Genomics Service Unit (LMU Biocenter, Munich). TPS and TPP domains of the proteins were identified by blasting the proteins against the "non-redundant protein sequences" (nr) database (NCBI) and by checking for conserved domains. For the phylogenetic comparison the TPS domains of 52 TPS1, 59 TPS2 and 45 TSL1 proteins were aligned using CLC Main Workbench (Qiagen). A maximum likelihood phylogeny was produced using CLC Main Workbench (Qiagen) with the construction method neighbor joining and 1000 replicates. For the comparison of the N-terminus of TSL1 proteins, a selection of the N-terminal sequences up to the start of the TPS domain was aligned. The maximum length of a predicted secretion peptide was taken and the length of the sequences was adapted so that the alignment will end three amino acids after that point. Secretion peptide prediction was performed using SignalP4.1[201] with settings for eukaryotes.

4.2.3.19 Identification of RNAi target regions in *R. irregularis* effector candidate genes

200 bp regions with a GC-content of about 40% were selected and off-target prediction was carried out with si-Fi3 (http://labtools.ipk-gatersleben.de/) or Next-RNAi[202].

5 References

1 Ferguson, B. A., Dreisbach, T. A., Parks, C. G., Filip, G. M. & Schmitt, C. L. Coarse-scale population structure of pathogenic *Armillaria* species in a mixed-conifer forest in the Blue Mountains of northeast Oregon. *Canadian Journal of Forest Research* **33**, 612-623, (2003).

2 Hall, N. Advanced sequencing technologies and their wider impact in microbiology. *Journal of Experimental Biology* **210**, 1518, (2007).

3 Medini, D., Donati, C., Tettelin, H., Masignani, V. & Rappuoli, R. The microbial pan-genome. *Curr Opin Genet Dev* **15**, 589-594, (2005).

4 Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., Angiuoli, S. V., Crabtree, J., Jones, A. L., Durkin, A. S., Deboy, R. T., Davidsen, T. M., Mora, M., Scarselli, M., Margarit y Ros, I., Peterson, J. D., Hauser, C. R., Sundaram, J. P., Nelson, W. C., Madupu, R., Brinkac, L. M., Dodson, R. J., Rosovitz, M. J., Sullivan, S. A., Daugherty, S. C., Haft, D. H., Selengut, J., Gwinn, M. L., Zhou, L., Zafar, N., Khouri, H., Radune, D., Dimitrov, G., Watkins, K., O'Connor, K. J., Smith, S., Utterback, T. R., White, O., Rubens, C. E., Grandi, G., Madoff, L. C., Kasper, D. L., Telford, J. L., Wessels, M. R., Rappuoli, R. & Fraser, C. M. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A* **102**, 13950-13955, (2005).

5 Delaux, P. M., Varala, K., Edger, P. P., Coruzzi, G. M., Pires, J. C. & Ane, J. M. Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. *PLoS Genet* **10**, e1004487, (2014).

6 Spatafora, J. W., Chang, Y., Benny, G. L., Lazarus, K., Smith, M. E., Berbee, M. L., Bonito, G., Corradi, N., Grigoriev, I., Gryganskyi, A., James, T. Y., O'Donnell, K., Roberson, R. W., Taylor, T. N., Uehling, J., Vilgalys, R., White, M. M. & Stajich, J. E. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* **108**, 1028-1046, (2016).

7 Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C. A., Hutchison, C. A., Slocombe, P. M. & Smith, M. Nucleotide sequence of bacteriophage phiX174 DNA. *Nature* **265**, 687-695, (1977).

8 Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M. & et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496-512, (1995).

9 Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H. & Oliver, S. G. Life with 6000 genes. *Science* **274**, 546, 563-547, (1996).

10 Consortium, C. e. S. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012-2018, (1998).

11 Arabidopsis Genome Initiative, T. & Copenhaver, G. *Analysis of the genome sequence of the flowering plant Arabidopsis thaliana* Vol. 408 (2000).

12 National Human Genome Research Institute. *The Cost of Sequencing a Human Genome*, <https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/> (2016).

13 Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A. & Waterston, R. H. DNA sequencing at 40: past, present and future. *Nature* **550**, 345-353, (2017).

14 Wetterstrand, K. A. *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)*, <http://www.genome.gov/sequencingcostsdata> (

15 Illumina. *Illumina sequencing platforms*, <https://emea.illumina.com/systems/sequencing-platforms.html> (2018).

16 Oxford Nanopore Technologies. *How does nanopore DNA/RNA sequencing work?*, <https://nanoporetech.com/how-it-works> (2018).
17 Wang, Y., Yang, Q. & Wang, Z. The evolution of nanopore sequencing. *Front Genet* **5**, 449, (2014).
18 Oxford Nanopore Technologies. *World first: continuous DNA sequence of more than a million bases achieved with nanopore sequencing*, <https://nanoporetech.com/about-us/news/world-first-continuous-dna-sequence-more-million-bases-achieved-nanopore-sequencing> (2017).
19 Jain, M., Tyson, J. R., Loose, M., Ip, C. L. C., Eccles, D. A., O'Grady, J., Malla, S., Leggett, R. M., Wallerman, O., Jansen, H. J., Zalunin, V., Birney, E., Brown, B. L., Snutch, T. P., Olsen, H. E., Min, I. O. N. A. & Reference, C. MinION Analysis and Reference Consortium: Phase 2 data release and analysis of R9.0 chemistry. *F1000Res* **6**, 760, (2017).
20 Wescoe, Z. L., Schreiber, J. & Akeson, M. Nanopores discriminate among five C5-cytosine variants in DNA. *J Am Chem Soc* **136**, 16582-16587, (2014).
21 Laszlo, A. H., Derrington, I. M., Brinkerhoff, H., Langford, K. W., Nova, I. C., Samson, J. M., Bartlett, J. J., Pavlenok, M. & Gundlach, J. H. Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc Natl Acad Sci U S A* **110**, 18904-18909, (2013).
22 Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. & Law, M. Comparison of next-generation sequencing systems. *J Biomed Biotechnol* **2012**, 251364, (2012).
23 Cap today. *Illumina, NextSeq 550 Sequencing System,Next-generation sequencing instruments, December 2017*, <http://www.captodayonline.com/productguides/instruments/next-generation-sequencing-instrument-december-2017/illumina-550-sequencing-system-sequencing-december-2017.html> (2017).
24 Illumina. *NextSeq® 550 Sequencing System*, <https://emea.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/nextseq-550-system-spec-sheet-770-2013-053.pdf> (2017).
25 Oxford Nanopore Technologies. *Human Genome on a MinION*, <https://nanoporetech.com/about-us/news/human-genome-minion> (2016).
26 Glenn, T. C. Field guide to next-generation DNA sequencers. *Mol Ecol Resour* **11**, 759-769, (2011).
27 Heather, J. M. & Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1-8, (2016).
28 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410, (1990).
29 NCBI. *GenBank and WGS Statistics*, <https://www.ncbi.nlm.nih.gov/genbank/statistics/> (2018).
30 National Institute of Health. *GenBank celebrates 25 years of service with two-day conference; leading scientists will discuss the DNA database at April 7-8 meeting*, <https://www.nih.gov/news-events/news-releases/genbank-celebrates-25-years-service-two-day-conference-leading-scientists-will-discuss-dna-database-april-7-8-meeting> (2008).
31 CLC bio, a. Q. C. Manual for CLC Genomics Workbench 7.5. Windows, Mac OS X and Linux. (2015).
32 Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* **13**, e1005595, (2017).
33 Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M. & Robles, M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674-3676, (2005).

34 Vallenet, D., Calteau, A., Cruveiller, S., Gachet, M., Lajus, A., Josso, A., Mercier, J., Renaux, A., Rollin, J., Rouy, Z., Roche, D., Scarpelli, C. & Medigue, C. MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes. *Nucleic Acids Res* **45**, D517-D528, (2017).

35 Saier, M. H., Reddy, V. S., Tsu, B. V., Ahmed, M. S., Li, C. & Moreno-Hagelsieb, G. The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Research* **44**, D372-D379, (2016).

36 Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F. & Xu, Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **40**, W445-451, (2012).

37 Oldroyd, G. E., Murray, J. D., Poole, P. S. & Downie, J. A. The rules of engagement in the legume-rhizobial symbiosis. *Annu Rev Genet* **45**, 119-144, (2011).

38 Poole, P., Ramachandran, V. & Terpolilli, J. Rhizobia: from saprophytes to endosymbionts. *Nat Rev Microbiol* **16**, 291-303, (2018).

39 Gossmann, J. A., Markmann, K., Brachmann, A., Rose, L. E. & Parniske, M. Polymorphic infection and organogenesis patterns induced by a *Rhizobium leguminosarum* isolate from *Lotus* root nodules are determined by the host genotype. *New Phytol* **196**, 561-573, (2012).

40 Palmer, K. M. & Young, J. P. Higher diversity of *Rhizobium leguminosarum* biovar viciae populations in arable soils than in grass soils. *Appl Environ Microbiol* **66**, 2445-2450, (2000).

41 Palmer, K. M., Turner, S. L. & Young, J. P. Sequence diversity of the plasmid replication gene *repC* in the *Rhizobiaceae*. *Plasmid* **44**, 209-219, (2000).

42 Smith, S. E. & Read, D. in *Mycorrhizal Symbiosis (Third Edition)* 13-41 (Academic Press, 2008).

43 Brundrett, M. *Mycorrhizal associations and other means of nutrition of vascular plants: Understanding the global diversity of host plants by resolving conflicting information and developing reliable means of diagnosis*. Vol. 320 (2009).

44 AMF phylogeny. *Glomeromycota Species List*, <http://www.amf-phylogeny.com/amphylo_species.html> (2018).

45 Selosse, M. A., Strullu-Derrien, C., Martin, F. M., Kamoun, S. & Kenrick, P. Plants, fungi and oomycetes: a 400-million year affair that shapes the biosphere. *New Phytol* **206**, 501-506, (2015).

46 Redecker, D., Kodner, R. & Graham, L. E. Glomalean fungi from the Ordovician. *Science* **289**, 1920-1921, (2000).

47 Strullu-Derrien, C., Kenrick, P., Pressel, S., Duckett, J. G., Rioult, J. P. & Strullu, D. G. Fungal associations in *Horneophyton ligneri* from the Rhynie Chert (c. 407 million year old) closely resemble those in extant lower land plants: novel insights into ancestral plant-fungus symbioses. *New Phytol* **203**, 964-979, (2014).

48 Desiro, A., Duckett, J. G., Pressel, S., Villarreal, J. C. & Bidartondo, M. I. Fungal symbioses in hornworts: a chequered history. *Proc Biol Sci* **280**, 20130207, (2013).

49 Wang, B., Yeun, L. H., Xue, J. Y., Liu, Y., Ane, J. M. & Qiu, Y. L. Presence of three mycorrhizal genes in the common ancestor of land plants suggests a key role of mycorrhizas in the colonization of land by plants. *New Phytol* **186**, 514-525, (2010).

50 Parniske, M. Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nat Rev Microbiol* **6**, 763-775, (2008).

51 Smith, S. E. & Read, D. in *Mycorrhizal Symbiosis (Third Edition)* 611-XVIII (Academic Press, 2008).

52 Jung, S. C., Martinez-Medina, A., Lopez-Raez, J. A. & Pozo, M. J. Mycorrhiza-induced resistance and priming of plant defenses. *J Chem Ecol* **38**, 651-664, (2012).

53 Gianinazzi, S., Gollotte, A., Binet, M. N., van Tuinen, D., Redecker, D. & Wipf, D. Agroecology: the key role of arbuscular mycorrhizas in ecosystem services. *Mycorrhiza* **20**, 519-530, (2010).

54 Smith, S. E. & Read, D. in *Mycorrhizal Symbiosis (Third Edition)* 117-144 (Academic Press, 2008).
55 Gutjahr, C. & Parniske, M. Cell and developmental biology of arbuscular mycorrhiza symbiosis. *Annu Rev Cell Dev Biol* **29**, 593-617, (2013).
56 Kosuta, S., Chabaud, M., Lougnon, G., Gough, C., Denarie, J., Barker, D. G. & Becard, G. A diffusible factor from arbuscular mycorrhizal fungi induces symbiosis-specific MtENOD11 expression in roots of Medicago truncatula. *Plant Physiol* **131**, 952-962, (2003).
57 de Wit, P. J. How plants recognize pathogens and defend themselves. *Cell Mol Life Sci* **64**, 2726-2732, (2007).
58 Toruno, T. Y., Stergiopoulos, I. & Coaker, G. Plant-Pathogen Effectors: Cellular Probes Interfering with Plant Defenses in Spatial and Temporal Manners. *Annu Rev Phytopathol* **54**, 419-441, (2016).
59 Tisserant, E., Malbreil, M., Kuo, A., Kohler, A., Symeonidi, A., Balestrini, R., Charron, P., Duensing, N., Frei dit Frey, N., Gianinazzi-Pearson, V., Gilbert, L. B., Handa, Y., Herr, J. R., Hijri, M., Koul, R., Kawaguchi, M., Krajinski, F., Lammers, P. J., Masclaux, F. G., Murat, C., Morin, E., Ndikumana, S., Pagni, M., Petitpierre, D., Requena, N., Rosikiewicz, P., Riley, R., Saito, K., San Clemente, H., Shapiro, H., van Tuinen, D., Becard, G., Bonfante, P., Paszkowski, U., Shachar-Hill, Y. Y., Tuskan, G. A., Young, J. P., Sanders, I. R., Henrissat, B., Rensing, S. A., Grigoriev, I. V., Corradi, N., Roux, C. & Martin, F. Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proc Natl Acad Sci U S A* **110**, 20117-20122, (2013).
60 Tang, N., San Clemente, H., Roy, S., Becard, G., Zhao, B. & Roux, C. A Survey of the Gene Repertoire of *Gigaspora rosea* Unravels Conserved Features among Glomeromycota for Obligate Biotrophy. *Front Microbiol* **7**, 233, (2016).
61 Sedzielewska Toro, K. & Brachmann, A. The effector candidate repertoire of the arbuscular mycorrhizal fungus *Rhizophagus clarus*. *BMC Genomics* **17**, 101, (2016).
62 Lin, K., Limpens, E., Zhang, Z., Ivanov, S., Saunders, D. G., Mu, D., Pang, E., Cao, H., Cha, H., Lin, T., Zhou, Q., Shang, Y., Li, Y., Sharma, T., van Velzen, R., de Ruijter, N., Aanen, D. K., Win, J., Kamoun, S., Bisseling, T., Geurts, R. & Huang, S. Single nucleus genome sequencing reveals high similarity among nuclei of an endomycorrhizal fungus. *PLoS Genet* **10**, e1004078, (2014).
63 Kloppholz, S., Kuhn, H. & Requena, N. A secreted fungal effector of *Glomus intraradices* promotes symbiotic biotrophy. *Curr Biol* **21**, 1204-1209, (2011).
64 Tsuzuki, S., Handa, Y., Takeda, N. & Kawaguchi, M. Strigolactone-Induced Putative Secreted Protein 1 Is Required for the Establishment of Symbiosis by the Arbuscular Mycorrhizal Fungus *Rhizophagus irregularis*. *Mol Plant Microbe Interact* **29**, 277-286, (2016).
65 Genre, A., Chabaud, M., Timmers, T., Bonfante, P. & Barker, D. G. Arbuscular mycorrhizal fungi elicit a novel intracellular apparatus in *Medicago truncatula* root epidermal cells before infection. *Plant Cell* **17**, 3489-3499, (2005).
66 Gus-Mayer, S., Naton, B., Hahlbrock, K. & Schmelzer, E. Local mechanical stimulation induces components of the pathogen defense response in parsley. *Proc Natl Acad Sci U S A* **95**, 8398-8403, (1998).
67 Demchenko, K., Winzer, T., Stougaard, J., Parniske, M. & Pawlowski, K. Distinct roles of *Lotus japonicus* SYMRK and SYM15 in root colonization and arbuscule formation. *New Phytologist* **163**, 381-392, (2004).
68 Smith, S. E. & Read, D. in *Mycorrhizal Symbiosis (Third Edition)* 42-90 (Academic Press, 2008).
69 Balestrini, R. & Bonfante, P. The interface compartment in arbuscular mycorrhizae: A special type of plant cell wall? *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology* **139**, 8-15, (2005).
70 Balestrini, R. & Bonfante, P. Cell wall remodeling in mycorrhizal symbiosis: a way towards biotrophism. *Front Plant Sci* **5**, 237, (2014).

