UNIVERSITY OF LIVERPOOL

Diversity and Roles of Mycorrhizal Fungi in the Bee Orchid *Ophrys apifera*

By

Wazeera Rashid Abdullah

April 2018

A Thesis submitted to the University of Liverpool in fulfilment of the requirement for the degree of Doctor in Philosophy

Table of Contents **Page No.**

List of Tables

List of Figures

Acknowledgements

I would like to express my honest thanks to my supervisor Dr. Meriel Jones for her incredible support and I am extremely grateful for her valuable concern, extreme patience and advice about this thesis. I would also like to thank my second supervisor Dr. James Hartwell for his useful advice.

Thanks to both my assessors Dr. Heather Allison and Dr. Raj Whitlock for their advice and support. I would also like to thanks Dr. Raj Whitlock for the advice he gave me about how to extract RNA from plants. Special thanks to our wonderful technician Jean Wood for her technical assistance in Lab G. I am also extremely grateful to Dr. Luca Lenzi Centre of Genomic Research, University of Liverpool, for helping me to learn basic bioinformatics using command line in Linux for analysis DNA data.

Thanks to Mr Mathew Gemmell for analysing RNA-seq, and also Dr. Sam Haldenby who put me in the right way for analysing data in the beginning, both members of the team in Centre of Genomic Research, University of Liverpool.

Thanks to my friends in Lab G Aleya, Nirja, Ezwan, Amir, Hanaa, Eiman, Batool, Stefany, and all other colleagues in the lab and out of lab who helped me and made my work more enjoyable.

I wish to acknowledge the financial support from my sponsor HCDP-Kurdistan region/ Iraq (Ministry of Higher Education and Scientific Research).

Finally, thanks to my lovely Mum and I would to say, you are everything in my life and dad (I really miss you). Thanks to all my sisters, brothers, nephews and nieces who really provide their love, encouragement and help.

Abbreviations

AGs	Anastomosis groups
AM	Arbuscular mycorrhiza
ANOVA	Analysis of Variance
ATP	Adenosine triphosphate
Basionym	The original name on which a new name is based
BLAST	Basic Local Alignment Search Tool
bp	base pair
cDNA	Complementary DNA
CGR	Centre of Genomic Research
CMA	Corn Meal Agar
COMA	Coconut Milk Agar
CTAB	Hexadecyltrimethylammonium bromide
2D	Two-dimensional
DE	Differentially expressed
DE	Discovery Environment (CyVerse iPlant)
FDR	False Discovery Rate
FIM	Fungal Isolation Medium
FPKM	Fragments mapped per kilobase of exon per million reads mapped
HTS	High-Throughput sequencing
IAA	Indoleacetic Acid
ID	Identification
ITS	Internal Transcribed Spacer
kPa	kilopascal
Kruskal-Wallis	Or H test (named after William Kruskal and W. Allen Wallis).
LB	Luria broth
M	Molar
MEA	Malt Extract Agar
MEGA	Molecular Evolutionary Genetics Analysis
MSA	Multiple Sequence Alignment
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
OM	Orchid culture medium
OMA	Oat Meal Agar
OTU	Operational Taxonomic Unit
PacBio	Pacific Biosciences
PCA	Principal Component Analysis
PCoA	Principal Coordinates Analysis
PDA	Potato Dextrose Agar
PLB	Protocorm-like body
QIIME	Quantitative Insight Into Microbial Ecology
qPCR	quantitative polymerase chain reaction
RIN	RNA integrity number
rRNA	Ribosomal RNA
SNP	Single-nucleotide polymorphism
SOAP	Short Oligonucleotide Analysis Package
SOLID	Sequencing by Oligonucleotide Ligation and Detection
WA	Water Agar
www	World Wide Web
UV	Ultraviolet

ABSTRACT

Mycorrhiza is a crucial symbiotic association between the roots of plants and fungi within the soil environment. The specificity in these partnerships, particularly of the fungal partner is still poorly understood. Orchids require symbiotic fungi in a mycorrhizal association for seed germination and establishment, typically species within the *Tulasnellaceae* and *Ceratobasidiaceae*. The distribution of suitable mycorrhizal fungi will therefore affect the distribution of orchid plants. The bee orchid (*Ophrys apifera*) is considered common in southern England but is less frequent in the north. It has a conspicuous flower spike with individual flowers resembling bees. The species is surprisingly common in urban reclaimed ground where rubble provides suitable well-drained, alkaline and low nutrient conditions.

Aboveground organs of *Ophrys apifer*a were recorded over a 3-year period to understand the life history better. The basal leaf rosette starting appeared in September, flower shoot initiated at the end of April, and flowers opened in May and all organs above ground died in summer (July and August). There was a positive correlation between leaf length and leaf area, and there is no significant correlations between maximum number of leaves and maximum number of flowers. There was no obvious effect of temperature and rainfall, possibly due to stability of the weather during this period.

Root and soil samples were analysed to identify potential mycorrhiza fungi. Culture-based techniques provided limited information, although this included one culture of a likely mycorrhizal genus. Samples from three sites of Liverpool University campus were investigated using Illumina amplicon sequencing of the ITS region. DNA of *Tulasnellaceae* and *Ceratobasidaceae* were detected in the soil as anticipated. In addition *Sebacinaceae, Thelephoraceae, Clavariaceae, Psathyrellaceae* were also detected in the soil samples. There were significant differences between the fungal communities of the three sites.

A transcriptome analysis to provide insight into the orchid physiology was undertaken using protocorms derived from seed germinated *in vitro* and tissues of root and corm from the field. Determination of differentially expressed genes/transcripts showed specific ones were up-regulated in protocorm (e.g. related to auxin transport, carbohydrate metabolism) while others, (e.g. related to transport processes) were detected up-regulated in corm and root. These outcomes of these studies on *Basidiomycota* fungi and the *Ophrys apifera* host provide significant insight into the orchid mycorrhizal relationship.

1 Chapter one: Literature review:

1.1 Mycorrhiza:

Mycorrhiza is a crucial symbiotic association between the roots of plants and fungi within the soil environment. The term mycorrhiza (literally ‘fungus-root’) was originated by Frank (1885), who pointed out that these associations between fungi and plants are required for the nutrition of both partners. Most plant species are capable of this relationship with fungi from the *Glomeromycota, Ascomycota* or *Basidiomycota*. In general, mycorrhiza is a "balanced" mutualistic association between fungi and plants with a mutually beneficial exchange for growth and survival of both partners (Brundrett, 2004).

Arbuscular mycorrhiza and ectomycorrhiza are the most common types and are mutualistic involving over 90% of angiosperm species (Rasmussen and Rasmussen, 2009). Orchids, however, have developed their own mycorrhizal form, which can be parasitic to the fungal partner. Some of these relationships are obligate, for either fungal or plant partner, which has been a barrier to study (Ogura-Tsujita et al., 2012). The fungi reinforce the plant's ability to obtain nutrients from the soil, particularly phosphorus, nitrogen, and water while absorbing carbohydrates that the plant produces in exchange. The plant in the relationship delivers carbon compounds as soluble sugars and demand from the fungal partner can be met by increased photosynthetic productivity of the plant.

In addition, some mycorrhizal fungi have the enzyme capacity to obtain carbon resources from the soil or via cell wall-degrading activity (Nehls, 2008). Mycorrhizal fungi enhance their partners’ acquisition of phosphorous from different soil pools because they have the power to access phosphorous in its diverse forms (Cairney, 2011). Generally, both the plants and the fungi profit from these interactions. At least seven specific types of mycorrhizal relationship have been characterized based on the distinctive structural and physiological characteristics of the plant-fungal partners involved (Table 1.1) (Smith and Read, 2008). They are considered ectomycorrhiza if the fungal hyphae grow around the plant root and between the plant cells, and endomycorrhiza if the fungus penetrates the plant cells.

Table 1. 1: Characteristics of the mycorrhizal types (Smith and Read, 2008).

Note:- (*) Its means all orchids are achlorophyllous in the early seedling stage and most orchid species are green as adults.

(?+) Entries in brackets indicate rare conditions. - = absent, += present, + or - = present or absent.

	Types of mycorrhiza						
Fungi	AM	Ectomycorrhiza	Ectendomycorrhiza	Arbutoid	Monotropoid	Ericoid	Orchid
Septate	–	+	+	+	+	+	+
Aseptate	+	–	–	–	–	–	–
Intracellular colonization	+	–	+	+	+	+	+
Fungal sheath	–	+	+ or –	+ or –	+	–	–
Hartig net	–	+	+	+	+	–	–
Vesicles	+ or –	–	–	–	–	–	–
Achlorophyllous	–(?+)	–	–	–	+	–	+(*)
Fungi	*Glomeromycota*	Most *Basidiomycota*, but some *Ascomycota* and *Glomeromycota*		*Basidiomycota*		*Ascomycota*	*Basidiomycota*
Plant host	Vascular and non-vascular plants	*Gymnosperm*	Vascular plants	*Ericaceae*	*Monotropoideae*	*Ericaceae* *Bryophyta*	*Orchidaceae*

The specificity in these partnerships, particularly of the fungal partner, is one area where there are large gaps in knowledge, and where molecular ecology methods can be applied. The control of the symbiosis, so that it remains mutualistic rather than one partner becoming parasitic, is another area that is only partially understood and where a greater understanding of the physiology is needed (Bailarote et al., 2012).

1.1.1Arbuscular mycorrhiza (AM) or Vesicular-arbuscular mycorrhiza (VAM):
AM is the most widespread underground symbiosis and is an endomycorrhiza. It is formed by 70-90% of land plant species. The name AM come from characteristic structures namely arbuscules, (literally 'bush or 'little tree'). These structures formed inside the inner cortex cells, and are the main site of exchange nutrient between the partners (Parniske, 2008).

The *Glomeromycota* has 284 living species. It forms the mycorrhiza called arbuscular mycorrhiza (AM) which is widely distributed in the plant kingdom. All AM are obligate symbionts and are not able to live alone without plants. They rely on a living photoautotrophic partner to produce the next generation of spores (Parniske, 2008).

1.1.2 Ectomycorrhiza:
This is considered the most advanced symbiotic association between higher plants and fungi. It is present in around 3% of seed plants, especially forest trees. The fungi form a sheath around the root system which is around 100 μm thick. Fungal hyphae penetrate the outer cell layers of the root producing the Hartig net (Moore et al., 2011) .

The fungi involved in this association are *Basidiomycota* and *Ascomycota*. The Hartig net is a hyphal network which extends into the root, penetrating between the cells of the root to form a complex intercellular system, and this network is a place of nutrient exchange between the partners (Smith and Read, 1997). Ectomycorrhiza can link groups of trees by submerged mycelium acting as a wood-wide-web. The fungi depend on the plant to supply carbon, and most of these fungi are saprotrophic but cannot use cellulose or lignin. It can provide mineral ions to the plant, such as phosphate and ammonium ions (Moore et al., 2011).

1.1.3 Ectendomycorrhiza:

The term ectendomycorrhiza is used especially for mycorrhizal roots which show features of both endomycorrhiza and ectomycorrhiza, based entirely on visible structures (Moore et al., 2011). The family *Pinus* is one of the few where this type occurs.

There are therefore many of the same characteristics between ectendomycorrhiza and endomycorrhiza, including extensive intracellular penetration. In appearance in *Pinus* roots there is a Hartig net with fungal hyphae between the plant cortical cells, but fungal structures are also found within living plant cells. The association is stable over many months, and ectendomycorrhiza stimulate the growth of short roots, same as ectomycorrhiza association (Moore et al., 2011).

1.1.4 Ericoid mycorrhiza, Arbutoid mycorrhiza, and Monotropoid mycorrhiza:

They are all variants around a theme. The plant family Ericaceae has a symbiotic association with *Ascomycota* and *Basidiomycota* to form ericoid mycorrhiza, but with *Basidiomycota* to form arbutoid mycorrhiza.

A characteristic structural feature in ericoid mycorrhiza is a loose network of hyphae covering the exterior of the fine plant roots and fungal coils within the epidermal cells. The hyphae are capable of penetrating the epidermal cells at several points, and then these cells become filled with coils of hyphae. The plant membrane separates the intracellular fungal symbiont from the plant cytoplasm. The nutrient exchange occurs within these coils (Moore et al., 2011). The main diagnostic features in arbutoid mycorrhiza are the presence of both a Hartig net and intracellular coils within a root surrounded by a fungal sheath.

The sub-family Monotropoideae within the Ericaceae (previously known as the family *Monotropaceae*) involve a small number of genera of exclusively achlorophyllous plants. The seeds of *Monotropa* are small and cannot germinate without fungi because they need an external supply of organic compounds (Smith and Read, 1997) as well as requiring a fungal symbiont throughout the rest of their lives. The fungi themselves obtain carbon from mycorrhizal association with other, photosynthetic, plants. The fungal sheath that surrounds monotropoid roots are multilayered and sometimes squeeze the sheath of fungal hyphae between the boundaries in which the layers are demarcated by tannin deposits. A Hartig net of

hyphae surround the root tip and do not penetrate the cortex, although single hyphae grow into epidermal cells to facilitate nutrient transfer (Smith and Read, 1997).

1.1.5 Orchid mycorrhiza:

For over 100 years, the members of the family *Orchidaceae* have been known to be mycorrhizal (Smith and Read, 1997) in association with members of the *Basidiomycota*. Orchid seeds are very small and do not have sufficient nutrients to support germination. Therefore the early seedling stage is obligatately mycorrhizal and in the natural environment the seeds cannot germinate without a suitable fungus. However, in the laboratory seedlings can be germinated *in vitro* through using artificial media with or without fungi, and this technique is widely used for propagation of orchids (Moore et al., 2011).

All orchids therefore have a non-photosynthetic stage in seedling development, depending on a supply of carbohydrates which is provided by mycorrhizal fungi under natural conditions. Most orchid plants are chlorophyllous in the adult stages so show only partial myco-heterotropy where carbon flows from the fungus to the plant (Taylor et al., 2002). However, around 200 species remain achlorophyllous, and are thus fully myco-heterotrophic throughout their life-cycle (Haselwandter et al., 2006). In some situations and specific species, the plant may therefore be parasitic on the fungus, such as the obligate myco-heterotrophic orchid *Rhizanthella gardneri* that lacks the ability to photosynthesize (Bougoure et al., 2010).

Myco-heterotrophic orchid plants have the unusual mycorrhizal association with fungal partners in their nutrient. They have been investigated that fungal symbionts with orchid *Cymbidium* have been shifted their association according to the evolution of nutritional mode in plant partners. For example, molecular identification of mycorrhizal fungi showed that fungal partners can shift from saprotrophic to ectomycorrhizal. Thus, often the autotroph is dependent on saprobic *Tulasnellaceae*, the mixotrophs related to the *Tulasnellaceae* and ectomycorrhizal groups including the *Sebacinales, Russulaceae, Thelephoraceae,* and *Clavulinaceae,* and the myco-heterotrophs are specifically associated with ectomycorrhizal *Sebacinales*. This is correlated to evolution of nutritional mode in a plant from autotroph which depends on saprotrophic fungi (*Tulasnellaceae*) to myco-heterotrophic via mixotrophic which rely on both *Tulasnellaceae* and ectomycorrhizal. This gradual shift of the fungal

mycorrhiza during a phase of coexistence may have a significant impact on the evolution of mycoheterotrophic plants (Ogura-Tsujita et al., 2012).

1.1.5.1 Orchid mycorrhizal interaction:

Establishment of a mycorrhizal association is essential for seed germination and seedling establishment. The infection of an orchid seed by fungi will happen after the embryo takes up water and swells, breaking the seed coat. Development is limited to an intermediate stage called a protocorm and formation of a few epidermal hairs before the growth stops in the absence of fungi (Yam and Arditti, 2009).

In nature the protocorm will only develop further if mycorrhizal fungi can colonise the epidermal hairs, and spread hyphae from cell to cell forming a dense region of cortical cells containing coils of hyphae known as pelotons (Smith and Read, 2008). In this way, the peloton increases the surface area between the orchid and fungus for the exchange of nutrients between both partners, supplying minerals, water, and carbon (Herrera et al., 2017).

The life-span of an intracellular peloton is short**.** After only a few days it degenerates. The plant cell remains active throughout this process, while the surviving fungal hyphae re-colonise the orchid cells. This cycle of re-infection occurs several times in each cell. More than one fungus can produce pelotons at the same time in the same orchid tissue (Moore et al., 2011, Smith and Read, 2008).

In *Ophrys* the sequence of organ growth is different between asymbiotic culture and field. *In vitro*, leaves appear first and then roots, while in the field the roots develop first and then the leaves. This seems to be the result of asymbiotic culture allowing premature leaf development and an early change-over to photosynthesis (Rasmussen, 1995).

1.2 *Rhizoctonia* systematics:

The fungi that form mycorrhizal relationships with orchids are classified within the *Basidiomycota* and specifically within the genus *Rhizoctonia,* or *Rhizoctonia-like* fungi. They involve anamorphic genera such as *Rhizoctonia*, *Ceratorhiza, Moniliopsis,* and *Epulorhiza* and a variety of teleomorphs including *Ceratobasidium, Sebacina*, *Tulasnella,* and *Thanatephorus* (Roberts, 1999). Morphological identification is based on the comparison of mycelial structures on culture media *in*

vitro including the colour of the colony, septate hyphae, and angles of the lateral branching. Recently, criteria of the taxonomy of *Rhizoctonia* have been described related to the number of nuclei in their cells, whether uni, di, or multinucleate (Sneh et al., 2013).

1.2.1 *Rhizoctonia* as a pathogen:

Rhizoctonia species are pathogens for many plants. It leads to root rot diseases and failure for the plant. There has been extensive research on *Rhizoctonia* pathogens, especially in plants that are economically important such as potatoes and lettuce. Well studied species include *R. solani*, the most common species with a world-wide distribution, affecting more than 500 plant species (Tomaso-Peterson and Trevathan, 2007). *R. solani* (teleomorph = *Thanatephorus cucumeris*) is multinucleate and able to survive as sclerotia, or mycelia. It is considered a virulent pathogen to many plant hosts (Sneh et al., 2013). *R. zeae* is pathogenic to corn (*Zea mays* L.) and (with *R. oryzae*) is isolated from many warm and cool-season turf-grasses as the cause of *Rhizoctonia* sheath spot of turf-grasses (Tomaso-Peterson and Trevathan, 2007). *R. cerealis* causes sharp eye spot disease of wheat. Moreover, *R. croccorum* (teleomorph = *Helicobasidium purpureum*) has been reported to cause diseases in crops including violet root rot of carrot, sugar beet and saffron (see table 1.2) (García et al., 2006).

1.2.2 *Rhizoctonia* as a mycorrhizal symbiont:

R. solani has also been isolated from mature and healthy mycorrhizal orchid roots. Many *Rhizoctonia* orchid mycorrhizal species are saprotrophic and have the ability to undertake carbohydrate degradation (see table 1.2) (Moore et al., 2011). The focus of recent research has been on saprophytic and symbiotic *Rhizoctonia* originating from wood and leaves of angiosperms and orchid mycorrhiza, respectively (García et al., 2006). The *Tulasnella calospora* species group is a *Rhizoctonia*–like fungus that has been isolated from *Liparis japonica* (terrestrial photosynthetic orchid). This orchid has declined sharply in China and its mycorrhizal fungi and ecology are therefore important for its future conservation (Ding et al., 2014).

Table 1. 2: The form genus Rhizoctonia: Examples of anamorphic and teleomorphic genera. Adapted from García et al., 2006.

Anamorph	**Basionym**	**Teleomorph**	**Basionym**	**Method of Identification**	**Host plant**	**References**
Epulorhiza	*Epulorhiza epiphytica* *(R. repens).*	*Tulasnella*	*T. violea*	Morphologically by culture of roots on media (PDA and CMA)	Isolated from roots of two Brazilian epiphytic orchids (*Epidendrum rigidum* and *Polystachia concreta)*	(Pereira et al., 2003).
Opadorhiza	*R. globularis*	*Sebacina*	*S. vermifera*	Molecular phylogenetic analyses based on nuclear rDNA	Associated with many mycorrhizal types including ectomycorrhiza as well as orchid and ericoid. Isolated from roots of eight species of Orchidaceae such as *Eriochilus scaber, Cyrtostylis reniformis,* and *Caladenia catenata* and one non-orchid *Phyllanthus calycinus (*Euphorbiaceae*)*	(Weiss et al., 2004).
Moliniopsis	*R. solani*	*Thanatephorous*	*T. cucumeris*	Morphologically by culture roots on media (PDA).	Important ecological role as crop (potatoes) pathogens, orchid mycorrhizal symbionts, saprotrophs and endophytes (*Orchis purpurella*)	Sneh *et al*., 2013; Roberts,1999; (Parry, 1990); (Agrios, 2005).

1.3 Isolation of orchid symbiont fungi:

Traditional isolation methods used to isolate orchid mycorrhizal fungi involved plating fragments of surface-sterilized roots or separated single fungal pelotons on culture media. Even though these have many problems, such as with fungal and bacteria contamination, and difficulties with isolating mycorrhizal fungi that are slow growing (Zhu et al., 2008) they can be successful. For example, mycorrhizal fungi have been isolated from orchid roots using 70% ethanol for surface sterilization and then crushing the roots in a petri dish using a sterile glass rod to disperse the cells of the root cortex. This was followed with a pour plate technique using a medium containing 0.05% sucrose as carbon source and incubation at 20 °C in the dark. The fungal colonies that emerged were transferred to potato dextrose agar (PDA) (Yamato et al., 2005).

Zhu et al. (2008) isolated mycorrhizal fungi from peeled roots of *Cremastra appendiculata*. The peeled roots were treated with sterile distilled water followed by streptomycin and pencillin for surface sterilization. A suspension of cortical cells from these roots in sterile distilled water was dropped onto PDA supplemented with streptomycin and penicillin. Plates were incubated at 18 °C in the dark. Fungal hyphae that emerged were picked off and transferred to a new medium, PDA (Zhu et al., 2008).

1.4 Molecular Identification of fungi:

Nowadays, molecular methods are used widely to identify organisms, including fungi. The rRNA gene sequences of the internal transcribed spacer (ITS) regions and large subunit (LSU), as shown in Figure 1.1, are widely used for taxonomic identification within the fungi. This is because it is a fast and evolving portion of the rRNA cistron that is easy for amplification and has a suitably large barcode gap (Raja et al., 2017). The ITS1, 5.8S, and ITS2 regions are highly variable, extremely conserved, and variable to semi-conserved respectively (Martin and Rygiewicz, 2005).

Generally, fungal DNA is extracted followed by a PCR technique to amplify the ITS sequences for identification using pairs of primers selected for their wide or narrow taxonomic range (Yamato et al., 2005, Bougoure et al., 2005). Appropriate primers

can amplify a wide range of fungal targets but there can be a problem in excluding plant host sequences in DNA isolated from mixed sources such as plants or soil (Martin and Rygiewicz, 2005).

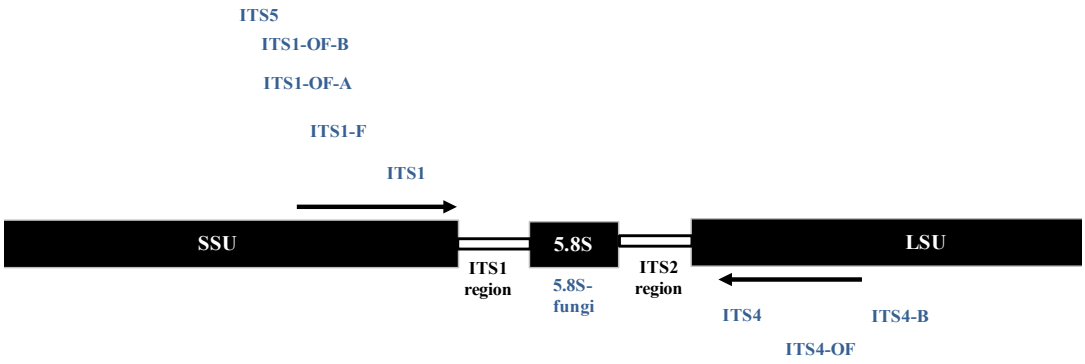


Figure 1.1: The ribosomal rRNA region. Small ribosomal rRNA subunit (SSU), large ribosomal rRNA subunit (LSU), internal transcribed spacer regions (ITS1 and ITS2). Arrows indicate direction when amplifying the DNA sample using PCR (direction shown in black and PCR primers in blue). ITS1 (forward) and ITS4 (reverse) are universal primers (White et al., 1990). ITS1-F (forward) and ITS4-B (reverse) (Gardes and Bruns 1993) and also ITS1-OF-A/ITS1-OF-B (forward) and ITS4-OF (reverse) (Taylor and McCormick 2008) were used for Basidiomycota. ITS5 (forward) (White et al., 1990) and 5.8S-fungi (reverse) (Epp et al., 2012) were used for Illumina sequencing. See chapter 4 for more details.

1.4.1 Barcoding sequences for fungi:

With the advent of molecular methods for identification of species, one valuable aim is to have a set of one or more PCR primers that could identify all species. This developed as the barcode concept (Hebert et al., 2003). A DNA barcode marker is a short nucleotide sequence in one to three loci used as a catalogue for identifying unknown samples from known classifications, and it is highly efficient, rapid and accurate for identified taxonomic groups (Toju et al., 2012).

It is particularly useful in mycology and ecology as many fungi cannot be cultured in isolation and most can only be identified clearly based on morphological characters such as fruiting bodies that are not always available (Toju et al., 2012). A desirable marker should be reliably amplified with universal primers and short enough to be easily sequenced with modern technologies. Also, it should provide sufficient interspecific variation when compared with intraspecific variation. The barcode must be standardized, available in large databases of sequences and be present in most of the taxa of interest (Lewis et al., 2011). Once this is available, as well as barcodes

being useful to identify isolated fungi in pure culture, they can also be applied to sequence data generated from DNA mixtures, such as from soil (see 1.5 below). This so-called meta-barcoding gives a new perspective on the presence and diversity of life in environments. This will be discussed further in Chapter 4.

1.4.2 Molecular methods in *Rhizoctonia* classification:

The literature on the identification of *Rhizoctonia*-like fungi shows a variety of approaches, including the use of cytomorphological features of teleomorphs and anamorphs, anastomosis group and ultrastructure of cell wall and septal pore features as well as molecular techniques (Nontachaiyapoom et al., 2010). Anastomosis group (AG), where if two isolates have the same AG their hyphae can fuse, also provides information especially useful in form species such as *Rhizoctonia*. The multinucleate *R. solani* and binucleate *R. spp* have been subdivided into 11 and 21 anastomosis groups respectively based on affinity and fusion of interacting hyphae of paired cultures (Hietala et al., 1994). Molecular techniques are one of the best methods among those mentioned, because it can distinguish between strains, creation of the teleomorph is not essential and it is relatively rapid (Nontachaiyapoom et al., 2010). The application of molecular methods to complex taxonomic situations such as in *Rhizoctonia* provides a practical way forwards to advance understanding of the classification of *Rhizoctonia spp.* and identification of isolates (Otero et al., 2002a).

Sequence analysis of both ribosomal ribonucleic acid (rRNA) genes and the ITS region have been used for investigating the diversity of *Rhizoctonia* isolates (Gonzalez et al., 2001). Analyses of these genes has shown the genetic relatedness of *Rhizoctonia* isolates and also confirmed some of the anastomosis categories (Tewoldemedhin, 2005). However, sequence analysis of the 5.8S ribosomal RNA gene of *Rhizoctonia* was considered not useful for genetic comparisons in *Rhizoctonia* because of limited sequence variation in this region (Gonzalez et al., 2001).

Sequences of the ITS region of certain AGs have been investigated more intensively than others, because of their relative importance as plant pathogens or as mycorrhizal endophytes in orchids, such as *Epulorhiza* (synonym *Rhizoctonia*, teleomorph *Tulasnella*). Therefore, rDNA-ITS data on these isolates are available in GenBank (Sharon et al., 2008). Also, the mitochondrial ribosomal large subunit (LSU) DNA

has been used to identify *Tulasnella* isolated from a single peloton obtained from fresh roots of *Dactylorhiza majalis* and it has been cultured from roots of *Neuwiedia reratrifolia* and sequenced and compared with sequences determined from reference fungal collections and published sequences. The fungi have been referred to the *Heterobasidiomycetous* order *Tulasnellales* and *Ceratobasidiales* (Kristiansen et al., 2004).

Two different approaches have been applied to investigate the specificity of orchid mycorrhizal association: (1) seed germination *in vitro*, (2) systematics of fungi which are present in roots of mature orchid plants. These will be discussed further in Chapter 4.

1.5 Next Generation Sequencing (NGS) Technologies:

DNA sequencing was initially only applied to individual genes or a small number of either model organisms or ones of great interest (e.g. humans) because of cost and technical constraints. NGS has allowed expansion to the complete range and complexity of life on this planet. As a consequence, genomic analysis is becoming a significant tool for understanding ecological diversity, functional and evolutionary history and especially for studies involving environmental samples originating from soil, water, and sediments, as well as the microbial communities of various host organisms (Shokralla et al., 2012). High throughput meta-barcoding has revealed previously invisible fungal, bacterial, and virus communities (Schmidt et al., 2013). Many studies of soil fungi depend on next generation sequencing technologies that have made these approaches possible because of the increased facility to generate DNA libraries for unknown specimens (Schmidt et al., 2013).

Sanger sequencing was developed in the 1980s and dominated the first era of sequencing, but because of the high cost, lack of automation and length of its sequencing-by-synthesis output is becoming less used (Kumar et al., 2016a). Currently, the newer next generation developments in sequencing technologies can be categorized into three types: sequencing by synthesis (Roche 454 Pyrosequencing, Ion Torrent, Illumina), sequencing by ligation (SOLiD and Complete Genomics), and single-molecule sequencing (Helicos, PacBio) (Egan et al., 2012, Hodkinson and Grice, 2015, Maheswari and Ravi, 2016, Goodwin et al., 2016). The pace of change in these technologies is illustrated by the Roche 454 platform. It was introduced in

2005 but production of new machines ended in 2013 and support ended in 2017 (Goodwin et al., 2016).

The methods allow sequencing DNA from multiple templates in parallel. Identification is based on comparing the sequences obtained with reference libraries of known species, for identification of environmental specimens with varying degrees of confidence (Shokralla et al., 2012). Developments in the computational analysis have helped in biodiversity measurements based on DNA clustering, annotation, and phylogenetic aspects (Hajibabaei et al., 2011).

The Illumina HiSeq is a platform that has become the most broadly used among second-generation sequencing technologies as it is lower cost and provides more reads (up to 5 billion reads per run) (Illumina, 2017) compared to the Roche 454 (~ 1 million reads per run) (Egan et al., 2012). However, a perceived limitation of these ultra-high-throughput technologies such as Illumina is that assembly is difficult because of their short read lengths and errors that occur during sequencing (Shokralla et al., 2012).

1.5.1 Illumina sequencing:

The Illumina sequencing approach is based on solid-phase bridge amplification in which both ends of the DNA fragments are ligated to adapters within miniature flow cells (Egan et al., 2012). A mixture of primers, DNA polymerase, and modified nucleotides are added sequentially to each flow cell channel (Goodwin et al., 2016). One end of the DNA fragment is attached to the surface of the flow cell. A bridge of adapters will be created from hybridizing forward or reverse primers. This eases amplification and generates amplicons which remain attached to the surface (Egan et al., 2012). Thus, it forms millions of clusters of identical DNA.

In the sequencing step, fragments in each cluster are provided with DNA polymerase and four differentially labelled fluorescent nucleotides that have their 3′-OH chemically inactivated. This blocking modification ensures that only a single base will be incorporated per flow cycle. After base incorporation, inactivated bases are washed away (Shokralla et al., 2012). The activated nucleotide in each cluster is measured by total internal reflection fluorescence using either two or four laser channels (Goodwin et al., 2016). The fluorophores are then cleaved and washed from

the flow cells and the 3′-OH group is regenerated. A new cycle begins with the addition, elongation, and cleavage of new nucleotides. This cycle is repeated until the DNA fragment has been synthesized to its target length.

1.5.2 NGS application:

NGS platforms are useful for applications such as resequencing of individual genes-of-interest or entire genomes. The large amount of data generated in each run has allowed studies that were previously too difficult and expensive. These include *de novo* assemblies of bacteria and eukaryotic genomes and the transcriptomes of cells, tissues, and organisms (RNA-seq) (Metzker, 2010). Metagenomic studies have become possible with NGS methods. Different platforms are most suitable for particular applications. For example, the Illumina platform is good for gene variant discovery because of the massive volumes of high-quality bases produced per run.

In plant biology, NGS has made it possible to speed applied research by providing the data needed to analyze interactions between genes to produce genetic networks. Through allowing sequencing of any plant genome and transcriptome it has become possible to identify enzymes that contribute to secondary metabolites, with possible medical applications (Johnson et al., 2012). Also, plants' genome-wide transcriptomic responses to biotic and abiotic environmental stress can be studied, seeking applications in species of economic interest. Moreover, the huge databases resulting from NGS provide data that can be used to answer questions relating to plant evolution and diversification (Johnson et al., 2012).

In order to analyse this new DNA sequence data, many computing resources or tools have been developed for sequence assembly, interpretation and data storage. This will be discussed further in Chapter 5.

1.5.3 Transcriptomes in non-model plants:

Gene expression studies have been restricted to studies of single genes, especially outside genetic model organisms, using methods that have gradually advanced from northern analysis to qPCR. Microarrays were the first method for observing expression of many genes simultaneously (Axtner and Sommer, 2009, Naurin et al., 2011). With NGS technologies it has become realistic to analyse the whole genome or whole transcriptome in non-model organisms and a standard practice in molecular ecological studies (Wolf, 2013). In the few model plants, the sequencing reads can be

mapped directly to the genomes for information about the expression of genes. For most plants, there is no genome information and the sequenced reads have to be assembled prior to assessing their expression (Han et al., 2016). The ability to sequence whole genomes, and their entire transcribed regions as whole transcriptomes in a rapid and cost effective way is transforming the life sciences (Johnson et al., 2012).

1.6 Family *Orchidaceae*:

The Orchidaceae is the world's largest and most diverse family of angiosperm plants. It estimated to contain 26,000 to 35,000 species in over 750 genera (Chutima, 2012). Orchids are economically important in horticulture because of the diversity and beauty of the flowers and their ability to survive for many weeks. The pods of *Vanilla planifolia* are significant for the commercial production of vanilla flavouring (Prince and Gunson, 1994). In addition, some orchid products are believed to have medicinal properties and are still used as aphrodisiacs, treatments for sores, emetics, and vermifuges (Smith and Read, 2010). Finally, but most importantly, orchids are also iconic species within natural populations with many species that are rare or widely dispersed.

Orchids are distributed across all continents (apart from Antarctica) and grow in many different habitats. Three main groups of habitats can be distinguished: epiphytic, lithophytic, and terrestrial (Smith and Read, 2010). This project focuses on a terrestrial orchid *Ophrys apifera* (bee orchid), illustrated in Figure 1.2.

1.6.1Terrestrial orchids:

Terrestrial orchids have been estimated to be about 12% of the total number of orchid species and grow in soil (Chutima, 2012). According to distribution studies, there are about 400 species in North Asia and more than 300 in Europe–North African–Near East region. Also, in Australia, there are 300 species and over 100 in New Zealand, most of which are endemic species (Frericks, 2014).

Flowering patterns are irregular in terrestrial orchid species because of a complex interaction of biotic and abiotic factors (Kindlmann and Balounová, 2001). Many terrestrial orchids can adopt different strategies for carbon acquisition during their lives. They range from green and photosynthetic, to achlorophyllous and fully

mycoheterotrophic at a mature stage (Smith and Read, 1997). Some of these green orchids remain partially parasitic and at the adult stage take more than 50% of their carbon from fungi (Girlanda et al., 2011, Selosse and Roy, 2009, Stöckel et al., 2011), and this combination is termed mixotrophy. In this case, the plant continues to depend on fungi as a carbon source to complement or exchange their carbon acquisition from photosynthetic processes (Girlanda et al., 2011, Selosse and Roy, 2009).

Globally, terrestrial orchids are more threatened than epiphytic orchids (Lehnebach et al., 2005), because of their great dependence on soil-borne mycorrhizae. Half of the orchid extinctions have been terrestrial species (Swarts and Dixon, 2009) with major threats being land clearing for agriculture, mining and urban development, weed invasion, habitat degradation, grazing, altered environmental conditions, habitat fragmentation, climate change and collection of plants. Habitat loss and alteration are serious threats to orchid populations and survival of species (Batty et al., 2002, Koopowitz and Lavarack, 2003).

Most European orchids are terrestrial and rooting in the soil gives protection and allows them to tolerate the winter (Rademacher, 2012). Terrestrial orchid seeds have the characteristic of microspermy which is that they produce very numerous, tiny seeds (Arditti and Ghani, 2000). Each consists of a globular embryo surrounded by a thin seed coat (testa), but importantly without endosperm. Therefore, it needs an external carbon source and nutrients to germinate (Rasmussen, 1995).

Although many seeds of orchid are scattered, typically through the air, it has been calculated that roughly 1 to 45% of orchid seeds germinate in soil and develop into mature plants. The persistence of viable seeds in the soil for most terrestrial orchids is between 1-4 years (Ariza, 2013, Batty et al., 2002, Brundrett, 2007). It has been considered that population sizes are variable because of periodic vegetative dormancy (i.e. individuals do not appear aboveground in one or more entire growing seasons, then re-emerge in subsequent years), (Lesica and Steele, 1994, Shefferson, 2009, Shefferson et al., 2012, Wells and Cox, 1991).

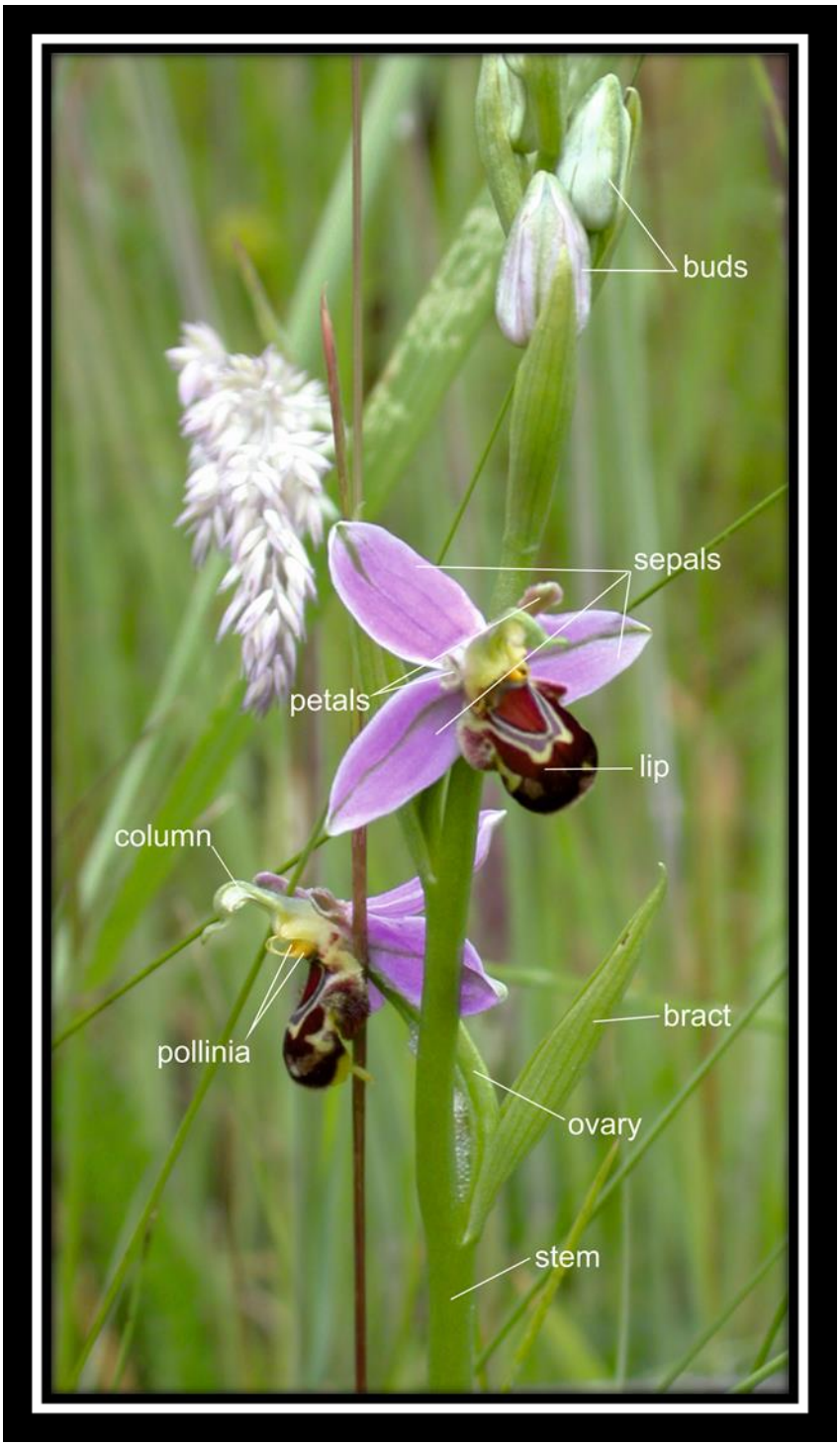


Figure 1. 2: Flower stem of bee orchid (*Ophrys apifera*) in Liverpool. Height is approximately 30 cm. Column (incorporating male and female organs). Petals (lip or labellum) help with pollination through attracting insects. Pollinia contain thousands of pollen grain to fertilise the large number of minute seeds within each ovary (Photograph, M. G. Jones).

True roots in orchid abort at an early stage of development and instead rhizomes as well as root tubers are formed (Rademacher, 2012). The leaf veins are parallel and used for transporting nutrients and water. The characteristic inflorescence (see Fig 1.2), usually involves flowers opening from upwards along the stem, although some species have solitary flowers. *Ophrys apifera* prefers a calcareous dry meadow over loess and limestone. Their flowering time is late and it can better endure spring frost comparing to other *Ophrys* species (Rademacher, 2012). Terrestrial orchids are generally considered to exist underground through the year as corms or tubers linked with a mycorrhizal fungal network into the soil. Then they develop through the soil, developing leaves to capture light for photosynthesis and a flower shoot to reproduce. However, orchids can be difficult to find because of their lifestyle in hiding (Pierce and Belotti, 2011).

1.7 Aims of this thesis:

The aim of this thesis was to better define the life and mycorrhizal symbiosis in Ophrys apifera. The study was therefore divided into three main parts:

Firstly, to understand the above ground growth dynamics and life-cycle of *Ophrys apifera* over the annual seasons through morphometric measurement to define the annual timing of leaf production and flowering. It was also important to identify whether individuals flowered in successive years. A small, individually labelled, population of plants was used for this study over 3 years. Thus, measurements were made of leaf length, width, and area, length of flower shoot, number of flowers per shoot and the number of leaves per plant It is important to consider environmental factors and data on rainfall and temperature was therefore obtained. This is further described in Chapter 2.

Secondly, to identify fungi both associated with the orchid root and in soil surrounding the plants using several methods to provide a rounded view. This included isolation of fungal cultures from soil and roots as well as isolation of DNA from these sources. The rRNA region was analysed to identify the fungi present using universal and phylum-specific primers. This was finally performed taking a microbiome approach using Illumina rRNA amplicon sequencing. Bioinformatic

approaches were used to analyse the sequencing data and to explore the taxonomic relationships. These are presented in Chapter 4.

The final aim was to obtain more understanding of orchid mycorrhizal physiology at a molecular level using RNA-seq to identify genes differentially expressed in tissues (protocorm, corm, and root). It was important to have tissue from plants grown in the absence of mycorrhiza. In order to obtain suitable tissues for comparison, protocorms were grown aseptically from seeds and this is described further in Chapter 3. In addition, tissues were obtained from plants growing in the natural environment. Methods to obtain RNA of suitable quality for sequencing were tested to identify suitable methods. Finally a comparison could be made between aseptic protocorms, and corm and root tissues from the field. The results shed light on the differentially expressed genes in the under-ground organs in *Ophrys apifera* and are in Chapter 5.

2 Chapter two: Life history of *Ophrys apifera.*

2.1 Introduction: -

The goal of this chapter is to understand the above-ground dynamics of *Ophrys apifera* over the seasons and plant life-cycle. A small population in the university campus was studied using photographs to record leaves and flower shoots throughout the year.

2.1.1 Life history of orchid plant:

Mature orchid plants have three annual stages in their life cycle: vegetative rosette, flowering stalk, and dormant underground stage. In nature, the germination of orchid seed and seedling establishment are dependent on mycorrhizal fungi to provide the carbon sources for energy via pelotons, which are a coil of hyphae within the cells (Yam and Arditti, 2009). This is because of the limited nutrient reserves within each orchid seed.

The embryo is ovoid surrounded by a few hundred cells and then enclosed in the testa. The seed contains very limited food reserves in the form of oil and starch since it lacks a tissue equivalent to the endosperm (Nadarajan et al., 2011). So, an external supply of nutrition is required (Brown et al., 1981). In nature, this is supplied by mycorrhizal fungi but seeds may germinate and develop normally in artificial media (Hadley and Harvais, 1968), into protocorms that are initially small and spherical in shape and can store reserves (Zhu et al., 2008).

Since young seedlings of terrestrial orchids live underground, knowledge of this part of the life history is limited. There are several types of underground storage organs: such as corms and rhizomes that are important for storage of starch, calcium oxalate crystals, and water. Rhizomes are modified stems that grow horizontally and like corms have buds which help rhizomes to develop from it, while corms are stem structures and do not grow horizontally. The plant has two corms, a new one for survival during a current dormant season and an old one from a previous season (Dressler, 1993). They play an important role as a reserve of energy and water for the vegetative season and to prevent freezing of the plant tissue during cold winter months. Borsos (1990) has published details about storage organs in several European orchid species as well as the location of storage material. Corms and rhizomes give a protected environment for new buds through dormant periods until

they are ready to grow into new green vegetative organs above the soil surface (Borsos, 1990).

In Mediterranean orchids such as the genus *Ophrys* (see section 2.1.2), the leaf rosette usually starts to grow from the beginning of autumn and develops in the winter period (see Figure 2.1). Other species may delay emergence above ground until the late summer such as *Spiranthes spiralis*. An exception is the species *Goodyera repens*, found throughout Europe, which is most common in Scandinavia and is evergreen (Breitkopf, 2011). The cold winter induces some of these wintergreen species to produce more leaves in spring (Mrkvicka, 1992).