71 Bonfante-Fasolo, P., Faccio, A., Perotto, S. & Schubert, A. Correlation between chitin distribution and cell wall morphology in the mycorrhizal fungus *Glomus versiforme*. *Mycological Research* **94**, 157-165, (1990).

72 Calabrese, S., Perez-Tienda, J., Ellerbeck, M., Arnould, C., Chatagnier, O., Boller, T., Schussler, A., Brachmann, A., Wipf, D., Ferrol, N. & Courty, P. E. GintAMT3 - a Low-Affinity Ammonium Transporter of the Arbuscular Mycorrhizal *Rhizophagus irregularis*. *Front Plant Sci* **7**, 679, (2016).

73 Harrison, M. J., Dewbre, G. R. & Liu, J. A phosphate transporter from *Medicago truncatula* involved in the acquisition of phosphate released by arbuscular mycorrhizal fungi. *Plant Cell* **14**, 2413-2429, (2002).

74 Harrison, M. J. & van Buuren, M. L. A phosphate transporter from the mycorrhizal fungus *Glomus versiforme*. *Nature* **378**, 626-629, (1995).

75 Helber, N., Wippel, K., Sauer, N., Schaarschmidt, S., Hause, B. & Requena, N. A versatile monosaccharide transporter that operates in the arbuscular mycorrhizal fungus *Glomus sp* is crucial for the symbiotic relationship with plants. *Plant Cell* **23**, 3812-3823, (2011).

76 D. Douds, D., E. Pfeffer, P. & Shachar-Hill, Y. *Carbon Partitioning, Cost, and Metabolism of Arbuscular Mycorrhizas*. (2000).

77 Graham, J. H. *Assessing costs of arbuscular mycorrhizal symbiosis in agroecosystems*. (2000).

78 Schaarschmidt, S., Roitsch, T. & Hause, B. Arbuscular mycorrhiza induces gene expression of the apoplastic invertase *LIN6* in tomato (*Lycopersicon esculentum*) roots. *J Exp Bot* **57**, 4015-4023, (2006).

79 Manck-Gotzenberger, J. & Requena, N. Arbuscular mycorrhiza Symbiosis Induces a Major Transcriptional Reprogramming of the Potato SWEET Sugar Transporter Family. *Front Plant Sci* **7**, 487, (2016).

80 Wright, D. P., Read, D. J. & Scholes, J. D. Mycorrhizal sink strength influences whole plant carbon balance of *Trifolium repens* L. *Plant, Cell and Environment* **21**, 881-891, (1998).

81 Schussler, A., Martin, H., Cohen, D., Fitz, M. & Wipf, D. Characterization of a carbohydrate transporter from symbiotic glomeromycotan fungi. *Nature* **444**, 933-936, (2006).

82 Shachar-Hill, Y., Pfeffer, P. E., Douds, D., Osman, S. F., Doner, L. W. & Ratcliffe, R. G. Partitioning of intermediary carbon metabolism in vesicular-arbuscular mycorrhizal leek. *Plant Physiol* **108**, 7-15, (1995).

83 Pfeffer, P. E., Douds Jr, D. D., Becard, G. & Shachar-Hill, Y. Carbon uptake and the metabolism and transport of lipids in an arbuscular mycorrhiza. *Plant Physiol* **120**, 587-598, (1999).

84 Bago, B., Zipfel, W., Williams, R. M., Jun, J., Arreola, R., Lammers, P. J., Pfeffer, P. E. & Shachar-Hill, Y. Translocation and utilization of fungal storage lipid in the arbuscular mycorrhizal symbiosis. *Plant Physiol* **128**, 108-124, (2002).

85 Bago, B., Pfeffer, P. E., Abubaker, J., Jun, J., Allen, J. W., Brouillette, J., Douds, D. D., Lammers, P. J. & Shachar-Hill, Y. Carbon export from arbuscular mycorrhizal roots involves the translocation of carbohydrate as well as lipid. *Plant Physiol* **131**, 1496-1507, (2003).

86 Bécard, G., Doner, L. W., Rolin, D. B., Douds, D. D. & Pfeffer, P. E. Identification and quantification of trehalose in vesicular-arbuscular mycorrhizal fungi by in vivo13C NMR and HPLC analyses*. *New Phytologist* **118**, 547-552, (1991).

87 Keymer, A., Pimprikar, P., Wewer, V., Huber, C., Brands, M., Bucerius, S. L., Delaux, P. M., Klingl, V., Ropenack-Lahaye, E. V., Wang, T. L., Eisenreich, W., Dormann, P., Parniske, M. & Gutjahr, C. Lipid transfer from plants to arbuscular mycorrhiza fungi. *Elife* **6**, (2017).

88 Wewer, V., Brands, M. & Dormann, P. Fatty acid synthesis and lipid metabolism in the obligate biotrophic fungus *Rhizophagus irregularis* during mycorrhization of *Lotus japonicus*. *Plant J* **79**, 398-412, (2014).
89 Salvioli, A., Ghignone, S., Novero, M., Navazio, L., Venice, F., Bagnaresi, P. & Bonfante, P. Symbiosis with an endobacterium increases the fitness of a mycorrhizal fungus, raising its bioenergetic potential. *ISME J* **10**, 130-144, (2016).
90 Gutjahr, C., Radovanovic, D., Geoffroy, J., Zhang, Q., Siegler, H., Chiapello, M., Casieri, L., An, K., An, G., Guiderdoni, E., Kumar, C. S., Sundaresan, V., Harrison, M. J. & Paszkowski, U. The half-size ABC transporters STR1 and STR2 are indispensable for mycorrhizal arbuscule formation in rice. *Plant J* **69**, 906-920, (2012).
91 Yang, S. Y., Gronlund, M., Jakobsen, I., Grotemeyer, M. S., Rentsch, D., Miyao, A., Hirochika, H., Kumar, C. S., Sundaresan, V., Salamin, N., Catausan, S., Mattes, N., Heuer, S. & Paszkowski, U. Nonredundant regulation of rice arbuscular mycorrhizal symbiosis by two members of the phosphate transporter1 gene family. *Plant Cell* **24**, 4236-4251, (2012).
92 Maldonado-Mendoza, I. E., Dewbre, G. R. & Harrison, M. J. A phosphate transporter gene from the extra-radical mycelium of an arbuscular mycorrhizal fungus *Glomus intraradices* is regulated in response to phosphate in the environment. *Mol Plant Microbe Interact* **14**, 1140-1148, (2001).
93 Balestrini, R., Gomez-Ariza, J., Lanfranco, L. & Bonfante, P. Laser microdissection reveals that transcripts for five plant and one fungal phosphate transporter genes are contemporaneously present in arbusculated cells. *Mol Plant Microbe Interact* **20**, 1055-1062, (2007).
94 Xie, X., Lin, H., Peng, X., Xu, C., Sun, Z., Jiang, K., Huang, A., Wu, X., Tang, N., Salvioli, A., Bonfante, P. & Zhao, B. Arbuscular Mycorrhizal Symbiosis Requires a Phosphate Transceptor in the *Gigaspora margarita* Fungal Symbiont. *Mol Plant* **9**, 1583-1608, (2016).
95 Volpe, V., Giovannetti, M., Sun, X. G., Fiorilli, V. & Bonfante, P. The phosphate transporters LjPT4 and MtPT4 mediate early root responses to phosphate status in non mycorrhizal roots. *Plant Cell Environ* **39**, 660-671, (2016).
96 Paszkowski, U., Kroken, S., Roux, C. & Briggs, S. P. Rice phosphate transporters include an evolutionarily divergent gene specifically activated in arbuscular mycorrhizal symbiosis. *Proc Natl Acad Sci U S A* **99**, 13324-13329, (2002).
97 Cappellazzo, G., Lanfranco, L., Fitz, M., Wipf, D. & Bonfante, P. Characterization of an amino acid permease from the endomycorrhizal fungus *Glomus mosseae*. *Plant Physiol* **147**, 429-437, (2008).
98 Lopez-Pedrosa, A., Gonzalez-Guerrero, M., Valderas, A., Azcon-Aguilar, C. & Ferrol, N. *GintAMT1* encodes a functional high-affinity ammonium transporter that is expressed in the extraradical mycelium of *Glomus intraradices*. *Fungal Genet Biol* **43**, 102-110, (2006).
99 Perez-Tienda, J., Testillano, P. S., Balestrini, R., Fiorilli, V., Azcon-Aguilar, C. & Ferrol, N. GintAMT2, a new member of the ammonium transporter family in the arbuscular mycorrhizal fungus *Glomus intraradices*. *Fungal Genet Biol* **48**, 1044-1055, (2011).
100 Govindarajulu, M., Pfeffer, P. E., Jin, H., Abubaker, J., Douds, D. D., Allen, J. W., Bucking, H., Lammers, P. J. & Shachar-Hill, Y. Nitrogen transfer in the arbuscular mycorrhizal symbiosis. *Nature* **435**, 819-823, (2005).
101 Jin, H., Pfeffer, P. E., Douds, D. D., Piotrowski, E., Lammers, P. J. & Shachar-Hill, Y. The uptake, metabolism, transport and transfer of nitrogen in an arbuscular mycorrhizal symbiosis. *New Phytol* **168**, 687-696, (2005).
102 Breuillin-Sessoms, F., Floss, D. S., Gomez, S. K., Pumplin, N., Ding, Y., Levesque-Tremblay, V., Noar, R. D., Daniels, D. A., Bravo, A., Eaglesham, J. B., Benedito, V. A., Udvardi, M. K. & Harrison, M. J. Suppression of Arbuscule Degeneration in

Medicago truncatula phosphate transporter4 Mutants is Dependent on the Ammonium Transporter 2 Family Protein AMT2;3. *Plant Cell* **27**, 1352-1366, (2015).

103 Kobae, Y., Tamura, Y., Takai, S., Banba, M. & Hata, S. Localized expression of arbuscular mycorrhiza-inducible ammonium transporters in soybean. *Plant Cell Physiol* **51**, 1411-1415, (2010).

104 Sanders, I. R. & Croll, D. Arbuscular mycorrhiza: the challenge to understand the genetics of the fungal partner. *Annu Rev Genet* **44**, 271-292, (2010).

105 Giovannetti, M., Azzolini, D. & Citernesi, A. S. Anastomosis formation and nuclear and protoplasmic exchange in arbuscular mycorrhizal fungi. *Appl Environ Microbiol* **65**, 5571-5575, (1999).

106 Corradi, N. & Brachmann, A. Fungal Mating in the Most Widespread Plant Symbionts? *Trends Plant Sci* **22**, 175-183, (2017).

107 Sedzielewska, K. A., Fuchs, J., Temsch, E. M., Baronian, K., Watzke, R. & Kunze, G. Estimation of the *Glomus intraradices* nuclear DNA content. *New Phytol* **192**, 794-797, (2011).

108 Ropars, J., Sedzielewska Toro, K., Noel, J., Pelin, A., Charron, P., Farinelli, L., Marton, T., Kruger, M., Fuchs, J., Brachmann, A. & Corradi, N. Evidence for the sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. *Nat Microbiol* **1**, 16033, (2016).

109 Halary, S., Malik, S. B., Lildhar, L., Slamovits, C. H., Hijri, M. & Corradi, N. Conserved meiotic machinery in *Glomus spp.*, a putatively ancient asexual fungal lineage. *Genome Biol Evol* **3**, 950-958, (2011).

110 Riley, R., Charron, P., Idnurm, A., Farinelli, L., Dalpe, Y., Martin, F. & Corradi, N. Extreme diversification of the mating type-high-mobility group (*MATA-HMG*) gene family in a plant-associated arbuscular mycorrhizal fungus. *New Phytol* **201**, 254-268, (2014).

111 Lee, S. C., Ni, M., Li, W., Shertz, C. & Heitman, J. The evolution of sex: a perspective from the fungal kingdom. *Microbiol Mol Biol Rev* **74**, 298-340, (2010).

112 Redecker, D., Schussler, A., Stockinger, H., Sturmer, S. L., Morton, J. B. & Walker, C. An evidence-based consensus for the classification of arbuscular mycorrhizal fungi (Glomeromycota). *Mycorrhiza* **23**, 515-531, (2013).

113 Kluge, M., Mollenhauer, D. & Mollenhauer, R. Photosynthetic carbon assimilation in *Geosiphon pyriforme* (Kutzing) F. v. Wettstein, an endosymbiotic association of fungus and cyanobacterium. *Planta* **185**, 311-315, (1991).

114 Tisserant, E., Kohler, A., Dozolme-Seddas, P., Balestrini, R., Benabdellah, K., Colard, A., Croll, D., Da Silva, C., Gomez, S. K., Koul, R., Ferrol, N., Fiorilli, V., Formey, D., Franken, P., Helber, N., Hijri, M., Lanfranco, L., Lindquist, E., Liu, Y., Malbreil, M., Morin, E., Poulain, J., Shapiro, H., van Tuinen, D., Waschke, A., Azcon-Aguilar, C., Becard, G., Bonfante, P., Harrison, M. J., Kuster, H., Lammers, P., Paszkowski, U., Requena, N., Rensing, S. A., Roux, C., Sanders, I. R., Shachar-Hill, Y., Tuskan, G., Young, J. P., Gianinazzi-Pearson, V. & Martin, F. The transcriptome of the arbuscular mycorrhizal fungus *Glomus intraradices* (DAOM 197198) reveals functional tradeoffs in an obligate symbiont. *New Phytol* **193**, 755-769, (2012).

115 Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212, (2015).

116 Kumar, N., Lad, G., Giuntini, E., Kaye, M. E., Udomwong, P., Shamsani, N. J., Young, J. P. & Bailly, X. Bacterial genospecies that are not ecologically coherent: population genomics of *Rhizobium leguminosarum*. *Open Biol* **5**, 140133, (2015).

117 Laetsch, D. & Blaxter, M. *BlobTools: Interrogation of genome assemblies [version 1; referees: 1 approved with reservations]*. Vol. 6 (2017).

118 Hane, J. K., Anderson, J. P., Williams, A. H., Sperschneider, J. & Singh, K. B. Genome sequencing and comparative genomics of the broad host-range pathogen Rhizoctonia solani AG8. (2014).