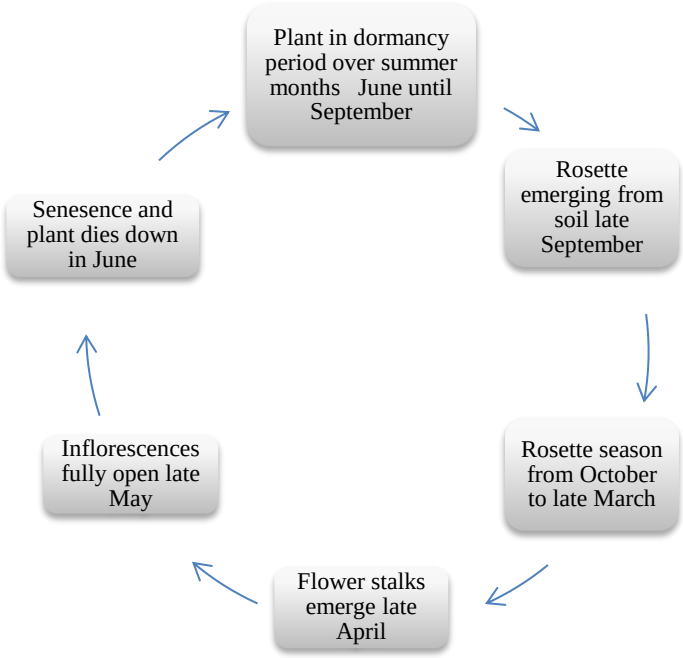


Figure 2. 1: Overview of annual life cycle of *Ophrys apifera* summarized from observation during this study.

The typically mycophagy stage occurs in the wet autumn and early winter when dead leaf and plant material is available, including for the fungi living in symbiosis with orchids. Fuchs and Ziegenspeck (1924) found that in species of *Dactylorhiza* the pelotons of the fungi are alive and contained glycogen in October and November. Also, they showed that the pelotons can die and lyse but this is followed by re-infection. In December some of these pelotons are dead (Fuchs and Ziegenspeck, 1924). However, the timing of this stage is different according to species. For example, *Goodyera repens* was highly infected by fungi from December to May

(Alexander and Alexander, 1984).

In the genus *Ophrys* the protocorm formed after seed germination will produce an exogenous mycotrophic root in the early autumn. In the beginning of the dry season (beginning of July), the protocorm will form a short mycorrhizal network, which can form a tiny globose corm from the axil of one of the scale leaves (Rasmussen, 1995). By the summer period, only that corm will survive. In the second autumn, the apical bud on the top of corm develops to a short rhizome carrying one or a few roots and ending with a leafy shoot (see Figure 2.2). A similar development has been noted in *Ophrys holoserica* (known as *Ophrys fuciflora*) and *Aceras anthopophorum* (Rasmussen, 1995).

According to these observations, researchers suggested that in regions with more precipitation and wet winters orchid mycorrhiza is active throughout the year and that specifically wetting is the main factor affecting mycorrhizal development (Alexander and Alexander, 1984).

Many orchids flower in the early spring until the middle of summer and a few (such as *Epipactis spp*) from late summer to autumn (Breitkopf, 2011). The flower spike in *Ophrys apifera* appears during spring until early summer and after flowering, the plant leaves die down and stop above ground growth until September (Pedersen et al., 2007). The European peak time for flowering orchids is from March to May, with flowers often open for more than three weeks. If the conditions are unfavourable then flowering by individual plants may not happen for one or more years (Pedersen et al., 2007). Two important factors have an effect on the success of flowering; pollination by animals and mycorrhizal symbioses with fungi as have been pointed by Waterman and colleagues (2011).

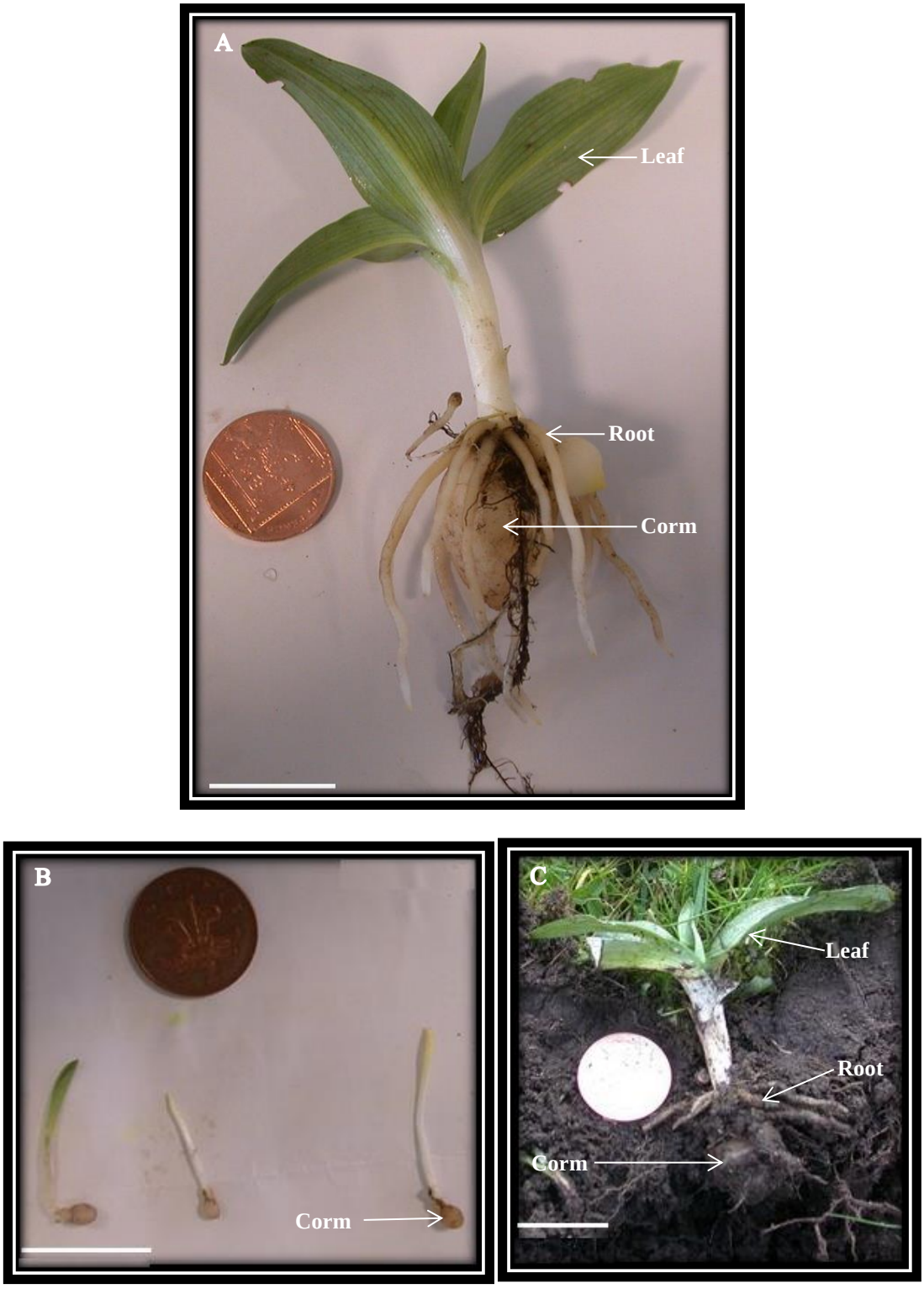


Figure 2. 2: *Ophrys apifera* plants. (A) March 2016 showing vegetative growth in spring, underground corms for current and future year's growth. Mature plant, at least 7 years old, transplanted from open ground prior to building work (see section 2.2.1). (B) November 2015 showing corms and leaf (Young plants developing in pots). Comparison with the corms in A and C illustrates the resources built up prior to conspicuous vegetative growth (C) October 2016 showing vegetative growth in autumn. Scale bar = 2.5 cm.

2.1.2 Genus *Ophrys*:

Ophrys has been well defined both morphologically and genetically within the *Orchidaceae* (Bateman et al., 2003). The genus has a distinct floral morphology and unique reproductive strategy through the flowers mimicing bees or other insects. *Ophrys* is a Eurasian genus found in the Mediterranean region, although a few species have extended to northern Europe. Generally, it is a resident of dry soils forming a basal rosette in late autumn that continues through winter and spring to die away through the flowering period. In the dry summer period, the plant survives as a corm (Rasmussen, 1995). The genus is often seen as a pioneer plant in places such as roadside edges (Pedersen et al., 2007).

Some of these species grow in small groups of individuals, for example *O. insectifera* is found in populations of several dozen individuals (Osiadacz and Kręciała, 2014). Nevertheless, estimating population is different according to the area. *O. apifera* can grow in small groups e.g. 25 in Poland reported by Pierce et al. (2006), or in large groups so that a population of hundreds has been recorded (Wells and Cox, 1991).

2.1.2.1 *Ophrys apifera*:

Ophrys apifera is one of the four *Ophrys* species that are found in the UK, along with *O. sphegodes*, *O. fuciflora*, and *O. insectifera* (Devey et al., 2009). This is the only species in genus *Ophrys* which practices self-pollination through the curved apex of the column (see Figure 1.2). It is found around the Mediterranean as far as the Levant to the east and the British Isles to the North West. It is adapted from full sunlight to light shade, dry to wet limy soil and typically includes the edges of forest environments (Pedersen et al., 2007). The primordium of a new corm appears in November, but growth is slow and delays until spring (March). Initial roots can occur in September, but the new roots form from the leaf sheath and appear in November (Möller, 1987). The flat leaf rosette grows close to the ground so that plants are not conspicuous until the inflorescence develops vertically.

The term ‘demography’ has come to be used to refer to spatial and temporal variation in abundance of plants. There are number of studies that have been published on demographics of orchids, such as a demographic study of a population of *Ophrys sphegodes* (Hutchings, 2010), effects of life-state on detectability in *Cleistes bifaria*

(Kéry and Gregg, 2003), and analysis of dormancy and survival in *Cypripedium reginae* (Kéry and Gregg, 2004).

Very few publications can be found about the demographic of populations or growth dynamics of *Ophrys apifera*. There are obvious difficulties with such studies of natural plant populations related to for example, difficulty in assessing the age of individuals and in studying the underground plant parts. Figure 2.2 (B) shows young plants found when removing the plant shown in Figure 2.2 (A) from its pot, indicating the difficulty of detecting young individuals. Wells and Cox (1989) have shown that the behaviour of the plant during flowering and vegetative stages depends on the age, size, and number of leaves in any growing season. There are large differences in the numbers of leaves among individual plants, likely reflecting differences in age. In general, *Ophrys apifera* individuals with more leaves, longer leaves, and a greater leaf area are more likely to flower than individual with fewer or shorter leaves (Wells and Cox, 1989).

Wells and Cox (1991) also indicated that it is dangerous to depend on one year's result to estimate population size. Throughout a ten-year research period, they found that firstly, some individual plants flowered for up to four years consecutively and were rarely monocarpic. Secondly, annual variation in the number of plants in flower varied from 6-57% and the flower shoot length and the number of flowers per inflorescence were correlated with summer rainfall periods of both the previous year and the year of study. Finally, they took into account damage to leaves by animals that varied year by year but was less than 10% throughout the 10-year study (Wells and Cox, 1991).

The current study focused on the above ground portion of *Ophrys apifera* over a three-year period. The aim of this study is to understand the above ground dynamics of vegetative and reproductive growth throughout the year. Also, to compare our result with the previous results reported before, to define whether the additional period of study changed understanding.

2.2 Materials and Method:

2.2.1 Study area - Biosciences plants:

One group of 17 plants was studied, shown in Figure 2.3. They have developed naturally in the lawn outside the Biosciences Building (University of Liverpool) in Liverpool city (see section 4.2). Each individual plant was numbered in order to return to the same plant at successive recordings. Small wire fences were placed around the plants to prevent disturbance from lawn mowing (see Figure 2.3). Vegetation around the plants was clipped back to facilitate data recording but there was otherwise no treatment of the plants.

Figure 2. 3: Population of 17 *Ophrys apifera* plants outside the Biosciences Building, 11th November 2017, University of Liverpool.

2.2.2 Data recording

Photographs of the basal rosette of each plant were taken 4th February 2014 to 23th July 2016 using a Nikon 995 camera and tripod. A scale (two pence coin, diameter 2.5 cm) was included in each photograph. Counts of the number of leaves in the basal rosette were recorded. The state of the leaves (green and alive, dead or damaged) was recorded as well.

The imageJ program (Version 1.47) was used to measure the photographs to provide leaf length, width and area. Only green leaf surface was included in these measurements. The length of the flower shoot (April to June) and the number of open

flowers (May to June) were recorded each year and used as an estimate of flower production. A ruler was used to measure the length of the shoot.

2.2.3 Environmental data:

Information about weather (rainfall and temperature) was obtained from the National Meteorological Library & Archive (http://www.metoffice.gov.uk) by contacting them. Crosby weather station was chosen as it is the nearest station to University of Liverpool campus, around 7.5 – 8 miles away, and at an altitude of 9.0 m above mean sea level, Latitude = 53:50N Longitude = 03:06W. The data set provided daily rainfall in mm and the temperature (maximum, minimum, and mean) in ºC over four years (2013-2016).

The 30 year average (1981-2010) of annual temperature and rainfall were obtained from https://www.metoffice.gov.uk/public/weather/climate/gctb66ydw.

2.3 Results:

2.3.1 Bee orchid growth cycle:

2.3.1.1 Vegetative development: The data were obtained from February 2014 (Day 0 = 4/02/2014) onwards. The annual growth cycle of the plants is shown in Figures 2.4-A. On 11th of October 2014 and 19th of September 2015 new leaves begin to emerge to form a basal rosette. Growth continues throughout the winter and early spring, reaching the maximum leaf area by March in 2015 - 2016. From April and May onwards green photosynthetic leaf area decreases. The leaves visibly yellow and turn black from the tip. All the above ground vegetation dies prior to summer, leaving only dead leaves and flower stalks visible from the end of June until mid-October 2014, and from the end of the June until end of the September 2015. This adaptation of summer dormancy to a Mediterranean climate of hot, dry summer is maintained into the wetter, cooler summers of the UK.

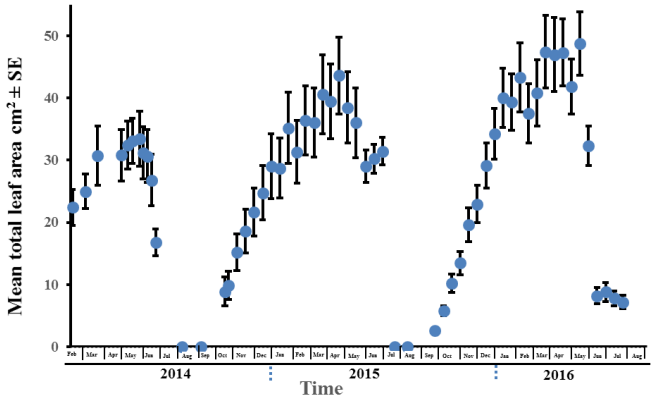
Leaf area increases through both emergence of new leaves and elongation of the leaves. The scatterplot in Figure 2.4-C shows a good correlation between these two measure ($R^2 = 0.89$).

Figure 2.4-B also indicates that the mean total leaf area per plant was different each year with the maximum increasing over the 3 years monitored. The mean total leaf area per plant was lower in 2014 than in 2015 and 2016. Figure 2.4-C shows the relationship between growth of the leaf length and the leaf area during 2014 to 2016.

Environment conditions affect plant performance, such as the amount of precipitation and temperature. Over the study period, the amount of rainfall in each year was different (see Figure 2.5-A). It was less in 2013 than 2014 and 2015 and all were below the 30 year average (1981-2010). In contrast, there is no dramatic different between annual temperature ranges during the study period and 30 years average (1981-2010), as shown in Figure 2.5-B.

A

B



C

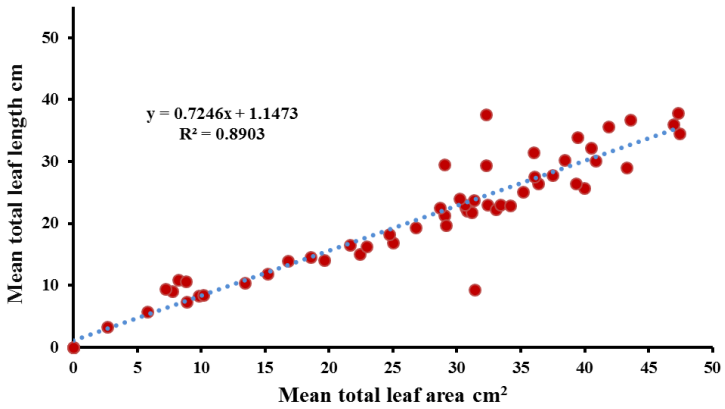


Figure 2. 4: Development of leaf area. (A) Plant 13 in 2015. Leaves emerge from underground corm in autumn and grow through winter. In late spring resources are re-directed from leaves to the flower shoot, first visible in May. (B) Mean total leaf area per plant from Feb 2014 to September 2016. N = 17. Error bars represent the standard deviation of the mean. (C) Growth of the leaves. Each point depicts the mean of total leaf area (dark red) and mean total leaf length per plant with data from 2014-2016. Scale bar = 2.5 cm.

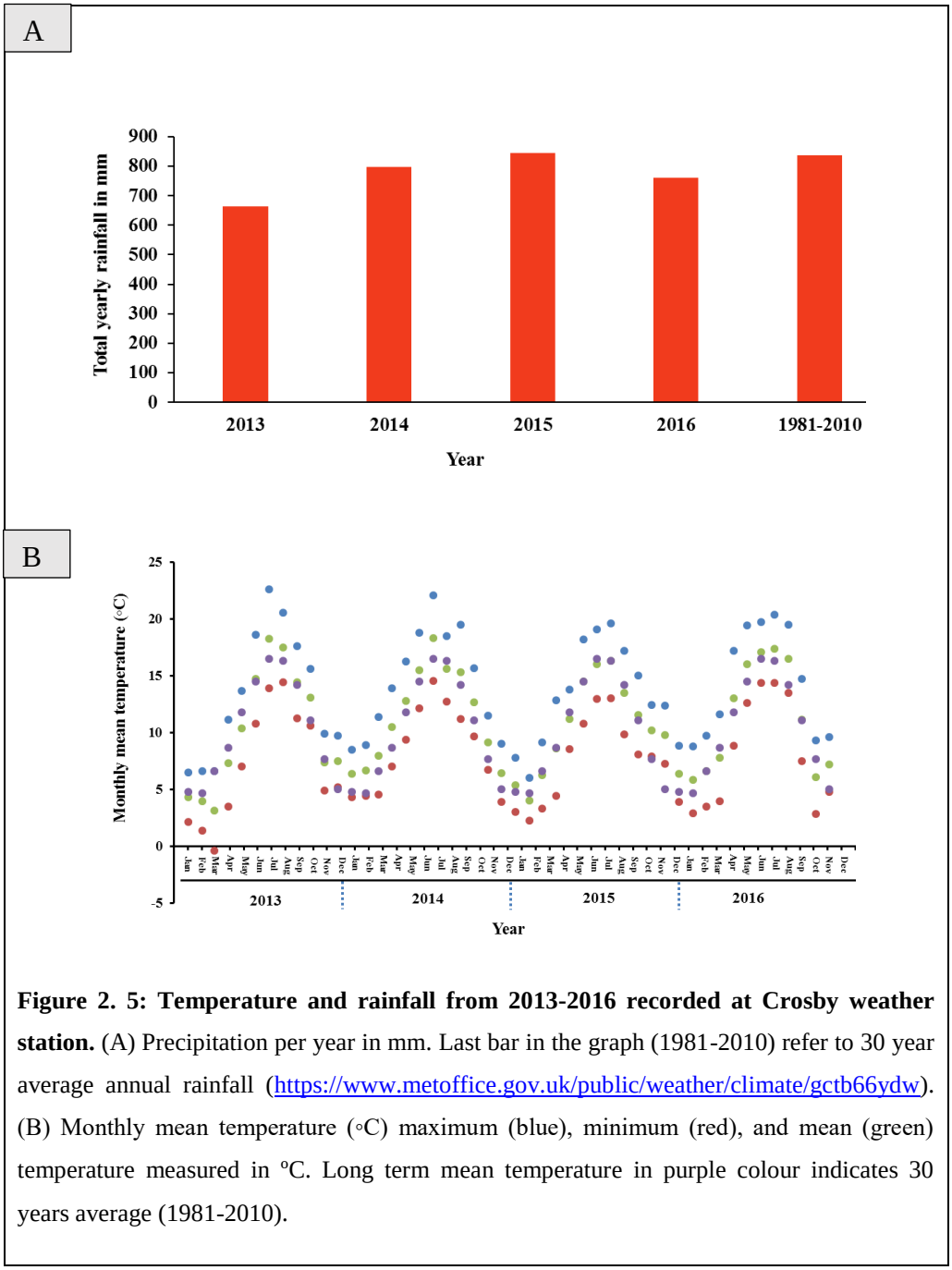


Figure 2. 5: Temperature and rainfall from 2013-2016 recorded at Crosby weather station. (A) Precipitation per year in mm. Last bar in the graph (1981-2010) refer to 30 year average annual rainfall (https://www.metoffice.gov.uk/public/weather/climate/gctb66ydw). (B) Monthly mean temperature (◦C) maximum (blue), minimum (red), and mean (green) temperature measured in ºC. Long term mean temperature in purple colour indicates 30 years average (1981-2010).

2.3.1.2: Flower shoots development:

The flower shoots elongated rapidly once flowering started, as indicated in Figure 2.6 A. This began at the beginning of May in 2014 but from end of the April in 2015 and 2016. After reaching a maximum length in June further shoot extension ceases. Flowers open successively from the lowest one upwards. In all three years, all flower shoots had died by the end of June. It is clear that the mean flower shoot length was longer in 2016 than 2015 and 2014. The leaves die as the flower shoot develops as evidenced by the blackened leaves at the base of the flower shoot that can be seen in

Figure 2.6 B by comparing the basal leaves in the photograph from 28th May, where they are green, with that of 26th June where they are completely black.

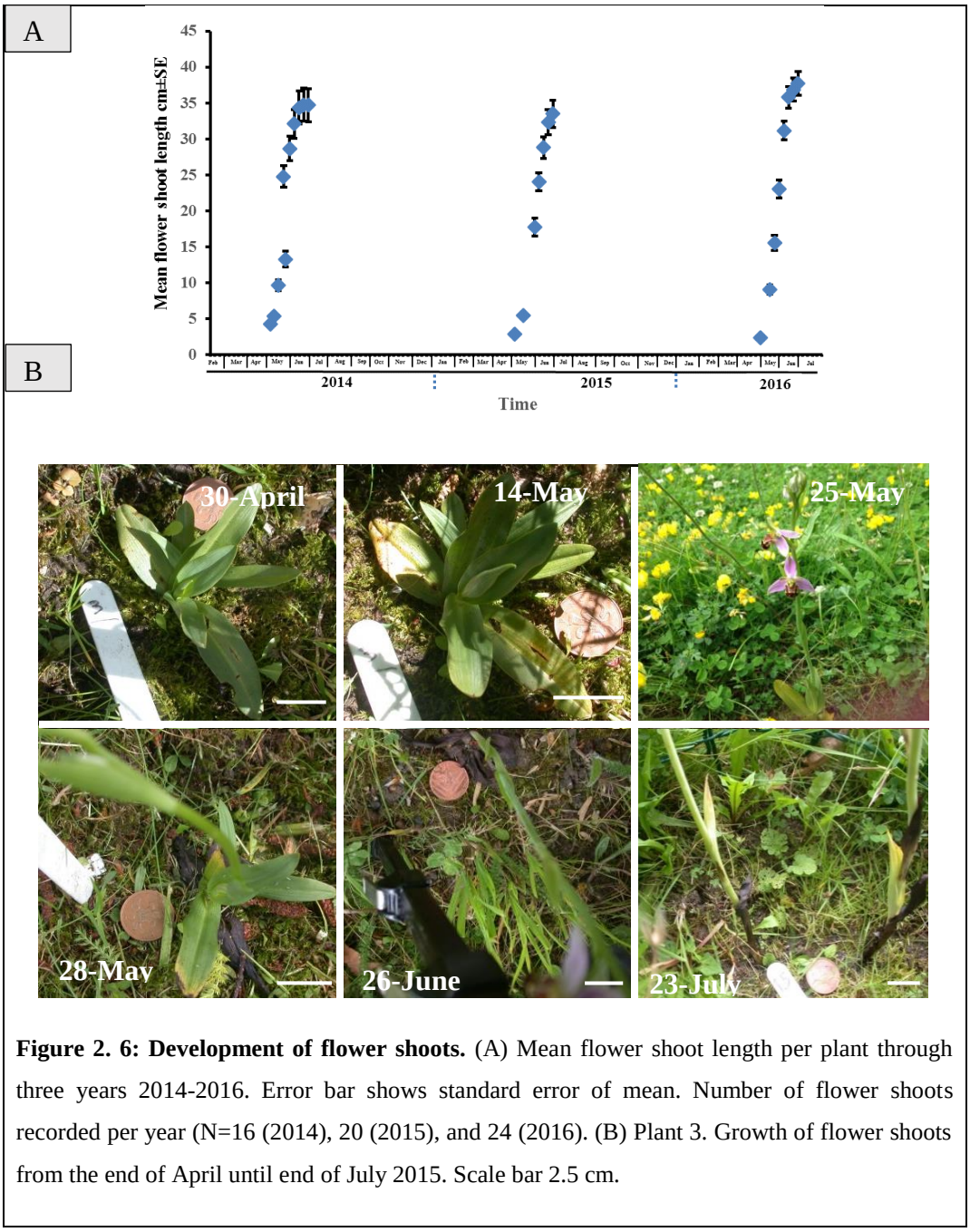


Figure 2. 6: Development of flower shoots. (A) Mean flower shoot length per plant through three years 2014-2016. Error bar shows standard error of mean. Number of flower shoots recorded per year (N=16 (2014), 20 (2015), and 24 (2016). (B) Plant 3. Growth of flower shoots from the end of April until end of July 2015. Scale bar 2.5 cm.

Table 2.1 shows that the number of flower shoots per plant can be more than one. For example plant 7 in 2014 has two flower shoots, and in 2015 has three. It also reveals that some plant do not have flower shoot in one year but flowered in the year after such as plant 8.

Table 2. 1: Flower shoots in each year. The number in the brackets refers to the number of flower shoot for each plant.

Year	Plants with more than one flower shoot	Plants without flower shoot
2014	Plant 7 (2)	Plant 8
2015	Plant 7 (3), Plant 15 (2), Plant 16 (2), and Plant 18 (2)	Plant 10 and Plant 17
2016	Plant 7 (3), Plant 13 (2), Plant 15 (2), Plant 16 (2), and Plant 18 (2)	

2.3.1.3: Relationship between vegetative and reproductive development:

Figure 2.7 shows that although the maximum mean number of leaves and flowers per plant differs in successive years, the growth dynamics were the same. Leaves appeared successively from the autumn with a linear increase in mean number. The average rosette contained a maximum of 5 - 6 leaves in 2014, but up to 8 in 2015 and 2016. The rapid decreases in mean leaf number after this maximum occur as flowering begins.

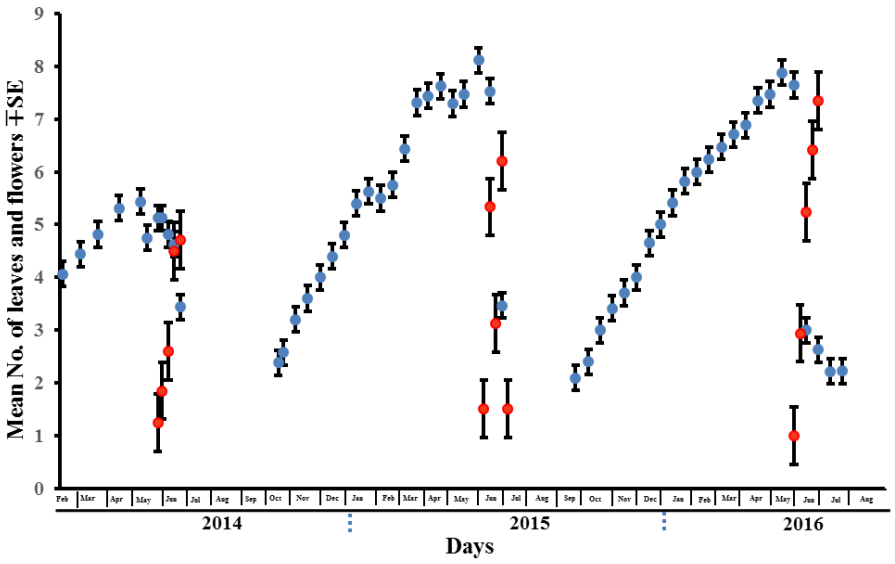


Figure 2. 7: Mean numbers of leaves and flowers per plant over the period 2014-2016. (N=17) Mean numbers of leaves (blue) and flowers (red) per plant. Error bar is standard error of the mean

Figure 2.7 also shows clearly that flowering occurs when leaf growth, in terms of leaf number as well as area, has reached a maximum. The mean number of open flowers is shown. However, as shown in Figure 2.8, there is no significant correlation (R^2 = 0.18) between the maximum number of leaves and maximum number of flowers produced by each plant.

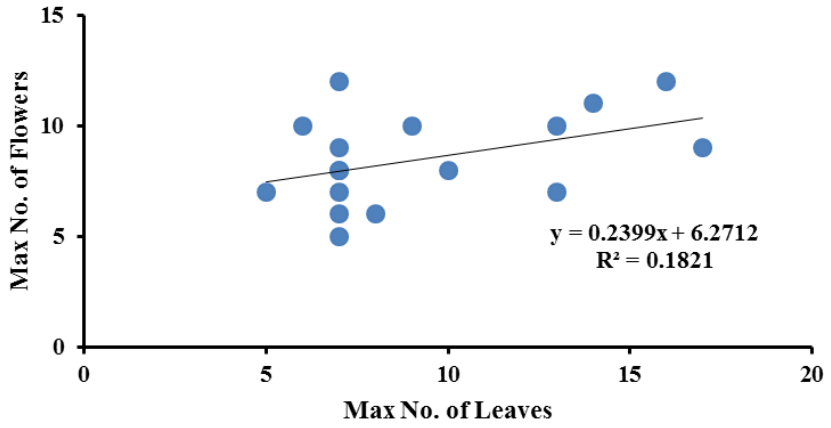


Figure 2. 8: Relationship between the maximum number of flowers and maximum number of leaves produced by each plant during the 3 years 2014 – 2016

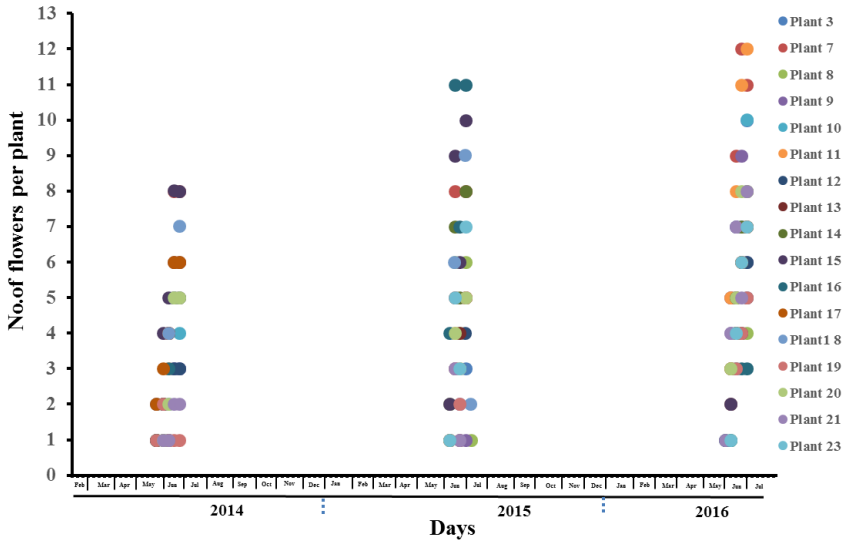


Figure 2. 9: Numbers of flowers opening plant 2014- 2016. Each color indicates a different plant. The blank space between each year shows time with only vegetative growth (late September-late April).

Development of the flower shoot and opening of the flowers happened over a short time period. The flowers were open for a month from late May to the end of June, differing slightly from year to year as shown in Figure 2.9. Table 2.2 gives the dates when the first flower opened and when the last flower was still alive.

Table 2. 2: Date of that the first appeared and the last flower survived.

Year	First flower opened	Last flower opened
2014	21-May	25-June
2015	5-June	3-July
2016	28-May	25-June

Although most of the plants studied flowered each year, five did not flower every year (see Table 2.3). In 2016 all plants flowered. The number of flowers increased over the years from 2014 to 2016.

Table 2. 3: Flowering in the plants studied 2014 – 2016. The maximum number of flowers opened on each plant is shown. A blank box means that the plant did not flower in that year.

Plant ID	2014	2015	2016
plant 3	3	3	10
plant 7	8	8	12
plant 8		6	4
plant 9	4		10
plant 10	4		10
plant 11	5	5	12
plant 12	3	4	6
plant 13	6	5	7
plant 14	1	8	8
plant 15	8	10	5
plant 16	5	11	3
plant 17	6		7
plant 18	7	9	8
plant 19	2	5	5
plant 20	5	5	8
plant 21	2	3	8
plant 23		7	7

2.4 Discussion:

This is the first detailed record of the above ground annual growth cycle of *O. apifera*. This cycle starts from a production of leaves above ground in September for a few months in a basal rosette. New leaves are added until there are, on average 5-9. Leaf production continues until the end of May when leaf number stops increasing and there is then a dramatic change to reproductive growth. The flower shoots initially has rapid elongation, so that in the period of two months reaches their maximum extent of around 35 cm. At the same time, the leaves turn yellow and die indicating a reallocation of resources from vegetative to reproductive growth.

Flowers open progressively along the stalk from 21-May to 3-July. The flowers and flower stalk then also die, leaving only the below ground portion of the plant from July to September. The date that the first leaf emerged differed between the three years (see table). This agrees with previous findings of Wells and Cox (1991).

The maximum number of leaves has no statistically significant relation to the maximum number of flowers, whereas Wells and Cox (1989) reported that a plant with more leaves has more probability to flower.

However, the number of leaves might have influenced the number of flowers and lack of correlation between them can be attributed to small population size (17) compared with 182-222 studied by (Wells and Cox, 1991). Further data collection is required to determine how rainfall and temperature affect the development of leaves, flower shoot, and flowering. The restricted time of research was the reason for not collecting data more than three years.

There is a difference between 2014 and the subsequent two years, in terms of both mean leaf number, area and number of flowers per plant. The environmental data for rainfall and temperature (Figure 2.5) do not show an obvious difference between these years. However, the plants did not suffer from leaf damage caused by mowing as part of the University's routine lawn maintenance once the plants were fenced at the start of this study. Although the leaf rosette has a strong tendency to grow horizontally, the ends of leaves are removed by the lawn mowers and whole leaves can be pulled away. Any flower shoots are cut off. This will have occurred prior to February 2014. The similarity between performance in 2015 and 2016 supports this

idea. It would be interesting to assess whether the additional leaf area retained after fencing also allowed a larger underground corm to develop that could support increased plant performance. However, this would require destructive harvesting among a larger *O. apifera* population. This additional leaf area retained may have supported the additional vegetative and reproductive growth in 2015 and 2016.

Most terrestrial orchids flower from early spring to the middle of summer, while in *Epipactis* spp and a few of other species flowering is from late summer to autumn (Breitkopf, 2011).

There was no obvious relationship between average temperature or precipitation on plant performance and this agrees with the previous finding by Janečková et al. (2006) that found flowering was influenced by leaf area, not weather and concluded that this could be due to low weather variability during the period of 5 years studies on five *Dactylorhiza majalis* populations. However, Tali (2002) noticed fluctuation of plant height (*Orchis ustulata*) between years was more than between populations and attributed this fluctuation to spring weather.

3 Chapter three: Seed germination of *Ophrys apifera in vitro*.

3.1 Introduction: -

The aim of this chapter was to obtain *Ophrys apifera* tissue grown in aseptic conditions to compare with field grown plants using transcriptomic methods (Chapter 5). This involved using sterilization methods and media to support seed development for long enough to achieve this goal. In addition, there was an unsuccessful attempt to co-culture seeds with isolates of *Tulasnella* to investigate germination under symbiotic conditions.

3.1.1 Orchid seeds:

Orchid seeds differ from other angiosperms in that they have very limited nutrient reserves to support the development of the embryo and germination (Porras-Alfaro and Bayman, 2007). It means that the seeds are small, very light and produced in large numbers (Arditti and Ghani, 2000). Typically, more than 4 million seeds are present in an orchid seed capsule. The seeds themselves are typically 0.01 mm in width, 0.05 to 6mm in length and weigh 0.34 to 24 μg per seed (Arditti and Ghani, 2000). The development of these seeds structural, anatomical, morphological, physiological, ecological, symbiotic and horticultural aspects have been discussed by researchers (Weatherhead et al., 1986, Mutsuura et al., 1962, Arditti, 1992). The consequence of the small size is reduced reserves to support germination leading to dependence on mycorrhizal fungi for successful germination in nature.

3.1.1.1 Orchid seeds structure:

The individual seed (see Figure 3.1A) contains a spherical embryo which is surrounded by a cell layer (testa) for protection against any infections by bacteria or fungi. The testa also can have secondary thickening and provide an air-filled structure for wind dispersal. Within the embryo, there is a seed reserve body containing fat and proteins (Pedersen et al., 2007).

3.1.1.2 Interaction between orchid seeds and fungi during germination:

The interaction between orchid mycorrhiza and orchid seeds is special in that germination will not occur in the absence of the fungal symbiont (Arditti, 1992), although they can be germinated *in vitro* without a fungal partner using specific media (Knudson, 1922). Some species of orchids have a specific interaction with

certain species of fungus, while others have relations with more than one species of fungi (Watkinson, 2002).

After germination, the embryo can develop into a unique structure called a protocorm (see Figure 3.1 B) which is found only in orchids (Chang et al., 2011). The protocorm will develop rhizoids (see Figure 3.1 C) and after a few days or weeks the protocorm develops a shoot apex with leaf primordium (see Figure 3.1 D and E) and will form roots (Watkinson, 2002).

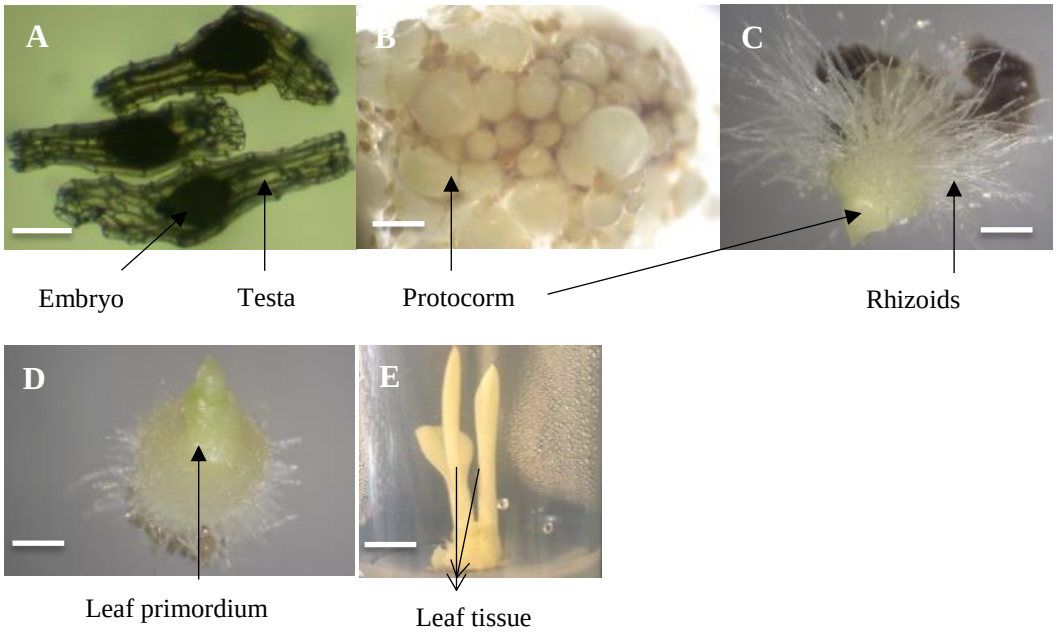


Figure 3.1: Stages used to assess germination of seed and development in O. apifera. (A) Seeds. (B) Protocorm development after 11weeks (C, D) Protocorm development after 17 weeks (E) Development of leaves after 11 months Scale bar: A=100 µm, B, C, D= 1 mm and E= 10 mm. (All photographs, W. Abdullah).

3.1.2 Laboratory culture of orchid seeds:

Many studies on the first stage of orchid seedling germination and growth have been undertaken in the laboratory. Studying seed germination in terrestrial orchids asymbiotically using artificial media, is difficult and needs very specific ingredients and therefore many researchers use symbiotic fungi for propagation as Haas (1977). In contrast, asymbiotic seed germination in the epiphytic orchids is generally easy and successful (Otero Ospina and Bayman, 2009).

This will not happen in nature and in soil fungi are probably essential, but there are few studies on the earlier stages of orchid development under field condition (Dearnaley, 2007).

The most important factors contributing to better germination *in vitro* are sterilization and chemical composition of media. Therefore, it is important to understand the disinfection treatment to sterilize surfaces and the influences of this treatment on seed germination (Jevsnik and Luthar, 2015).

3.1.2.1 Decontamination of capsules and seeds:

An essential step to reduce contamination before *in vitro* cultivation of orchid seeds is surface sterilization of the seed capsule. Many chemicals have been used to avoid or reduce microbial contamination. Although sodium hypochlorite (NaClO) 3-6% has been used recently, calcium hypochlorite ($Ca(ClO)_2$) 10% (w/v) concentration is also a common sterilizing agent used for disinfection (Arditti, 1977, Bergman, 1996). Studies have indicated that not only is $Ca(ClO)_2$ less damaging to the seeds, but also could break the dormancy of the seed compared with NaClO and thus accelerate germination (Knudson, 1948, Sweet and Bolton, 1979, Waes and Debergh, 1986a). Despite these studies, NaClO is used extensively.

Ethanol, hydrogen peroxide and mercuric chloride are less often used for disinfection (Arditti, 1977). Brown et al. (1981) and Yanagawa et al. (1995) reported that adding decontaminating chemicals or fungicides and bacteriocides such as benzimidazole directly to the medium are harmful and inhibit seed germination.

Surface sterilization (seed or capsules) with 0.3% NaClO for 45 minutes gave optimum conditions for germination in *O. holoserica* and *O. sphegodes* (Frosch, 1982). A shorter treatment of 0.2% NaClO for 10 minutes followed by 3-5% hydrogen peroxide (H_2O_2) for 10 minutes' treatment also gave successful germination (Lucke, 1971). However, $Ca(ClO)_2$ for 4 hours was used to treat seeds of *O. apifera* and *O. sphegodes* for best results (Waes and Debergh, 1986a). Surface disinfection prior to *in vitro* culture of mature seeds and immature seeds in *Ophrys* species were reported by Kitsaki et al. (2004). They used a different technique for surface sterilizing seeds by enclosing them in bags of Whatman No.3 paper and flooding with distilled water for 3 days at 4 °C, followed by surface sterilization in

10% $Ca(ClO)_2$ containing 0.1% Tween 20 for 15 min, then rinsed three times in sterile distilled water. Capsules of immature seeds were surface sterilized in the same way but immersed in the $Ca(ClO)_2$ Tween 20 for 40 min (Kitsaki et al., 2004). Recently, other chlorine-based disinfection solutions have been used for surface sterilizing seeds such as sodium dichloroisocyanuric acid (dichlor) and was reported to be 100% efficient for four epiphytic orchid species (Jevsnik and Luthar, 2015).

In this study, Domestos (with NaOCl and NaOH content below 5%) has been used for surface sterilization of capsules followed by three separate washes in sterile distilled water before *in vitro* culture.

3.1.2.2 Culture medium:

In nature, plants obtain mineral nutrients from the soil and carbon from carbon dioxide in the atmosphere through photosynthesis. Culture medium has to replace these sources for *in vitro* plant cells to grow. The components are macronutrients including ions of nitrogen (N), potassium (K), calcium (Ca), phosphorus (P), magnesium (Mg) and sulphur (S) and micronutrients such as iron (Fe), nickel (Ni), chlorine (Cl), manganese (Mn), zinc (Zn), boron (B), copper (Cu), and molybdenum (Mo).

In addition, a medium can contain vitamins, amino acids or other nitrogen compounds, sugars, other organic compounds, solidifying agents, and growth regulators (George et al., 2008, Goh, 2016). For orchid *in vitro* culture purposes, the common micro-propagation media are based on MS medium, Vacin and Went medium, and Knudson C medium (Arditti, 2009, George et al., 2008, Goh, 2016, Murashige and Skoog, 1962, Pant, 2013). However, media containing plant extracts have also been used, such as coconut milk-enriched medium and pineapple-enriched medium (Kitsaki et al., 2004).

The literature on orchid culture media shows development of techniques suitable for the orchid seeds and less suitable for fungi, and other media suitable for the growth of mycorrhizal fungi (Anderson, 1991, Beardmore and Pegg, 1981, Breddy, 1991, Clements and Ellyard, 1979, Clements et al., 1986, Rasmussen, 1990). The role of several substances in seed germination and promotion of seedling growth has been studied.

Carbon sources: This is one of the most important substances in the medium for germination and growth of protocorms and both sucrose and glucose have been used. Huh et al. (2016) have shown that ¼ MS medium supplemented with 10g/L sucrose can improve seed germination *in vitro* and significantly increased the rate of both protocorm germination and seedling development.

It has been reported that trehalose could have the same effect as sucrose and a sugar signal is involved (Müller et al., 1999). Therefore, it can be used as carbon source for growth of orchid protocorms. The trehalose is translocated along the fungal hyphae to the protocorm as an energy source during symbiotic orchid seed germination (Liu et al., 2006). Nevertheless, the effect of this sugar on seed germination *in vitro* is very poorly investigated by researchers (Jheng et al., 2006, Liu et al., 2006, Ponert, 2009).

Nitrogen sources: The exact type of nitrogen source best for orchid seed germination is still unclear with both inorganic and organic forms effective. Many investigators have used inorganic nitrogen in the form of nitrate (NO_3^-) in media for temperate orchids (Ponert et al., 2013). Organic nitrogen in the form of amino acids is also a very effective substrate in media for orchid seed germination in both terrestrial and epiphytic orchids (Ponert et al., 2013, Rasmussen, 1995). The amino acids are usually provided as complex mixtures derived from milk (casamino acids), meat (peptone) or yeast extract (Arditti and Ernst, 1993, Szendrák, 1997a).