119 Malik, S. B., Pightling, A. W., Stefaniak, L. M., Schurko, A. M. & Logsdon, J. M., Jr. An expanded inventory of conserved meiotic genes provides evidence for sex in *Trichomonas vaginalis*. *PLoS One* **3**, e2879, (2007).
120 Halary, S., Malik, S. B., Lildhar, L., Slamovits, C. H., Hijri, M. & Corradi, N. Conserved meiotic machinery in Glomus spp., a putatively ancient asexual fungal lineage. *Genome Biol Evol* **3**, 950-958, (2011).
121 Ellerbeck, M., Schussler, A., Brucker, D., Dafinger, C., Loos, F. & Brachmann, A. Characterization of three ammonium transporters of the glomeromycotan fungus Geosiphon pyriformis. *Eukaryot Cell* **12**, 1554-1562, (2013).
122 Benedetto, A., Magurno, F., Bonfante, P. & Lanfranco, L. Expression profiles of a phosphate transporter gene (GmosPT) from the endomycorrhizal fungus Glomus mosseae. *Mycorrhiza* **15**, 620-627, (2005).
123 Li, T., Hu, Y. J., Hao, Z. P., Li, H., Wang, Y. S. & Chen, B. D. First cloning and characterization of two functional aquaporin genes from an arbuscular mycorrhizal fungus Glomus intraradices. *New Phytol* **197**, 617-630, (2013).
124 De Virgilio, C., Burckert, N., Bell, W., Jeno, P., Boller, T. & Wiemken, A. Disruption of *TPS2*, the gene encoding the 100-kDa subunit of the trehalose-6-phosphate synthase/phosphatase complex in *Saccharomyces cerevisiae*, causes accumulation of trehalose-6-phosphate and loss of trehalose-6-phosphate phosphatase activity. *Eur J Biochem* **212**, 315-323, (1993).
125 Bell, W., Klaassen, P., Ohnacker, M., Boller, T., Herweijer, M., Schoppink, P., Van der Zee, P. & Wiemken, A. Characterization of the 56-kDa subunit of yeast trehalose-6-phosphate synthase and cloning of its gene reveal its identity with the product of *CIF1*, a regulator of carbon catabolite inactivation. *Eur J Biochem* **209**, 951-959, (1992).
126 Vuorio, O. E., Kalkkinen, N. & Londesborough, J. Cloning of two related genes encoding the 56-kDa and 123-kDa subunits of trehalose synthase from the yeast *Saccharomyces cerevisiae*. *Eur J Biochem* **216**, 849-861, (1993).
127 Avonce, N., Wuyts, J., Verschooten, K., Vandesteene, L. & Van Dijck, P. The Cytophaga hutchinsonii ChTPSP: First characterized bifunctional TPS-TPP protein as putative ancestor of all eukaryotic trehalose biosynthesis proteins. *Mol Biol Evol* **27**, 359-369, (2010).
128 Kamoun, S. A catalogue of the effector secretome of plant pathogenic oomycetes. *Annu Rev Phytopathol* **44**, 41-60, (2006).
129 Koren, S., Schatz, M. C., Walenz, B. P., Martin, J., Howard, J. T., Ganapathy, G., Wang, Z., Rasko, D. A., McCombie, W. R., Jarvis, E. D. & Adam, M. P. Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat Biotechnol* **30**, 693-700, (2012).
130 Goodwin, S., Gurtowski, J., Ethe-Sayers, S., Deshpande, P., Schatz, M. C. & McCombie, W. R. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res* **25**, 1750-1756, (2015).
131 Limpens, E. & Geurts, R. Plant-driven genome selection of arbuscular mycorrhizal fungi. *Mol Plant Pathol* **15**, 531-534, (2014).
132 Martin, H. *Isolierung symbiosespezifischer Gene aus Geosiphon pyriformis und funktionelle Charakterisierung des ersten Glomeromycota-Zuckertransporters*, TU Darmstadt, (2006).
133 Lanfranco, L. & Young, J. P. Genetic and genomic glimpses of the elusive arbuscular mycorrhizal fungi. *Curr. Opin. Plant. Biol.* **15**, 454-461, (2012).
134 Sanders, I. R. & Croll, D. Arbuscular mycorrhiza: the challenge to understand the genetics of the fungal partner. *Annu Rev Genet* **44**, 271-292, (2010).
135 Young, J. P. Genome diversity in arbuscular mycorrhizal fungi. *Curr. Opin. Plant Biol.* **26**, 113-119, (2015).
136 Judson, O. P. & Normark, B. B. Ancient asexual scandals. *Trends Ecol. Evol.* **11**, 41-46, (1996).

137 Ropars, J., Sedzielewska Toro, K., Noel, J., Pelin, A., Charron, P., Farinelli, L., Marton, T., Kruger, M., Fuchs, J., Brachmann, A. A.-O. & Corradi, N. Evidence for the sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. (2016).

138 Wu, S., Zhu, Z., Fu, L., Niu, B. & Li, W. WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics* **12**, 444, (2011).

139 Spanu, P. D., Abbott, J. C., Amselem, J., Burgis, T. A., Soanes, D. M., Stuber, K., Ver Loren van Themaat, E., Brown, J. K., Butcher, S. A., Gurr, S. J., Lebrun, M. H., Ridout, C. J., Schulze-Lefert, P., Talbot, N. J., Ahmadinejad, N., Ametz, C., Barton, G. R., Benjdia, M., Bidzinski, P., Bindschedler, L. V., Both, M., Brewer, M. T., Cadle-Davidson, L., Cadle-Davidson, M. M., Collemare, J., Cramer, R., Frenkel, O., Godfrey, D., Harriman, J., Hoede, C., King, B. C., Klages, S., Kleemann, J., Knoll, D., Koti, P. S., Kreplak, J., Lopez-Ruiz, F. J., Lu, X., Maekawa, T., Mahanil, S., Micali, C., Milgroom, M. G., Montana, G., Noir, S., O'Connell, R. J., Oberhaensli, S., Parlange, F., Pedersen, C., Quesneville, H., Reinhardt, R., Rott, M., Sacristan, S., Schmidt, S. M., Schon, M., Skamnioti, P., Sommer, H., Stephens, A., Takahara, H., Thordal-Christensen, H., Vigouroux, M., Wessling, R., Wicker, T. & Panstruga, R. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* **330**, 1543-1546, (2010).

140 Duplessis, S., Cuomo, C. A., Lin, Y. C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., Joly, D. L., Hacquard, S., Amselem, J., Cantarel, B. L., Chiu, R., Coutinho, P. M., Feau, N., Field, M., Frey, P., Gelhaye, E., Goldberg, J., Grabherr, M. G., Kodira, C. D., Kohler, A., Kues, U., Lindquist, E. A., Lucas, S. M., Mago, R., Mauceli, E., Morin, E., Murat, C., Pangilinan, J. L., Park, R., Pearson, M., Quesneville, H., Rouhier, N., Sakthikumar, S., Salamov, A. A., Schmutz, J., Selles, B., Shapiro, H., Tanguay, P., Tuskan, G. A., Henrissat, B., Van de Peer, Y., Rouze, P., Ellis, J. G., Dodds, P. N., Schein, J. E., Zhong, S., Hamelin, R. C., Grigoriev, I. V., Szabo, L. J. & Martin, F. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc Natl Acad Sci U S A* **108**, 9166-9171, (2011).

141 Martin, F., Aerts, A., Ahren, D., Brun, A., Danchin, E. G., Duchaussoy, F., Gibon, J., Kohler, A., Lindquist, E., Pereda, V., Salamov, A., Shapiro, H. J., Wuyts, J., Blaudez, D., Buee, M., Brokstein, P., Canback, B., Cohen, D., Courty, P. E., Coutinho, P. M., Delaruelle, C., Detter, J. C., Deveau, A., DiFazio, S., Duplessis, S., Fraissinet-Tachet, L., Lucic, E., Frey-Klett, P., Fourrey, C., Feussner, I., Gay, G., Grimwood, J., Hoegger, P. J., Jain, P., Kilaru, S., Labbe, J., Lin, Y. C., Legue, V., Le Tacon, F., Marmeisse, R., Melayah, D., Montanini, B., Muratet, M., Nehls, U., Niculita-Hirzel, H., Oudot-Le Secq, M. P., Peter, M., Quesneville, H., Rajashekar, B., Reich, M., Rouhier, N., Schmutz, J., Yin, T., Chalot, M., Henrissat, B., Kues, U., Lucas, S., Van de Peer, Y., Podila, G. K., Polle, A., Pukkila, P. J., Richardson, P. M., Rouze, P., Sanders, I. R., Stajich, J. E., Tunlid, A., Tuskan, G. & Grigoriev, I. V. The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **452**, 88-92, (2008).

142 Martin, F., Kohler, A., Murat, C., Balestrini, R., Coutinho, P. M., Jaillon, O., Montanini, B., Morin, E., Noel, B., Percudani, R., Porcel, B., Rubini, A., Amicucci, A., Amselem, J., Anthouard, V., Arcioni, S., Artiguenave, F., Aury, J. M., Ballario, P., Bolchi, A., Brenna, A., Brun, A., Buee, M., Cantarel, B., Chevalier, G., Couloux, A., Da Silva, C., Denoeud, F., Duplessis, S., Ghignone, S., Hilselberger, B., Iotti, M., Marcais, B., Mello, A., Miranda, M., Pacioni, G., Quesneville, H., Riccioni, C., Ruotolo, R., Splivallo, R., Stocchi, V., Tisserant, E., Viscomi, A. R., Zambonelli, A., Zampieri, E., Henrissat, B., Lebrun, M. H., Paolocci, F., Bonfante, P., Ottonello, S. & Wincker, P. Perigord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* **464**, 1033-1038, (2010).

143 Lombard, V., Bernard, T., Rancurel, C., Brumer, H., Coutinho, P. M. & Henrissat, B. A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochem J* **432**, 437-444, (2010).

144 Malinovsky, F. G., Fangel, J. U. & Willats, W. G. T. The role of the cell wall in plant immunity. *Frontiers in Plant Science* **5**, (2014).
145 Silipo, A., Erbs, G., Shinya, T., Dow, J. M., Parrilli, M., Lanzetta, R., Shibuya, N., Newman, M. A. & Molinaro, A. Glyco-conjugates as elicitors or suppressors of plant innate immunity. *Glycobiology* **20**, 406-419, (2010).
146 Shibuya, N. & Minami, E. Oligosaccharide signalling for defence responses in plant. *Physiological and Molecular Plant Pathology* **59**, 223-233, (2001).
147 Lee, J., Park, J. S., Moon, J. Y., Kim, K. Y. & Moon, H. M. The influence of glycosylation on secretion, stability, and immunogenicity of recombinant HBV pre-S antigen synthesized in *Saccharomyces cerevisiae*. *Biochem Biophys Res Commun* **303**, 427-432, (2003).
148 Zhao, Z., Liu, H., Wang, C. & Xu, J. R. Correction: Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics* **15**, 6, (2014).
149 Guether, M., Neuhauser, B., Balestrini, R., Dynowski, M., Ludewig, U. & Bonfante, P. A mycorrhizal-specific ammonium transporter from Lotus japonicus acquires nitrogen released by arbuscular mycorrhizal fungi. *Plant Physiol* **150**, 73-83, (2009).
150 Schüßler, A., Bonfante, P., Schnepf, E., Mollenhauer, D. & Kluge, M. Characterization of the *Geosiphon pyriforme* symbiosome by affinity techniques: confocal laser scanning microscopy (CLSM) and electron microscopy. *Protoplasma* **190**, 53-67, (1996).
151 Wente, S. R. & Rout, M. P. The nuclear pore complex and nuclear transport. *Cold Spring Harb Perspect Biol* **2**, a000562, (2010).
152 Wang, Y. Y., Liu, B., Zhang, X. Y., Zhou, Q. M., Zhang, T., Li, H., Yu, Y. F., Zhang, X. L., Hao, X. Y., Wang, M., Wang, L. & Wei, J. C. Genome characteristics reveal the impact of lichenization on lichen-forming fungus *Endocarpon pusillum* Hedwig (Verrucariales, Ascomycota). *BMC Genomics* **15**, 34, (2014).
153 Linding, R., Russell, R. B., Neduva, V. & Gibson, T. J. GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res* **31**, 3701-3708, (2003).
154 Afsar Minhas, F. U. A., Ross, E. D. & Ben-Hur, A. Amino acid composition predicts prion activity. *PLoS Comput Biol* **13**, e1005465, (2017).
155 Maurer-Stroh, S., Debulpaep, M., Kuemmerer, N., Lopez de la Paz, M., Martins, I. C., Reumers, J., Morris, K. L., Copland, A., Serpell, L., Serrano, L., Schymkowitz, J. W. & Rousseau, F. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat Methods* **7**, 237-242, (2010).
156 Blanco, L. P., Evans, M. L., Smith, D. R., Badtke, M. P. & Chapman, M. R. Diversity, biogenesis and function of microbial amyloids. *Trends Microbiol* **20**, 66-73, (2012).
157 Uptain, S. M. & Lindquist, S. Prions as protein-based genetic elements. *Annu Rev Microbiol* **56**, 703-741, (2002).
158 Teertstra, W. R., van der Velden, G. J., de Jong, J. F., Kruijtzer, J. A., Liskamp, R. M., Kroon-Batenburg, L. M., Muller, W. H., Gebbink, M. F. & Wosten, H. A. The filament-specific Rep1-1 repellent of the phytopathogen *Ustilago maydis* forms functional surface-active amyloid-like fibrils. *J Biol Chem* **284**, 9153-9159, (2009).
159 Dittmann, E., Fewer, D. P. & Neilan, B. A. Cyanobacterial toxins: biosynthetic routes and evolutionary roots. *FEMS Microbiol Rev* **37**, 23-43, (2013).
160 Jones, J. D. & Dangl, J. L. The plant immune system. *Nature* **444**, 323-329, (2006).
161 Pliego, C., Nowara, D., Bonciani, G., Gheorghe, D. M., Xu, R., Surana, P., Whigham, E., Nettleton, D., Bogdanove, A. J., Wise, R. P., Schweizer, P., Bindschedler, L. V. & Spanu, P. D. Host-induced gene silencing in barley powdery mildew reveals a class of ribonuclease-like effectors. *Mol Plant Microbe Interact* **26**, 633-642, (2013).
162 Cheng, W., Song, X. S., Li, H. P., Cao, L. H., Sun, K., Qiu, X. L., Xu, Y. B., Yang, P., Huang, T., Zhang, J. B., Qu, B. & Liao, Y. C. Host-induced gene silencing of an essential chitin synthase gene confers durable resistance to Fusarium head blight and seedling blight in wheat. *Plant Biotechnol J* **13**, 1335-1345, (2015).

163 Koch, A., Biedenkopf, D., Furch, A., Weber, L., Rossbach, O., Abdellatef, E., Linicus, L., Johannsmeier, J., Jelonek, L., Goesmann, A., Cardoza, V., McMillan, J., Mentzel, T. & Kogel, K. H. An RNAi-Based Control of F*usarium graminearum* Infections Through Spraying of Long dsRNAs Involves a Plant Passage and Is Controlled by the Fungal Silencing Machinery. *PLoS Pathog* **12**, e1005901, (2016).

164 Paul, M. J., Primavesi, L. F., Jhurreea, D. & Zhang, Y. Trehalose metabolism and signaling. *Annu Rev Plant Biol* **59**, 417-441, (2008).

165 Bago, B., Pfeffer, P. E., Douds, D. D., Jr., Brouillette, J., Becard, G. & Shachar-Hill, Y. Carbon metabolism in spores of the arbuscular mycorrhizal fungus *Glomus intraradices* as revealed by nuclear magnetic resonance spectroscopy. *Plant Physiol* **121**, 263-272, (1999).

166 Lowe, R. G., Lord, M., Rybak, K., Trengove, R. D., Oliver, R. P. & Solomon, P. S. Trehalose biosynthesis is involved in sporulation of *Stagonospora nodorum*. *Fungal Genet Biol* **46**, 381-389, (2009).

167 Schellenbaum, L., Muller, J., Boller, T., Wiemken, A. & Schuepp, H. Effects of drought on non-mycorrhizal and mycorrhizal maize: changes in the pools of non-structural carbohydrates, in the activities of invertase and trehalase, and in the pools of amino acids and imino acids. *New Phytologist* **138**, 59-66, (1998).

168 Kobae, Y. & Hata, S. Dynamics of periarbuscular membranes visualized with a fluorescent phosphate transporter in arbuscular mycorrhizal roots of rice. *Plant Cell Physiol* **51**, 341-353, (2010).

169 Baier, M. C., Keck, M., Godde, V., Niehaus, K., Kuster, H. & Hohnjec, N. Knockdown of the symbiotic sucrose synthase MtSucS1 affects arbuscule maturation and maintenance in mycorrhizal roots of *Medicago truncatula*. *Plant Physiol* **152**, 1000-1014, (2010).

170 Park, H. J., Floss, D. S., Levesque-Tremblay, V., Bravo, A. & Harrison, M. J. Hyphal branching during arbuscule development requires *Reduced Arbuscular Mycorrhiza1*. *Plant Physiol* **169**, 2774-2788, (2015).

171 Wang, E., Schornack, S., Marsh, J. F., Gobbato, E., Schwessinger, B., Eastmond, P., Schultze, M., Kamoun, S. & Oldroyd, G. E. A common signaling process that promotes mycorrhizal and oomycete colonization of plants. *Curr Biol* **22**, 2242-2246, (2012).

172 Javot, H., Penmetsa, R. V., Terzaghi, N., Cook, D. R. & Harrison, M. J. A *Medicago truncatula* phosphate transporter indispensable for the arbuscular mycorrhizal symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 1720-1725, (2007).

173 Xie, X., Huang, W., Liu, F., Tang, N., Liu, Y., Lin, H. & Zhao, B. Functional analysis of the novel mycorrhiza-specific phosphate transporter AsPT1 and PHT1 family from *Astragalus sinicus* during the arbuscular mycorrhizal symbiosis. *New Phytol* **198**, 836-852, (2013).

174 Durand, M., Mainson, D., Porcheron, B., Maurousset, L., Lemoine, R. & Pourtau, N. Carbon source-sink relationship in *Arabidopsis thaliana*: the role of sucrose transporters. *Planta* **247**, 587-611, (2018).

175 Yadav, U. P., Ivakov, A., Feil, R., Duan, G. Y., Walther, D., Giavalisco, P., Piques, M., Carillo, P., Hubberten, H. M., Stitt, M. & Lunn, J. E. The sucrose-trehalose 6-phosphate (Tre6P) nexus: specificity and mechanisms of sucrose signalling by Tre6P. *J Exp Bot* **65**, 1051-1068, (2014).

176 Thevelein, J. M. & Hohmann, S. Trehalose synthase: guard to the gate of glycolysis in yeast? *Trends Biochem Sci* **20**, 3-10, (1995).

177 Elbein, A. D., Pan, Y. T., Pastuszak, I. & Carroll, D. New insights on trehalose: a multifunctional molecule. *Glycobiology* **13**, 17R-27R, (2003).

178 Bell, W., Sun, W., Hohmann, S., Wera, S., Reinders, A., De Virgilio, C., Wiemken, A. & Thevelein, J. M. Composition and functional analysis of the *Saccharomyces cerevisiae* trehalose synthase complex. *J Biol Chem* **273**, 33311-33319, (1998).

179 Binder, A., Lambert, J., Morbitzer, R., Popp, C., Ott, T., Lahaye, T. & Parniske, M. A modular plasmid assembly kit for multigene expression, gene silencing and silencing rescue in plants. (2014).
180 Chabot, S., Becard, G. & Piche, Y. *Life Cycle of Glomus intraradix in Root Organ Culture*. Vol. 84 (1992).
181 *In Vitro Culture of Mycorrhizas*. Vol. 1 (Springer-Verlag Berlin Heidelberg, 2005).
182 Jacobs, K. A., Collins-Racie, L. A., Colbert, M., Duckett, M., Golden-Fleet, M., Kelleher, K., Kriz, R., LaVallie, E. R., Merberg, D., Spaulding, V., Stover, J., Williamson, M. J. & McCoy, J. M. A genetic selection for isolating cDNAs encoding secreted proteins. *Gene* **198**, 289-296, (1997).
183 Broughton, W. J. & Dilworth, M. J. Control of leghaemoglobin synthesis in snake beans. *Biochem J* **125**, 1075-1080, (1971).
184 Alberti, S., Gitler, A. D. & Lindquist, S. A suite of Gateway cloning vectors for high-throughput genetic analysis in *Saccharomyces cerevisiae*. *Yeast* **24**, 913-919, (2007).
185 Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P. & Cardona, A. Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682, (2012).
186 Froger, A. & Hall, J. E. Transformation of plasmid DNA into *E. coli* using the heat shock method. *J Vis Exp*, 253, (2007).
187 Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* **2**, 31-34, (2007).
188 Amberg, D. C. B., D. J.; Strathern, J- N-. *Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual*. 2005 Edition edn, (Cold Spring, 2005).
189 Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R. & Leunissen, J. A. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* **35**, W71-74, (2007).
190 Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *Plos Computational Biology* **13**, (2017).
191 Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F. F. & Stevens, R. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research* **42**, D206-D214, (2014).
192 Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A. & Zagnitko, O. The RAST server: Rapid annotations using subsystems technology. *Bmc Genomics* **9**, (2008).
193 Vallenet, D., Calteau, A., Cruveiller, S., Gachet, M., Lajus, A., Josso, A., Mercier, J., Renaux, A., Rollin, J., Rouy, Z., Roche, D., Scarpelli, C. & Medigue, C. MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes. *Nucleic Acids Research* **45**, D517-D528, (2017).
194 Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics* **27**, 1009-1010, (2011).
195 Alikhan, N. F., Petty, N. K., Ben Zakour, N. L. & Beatson, S. A. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **12**, 402, (2011).
196 Torres-Cortes, G., Ghignone, S., Bonfante, P. & Schussler, A. Mosaic genome of endobacteria in arbuscular mycorrhizal fungi: Transkingdom gene transfer in an ancient mycoplasma-fungus association. *Proc Natl Acad Sci U S A* **112**, 7785-7790, (2015).

197 R: A Language and Environment for Statistical Computing (Vienna, Austria, 2008).
198 Csardi, G. N., T. The igraph software package for complex network research. *InterJournal* **Complex Systems**, 1695, (2005).
199 Graves, S. P., Hans-Peter; Selzer, L. multcompView: Visualizations of Paired Comparisons. (2012).
200 Grigoriev, I. V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., Kuo, A., Minovitsky, S., Nikitin, R., Ohm, R. A., Otillar, R., Poliakov, A., Ratnere, I., Riley, R., Smirnova, T., Rokhsar, D. & Dubchak, I. The Genome Portal of the Department of Energy Joint Genome Institute. *Nucleic Acids Research* **40**, D26-D32, (2012).
201 Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**, 785-786, (2011).
202 Horn, T., Sandmann, T. & Boutros, M. Design and evaluation of genome-wide libraries for RNA interference screens. *Genome Biology* **11**, R61, (2010).

6 Supplementary information

List of supplementary figures

List of supplementary tables

List of electronic supplementary information

For retrieving electronic supplementary information please contact anne.hoffrichter.dissertation@gmail.com.

Supplementary Figures

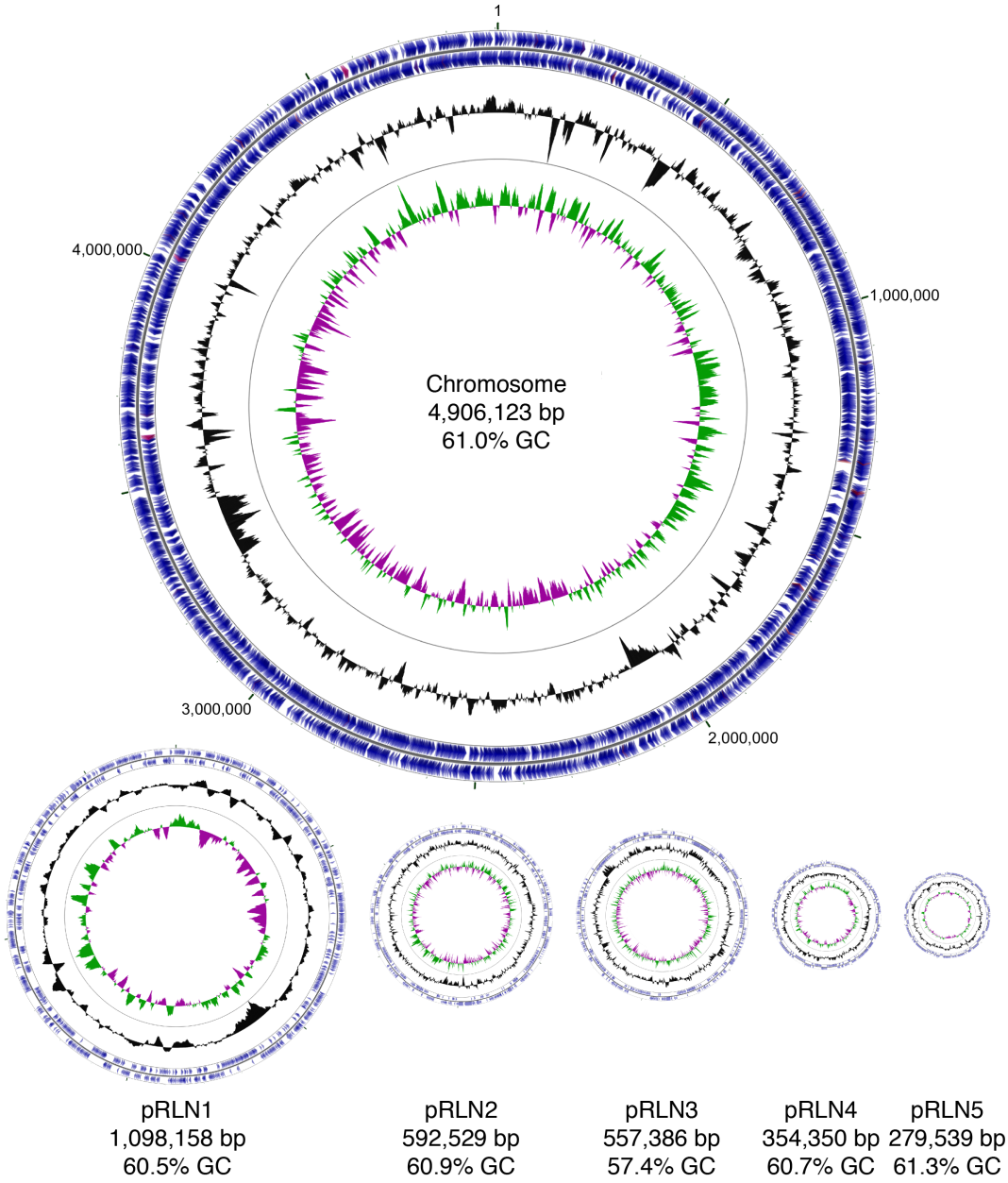

Fig. S1: The chromosome and five plasmids of RI Norway. The plasmids are depicted to scale with the chromosome one-half of this scale. The outermost circles show protein encoding genes (blue) and rRNA and tRNA genes (red) in clockwise and counter-clockwise orientation. The inner circles indicate deviations in GC content (black) and GC skew (green/purple). Plasmid maps were generated using GCView (Stothard & Wishart 2005).

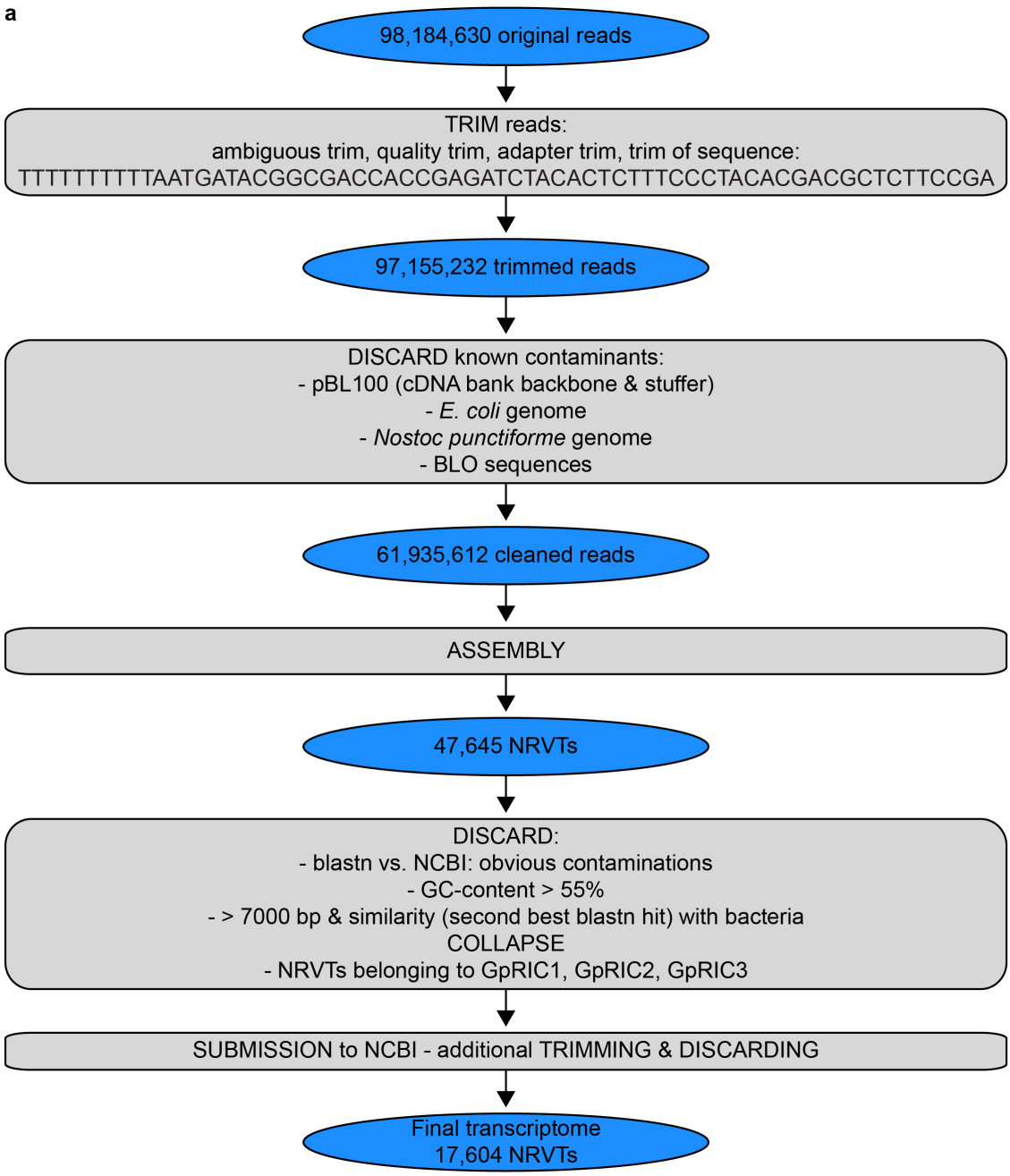

Fig. S2: Trimming and filtering of reads, assembly and filtering of contigs. (a) Pipeling of filtering reads and contigs from the original reads to the final transcriptome.