3.1.2.3 Growth regulators:

Auxin and cytokinin are the hormones (growth regulators) that have the major roles in regulating growth and physiological function in seed germination (Rasmussen, 1995). The cytokinin kinetin is most often used in media for orchid species, for its role in stimulating germination of immature seeds (Pierce and Cerabolini, 2011). Kinetin has a positive effect on growth at levels up to around 1 mg/l (Harvais, 1982, Steele, 1996, Waes and Debergh, 1986b), although Ponret (2011) demonstrated that no differences were found in germination of *Dactylrhiza majalis* over the range 0-1 mg/l. A concentration of 100 mg/l is possibly toxic resulting in no protocorm growth (Waes and Debergh, 1986b). The auxin indole-3-acetic acid (IAA) is also used in

orchid culture media. This hormone stimulates cell elongation to improve plant growth and produce root initiation (Davies, 2010, El-Tarabily, 2008).

Coconut milk (water or liquid of coconut) has frequently been used for *in vitro* orchid seed germination studies (Kitsaki et al., 2004). This additive contains nitrogenous compounds, inorganic ions, amino acids, enzymes, vitamins, sugar, and plant hormones with the amounts changing according to the age and ripeness of the coconut (Arditti and Ernst, 1993). Coconut water has been used in several media for terrestrial orchids not only in micropropagation but also for *in vitro* seed germination (Szendrák, 1997a).

3.1.4 Culture of seedlings asymbiotically:

Successful asymbiotic *in vitro* germination and culture of temperate orchids have been reported using several media. These include Pfeffer-glucose medium, Norstog medium, Ramin medium, Mead and Bulard medium and Curtis medium (Ponert et al., 2011). These media are different from other media used in plant tissue culture since the mineral ion concentration is less because these plants are more sensitive to high salt concentration (Fast, 1982, Harvais, 1974, Van Waes, 1984). Veyret (1969) found that *Ophrys* species are generally difficult to germinate *in vitro* because of problems with the plantlet acclimatization at the end of the culture period, although successful germination of up to 90% of seeds has been reported by (Ponert et al., 2011). Pfeffer′s medium plus glucose has been used to germinate *O. apifera*, and a medium with organic nitrogen is recommended rather than one with inorganic nitrogen (Hadley, 1982, Waes and Debergh, 1986a).

Previous research has shown that asymbiotic germination *in vitro* and then transfer to pots and to a greenhouse is a very effective method for producing healthy seedlings, giving a higher rate of germination than in nature (Znaniecka et al., 2005).

3.1.5 Other factors that affect seed germination *in vitro*:

Light, temperature, and pH can have a great impact on orchid seed germination.

3.1.5.1 Effects of temperature on the seed germination:

Terrestrial orchids can be classified into two categories depending on their response to temperature: firstly, a group where seeds can be germinated *in vitro* immediately

after harvesting when incubated between 17 and 25°C and a second group which needs cold stratification to stimulate germination (Ballard, 1987, Coke, 1990, Ponert et al., 2013, Ponert et al., 2011, Rasmussen, 1992, Rasmussen, 1995).

Previous research on asymbiotic germination of European terrestrial orchid seed using different temperatures showed that all germinated well at 17°C, but at 23°C some had problems germinating and *Ophrys* did not produce leaves (Ponert et al., 2011). They also showed that a high temperature has a harmful effect on germination, as also observed by Hass (1977). *Ophrys* species thus seem to fall into the first group for germination temperature.

3.1.5.2 Effects of light on seed germination:

The role of light in seed germination has also been evaluated. Darkness is generally best, although some species respond to dim light or light-dark photoperiods are needed. Until the true leaves appear and a small stem develops darkness is required and if they are moved to light too early the protocorms will turn brown and die. After the development of the small stems, the little plantlets can be moved into the light and soon, usually within two weeks, the apical region turned green (Szendrák, 1997a).

Light has been shown to inhibit germination in some terrestrial orchid in the genus *Dactylorhiza*, although germination of *Dactylorhiza majalis* seeds was higher when incubated in the light and subsequently moved to the dark (Rasmussen, 1990). Seedling growth of protocorms in light and high temperatures can give some problems such as becoming brownish with exudation of phenolic compounds. This can be solved by moving cultures into darkness or subculturing frequently to a new medium (Ponert et al., 2011). In a study on the effect of different photoperiods on asymbiotic seed germination *in vitro* of *Habenaria macroceratitis*, a terrestrial orchid species, there was no effect of light on the initial stage of seed germination, but a significant impact on the development of the protocorm (Stewart and Kane, 2006).

3.1.5.3 Effect of seed maturity on germination *in vitro*:

Mature orchid seeds can be stored for a long-term under refrigeration if care is taken to not damage them (Kitsaki et al., 2004). Following storage, it is necessary to

decontaminate the seed surface and break the seed dormancy. On the other hand, immature seeds are ready to culture immediately *in vitro* with no need to treat them for dormancy. However, these immature seeds cannot be dried or stored long term for culture later (Steele, 1996).

3.1.6 Germinating genus *Ophrys in vitro*:

Development of the seed into a small protocorm is slow. After sowing *Ophrys* seeds in August, the initial protocorms were visible in March the following year (Lucke, 1971). Germination rates and proportions are relatively different from study to study (Rasmussen, 1995). Within a month, about 90% of *O. insectifera* immature seeds germinated after they had been removed from capsules 5-7 weeks after pollination (Borriss and Albrecht, 1969, Ponert et al., 2013). However, in *O. sphegodes*, immature seeds germinated best if removed from capsules 8-9 weeks after pollination (Lucke, 1971). About 60% of seeds germinated when using full strength Curtis medium for both *O. fusca* and *O. lutea* within 2 months (Borsos, 1990).

Hoppe and Hoppe, (1988) indicated that a solid orchid medium (OM) was better for differentiation in tissue culture of *O. apifera* than a liquid medium. They tested both media and found the liquid medium was much faster to contaminate than a solid medium. Therefore, only solid medium was used for differentiation. A low level of casamino acid in the orchid medium helped the formation of protocorm, shoot, and leaves (Hoppe and Hoppe, 1988).

3.2 Materials and Method:

3.2.1 Collection of orchid capsules:

Seed capsules were collected randomly at the end of June 2015 and 2016 from plants outside the Biosciences Building and growing in pots on the Life Sciences Building roof (see section 4.2.1 for more about the plants). White immature seeds from green capsules were used in the seed germination experiment.

3.2.2 Equipment:

These experiments were carried out using a laminar air flow cabinet to provide aseptic conditions for working. Media and all other equipment (forceps, needle, spatula, tissue, flask, distilled water, and tips) were wrapped with foil and sterilized by autoclaving.

3.2.3 Culture medium composition:

A modified version of the *Ophrys* seed culture medium (OM) described by Hoppe and Hoppe (1988) was used for seed germination. After the media were autoclaved, the vitamin solution and trace elements solution were added as shown in Table 3.1. In addition, the growth regulators indole-3 acetic acid and kinetin were sterilized using 0.45 µm filters (Merck KGaA, Germany) and added to the sterilized medium.

3.2.4 Seed isolation:

The surface of the capsule was sterilized in 10 % Domestos bleach for 15 min, and subsequently washed three times with sterilized distilled water. After that, the seeds were obtained by cutting the capsule longitudinally using a sterile blade and forceps and transferred to the growth medium OM or Oat Meal Agar (OMA). The seeds from one capsule were placed at four points on three plates.

3.2.5 Culture Condition:

The seeds were initially incubated in a growth room at 22 °C in the dark. Plates were covered with aluminium foil but after two months were exposed to the light growth room (22 °C, 12-hour L/D). They were observed for contamination and germination, and stages in seed germination were photographed using a stereomicroscope (Novex B) and digital camera (Canon, EOS D2500). After 9 months, some protocorms were

harvested, and frozen immediately in liquid nitrogen and then kept at – 80 °C for extraction of RNA (see Chapter 4).

In addition, some seeds were incubated from the start in tubes containing 20 ml OM medium. Tubes were covered with a single layer of foil and incubated in a culture room (22 °C, 12-h photoperiod). After 2 month incubation, some of these seeds became swollen and started to grow into small white protocorms although others did not germinate.

Table 3. 1: Composition of media used for seed germination of *Ophrys apifera in vitro*.

Media	Composition (Per litre)
Potato Dextrose Agar (PDA)	Potato Dextrose Powder (24g) (Sigma Aldrich, USA), Agar (8g), pH (5.6±2).
Oat Meal Agar (OMA)	Basic Oat Agar (8.6g) or oatmeal (60g), agar (12.5g) Sigma-Aldrich, UK (Waller et al., 2002).
Orchid Medium (OM) modified from (Hoppe & Hoppe, 1988).	KH_2HPO_4 (0.3g) $MgSO_4\ _7H_2O$ (0.1g) Fe EDTA (30mg) glycine (2mg) caseamino acid (1g) myo-inositol (1.1g) sucrose (20g) agar (6g). indole-3 acetic acid (0.5mg) kinetin (0.5mg) Trace Elements Solution ($1ml^{-1}$) Vitamins Solution ($2ml^{-1}$)

Table 3. 2: Vitamin and Trace Elements solutions.

Solutions	Composition
Vitamin Solution	para-aminobenzoic acid 160 mg inositol 160 mg pantothenic acid (Ca salt) 240 mg nicotinic acid 40 mg, pyridoxine 100 mg riboflavin 40 mg choline chloride 560 mg putrescine (HCl) 800 mg biotin stock (20 mg /100 ml) 4 ml dH_2O 396 ml.
Trace Elements Solution	sodium tetraborate 40 mg cupric sulphate 400 mg ferric orthophosphate 800 mg manganese (III) sulphate ($MnSO_4$) 800 mg sodium molybdate ($NaMoO_4$) 800 mg zinc sulphate ($ZnSO_4$) 8 g dH_2O1L.
Both solutions were sterilized using autoclave at 121^{o}C, 108 kPa and 30 min and store at 4 $^{\circ}$C in dark.	

3.2.6 Inoculation of fungi into seed germination cultures:

Tulasnella isolates obtained from roots (see section 4.2.2.3.2) were grown on PDA medium at 25 ^{o}C for two weeks in darkness, and then plugs (diameter 1.5cm) with active mycelium from the colony margin were transferred to a petri dish containing 20 ml sterile OM medium and four groups of seeds. These were covered with foil and incubated at 22 ^{o}C for one year, assessing growth and contamination every 2 months under stereomicroscope. Some of the seeds were sub-cultivated after 3 months onto new 20 ml OM medium in tubes in order to reduce dryness and browning.

3.3 Results:

3.3.1 Developmental stages of seed germination *in vitro*:

Seeds germinated asymbiotically on both OM and OMA. Data were collected and recorded after two months of incubation and continued for another nine months. About 2 months after culture initiation, seeds (Figure 3.2.A) had developed into undifferentiated protocorms (Figure 3.2.B). After 3 months, the testa had ruptured to release the protocorms. These protocorms were spherical in shape and white in color signifying success in germinating the seeds *in vitro* using artificial media (Figure 3.2.B and C).

Over the following six months, rhizoids and initial shoot were observed (Figure 3.2.D, E, and H), followed by differentiation to form a shoot and leaves at nine months (Figure 3.2.G and F). However, although leaves and shoots formed (Figure 3.2.I), photosynthesis did not start. The leaves were white, not green, turned a brown colour and then died. Thus the *O. apifera* seedlings grew well to the stage of producing a shoot and leaves but did not produce any green leaves (white) or roots.

3.3.2 Effects of media on *in vitro* seed germination and protocorm formation:

Ninety-one green capsules were used for seed germination *in vitro*, in the two years 2015 and 2016 with two media, OM and OMA. Each capsule was cultured in three replicates. Table 3.3 shows the number of plates containing group of seeds that were either viable, dead, or had developed into protocorms over a 12-month period. Germination was estimated visually. The proportion of plates with groups of protocorm formation was higher in 2016 than 2015 on both media. Figure 3.3 shows a typical situation where some seeds are developing into protocorms while others are dead.

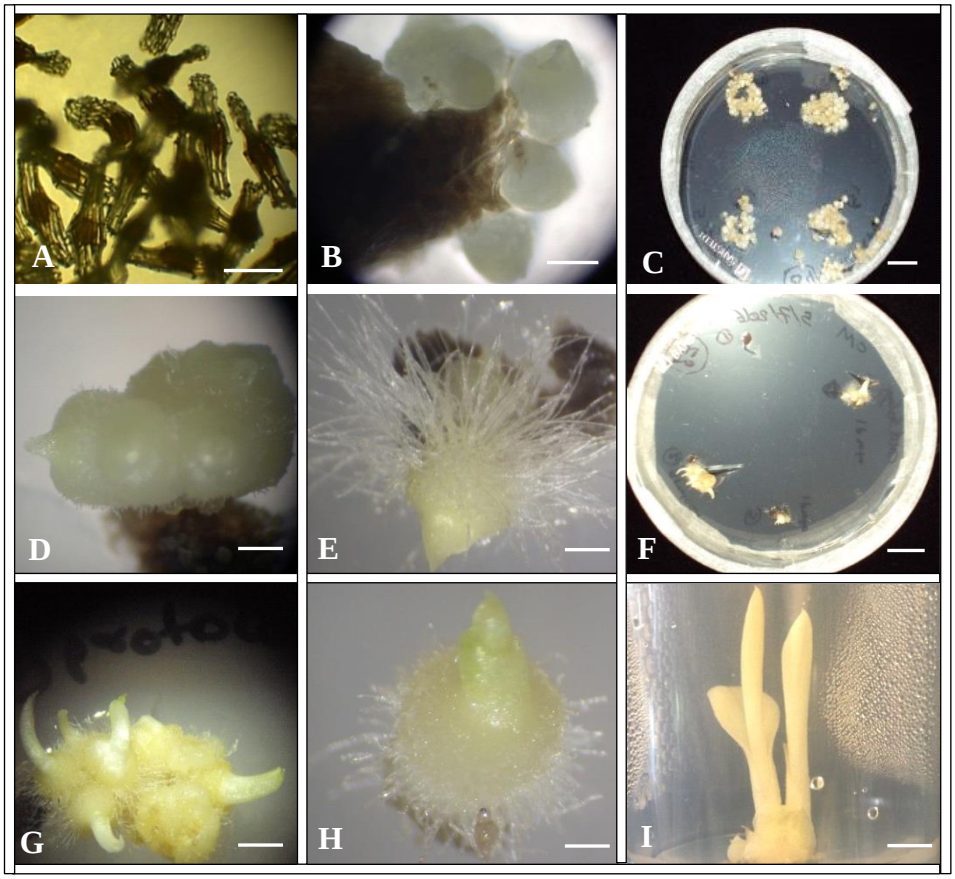


Figure 3. 2: The terrestrial species *O. apifera*, showing the stages of development from seeds on OM culture medium in aseptic condition. (A). 2 months first stages of seed germination, (B). Protocorms at 3 months, (C). Groups of protocorms at 3 months (D, E, and H). Hairs of rhizoid and shoot initial at 6 months, (F-G). Development of shoot and leaves at 9 months. (I). Shoot and leaves at 12 months after transfer to tubes. Scale bar: A and B 100 µm; C 0.5 cm; F 0.8 cm; D, E and H 1mm; G and I 1 cm.

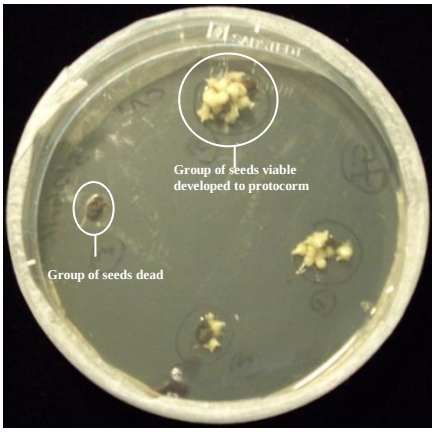


Figure 3. 3: Groups of *O. apifera* seeds developing to protocorms and others where the seeds have died after 6 months incubation.

Table 3. 3: Growth of seeds and development of protocorm *in vitro*. Plates were incubated for one year. The different number of plates depends on availability of capsules in each year.

Year of culturing (end June to beginning July)	***Growth medium***	***Total number of plates***	***Number of plates contain seed groups***		
			Some seed in group viable	**Groups where protocorms developed**	**All seeds in group dead**
2015 (June 2014-July 2015)	OM	156	68	8	88
	OMA	7	0	0	7
2016 (June 2015-July 2016)	OM	26	19	13	7
	OMA	30	18	11	12
2016 (with fungus B)	OM	20	14	0	6
	OMA	16	15	0	1
2016 (with fungus R3)	OM	1	1	0	0
	OMA	17	14	0	3

The Table 3.4 indicates the number of tubes that contained seed groups (viable, developed to protocorm, or dead). As can be seen the total number of tubes in 2015 (160) was higher than 2016 (43), but the number of entirely dead seed groups was higher in 2015 and no protocorms formed. In 2016 a greater proportion of seed groups were viable and nine groups have developed to the protocorm.

Overall, all protocorms that formed in 2016 were harvested for transcriptomics RNA in chapter 5 whether OM or OMA and from plates or tubes.

Table 3.4 Sub-culturing seed groups in tubes without fungus in aseptic condition.

Year of culturing (end June to beginning July)	***Growth medium***	***Total number of tubes***	***Number of tubes contain seed groups***		
			Some seed in group viable	**Groups where protocorms develop**	**All seed in group dead**
2015 (Tubes)	OM	160	68	0	92
2016 (Tubes)	OM	43	23	9	20

3.4 Discussion:

3.4.1 *In vitro* cultivation:

The results showed that it is possible to germinate seeds *in vitro* without the presence of the fungus that develop to produce protocorm, shoot, and leaves in *Ophrys apifera,* as reported previously for two genera of European orchids *Ophrys* and *Dactylorhiza* (Ponert et al., 2011). However, leaves were not produced.

The growth medium, essentially as used by Hoppe and Hoppe, (1988) including the additive of indoleacetic acid and kinetin to support the plant growth. Interestingly, oatmeal medium also worked well with asymbiotic germination, although, it was previously reported that germination of fresh seeds of *Ophrys* on oat medium was poor in association with fungi (Rasmussen, 1995). It has been suggested that orchid seeds not only need a fungus but also unknown elements supplied by organic additives (e.g. yeast extract or potatoes extract) to stimulate germination, and the combination of them in the medium increases the percentage of germination (Rasmussen, 1995).

3.4.2 Factors that impact on germination:

A suitable medium and incubation conditions are the most important factors for successful asymbiotic and symbiotic germination *in vitro*. Previous research on the effect of temperature on asymbiotic germination of several species of European terrestrial orchid showed that all germinated well at 17°C, but had some problems with growth at 23°C and *Ophrys* did not produce corms (Ponert et al., 2011). A harmful effect of high temperature on germination plant has also been referred by Hass (1977). The temperature used in our study (22 °C) may, therefore, have not been optimal, even though a proportion of seeds germinated and developed.

4 Chapter four: Isolation and identification of orchid mycorrhizal fungi:

4.1 Introduction: -

Orchid mycorrhizal fungi are extremely important for the growth of orchids in nature. Therefore, the objective of this chapter is to isolate and identify fungi from the root of the *Ophrys apifera* and the soil around the roots. In order to overcome the challenges associated with isolating microorganisms and sequencing DNA from roots and soil, several methods were used. Finally, Illumina amplicon sequencing of DNA from soil samples was undertaken and by using bioinformatics, the DNA sequences were analyzed. In addition, a fungal isolate was obtained from the roots and its growth characteristics studied.

4.1.1 Overview:

The total number of fungal species has been estimated to be about 1.5 million, but only around 80,000 – 120,000 species have been described (Hawksworth, 2001, Hawksworth, 1991). More recent estimates of the total number, making use of environmental molecular ecological studies from a diverse range of habitats has, if anything, increased this estimate of total species (Hawksworth and Luecking, 2017). This is therefore one of the least explored biodiversity resources on the planet. For mycorrhizal fungi, the presence of fruiting bodies is an unreliable method to identify the underground form associated with roots (Gardes and Bruns, 1993). Molecular microbial ecology has had a huge impact on the appreciation of the genetic diversity of soil microorganisms which had escaped the notice of traditional observational and laboratory cultivation methods (Rondon et al., 2000).

4.1.2 Isolation of fungi from orchid roots:

The majority of orchid mycorrhizal fungi are *Basidiomycota* and most fungi isolated from mycorrhizal orchid roots have been assigned to the taxon *Rhizoctonia* or *Rhizoctonia*-like fungi (Rasmussen, 2002), (see also Chapter 1, section 1.2.1). There are other approaches to identify fungal symbionts, namely isolation of the fungus into a culture or DNA sequencing. Both approaches have been applied to terrestrial orchids. The roots of most terrestrial orchid species are normally myco-heterotrophic, even as adults and several have been studied because of their interesting relationships with fungi (Brundrett, 2006). DNA has been isolated directly from orchid roots or from pelotons. PCR primers designed to amplify

sequences characteristic of the *Basidiomycota* have then been used to test for the presence of fungi. The primers typically target regions within the rRNA ITS region. This has been used in many orchids such as in North America *Cephalanthera* (Taylor and Bruns, 1997), *Corallorhiza sp.* (Taylor and Bruns, 1999) and *Hexalectris* (Taylor et al., 2003). In Denmark, this has been used on *Dactylorhiza* (Kristiansen et al., 2001) and *Neottia* in the United Kingdom, Germany and in France.

4.1.3 Identification fungi from orchid roots:

Numerous fungi have been isolated from the roots of orchids and have been identified as *Rhizoctonia*-like fungi. The members of this group are unable to form asexual spores and all have similar distinctive vegetative characters (see Chapter 1, section 1.4.2). Six *Epulorhiza* species were described depending on the shape and the dimensions of monilioid cells (a constriction at the branch point, and a septum in the branch hypha near its point of origin, with chains of inflated hyphae): *E. repens* (Moore, 1987); *E. albertensis*, *E. anaticula* and *E. inquiline* (Currah et al., 1990, Currah et al., 1997); *E. calendulina* (Zelmer and Currah, 1995); and *E. epiphytica* (Pereira et al., 2003). Also, six *Tulasnella* species with *Rhizoctonia*-like anamorphs have been described as orchid mycorrhiza: *Tulasnella allantospora*; *T. asymmetrica; T. calospora* (anamorph = *E. repens*); *T. cruciata*; *T. irregularis*; and *T. violea*. Three *Rhizoctonia*-like fungi from the roots of adult orchids from Brazil have been identified by amplifying the ITS region and comparison with other sequences found in the NCBI database (Nogueira et al., 2014).

Many attempts have been made in order to isolate and identify mycorrhizal fungi from orchids using the ITS region. For example, eight isolates of *Ceratorhiza pernacatena* from the family *Ceratobasidiaceae* have been isolated by sampling roots from seven plants of *P. praeclara* from a single population in Manitoba, Canada and characterized by Zelmer and Currah (Zelmer and Currah, 1995). Also, they isolated *Tulasnellaceae* and fifteen *Ceratobasidiaceae* observed in 11 plants of the same orchid (Zelmer et al., 1996), while Sharma et al. (2003) assessed mycorrhizal fungi from 21 plants of *P. praeclara* in Minnesota and Missouri. Moreover, 78 isolates of fungi have been found that belongs to family *Ceratobasidiaceae* and nine isolates belonged to family *Tulasnellaceae* from a rare North American orchid (Tovar, 2015). Amplification of rRNA from mycorrhiza of

terrestrial orchids of Europe and North America and a tropical mycorrhizal fungus from Asia have been sequenced (Kristiansen et al., 2001).

4.1.4 Methods for isolation and estimation of fungi in soil:

Isolation of fungi from soil using culture media is problematic because of the small number of species that can be established *in vitro*. Nevertheless, laboratory cultivation methods have been used extensively for qualitative and quantitative estimation of fungi in soil. The dilution plate method (Johnson et al., 1959) has been used for quantitative estimation of soil fungi e.g. Rao (1970). This involves preparing a soil suspension and, after suitable dilution, plating this onto petri dishes. After incubation, the fungal colonies are counted, and identified when possible. A high count can indicate something as simple as the presence of one organism that has sporulated, and may not relate to a role played in the soil.

A related method that has been widely adopted is the soil plate method (Warcup, 1950). In this, a very small amount of soil is placed in a sterile Petri dish and a cooled agar medium is added. After incubation, the number of fungal colonies can be counted and individuals picked off and transferred to fresh sterile medium for further study.

4.1.5 Growth of fungi in culture:

4.1.5.1 Effect of different culture media and temperature

Many media are used for growing fungi with effects on the colour and morphology of the colony. In addition, growth in culture and sporulation, a key characteristic for morphological identification, may require specific conditions. It is important to try several kinds of media to identify fungi in culture. Many fungi succeed on Potato Dextrose Agar (PDA), while for other fungi this is too rich and results in excessive mycelial growth instead of forming spores (Hibbett and Thorn, 2001).

Even though Corn Meal Agar (CMA) is a weak medium compared to PDA, numerous fungi that can be isolated from soil grow well on CMA. The fungi that can assimilate cellulose keep their ability do so when grown on a weak medium like Water Agar (WA) or Potato Carrot Agar (PCA) plus sterile filter paper (Hibbett and Thorn, 2001).

It has been difficult to isolate many fungi from infected tissues, infested soil or decaying organic material because of rapid development of fungi or bacteria with faster growth rates. Thus, for successful isolation of these kinds of fungi, it is necessary to surface sterilize the material and use selective media such as Water Agar that either slow or inhibit the growth of antagonistic organisms (Hibbett and Thorn, 2001). The growth rate slows which makes it easier to isolate the target fungi. Also, Antibiotic Agar is another way to isolate fungi from samples, using streptomycin sulfate and penicillin. In addition, a high pH in the medium can allow fungal growth but inhibits the growth of bacteria (Hibbett and Thorn, 2001).

Many studies have been carried out on morphological characteristics of fungal colonies, for example transferring 9 mm diameter agar plugs from the border of colonies into petri dishes containing PDA, CMA, MEA, or OMA. After that, colony diameter, colour, and aerial mycelium can be recorded along with growth rates. As an example, Nontachaiyapoom et al. (2010) studied morphological characteristics of fungi by culturing isolates on PDA at 30°C in the dark and measured diameters of fungal colonies every 2-3 days for fast-growing fungal isolates, and every week for at least 3 weeks for slow-growing fungal isolates.

4.1.6 Isolation of DNA of fungi from soil:

The soil is extremely complex in structure and a heterogeneous environment. Its structure and the soil microorganisms are characters which determine the functions of the soil in relation to water retention and transmission as well as C, N, P and K absorption and transformations to maintain soil fertility (Blaud et al., 2017).

Studying soil microbial communities has become more reliable after resolving problems related to isolating DNA of suitable quality for PCR and sequencing methods (Becker, 2011). However, DNA isolation can still be challenging. Two significant problems when isolating DNA from soil are ensuring breakage of cells and the presence of materials such as humic acid that co-extract with the DNA and inhibit PCR reactions. It has been found that these problems can be solved using a combination of physical and chemical treatments (Frostegård et al., 1999).

Physical methods are used for cell disruption, such as grinding, with or without freezing in liquid nitrogen, to break up mycelia (Al Nagi et al., 2013).

Homogenization is another method that has been applied to extract fungal DNA, by glass-bead-beating (Griffin et al., 2002). Many chemical methods have been used to remove humic acids from soil involving polyvinylpolypyrrolidone (PVPP), e.g. Cullen and Hirsch (1998), as well as application of additional chemicals such as 1% sodium dodecylsulphate, enzymes and 1 M salt concentrations, with additional heating and shaking (Edgcomb et al., 1999, Melo et al., 2006).

Protocols are available using chemical (SDS) or enzymatic lysis, in conjunction with the physical disruption that yields a high quantity and quality of DNA suitable for PCR amplification (e.g. Al Nagi et al., 2013). Commercial kits are also now available. For example, SoilMaster^{TM} DNA extraction kit (EPICENTER Biotechnologies) and PowerClean® Soil DNA Clean-Up Kit (MoBio Laboratories) have been used to extract DNA from forest soil. Both yielded DNA from the soil, but the Soil Master^{TM} DNA extraction kit needed a clean-up step and then re-amplification for a PCR product which the PowerClean® Soil DNA Clean-Up Kit did not require. However, all these methods are varied in their ability to extract DNA from different organisms in the soil (Becker, 2011).

4.1.7 Internal Transcribed Spacer (ITS):

The ITS region is part of the nuclear ribosomal DNA and situated between the SSU 18S and the LSU 28S rRNA coding regions (see Chapter 1 Figure 1.1). It is divided into ITS1 and ITS2, which are separated by the 5.8S rRNA coding gene. ITS1 is more variable than ITS2 as reported by several researchers (Froeschke and von der Heyden, 2014)

PCR of the ITS region has been applied to mycorrhizal fungi from plant roots since the 1990 (Gardes and Bruns, 1993). Often, DNA of plant and fungi are mixed, and therefore fungal DNA needs to be amplified specifically.

4.1.7.1 ITS primers and limitation:

Following the pioneering work of White et al. (1990) this region became established for molecular identification of fungi. This relied on the accumulation of DNA sequence records for this region from many fungal isolates, based on samples where identify was confirmed through conventional taxonomy. In addition, new primers were optimized for specific fungal species or groups. When the Fungal Barcoding

Consortium assessed the potential of candidate DNA regions, ITS stood out as the region most likely to give a successful identification and has been formally adopted by the Consortium for the Barcode of Life (www.barcodeoflife.org). The high variability of the ITS region can differentiate species or operational taxonomic units (Kumar et al., 2011) within closely related species (Tover, 2015). In addition, identification can be successful from small, diluted or degraded DNA samples (Gardes and Bruns, 1993).

This marker is found in all fungal cells and in multiple copies. It can be recovered by relatively strong primers with a constant record of reliability, across most fungal groups (Schoch et al., 2012). However, in some cases the individual copies within a genome do not match, causing difficulty in sequencing. Also, several databases of ITS sequencing do not meet formal barcode standards or are from unarchived strains (Lewis et al., 2011). In addition, the identification of fungal species can lack discrimination because of dependence on only this sequence locus, and the fine-scale of microsatellite markers have been proposed for studies of the ecological and evolutionary interaction between orchids and their mycorrhizal symbionts (Ruibal et al., 2013).

The standard ITS primers (ITS1, ITS4, ITS1F, and ITS4B) can be used to amplify fungal DNA from pure cultures (Table 4.1), but fungal DNA amplification directly from root tissue can fail, probably because plant DNA is at much higher concentration than fungal DNA. Other limitations of the standard ITS primers are that they not able to amplify some groups of fungi such as *Tulasnella*, an important orchid mycobiont (Bidartondo et al., 2003).

A primer pair for the ITS and 5.8S ribosomal DNA region (ITS1-OF/ITS4-OF) was developed to amplify all *Basidiomycota* with the objective to give a less biased view of orchid associations within the *Basidiomycota*. These are effective primers for all *Basidiomycota* and have been recommended for use with unknown orchid symbionts. Twenty-seven isolates of *Rhizoctonia*-like fungi have been identified from the root of three genera of orchids (*Paphiopedilum, Cymbidium,* and *Dendrobium*) by their use (Taylor and McCormick, 2008).

The alternative strategy, of designing primers that are more specific for groups known to be orchid symbionts has also been tried. Specific primers (Tul1/Tul4 and CeTh1/CeTh4) have been designed for the ITS region in *Ceratobasidium*, *Thanatephorus*, and *Tulasnella* (see Table 4.1). *Tulasnella* J. Schröt. (Family *Tulasnellaceae*, order *Cantharellales*) is an endophytic fungus that occurs in the roots, stems, or protocorms of a group of orchids. The actual number of *Tulasnella* species related to orchid genera has not been wholly resolved.

Table 4.1: Specific primers for identification of Basidiomycota fungi that have been used in previous studies.

Name	Fungi amplified	Reference
ITS1-OF/ ITS4-OF	All *Basidiomycota*	Taylor and McCormick 2008
ITS1/ITS4	*Basidiomycota*	White et al 1990
ITS1F/ITS4B	*Basidiomycota*	Gardes and Bruns 1993
Tul1/Tul4	*Tulasnella*	Porras-Alfaro and Bayman, 2007
CeTh1/CeTh4	*Ceratobasidium and Thanatephorus*	Porras-Alfaro and Bayman, 2007

4.1.8 Phylogenetic trees:

Phylogeny is the area of research concerned with finding evolutionary history and genetic relationships among various organisms or species. Modern phylogenetic analysis depends on information extracted from genetic material (e.g. DNA, RNA, or protein sequences). The result of these relationships is known as a phylogenetic tree (Potter, 2008).

Sequence alignment has been used in order to distinguish regions that may correspond to function, structure, or evolutionary relatedness. This generates multiple sequence alignments (MSA) aligning data from all individuals or samples in the study that are used to create a phylogenetic tree (Potter, 2008). The accuracy of the phylogenetic tree relies on an optimal sequence alignment. This requires a dynamic programming technique, and the most commonly used are heuristics methods, called progressive alignment. These include methods such as ClustalW2, MAFFT (McWilliam et al., 2013), Clustal Omega and T-Coffee (Notredame et al., 2000). In contrast, there are iterative methods such as MUSCLE (Edgar, 2004), which are more accurate than progressive methods. Both work similarly, but a

progressive alignment cannot reconsider decisions as it proceeds while an iterative method has the ability to return to the previous calculated pairwise alignments or sub-MSAs and correct errors. For example, MUSCLE is one of the iterative methods that are considered fast and accurate when compared with T-Coffee, MAFFT, and CLUSTALW (Felsenstein and Felenstein, 2004). MUSCLE was used for alignment to provide a phylogenetic tree in this project.

Many approaches have been used for estimating phylogenies such as neighbour-joining, maximum parsimony, UPGMA, Bayesian phylogenetic inference, maximum likelihood and distance matrix methods (Hellmuth et al., 2015, Minh et al., 2013, Nguyen et al., 2014). Molecular Evolutionary Genetics Analysis (MEGA) is a computer suite for conducting the analysis of molecular evolution and for constructing phylogenetic trees (www.megasoftware.net; (Kumar et al., 2016b)). It provides many methods and tools to use for phylogenomics (Tamura et al., 2013). It has been updated and expanded many times and MEGA7 is the most recent version.

4.1.9 High throughput sequencing:

With the development of high throughput sequencing (HTS), a view of the entire microbiome can be obtained from a single sample (Turnbaugh et al., 2007, Hamady and Knight, 2009). Illumina Miseq is the most common platform in current use for microbiome analyses from many different kinds of sample (e.g. water, soil, arthropods and human) (Zhou et al., 2016, Rubin et al., 2014, Nelson et al., 2014, Kennedy et al., 2014, Jervis-Bardy et al., 2015, Cong et al., 2015).

The principle of the Illumina platform was explained in Chapter 1, Section 1.5.1. Although Illumina technology is interesting, it is not able to provide a sequence for the whole ITS region (ITS1, 5.8S, and ITS2) with overlapping paired-end reads because the longest reads are only 2 x 300 bp. As a result, this is sufficient to sequence ITS1 and ITS2 of most fungal species, but it does not allow overlap of these reads in the 5.8S region (Bálint et al., 2014).

The Illumina MiSeq platform provides sequencing in more depth at a low price than the earlier 454 sequencing and this gives a deeper characterization of fungal communities. To understand these communities requires analysis of a large number of samples per study. Therefore, increasing the number of replicates of each sample

will allow evaluation of biases that are inherent in metabarcoding such as the randomness of individual PCR reactions and statistical power of data analyses (Schmidt et al., 2013).

In terms of computing analysis, various data processing pipelines have been suggested to process meta-barcoding data such as RDP (Cole et al., 2009), MOTHUR (Schloss et al., 2009), QIIME (Caporaso et al., 2010), PANGEA (Giongo et al., 2010), WATERS (Hartman et al., 2010), and CANGS (Pandey et al., 2010), but most of these tools are only suitable for prokaryotic sequences and are not straightforward to use with the fungal ITS metabarcode, although limited support for fungi already exist, for example within QIIME. The fungal ITS region has been selected by researchers (Schoch et al., 2012, Nilsson et al., 2010), because it has hypervariable and highly conserved regions. In order to not distort BLAST assignments during data processing, ITS regions have to be separated from the surrounding conserved regions (Nilsson et al., 2010). There are also some other pipelines which are specific for ITS fungi metabarcoding but were developed for 454 rather than Illumina sequences, such as CLOTU (Kumar et al., 2011), SCATA (http://scata.mykopat.slu.se/), PLUTOF (Abarenkov et al., 2010). Nevertheless, Illumina MiSeq has become widely used by fungal ecologists (Martin, 2016) and was used in this study with the QIIME pipeline.

It has been applied for sequencing of fungal ITS2 amplicons to study the fine-scale spatial distribution of orchid mycorrhizal fungi from the soil of two orchid-rich Mediterranean grasslands (Voyron et al., 2017). Much work of HTS has been focused on ECM and AM symbioses in forest and agriculture, although some knowledge on establishment and functioning of orchid mycorrhizal interaction has been obtained (Perotto et al., 2014, Fochi et al., 2017, Balestrini and Lumini, 2017).

4.1.9.1 QIIME:

Quantitative Insight Into Microbial Ecology (QIIME, pronounced ‘chime’ http://qiime.org) is an open-source software pipeline for analysis of metagenomic data and visualization from raw data input to databases. It contains a number of scripts written in the Python programming language and some scripts to incorporate other software such as BLAST. It can analyze raw data for functionalities such as

operational taxonomic units (OTU) and carry out sequence alignment to produce a taxon-based analysis of diversity within and between samples (Caporaso et al., 2010).

4.1.9.1.1 Alpha diversity:

Three levels of species diversity have been defined by (Whittaker, 1972). Alpha diversity refers to species diversity in sites or habitats (within an individual), beta diversity compares species among sites or habitats, and gamma diversity represents the diversity across an entire landscape. Alpha and beta diversity determine gamma diversity.

Alpha and beta diversities are two major tools within the QIIME package for estimating diversity. They are independent of each other in terms of community weights (Jost, 2006). The main step that to consider in alpha diversity is the number of sequences in each dataset that will have an effect on the diversity estimation of species (e.g. fewer sequences lead to lower species diversity). Therefore, rarefaction is the first step that must be used for normalizing the datasets to account for uneven sequencing depth. The input file for this process in the QIIME pipeline is the OTU table and the output is a folder containing many tables representing random subsamples from the original OTU table.

However, there are several different diversity metrics for calculating alpha diversity such as PD_whole_tree, chao1, and observed_otus. PD_whole_tree and observed_otus metrics are using in phylogenetic distances to perform alpha diversity, while non-phylogenetic measures can use the chao1 metric or can use multiple metrics in single file for alpha diversity **(**http://qiime.org/scripts/alpha_diversity.html). The output of the alpha diversity process will be rarefaction_plots in the html file.

Chao1 is an available metric in QIIME and was used in this project (Robert and Jonathan, 1994, Chao, 1984).

4.1.9.1.2 Beta diversity:

Beta diversity can use for the comparison between samples. Also, it has different metrics for calculating a distance between a pair of samples such as weighted

UniFrac, unweighted UniFrac (Hamady & Knight, 2009), and for non-phylogenetic distance, Euclidean distance or Bray-Curtis (Bray and Curtis, 1957). There are two optional input files for this step, using the OTU table or the rarefaction file from alpha diversity. In both ways, phylogenetic beta diversity should use UniFrac metrics and the phylogenetic tree file as input plus the OTU table. The output is a distance matrix visualized as a dissimilarity value for each pairwise comparison.

Statistical tests such as ANOVA, Kruskal-Wallis, G-test, and parametric T-test have been used for comparing the diversity values of alpha or beta diversity to analyze microbiome data. For example, ANOVA and Kruskal-Wallis have been used for comparing two growing environments (*ex situ* and *in situ*) to compare alpha and beta diversity of orchid mycorrhizal fungi. This indicated that both environments had similar fungal symbionts with terrestrial orchid *Platanthera chapmanii* (Kaur et al., 2018).

4.1.9.2 Limitations of technology that affect microbiome analysis:

The main limitation of HTS technology is distortion in the microbiome profile. Contamination from laboratory procedures and molecular laboratory reagents can impact on the microbiome data that are produced (Salter et al., 2014). Controls and samples to account for any contamination or false-positive signals can be included (Chaisiri, 2016). Also, it may be possible to include a positive control which is an artificial microbial community to reduce noise and identify biases in microbiome projects (Chaisiri, 2016).

4.2 Materials and Method:

4.2.1 The study sites

There were three sample sites in grass-covered areas on the campus of the University of Liverpool, as shown in Figure 4.1. All were mown as routine by the University's Facilities Management, with the mowings removed, but were otherwise not subject to maintenance treatments. *O.apifera* leaves form a flattened rosette so can continue vegetative growth within a mowing regime, although leaves can be damaged.

Site A (corner of Crown Street and West Derby Road) was within an area that, following demolition of housing in the 1970s, had been subject to minimal human intervention apart from grass mowing. Soil was allowed to form naturally over the building rubble following addition of pulverised domestic rubbish (personal communications, AD Bradshaw, MG Jones). Prior to new building construction on this location from 2009 onwards (before the start of this research project), *O.apifera* plants from the site were transferred to pots on the roof of the Life Sciences Building. These plants were maintained with their original soil and surrounding vegetation. They are referred to as Roof Plants samples.

Site B was within the original Liverpool Royal Infirmary complex. This building was constructed in the 1890s, closed in 1978 and gradually restored to use by the University of Liverpool from the mid-1990s. The grassed areas are now mown regularly but not otherwise maintained. The location chosen for study was adjacent to Dover Street and thus samples are referred to as Dover Street.

Site C was immediately outside the Biosciences Building within the same area as site A. Following completion of this building in 2003, grass turf was laid. This is mown but not fertilised and has developed a very diverse sward. For the duration of this study selected plants were enclosed with low fencing to prevent mowing and allow the plants to flower. This also facilitated identification of individuals. Surrounding vegetation was cut by hand during the study to maintain appearance. Samples from this site are called Ground Biosciences.

4.2.2 Isolation of fungi from root and soil:

The strategy for isolating fungi from soil and roots is summarized in Figure 4.2.

4.2.2.1 Sample collection

Roots and soil were obtained from 20 orchid plants at different times of the year starting from September 2014 and ending in August 2016. One or two roots and 10 g soil were removed at a depth of 2-8 cm with a sterile trowel which was sterilized with 70% ethanol. The root and soil samples were stored in sterile universal tubes, returned to the laboratory, and used immediately or kept at -80 ◦C for later use.

Figure 4.1: Location study sites. Sites A, B and C on north campus University of Liverpool (University campus map, www.liverpool.ac.uk/maps). (A) Site A in 2009, prior to new building construction. (B) Site B, adjacent to Dover Street. (C) Site C, ground outside the Biosciences Building.

4.2.2.2 Culture medium:

Four media were used to isolate fungi from soil and roots as shown in Table 4.2 namely PDA and MEA from (Sigma-Aldrich (UK)), FIM (Clements et al., 1986), and WA (Zelmer and Currah, 1995). All media were autoclaved at 121ºC, 108 kPa and 30 min to sterilise. Three antibiotics were used namely tetracycline hydrochloride and streptomycin sulphate (Duchefa biochemie, Netherlands), and

penicillin G potassium salt (Sigma Aldrich) as stock solutions of 50 mg / ml following filter-sterilisation using filter disc size 0.45 µm (Millipore Express PES Membrane, MILLEX-HP Filter Unit) and then stored at – 20 ◦C. These antibiotics were added to petri dishes containing the sterilized medium under aseptic conditions to a final concentration of 5mg/ml in order to minimise bacterial contamination.

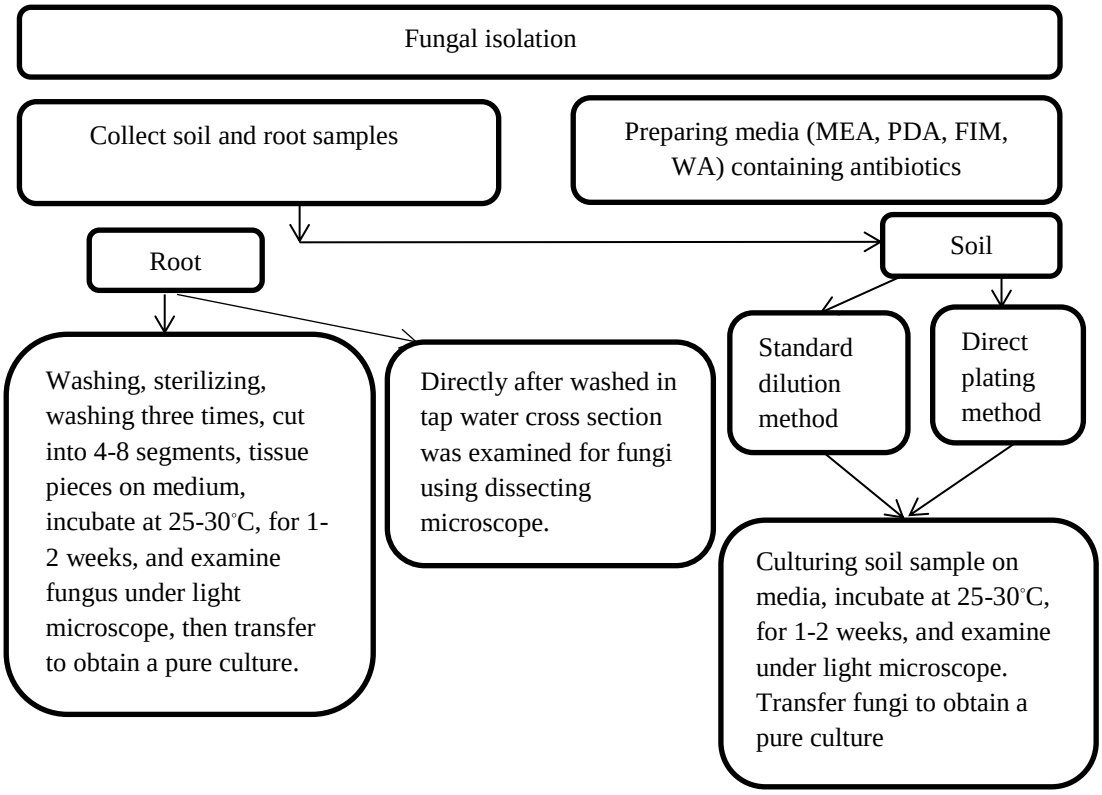


Figure 4. 2: Summaries of methods used to isolate fungi from root and soil.