b

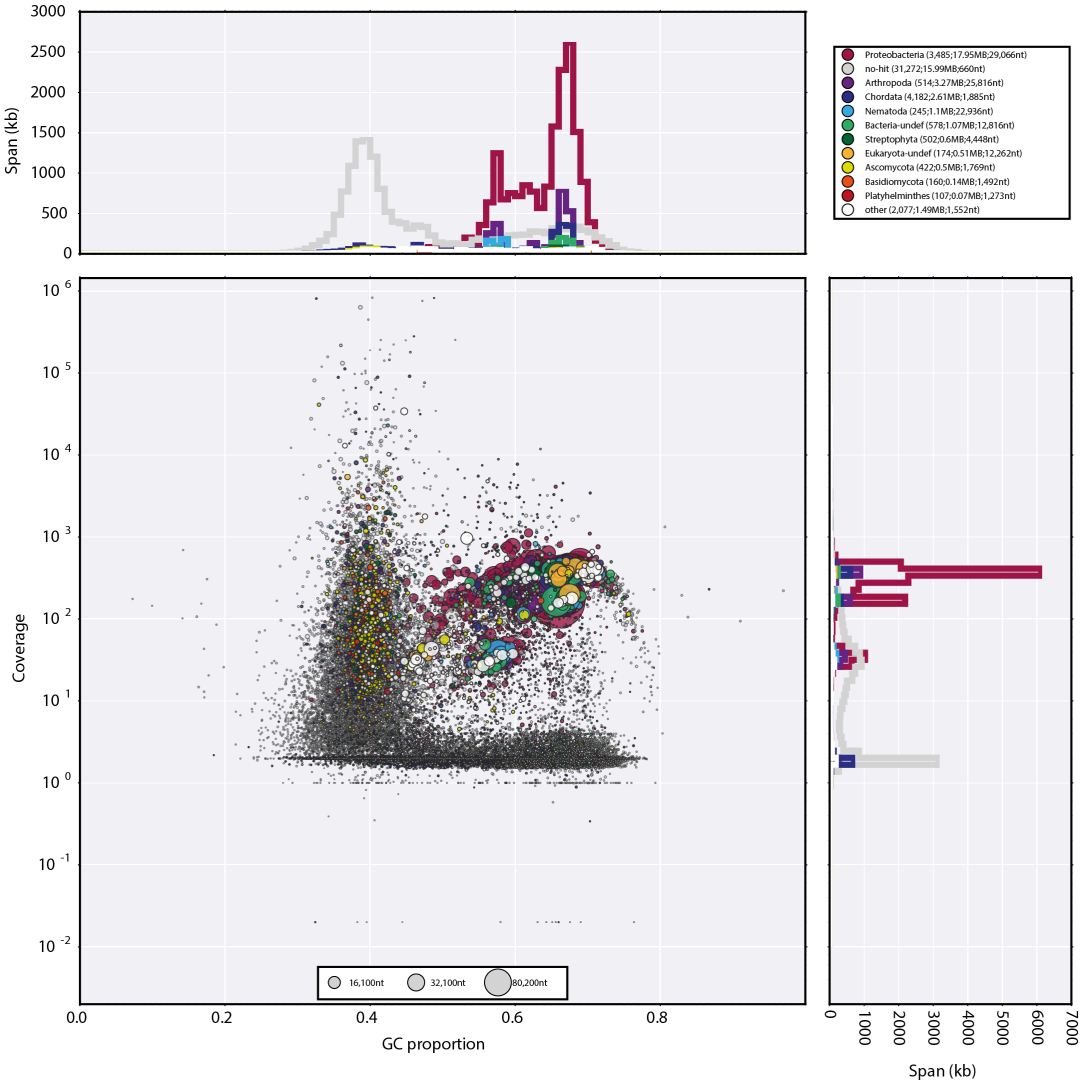

Fig. S2: Trimming and filtering of reads, assembly and filtering of contigs. (b) Blobplot analysis of the original assembly shows high amount od bacterial contamination with a GC-content around 60%. Each “Blob” represents one contig.

c

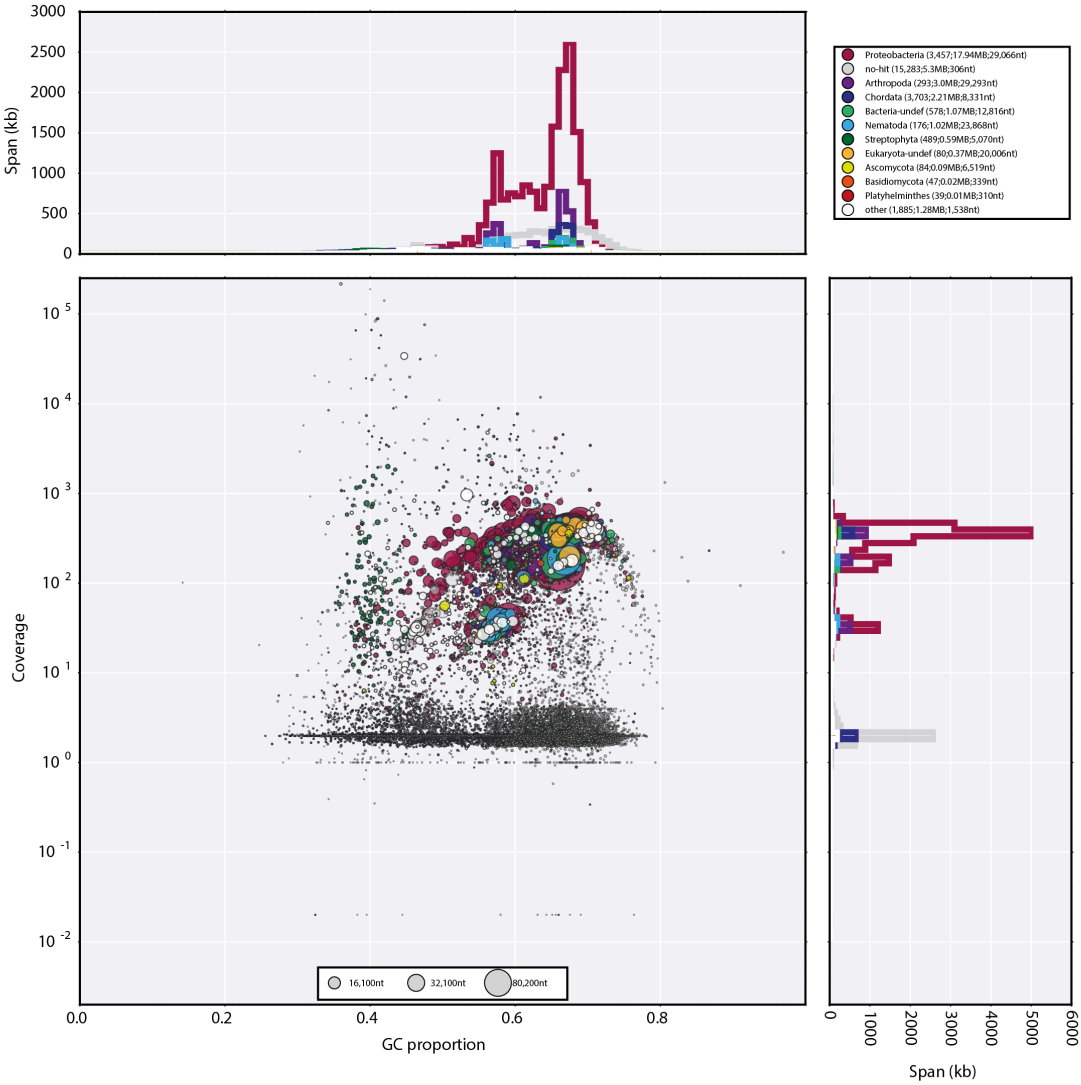

Fig. S2: Trimming and filtering of reads, assembly and filtering of contigs. (c) Blobplot analysis of all contigs that were removed due to origin, GC, or length.

d

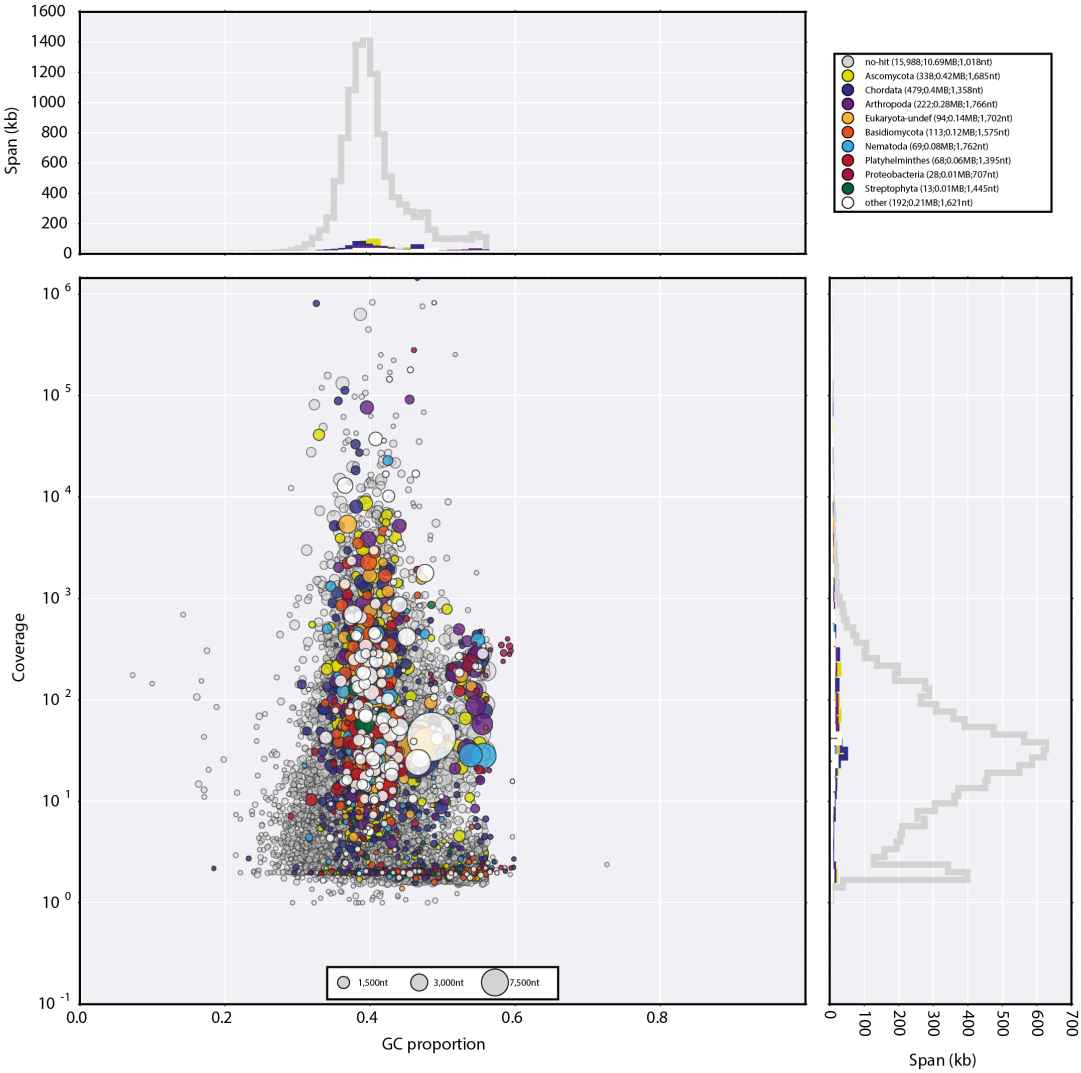

Fig. S2: Trimming and filtering of reads, assembly and filtering of contigs. (d) Blobplot analysis of NRVTS in the final transcriptome.

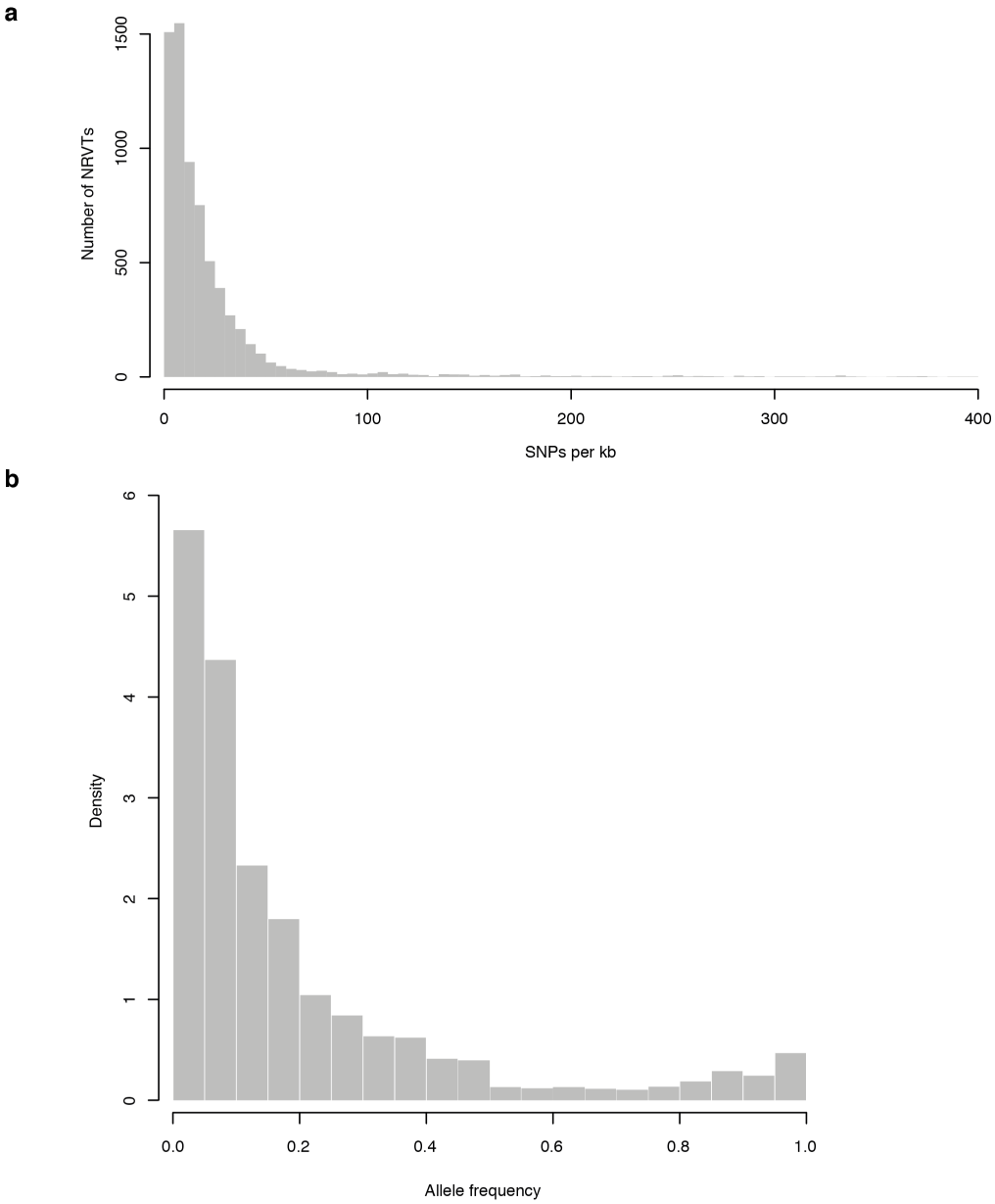

Fig. S3: SNP distribution and allele frequency in *G. pyriformis* transcriptome. (a) Number of NRVTs that contain different amounts of SNPs. (b) Occurrence of allele frequency of SNPs.