4.2.2.3.1 Microscopic examination:

Roots were washed under tap water and a small part was cut in cross section and examined under a compound microscope (Novex B LED series, 110-240V / Holland).

4.2.2.3.2 Isolating fungi from roots:

Part of the root was sterilized, first in 3% sodium hypochlorite for 1 min, and then in 70% ethanol for 2 min and finally washed three times in sterile distilled water. The root was cut into 4-8 pieces and placed on MEA, PDA, FIM, and WA (Table 4.2) containing antibiotics, then incubated at 25-30 ºC for 1-2 weeks in the dark.

Following stereo microscopy examination, any fungal colonies were transferred to PDA with antibiotics for identification.

Table 4.2: Media used for isolating fungi from root and soil.

Media	Composition (per litre)	Source	Reference
Malt Extract Agar (MEA)	Malt Extract Powder (30g)	Sigma-Aldrich UK	(Waller et al., 2002)
Potato Dextrose Agar (PDA)	Potato Dextrose Broth (24g), Agar (8g)	Sigma-Aldrich UK	(Amiri et al., 2009)
Water Agar (WA)	Agar (10 g), Water.		(Zelmer and Currah, 1995).
Fungal Isolation Medium (FIM)	$Ca(NO_3)_2$. $4H_2O$(0.5g), KCl (0.1g), KH_2PO_4(0.2g), $MgSO_47H_2O$ (0.1g), Yeast Extract Powder (0.1g), Sucrose (5g), Agar (10g).		(Clements et al., 1986)
Coconut Milk Agar (COMA)	5% Coconut Milk and 2% Agar.	Lidl, UK; Sigma-Aldrich UK	(Davis et al., 1987, Dyer and McCammon, 1994).
Corn Meal Agar (CMA)	17 g	Sigma-Aldrich, UK	(Gonçalves et al., 2006)
Oat Meal Agar (OMA)	Basic Oat Agar (8.6g) or Oatmeal (60g), Agar (12.5g)	Sigma-Aldrich, UK	(Waller et al., 2002)
Luria-Bertani agar (LB)	37 g containing yeast extract (5.0g), peptone from casein (10.0 g), sodium chloride (10.0 g), agar (12.0 g).	Sigma-Aldrich, UK	(Mamiatis et al., 1985)

4.2.2.3.3 Culturing fungi from soil:

Two isolation methods were applied, standard dilution plate and direct plating. Three dilutions, 1:10, 1:100, and 1:1000, were used for each sample of soil suspension initially 0.1g /ml. Thus, to each plate was added 0.1ml of soil suspension with the three antibiotics plus 20 ml PDA then gently swirled. Plates were incubated at 30 ◦C for at least two weeks. There were three replicates for each condition. For the direct plating method, a small amount (25 mg) of soil was spread on semi melt PDA containing the three antibiotics. Plates were assessed under a dissecting microscope (Novex B / Holland) every two weeks and single colonies of fungi were picked off to new medium (PDA with antibiotics) to allow isolated colonies to develop.

4.2.3 Fungal molecular characterization:

Fungal molecular methods include DNA isolation, PCR amplification and sequencing DNA fragments (Figure 4.3).

4.2.3.1 Extraction of DNA from fungal cultures:

Genomic DNA was extracted from the fungal cultures using the procedure of (Doyle, 1991, Jeewon et al., 2004, Lacap et al., 2003). A small portion of mycelium (2 x 0.5 cm square regions) was scraped with a sterile spatula and transferred to 1.5 ml Eppendorf tube with disruption Bio-beads (180 µm) and 1 ml CTAB buffer (2 g CTAB, 8.12 g NaCl, 1.2 g Tris-HCl, 0.37 g EDTA, pH 8.0 dissolved in 100 ml dH_2O and then autoclaved). The mycelium was disrupted using a PowerLyzer® 24 Bench Top Bead-Based Homogenizer (MO-Bio) at 3,000 rpm for 90 sec. The suspension was then incubated at 65℃ for 10 minutes and then centrifuged at 15000 *g* for 2 min.

The supernatant of 700 µl was transferred to a new 2 ml tube containing 2 µl RNase (100 mg /ml) and incubated at 37℃ for 30 min. Phenol: chloroform: isoamyl alcohol (v/v/v, 25:24:1) was added to the tube and vortexed briefly to mix. The mixture was then centrifuged at maximum speed for 2 min. Then 600 µl of the aqueous phase was transferred to a new 2 ml tube and the phenol:chloroform:isoamyl alcohol extraction repeated. An equal volume of isopropanol was then added to the aqueous phase and the mixture inverted few times before incubating on ice for 5 minutes. After that, the mixture was centrifuged at maximum speed for 5 min and the liquid was removed, retaining the pellet. The pellet was washed twice with 70% ethanol and then air dried on the bench for 5 min and then re-suspended with 100-150 µl nuclease-free water. Then, the DNA was used for PCR or stored at -20℃.

4.2.3.1.1 DNA quantification:

DNA concentration was measured using 1 µl samples at 260 nm using NanoDrop-1000 (Thermo Scientific).

4.2.3.1.2 PCR reaction:

PCR was used to amplify the characteristic ITS region using two sets of primers, ITS1with ITS4 and ITS1-OF-A / ITS1-OF-B with ITS4-OF (Table 4.3) (Taylor and McCormick, 2008). Each reaction contained 10 µl of 2x BioMix™ Red (Bioline), 2 µl (25 µM) forward and 2 µl (25 µM) reverse primers, 4 µl of dH_2O, and 2 µl DNA

templates (0.1µg). The PCR condition was: Pre-denaturation at 95℃ for 5 min, followed by 35 cycles of a denaturation step at 95 ºC for 30 sec., annealing temperature for 1 min (see Table 4.3) and extension at 72℃ for 1 min, finishing with one step of extension for 6 min at 72℃, and then maintaining the products at 4 ºC until analysis.

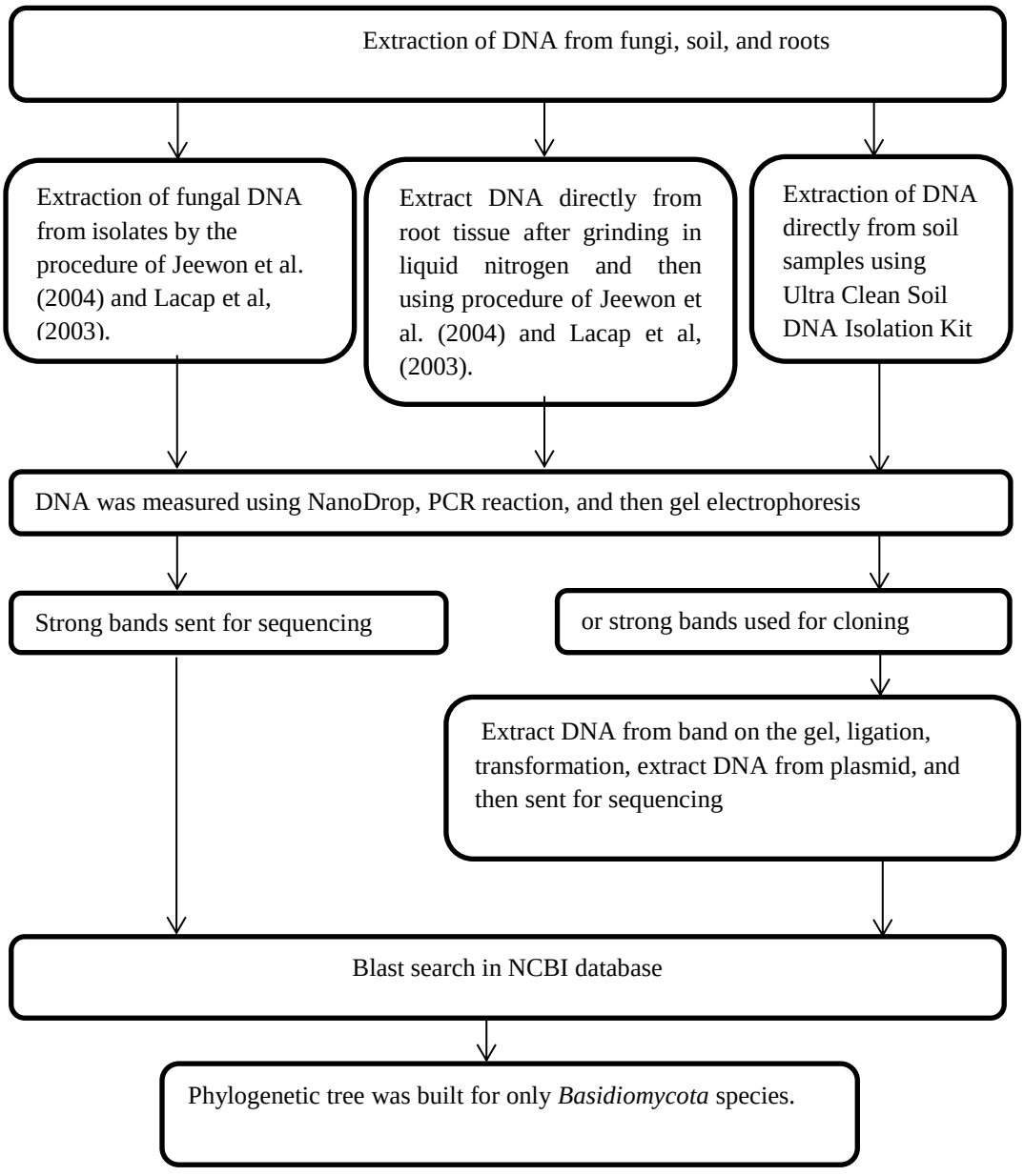


Figure 4.3: An overview of methods used for molecular characterization of fungi directly or indirectly.

4.2.3.1.3 Gel electrophoresis:

Samples (5 µl) were separated on a 1% agarose gel containing TAE (0.4 M Tris-acetate, 1mM EDTA) buffer. Midori Green 3 (Nippon Genetics Europe GmbH, Germany) was used for staining the gel. PCR products were separated on the gels at 100 V for 45 min. A 1 kb ladder (5 µl, Bioline 1kb hyperladder) was used as a marker and images were captured under UV (U-Genius, Syngene, UK). DNA bands which were more bright and clear were chosen to cut out and send for sequencing (Source Bioscience, Nottingham, UK). BLAST searches were used to seek matches with known fungal species.

Table 4. 3: General fungal ITS primer sequences used for amplification of DNA fragments in this chapter. F-forward primer; R = reverse primer.

Primer	Direction	Sequence	Annealing temp (℃)	References
ITS1-OF-A ITS1-OF-B	Mix as Forward	5-AACTCGGCCATTTAGAGGAAGT-3 5-AACTGGTCATTTAGAGGAAGT-3	54.7	(Taylor and McCormick, 2008)
ITS4-OF	R	3-GTTACTAGGGGAATCCTTGTT-5		(Taylor and McCormick, 2008)
ITS1	F	5-TCCGTAGGTGAACCTGCGG-3	55	(White et al., 1990)
ITS4	R	3-TCCTCCGCTTATTGATATGC-5		(White et al., 1990)
ITS5	F	5-GGAAGTAAAAGTCGTAACAAGG-3	55	(White et al., 1990)
5.8S-fungi	R	3-CAAGAGATCCGTTGTTGAAAGTT-5		(Epp et al., 2012)

4.2.3.1.4 Effect of culture media on growth of fungi:

R3 (*Uncultured Tulasnellaceae*) was isolated from *O. apifera* root through culture as described in section 4.2.2.3.2, and B was obtained as isolate B1 from the Hardy Orchid Society (www.hardyorchidsociety.org.uk/HOS%201012/Cultivation.html **).**

Isolates B and R3 were tested on five media: MEA, WA, COMA, OMA, and CMA (Table 4.2). Each was prepared separately and then autoclaved at 121^{o}C for 30 min to sterilize, and supplemented with antibiotics as described in section 4.2.2.2. After initial incubation of the isolates on these media, a cork borer (0.5 cm) was used to cut samples from the edge of the colony. Ten replicates were incubated for each medium at 26 ◦C and 20 ◦C for 3-4 weeks in the dark. Colony diameter was recorded every 7 days using ruler.

4.2.3.2 Identification of fungi from soil and roots using non-culture methods:

Non-culture methods involved isolation of DNA from orchid roots and soil. Several methods were tested to obtain DNA in sufficient quantity and quality for sequence analysis. PCR followed by sequencing was attempted using general fungal primers from Table 4.3 and also primers specific to orchid mycorrhizal fungi (Table 4.4). When this was not entirely successful, PCR products were cloned prior to sequencing. Finally Illumina MiSeq was used for sequencing.

4.2.3.2.1 Extraction of DNA directly from root samples:

Root tissue from site B which had been stored at -80 ºC before DNA extraction was used. A portion of the tissue was ground in liquid nitrogen using a mortar and pestle to produce a powder and then 1 ml CTAB buffer was added. All other steps were followed the procedure in section 4.2.3.1.

4.2.3.2.1.1 PCR reaction using DNA isolated from root samples:

The PCR reaction mixture (20 µl) contained 10 µl 2x BioMix^{TM} Red (Bioline), 2 µl (100 µM) forward and 2 µl (100 µM) reverse primers (listed in Table 4.4), 4 µl dH_2O, and 2 µl DNA templates (0.1µg). The PCR conditions were: Pre-denaturation at 95℃ for 5 min, followed by 35 cycles of a denaturation step at 95 ºC for 30 sec., gradient annealing temperature was used for 1 min (see annealing temperature for each primer pair in Table 4.4) and extension at 72℃ for 1min finishing with one step of extension for 6 min at 72℃, and then maintaining the products at 4 ºC until analysis. The PCR products were separated on 1% (w/v) agarose gels and stained with 3 µl Midori Green. Gels were viewed under UV light and images recorded using a U-Genius Imager (Syngene Imager). DNA bands were chosen and sent for sequencing (Source Bioscience, Nottingham, UK).

4.2.3.2.2 Extraction of DNA from soil samples:

Soil samples from sites B and C and the pots of the roof plants were collected and stored at – 80 ºC prior to DNA extraction (Table 4.5). The soil samples (0.5 g), frozen in liquid nitrogen, were ground using mortar and pestle previously sterilized in the autoclave to produce a fine powder that was stored at – 80 ◦C until it was used for DNA extraction. After testing several methods (see results section 4.3.3), the

UltraClean Soil DNA Isolation Kit (MO BIO, version 03222005) was selected to isolate DNA from soil following the manufacture's protocol.

Table 4. 4: Orchid mycorrhizal specific primer pairs. F - forward primer; R = reverse primer.

Primer pairs	Primer		Sequences / Amplicon size (bp)	Annealing temp (◦C)	Reference
Tul1/Tul4	Tul1	F	5-ACGTTAAGGTGCTCTGGT**C**GAGG-3 /(23)	61.65	(Porras-Alfaro and Bayman, 2007)
	Tul4	R	3-ATGAGGTCATGCGT**T**GTAGTA CC-5 / (23)		
Tul1/Tul4	Tul1	F	5-ACGTTAAGGTGCTCTGGTTGAGG-3 / (23)	60.6	
	Tul4	R	3-ATGAGGTCATGCGTTGTAGTA CC-5 / (23)		
MLin3/ML6	MLin 3	F	5-CGACACAGGTTCGTAGGTAG-3 / (20)	55.15	(Porras-Alfaro and Bayman, 2007)
	ML6	R	3-CAGTAGAAGCTGCATAGGGTC-5 / (21)		
ML7/ML8	ML7	F	5-GACCCTATGCAGCTTCTACTG-3 / (21)	53.85	(Porras-Alfaro and Bayman, 2007)
	ML8	R	3-TTATCCCTAGCGTAACTTTTATC-5 / (23)		
ITS1/ITS4-tul	ITS1	F	5-TCCGTAGGTGAACCTGCGG-3 / (19)	59.05	(White et al., 1990)
	ITS4-Tul	R	3-CCGCCAGATTCACACATTGA-5 / (20)		(Taylor and McCormick, 2008)

4.2.3.2.2.1 DNA quantitation: The DNA concentration was measured using both NanoDrop^{TM}1000 Spectrophotometer (Thermo, USA), and Qubit (Thermo Fisher Scientific, USA).

4.2.3.2.2.2 Amplification of DNA template prior to cloning:

The PCR reaction mixture contained: 2x BioMix^{TM} Red (Bioline) 12.5 µl, 1.5 µl (25 µM) of each primer ITS1-OF-A forward +ITS1-OF-B forward / 1.5 µl (25 µM) ITS4-OF reverse (Table 4.3), bovine serum albumin 2.5 µl (10 mg / ml, purified BSA, Sigma Aldrich), sterile distilled water 2.5 µl, and DNA template 3 µl (0.1µg). The PCR conditions were as follows: pre-denaturation at 95◦C for 5 min, followed by 35 cycles of denaturation at 95◦C for 30 sec, annealing at 54.7◦C for 1 min, and extension at 72◦C for 1 min, followed by a final extension for 5 min at 72◦C. Samples were then held at 4 ◦C until tested on gel electrophoresis.

4.2.3.2.2.3 Agarose gel analysis of DNA:

The PCR products (25 µl) were separated at 100 V for 1 h on 1% agarose (w/v Bioline) dissolved in 1x TAE buffer, with 3 µl Midori Green to visualize. A DNA

ladder (1 kb, Bioline, 5 µl), was used as a marker. The DNA fragments were visualized under UV light and recorded using a U-Genius Imager (Syngene Imager).

4.2.3.2.2.4 Cloning DNA fragments:

A- Extraction of DNA from the gel:

After obtaining a suitable sized band on the electrophoresis gel following DNA amplification by PCR, the band was cut from the gel while illuminated with UV light using a sterile scalpel. The MinElute Gel Extraction Kit (QIAGEN® kit) was used to extract the DNA from the gel, according to the manufacturer's protocol. The DNA was re-suspended in 10 µl SDW.

The DNA quantity was measured using NanoDrop-1000 (Thermo Scientific). Finally, the eluted DNA was tested on 1% agarose gel in 100 ml 1X TAE buffer. The Midori Green (3 µl) was used for staining to visualize the DNA under UV light using the U-Genius (Syngene Imager) system.

B- Ligation of DNA fragments into pGEM-T Easy Vector:

The vector pGEM-T Easy (Promega) which has an ampicillin resistance gene to select and maintain the plasmid within *E. coli* was used for cloning the PCR products. The ligation reagent (10 µl) was prepared in a 1 ml Eppendorf tube containing 5 µl of 2X Rapid Ligation Buffer, pGEM®-T or pGEM®-T Easy Vector (1 µl, 50 ng), DNA insert 3 µl, and T4 DNA Ligase (3 Weiss units/µl) 1 µl. The ligation was performed according to manufacturer's protocol, and incubated at 4°C overnight.

C- Transformation using ligation reaction product:

The transformation was prepared according to manufacturer's instructions. A volume of 50 µl competent cells (DH5α, Thermo) from frozen stock and LB liquid medium (950 µl) was added to tubes containing the ligation reaction (10 µl). They were incubated in orbital incubator (200 rpm) at 37°C for 1.5 hour.

During this time, the LB solid medium 250 ml was prepared. This contained 250 µl ampicillin (50 µg /ml), X-Gal 250 µl (20 mg / ml), and IPTG 500 µl (100 mM). Petri dishes containing 20 ml each of LBA medium were poured and left for half an hour to solidify before use. Finally, 100-200 µl of the transformation mixture was spread

over the entire surface of each plate and incubated overnight at 37˚C.

D- Extraction of plasmid DNA:

After incubation, the plates contained blue and white colonies. The white colonies should have the inserted DNA in the plasmid and this was confirmed using PCR. So, white colonies were inoculated into universal tubes (50 ml) containing 5 ml LB broth with 40 µl (50 µg /ml) ampicillin. The tubes were incubated overnight at 37 ◦C. Then 1.5 ml was transferred to an Eppendorf tube and centrifuged for 1 min at 13000 *g*. The liquid was removed gently without disturbing the pellet in the bottom.

The manufacturer's protocol for the GeneJET Plasmid Miniprep Kit (Thermo Scientific) was then followed to isolate the plasmid. Re-suspension solution (250 µl) at 4 ◦C was added to the pellet and vortexed to mix. At the end of the protocol, the DNA plasmid was re-suspended in 15-20 µl sterile distilled water and stored at -20 ◦C.

E-Testing DNA plasmid using PCR:

The DNA plasmid was tested using PCR. Reactions contained 5 µl BioMix Red (Bioline, UK), forward primers (ITS1-OF-A+ITS1-OF-B 1 µl (25 µM)), and reverse ITS4-OF 0.5 µl (25 µM), sterile distilled water 2.5 µl, and 1 µl DNA plasmid. The PCR condition was as followed: pre-denaturation at 95◦C for 5 min, followed by 35 cycles of denaturation at 95◦C for 30 sec, annealing at 54.7˚C for 1 min, extension at 72◦C for 1 min, followed by a final extension for 5 min at 72◦C. Then, 9 µl of the PCR product was separated on a 1% agarose gel and stained with Midori Green and assessed under UV. All DNA plasmids containing an insert were sent for sequencing (Source Biosciences Company). A BLAST search against the NCBI database was used to match with known fungal species.

4.2.4 Sequence alignment and phylogenetic analyses:

All fungal DNA sequences obtained from soil and root either from culture or non-culture methods were submitted to a Blast search (http://www.ncbi.nlm.nih.gov/genbank/). Alignments were performed using MUSCLE within the Mega7 suite prior to building a phylogenetic tree (Kumar et al., 2016b). After alignment and trimming of the sequences at least 98% similarity was

chosen for identification. Sequences from *Tulasnellaceae, Ceratobasidaceae,* one species of *Sebacina,* and one species of *Ascomycota* were downloaded from NCBI and included in the analysis of fungi from soil and root. The Neighbour-Joining method was used and the branch was calculated by bootstrap testing (1000 replicates are shown next to branches) (Saitou and Nei, 1987). The Jukes-Cantor method was chosen for the evolutionary distances (Jukes et al., 1969).

4.2.5 Next generation sequencing:

4.2.5.1 Samples used:

DNA was extracted from the 61 samples listed in Table 4.5. The DNA extraction and quantification were described in Section 4.2.3.2.2 and 4.2.3.2.2.1.

4.2.5.2 PCR amplification:

All soil samples were tested using PCR before sending them to the Centre for Genomic Research (CGR), Liverpool for sequencing. The 10 µl reactions contained: BioMix Red (Bioline, UK) 5 µl, primers ITS5-forward 1 µl (100 µM) and 5.8S-Fungi-reverse 1 µl (100 µM), Bovine serum albumin (Purified BSA, Sigma Aldrich) 1 µl (10 mg/ml), sterile distilled water 1 µl, and DNA template 1 µl (0.1µg). The PCR conditions were as follows: pre-denaturation at 95◦C for 5 min, followed by 25 cycles of denaturation at 95◦C for 30 sec, annealing at 55◦C for 1min, and extension at 72◦C for 1min, followed by a final extension for 5 min at 72◦C and then incubated at 4 ◦C until analysed.

4.2.5.3 Agarose gel electrophoresis:

A 2% agarose gel was prepared in 1X TAE buffer. DNA was stained with 5 µl Midori Green (NIPPON Genetics EUROPE). Then, 10 µl of PCR product was loaded into the agarose gel wells. A 100 bp size ladder (Bioline) was used. Gels were run at 100 volts for 1 hour. The DNA bands within the gels were visualized under UV light and the image of the gel was taken using the U-Genius (Syngene Imager) machine.

4.2.5.4 Sample selected for Illumina sequencing:

After extraction of DNA from soil using Ultra Clean Soil DNA Isolation Kit and from one root sample using Fungal protocol (section 4.2.3.1),these extracts were selected for Illumina sequencing.

4.2.5.5 Library preparation of DNA from soil samples:

A 96 well plate was prepared containing 1 µl in each well of the DNA samples from soil and root selected for Illumina sequencing. The primers were used ITS5 and 5.8S-fungi, with master mix. The CGR performed two rounds of PCR, the first round 10 cycles and the second round 15 cycles using conditions as in section 4.2.5.2. Added water 10 µl for clean up and to increase the volume of first round PCR and second round was without any addition. The PCR products were then sequenced using the Illumina MiSeq platform.

4.2.5.6 Bioinformatics:

Three samples (S3, S6, and S7), one from each site, did not give data after Illumina sequencing. After obtaining the raw data sequences from the Centre for Genomic Research, it was processed as summarized in Figure 4.4.

4.2.5.6.1 Pre-processing data:

4.2.5.6.1.1 Quality filtering of paired-end raw data sequences:

Quality filtering of the raw data was already done by the CGR using Cutadapt v.1.2.1 (Martin, 2011), and Sickle v.1.200 (Joshi and Fass, 2011), before down stream analysis. The reads in fastq format were trimmed to remove those with low quality scores (lower than 20), adapters from Illumina sequencing and any with a length shorter than 10 bp.

There were three files for each sample, forward (R1) and reverse (R2) reads of the paired-end, and R0 (singlet reads). The latter were excluded from further steps. Table 4.6 demonstrates that two of these files (R1 and R2) were ready for downstream analysis and each of them contained sample name, for example (1-S1), barcode TAAGGCGA-CTCTCTAT, lane 1 (L001), forward read (R1) or reverse (R2), file 1 (001), file in the fastq format (fastq), and compressed in gzip (.gz).

Table 4. 5: List of samples used for non-culture methods.

Site	ID Samples	Plant	Date collection	DNA concentration in ng/µl	Ratio	
					260/280	260/230
Dover St	S1	1	22/03/2015	113.26	1.76	1.57
	S2	1	22/03/2015	98.92	1.74	1.42
	S14	1	22/03/2015	71.62	1.85	2.12
	S17	1	22/03/2015	61.91	1.77	1.54
	S26	2	22/03/2015	94.93	1.83	1.85
	S27	2	22/03/2015	103.88	1.81	1.76
	S50	2	22/03/2015	118.85	1.80	1.66
	S33	3	22/03/2015	82.95	1.89	2.26
	S40	3	22/03/2015	84.10	1.83	1.64
	S48	3	22/03/2015	130.59	1.89	1.89
	S61	4	22/03/2015	164.49	1.85	1.92
	S13	1	14/10/2015	78.15	1.88	2.29
	S22	2	14/10/2015	108.79	1.79	1.85
	S24	3	14/10/2015	107.58	1.84	1.83
	S38	3	14/10/2015	112.16	1.83	1.73
	S52	3	14/10/2015	143.61	1.85	1.88
	S59	4	14/10/2015	217.26	1.86	1.97
	S62	4	14/10/2015	123.37	1.73	1.60
	S54	2	17/12/2015	151.80	1.77	1.61
	S29	2	17/12/2015	34.13	1.76	1.64
	S37	3	17/12/2015	41.02	1.67	1.33
	S39	3	17/12/2015	124.47	1.84	1.69
	S32	3	17/12/2015	99.93	1.92	1.75
	S63	3	17/12/2015	138.61	1.67	1.40
	S16	6	17/12/2015	103.30	1.80	1.91
	S19	6	17/12/2015	132.13	1.89	2.29
	S18	9	17/12/2015	102.60	1.89	1.85
	S4	1	05/03/2016	47.7	1.87	1.40
	S23	2	05/03/2016	89.18	1.86	1.69
	S30	2	05/03/2016	49.56	1.85	1.27
	S57	2	05/03/2016	92.40	1.73	1.64
	S25	3	05/03/2016	76.27	1.78	1.66
	S31	3	05/03/2016	77.98	1.83	1.53
	S34	3	05/08/2016	102.71	1.83	1.80
	S55	3	05/08/2016	142.59	1.84	1.82
Biosciences	S46	4	19/04/2015	139.01	1.88	1.85
	S8	4	08/05/2015	89.19	1.82	1.62
	S47	4	08/05/2015	150.11	1.89	2.03
	S9	1	19/09/2015	119.60	1.85	1.94
	S45	1	19/09/2015	112.79	1.87	1.75
	S10	1	27/09/2015	44.64	1.72	1.43
	S60	1	27/09/2015	157.88	1.83	1.85
	S44	2	27/09/2015	124.87	1.85	1.74
	S11	-2	27/09/2015	90.98	1.88	2.64
	S43	3	27/09/2015	97.29	1.87	1.63
	S20	3	27/09/2015	158.25	1.90	1.83
	S35	3	05/08/2016	81.61	1.88	1.62
	S53	3	05/08/2016	161.95	1.88	1.95
Roof	S21	22	21/10/2014	81.31	1.88	1.78
	S28	22	21/10/2014	110.36	1.85	1.92
	S41	22	21/10/2014	91.88	1.87	1.89
	S42	22	21/10/2014	122.95	1.96	2.07
	S5	4	19/03/2015	123.41	1.87	2.01
	S49	4	19/03/2015	131.09	1.81	1.83
	S15	1	25/06/2015	119.32	1.90	2.03
	S12	2	25/06/2015	35.56	1.95	4.09
	S56	1	21/10/2015	148.59	1.83	1.96
	S58	1	21/10/2015	154.67	1.88	1.97
	S36	2	05/08/2016	94.52	1.93	2.34
	S51	2	05/08/2016	75.52	1.90	1.95
Root	S64	2	05/08/2016	48.90	1.82	1.47

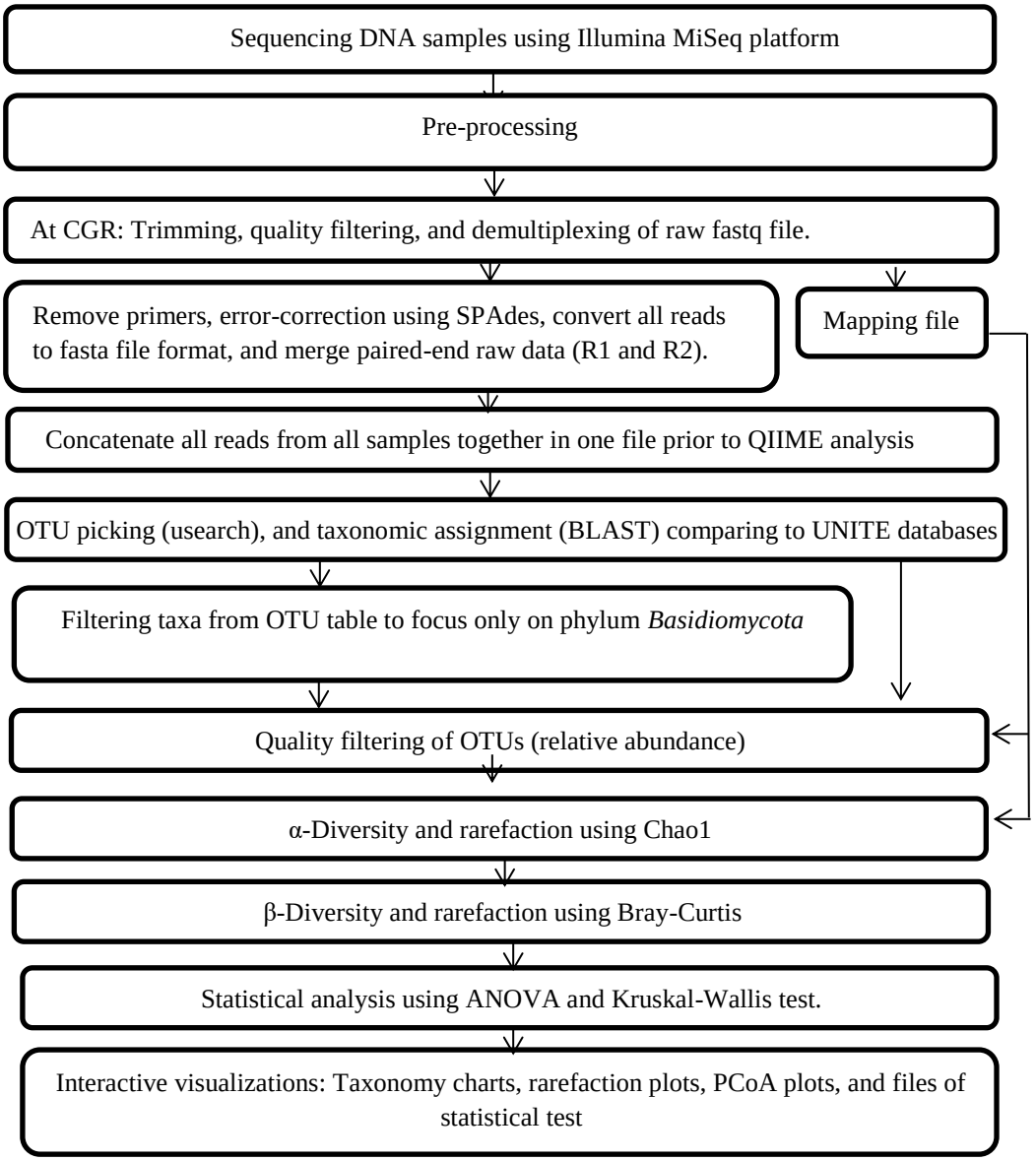


Figure 4. 4: Workflow of bioinformatics analysis of data filtering with QIIME

Table 4. 6: The format of the two files that were generated by the CGR ready for processing.

1-S1_TAAGGCGA-CTCTCTAT_L001_R1_001.fastq.gz

1-S1_TAAGGCGA-CTCTCTAT_L001_R2_001.fastq.gz

A statistical summary and boxplot were produced by CGR to show the number of reads in each sample (Figure 4.5), as well as the distribution of trimmed reads length in each paired-end (Figure 4.6).

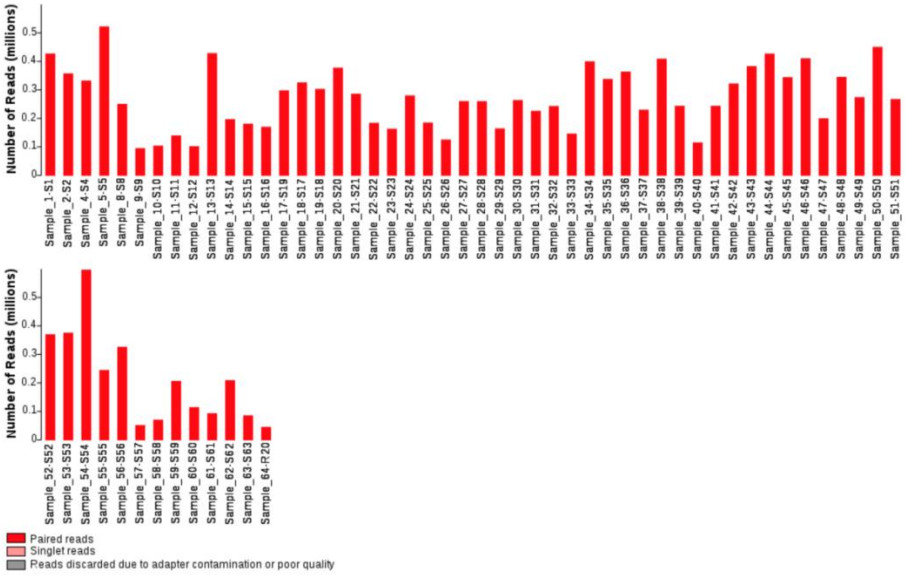


Figure 4. 5: Total number of reads per sample provided by CGR.

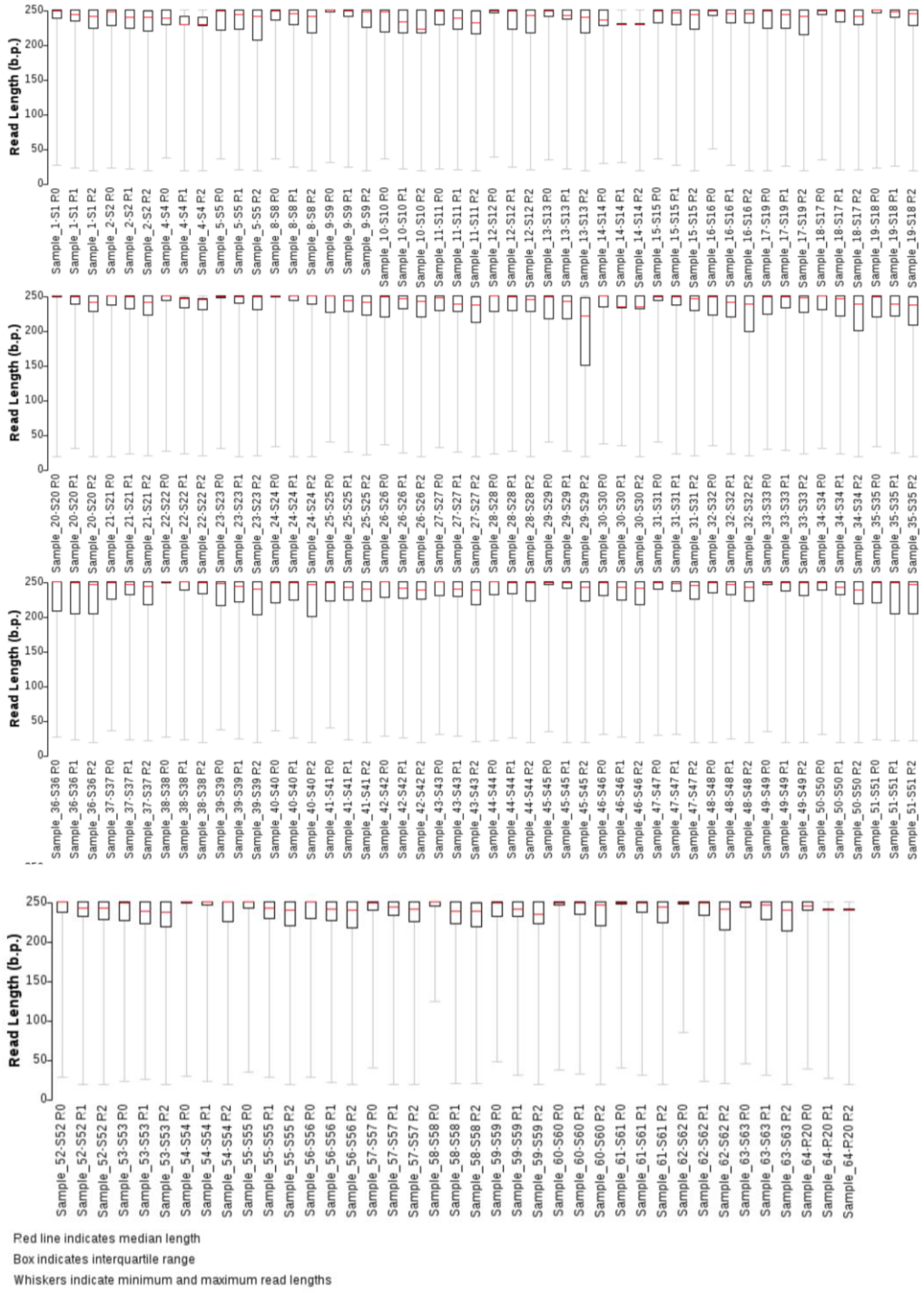


Figure 4. 6: Boxplot showing the length distribution of trimmed read R1 (forward), R2 (reverse), and R0 (singlet) reads. A long tail means that the sample is more distributed (Figure provided by CGR).

4.2.5.6.2 Data analysis:

QIIME version 1.9.0 (http://qiime.org, Caporaso et al. (2010)), was used for analysis of the read sequences to define Operational Taxonomic Units (OTUs) using the OTU picking method with USEARCH61 (Edgar, 2010). All sequences were clustered at 97% similarity (pick_otus.py), and in this step, reads were binned against a reference collection of the UNITE database (version 7.2 release date June 28, 2017) (Kõljalg et al., 2013).

This is a curated rDNA sequence database for *Ascomycota* and *Basidiomycota*. It only includes high-quality sequences of well-identified fungi. Any reads that did not match with the reference database were subsequently clustered using a de novo method against each other with the same similarity threshold. Then, the output file was used to pick a set of representative sequences (pick_rep_set.py), and the output file was then assigned to taxonomy using a BLAST algorithm through QIIME, and the UNITE database as a reference on QIIME (assign_taxonomy.py). All other steps of analysis within QIIME were performed under non-phylogenetic constraints. An OTU table was made (make_otu_table.py) using the output file from previous step in BIOM format (Biological Observation Matrix - http://biom-format.org).

At this stage of the analysis, further steps of downstream analysis were continued in two directions. Firstly, to continue steps to summarize the taxonomy data for each sample and plot the results, as well as alpha (within a sample) and beta diversity (between sample) using BIOM file. Secondly, filtering BIOM file to get only phylum *Basidiomycota* using this script (filter_taxa_from_otu_table.py), and then further steps but using output file from this filtering.

Relative abundance was used with the mapping file to transform the OTU table to a fungal taxonomy table with raw reads counts of each OTU and bar charts. Also, alpha diversity was performed using four different scripts in QIIME (multiple_rarfactions.py, alpha_diversity.py, collate_alpha.py, and make_rarefaction_plots.py).

The reads in the OTU table were normalized by rarefaction depth per sample at 100,000 to prior comparative analysis. The rarefaction plots from alpha diversity

subsampling at 70,000 sequences depth was applied for beta diversity (multiple_rarefactions_even_depth.py). As a consequence, 14 samples were discarded, but this loss will not cause any major issue for the analysis because they are replicated in other different samples. Therefore, a new mapping file was created to exclude 14 samples from the list, and accordingly four columns header (BarcodeSequence, Description, SampleID, and Site) was determined through richness and diversity. Bray-Curtis is a dissimilarity measure between two samples (Bray and Curtis, 1957), and it is a popular metric that has been used in ecological studies. Finally, Beta diversity generated 2D principal coordinate analysis (PCoA) plots (make_2d_plots.py).

After filtering, the OTU table was normalized at several rarefaction depths (10,000, 70,000, 50,000, and 150,000) before comparative analysis to optimize the best value. Rarefaction subsampling at 70,000 sequences depth was applied as the highest percentage of samples OTU recovery rather than another depth. Beta diversity at rarefaction of 20,000 was used for comparing samples.

4.2.5.6.3 Statistical tests:

Non-parametric Kruskal-Wallis (KW) tests with ANOVA test were performed for comparing the alpha diversity of fungal OTUs within groups. The input files were the BIOM file and mapping file after rarefaction using the script (group_significance.py) from QIIME, or using the BIOM file after filtering and mapping file after rarefaction.

4.3 Result:

4.3.1 Microscopic examination of orchid plant:

Cross sections of roots collected from plants (Figure 4.7-A), were examined after lacto phenol staining on glass slides under cover slips. Light microscopy showed that fungal hyphae were clearly found in the cortical cells of root tissue (Figure 4.7-B).

4.3.2 Culture of fungi from roots:

Parts of root tissue were incubated on PDA medium and resulted in the growth of fungal colonies (Figure 4.7-C). To assess these fungi, a compound microscope was used (Novex, Holand) and monilioid cells were found as shown in Figure 4.7-D. These cells are typically seen in *Rhizoctonia*.

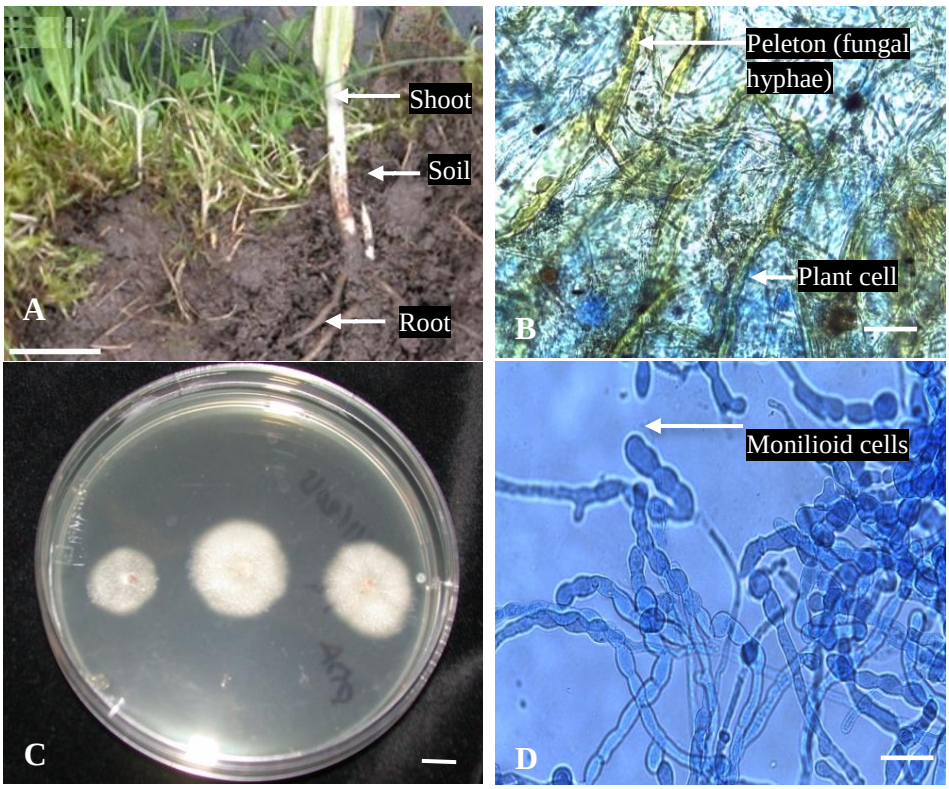


Figure 4.7: (A), Bee orchid showing roots and surrounding soil of Dover Street. (B), fungal hyphae (in yellow colour) associated with root cells (in blue colour). (C), Fungal R3 in pure culture. (D), Slide of fungus from pure culture. Scale bars: A 1.7 cm; B and D 100 µm; C 0.5 cm.

4.3.2.1Growth of two isolates on different culture media and temperature:

The growth of isolate B covered the whole plate on all media after three weeks at 26◦C (colony diameter 8.4 cm) and grew faster on COMA and OMA reaching the edge of the Petri dish in 14 days. It grew faster at 20◦C, covering the whole plate just after 14 days for all media as shown in Figure 4.8, 4.9, 4.10, and 4.11.