a

b

KOG protein functional class	Glomeromycota					Ascomycota				Basidiomycota				Zygomycota	
	G. pyriformis	*G. margarita*	*G. rosea*	*R. irregularis*	*R. irregularis* genome	*B. graminis*	*C. albicans*	*M. grisea*	*S. cerevisiae*	*C. cinerea*	*L. bicolor*	*M. larici -populina*	*U. maydis*	*M. elongata*	*R. oryzae*
Signal transduction mechanisms	7.7%	19.5%	36.0%	23.4%	18.8%	7.8%	7.0%	7.3%	6.6%	9.1%	9.7%	8.4%	7.4%	9.7%	10.2%
Posttranslational modification, protein turnover, chaperones	9.6%	10.2%	19.2%	11.7%	11.3%	10.1%	8.9%	7.9%	9.3%	8.7%	8.6%	9.4%	8.3%	9.0%	8.6%
General function prediction only	11.9%	11.2%	12.3%	13.3%	13.8%	11.3%	11.7%	14.3%	11.5%	12.3%	11.6%	10.6%	12.4%	13.3%	10.9%
Cell wall/membrane/envelope biogenesis	1.6%	4.0%	11.7%	5.8%	5.0%	1.2%	1.0%	1.3%	1.4%	1.6%	1.6%	1.1%	1.2%	2.2%	1.7%
RNA processing and modification	5.7%	5.2%	6.5%	2.9%	2.9%	5.1%	4.8%	3.8%	5.2%	4.4%	4.5%	4.9%	4.9%	4.8%	3.9%
Function unknown	6.7%	5.4%	5.0%	4.1%	5.7%	6.7%	5.8%	5.1%	5.9%	5.1%	5.9%	6.0%	6.0%	6.1%	5.1%
Lipid transport and metabolism	6.8%	3.8%	4.4%	3.2%	3.3%	4.2%	4.0%	5.9%	3.0%	4.2%	3.9%	4.6%	5.5%	4.0%	4.4%
Carbohydrate transport and metabolism	2.5%	2.9%	4.4%	1.8%	2.4%	3.1%	3.5%	4.5%	4.0%	3.7%	3.7%	4.4%	3.5%	2.9%	3.9%
Intracellular trafficking, secretion, and vesicular transport	5.6%	4.5%	4.2%	3.8%	3.5%	7.2%	7.0%	6.5%	7.3%	7.7%	6.3%	6.8%	6.6%	6.4%	6.4%
Translation, ribosomal structure and biogenesis	7.5%	3.7%	4.2%	3.9%	3.4%	7.2%	7.2%	5.3%	8.4%	5.5%	5.1%	5.9%	6.5%	5.2%	7.4%
Transcription	6.8%	4.5%	4.2%	4.3%	5.1%	6.3%	7.4%	5.5%	6.7%	6.9%	6.3%	6.6%	7.1%	9.7%	7.5%
Secondary metabolites biosynthesis, transport and catabolism	2.0%	3.0%	4.1%	2.7%	2.4%	1.4%	2.1%	5.1%	1.4%	4.2%	3.4%	2.5%	2.7%	2.0%	1.7%
Nuclear structure	0.4%	1.7%	3.9%	0.3%	1.0%	0.6%	0.7%	0.5%	0.7%	0.5%	0.5%	0.5%	0.6%	0.9%	0.6%
Replication, recombination and repair	4.1%	3.5%	3.6%	3.3%	3.2%	4.1%	3.7%	3.0%	3.8%	3.5%	5.2%	4.3%	3.6%	2.6%	3.1%
Cell cycle control, cell division, chromosome partitioning	2.7%	3.2%	3.1%	2.9%	4.3%	3.8%	3.8%	2.7%	3.9%	3.6%	4.2%	4.2%	3.2%	3.2%	3.6%
Energy production and conversion	3.5%	2.7%	2.9%	2.8%	2.4%	4.3%	4.8%	5.3%	4.5%	4.0%	4.1%	4.2%	4.7%	5.0%	4.4%
Amino acid transport and metabolism	3.6%	2.5%	2.8%	2.0%	1.8%	4.1%	4.8%	4.3%	4.8%	3.2%	3.2%	3.9%	4.3%	3.3%	3.8%
Cytoskeleton	3.1%	2.2%	2.4%	2.0%	2.2%	3.6%	3.1%	3.8%	2.6%	4.4%	4.0%	3.3%	3.3%	3.1%	3.8%
Inorganic ion transport and metabolism	2.0%	1.7%	2.1%	1.8%	2.1%	2.0%	2.6%	2.6%	2.4%	2.1%	2.1%	2.3%	2.1%	2.4%	2.5%
Chromatin structure and dynamics	1.8%	1.5%	1.5%	1.6%	2.7%	2.1%	2.1%	1.6%	2.1%	2.0%	2.2%	2.3%	1.9%	1.4%	3.9%
Coenzyme transport and metabolism	1.3%	1.1%	1.2%	0.8%	0.6%	1.7%	1.6%	1.6%	1.9%	1.3%	1.5%	1.6%	1.6%	1.0%	1.1%
Defense mechanisms	0.9%	0.8%	1.1%	0.6%	1.0%	0.4%	0.8%	0.7%	0.6%	0.8%	0.7%	0.6%	0.8%	0.6%	0.6%
Nucleotide transport and metabolism	1.9%	0.8%	0.7%	0.7%	0.6%	1.5%	1.5%	1.1%	1.8%	1.0%	1.5%	1.3%	1.4%	0.9%	1.1%
Extracellular structures	0.3%	0.1%	0.1%	0.1%	0.4%	0.1%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.3%	0.2%	0.2%
Cell motility	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.1%	0.0%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%

Fig. S4: Comparison of *G. pyriformis* proteins to other fungal proteins, and KOG distribution. (a) Number of NRVTs with KOG hits; all NRVTs compared to other fungal proteomes and KOG hits in these subsets. (b) KOG functional class distribution in *G. pyriformis* compared to other funig.

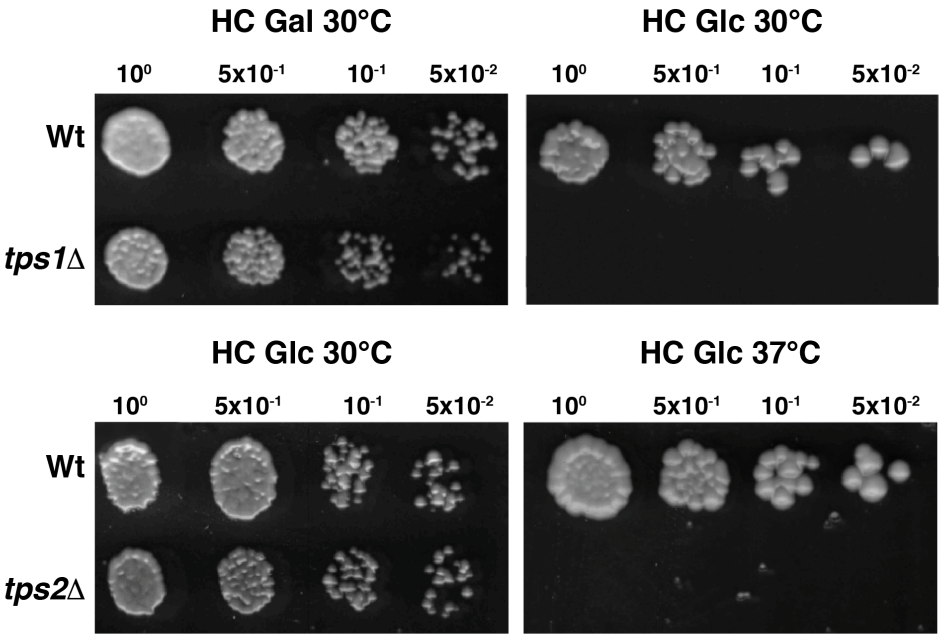

Fig. S5: Phenotypes of deletion mutants *tps1*Δ and *tps2*Δ on glucose or at 37°C. Yeast were spotted at different cell densities (10^0 corresponds to $1.5x10^6$ cells/ml), 2 μl of each dilution were spotted.

Supplementary tables

Table S1: List of NRVTs identified as HMG-genes in *G. pyriformis*.

NRVT	HMG	E-value	Score
Contig_10772	P25042.1	9.98E-10	49.3
Contig_11291	BAC66503.1	7.62E-09	43.9
Contig_11821	AAF00498.1	9.57E-12	49.7
Contig_12164	CAA06843.1	0.001	30
Contig_1226	CAD21099.1	0.012	27.7
Contig_12300	CAD21099.1	3.38E-14	64.7
Contig_13149	BAC66503.1	0.000125	30.8
Contig_137	AAK83344.1	0.006	29.6
Contig_13858	P36631.2	1.47E-11	55.1
Contig_13859	AAL30836.1	1.36E-05	36.2
Contig_14826	BAC66503.1	0.000118	33.1
Contig_15623	AET35422.1	0.00071	28.9
Contig_1569	AAC71053.1	0.007	28.1
Contig_16229	AET35422.1	8.05E-07	39.7
Contig_17299	C7U331.1	0.000821	30
Contig_17611	Q99101	5.93E-08	43.1
Contig_17875	AET35419.1	2.78E-05	31.2
Contig_18462	AAG42812.1	0.000538	28.5
Contig_19567	BAC66503.1	0.001	28.5
Contig_21333	P36981.2	1.42E-09	45.1
Contig_21363	AAG42810.1	1.70E-06	37
Contig_2140	C7U331.1	1.48E-05	34.7
Contig_29916	CAD21099.1	0.000341	28.1
Contig_3232	AAG42812.1	0.000623	27.3
Contig_3344	Q99101	2.57E-05	36.6
Contig_4059	P36981.2	0.001	30.4
Contig_4235	P36981.2	2.25E-06	39.7
Contig_5635	AAF00498.1	9.10E-11	50.4
Contig_7334	Q02991.1	0.000183	30.4
Contig_7834	P25042.1	0.002	30.4
Contig_8150	Q02991.1	2.93E-07	42.7
Contig_8190	P25042.1	0.003	31.2
Contig_8621	AAF00498.1	0.000442	30.4
Contig_8948	P35693.2	0.002	31.2
Contig_9242	AAK83344.1	4.77E-15	63.2
Contig_9620	C7U331.1	1.99E-26	90.5

Table S2: List of NRVTs identified as CAZymes in *G. pyriformis*. * indicates that more than one class was assigned to the respective NRVT.

Query	Subject	E-value	Subject Start	Subject End	Query Start	Query End	Covered fraction
Contig_3214*	AA11	1.10E-25	26	146	59	187	0.628
Contig_4653	AA11	3.20E-12	44	149	129	228	0.550
Contig_9826	AA11	2.00E-11	21	145	85	199	0.649
Contig_2942	AA2	3.30E-64	3	254	108	360	0.984
Contig_10566	AA5	2.80E-104	582	1018	5	495	0.340
Contig_135	AA5	3.50E-126	556	1027	26	525	0.368
Contig_11795	AA6	6.40E-76	1	193	3	195	0.985
Contig_1224	AA7	1.30E-18	6	182	48	227	0.384
Contig_1725	AA7	1.80E-67	3	455	35	482	0.987
Contig_10623*	CBM18	1.40E-10	1	36	15	52	0.921
Contig_10623*	CBM18	7.20E-13	2	36	96	130	0.895
Contig_3858*	CBM18	1.40E-08	2	36	28	63	0.895
Contig_3214*	CBM19	5.30E-07	6	44	317	358	0.844
Contig_6888	CBM4	0.00011	5	77	17	91	0.571
Contig_34302	CBM50	4.20E-14	1	39	26	68	0.950
Contig_6639	CBM50	3.60E-15	1	40	111	161	0.975
Contig_12349	CE1	2.30E-13	14	207	3	207	0.850
Contig_6151	CE1	2.50E-09	32	208	1	187	0.775
Contig_9814	CE1	1.20E-06	3	216	40	255	0.938
Contig_12436	CE10	1.70E-30	98	298	1	197	0.587
Contig_1282	CE10	4.20E-55	93	333	142	381	0.704
Contig_5626	CE10	5.00E-34	136	313	10	183	0.519
Contig_28453	CE14	3.70E-11	3	73	2	61	0.565
Contig_12461	CE16	7.70E-18	3	266	50	310	0.985
Contig_18138	CE16	1.80E-06	105	203	10	103	0.367
Contig_2415	CE16	2.20E-18	1	215	32	234	0.801
Contig_3345	CE16	3.60E-13	1	216	32	234	0.805
Contig_7312	CE16	1.20E-22	1	266	50	309	0.993
Contig_9417	CE16	1.40E-28	32	264	62	300	0.869
Contig_17173	CE4	4.70E-15	54	125	2	74	0.546
Contig_3524	CE4	2.10E-27	9	123	133	249	0.877
Contig_3858*	CE4	2.20E-31	6	126	125	251	0.923
Contig_42234	CE4	3.60E-11	23	68	54	99	0.346
Contig_7677	GH109	2.00E-10	2	118	44	155	0.921
Contig_17859	GH125	3.40E-63	4	228	1	244	0.557
Contig_8355	GH125	3.20E-155	1	402	76	499	0.998
Contig_6070	GH13	1.00E-27	130	292	3	168	0.542
Contig_712	GH13	2.40E-23	22	280	69	357	0.863
Contig_2022	GH133	2.20E-150	6	371	233	691	0.981

Query	Subject	E-value	Subject Start	Subject End	Query Start	Query End	Covered fraction
Contig_1098	GH15	1.00E-80	3	360	88	465	0.989
Contig_16539	GH16	1.60E-24	22	159	19	154	0.725
Contig_4024	GH16	4.90E-27	2	188	87	316	0.984
Contig_2309	GH17	1.60E-28	15	306	138	389	0.936
Contig_3086	GH17	2.20E-29	22	307	125	367	0.916
Contig_25898	GH18	8.80E-26	94	193	1	134	0.334
Contig_1913	GH20	7.70E-93	8	336	208	575	0.973
Contig_7223	GH20	1.60E-44	155	336	2	190	0.537
Contig_32098	GH23	1.60E-07	66	114	2	76	0.356
Contig_7580	GH27	4.00E-43	179	346	1	152	0.445
Contig_9721	GH31	5.70E-90	199	427	2	233	0.534
Contig_12205	GH37	6.40E-90	133	418	1	324	0.580
Contig_5851	GH37	2.70E-101	168	489	1	362	0.654
Contig_10165	GH38	4.30E-91	3	260	267	523	0.955
Contig_13893	GH47	3.50E-101	122	446	13	354	0.726
Contig_8157	GH47	1.10E-79	38	249	2	210	0.473
Contig_20073	GH5	9.40E-06	164	261	10	104	0.353
Contig_3454	GH5	4.10E-45	27	275	87	397	0.902
Contig_8780	GH63	6.60E-08	318	492	17	146	0.305
Contig_7110	GH99	1.50E-87	68	314	1	255	0.737
Contig_12167	GT2	1.80E-09	99	164	1	67	0.387
Contig_15162	GT2	1.50E-17	73	164	32	124	0.542
Contig_2881	GT2	3.00E-06	106	164	4	63	0.345
Contig_3294	GT2	1.30E-35	1	166	71	250	0.982
Contig_3933	GT2	2.20E-46	1	168	9	179	0.994
Contig_42559	GT2	1.50E-11	3	58	17	71	0.327
Contig_5714	GT2	2.70E-29	1	130	15	146	0.768
Contig_10332	GT22	1.30E-62	4	360	13	465	0.915
Contig_3304	GT3	2.50E-140	316	623	1	315	0.482
Contig_12481	GT31	1.20E-31	19	191	9	179	0.896
Contig_9054	GT39	5.50E-29	121	223	1	105	0.457
Contig_15245	GT4	4.20E-10	51	107	7	69	0.350
Contig_7597	GT41	4.70E-10	17	243	170	457	0.321
Contig_2929	GT49	7.30E-07	220	327	14	121	0.318
Contig_10337	GT62	3.40E-50	117	268	12	158	0.563
Contig_6701	GT8	6.20E-53	30	254	17	259	0.872

Table S3: List of Missing Glomeromycota core genes (MGCGs) and identified NRVT in *G. pyriformis*. *Gp*: *G. pyriformis*, *Gm*: *G. margarita*, *Gr*: *G. rosea*, *Ri*: *R. irregularis*

Gene ID	Name	Description	*Gp*	*Gm*	*Gr*	*Ri*
Thiamine Metabolism/Transport						
YGR144W	THI4	Thiazole synthase	Missing	Missing	Missing	Missing
YPL214C	THI6	Bifunctional enzyme with thiamine-phosphate pyrophosphorylase and 4-methyl-5-beta-hydroxyethylthiazole kinase activities	Missing	Missing	Missing	Missing
YLR237W	THI7	Plasma membrane transporter responsible for the uptake of thiamine	Missing	Missing	Missing	Missing
YOL055C	THI20	Multifunctional protein with hydroxymethylpyrimidine phosphate (HMP-P) kinase and thiaminase activities	Missing	Missing	Missing	Missing
YPL258C	THI21	Hydroxymethylpyrimidine phosphate kinase	Missing	Missing	Missing	Missing
YPR121W	THI22	hydroxymethylpyrimidine phosphate kinases	Missing	Missing	Missing	Missing
YOR192C	THI72	Transporter of thiamine or related compound	Missing	Missing	Missing	Missing
YOR071C	NRT1	High-affinity nicotinamide riboside transporter	Missing	Missing	Missing	Missing
Allantoine Metabolism/Transport						
YIR028W	DAL4	Allantoin permease	Missing	Missing	Missing	Missing
Alcohol Metabolism/Fermentation						
YGL256W	ADH4	Alcohol dehydrogenase isoenzyme type IV	Missing	Missing	Missing	Missing
YOL165C	AAD15	Putative aryl-alcohol dehydrogenase	Missing	Missing	Missing	Present
Uracil Metabolism/Transport						
YBL042C	FUI1	High affinity uridine permease	Missing	Missing	Missing	Missing
YBR021W	FUR4	Uracil permease	Missing	Missing	Missing	Missing
Detoxification/Stress Response						
YER185W	PUG1	Plasma membrane protein with roles in the uptake of protoprophyrin IX and the efflux of heme	Missing	Missing	Missing	Missing
YGR213C	RTA1	Protein involved in 7-ami-cholesterol resistance	Missing	Missing	Missing	Missing
YGR234W	YHB1	Nitric oxide oxidoreductase	Missing	Missing	Missing	Missing
YIL053W	RHR2	DL-glycerol-3-phosphatase	Missing	Missing	Missing	Missing
YPR201W	ARR3	Arsenite transporter	Missing	Missing	Missing	Missing
YGL196W	DSD1	D-serine dehydratase (aka D-serine ammonia-lyase)	Missing	Missing	Missing	Missing
YHR044C	DOG1	2-deoxyglucose-6-phosphate phosphatase	Missing	Missing	Missing	Missing
YHR043C	DOG2	2-deoxyglucose-6-phosphate phosphatase	Missing	Missing	Missing	Missing
Chaperones						
YBR227C	MCX1	Mitochondrial matrix protein	Missing	Present	Missing	Missing
Proteases/Peptidases						
YHR132C	ECM14	Putative metalloprotease	Missing	Missing	Missing	Missing
Aromatic Aminoacid Metabolism						
YGL202W	ARO8	Aromatic ami-transferase I	Missing	Missing	Missing	Missing
YHR137W	ARO9	Aromatic ami-transferase II	Missing	Missing	Missing	Missing
Channels/Transporters						
YJL093C	TOK1	Outward-rectifier potassium channel	Missing	Missing	Missing	Missing
YBR296C	PHO89	Na+/Pi cotransporter	Missing	Missing	Missing	Present

Gene ID	Name	Description	*Gp*	*Gm*	*Gr*	*Ri*
YKL221W	MCH2	monocarboxylate permeases	Missing	Missing	Missing	Missing
Mating Type/Cell Cycle/Budding						
YIL140W	AXL2	Integral plasma membrane protein required for axial budding in haploid cells	Missing	Missing	Missing	Missing
ER Quality Control						
YPL096W	PNG1	Conserved peptide N-glycanase required for deglycosylation of misfolded glycoproteins during proteasome-dependent degradation	Missing	Missing	Missing	Missing
YBR015C	MNN2	Alpha-1,2-man-syltransferase	Missing	Missing	Missing	Missing
YJL186W	MNN5	Alpha-1,2-man-syltransferase	Missing	Missing	Missing	Missing
Others						
YLL057C	JLP1	Fe(II)-dependent sulfonate/alpha-ketoglutarate dioxygenase, involved in sulfonate catabolism for use as a sulfur source	Missing	Missing	Missing	Missing
YNL229C	URE2	Nitrogen catabolite repression transcriptional regulator	Contig_8671	Missing	Missing	Present
YLR047C	FRE8	Iron/copper reductases, involved in iron homeostasis	Missing	Missing	Missing	Missing
YLR278C	YLR278C	Zinc-cluster protein	Missing	Missing	Missing	Missing
YIL162W	SUC2	Invertase	Missing	Missing	Missing	Missing
YKL182W	FAS1	Fatty Acid Synthase	Missing	Missing	Missing	Missing
YPL231W	FAS2	Fatty Acid Synthase	Missing	Missing	Missing	Missing

Table S4: List of primers used in this work.

Name	Sequence (5'->3')	Information	Tm1 (nearest Neighbor)	Tm2 (nearest Neighbor)
HIGSP004	CTCCCAAACCAGCCATAATC	qPCR Ri NLS_98735 RV	50.3°C	
HIGSP006	GACGTGGAAAAGGCACCATA	qPCR Ri beta-tubulin RV	51.6°C	
HIGSP007	CGTTCAAGCAAGTTGTCGAG	qPCR Ri EX_25359 FW	53.5°C	
HIGSP008	AACCATTCTACCAGCCCAAG	qPCR Ri EX_25359 RV	51.4°C	
HIGSP009	AGCGGGTGGATGTGGTATAA	qPCR Ri EX_27056 FW	51.9°C	
HIGSP010	CCAGTGCCTGTTGATCTGAA	qPCR Ri EX_27056 RV	52.3°C	
HIGSP011	TGGTCTCGTCCTTTTGAAGC	qPCR Ri NLS_30765 FW	52.5°C	
HIGSP012	CACGATCAAGACATGCGAGA	qPCR Ri NLS_30765 RV	54.0°C	
HIGSP013	CGATATGAACTCATTGCCAAC	qPCR Ri NLS_334409 FW	50.6°C	
HIGSP014	TGGGGCTTAGGAACAGAGAA	qPCR Ri NLS_334409 RV	52.3°C	
HIGSP015	TGCCACCGAACCTTTACCTA	qPCR Ri RCP_335225 FW	52.2°C	
HIGSP016	TTGGACCAGAACCTACAGCA	qPCR Ri RCP_335225 RV	52.1°C	
HIGSP017	GAGGCACTGCTCAATTAAACG	qPCR Ri RCP_340423 FW	53.5°C	
HIGSP018	GCTGGCTGATATGCTTTTCC	qPCR Ri RCP_340423 RV	52.4°C	
HIGSP019	GATCCGCCAGGTGAATTATC	qPCR Ri NLS_343100 FW	50.5°C	
HIGSP020	GAAGGCAAAGAAGGTTCACG	qPCR Ri NLS_343100 RV	52.0°C	
HIGSP021	CACATGCAGCTAGTTGTGAGAG	qPCR Ri EX_54675 FW	55.9°C	
HIGSP022	AATCGCGTTCTCCTTTTTCC	qPCR Ri EX_54675 RV	51.6°C	
HIGSP023	GACATCGCACGAATATGTGG	qPCR RiTSL1/NLS_7749 FW	51.9°C	
HIGSP024	CGTTGGCACACTCCTGATAA	qPCR RiTSL1/NLS_7749 RV	52.4°C	
HIGSP025	TGGTAAATGTGGAGCGACAG	qPCR Ri RCP_84949 FW	52.4°C	
HIGSP026	ACATCCTGTACCGCAATGGT	qPCR Ri RCP_84949 RV	52.3°C	
HIGSP027	TACGTAATGTGGCAGGACCA	qPCR Ri RCP_230436 FW	52.1°C	
HIGSP028	TTACGTCGGGCTCACTGAAT	qPCR Ri RCP_230436 RV	53.5°C	
HIGSP029	CGCGAGTTGAAGTCGAAGA	qPCR Ri NLS_98735 FW	54.1°C	
HIGSP030	GCCATACCGCTCATATTGCT	qPCR Ri EF1-alpha FW	53.4°C	
HIGSP031	TTAACGATAGCGGCATCTCC	qPCR Ri EF1-alpha RV	53.3°C	
HIGSP032	CTCCAACTTATGGCGATCTCA	qPCR Ri beta-tubulin FW	53.4°C	
HIGSP033	ATGGTCTCCCACCGTTAGATTATGACG GTAC	Cloning RNAi target1 Ri NLS_7749 +BsaI CACC FW	40.7°C	61.6°C
HIGSP034	ATGGTCTCACCTTAAATGTACCTCATT CCTAG	Cloning RNAi target1 Ri NLS_7749 +BsaI CCTT RV	43.6°C	60.2°C
HIGSP035	ATGGTCTCTCACCTTCAAGTGCTTGGC AG	Cloning RNAi target2 Ri NLS_7749 +BsaI CACC FW	45.6°C	64.1°C
HIGSP036	ATGGTCTCGCCTTATCTTCATCTGTTC GATC	Cloning RNAi target2 Ri NLS_7749 +BsaI CCTT RV	43.8°C	63.1°C
HIGSP039	ATGGTCTCTCACCCAATCTTGTCAATT AGCTG	Cloning RNAi target2 Ri RCP_340423 +BsaI CACC FW	45.1°C	62.8°C
HIGSP040	ATGGTCTCGCCTTAACAGCAGCAATCA TAAC	Cloning RNAi target2 Ri RCP_340423 +BsaI CCTT RV	44.8°C	63.7°C
HIGSP041	ATGGTCTCACACCGTTTACACTTCTGC TTAATA	Cloning RNAi target1 Ri NLS_98735 +BsaI CACC FW	43.6°C	61.9°C
HIGSP042	ATGGTCTCACCTTCTAAAACGCGAATC GTG	Cloning RNAi target1 Ri NLS_98735 +BsaI CCTT RV	47.3°C	63.2°C
HIGSP043	AAGGTCTCTCACCAAGTATCGTAGTTA TTCATC	Cloning RNAi target2 Ri NLS_98735 +BsaI CACC FW	42.9°C	60.8°C

Name	Sequence (5'->3')	Information	Tm1 (nearest Neighbor)	Tm2 (nearest Neighbor)
HIGSP044	ATGGTCTCGCCTTCATCAATGTCATACCTTC	Cloning RNAi target2 Ri NLS_98735 +BsaI CCTT RV	41.4°C	62.3°C
HIGSP045	TGTTGTTGATGTGATTACAG	Sequencing RNAi vector pLjUbi FW	45.6°C	
HIGSP046	TCTCTCTTTGAAACCGTTG	Sequencing RNAi vector insert RV/ Test-PCR hairpin detection RV	48.0°C	
HIGSP047	TCGGAAGTGATAAAGTTATG	Sequencing RNAi vector insert FW/ Test-PCR hairpin detection	46.1°C	
HIGSP048	TTGCGGACTCTAGCATG	Sequencing RNAi vector 35S term RV	49.2°C	
HIGSP051	GAATAAAGGGGCCAAAATCG	qPCR LjPT4 3'_F2 (pers. Comm. Caroline Gutjahr)	49.6°C	
HIGSP052	GCTGTATCCTATCCCCATGC	qPCR LjPT4 3'_R2 (pers. Comm. Caroline Gutjahr)	51.8°C	
HIGSP053	TGGTTCAACTTTTCGTTCCA	qPCR LjAMT2.2 F (pers. Comm. Caroline Gutjahr)	49.3°C	
HIGSP054	CTTATCACCCTGACCCCAGA	qPCR LjAMT2.2 R (pers. Comm. Caroline Gutjahr)	51.8°C	
HIGSP055	CACGTTGTTAGGACCCCAAT	qPCR LjSbtM1 3'_F (pers. Comm. Caroline Gutjahr)	50.8°C	
HIGSP056	TTGAGCAGCACCCTCTCTATC	qPCR LjSbtM1 3'_R (pers. Comm. Caroline Gutjahr)	55.0°C	
HIGSP057	TCATCTGTCCTTGGGGTCAT	qPCR LjBcp1_F (pers. Comm. Caroline Gutjahr)	51.2°C	
HIGSP058	CAGCTGCAGAAGTTGCATTT	qPCR LjBcp1_R (pers. Comm. Caroline Gutjahr)	53.3°C	
HIGSP059	ATGCAGATCTTCGTCAAGACCTTG	qPCR LjUbi F (pers. Comm. Caroline Gutjahr)	57.4°C	
HIGSP060	ACCTCCCCTCAGACGAAG	qPCR LjUbi R (pers. Comm. Caroline Gutjahr)	51.5°C	
HIGSP061	CAGGTCTCACACCATGTCACCGCCTTT	Cloning RiTSL1+SP part1 +BsaI CACC FW	40.0°C	62.4°C
HIGSP065	CGGGTCTCGGTTTCAAAAGCCAACATC	Cloning RiTSL1 part1 +BsaI GTTT RV	43.2°C	61.1°C
HIGSP066	CGGGTCTCCAAACTTTCCTCACGATG	Cloning RiTSL1 part2 +BsaI AAAC FW	43.3°C	60.6°C
HIGSP068	AGGGTCTCTCACCATGTTTGATATTGGTAAAGACA	Cloning RiTSL1-SP part1 +BsaI CACC FW	40.9°C	62.4°C
HIGSP081	GCGGTCTCACCTTATCGTTACTAGCTAATATT	Cloning RiTSL1 -STOP codon part2 +BsaI CCTT RV	41.4°C	61.8°C
HIGSP096	TGCGGATTTCTTACATGAGC	qPCR Ri NLS_26232 FW	51.6°C	
HIGSP097	GGTTCGTTACTGGATCTTGAGTG	qPCR Ri NLS_26232 RV	54.1°C	
HIGSP098	CCTATAGCGTTGTTTCAAGACG	qPCR Ri NLS_320155 FW	54.0°C	
HIGSP099	CTTTCGAGAAAATCCAATGTCC	qPCR Ri NLS_320155 RV	51.1°C	
HIGSP100	TCCGGTCATAGTTTGTGGTG	qPCR Ri NLS_32853 FW	51.0°C	
HIGSP101	TGAACTTCGACCATCATAGCC	qPCR Ri NLS_32853 RV	52.7°C	
HIGSP102	GGTGATGCAAAAGGAGGAAC	qPCR Ri RCP+NLS_349824 FW	50.4°C	
HIGSP103	GCTGGTTGACCATTTGTTCG	qPCR Ri RCP+NLS_349824 RV	51.7°C	
HIGSP104	GATCAAGGCACGTGATGATG	qPCR Ri RCP_349288 FW	51.8°C	
HIGSP105	CGGCAGGCTTCGATATTAAC	qPCR Ri RCP_349288 RV	52.7°C	
HIGSP106	TGTCCCGTCACATCTTTCAG	qPCR Ri SRC_343180 FW	51.8°C	
HIGSP107	CATGGAACAGCCCAAAACAG	qPCR Ri SRC_343180 RV	51.5°C	
HIGSP108	CTTAAACGTGGCACTTCCATC	qPCR Ri SCR_339199 FW	52.1°C	
HIGSP109	TTGCGTGCTTGGAAAAGAC	qPCR Ri SCR_339199 RV	51.2°C	
HIGSP142	ATGGTCTCACACCCCACGACACCATTAA	Cloning RNAi target1 Ri SCR+RCP_84949 +BsaI CACC FW	39.2°C	60.5°C
HIGSP143	ATGGTCTCACCTTGGACCGGTAATATCAG	Cloning RNAi target1 Ri SCR+RCP_84949 +BsaI CCTT RV	40.6°C	60.6°C
HIGSP144	ATGGTCTCACACCCAAGAGGAATTCGC	Cloning RNAi target1 Ri SCR_339199 +BsaI CACC FW	38°C	61.1°C