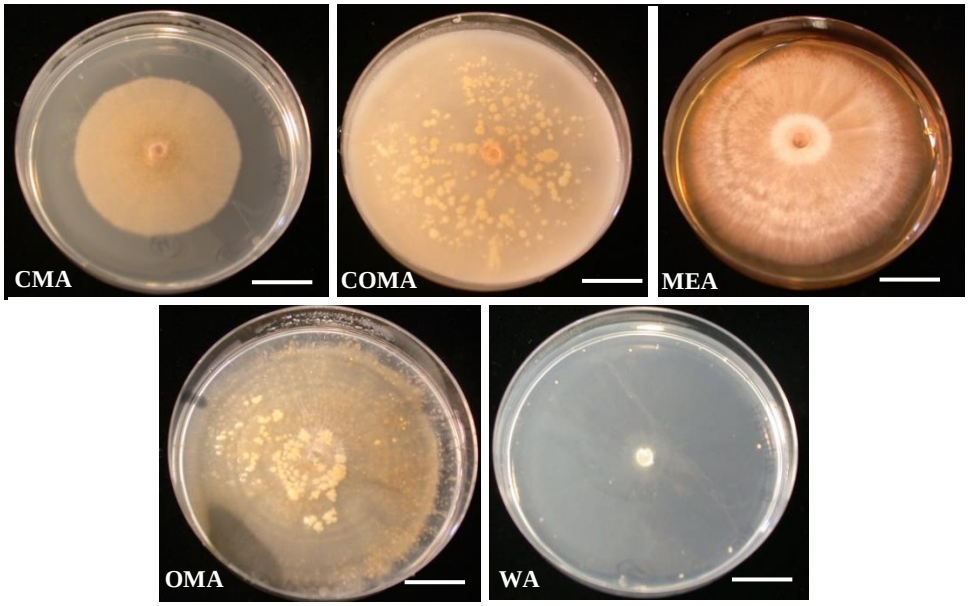


Figure 4. 8: Isolate B (Unidentified *Ceratobasidaceae*) on CMA, COMA, MEA, OMA, and WA at 26◦C for 21 days. Scale bar 1 cm.

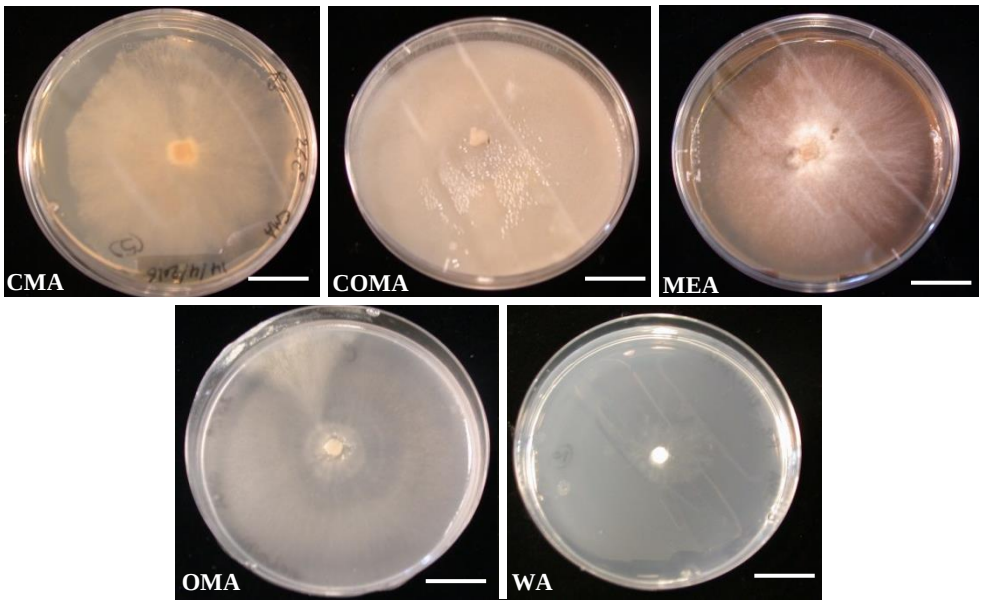


Figure 4. 9: Isolate R3 (Unidentified *Tulasnellaceae*)CMA, COMA, MEA, OMA, and WA at 26◦C for 21 days. Scale bar 1 cm.

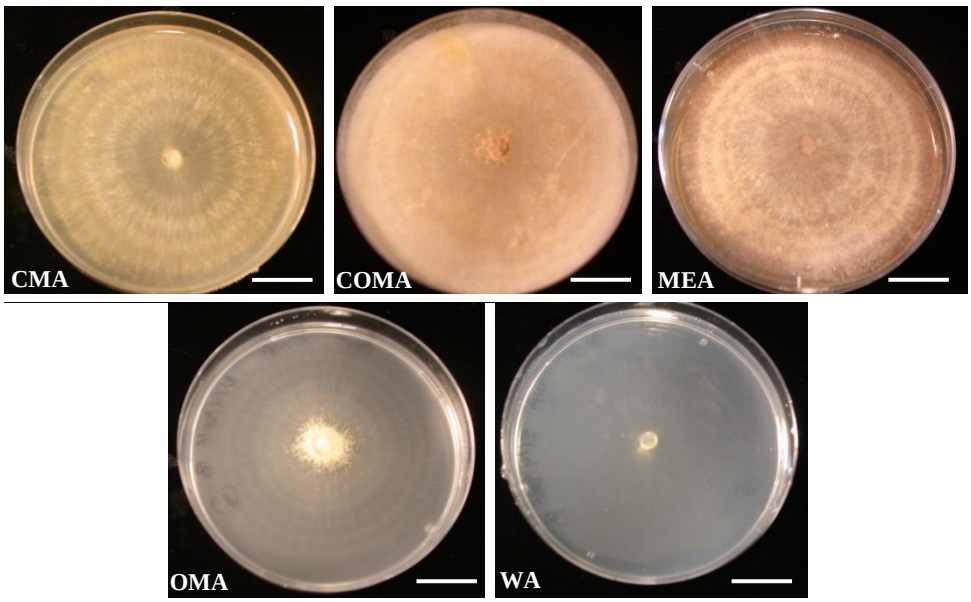


Figure 4. 10: Isolate B (Unidentified *Ceratobasidaceae*) on CMA, COMA, MEA, OMA, and WA at 20◦C for 21 days. Scale bar 1 cm.

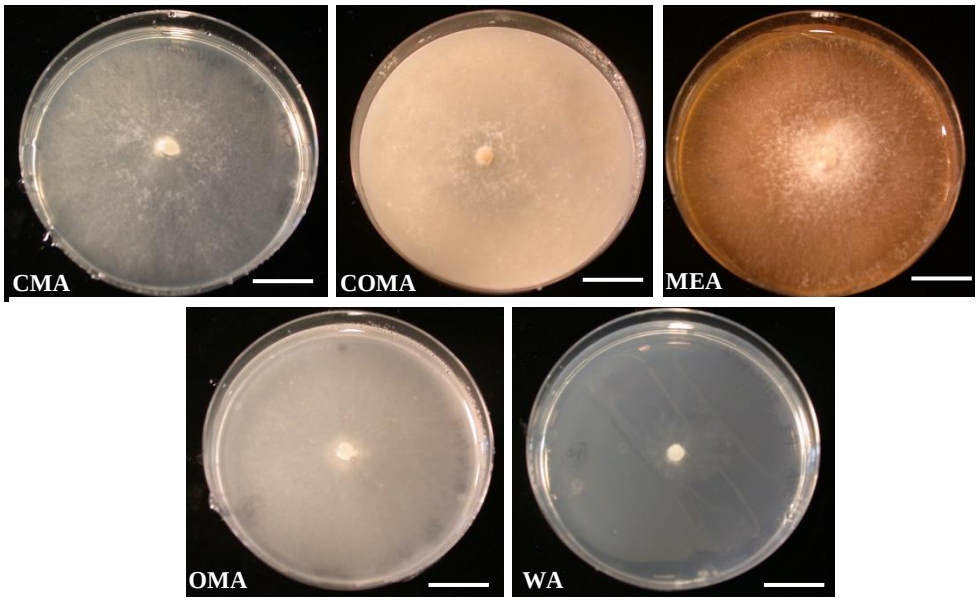


Figure 4.11: Isolate R3 (Unidentified *Tulasnellaceae*) CMA, COMA, MEA, OMA, and WA at 20◦C for 21 days. Scale bar 1 cm.

Growth of isolate R3 was slightly different and took a longer time. When incubated at 26◦C even after 37 days growth had not reached the edge of the plate for Figure 4.9, but at 20◦C this isolate also grew faster (Figure 4.11) and again was faster on COMA and OMA than the other media (Figure 4.12).

Overall, growth was more rapid with incubation at 20◦C of both isolates than at 26◦C. Isolate R3 grew poorly on WA when compared with iaolate B (compare Figure 4.12 A and B).

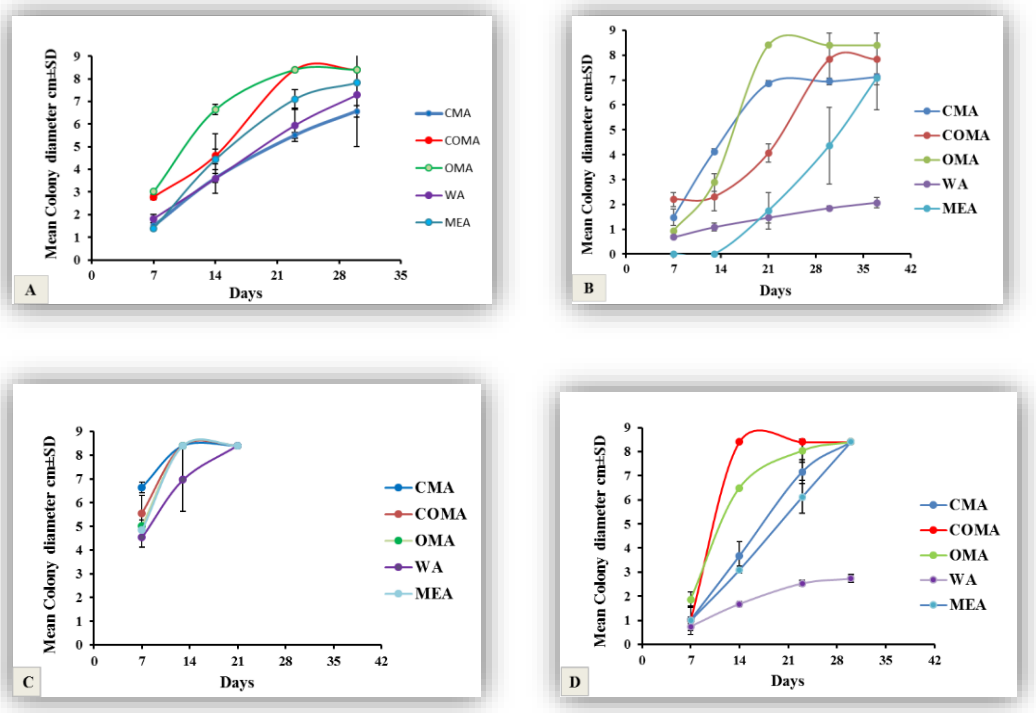


Figure 4. 12: Mean colony diameter of isolates B and R3 on CMA, COMA, OMA, MEA, and WA ±SD. A: isolate B, 26◦C; B: isolate R3, 26◦C; C: isolate B, 20◦C; D: isolate R3, 20◦C. N =10 for each medium.

4.3.2.2 Result of fungi from culture and non culture roots:

Many attempts to culture fungi were unsuccessful or yielded isolates outside the *Basidiomycota*. In total 19 isolates of fungi *Ascomycota*, 4 of *Mucoromycota* (*incertae sedis*), and 3 identified as uncultured environmental samples were identified after culture from soil or roots. Finally, three isolates (R1, R2, and R3) were obtained from roots and identified as *Rhizoctonia-like* fungi of the *Basidiomycota* group *Tulasnellaceae* (Table 4.7). They were identified by their ITS sequences.

DNA was also extracted directly from root (section 4.2.3.1) resulting in two sequences (R7OF and POF, Table 4.7) within the *Basidiomycota* (*Tulasnellaceae).* After that, two isolates (R2, R3), and the DNA extracted P (POF and PMLIN3) were tested with five pairs of primers that had been designed specifically to amplify DNA from this family (see Section 4.2.3.2.1.1). This could therefore identify the family typically found as an orchid symbiont.

Table 4. 7: Identification from DNA sequence obtained from root, and fungal isolates. BLAST results show the top hit matching the sequencing (NCBI accession number, query coverage in % and E-value are given).

Source Isolate or Root	Site	Method	Closest BLAST match	NCBI Accession number match	Query Coverage %	E-value
P-OF	Dover Street	Direct using liquid nitrogen from root	*Uncultured Tulasnellaceae*	KC243933	95	0
R1-OF	Roof plant	Indirect from culture	*Uncultured Tulasnellaceae*	KC243933	96	0
R2ML7	Dover Street	Indirect from culture	*Uncultured Tulasnella*	GQ907057	16	6e-19
R3-OF	Dover Street	Indirect from culture	*Uncultured Tulasnellaceae*	KJ188463	91	0
R7OF	Dover Street	Direct using liquid nitrogen from root	*Uncultured Tulasnella*	Ku902973	11	9e-24
PMLIN3	Dover Street	Direct using liquid nitrogen from root	*Uncultured Tulasnellaceae*	GQ250127	57	1e-100
R3 MLIN3	Dover Street	Indirect from culture	*Uncultured Tulasnellaceae*	GQ250149	29	8e-43

4.3.3 Isolation of DNA from soil:

In an attempt to extract DNA from soil, several methods were trialed before selecting the Ultra Clean Soil DNA Isolation Kit. All methods were applied to one soil sample S6 as shown in Table 4.8. Although fungal DNA was extracted using all these methods, only the Ultra Clean Soil DNA Isolation Kit provided DNA that could be amplified by PCR, as seen in Figure 4.13.

Table 4. 8: Comparing between methods used for extraction of DNA from soil sample S6.

Name of protocol	DNA concentration in ng/µl	Ratio		Reference
		260/280	260/230	
Rapid method	70.33	1.34	0.8	(Griffiths et al., 2000)
Fungal protocol (see section 4.2.3.1 Extraction of DNA from fungal cultures)	520.36	1.28	0.98	(Doyle, 1991, Jeewon et al., 2004, Lacap et al., 2003)
High-throughput method	46.95	1.35	0.62	(Whitlock et al., 2008)
Ultra Clean Soil DNA Isolation Kit	45.37	1.83	1.64	(MO BIO, version 03222005)

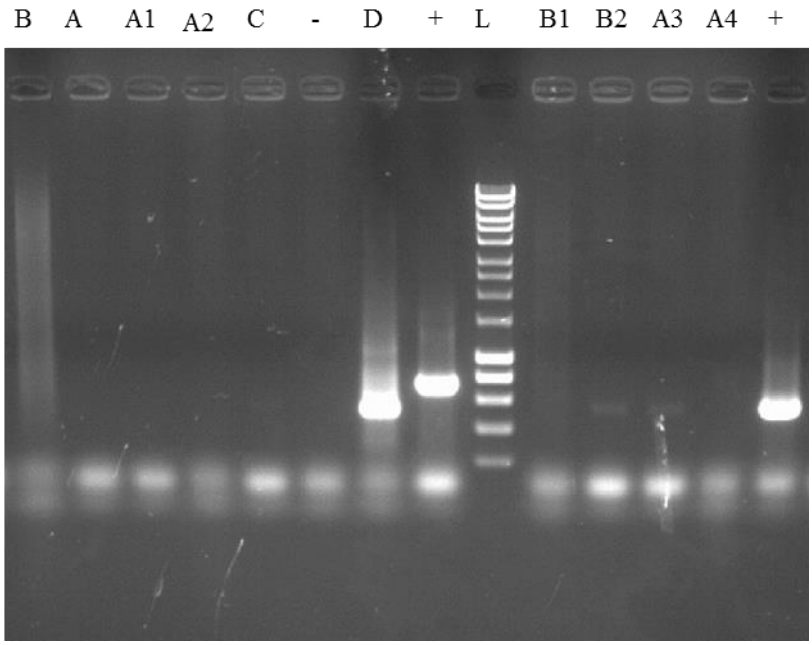


Figure 4. 13: PCR amplification following DNA isolation using four different methods from one soil sample (S6, Table 4.6). The PCR primers used were ITS1-forward and ITS4-reverse. A, A2 and A3: Rapid method; A1: Rapid method using PEG+isopropanol with filter; A4: Rapid method using filter; B and B2: Fungal protocol without filter; B1: Fungal protocol using filter; C: High-throughput method using filter; (-): negative control (water); D: Ultra Clean Soil DNA Isolation Kit; L: 1 kb ladder; (+ middle): Positive control of mushroom (*Agaricus bisporus*); and (+ right side): Positive control of *Aspergillus nidulans.*

4.3.3.1 Cloning of DNA isolated from soil:

Cloning techniques were used with DNA from one sample of soil (sample 6, sourced from roof plant) to assess fungal diversity. Following PCR, product purification, ligation and testing, twenty clones containing an insert of the appropriate size in the

plasmid were obtained. The inserts were sequenced (Table 4.9), showing that 16 came from the *Basidiomycota* (group: *Trichosporonaceae*), 2 from the *Ascomycota* and 2 matched to an environmental sample from an uncultured fungus.

Table 4.9: Sequences obtained from soil sample (6) using direct cloning method. BLAST results showing the top hit matching the sequence (Accession number, query coverage in % and E-value are given).

Clone ID	Closest BLAST match	Phylum	NCBI Accession number	Query Coverage %	E-value
B1	*Apiotrichum dulcitum*	*Basidiomycota*	KY101666.1	94%	0
B2	*Apiotrichum dulcitum*	*Basidiomycota*	KY101666.1	98%	0
B3	*Saitozyma podzolica*	*Basidiomycota*	KY102915.1	55%	0
B4	*Apiotrichum dulcitum*	*Basidiomycota*	KY101666.1	46%	0
B5, 6, 9, 10, 11, 12 and 16	*Apiotrichum dulcitum*	*Basidiomycota*	KY101666.1	44%	0
B7	*Apiotrichum dulcitum*	*Basidiomycota*	KY101666.1	50%	0
B8	*Apiotrichum dulcitum*	*Basidiomycota*	KY101666.1	42%	0
B13	*Apiotrichum dulcitum*	*Basidiomycota*	KY101666.1	58%	0
B14	*Apiotrichum dulcitum*	*Basidiomycota*	KY101666.1	66%	0
B15	*Apiotrichum dulcitum*	*Basidiomycota*	KY101666.1	61%	0
B17	*Ilyonectria liliigena*	*Ascomycota*	LT719143.1	100%	0
B18	*Uncultured Chaetomiaceae*	*Ascomycota*	GU055625.1	64%	0
B19	*Uncultured fungus*	environmental samples	AJ875386.1	52%	0
B20	*Uncultured fungus*	environmental samples	JF432993.1	100%	0

4.3.4 Phylogenetic analysis of DNA sequences from root, cultured *Basidiomycota* and soil obtained through cloning:

Most of the fungi isolated from roots in our study were closely related to each other and formed a well-supported group within *Tulasnellacea*, as shown in Figure 4.14. Sequences of R3MLIN3 and PMLIN3 from the MLin3/ML6 region were supported by a bootstrap of 99% and with R2ML7 from the ML7/ML8 region were related by 100% subclade (monophyletic group). Four subclades of R1, R3, P, and R7OF were amplified from the ITS1-OF/ITS4-OF region and were closely related to each other and related to the clade group of previous regions (MLin3/ML6 and ML7/ML8) by a bootstrap of 82%.

Isolate B (used as a control) grouped with *Ceratobasidiaceae* from a vouchered mycorrhizal specimen. More likely at genus level and are well supported by a bootstrap of 62% and with the group of *Ceratobasidium* at 45%.

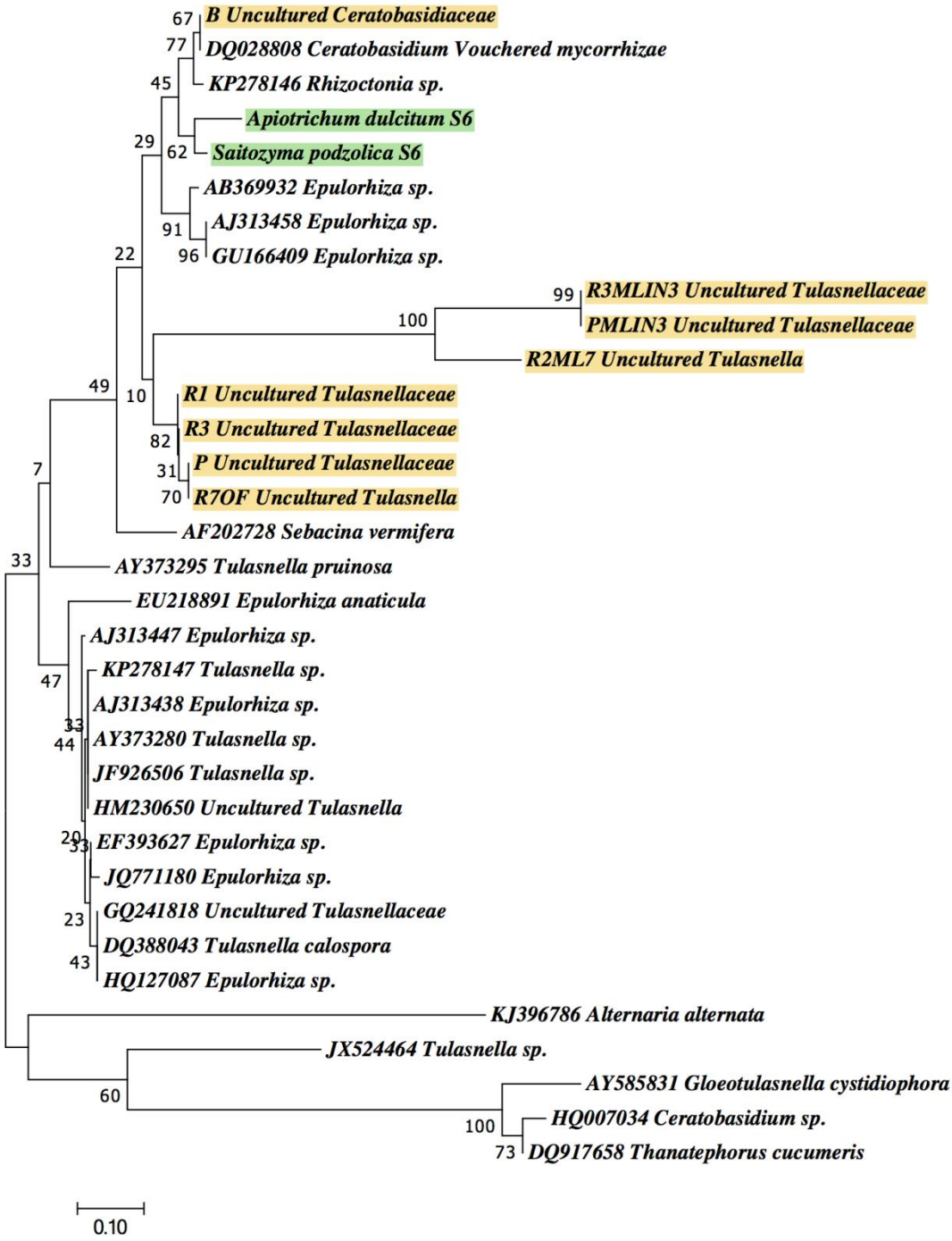


Figure 4. 14: Phylogenetic tree demonstrating relationships derived from sequences of the ITS region of DNA isolated from fungi and root (Table 4.7) with sequences of fungi obtained from GenBank. Numbers alongside branches correspond to values in 1000 bootstrap trials. The species highlighted in green were isolated from soil sample 6 using the cloning technique and the species in yellow indicates fungi isolated from root using both culture and non-culturing methods.

4.3.5 High throughput analysis of fungal DNA sequences from soil:

In order to obtain a broader picture of the fungal flora around the *O.apifera* plants, high throughput sequencing using the Illumina MiSeq platform was undertaken. Prior to submission for sequencing, all samples were tested for amplification by the selected primers (see Figure 4.15). The amount and purity of the samples was also estimated (Table 4.5). The results were satisfactory, although 3 samples subsequently did not provide any sequence data.

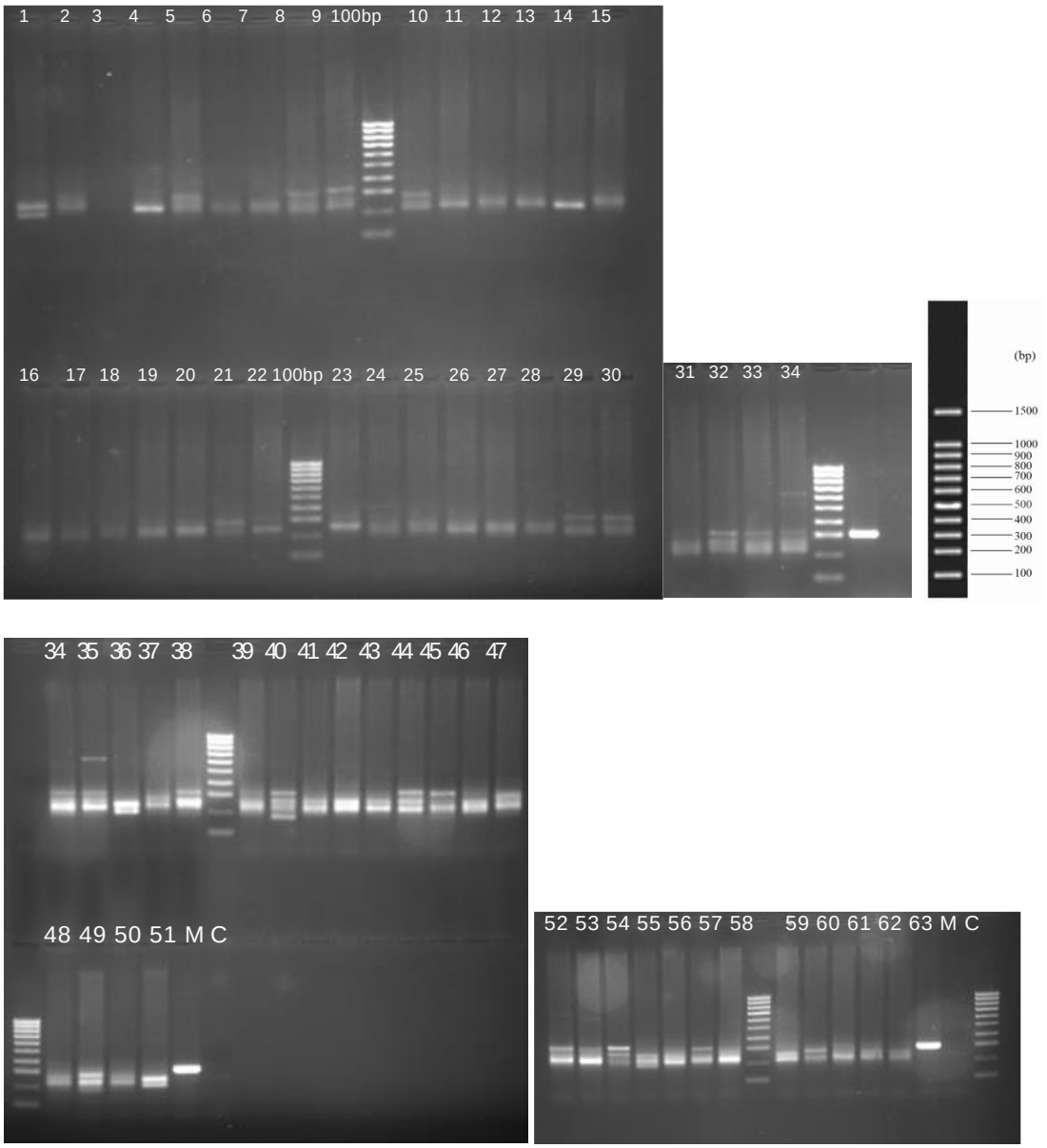


Figure 4.15: DNA bands from soil samples (1-63) using primers ITS5-forward and 5.8S-Fungi-reverse M: positive control (mushroom, *Agaricus biosporus*). C: negative control (water). Ladder (100 bp).

The total number of reads from the full set of samples, after quality filtering by CGR, was 15,773,682. Following several steps of pre-processing, the raw sequenced data was ready to be analyzed using the QIIME pipeline.

4.3.5.1 Taxonomy of fungi at phylum level:

Table 4.10 summarizes the relative abundance of sequences at the phylum level, before filtering. The *Ascomycota* comprised 54.6 %, the highest percentage in the entire fungal community. *Basidiomycota* made up 23.6 % in relative abundance. Other phyla occupied a lower percentage between 7.7 % and 0.1 %, with the *Glomeromycota* at 1.6 % relative abundance. Unidentified fungi comprised 5.9 %. This is reflected for each sample in Figure 4.16.

Table 4.10: Total percentage of relative abundance at phylum level for all samples.

Phylum	**Total in %**
Ascomycota	54.6
Basidiomycota	23.6
Chytridiomycota	3.5
Glomeromycota	1.6
Kickxellomycota	0.1
Monoblepharidomycota	0.1
Mortierellomycota	7.7
Mucoromycota	1.0
Neocallimastigomycota	0.1
Olpidiomycota	0.2
Rozellomycota	1.1
Unidentified	5.9

4.3.5.2 Taxonomy of Basidiomycota according to relative abundance:

Basidiomycota were filtered from other phyla of fungi for each level of taxonomy (class, order, family, genus, and species). The results are shown in Figures 4.16-4.21 and Table 4.11. The abundance at class level for each sample is shown in Figure 4.17. The relative abundance of *Agaricomycetes* made up 79.8 % of the entire *Basidiomycota* fungi in all samples, following by *Tremellomycetes* 11.9 % of relative abundance, while unidentified *Basidiomycota* occupied 1.5 %. The percentage of other classes was ranging from 1.8 % to 0.0 %.

At order level (see Figure 4.18 for abundance in each sample), *Agaricales* had the highest percentage at 51.2 %, followed by *Cantharellales* at 12.8 % relative abundance. Interestingly, a high value of this order (99.6%) was found in sample 64 (which is from root). *Trichosporonales and Thelephorales* occupied 6.1 % and 4.9 % respectively, while *Trechisporales* were at 3.3% with *Sebacinales* at only 1.3 % of the entire *Basidiomycota* fungal communities. The total percentage of relative abundance at order level in each sample is shown in Appendix 4.1.

As can be seen from Table 4.11 the *Clavariaceae* were the pre-dominant family of the *Agaricales* and the *Tulasenellaceae* of the order *Cantharellales* with 18.7 % and 11.5 % respectively. Also, another *Agaricales* family that had a high value of relative abundance was the *Psathyrellaceae* (9.2 %) as well as *Trichosporonaceae* (6.1 %) from the Trichosporales.

Focusing on the families associated with orchid mycorrhiza shown in Table 4.11, it is clear that the entire family *Tulasnellaceae* comprised a high proportion of abundance (11.5%), although genera were not identified. However, in the family *Ceratobasidaceae* at genus level, two genera *Ceratobasidium*, and *Clavulina* were identified along with a distinct group of unidentified genera. Family *Sebacinaceae* had one genus, *Sebacina* 0.2 % but the species could not be identified (see also Figure 4.21).

Overall, at species level many could not be identified through the data analysis.

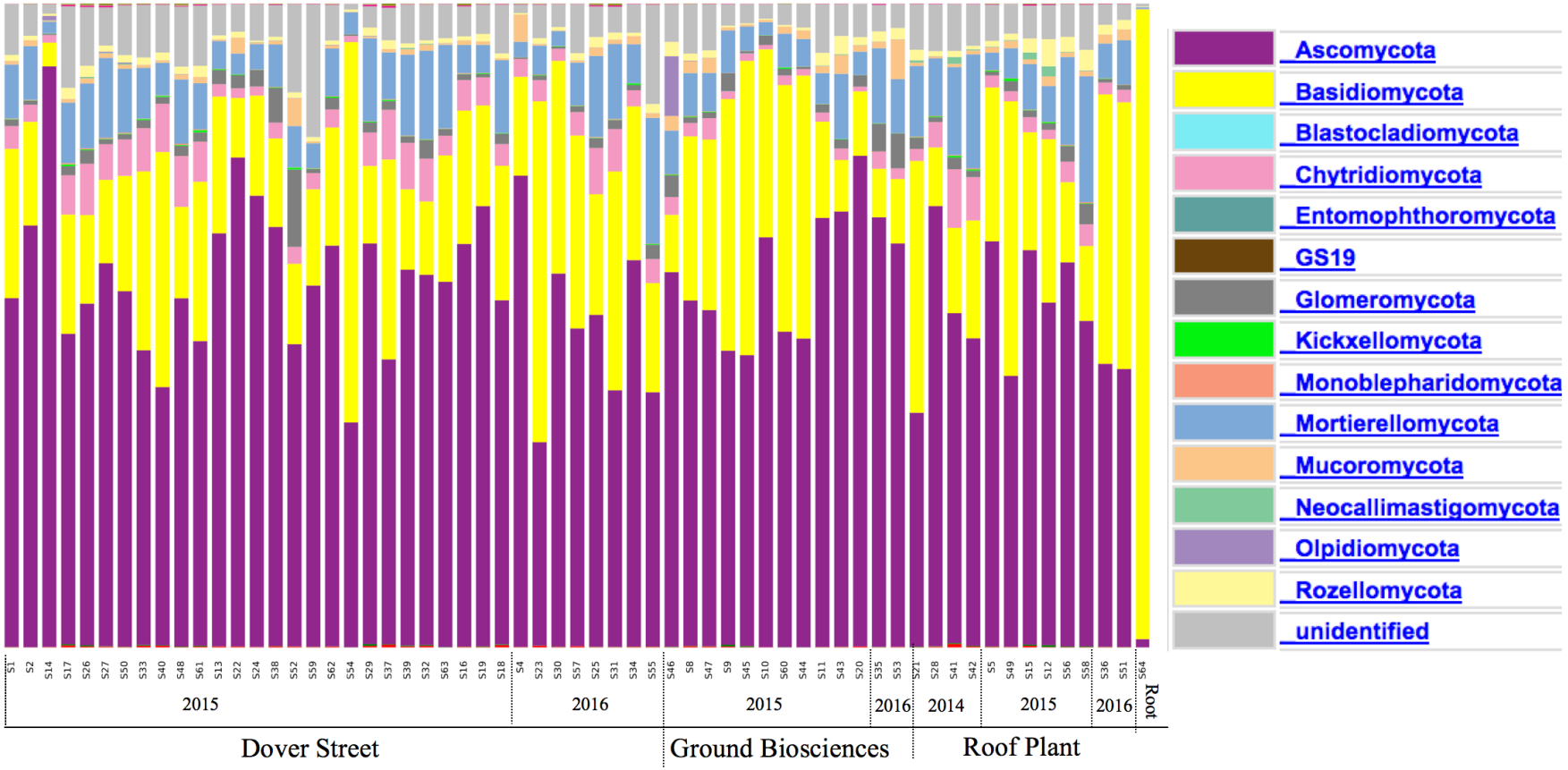


Figure 4. 16: The relative abundance of fungal OTUs taxonomy at phylum level before filtering. Samples are soil (S1-S63 from three sites) and one root sample (S64 from roof plant).

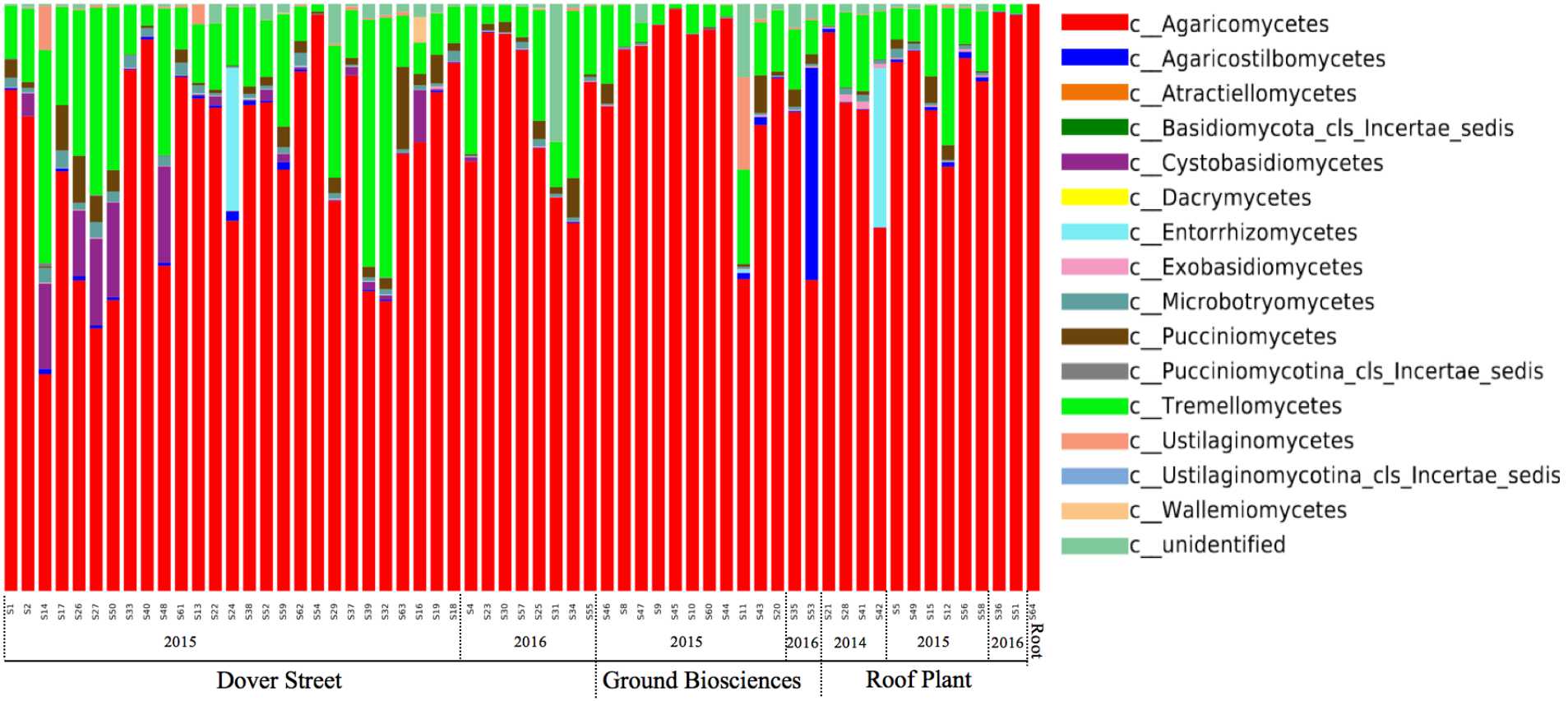


Figure 4. 17: The relative abundance of Basidiomycota OTUs taxonomy at class level after filtering.

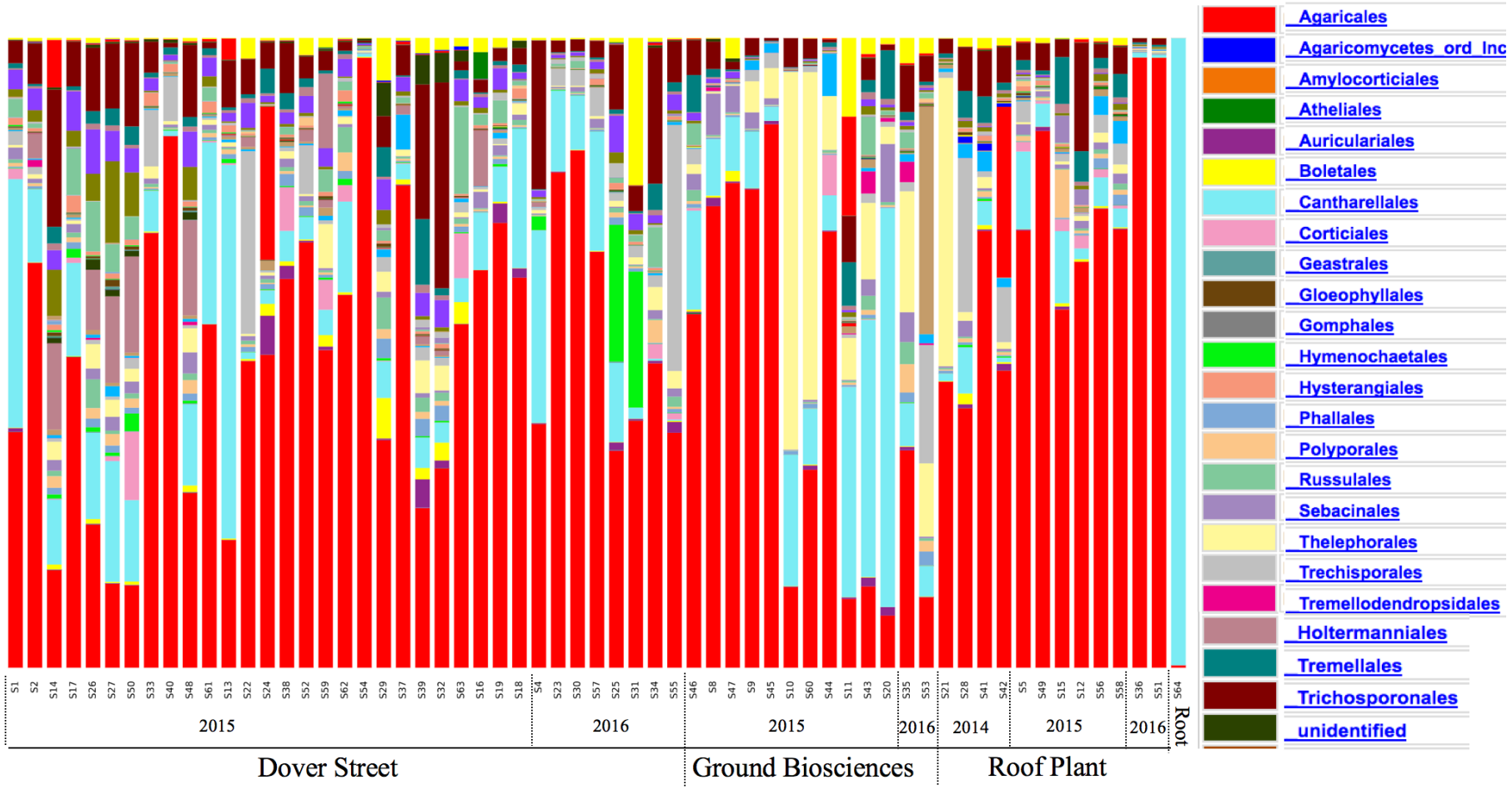


Figure 4. 18: Relative abundance of Basidiomycota OTUs taxonomy at order level after filtering.

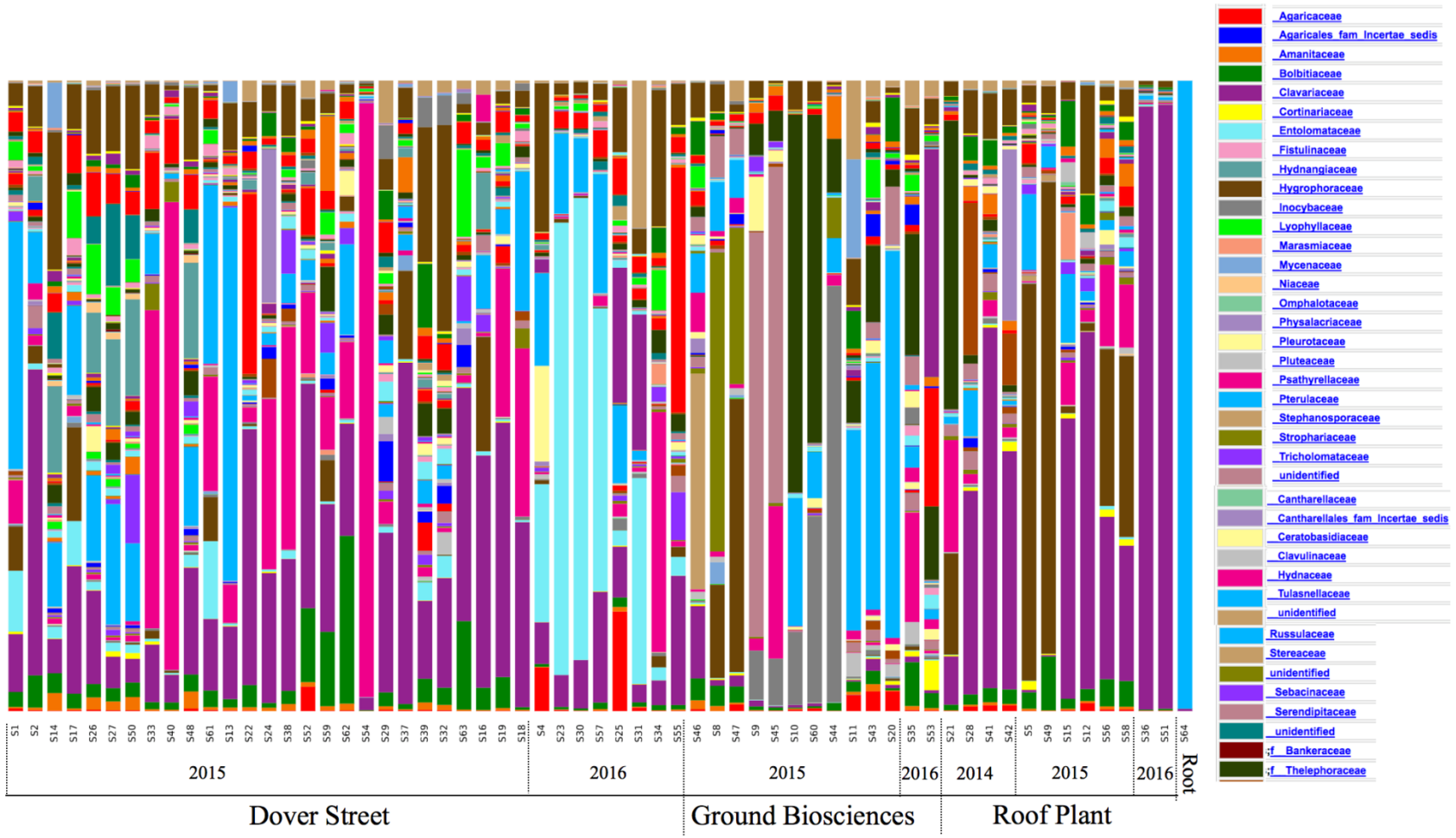


Figure 4. 19: Relative abundance of Basidiomycota OTUs taxonomy at family level after filtering.

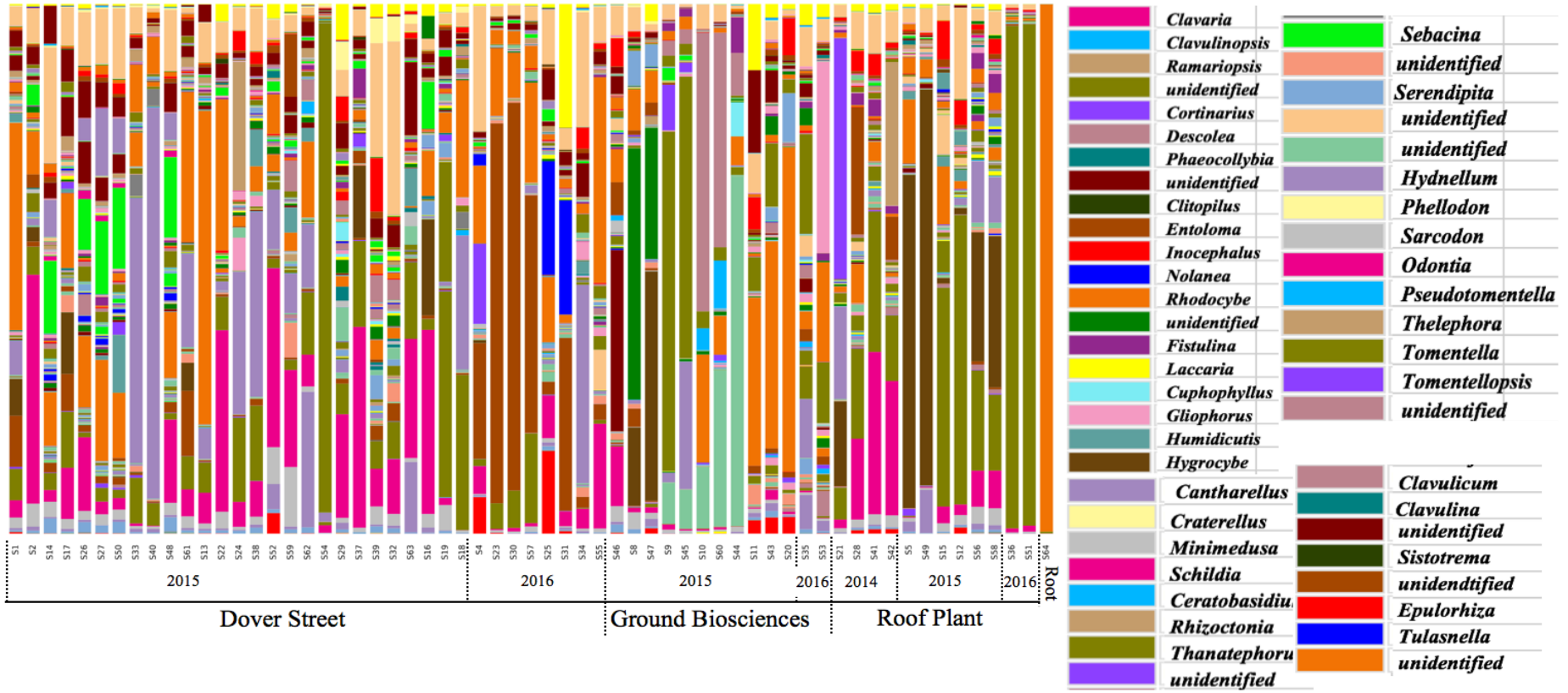


Figure 4. 20: Relative abundance of Basidiomycota OTUs taxonomy at genus level after filtering.

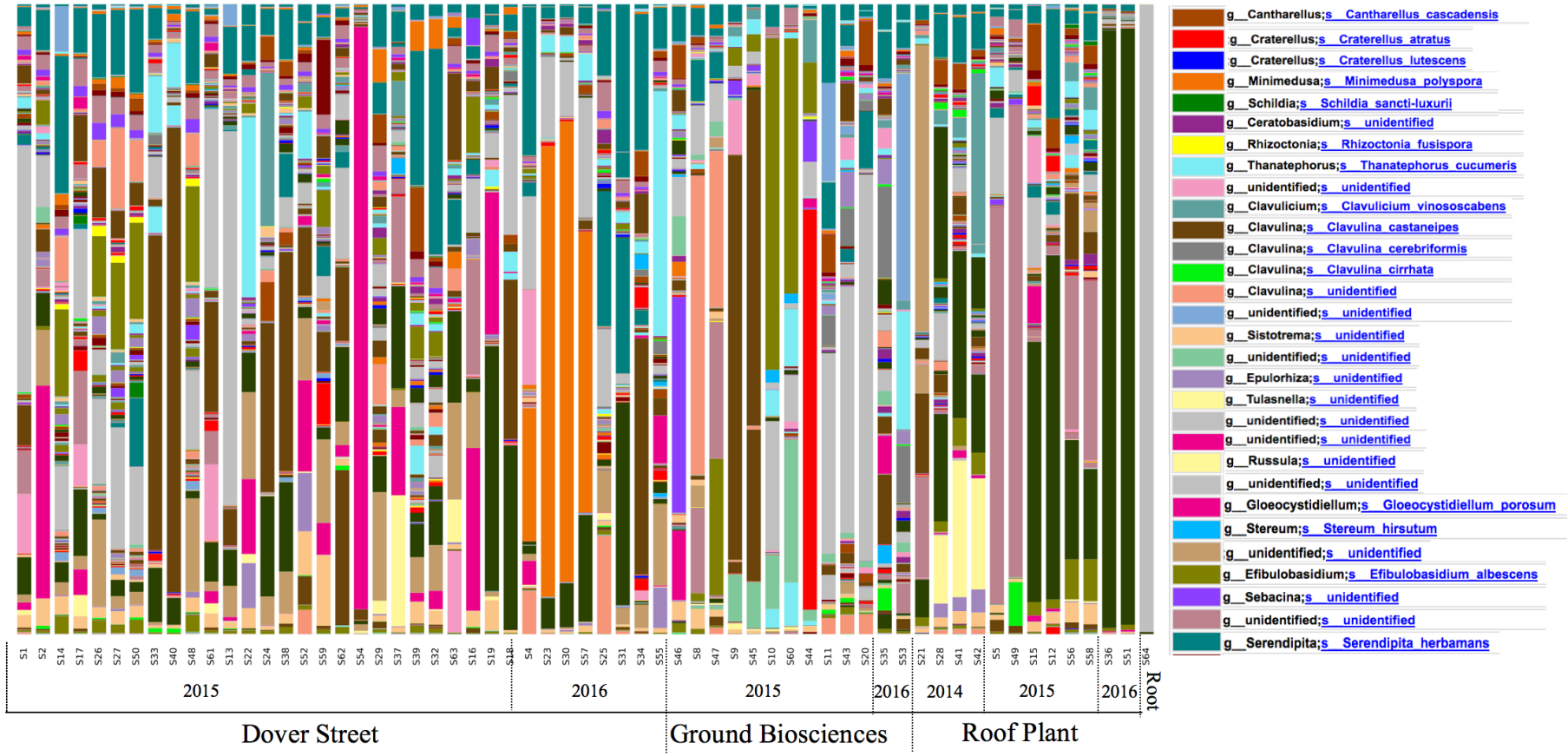


Figure 4. 21: Relative abundance of Basidiomycota OTUs taxonomy at species level for the order *Cantharales* after filtering.

Table 4. 11: Taxonomy relative abundance of *Basidiomycota* at levels of taxa (class, order, family, genus, and species) from all soil samples. The taxa highlighted in yellow refer to the species that are considered associated with orchids. All taxonomy from class to species for other not mentioned here in Appendix 4.2.

Class	Total %	Order	Total %	Family	Total %	Genus	Total %	Species
Agaricomycetes	79.8	*Agaricales*	51.2	*Clavariaceae*	18.7	*Clavaria*	8.1	*californica, falcate, incarnate, tenuipes*
						Unidentified	10.5	
				Psathyrellaceae	9.2	*Coprinellus*	0.2	*Micaceus and verrucipermus*
						Coprinopsis	6.6	*Atramentaria, pachyderma, and spelaiophila*
						Psathyrella	0.1	*globosivelata*
						Unidentified	2.2	
				Hygrophoraceae	5.7	*Hygrocybe*	5.4	*nigrescens*
						Neohygrophorus	0.3	*angelesianus*
				Entolomataceae	5.0	*Entoloma*	5.0	*Graphitipes, rusticoides, sericeum, and sphagneti*
				Bolbitiaceae	2.9	*Conocybe*	1.3	*Apala, nigrescens, striaepes, and velutipes*
						Unidentified	1.6	
				Unidentified	2.7			
				Inocybaceae	2.2	*Inocybe*	2.2	*Amethystine, griseovelata, and subcarpta,*
				Strophoariaceae	1.6	*Hebeloma*	1.2	*cavipes*
						Hypholoma	0.1	*fasciculare*
						Phollota	0.1	
						Psilocybe	0.2	*montana*
		Cantharellales	12.8	*Tulasenellaceae*	11.5	*Unidentified*	11.5	
				Ceratobasidiaceae	0.6	*Ceratobasidium*	0.1	*Unidentified*
						Unidentified	0.5	
						Clavulina	0.1	*castaneipes*
		Thelephorales	4.9	*Thelephoraceae*	4.9	*Pseudotomentella*	0.2	*Unidentified*
						Tomentella	1.6	*agbassaensis, amyloapiculata, fuscocinerea, and lilacinogrisea*
						Tomentellopsis	0.7	*Unidentified*
						Unidentified	2.3	
		Trechisporales	3.3	*Unidentified*	2.6			

		Sebacinales	1.3	*Sebacinaceae*	0.2	*Sebacina*	0.2	*Unidentified*
				Serendipitaceae	0.9	*Serendipita*	0.7	*herbamans*
						Unidentified	0.3	
				Unidentified	0.2			
		Hymenochaetales	1.0	*Unidentified*	0.7			
Agaricostillbomycetes	0.9	*Agaricostilbales*	0.9	*Chionosphaeraceae*	0.9	*Kurtzmanomyces*	0.8	*shapotouensis and tardus,*
Cystobasidiomycetes	1.6	*Cyphobasidiales*	1.5	*Unidentified*	1.5			
Entorrhizomycetes	0.9	*Entorrhizales*	0.9	*Entorrhizaceae*	0.9	*Entorrhiza*	0.9	*cypericola*
Exobasidiomycetes	0.1	*Entylomatales*	0.1	*Entylomataceae*	0.1	*Entyloma*	0.1	
		Georgefischeriales	0.1	*Tilletiariaceae*	0.1	*Phragmotaenium*	0.1	*oryzicola*
Microbotryomyctes	0.8	*Leucosporidiales*	0.2	*Leucosporidiaceae*	0.1	*Mastigobasidium*	0.1	*intermedium*
		Sporidiobolales	0.5	*Incertae sedis*	0.5	*Rhodotorula*	0.5	*erulica*
Pucciniomycetes	1.8	*Helicobasidiales*	0.1	*Helicobasidiaceae*	0.1	*Unidentified*	0.1	
		Platygloeales	1.4	*Eocronartiaceae*	1.4	*Eocronartium*	1.4	*Unidentified*
		Pucciniales	0.2	*Pucciniastraceae*	0.1	*Thekopsora*	0.1	*nipponica*
		Unidentified	0.2					
Unidentified	1.5							
Tremellomycetes	11.9	*Cystofilobasidiales*	1.1	*Cystofilobasidaceae*	1.0	*Cystofilobasidium*	0.1	*capitatum and macerans*
						Guehomyces	0.8	*pullulans*
						Phaffia	0.1	*Unidentified*
				Unidentified	0.2			
		Filobasidiales	1.8	*Filobasidiaceae*	0.1	*Goffeauzyma*	0.1	*gastrica*
				Piskurozymaceae	1.7	*Solicoccozyma*	1.7	*terrea and terricola*
				Unidentified	0.1			
		Holtermanniales	0.6	*Incertae sedis*	0.6	*Holtermanniella*	0.6	*takashimae*
		Tremellales	1.8	*Bulleraceae*	1.3	*Bullera*	1.3	*Unidentified*
				Bulleribasidiaceae	0.3	*Vishniacozyma*	0.3	*victoriae*
				Tremellaceae	0.2	*Cryptococcus*	0.1	*Unidentified*
						Tremella	0.1	*giraffa*
		Trichosporonales	6.1	*Trichosporonaceae*	6.1	*Apiotrichum*	6.1	*gamsii and veenhuisii*
		Unidentified	0.4					

4.3.5.3 Alpha diversity:

After relative abundances had been determined, the alpha diversity was calculated to determine similarities among samples. One of the tests was rarefaction curves (at 100000 sequences per sample before filtering) using the richness estimator, Chao1, as shown in Figure 4.22. This depends on the sampling depth (e.g. a sample which incompletely represents community will appear less diverse than when full sampled).

A rarefaction curve is represented as a plot of the number of species in the sample and the richness of the samples. Therefore, for comparing alpha diversity among samples of different sizes, it is necessary to count the re-sampling from the data. Thus, Figure 4.22 shows that sample S25 (pale brown, at the top of the other curves) has the highest species richness indicating there is more alpha diversity than the other soil samples. In contrast sample S64, (orange, the lowest of the curves) has substantially less species richness and depth than the other samples, fitting its origin from a root. Other samples have different levels of plateaus. This rarefaction curve does not include all samples (14 samples outside the range of sequences between 10000 and 100000) were discarded.

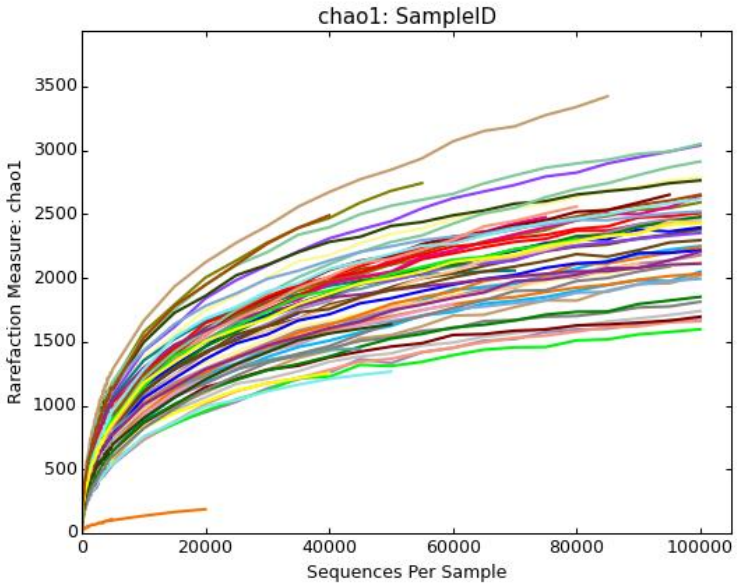


Figure 4. 22: Rarefaction curves of community richness before filtering by using Chao1 metrics. It shows an estimate measure for sequences per sample for all OTUs.

However, after filtering the rarefaction curves plot was different and the number of sequences decreased so that the range of sequences choice was 5000-70000. As can be seen in Figure 4.23, only four samples have extended to 70000 sequences, so that the highest alpha diversity was in samples S45 and S54 (orange and pale brown, respectively) and less in samples S36 (dark green) and S44 (bright green).

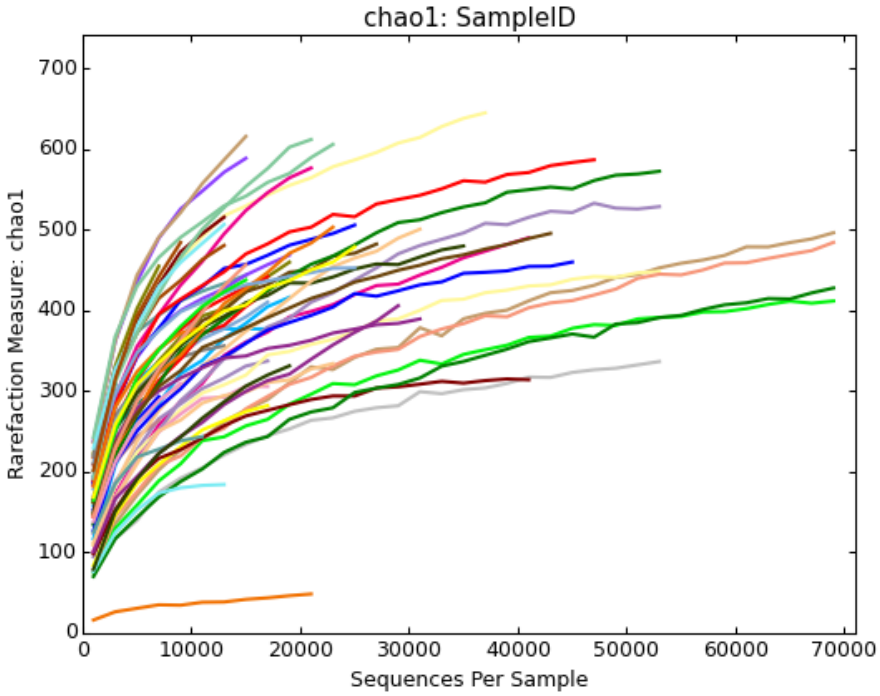


Figure 4. 23: Rarefaction curves of community richness after filtering using Chao1 metrics. It shows an estimate measure for sequences per sample for all OTUs.

4.3.5.4 Beta diversity:

Principal Coordinates Analysis (PCoA) was performed using Bray-Curtis beta diversity to show the dissimilarities of sample groups before and after filtering in a two-dimensional plot per site (Figure 4.24). The left side of this figure relates to data before filtering and it can be seen that separation of groups in PC1 vs PC2 explained 12.02 % to 8.63 % of variation respectively. Prior to filtering, samples were generally grouped according to location, although two samples (S5, S49) from the roof plants (R) were grouped with the Ground Biosciences group. The case with PC3 vs PC2 was different, with less percent of variation explained (7.33 % PC3, 8.63 % PC2). In addition, the Dover Street group was mixed with Roof plant group,

indicating that there is less dissimilarity between them, while the Ground Biosciences samples were still a distinct group. The plot of PC1 vs PC3 showed separation of the three sample groups in a similar way to PC1 vs PC2.

After filtering the data to focus on the phylum Basidiomycota only, the PCoA plots explained similar amounts of the variation. However, they were rather different in terms of distribution of samples according to the sites (Figure 4.24, right hand side). It is obvious that although groupings can be seen according to sample site (e.g. PC1 vs PC2), these are less distinct. The similarities of the three sites are more apparent.

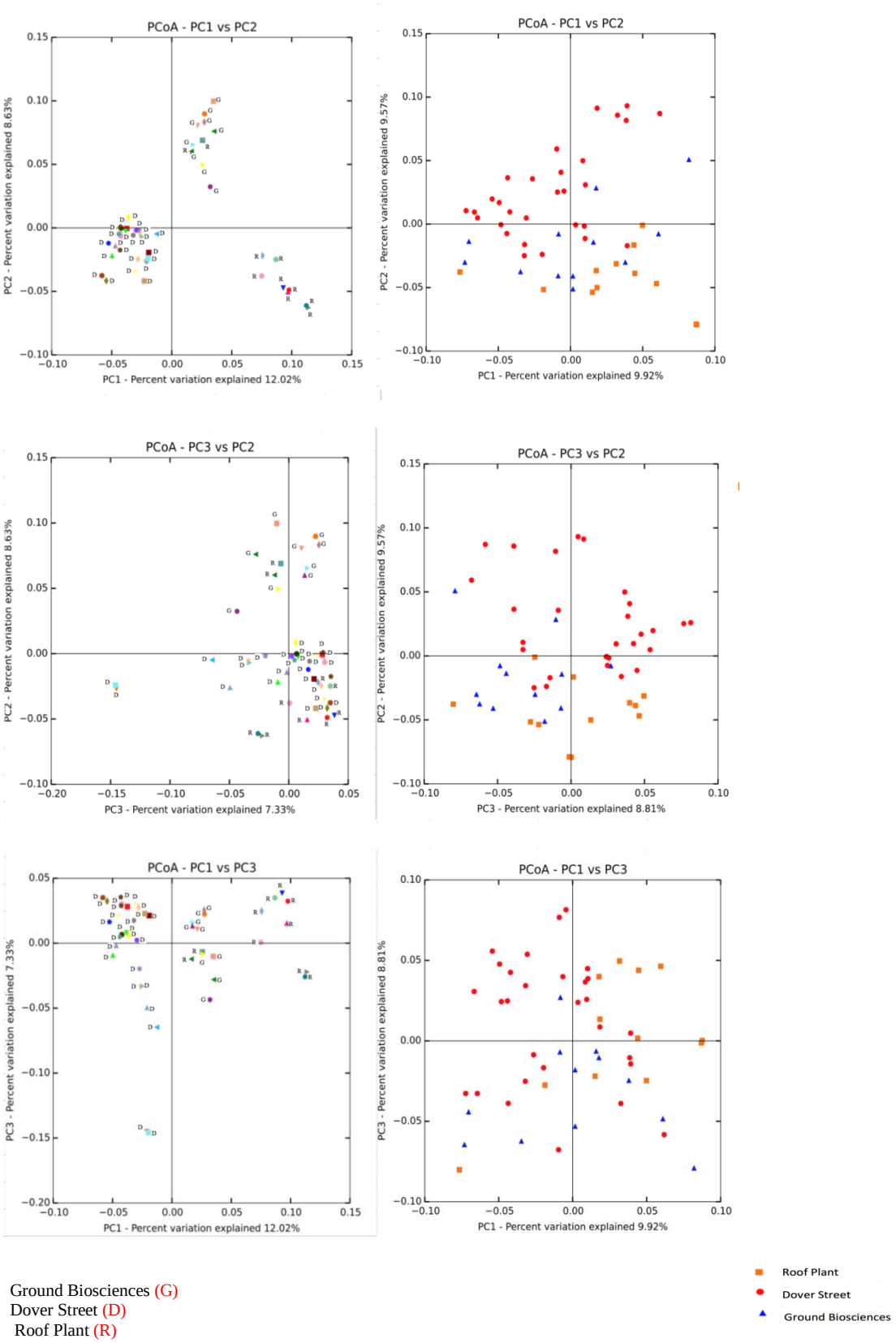


Figure 4. 24: PCoA plots before and after filtering using Bray-Curtis metrics showing three groups of site distance. Organized according to filtering before (left panel), and after (right panel).

4.3.5.5 Statistical test:

The impact of species taxa on diversity was tested using the Kruskal-Wallis test. This identified that there were 120 *de novo* significant differences in species taxa between the three sites (see Appendix 4.3). As can be seen in Table 4.12, showing the top 26 significant taxa, there is a significant difference in the species present in each site. There was a significant different between groups in *Tulasnellaceae, Sebacinaceae,* and *Ceratobasiaceae* with FDR P-value ≤ 5%.

Table 4.12: Top 26 significant taxa supported by Kruskal-Wallis statistical test from beta diversity according taxonomic level.

Class	Order	Family	species	Roof plant-Mean	Dover Street-mean	Ground Biosciences-mean	P-Value	FDR_P
Agaricomycetes	*Agaricales*	*Entolomataceae*	*Entoloma terreum*	0.333333333	13.93333333	0.076923077	5.21E-10	1.11E-06
Agaricomycetes	*Agaricales*	*Stephanosporaceae*	*unidentified unidentifi*	0	8.2	0	7.25E-10	1.11E-06
Tremellomycetes	*Tremellales*	*Tremellaceae*	*Cryptococcus unident*	0	8.4	0.076923077	1.26E-09	1.11E-06
Tremellomycetes	*Filobasidiales*	*Piskurozymaceae*	*Solicoccozyma terrico*	17.16666667	426.6333333	55.46153846	1.43E-09	1.11E-06
Agaricomycetes	*Phallales*	*Phallaceae*	*Clathrus uber*	0	7.966666667	0.153846154	1.59E-09	1.11E-06
Agaricomycetes	*Agaricales*	*Clavariaceae*	*Clavaria tenuipes*	8.416666667	1016.366667	0.307692308	2.75E-09	1.60E-06
Tremellomycetes	*Filobasidiales*	*Piskurozymaceae*	*Solicoccozyma terrico*	8.666666667	176.2666667	27.23076923	3.32E-09	1.62E-06
Agaricomycetes	*unidentified*	*unidentified*	*unidentified unidentifi*	99.5	0.033333333	0	3.70E-09	1.62E-06
Agaricomycetes	*Thelephorales*	*Thelephoraceae*	*Tomentella unidentifie*	11.41666667	0	0	7.30E-09	2.56E-06
Tremellomycetes	*Tremellales*	*Tremellaceae*	*Cryptococcus unident*	13.75	0	0	7.31E-09	2.56E-06
Agaricomycetes	*Agaricales*	*Clavariaceae*	*Clavaria argillacea*	0.25	0	28.46153846	2.29E-08	7.26E-06
Agaricomycetes	*Cantharellales*	*Clavulinaceae*	*Clavulicium vinososca*	1.25	12.86666667	0.153846154	4.84E-08	1.41E-05
Agaricomycetes	*Phallales*	*Phallaceae*	*Clathrus ruber*	0.083333333	9	0.230769231	5.52E-08	1.49E-05
Agaricomycetes	*Agaricales*	*Bolbitiaceae*	*Conocybe hornana*	8.583333333	0.233333333	0	6.15E-08	1.54E-05
Agaricomycetes	*Russulales*	*Peniophoraceae*	*Peniophora unidentifi*	1.5	22.6	6.076923077	6.62E-08	1.54E-05
Agaricomycetes	*Russulales*	*Russulaceae*	*Russula myrmecobrom*	1.75	0	0	7.97E-08	1.74E-05
Agaricomycetes	*Russulales*	*Russulaceae*	*unidentified unidentifi*	0.416666667	0.066666667	12.30769231	1.51E-07	3.11E-05
Agaricomycetes	*Thelephorales*	*Thelephoraceae*	*unidentified unidentifi*	0.083333333	0	198.0769231	2.61E-07	5.06E-05
Tremellomycetes	*Cystofilobasidiales*	*Cystofilobasidiaceae*	*Guehomyces pullulans*	9.583333333	260.4	15.92307692	3.02E-07	5.37E-05
Agaricomycetes	*Polyporales*	*Fomitopsidaceae*	*Antrodia hyalina*	0	13.73333333	0.076923077	3.14E-07	5.37E-05
Microbotryomycete	*Leucosporidiales*	*unidentified*	*unidentified unidentifi*	0	4.066666667	0.076923077	3.23E-07	5.37E-05
Agaricomycetes	*Agaricales*	*Bolbitiaceae*	*Conocybe hornana*	4	0.1	5.923076923	5.17E-07	7.91E-05
Agaricostilbomycet	*Agaricostilbales*	*Chionosphaeraceae*	*Kurtzmanomyces shap*	69.16666667	0.033333333	0.153846154	5.21E-07	7.91E-05
Tremellomycetes	*unidentified*	*unidentified*	*unidentified unidentifi*	0.583333333	8.666666667	0	5.43E-07	7.91E-05
Agaricomycetes	*Agaricales*	*unidentified*	*unidentified unidentifi*	2.75	0.033333333	0	6.01E-07	8.41E-05
Agaricomycetes	*Agaricales*	*Entolomataceae*	*Entoloma sericeum*	6.833333333	29.86666667	5	7.07E-07	9.50E-05

4.4 Discussion:

4.4.1 Identification of fungi morphologically:

Microscopic observation showed a number of fungal pelotons in the root tissue (see Figure 4.7) and these results were vital morphological evidence for the relationship between the orchid plant and their fungal partner associates. This agrees with previous findings in the literature (Pecoraro et al., 2015), who found fungal pelotons in the root cells of *O. bertolonii* using the microscope.

4.4.2 Molecular identification of root fungal associates:

Although isolation of the peloton directly from the orchid tissue is the preferable method (Warcup and Talbot, 1967, Taylor and Bruns, 1997, Rasmussen, 2002), in many orchids this is difficult because the roots do not have sufficient mycorrhizal infections (Otero et al., 2007). Also, it is not possible to distinguish readily between live and dead pelotons and other contaminations when examined under the dissecting microscope. Therefore, the method described by (Otero et al., 2002b) was adopted to isolate fungi from roots onto culture media.

The value of molecular methods, provided that well-characterized sequences are available in databases, was demonstrated in that twenty-nine isolates have been identified using molecular methods among those isolated to the culture medium. Only three of these were *Basidiomycota* while the others were *Ascomycota* or uncharacterized environmental samples. The identity of the *Basidiomycota* could be established to family level based on sequencing several rRNA regions as described in section 4.3.2.2. In addition, it was demonstrated that DNA could also be isolated directly from orchid root suitable for amplification by PCR with primers targeted to the rDNA region. In both instances, sequencing indicated the presence of *Basidiomycota* belonging to the *Tulasnellaceae.* This is in a good agreement with Pecoraro et al. (2015), and with Shimura et al. (2009) from populations of the threatened orchid *C. macranthos var. rebunense*.

4.4.3 Effects of different temperature and culture media on growth of fungi:

The effect of temperature and culture media are variable on the growth of different fungi. Many woodland fungi are mesophilic, growing across the temperature range between 5 to 35°C. There are other fungi adapted to temperatures outside this range,

such as thermo-tolerant species (≤ 40 °C) and thermophilic species that grow optimally between 40 and 50°C and also psychro-tolerant species that can grow below 0 °C (Dix and Webster, 1995). Fungal isolates B and R3 grew within the range of 20-26°C, more rapidly at 20°C, indicating that they are mesophilic, as would be expected.

In term of culture media, the radial growth rate of the fungi that were isolated in this study varied depending on the nature of culture media used. The difference in growth rate between isolates B and R3 on water agar, a poor source of carbon and nitrogen, may relate to different strategies to search for new sources in a nutrient-poor environment. Isolate B, which is used to initiate symbiotic development of seeds of several orchid species, has been selected for vigorous growth and in culture for many years (www.hardyorchidsociety.org.uk/HOS%201012/Cultivation.html) while R3 was recently isolated from an orchid root. The latter is thus more typically dependent on a high—nutrient environment than the former.

4.4.4 Identification of fungi from soil:

4.4.4.1 DNA extraction:

The application of molecular methods to soil fungal communities relies on isolating suitable DNA (Becker, 2011). The quality and quantity of the DNA obtained can be varied, related both to fungal structure and chemical materials within the soil. Adequate amounts of DNA may be extracted but the material may be inadequate for further reactions because of the humic acids or enzymatic inhibitors that may be co-extracted in the DNA (Zielińska et al., 2017). As a consequence, there is no single best isolation method. Several methods for DNA extraction from soil were therefore applied on one soil sample to compare the quantity and quality of the DNA yield and its suitability as a PCR template. Although all methods yielded DNA, it was in many cases unsuitable as a PCR template. The UltraClean Soil DNA Isolation Kit was picked for extraction of DNA from the soil because it yielded DNA that could be amplified by PCR.

4.4.4.2 Cloning to estimate fungal diversity within soil:

Cloning is a well-established technique to estimate diversity within microbial communities. Testing this approach for soil fungi yielded 22 clones from one roof

plant soil sample that were directly amplified using PCR and further analyzed by sequencing using Blast program at the NCBI. This identified 5 different species and also matches to two uncultured fungi from environmental samples. The PCR primers were selected to be directed towards the *Basidiomycota*, and a majority of clones were from this phylum. This indicated that this approach was possible, but also identified some of the drawbacks. The workload to apply this approach to a significant number of samples is substantial. In addition, the Basidiomycota fungi identified were the yeasts *Apotrichum dulcitum* and *Saitosyma podzolica* from the family *Trichosporonaceae* and *Trimorphomycetaceae*.

This strong bias toward yeast species could have reflected the soil community within the plant pots, the incompleteness of the database or errors in PCR related amplification (Smit et al., 1999) but did not match with the families generally identified as orchid mycorrhizal fungi. As a consequence, application of NGS techniques that should provide large samples of sequences was explored (see section 4.4.4.4 below)

4.4.4.3Phylogenetic analysis:

Phylogenetic tree construction using ITS sequences from the fungi isolated from orchid root, as well as sequences amplified from DNA extracted directly from the orchid root showed that the isolates grouped within the *Tulasnellacea* (Figure 4.14). This adds to the evidence that species belonging to this family are involved in mycorrhizal symbiosis with the *O.apifera*. This family has frequently been reported as a mycorrhizal fungus and seems to be common specifically in orchid (Brundrett, 2006, Bayman and Otero, 2006). The sequences derived from soil and root samples are closely related to each other. This agrees with previous findings in the literature (Voyron et al., 2017).

4.4.4.4 Identification of fungi from soil using Illumina:

The Illumina MiSeq platform was used to carry out analysis of the fungal community in the soil from around *O. apifera.* Most of the fungal symbiosis research is focused on the taxonomy of fungi in other species of orchid and their roots, such as the study of fungal biodiversity in Mediterranean orchids (Pellegrino et al., 2016) and from Aegean islands (Gezgin and Eltem, 2009).

Therefore, a fungal investigation of soil samples around bee orchid was performed using high throughput sequencing. QIIME was used to further investigate the raw sequence data obtained from CGR for picking OTU, taxonomic assignment, alpha diversity, and beta diversity analysis.

Finally, PCoA two dimensional plots were performed to determine relationships between OTUs based on similarity and the Kruskall-Wallis statistical test was applied to assess the level of significant difference between samples. A total of 35 amplicons from Dover street collections, 13 amplicons from the Biosciences Building collections, and 12 amplicons from Roof plant, with one amplicon from the root of a Roof plants, were examined.

4.4.4.4.1 Dominant fungi in soil samples:

The main point here is to discover whether the mycorrhizal fungi that associated with bee orchid are present, and their distribution, thus focusing on *Basidiomycetes*. QIIME revealed the phylum *Ascomycota* as the most dominant followed by *Basidiomycota*. The proportions of fungal phyla were similar to those identified by (Schmidt et al., 2013) who found that *Ascomycota* (78.14%) comprised most of the sequences followed by *Basidiomycota* (10.24%), *Glomeromycota* (3.61%) and *Chytridiomycota* 0.36%. The present study found *Ascomycota* 54.6%, *Basidiomycota* 23.6%, *Glomeromycota* 1.6%, and *Chytridiomycota* 3.5%. The differences in the percentages, if significant, could be due to the different types of soil samples but also related to technical aspects such data loss during data filtering, which is due to the paired-end assembly and generality of the primers. There is also a difference between the specificity of metabarcoding primers that affects recovery of multiple fungal lineages (Toju et al., 2012).

After focusing on the *Basidiomycota* at the Order level, the order *Agaricales* was dominant (51.20%) followed by *Cantharellales* (12.80%), then *Trichosporales* (6.10%), and *Thelephorales* (4.90%). The most abundant class was *Agaricomycetes* with 79.8% which contains the *Cantharellales* 11.5% (*Tulasenellaceae and Ceratobasidiaceae*) and *Sebacinales* 1.3 % (*Sebacinaceae*). These fungi have been reported to be members of the *Rhizoctonia*-like fungi and interact largely with orchids (Dearnaley et al., 2012). A number of studies have confirmed that many

orchid species associate with *Tulasenellaceae* and, to a lesser extent, with *Ceratobasidiaceae* and *Sebacinaceae* such as in studies of roots from *Orchis militaris* and *Cypripedium spp* (Shimura et al., 2009). At the family level, the *Clavariaceae* (18.7%) is the most abundant, then *Tulasenellaceae* (11.5%), followed by *Psathrellaceae* (9.2%).

However, many sequences could not be identified. Thus, although at class level 1.5% remained unidentified, many of the families had unidentified genera, such as all genera of family *Tulasenellaceae* (11.5%). In addition, 0.5% unidentified was *Ceratobasidiaceae* as well as all species of genus *Sebacina*. There are at least two causes of this. Firstly, the UNITE database (Koljalg et al., 2013) was designed around sequences derived from Basidiomycota and Ascomycota that have been identified in ectomycorrhizal associations. The source material is often fruiting bodies from well-documented collections. However, sequences derived from cultures, soil and root samples are also included. The advantage of this database is the high quality, curated sequences, with the disadvantage of poor coverage of some groups. Secondly, information on many orchid mycorrhizal fungi at the species level is lacking. Nevertheless, analysis using a different source of fungal sequence data would provide an interesting comparison.

The most striking result to emerge is the root sample that has 99.6% *Tulasenellaceae.* This sample was collected at beginning of August 2016 which is the time of dormancy for bee orchid. Kohout et al., (2013), reported that *Tulasnellaceae* partners of the endangered *Pseudorchis albida* are more abundant and diverse in summer, perhaps to take benefit from the nutrient exchange as the green adult plant develops (Kohout et al., 2013). On a substrate containing low nitrogen and high carbohydrate, protocorms of *A. morio* formed healthy mycorrhiza (Beyrle et al., 1995)

4.4.4.4.2 Diversity over time and space in the soil fungal community

Some orchids need to associate with different fungi at different stages of development such as *Gastrodia elata* that requires *Rhizoctonia*-like fungi at an early stage of their protocorm development while at mature development the plant is colonized with *Armillaria* (Smith and Read, 1997). This may be because of not only the development of the orchid through the year but also the variable environmental

conditions expected to affect the mycorrhizal fungi in the soil. All orchids need mycorrhizal fungi in one of their stages of life whether photosynthetic or non-photosynthetic. Organic sources of nitrogen and phosphorous are important for fungi that grow in the soil (Smith and Read, 1997). It has been suggested that in the soil, fungal communities are similar in presence of spatial community structure over distances of only a few metres (Schmidt et al., 2013). It has been observed that an increase in fungal richness can be due to expanding the geographic region of sampling, and a higher number of species have been expected when the area available for fungal colonization is increased (Peay et al., 2007).

The samples analysed by high-throughput sequencing were collected in different months across a two-year period at the three sites. Multiple samples were analysed for several time points to provide biological replicates. Analysis of beta diversity using principal coordinate's analysis showed that the three sites could be distinguished on most metrics when all the data were included. However, once the data was filtered to focus on the *Basidiomycota* alone, the sites were less distinguishable (Figure 4.24). This overall similarity can also be seen when looking at the relative abundances of fungal groups in Figures 4.16 – 4.21. At the levels of phylum and class, there are few differences over time but at the level of order, family and species some distinct changes are visible.

The list of top 120 significantly different OTUs identified taxa that differed significantly between the three sites. This included the *Agaricomycetes* (*Entoloma terreum*), which was much more abundant at the Dover street site than the Roof or Biosciences Ground sites. At the family level, the family *Thelephoraceae* was more abundant in the Ground Biosciences than the Dover Street and Roof sites, while *Chionosphaeraceae* (*Kurtzmanomyces shapotouensis*) was higher in Roof site and lower in both other two sites. This latter species is again yeast, and its abundance in the containers on the roof matches with the identification of yeast species in the cloning experiment to assess fungal diversity. The order *Cantharellales* was lower in the Ground Biosciences and the Roof site compares to Dover Street. However, many significant species at each site were unidentified.

The Kruskall-Wallis test showed there was a significant difference between the three sites for *Tulasnellaceae* (*Epulorhiza*) and *Sebacinaceae* (*Sebacina*), although these were present at all three sites, while *Ceratobasidiaceae* (*Ceratobasidium*) was only present at the Ground Biosciences site. *Epulorhiza* was considered as a common and distinctive form-genus of Basidiomycetes that form a mycorrhizal association with orchid roots (Chutima, 2012) while *Tulasnella and Gloeotulasnella* have been isolated from roots of terrestrial orchids (Porras-Alfaro and Bayman, 2007, Lee, 2002, Kristiansen et al., 2001).

5 Chapter Five: Transcriptome study of protocorms, corms, and roots of *O. apifera.*

5.1 Introduction:

5.1.1 Transcriptomic study platforms for plant RNA-seq:

The transcriptomes of many non-model plant species have been assessed using several platforms. The Roche/ 454 platform has been used for several non-model plants (Alagna et al., 2009, Barakat et al., 2009, Guo et al., 2010, Novaes et al., 2008, Cole et al., 2009, Franssen et al., 2011), but has now been replaced by other technologies. One of these is Illumina so that non-model plant transcriptome studies have also been reported using this platform (Strickler et al., 2012a). *De novo* assembly can be more difficult using the Illumina platform because of their short reads comparing to the 454 platform (Pop and Salzberg, 2008)

5.1.1.1 Transcriptomes of Orchid plant

The first comprehensive sequence resource from orchids of the subfamily *Orchidoideae*, and sexually deceptive orchids was reported by Sedeek et al., (2013). They combined sequencing using the 454 of three species of *Ophrys* (*O. exaltata*, *O. sphegodes and O. garganica*) with EST Sanger sequencing and Illumina /Solexa data into a reference transcriptome to represent the maximum amount of genetic information profile for *Ophrys* (Sedeek et al., 2013).

A total of 121 917 unique putative *Ophrys* transcripts were found and then used for annotation and consequent analysis. This was more than for the earlier *Phalaenopsis* orchid transcriptome (42 863 transcripts) (Hsiao et al., 2011), and even more than in *Oncidium* (Chang et al., 2011). The *Ophrys* reference transcriptome has been generated from floral tissues (leaves, bracts, labella, sepals, petals and columns) (Sedeek et al., 2013), and consequently, they identified genes for pollinator attraction.

OrchidBase and Orchidstra (http://orchidbase.itps.ncku.edu.tw/ and http://orchidstra2.abrc.sinica.edu.tw) are orchid transcriptomic databases that have been established. OrchidBase is constructed from different species of *Phalaenopsis* orchid and from various tissues (seed, protocorm, vegetative tissue, leaf, cold-treated plantlet, pathogen-treated plantlet, inflorescence and flower buds). It includes new

information on *Orchidaceae* floral expressed sequences, and it a comprehensive collection of *Orchidaceae* floral transcriptomes (Tsai et al., 2013). In Orchidstra both genes and tissue-specific expression are categorized using analysis of RNA-seq (Su et al., 2011).

Niu et al. (2016) combined Illumina HiSeq2000 data and used Trinity for *de novo* assembly to characterize the transcriptomes from 11 different tissues (root, stem, leaf, flower buds, column, lip, petal, sepal and three developmental stages of seeds) of the orchid *Phalaenopsis equestris*. They are presented in three databases (RNA-Seq raw reads, sets of unigenes and predicted coding sequences and proteins, annotation results versus the Nonredundant (Nr) protein database) to compare with previously generated *Phalaenopsis* transcriptome data from the same tissues or various other tissues to understand morphology of orchid flowers (Niu et al., 2016).

Comparative gene expression of reproductive tissues in *Phalaenopsis aphrodite* has also been studied. As a result, it has been suggested that protocorm-like-body (PLB) regeneration (the clonal propagation used in the orchid industry) does not follow embryogenesis and the data showed that protocorms and PLBs have similar transcriptomic signatures, and are different from the zygotic embryos (Fang et al., 2016).

Transcriptome analysis has been applied using Illumina paired-end sequencing for tessellated and green leaves in the *Paphiopedilum* genus (*Paphiopedilum* spp. known as lady's slipper orchids) to study leaf physiology and evolutionary adaptation. In addition, it provided a large amount of sequence data for leaves and identified genes necessary for studying the functional regulation of leaf features at molecular levels (Li et al., 2014).

Moreover, the genome of *Phalaenopsis equestris* has been sequenced, and it was found that gene duplication has contributed to the evolution of CAM photosynthesis and they also found that families of MADS-box C/D-class, B-class, AP3, and AGL6-class genes might contribute to the morphology of orchid flowers (Cai et al., 2015).

Mature flower buds of *Apostasia shenzhenica* (subfamily *Apostasioideae*) were used to study development and evolution of orchid flowers (Tsai et al., 2013). Gene expression levels can be inferred independently of prior genomic knowledge of species. Therefore, this method has been applied to wide a range of research areas, such as the role of differential gene expression (DE) in phenotypic divergence and speciation (Vijay et al., 2013). The full scope of how RNA-seq can be applied in biology is still being explored (Ekblom and Galindo, 2011).

The application of transcriptome sequencing to orchids is still limited and has focused especially on flowers and on the *Phalaenopsis* and *Cymbidium* genera because they are important commercially and medicinally (Bhadauria, 2017). The current project aimed to study physiological differences between bee orchid grown in the presence and absence of its mycorrhizal fungus.

5.1.2 Illumina sequencing:

The Illumina HiSeq 2500 sequencing platform has been used for transcriptome sequencing and is an upgrade of the HiSeq2000. This platform can generate up to 120 Gb of data in 27 hours (Shokralla et al., 2012). In a further update, the Illumina HiSeq3000/4000 platform can generate up to 650 Gb of throughput and up to 2.5 billion reads for paired-end sequencing in ~3 days run time (Goodwin et al., 2016).
The Illumina HiSeq 4000 platform with version 1 chemistry was used for sequencing in the current study using sequencing-by-synthesis (SBS) technology to generate 2 x 150 bp paired-end reads (Centre for Genomic Research, University of Liverpool. Laboratory analysis was performed by Anita Lucaci).

5.1.2.1 Samples Selection:

Choosing appropriate plant tissues is important and is usually driven by the biological question of the transcriptomic study (Strickler et al., 2012a). Shock-freezing of samples in liquid nitrogen immediately after harvest is considered the most reliable method to prevent loss of RNA due to RNase activity (Wolf, 2013). Biological replicates of sampling should be included as it is important to determine the sample variation especially when the aim is to discover differential expression between two conditions (Auer and Doerge, 2010).

In the current study, RNA was extracted from three tissues with three replicates of each. In addition, these tissues originated from two different growth conditions (field and *in vitro*). The tissues selected were protocorms from *in vitro* cultures, and corms and roots from field (Ground Biosciences).

5.1.2.2 Factors that affect transcriptome sequencing:

There are many complicating factors that must be solved before sequencing plant transcriptomes. The first challenge is to isolate RNA of sufficient quality and quantity. The presence of RNases can be problematic in causing degradation of the RNA. Further, the primary and secondary plant metabolites (e.g., polyphenols, polysaccharides) can pose problems through irreversibly binding to the RNA (Johnson et al., 2012).

5.1.2.3 RNA extraction methods:

Many methods have been utilized since there is no standard method for all plant species. The presence of RNase, polysaccharides and secondary compounds cause significant problems. Therefore, many strategies and kits have been designed to solve these difficulties. These include the NucleoSpin® RNA Plant and Fungi kit (Nagtong et al., 2009), designed for the isolation of RNA from plant and fungal tissues and also, Spectrum^{TM} Plant Total RNA Kit (Chang et al., 2016), TRIzol reagent and HiPer^{TM} Plant RNA kit (Zaman et al., 2016), innuPREP Plant RNA Kit (Johnson et al., 2012), and both Qiagen RNeasy Plant Minikit and Qiagen hybrid method (MacKenzie et al., 1997). The most common kits that have been used for RNA isolation, and especially for difficult tissues, are innuPREP Plant RNA Kit and Qiagen RNeasy Plant Minikit (Kurbidaeva and Novokreshchenova, 2011, Park et al., 2014, Dobnik et al., 2013).