Name	Sequence (5'->3')	Information	Tm1 (nearest Neighbor)	Tm2 (nearest Neighbor)
HIGSP145	ATGGTCTCACCTTGGAAGTGCCACGT	Cloning RNAi target1 Ri SCR_339199 +BsaI CCTT RV	39.0°C	61.3°C
HIGSP146	ATGGTCTCACACCATGACCACCACTGCC	Cloning ScTPS2 +BsaI CACC FW	43.3°C	62.5°C
HIGSP147	ATGGTCTCACCTTAACCTTTGCGCCGG	Cloning ScTPS2 +BsaI CCTT RV	46.6°C	63.3°C
HIGSP148	TAGAAGACTATACGGGTCTCACACCATGACTACGGATAACG	Cloning ScTPS1 part1 +BsaI CACC FW	41.1°C	67.3°C
HIGSP151	TAGAAGACTATACGGGTCTCACCTTGTTTTTGGTGGCAGAG	Cloning ScTPS1 part2 +BsaI CCTT RV	42.8°C	67.8°C
HIGSP156	TAGAAGACTATACGGGTCTCACACCATGGTTTCTTCGGCA	Cloning RiTPS2 +BsaI CACC FW	42.1°C	68.1°C
HIGSP160	ATGGTCTCACACCTGATGCACTCAAGC	Cloning RNAi target1 Ri RCP_349288 +BsaI CACC FW	38.8°C	61.9°C
HIGSP161	ATGGTCTCACCTTAGCATCACGATCAG	Cloning RNAi target1 Ri RCP_349288 +BsaI CCTT RV	38.9°C	60.6°C
HIGSP162	ATGGTCTCACACCAAAATGGTTGTCCAAGG	Cloning RNAi target1 Ri SCR_343180 +BsaI CACC FW	42.1°C	60.9°C
HIGSP163	ATGGTCTCACCTTAAACAGCACCCGTTT	Cloning RNAi target1 Ri SCR_343180 +BsaI CCTT RV	42.1°C	60.5°C
HIGSP164	ATGGTCTCACACCCGTAGAGAATCAATCAA	Cloning RNAi target1 Ri RCP_230436 +BsaI CACC FW	40.6°C	60.9°C
HIGSP165	ATGGTCTCACCTTGCATCAACTATACTAGC	Cloning RNAi target1 Ri RCP_230436 +BsaI CCTT RV	40.5°C	60.9°C
HIGSP174	GGGTCTCACCTTTTCGTTCAGCTTCTG	Cloning RiTPS2 +BsaI CCTT RV	41.8°C	61.2°C
HIGSP179	CGGGTCTCGGAcACCAAACAAACATC	Cloning ScTPS1 part1 +BsaI mut RV	40.3°C	59.6°C
HIGSP180	GCGGTCTCCTgTCGTCCACCCG	Cloning ScTPS1 part2 +BsaI mut FW	42.0°C	61.8°C
HIGSP181	CAAGAGATGGTATGAATCTCGTGTC	Sequencing RTPS1 FW	55.6°C	
HIGSP182	TCCATCATGCCCCATGGCCTAG	Sequencing ScTPS2 FW	58.1°C	
HIGSP183	TGGAATAGCTGCAGCTGGGTC	Sequencing ScTPS2 RV	57.2°C	
HIGSP184	AAGACTCCCCGATGCTGCG	Sequencing RTPS2 FW	57.1°C	
HIGSP185	AGTTAGATCACGCCCAGAAACG	Sequencing RTPS2 RV	56.0°C	
HIGSP186	CGTTCCAGATCATGATTTTGAAGG	Sequencing RTSL1 FW	54.5°C	
HIGSP187	AACTATCGGACGTTCTTCCGC	Sequencing RTSL1 RV	55.4°C	
HIGSP188	TTGGAGGTCAGACCAACATC	qPCR RiTPS2 FW	50.3°C	
HIGSP189	ATCGGTTCGATCATCTCCAG	qPCR RiTPS2 RV	52.5°C	
HIGSP190	CCGGAATTTGGTTCATGG	qPCR RiTPS1FW	47.3°C	
HIGSP191	GACTTATCAACCGCAGATGG	qPCR RiTPS1 RV	51.3°C	
HIGSP192	CGGGTCTCACACCATGTCGCAATCTGTTAAC	Cloning RiTPS1 +BsaI CACC FW	44.5°C	63.9°C
HIGSP193	CGGGTCTCACCTTTAATTCTTTAGCTAATCTTTC	Cloning RiTPS1 -STOP codon +BsaI CCTT RV	43.4°C	61.7°C
HIGSP194	ACTGATCATATTGTGGGTGATGAC	Test-PCR DNAse treatment Lj Bcp1 FW	52.9°C	
HIGSP195	ATGCAAAGAAAGCTGTGATGAC	Test-PCR DNAse treatment Lj Bcp1 RV	53.1°C	
HIGSP196	CCGTTAGATTATGACGGTAC	Test-PCR hairpin detection in RiTSL1 target FW	48.2°C	
HIGSP197	ACAATCTCGATTCTCATTTTCTTTG	Cloning Sctps2Δ::KanMX cassette FW	53.3°C	
HIGSP198	AGAACAAGGAACAAAGTCCAAGC	Cloning Sctps1Δ::KanMX cassette FW	54.1°C	
HIGSP199	CTGCAGCGAGGAGCCGTAAT	KanMX confirmation primer RV	58.7°C	
HIGSP200	GGGAGAGAAAGAAAGAGAGAGAAAA	Cloning Sctps1Δ::KanMX cassette RV	55.1°C	
HIGSP201	GTAGTACCCTCTTTTACCTACCGCT	Cloning Sctps2Δ::KanMX cassette RV	56.5°C	
HIGSP213	TAACACCTAACCTCGATAGAGTTGC	Cloning Sctps3Δ::KanMX cassette FW	56.0°C	
HIGSP214	ACCACCTTTAGTGTTTTTCTTACCC	Cloning Sctps3Δ::KanMX cassette RV	53.2°C	

Name	Sequence (5'->3')	Information	Tm1 (nearest Neighbor)	Tm2 (nearest Neighbor)
HIGSP215	CATAGTCGAGCGGTCCGTTCTG	Confirmation Sctps1Δ::KanMX FW	59.0°C	
HIGSP216	CTTCTCGTTCTCTACGTAAGGAACTTG	Confirmation Sctps2Δ::KanMX FW	58.7°C	
HIGSP217	GGGTAGCCGTGCCTCGTCTG	Confirmation Sctsl1Δ::KanMX FW	58.7°C	
HIGSP218	GCAACGCAACGTACCCTGGGAC	Confirmation Sctps3Δ::KanMX FW	58.7°C	
HIGSP219	CGAGAAAAGCGGGTCACCCCCGCCCCTGCATTTTGATATGGCGTATTTGGCCTCGAGGAGAACTTCTAGTATATCCAC	Cloning SctslΔLEU2 cassette FW not underlined: LEU2 complementary binding site; underlined: ScTSL1 recombination site	58.8°C	78.4°C
HIGSP220	TTAGAATATTGGGCTTGAAGGTTATCTTACTATTTTCGCTGTTCATCGACTCGACTACGTCGTAAGGCCGTTTCT	Cloning SctslΔLEU2 cassette RV not underlined: LEU2 complementary binding site; underlined: ScTSL1 recombination site	60.9°C	75.3°C
HIGSP231	CTGCGGTCAAGATATTTCTTGAATCAG	Confirmation LEU2 RV	58.3°C	
HIGSP234	GCGAATTCATGTCACCGCCTTTGAGTATTCATAGAGTTGTAGTTGTTTC	Cloning full length RiTSL1 SP underlined1: EcoRI RS underlined2: complementary overlap with HIGSP235	35.9°C	69.2°C
HIGSP235	TTCTCGAGAGAAGCAGTATACGGAAGAAAAAGAGAAACAACTACAACTC	Cloning full length RiTSL1 SP underlined1: XhoI RS underlined2: complementary overlap with HIGSP234	35.9°C	69.1°C
HIGSP246	CCCAGCCATGTTCTCAAATG	qPCR RirG193020 Grx4 FW	51.5°C	
HIGSP247	AGTTCGGGGAATTTCTCAGC	qPCR RirG193020 Grx4 RV	52.3°C	
HIGSP248	GAAGCGAGACTTTGGAATGAC	qPCR RirG173060 Malate Synthase FW	52.4°C	
HIGSP249	CGGCAAGGATGGTTTCTATC	qPCR RirG173060 Malate Synthase RV	51.2°C	
HIGSP282	CGCCATTTGACCATTCATCCCCAGTTGACATTTCC	Cloning probe "GpRIC1" (with BKRSV overhang) FW	46.5°C	65.2°C
HIGSP283	CTCGTAGACTGCGTACCAGGAAGTTCTAATGGTGATTC	Cloning probe "GpRIC1" (with LUF overhang) RV	45.6°C	66.8°C
HIGSP284	CGCCATTTGACCATTCA	Cloning PCR on BKRSV overhang	46.3°C	
HIGSP285	CTCGTAGACTGCGTACCA	Cloning PCR on LUF overhang	51.0°C	
HIGSP286	TCTCTCTCTTGCTACTTC	Sequencing GpRIC1 5' -->	46.0°C	
HIGSP287	TTTTTAGGTTGCACTTCAC	Sequencing GpRIC1 <--3'	45.3°C	
HIGSP288	ATTTCAACTACTAGATACTC	Sequencing GpRIC1 3'-->	42.1°C	
HIGSP289	AAGCTTGACGTTTCTCCG	Sequencing GpRIC1 <--5'	50.5°C	
HIGSP290	GTTTCCAACGTTGTTACCG	Cloning probe "GpRIC2" FW	48.3°C	
HIGSP291	GTGACTTCAATGGCAACG	Cloning probe "GpRIC2" RV	48.2°C	
HIGSP292	CGCCATTTGACCATTCAGTTTCCAACGTTGTTACCG	Cloning probe "GpRIC2" (with BKRSV overhang) FW	48.3°C	66.0°C
HIGSP293	CTCGTAGACTGCGTACCAGTGACTTCAATGGCAACG	Cloning probe "GpRIC2" (with LUF overhang) RV	48.2°C	68.7°C
HIGSP294	CCGTCCCTATTATTGTTGCC	Cloning probe "GpRIC3" FW	50.4°C	
HIGSP295	GGCTGGTGGTGGAAACG	Cloning probe "GpRIC3" RV	50.2°C	
HIGSP296	CGCCATTTGACCATTCACCGTCCCTATTATTGTTGCC	Cloning probe "GpRIC3" (with BKRSV overhang) FW	50.4°C	66.6°C
HIGSP297	CTCGTAGACTGCGTACCAGGCTGGTGGTGGAAACG	Cloning probe "GpRIC3" (with LUF overhang) RV	50.2°C	69.7°C
HIGSP298	GGCATTTTCAACACCATCTA	Sequencing GpRIC2 3' -->	47.8°C	
HIGSP299	TAGATGGTGTTGAAAATGCC	Sequencing GpRIC2 <-- 3'	47.8°C	
HIGSP300	CGGGAAGAGACAGGAGGAA	Sequencing GpRIC3 3' -->	52.6°C	
HIGSP252	TAATACGACTCACTATAGGGGTTAGATTATGACGGTAC	PIGS RiTsl1-RNAiT1 +(underlined) T7 promoter sequence FW	40.7°C	62.7°C

Name	Sequence (5'->3')	Information	Tm1 (nearest Neighbor)	Tm2 (nearest Neighbor)
HIGSP253	TAATACGACTCACTATAGGGCTCATTCCTAGATTAG	PIGS RiTsl1-RNAiT1 +(underlined) T7 promoter sequence RV	35.6°C	63.1°C
HIGSP254	TAATACGACTCACTATAGGGGCAGTGCTTCAGCCGCTAC	PIGS GFP-RNAiT +(underlined) T7 promoter sequence FW	56.8°C	70°C
HIGSP255	TAATACGACTCACTATAGGGGAAGTCGATGCCCTTCAGCT	PIGS GFP-RNAiT +(underlined) T7 promoter sequence RV	55.2°C	68.9°C
GGP5	GCTCAACACATGAGCGAAACC	Confirmation hairpin cDNA RV (pers. comm. Andreas Binder)	54.7°C	
G114	AAACACAGATTATCATCACTAATTGGA	Confirmation hairpin cDNA FW (pers. comm. Andreas Binder)	53.2°C	

Table S5: List of vectors created in this work. Kan: Kanamycin, Amp: Ampicillin.

Name	Insert	Backbone	Resistance	Reference
HIGSV001	F1-2 pLjUbi:RiTSL1_RNAi_target1>F2-3 ins>F3-4 dy>F4-5 ins>F5-6 p35S:GFP	LIII fin - BB52	Kan	This study
HIGSV002	F1-2 pLjUbi:RiTSL1_RNAi_target2>F2-3 ins>F3-4 dy>F4-5 ins>F5-6 p35S:GFP	LIII fin - BB52	Kan	This study
HIGSV004	F1-2 pLjUbi:Ri_340423_RNA_target2>F2-3 ins>F3-4 dy>F4-5 ins>F5-6 p35S:GFP	LIII fin - BB52	Kan	This study
HIGSV005	F1-2 pLjUbi:Ri_98735_RNAi_target1>F2-3 ins>F3-4 dy>F4-5 ins>F5-6 p35S:GFP	LIII fin - BB52	Kan	This study
HIGSV006	F1-2 pLjUbi:Ri_98735_RNAi_target2>F2-3 ins>F3-4 dy>F4-5 ins>F5-6 p35S:GFP	LIII fin - BB52	Kan	This study
HIGSV007	F1-2 dy>F2-3 ins>F3-4 dy>F4-5 ins>F5-6 p35S:GFP	LIII fin - BB52	Kan	This study
HIGSV008	F 1-2: A-B pLjUbi>B-E Esp3I dy>E-F 35S Term>F-G dy	LIIc F 1-2 - BB30	Amp	This study
HIGSV009	R 3-4: A-C p35S>C-D GFP>D-E dy>E-F nos Term>F-G dy	LIIc R 3-4 - BB34	Amp	This study
HIGSV015	F1-2 dy>F2-3 dy>F3-4 dy>F4-5 ins>R5-6 p35S:Cerulean:35S Term	LIII fin - BB52	Kan	This study
HIGSV031	F1-2 pUbi:RiTSL1+SP-STOP-mCherry>F2-3 ins >F3-4 dy>F4-5 ins>R5-6 p35S:Cerulean:35S Term	LIII fin - BB52	Kan	This study
HIGSV032	F1-2 pLjUbi:RiTSL1ΔSP-STOP-mCherry:nos Term>F2-3 ins>F3-4 dy>F4-5 ins>R5-6 p35S:Cerulean:35S Term	LIII fin - BB52	Kan	This study
HIGSV054	F1-2 pLjUbi:Ri339199_RNAi_target1> R3-4 35S:GFP	HIGSV081	Kan	This study
HIGSV055	F1-2 pLjUbi:Ri349288_RNAi_target1 > R3-4 35S:GFP	HIGSV081	Kan	This study
HIGSV056	F1-2 pLjUbi:Ri230436_RNAi_target1 > R3-4 35S:GFP	HIGSV081	Kan	This study
HIGSV060	F1-2 pLjUbi:Ri84949_RNAi_target1 > R3-4 35S:GFP	HIGSV081	Kan	This study
HIGSV061	F1-2 pLjUbi:Ri343180_RNAi_target1 > R3-4 35S:GFP	HIGSV081	Kan	This study
HIGSV080/ pre RNAi	F1-2 pLjUbi:Esp3Idy:35S Term>F2-3 ins>R3-4 p35S:GFP:nos Term>F4-6 dy	LIII fin - BB52	Kan	This study
HIGSV081/ Fin. silencing	F1-2 pLjUbi:ccdb:35S Term>F2-3 ins>R3-4 p35S:GFP:nos Term>F4-6 dy	HIGSV080	Kan	This study
HIGSV095	RiTSL1+SP	426-GPD-ccdB-HA	Amp/URA3	This study
HIGSV096	RiTSL1ΔSP	426-GPD-ccdB-HA	Amp/URA3	This study
HIGSV097	RiRiTPS1	426-GPD-ccdB-HA	Amp/URA3	This study
HIGSV098	RiTPS2	426-GPD-ccdB-HA	Amp/URA3	This study
HIGSV099	ScTPS1	426-GPD-ccdB-HA	Amp/URA3	This study
HIGSV100	ScTPS2	426-GPD-ccdB-HA	Amp/URA3	This study
HIGSV101	dy	426-GPD-ccdB-HA	Amp/URA3	This study
E006	ScInv+SP	pSUC2T7M13ORI	Amp/TRP1	pers. comm. K. Sedzielewska Toro
E007	ScInvΔSP	pSUC2T7M13ORI	Amp/TRP1	pers. comm. K. Sedzielewska Toro
E015	SP(Rc-7749)	pSUC2T7M13ORI	Amp/TRP1	pers. comm. K. Sedzielewska Toro
HIGSV132	SP(Ri-7749)	pSUC2T7M13ORI	Amp/TRP1	This study

6.1 List of figures

6.2 List of tables

6.3 Curriculum vitae