For more difficult tissues, such as those known to contain high levels of polysaccharides and / or secondary metabolites, QIAGEN RNeasy Plant Mini Kit and TURBO DNA-*free* kits have been used following a CTAB- based method that was designed for pine trees and was successfully applied to *Vaccinium myrtillus* where a TRIzol-method was unsuccessful (Jaakola et al., 2001). Thus, the CTAB buffer was used to separate polysaccharides from nucleic acids, as well as using

polyvinylpyrrolidone (PVP) and β-mercaptoethanol for reducing oxidation of phenolic compounds. However, this has not been effective for all species (Jaakola et al., 2001).

In an evaluation of the methods used across seed plants focusing on more difficult species, Jordon-Thaden and his colleagues (2015), suggested the best RNA isolation approaches to obtain high-quality and adequate RNA for transcriptome sequencing was a combination of TRIzol supplemented with sarkosyl and the TURBO DNA-*free* kits. Another two options that used CTAB with or without a second stage of TRIzol, or another kit were less effective (Jordon-Thaden et al., 2015).

5.1.2.4 RNA assessment:

Successful analysis of gene expression depends on the quality and quantity of RNA. The critical elements are purity and integrity. Low quality RNA will compromise the result downstream, wasting time and resources. Some factors which have an effect on the quality of RNA are contamination with RNase, protein, and genomic DNA (Fleige and Pfaffl, 2006). Many techniques have been used to assess the purity and integrity of RNA such as optical density measurement by using NanoDrop (ND-3300, NanoDrop Technologies, USA), agarose gel electrophoresis or using Bioanalyzer 21000 (Agilent Technologies, USA) lab-on-chip technology (Fleige and Pfaffl, 2006). One advanced method is a combination between NanoDrop and RNA RiboGreen dye for ultra-sensitive quantification of RNA. It requires only 1-2 µl of RNA sample, and gives more detail about RNA integrity and other chemical contaminations (Fleige and Pfaffl, 2006).

5.1.2.5 rRNA removal or depletion:

Ribosomal RNA (rRNA) is the major portion of RNA within all cells, and therefore has to be removed prior to transcriptome library preparation (Wolf, 2013). RNA samples can be treated with the RiboMinus™ plant rRNA removal kit (Invitrogen, CA, USA) for rRNA removal, or any other commercial kit, for depletion, poly(A) capture or other hybridization-based method to obtain a sample with a highly-enriched mRNA (Ward et al., 2012). Ribo-Zero kits (Illumina) have been used in a hybridization-capture technique which removes >99% of cytoplasmic and mitochondrial rRNA from 1-5 µg of total RNA sample (Pease and Sooknanan,

2012). In this study 1 µg total RNA samples were treated with Ribo-Zero Magnetic kits (Plant root / seed) from Illumina by the CGR.

5.1.2.6 Reverse transcriptase polymerase chain reaction (RT-PCR):

RT-PCR is a technique used to make DNA copies of RNA. Reverse Transcriptase PCR uses an RNA template and makes DNA copies of it using the enzyme reverse transcriptase and then many copies of the DNA will amplified by normal PCR (Bustin and Nolan, 2013).

5.1.2.7 cDNA synthesis:

Generally sequencing platforms require the conversion of RNA to cDNA using an enzymatic reaction for reverse transcription and library preparation (Wolf, 2013). cDNA is usually fragmented into smaller pieces during library preparation as a template for sequencing and can be sequenced from single-end (SE) or paired-end (PE) reading from both sides. Paired-end sequencing is more beneficial for transcriptome assembly and isoform detection but the insert size must be short, generally <300 bp (Wolf, 2013). Commercial kits are used for this step such as TruSeq Stranded Total RNA kit (Huang et al., 2015) and ScriptSeq v2 RNAseq Library Preparation kit. The NEBNext® UltraTM Directional RNA Library Prep Kit for Illumina® #E7420 as first step and Illumina® KAPA Library Quantification Kit (Kapa Biosystems, Wilmington, USA) was performed for the final libraries in this study.

5.1.3 Transcriptome (RNA-seq) and initial data processing:

5.1.3.1 Raw data processing:

Raw data from sequencing was provided in a common output file in FASTQ format. These raw reads must be preprocessed before any subsequent handling such as mapping or assembly (Wolf, 2013). They have to be cleaned because they may contain adapters, empty reads, and low quality sequences that are generated through the sequencing process (Han et al., 2016, Chang et al., 2016). To address this, many programs have been used to remove Illumina adapter contamination and trimming such as Cutadapt (Martin and Wang, 2011), commercial software (eg, CLC Genomics Workbench) and free software like SeqTrim (Falgueras et al., 2010),

TagCleaner (Schmieder et al., 2010), AdapterRemoval (Schmieder et al., 2010), SOAP2 (Li et al., 2009), and HTSeq (Anders et al., 2014).

Cutadapt was used for this important task within the CGR's sequencing pipeline (Martin, 2011). A variety of file formats generated from next generation sequencing are supported by Cutadapt. Also, it can be searched for multiple adapters in one run to remove the best matched ones (Martin, 2011). This program has also been used for adapter removal in many other plant transcriptomic and genomic studies e.g. (Beletsky et al., 2017, Liu et al., 2016, Prabhudas et al., 2016).

Sickle is a sliding window that was used for quality trimming paired-end reads which have deteriorated towards both 3'- and 5'-end that could otherwise negatively impact on assembly or mapping of sequences. It is a tool to determine length thresholds when quality is sufficiently low or high enough to trim the 3'-end or 5'-end reads respectively (Moreno et al., 2016).

5.1.3.2 CyVerse (iPlant Collaborative)

CyVerse, formerly known as iPlant Collaborative, is a cyberinfrastructure funded by the National Science Foundation (http://www.iplantcollaborative.org) (Devisetty et al., 2016). It is particularly interesting to plant biologists (Goff et al., 2011). The purpose of this infrastructure is to provide different computational tools, high-performance and cloud computing services, and platforms for storing, sharing, and analyzing large and diverse biological datasets which are used for large scale science, domain experts and knowledge (Devisetty et al., 2016).

CyVerse's Discovery Environment (DE) is a web user interface to access the tools and computing resources (Devisetty et al., 2016). It is easy to use by users to run their analysis and computational survey and allows users to share their apps with collaborators and release them for public use without needing advanced skills of bioinformatics, command-line or linux (Devisetty et al., 2016), because most of the bioinformatics aspects of data management and analysis are handled in the DE.

There are three main features of DE, data store, apps, and analysis. Data store is cloud data storage which uses iRODS software infrastructure (http://www.irods.org) to support storage and transfer data (Oliver et al., 2013). Each user when registered

in CyVerse has 100 GB space for storage. File upload can use iDrop to upload data up to 5 GB (Oliver et al., 2013). The most common program used is Tuxedo Suite that includes RNA-seq analysis and all can run in command line: Bowtie, TopHat, Cufflinks, Cuffdiff, and CummeRbund (Ghosh and Chan, 2016).

Atmosphere is another part of the iPlant cyberinfrastructure and is an open-source, robust, configurable gateway that determines cloud infrastructure to cover computing needs for the plant science (Skidmore et al., 2011), and is a completely isolated operating system (Smith and Nair, 2005). Atmosphere tries to close the usability gap between a cloud provider and cloud users, particularly for researchers in plant science. Three logical layers are accompanied by a set of toolkits within the virtual machines (VMs) which are the cloud engine (toolkit configuration of the virtual machines), the middleware (communication), and the web frontend (interfacing with other parts) (Skidmore et al., 2011). The users are able to develop algorithms and install workflows, using less time and resources, and set up analyses. Also, it can provide access to the data from iPlant to their virtual machine as well as preserve the state of their VM instances, saving workflow, analysis and the entire system state (Skidmore et al., 2011).

A series of applications are preinstalled in DE. Several of these steps were used to analyze *Ophrys apifera* RNA-seq and CummeRbund. However, the result was unsatisfactory as the matching with reference was very low.

5.1.3.3Transcriptome assembly approaches:

The alignment of reads is one of the most important tasks in RNA-seq analysis (Garber et al., 2011). Therefore, computational methods are required for assembly of transcriptomes (Haas et al., 2013a). There are three categories of transcriptome assembly based on whether a reference genome is available or not, namely a reference-based strategy, a *de novo* strategy or a combined strategy (Martin and Wang, 2011). In the case of a reference-based strategy, a reference genome for the target transcriptome is available and assembly will be built upon it. When the transcriptomes do not have a reference genome, then '*de novo*' transcriptome assembly is the best way to assemble reads into transcripts. A combined strategy can be used involving both reference-based and *de novo* strategies for a more

comprehensive transcriptome as a reference- based assembly will take benefits of the high sensitivity and the other can detect alternatively spliced transcripts (Martin and Wang, 2011).

5.1.3.3.1 *De novo* transcriptome assembly

De novo assembly is a process which can put individual sequence reads together to produce long contiguous sequences ('contigs') (Paszkiewicz and Studholme, 2010). These assemblies are then used as a reference to map the paired-end reads. The number of reads that are aligned to contigs will be calculated and then used as a digital measure of gene expression (Vijay et al., 2013). Several assemblers have been designed for this, such as SOAPdenovo which is designed for rapidly assembling transcriptomes of large genomes sequenced as short-reads such as through the Illumina platforms (Johnson et al., 2012). Other assemblers have also been designed (e.g., Trinity, Oases, TransABySS) which are especially suitable for assembling transcriptomes where coverage varies from gene to gene because of differences in expression and alternative splicing (Johnson et al., 2012).

Trinity is a method for RNA-seq assembly and quickly became popular from 2011 onwards (Grabherr et al., 2011a). It is a useful and sensitive tool in non-model plant sequence assembly from the Illumina platform for transcriptome data and it helped to address the problem of short-reads (Li et al., 2013). The Trinity program combines three independent stages: Inchworm (initial assembly with more abundant k-mers), Chrysalis (de Bruijn graphs for each cluster using Inchworm contigs), and Butterfly (reconstructs transcripts from previous graphs and extracts full-length splicing isoforms using parallel computing) (Haas et al., 2013b).

5.1.3.4 Mapping assembly:

The mapping assembly for reads is based on using a reference sequenced genome, coding sequences, or a reference transcriptome as a template (Ghosh and Chan, 2016, Schmieder et al., 2010). In order to do an appropriate mapping, several points have to be considered. Firstly, choosing a suitable mapping tool especially when working with non-model species that will need to be aligned to a distant reference, such as the hybrid mapping strategy of stampy which is well equipped for differences of sequence distances up to 15% (Lunter and Goodson, 2011). Secondly, how to deal

with ambiguity in read mapping when alignment is with short reads (Wolf, 2013) because of increases in similarity between regions of the reference (e.g. by copy number variation, multigene families, repetitive domains) so that confidence in the reading mode at a particular site will decrease (Treangen and Salzberg, 2012). Finally, using mapping algorithms that can handle spliced read alignment is essential for mapping transcriptome data to a genomic reference.

Many packages are available for this purpose such as ERANGE (http://www.rna-seqblog.com) and ones for inference of alternative splicing are useful (Mortazavi et al., 2008). For example, study on the non-model plant *Rubus idaeus* L.(raspberry) involved both detection and calculation of differential gene expression relying on a normalised statistic, FPKM (fragments per kilobase of transcript per million reads mapped) (Ward et al., 2012). Consequently, the transcriptome data of *R. idaeus* L. was mapped to the genome of the closely related *Fragaria vesca* L. (woodland or alpine strawberry).

Burrows-Wheeler Transform is a common algorithm that is used in short read mapping tools for series matching that allows for rapid alignment and quality testing. It has been widely used as an aligner, for example in Bowtie (Langmead et al., 2009). TopHat is a spliced-read mapper that does not depend on a completely annotated genome and just needs the sequence of the raw genome as the backbone (Trapnell et al., 2009). Bowtie is described as performing ultrafast, memory-efficient alignment and is used for aligning short reads to large reference genomes (Langmead et al., 2009). This program is used with other bioinformatics algorithms such as TopHat (Trapnell et al., 2009), Cufflinks (Trapnell et al., 2012), and Bioconductor packages (Goff et al., 2012).

In this study the Illumina RNA-seq data obtained from *Ophrys apifera* samples was analyzed by using the assembled transcriptome as the reference for the alignment of the RNAseq reads. Alignment of read pairs was carried out using Bowtie version 1.2 (Langmead et al., 2009).

5.1.3.5 Transcript clustering:

RNA-seq is important for identifying genes that are differentially expressed among two or more condition (Oshlack et al., 2010). However, measuring differential

expression is a challenge when using a *de novo* assembled transcriptome since each gene can have multiple contigs that share sequence because the transcriptome assemblers create multiple versions among isoforms of the same gene.

It has been suggested that Corset software is the best method for obtaining gene-level counts after *de novo* transcriptome assembly (Davidson and Oshlack, 2014). It takes the multi-mapped contigs from a *de novo* assembly and maps reads to the transcriptome as input for Corset which will then cluster contigs with more than 10 supporting reads according to the shared reads and expression forms into genes (Davidson and Oshlack, 2014). Then, its output of mapped reads allows differential expression to be tested. The advantages of this software are that firstly, it works well compared to other clustering methods according to a variety metrics and it permits contigs and transcripts from different sources to be combined. Finally, it is easy to use and will run clustering and counting steps in one go.

5.1.3.6 Differential gene expression and gene annotation:

Gene expression analysis and gene annotation are the most important applications of RNA-seq. Gene expression can be quantified by mapping to the reference sequence, if one is available or if not the reads have to be assembled to reconstruct the transcriptome and then can be mapped to a reference genome for the identification of transcript levels (Trapnell et al., 2012, Wang et al., 2013a). Therefore, the read counts should be normalised for the detection of differential expression levels that account for sequencing to varied depths between lanes in the flow cell (Strickler et al., 2012b). The FPKM method can be used for calculating gene expression level, removing the amount of variation due to gene length and thus calculating gene expression which can be used directly for comparing between samples (Xiao et al., 2013).

Gene annotation is an important step in non-model organisms without reference genomes (Vijay et al., 2013). When there is an annotated reference genome, gene names can be assigned automatically, while in *de novo* assemblies, contigs do not give information about the sequenced gene, and their assignment to orthologous genes is not always clear (Vijay et al., 2013). Suffix-tree-based methods (NUCmer and PROmer) can be used for closely related species, while BLAST-based (e.g.

BLAST2GO, SATSUMA and PAPAYA) orthology detection can work well as an alternative at different levels of divergence (Wolf, 2013). Blast2Go is one approach that integrates various annotation approaches (Götz et al., 2008), and can be used to overcome the limitation of different data sets that have been used before (UniProt, SwissProt, Refseq, TAIR, Interpro and Rfam) (Pulman, 2014).

5.1.3.7 Gene function and interaction:

Getting a set of candidate genes is the ultimate return from a successful RNA-seq experiment along with annotation that indicates functions and interactions of the genes. The Gene Ontology (GO) cataloging system has been used for determination and identification of possible functions of differentially expressed genes in plants (Xiao et al., 2013) (http://www.geneontology.org/). Identification of gene functions can be divided into three main groups (cellular components, molecular functions and biological processes) and distributed into 55 classes such as cell, cell part, catalytic activity, organelle, cellular binding, metabolic processes, and response to stimulus (Xiao et al., 2013, Huang et al., 2015).

A gene ontology database using a controlled vocabulary is a valuable initiative for comparing gene functions cross species (Ashburner et al., 2000, Consortium, 2004). For example, functional annotations have led to more information about the complexity of floral structures and organ identity. This detail has broadened understanding of the interaction between MADS gene family members and floral morphology in the orchid (Xiao et al., 2013).

Other important tools have been designed for automatic functional annotation particularly for *de novo* assembled transcriptomes in model or non-model organisms such as Trinotate (https://trinotate.github.io/#OutputReport). This includes a number of different methods based around homology search for sequence data (BLAST, UniProt), to identify protein domains (HMMER, PFAM), predict signal peptides (signalP, tmHMM) and use comparison with several annotation databases (eggNOG, GO, Kegg databases). Then, all the functional annotation data obtained from these steps can be combined into a SQLite database which is fast for searching terms related to scientific hypotheses and can output an entire annotation report for the transcriptome.

5.2 Materials and methods:

5.2.1 Selection of plant material:

The plant materials were aseptic protocorms, obtained from seed germination *in vitro* (see chapter 3), and fresh roots and corms of bee orchid plants from Ground Biosciences. Whole corm of the plants was cut into pieces using a sterilized knife and each piece used for RNA extraction weighed 110 mg. In case of root samples, 2-3 roots were used per extraction because of their light weight (37-57 mg). The sample that was collected from the part between shoot and corm (SC) weighed 146 mg. Each sample was labelled, wrapped in aluminum foil and flash frozen in liquid nitrogen to freeze tissue before RNA degradation. The samples were then stored at -80°C, until ready for RNA extraction.

5.2.2 RNA extraction and purification:

Several methods were tested to obtain RNA suitable for sequencing. Finally, RNA was extracted from both protocorms and corms using a commercial kit (InnuPREP Plant RNA kit, Sigma, Analytic Jena, Germany). The TRIzol reagent protocol (Ambion RNA by Life Technologies) was used for extraction of root tissue to obtain higher levels of RNA.

A mortar and pestle baked in an oven at 180°C for at least 8 hours was used. The frozen tissues were ground to a powder in liquid nitrogen. After grinding, ~100 mg was transferred into 1.5 ml Eppendorf tubes kept in liquid nitrogen to avoid thawing. RNA samples were prepared following the manufacturer's instructions using the innuPREP Plant RNA kit and TRIzol reagent protocols.

A wash-up RNA step was used with the innuPREP Plant RNA kit and after the TRIzol reagent protocol to remove DNA, by adding a column DNAase treatment (RNase-Free DNase Set, Qiagen).

5.2.3 Quantity and quality assessment of RNA:

Samples were selected for sequencing depending on the quality and quantity of RNA. RNA was tested using a NanoDrop spectrophotometer (Thermo, USA) and Qubit Quant-iT™ RNA Assay Kit (Thermo Fisher Scientific, USA). Nine samples, three from each tissue were sent to the CGR, University of Liverpool, for RNA

sequencing. Before sequencing, all samples underwent further quality control using Qubit and a size distribution by the CGR using the Agilent 2100 Bioanalyser (Agilent Technologies).

5.2.4 rRNA depletion and library preparation:

Total RNA samples were treated using the Ribo-Zero Magnetic kit (plant root / seed) from Illumina (USA) according to the manufacturer's instruction. The NEBNext® Ultra™ Directional RNA-Seq library prep kit for Illumina®#E7420 was used for library preparation. The libraries were sequenced as one pool on one lane of the HiSeq4000 with version 1 chemistry using sequencing by synthesis providing 2x150 bp paired-end reads.

5.2.5 Data pre-processing:

The sequence data quality was assessed by Dr. Richard Gregory using an in-house pipeline. The output was indexed reads that were transformed by CASAVA version 1.8.2 (Illumina) to produce data in fastq format. The raw fastq files were trimmed to remove Illumina adapter sequences using Cutadapt version 1.2.1. The option -O 3 was used, so the 3' end of any reads which matched the adapter sequence for a minimum of 3 bp was trimmed away. The reads were further trimmed using Sickle version 1.200 with a window quality score of at least 20. Reads shorter than 10 bp were removed. After these steps, the CGR data processing was completed and RNA-seq in fastq format was released and ready for further analyses. After filtering, each sample has three reads files R1 (Forward reads), R2 (Reverse reads), and R0 (Unpaired reads).

5.2.6 Data analysis:

Initial attempts to use the tools within CyVerse to map the data were not very successful, giving poor coverage (see section 5.3.7.1). Therefore, the data was passed to the CGR Bioinformatics Team (Centre for Genomic Research) for *de novo* assembly, mapping and differential expression analysis.

The steps described in 5.2.6.1 were performed by Mathew Gemmell (research associate) of the CGR Bioinformatics Team, University of Liverpool.

5.2.6.1 *De novo* assembly:

There is no reference genome for genus *Ophrys*. Trinity v 2.4.0 (Grabherr et al., 2011a) with default parameters was used for transcriptome assembly.

5.2.6.2 Alignment of reads to transcriptome assembly:

The assembled transcriptome was used as the reference for read alignment. Reads were aligned to the *de novo* assembly using Bowtie v 1.2 (Langmead et al., 2009), accepting all valid alignments.

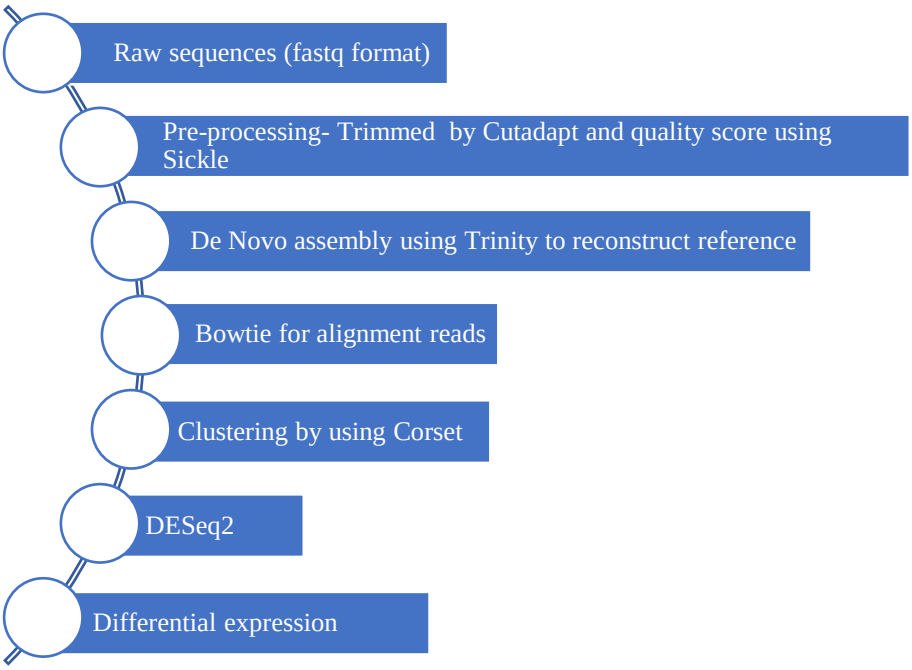


Figure 5.1 RNA-seq workflow performed by CGR.

5.2.6.3 Clustering alignment reads:

Clustering of transcripts was performed using Corset v 1.06 (Davidson and Oshlack, 2014) with default parameters because of the fragmented nature of the assembly and high redundancy. Contigs with less than 10 reads were removed to make the number of clusters more manageable. Then, the filtered contigs were clustered and categorized depending on the differential expression levels of the contigs.

5.2.6.4 Differential expression (DE):

5.2.6.4.1 Negative binomial distributions:

The expression values for each sample group were formulated as a parameter using a Generalised Linear Model (GLM) (Nelder and Baker, 1972), and the pairwise comparison was run according to the result of model fitting from the next step.

5.2.6.4.2 Model fitting and variance estimation for the data set:

Normalisation factors were calculated to correct for differences in library size between samples, which might otherwise cause bias in the differential expression analysis. Therefore, the median ratio method was applied in DESeq2, and again with default parameters (Anders and Huber, 2010). Also, both trended (dispersion-mean relationship) and genome-wise dispersion parameters were estimated. Significance testing was performed using 'Gene'-wise (i.e. transcript-wise) dispersion.

5.2.6.4.3 Testing log fold change (logFC):

The output of GLM parameters was used for log2 fold change of expression values (logFC) between sample groups per comparison. Then, it was tested in DESeq2 using a Wald test, and the P-values were adjusted for multiple testing using the False Discovery Rate (FDR) approach (Benjamini and Hochberg, 1995). A FDR-adjusted P-value < 5% was defined to identify significant DE transcripts.

5.2.6.4.4 Clustering analysis of detected DE clusters

The DE transcripts detected from each contrast (protocorms vs corms vs root) were clustered based on the logFC value using the k-means method and visualized with a heatmap.

5.2.6.5 Annotation:

Trinotate (trinotate.sourceforge.net), a Trinity tool, was used to annotate the transcripts. Data from the orchid *Apostasia shenzhenica*: ASH160606 (Zhang et al., 2017), was added as custom annotation.

5.3 Results

5.3.1 Preparation of RNA suitable for sequencing:

Several RNA extraction protocols were tested using protocorms, corms, roots and leaf tissues. For sequencing the RNA samples had to contain at least 100 ng/µl of total RNA (Bhargava et al., 2014), and a ratio of sample absorbance at 260 / 280 nm of ~2.0 was accepted as pure RNA (www.nanodrop.com). Also, RNA Integrity value (RIN) is important and a value greater than 8.0 is usually considering high quality RNA, but 7.0 is also accepted as good plant RNA samples for sequencing (Johnson et al., 2012, Ward et al., 2012).

Firstly, a combination method using CTAB and TRIzol with sarkosyl (option 2) was used to extract RNA from corms (Jordon-Thaden et al., 2015), but gave a very poor RNA yield of around 25-95 ng/µl. The rapid CTAB method (Gambino et al., 2008) was performed for extraction of RNA from leaf tissue, but again the yield was not good (14-18 ng/µl).

The RNeasy Plant Mini Kit (QIAGEN, Germany) was used for extracting RNA from leaf and the result was a higher yield of RNA, but this was only just sufficient 266-294 ng/µl.

TRIzol® reagent protocols have been designed to isolate high quality RNA, DNA and proteins rapidly from cell and tissue samples of animal, plant, yeast, or bacterial origin (Hummon et al., 2007, Chomczynski and Sacchi, 1987, Chomczynski, 1993). This was used for root samples and provided a good amount of RNA ~1000-2000 ng/µl (see Table 5.1) which was cleaned with a DNase treatment kit (RNase-Free DNase Set, Qiagen). Unfortunately, this method was not successful with corm and protocorms.

Finally, the innuPREP Plant RNA Kit (Analytic Jena, Germany) provided adequate quantity and quality of RNA from the protocorms and corms (Table 5.1).

Table 5. 1: RNA samples selected for RNA-seq.

Sample Name		Total Volume (µl)	Qubit (ng/ µl)	Nano drop (ng/µl)
Protocorms (Group 1)	Sample_1-P1	20	256	224
	Sample_2-P2	28	92.9	75.2
	Sample_3-P3	28	175	141
Roots (Group 2)	Sample_7-R3	22	710	2128
	Sample_8-R5	30	156	1248
	Sample_9-R6	32	166.3	998
Corms (Group 3)	Sample_12-SC3	24	271	228
	Sample_13-SC4	25	11	242.6
	Sample_14-Corm	25	85.6	109.51
SC= Tissue taken from upper of corm where it merges with the shoot.				

5.3.2 Quality and integrity of RNA samples performed by CGR:

Assessment of the integrity of the RNA samples using the Bioanalyser Agilent 2100 showed that most samples had good integrity. A RIN value of more than 8 was required to indicate a lack of degradation and the value for all samples was 8.6-10. However, the ratio of 28S:18S, at 1.2-2.2, was lower than the acceptable ratio of 2.0. Nevertheless, the RNA peaks were clearly more intense at the 28S region than 18S (Figure 5.2 A-5.2 K). Samples replicates were therefore taken forward to rRNA depletion before cDNA library creation. Representative Bioanalyser graphs of protocorm and root samples are in Figures 5.2 A (ladder), B-D (protocorms), E-G (roots), 5.2 H (ladder for corm samples only), I-K (corms).

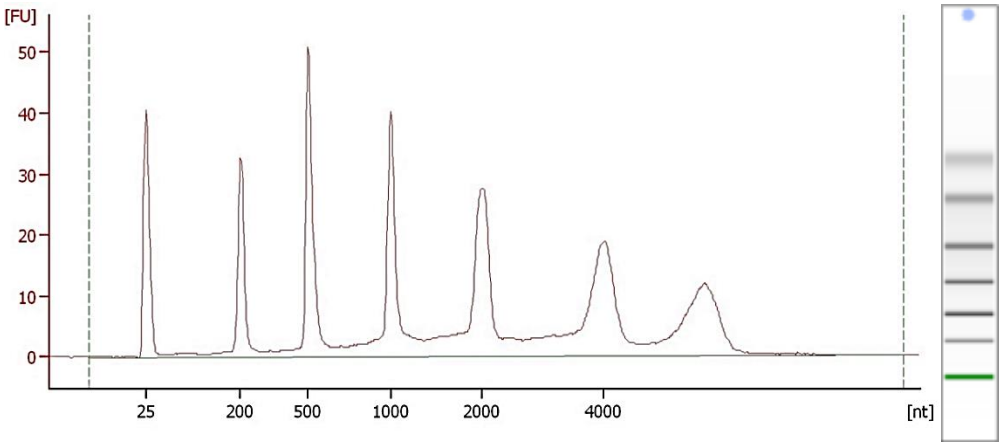


Figure 5.2 A: Electropherogram of ladder (RNA area: 471.2 and RNA concentration: 1000 pg/µl)A: Electropherogram of ladder (RNA area: 471.2 and RNA concentration: 1000 pg/µl). FU = fluorescence units; nt = nucleotide

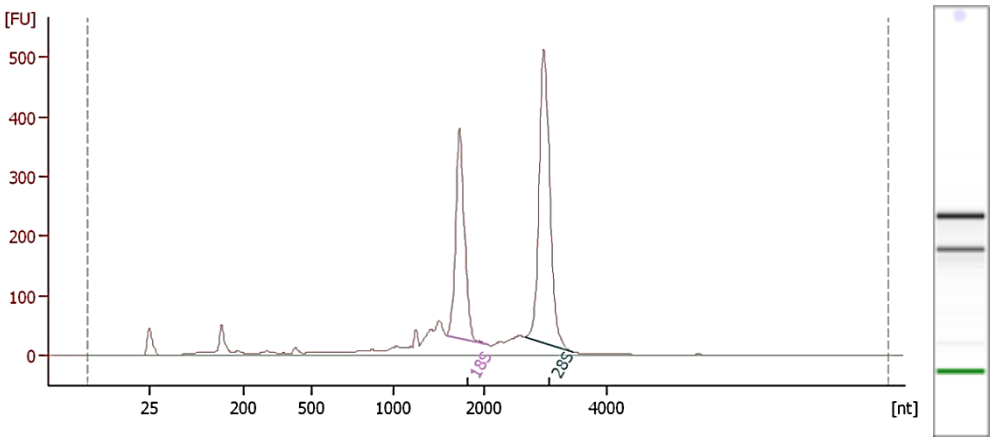


Figure 5.2 B: Electropherogram of total RNA for protocorm (Sample_1-P1). rRNA ratio (28S/18S) was 1.5, RIN = 9.10, RNA concentration: 4,805 pg/µl and RNA area: 2264.0.

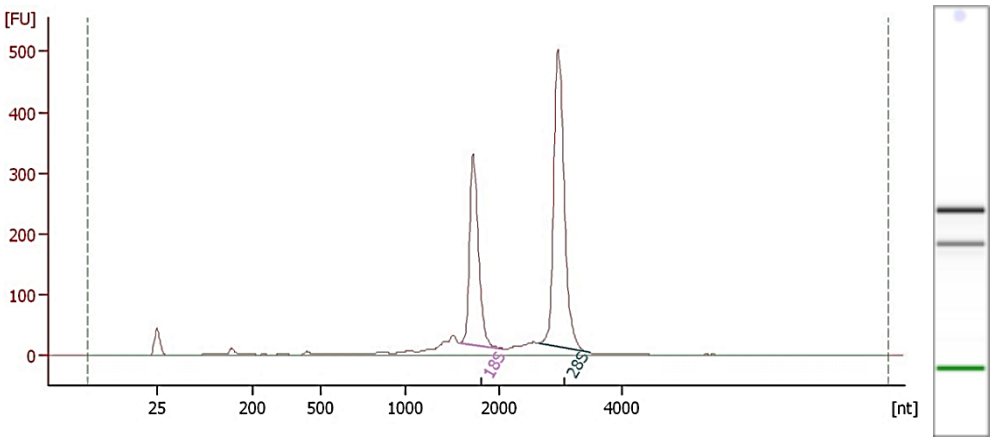


Figure 5.2 C: Electropherogram of total RNA for protocorm (Sample_2-P2). rRNA ratio (28S/18S) was 1.7, RIN = 9.80, RNA concentration: 3670 pg/µl and RNA area: 1729.1.

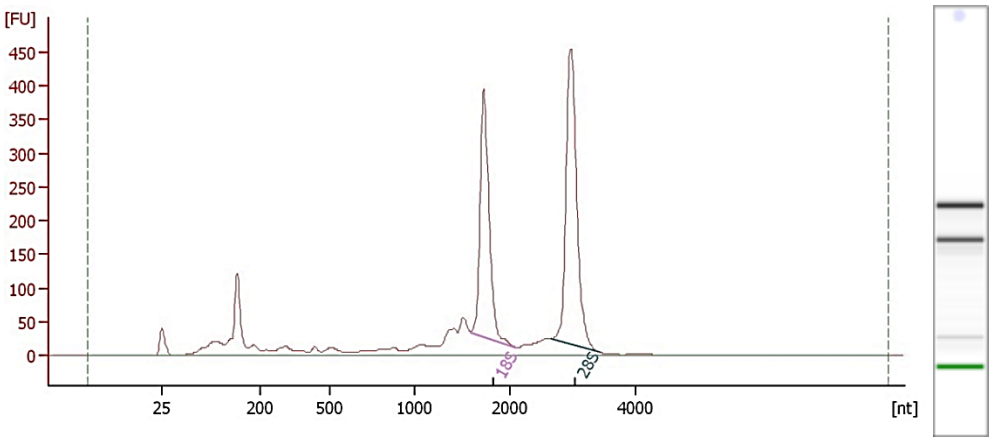


Figure 5.2 D: Electropherogram of total RNA for protocorm (Sample_3-P3). rRNA ratio (28S/18S) was 1.3, RIN = 8.80, RNA concentration: 4991 pg/µl and RNA area: 2351.6.

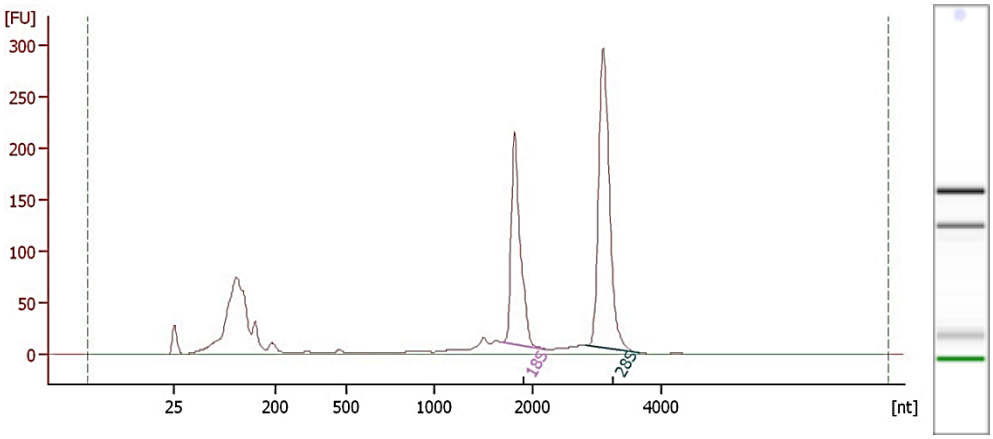


Figure 5.2 E: Electropherogram of total RNA for root (Sample_7-R3). rRNA ratio (28S/18S) was 1.6, RIN = 9.60, RNA concentration: 3014 pg/µl and RNA area: 1420.2.

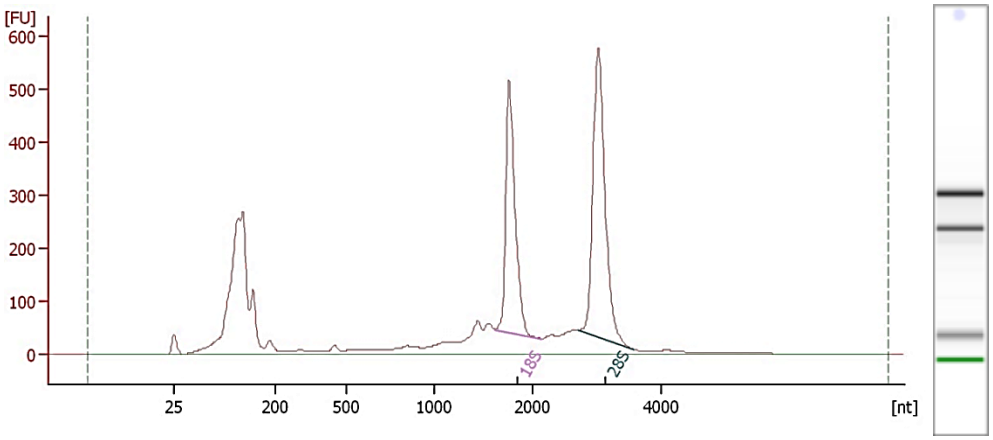


Figure 5.2 F: Electropherogram of total RNA for root (Sample_8-R5). rRNA ratio (28S/18S) was 1.4, RIN = 8.60, RNA concentration: 8834 pg/µl and RNA area: 4162.5.

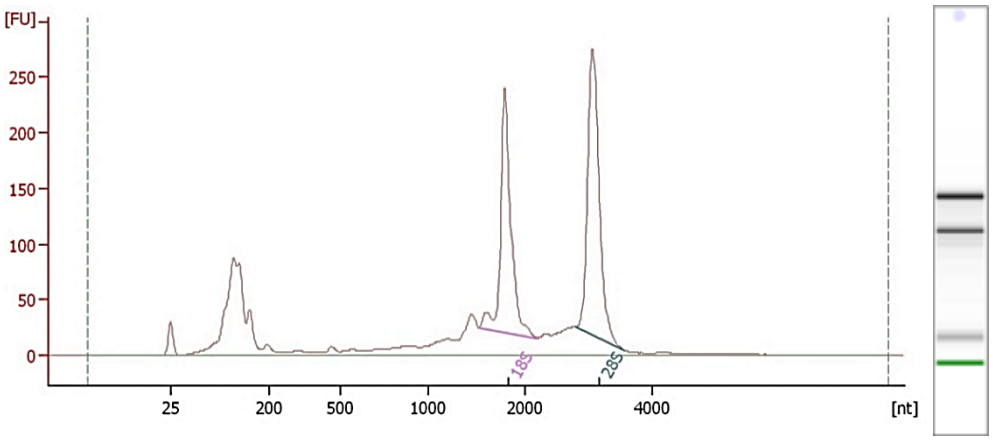


Figure 5.2 G: Electropherogram of total RNA for root (Sample_9-R6). rRNA ratio (28S/18S) was 1.2, RIN = 8.70, RNA concentration: 3962 pg/µl and RNA area: 1866.7.

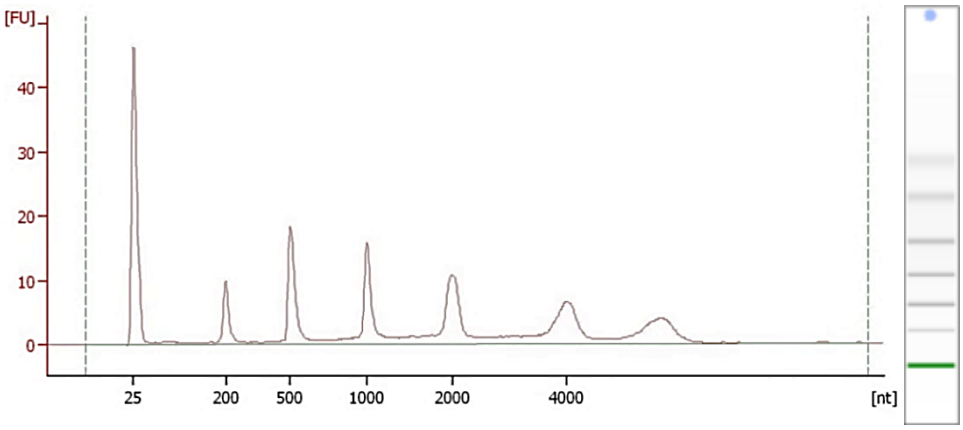


Figure 5.2 H: Electropherogram of ladder (RNA area: 171.9 and RNA concentration: 1000 pg/µl); FU = fluorescence units; nt = nucleotide. This ladder used for corm samples.

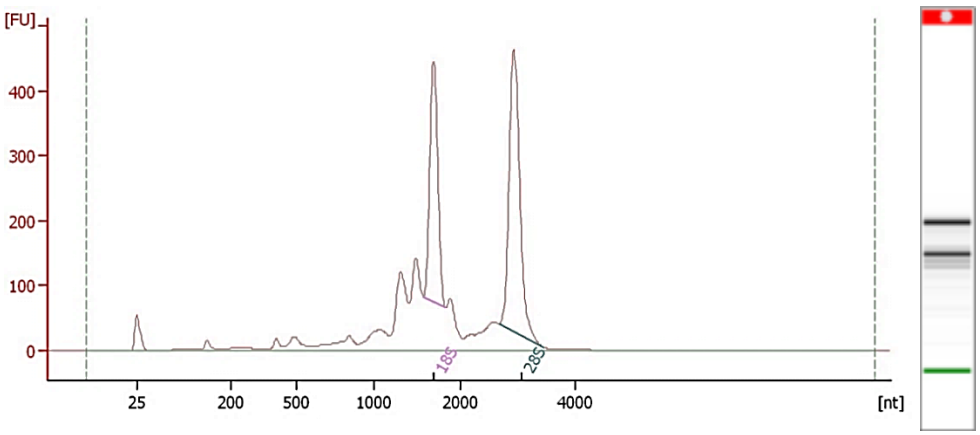


Figure 5.2 I: Electropherogram of total RNA for corm (Sample_12-SC3). rRNA ratio (28S/18S) was 1.4, RIN = N/A, RNA concentration: 16546 pg/µl and RNA area: 2844.5.

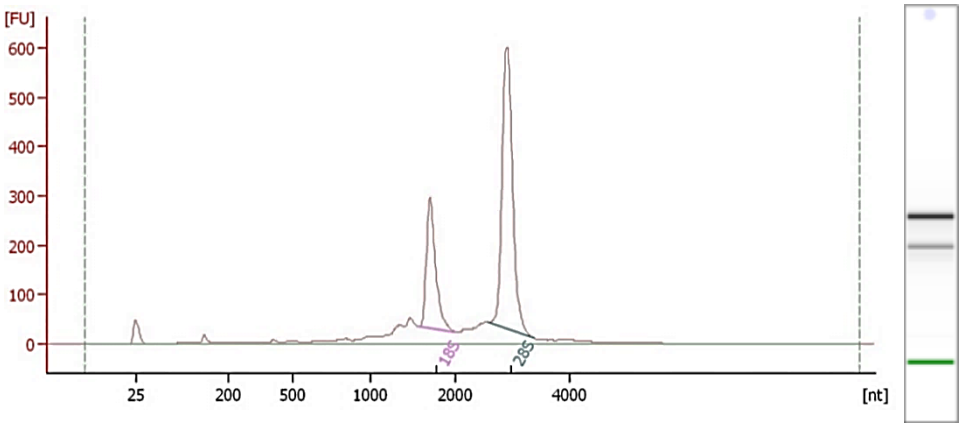


Figure 5.2 J: Electropherogram of total RNA for corm (Sample_13-SC4). rRNA ratio (28S/18S) was 2.2, RIN = 8.90, RNA concentration: 14033 pg/µl and RNA area: 2412.5.

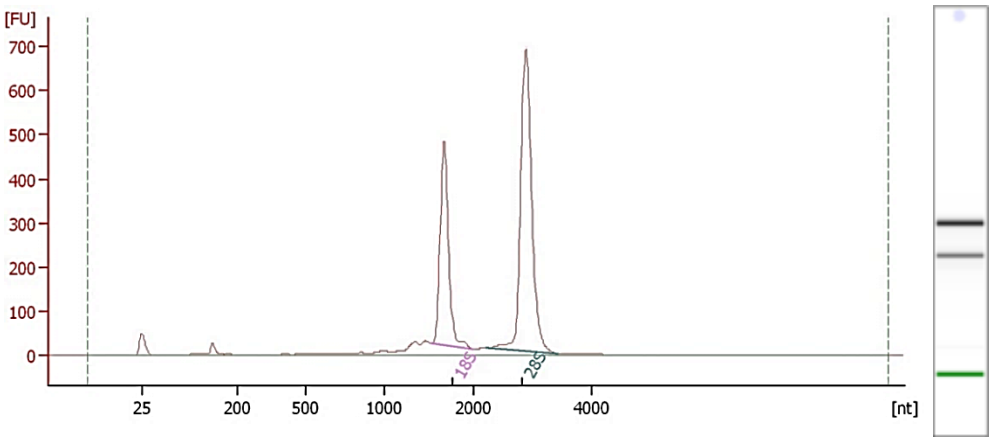


Figure 5.2 K: Electropherogram of total RNA for Corm (Sample_14-Corm). rRNA ratio (28S/18S) was 1.7, RIN = 10, RNA concentration: 14241 pg/µl and RNA area: 2448.3.

5.3.3 rRNA depletion

The RiboZero Plant Root kit was used to treat RNA for removal of rRNA from the total RNA samples. Only 1-5 µg of total RNA is required for library preparation. The amount of RNA is reduced after treatment and amount of depleted RNA obtained relies on the amount of RNA input, rRNA content of the sample, and the method used for depletion. Figures 5.3 A-J show that rRNA depletion from most samples was sufficient, but small peaks in the 18S region remained in samples protocorm 1-P1, root 8-R5, and corm 14-corm. Although the electropherogram traces were not ideal for the decision, it was continued with these samples and they were sent them for library preparation and sequencing.

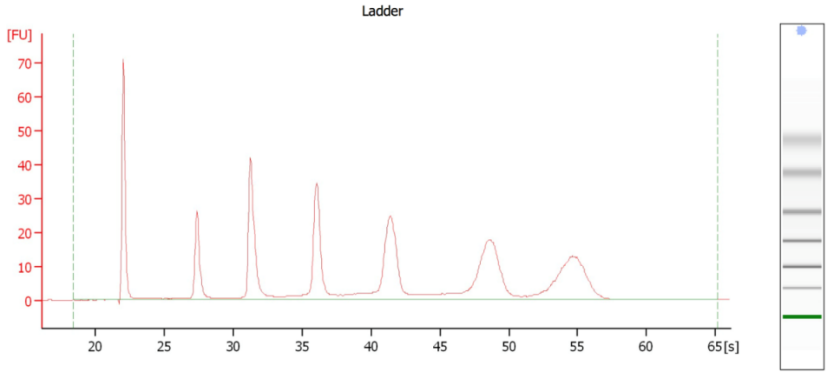


Figure 5. 3 A: Electropherogram of ladder (RNA area: 397.5 and RNA concentration: 1000 pg/µl); after RiboZero treatment. This ladder used for all samples.

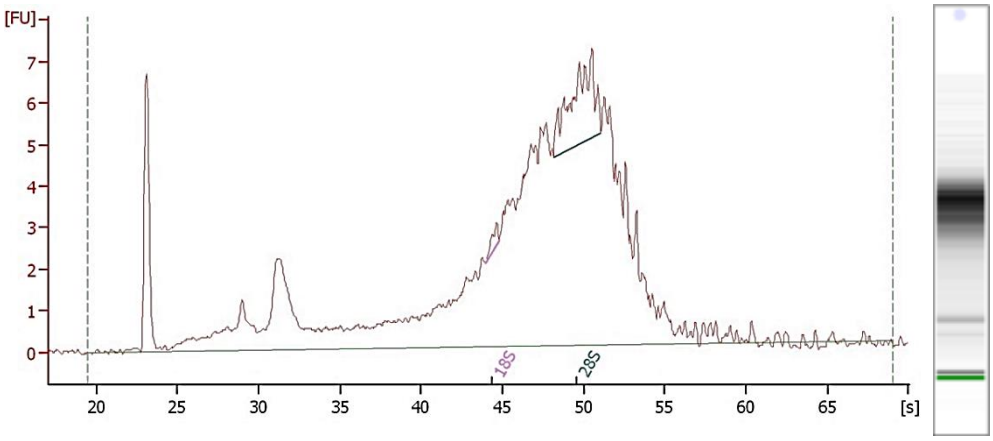


Figure 5.3 B: Electropherogram after rRNA depletion for protocorm (Sample_1-P1), after RiboZero treatment. Small peak is visible at 18S region. rRNA ratio (28S / 18S) was 12.4.

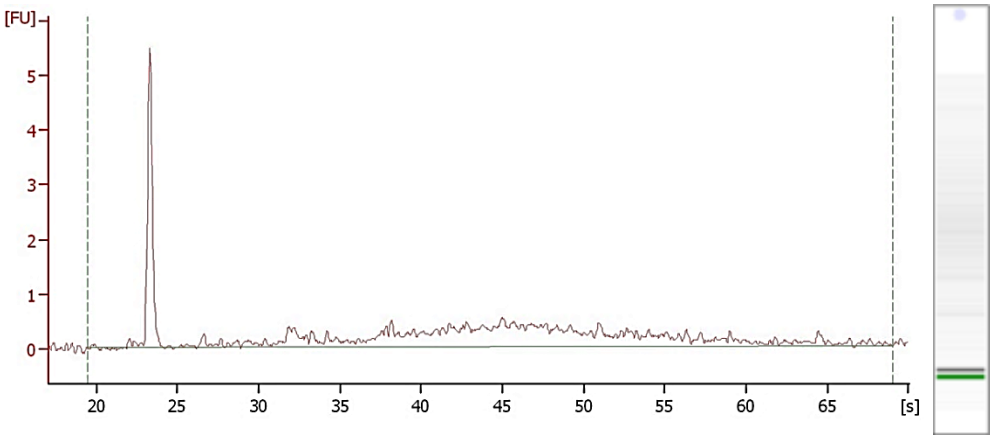


Figure 5.3 C: Electropherogram after rRNA depletion for protocorm (Sample_2-P2), after RiboZero treatment. No peaks are visible for both 28S and 18S regions. rRNA ratio (28S / 18S): 0.0

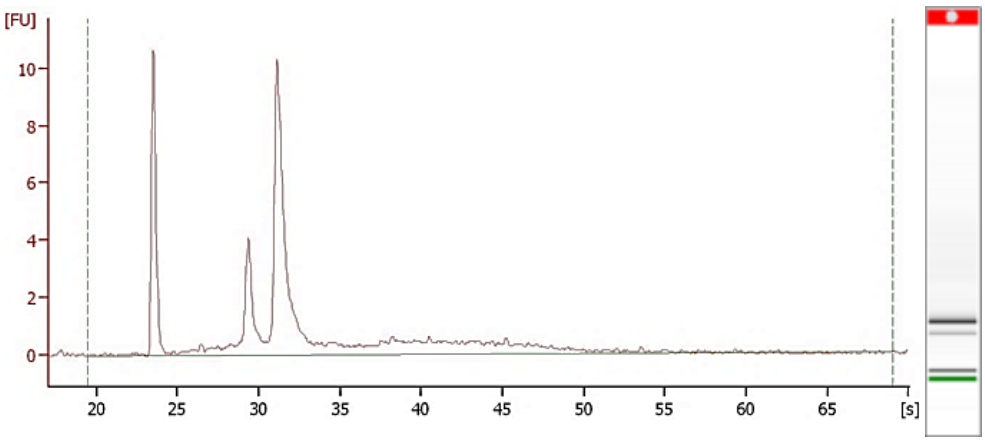


Figure 5.3 D: Electropherogram after rRNA depletion for protocorm (Sample_3-P3), after RiboZero treatment. No peaks are visible for both 28S and 18S regions. rRNA ratio (28S / 18S): 0.0

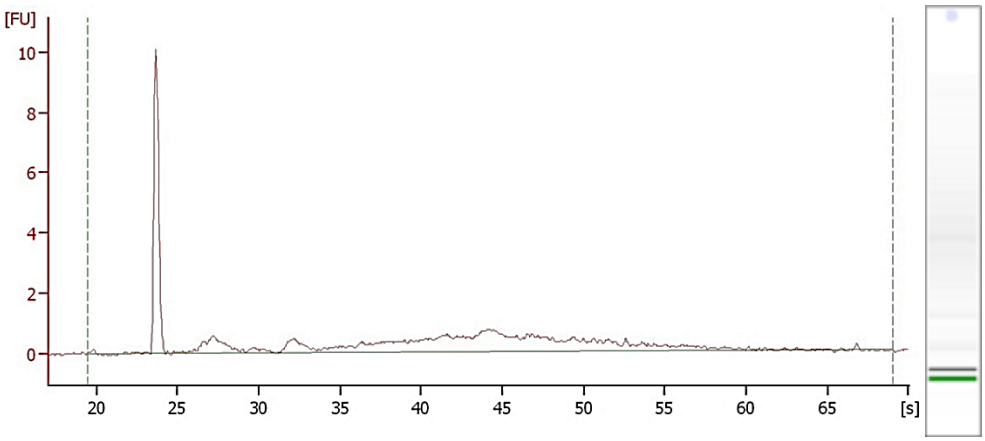


Figure 5.3 E: Electropherogram after rRNA depletion for root (Sample_7-R3), after RiboZero treatment. No peaks are visible for both 28S and 18S regions. rRNA ratio (28S / 18S): 0.0

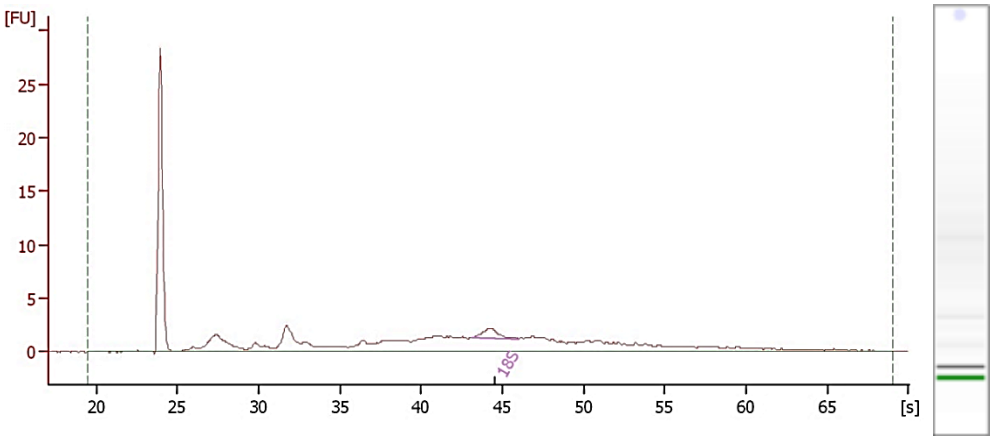


Figure 5.3 F: Electropherogram after rRNA depletion for root (Sample_8-R5), after RiboZero treatment. No peak is visible for 28S, but there is small peak at 18S regions. rRNA ratio (28S / 18S) 0.0

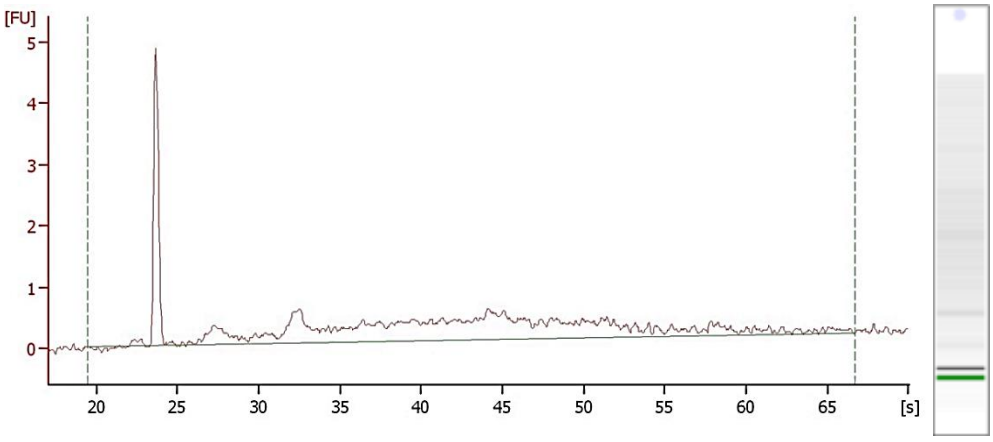


Figure 5.3 G: Electropherogram after rRNA depletion for root (Sample_9-R6), after RiboZero treatment. No peaks are visible for both 28S and 18S regions. rRNA ratio (28S / 18S): 0.0

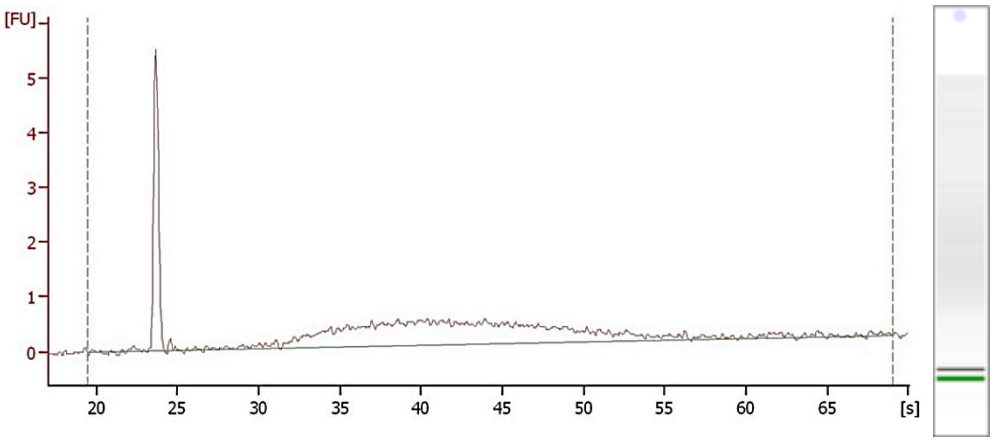


Figure 5.3 H: Electropherogram after rRNA depletion for corm (Sample_12-SC3), after RiboZero treatment. No peaks are visible for both 28S and 18S regions. rRNA ratio (28S / 18S): 0.0

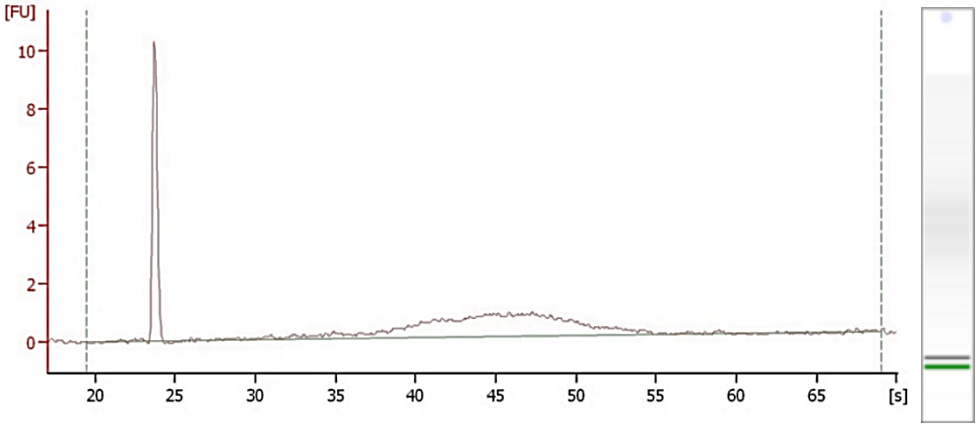


Figure 5.3 I: Electropherogram after rRNA depletion for corm (Sample_13-SC4), after RiboZero treatment. No peaks are visible for both 28S and 18S regions. rRNA ratio (28S / 18S): 0.0

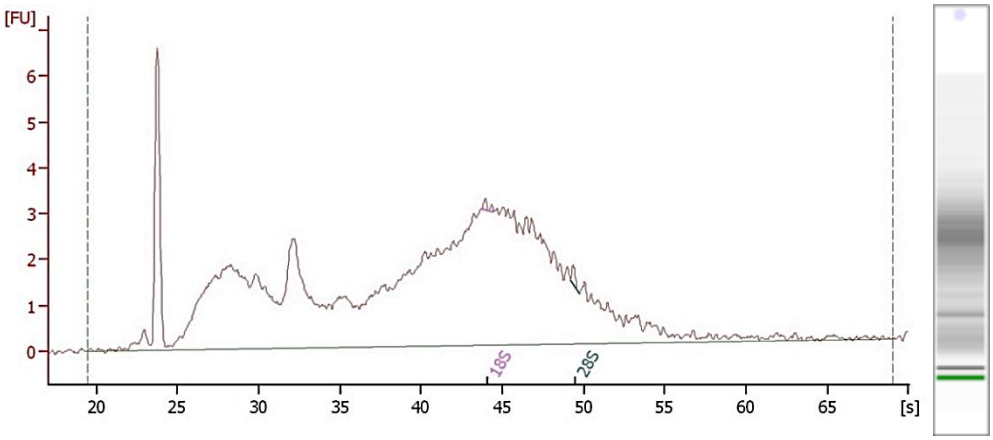


Figure 5.3 J: Electropherogram after rRNA depletion for corm (Sample_14-corm), after RiboZero treatment. There are peaks visible but not very clear for both 28S and 18S regions. rRNA ratio (28S / 18S) 1.3.

5.3.4 Initial pre-processing (trimming and filtering of raw reads)

Table 5.2 shows the results of the raw fastq for all samples after adapter trimming and filtering. All samples have average base quality scores of >36, which means a base call accuracy of 99.9%. More than 98% of total raw reads were recovered after trimming and filtering.

For example, in protocorms samples P1, 98.90%, P2 (98.20%), and P3 (99.29), in root samples R3 (99.25%), R5 (98.61%), and R6 (99.14%), and finally in corms, SC3 (99.00%), SC4 (98.77%), and Corm (98.58%).

The total number of reads is shown in Figure 5.4 below, obtained from each replicate of protocorm, root, and corm showing paired reads, singlet reads and discarded reads as well as poor quality or adapter contamination. The number of reads obtained from all samples under study was more than 60 million reads. The majority were paired reads.

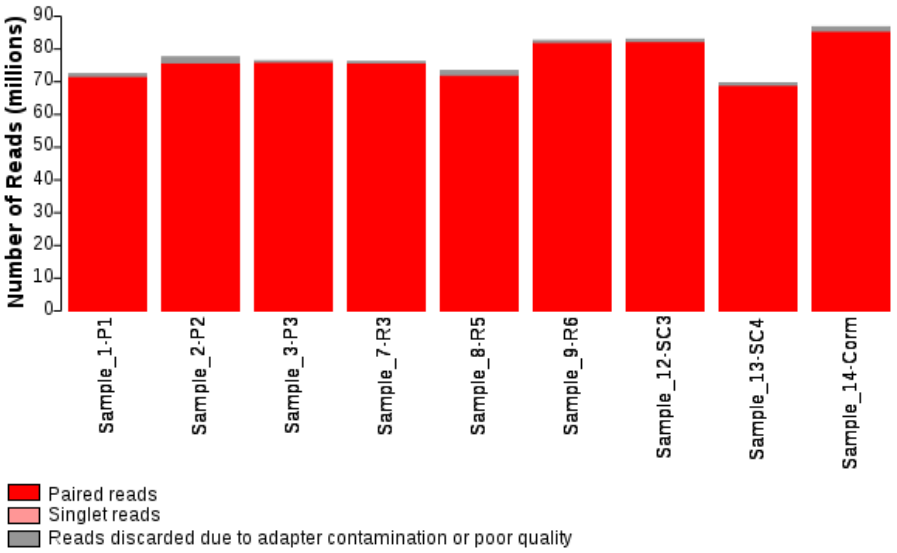


Figure 5.4 Total number of reads in millions obtained from each replicate of protocorm (Sample_1-P1, Sample_2-P2 and Sample_3-P3), root (Sample_7-R3, Sample_8-R5 and Sample_9-R6), and corm (Sample_12-SC3, Sample_13-SC4, and Sample_14-Corm).

The distribution of trimmed read length for forward, reverse and singlets is shown in Figure 5.5. Red lines show median read length within the interquartile range (box) while whiskers indicate minimum and maximum read lengths which are represented as lines extending vertically from the boxes. This shows that the median trimmed read length is near to the maximum read lengths.

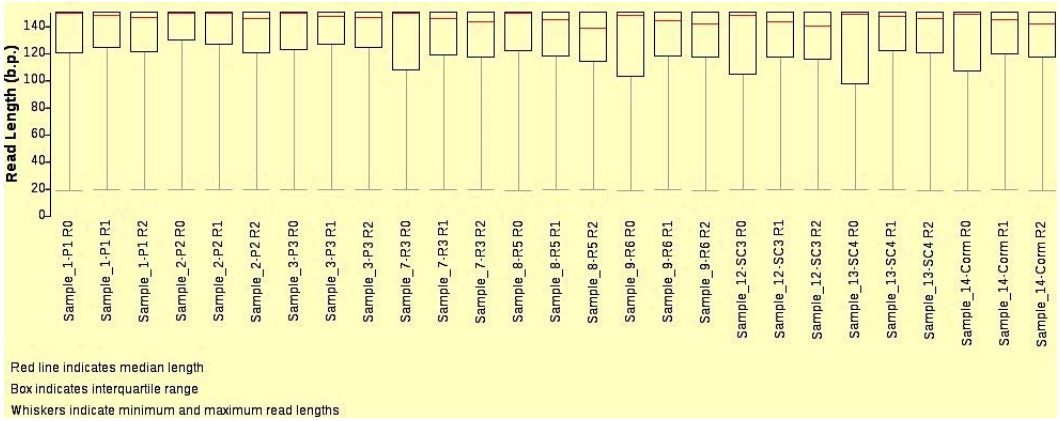


Figure 5. 5: Boxplot showing the distribution of trimmed read lengths for the forward (R1), reverse (R2) and singlet (R0) reads of all replicates for protocorm (Sample_1-P1, Sample_2-P2 and Sample_3-P3), root (Sample_7-R3, Sample_8-R5 and Sample_9-R6), and corm (Sample_12-SC3, Sample_13-SC4, and Sample_14-Corm).

Table 5. 2: Comparing Illumina reads of samples with their replicates after data trimmed and filtered.

Sample	Read		Total number of reads	Maximum read length (bp)	Average read length (bp)	Read length SD (bp)	Minimum base quality score	Maximum base quality score	Average base quality score	Base quality score SD
Sample_1-P1	forward	raw	36,015,391	150	150.0	0.0	2	41	39.4	3.8
		trimmed	35,419,668	150	135.9	19.5	2	41	39.5	3.6
	reverse	raw	36015391	150	150.0	0.0	2	41	38.5	5.4
		trimmed	35419668	150	133.6	22.1	2	41	38.8	4.9
Sample_1-P2	forward	raw	38576411	150	150.0	0.0	2	41	39.5	3.7
		trimmed	37471631	150	136.9	19.0	2	41	39.5	3.5
	reverse	raw	38576411	150	150.0	0.0	2	41	37.8	6.4
		trimmed	37471631	150	131.8	24.9	2	41	38.2	5.6
Sample_3-P3	forward	raw	37996662	150	150.0	0.0	2	41	39.4	3.8
		trimmed	37580675	150	137.2	17.9	2	41	39.5	3.6
	reverse	raw	37996662	150	150.0	0.0	2	41	38.7	5.0
		trimmed	37580675	150	135.6	19.9	2	41	38.9	4.7
Sample_7-R3	forward	raw	37876331	150	150.0	0.0	2	41	39.4	3.8
		trimmed	37478937	150	133.4	21.0	2	41	39.4	3.7
	reverse	raw	37876331	150	150.0	0.0	2	41	39.0	4.6
		trimmed	37478937	150	132.1	22.0	2	41	39.1	4.3
Sample_8-R5	forward	raw	36494121	150	150.0	0.0	2	41	39.4	3.8
		trimmed	35620159	150	132.9	21.4	2	41	39.5	3.5
	reverse	raw	36494121	150	150.0	0.0	2	41	38.0	6.1
		trimmed	35620159	150	128.7	25.5	2	41	38.4	5.4
Sample_9-R6	forward	raw	41089363	150	150.0	0.0	2	41	39.4	3.8
		trimmed	40596741	150	132.7	21.4	2	41	39.5	3.7
	reverse	raw	41089363	150	150.0	0.0	2	41	39.0	4.6
		trimmed	40596741	150	131.5	22.4	2	41	39.2	4.2
Sample_12-SC3	forward	raw	41261318	150	150.0	0.0	2	41	39.4	3.8
		trimmed	40698695	150	132.2	21.7	2	41	39.4	3.7
	reverse	raw	41261318	150	150.0	0.0	2	41	38.9	4.8
		trimmed	40698695	150	130.6	23.1	2	41	39.0	4.5
Sample_13-SC4	forward	raw	34640419	150	150.0	0.0	2	41	39.4	3.8
		trimmed	34074300	150	134.9	20.5	2	41	39.5	3.7
	reverse	raw	34640419	150	150.0	0.0	2	41	38.9	4.8
		trimmed	34074300	150	133.4	21.9	2	41	39.0	4.4
Sample_14-Corm	forward	raw	43143286	150	150.0	0.0	2	41	39.4	3.7
		trimmed	42294066	150	133.5	20.6	2	41	39.5	3.6
	reverse	raw	43143286	150	150.0	0.0	2	41	38.7	5.2
		trimmed	42294066	150	131.6	22.5	2	41	38.9	4.7

5.3.5 Gene/ transcript expression analysis

5.3.5.1 Sequence alignment:

The trimmed RNA-seq reads in fastq format required a reference genome for alignment. Initially, transcriptome data was mapped to the genome data of *Arabidopsis thaliana* and to the orchid *Phalaenopsis equetris* that were preinstalled in the CyVerse databases. TopHat 2.1.1 was used for alignment of reads. However, only 0.6-0.8% (*Arabidopsis thaliana*) and 0.9-1.4% (*Phalaenopsis equetris*) of the *Ophrys apifera* reads mapped to each, which showed very poor mapping coverage with both. An *Ophrys* transcriptome assembly (Sedeek et al., 2013) was downloaded from this website (https://doi.org/10.1371/journal.pone.0064621.s011) and used as a final attempted for RNA-seq read alignment using TopHat.

The result is shown in Table 5.3 that mapping coverage was aligned from about 17.1% to 34.7% of total input. It is obvious from the result that three samples (Sample_3-P3 (17.1%), Sample_9-R6 (18.2%) and Sample_8-R5 (19.9%) have low percent of coverage (Table 5.3). Generally, the result was low and therefore, Trinity was used to reconstruct reference from data and then mapped with it. This was more satisfactory.

As consequence, *de novo* assembly was undertaken within the Bioinformatics Team, CGR to produce better coverage and continue the data analysis. The results of this assembly are described from section 5.3.5.2 onwards.

Table 5. 3: Output result obtained from *Ophrys apifera* RNA-seq using TopHat2.1.1-PE (Discovery Environment, CyVerse) mapped with *Ophrys* transcriptome (Sedeek et al., 2013).

Samples	Read alignment rate (% of input)		Overall read alignment rate (%)	Concordant pair alignment rate (%)
	Left read	Right read		
Sample_1-P1	25.4	25.0	25.2	22.0
Sample_2-P2	35.4	34.0	34.7	29.5
Sample_3-P3	17.1	17.1	17.1	14.7
Sample_7-R3	21.9	33.1	27.2	4.5
Sample_8-R5	20.2	19.6	19.9	17.1
Sample_9-R6	18.3	18.0	18.2	15.9
Sample_12-SC3	29.7	33.6	31.4	6.1
Sample_13-SC4	33.9	21.7	27.4	4.6
Sample_14-Corm	33.6	29.1	31.3	6.0

5.3.5.2 De novo assembly using Trinity:

Trinity was used to build the reference for the transcriptome. The aim was to constructing k-mer graphs to define components and then reads were assigned to these graphs. The output of assembled transcriptome was used as reference for mapping at the next step (Table 5.4 and 5.5).

The raw assembly consisted of 1.92 million transcripts. Broadly, Trinity works by constructing k-mer graphs to define contiguous overlapping sequence (a "component"), then assigning reads to these graphs. In some cases, during assembly, putative incorrect components are split into non-overlapping "sub-components". Different "paths" across the final set of graphs can indicate splice variants (that share some, but not all, common sequence regions). In the Table 5.4 the values of N50 and N75 indicate length of contigs. The coverage of these assembled sequences was distributed between <500bp to 50kb as shown in Table 5.5. Thus, the higher coverage of 1,367,948 was at the region of <500bp and the lower coverage region at 10-50kb.

Table 5. 4: The output of Trinity assembly.

Assembly name	Assembled sequences	size range	N50 [1]	N75 [2]
Trinity	1,916,614	201-16783 bp	1,687 bp	914 bp

Table 5.5: Size distribution of assembled sequences**.**

Assembly name	<500bp	500bp-1kb	1-5kb	5-10kb	10-50kb
Trinity	1,367,948	301,324	239,413	7,660	269

5.3.5.3 Mapping reads to the assembled transcriptome:

Bowtie v 1.2 (Langmead et al., 2009) was used for mapping the reads to the transcriptome assembly, used as a reference. As can be seen in Table 5.6, the mapped percentage was around 38.7% to 40.8% in protocorm samples, while there was a lower mapping percent 31.9% to 34% in root samples, and in the corm samples it was 31.1% to 37.2%. Overall, these percentages were higher than obtained previously.

5.3.5.4 Clustering transcripts into clusters:

The number of assembled transcripts was 1.92 million and after clustering using Corset, the transcripts that clustered were around 516,247. This is based on the differential expression levels of contigs within clusters (Table 5.7). In this step, each

cluster has expression values per sample which were used for subsequent differential expression analysis.

Table 5.6: Reads mapping metrics showing the number and percentage of reads mapped to reference sequences.

Sample	Read Pairs	Mapped Pairs	% Mapped
Sample_1-P1	35,419,668	13,712,153	38.7
Sample_2-P2	37,471,631	14,656,770	39.1
Sample_3-P3	37,580,675	15,342,712	40.8
Sample_7-R3	37,478,937	12,425,856	33.6
Sample_8-R5	35,620,159	12,098,814	34.0
Sample_9-R6	40,596,741	12,936,924	31.9
Sample_12-SC3	40,698,695	12,658,542	31.1
Sample_13-SC4	34,074,300	12,683,001	37.2
Sample_14-Corm	42,294,066	14,386,381	34.0

Table 5.7: Results of Corset clustering demonstrated the number of clustered transcripts and clusters.

Assembled transcripts	Clustered transcripts	Clusters
1,916,614	516,247	271,338

5.3.5.5 Differential expression:

G1 represents the protocorm samples, which are grown in a controlled laboratory environment rather than field. In addition, each was produced by pooling several protocorms to obtain sufficient tissue for analysis. Greater uniformity within these protocorm samples is thus reasonable. G2 and G3 represent the root and corm samples respectively, which were grown in the field. For root samples, each was produced by combining 2-3 roots to obtain adequate amount of tissue for analysis. In case of corm, 100 mg from only one corm was used for each sample analysed.

Table 5.8: Information about samples used for differential expression contrasts.

Sample	Sample ID	Group	Reference group	Contrasts
P1	Sample_1-P1	G 1	Group 1	Group 1 vs Group 2
P2	Sample_2-P2			
P3	Sample_3-P3			
R3	Sample_7-R3	G2	Group 2	Group 1 vs Group 3
R5	Sample_8-R5			
R6	Sample_9-R6			
SC3	Sample_12-SC3	G3	Group 3	Group 2 vs Group 3
SC4	Sample_13-SC4			
Corm	Sample_14-Corm			

5.3.5.5.1 Testing for variation in DE:

5.3.5.5.1.1 Pairwise scatterplot:

The pairwise scatterplot in Figure 5.6 indicates that all tissue sources have a rough correlation within and between sample groups. For example, Figure 5.6-G1 shows the three replicates of the protocorms sample which has a strong relationship between them in term of read counts per gene as represented on a log10 scale in each box, while in group two (roots, Figure 5.6-G2) the difference in the number of genes expressed is greater but they also more correlated to each other. In Figure 5.6-G3, more genes are expressed and all are organized in a way that indicates a strong correlation among them. For statistical testing, the mean of groups has been used to compares the correlations and, as can be seen in Figure 5.6-G-means, there are positive relationships between groups.

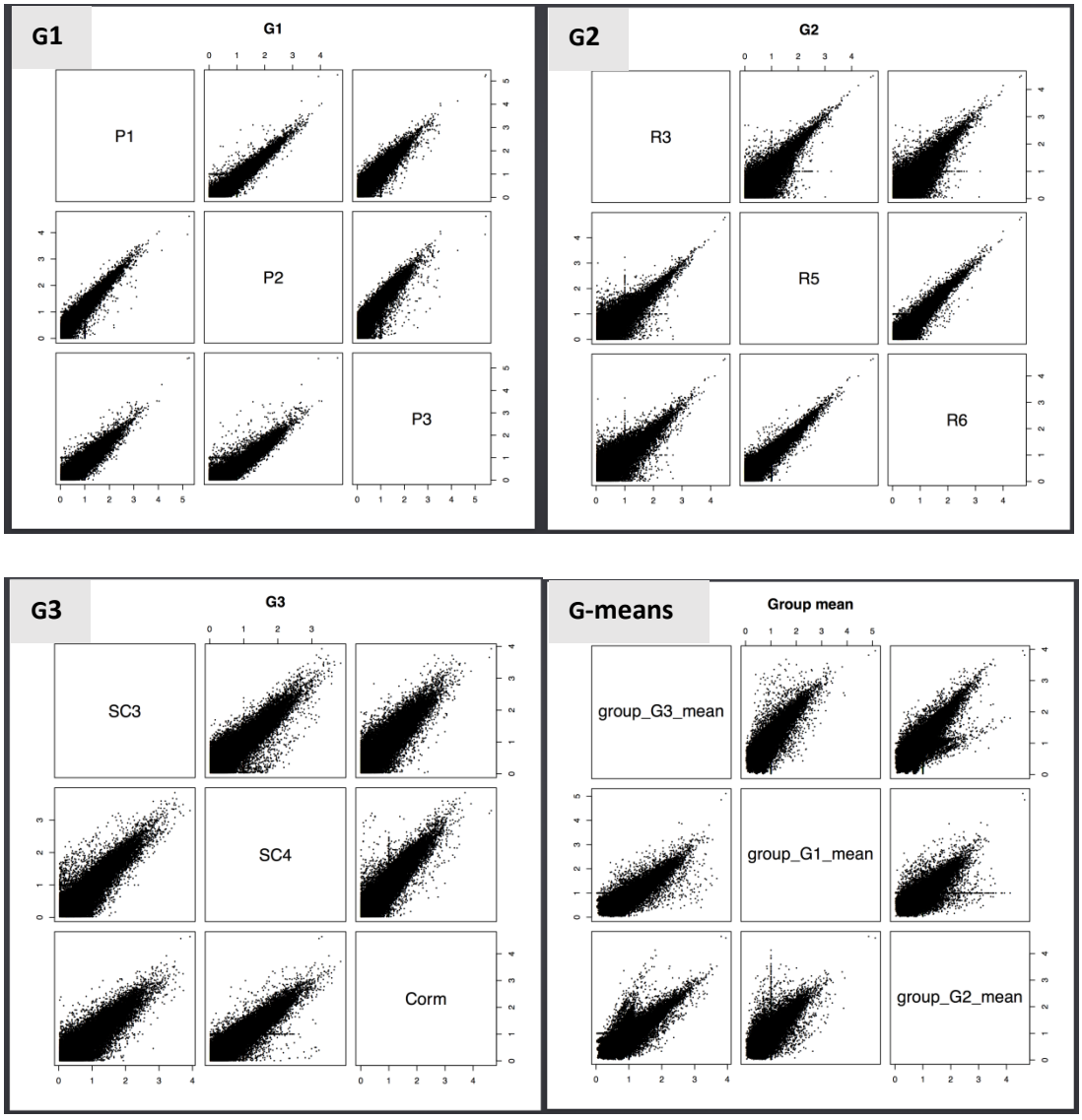


Figure 5. 6 Scatterplots showing the expression of genes for each groups of read counts per gene within a sample group (log10 scale), and between sample groups of read counts of the means to compare the variation. G1: Three samples of protocorms. G2: Three samples of roots. G3: Three samples of corms. G-means: group means of expression values between three groups.

5.3.5.5.1.2 Heatmap correlations of samples:

A heatmap was created from all expression values between sample groups and within a sample group. Figure 5.7 shows the correlation coefficient among all samples. Each block from bottom left to top right represents the correlations of samples within a sample group, and the remaining blocks are the correlation among sample groups. This shows that G1 (protocorm) has a higher correlation within samples compared to G2 and G3. An interesting point is that Sample-Corm in group 3 correlates stronger

to samples from group 1 than to the other two group 3 samples. Group 2 (roots) is the most distinct group.

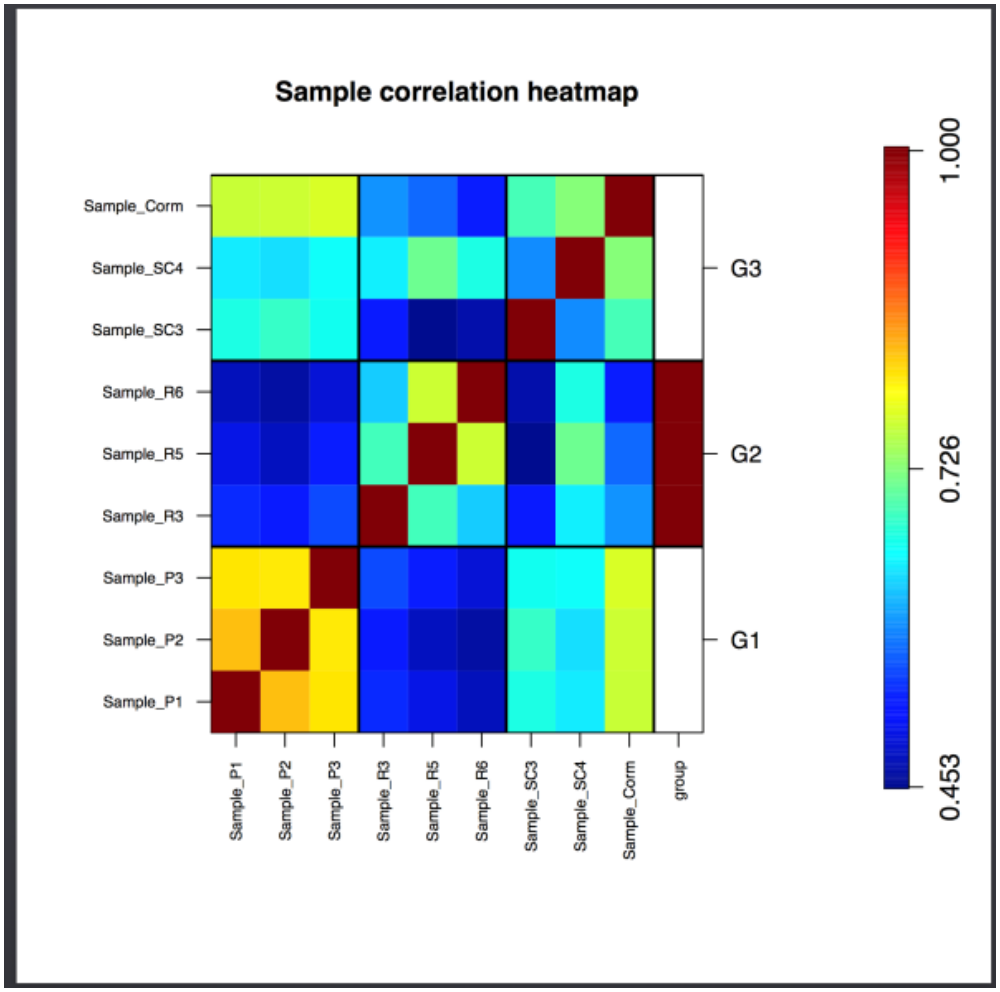


Figure 5.7: The heatmap of correlation coefficient between all samples.

5.3.5.5.1.3 Principle Component Analysis:

Principle components analysis, using log10 of read counts from the expression values of all transcripts, was undertaken. Figure 5.8-A shows the 1st and 2nd component indicating that the first component accounts for 80.9% of the variance in the samples. The three samples in each group are roughly clustered. The 2nd and 3rd components account for 9.1% and 2.7% of the variance respectively, and are shown in Figure 5.8-B. Samples from group 1 (protocorm) and group 3 (corm) are again shown to be quite similar, but with samples of group 2 (root) are shown to be very distinct. This result matches with the results of heatmap correlation (Figure 5.7)

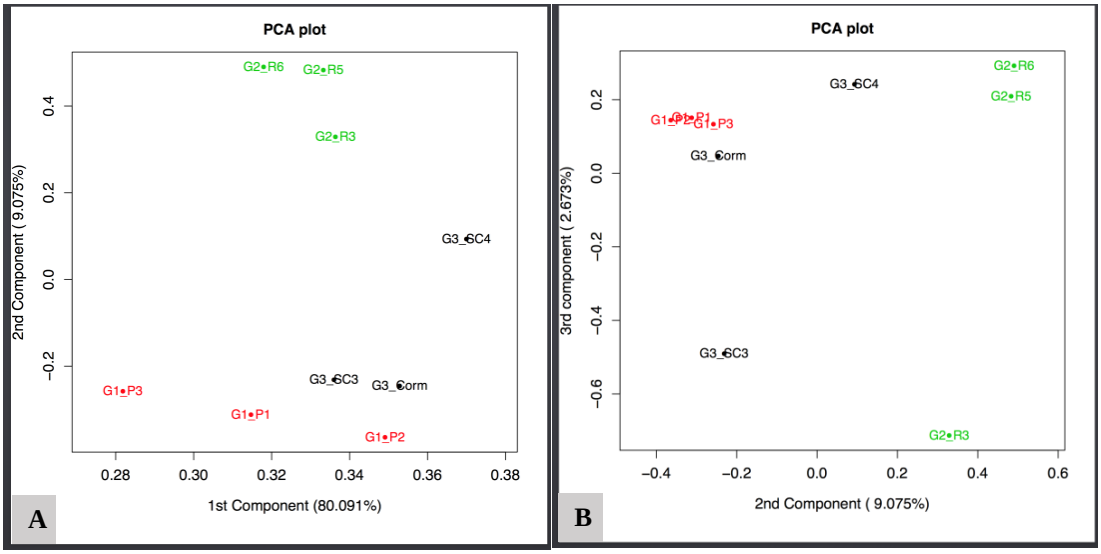


Figure 5.8: PCA plots using log10 of read counts showing the 1st and 2nd, and the 2nd and 3rd principle components from all libraries. The names of all sample are printed in the plot, G1 (protocorm, red), G2 (root, green), G3 (corm, black). A: 1st and 2nd components. B: 2nd and 3rd components.

5.3.5.5.1.4 P-value histogram:

The histogram of p-values for each contrast between the three sample groups (Figure 5.9) shows the frequency of significance or non-significance. Thus, any large bar appearing towards the right of the plot indicates less significance, and it can be attributed to transcripts with read count numbers of mainly zero across all samples. In contrast, large bars appearing towards the left of the plot means more significance and it indicates that DE transcripts exist in the corresponding contrast.

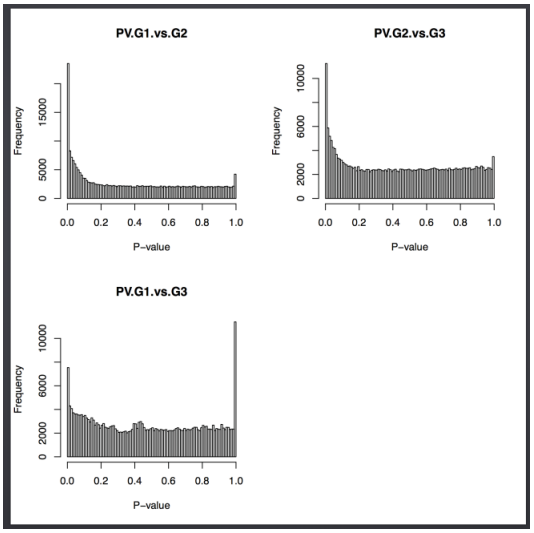


Figure 5. 9: Histogram of p-values from the Negative Binomial distribution test, which indicates that DE transcripts exist in the corresponding contrast. pV= p-values, G1 = protocorm samples, G2 = root samples, and G3 = corm samples.

5.3.5.5.1.5 Testing log fold change (logFC) for contrasts and detecting DE clusters:

There are differentially expressed genes in each of the three samples. Thus, a total of 14,272 genes were found to be differential expressed and regulated in protocorms and among them 2,818 genes were detected significantly up regulated, and 11,454 down regulated. In contrast, the two other groups G2 and G3 (root and corms) have a much lower number of differentially expressed genes, only around 1,630 and 2,241 respectively. Also, the number of DE genes detected as up regulated in root was 728 and down regulated 902, whereas 1,453 were up regulated in corms and 788 down regulated. Overall, protocorms had a higher number of differential expressed clusters than the two other groups as shown in Table 5.9 and Figure 5.10.

Table 5.9: Number of differentially expressed clusters for all tissues.

Name of Tissue	Total number DE	Number uP-DE	Number down-DE
Protocorm	14,272	2,818	11,454
Root	1,630	728	902
Corm	2,241	1,453	788

Plotting log2FC against log2CPM (counts per million mapped reads) were applied for all contrasts (Figure 5.10) shows significantly up and down DE genes (highlighted in red). Moreover, many of these genes were significantly different

between the protocorms and two other tissues (root and corm). Genes that were significantly DE were defined as those with a FDR-adjusted p-value < 5%, and were distinctly differentially expressed between the two conditions. Those genes on the positive log2FC axis were up regulated and down regulated on the negative log2FC axis per group.

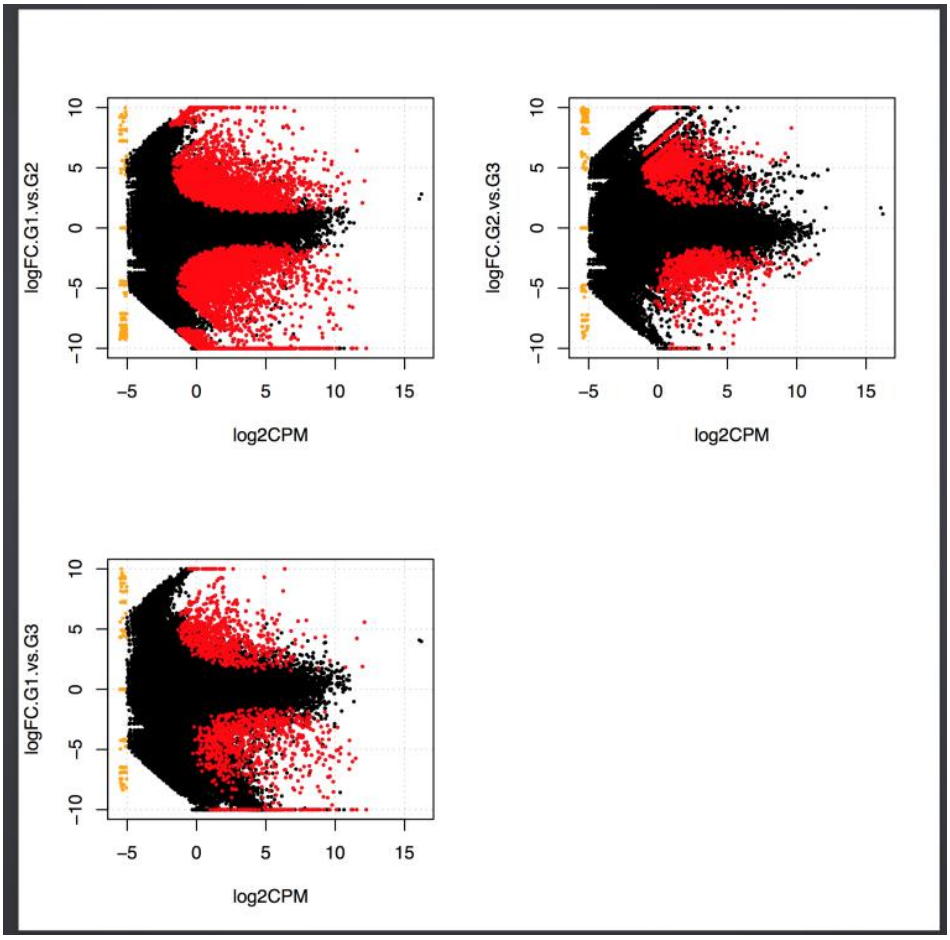


Figure 5. 10: The log2FC against log2CPM plots for all contrasts. Red dots represent significant DE transcripts, and the black dots indicate non-significant DE transcripts. G1 = protocorms; G2 = root; G3 = corm.

5.3.5.5.1.6 Clustering analysis of detected DE clusters:

Figure 5.11 shows a significant difference in DE of 15 clusters between the three groups of tissues as seen in the first column on the left. The yellow colors in the heatmap indicated up regulated and blue colour down regulated. It can be seen that 19% of G1 vs G2 are up regulated and 80% down regulated, while 64% of those from G2 vs G3 up regulated and 35% down regulated, and with G1 vs G3 about 44% are up and 55% down regulated.

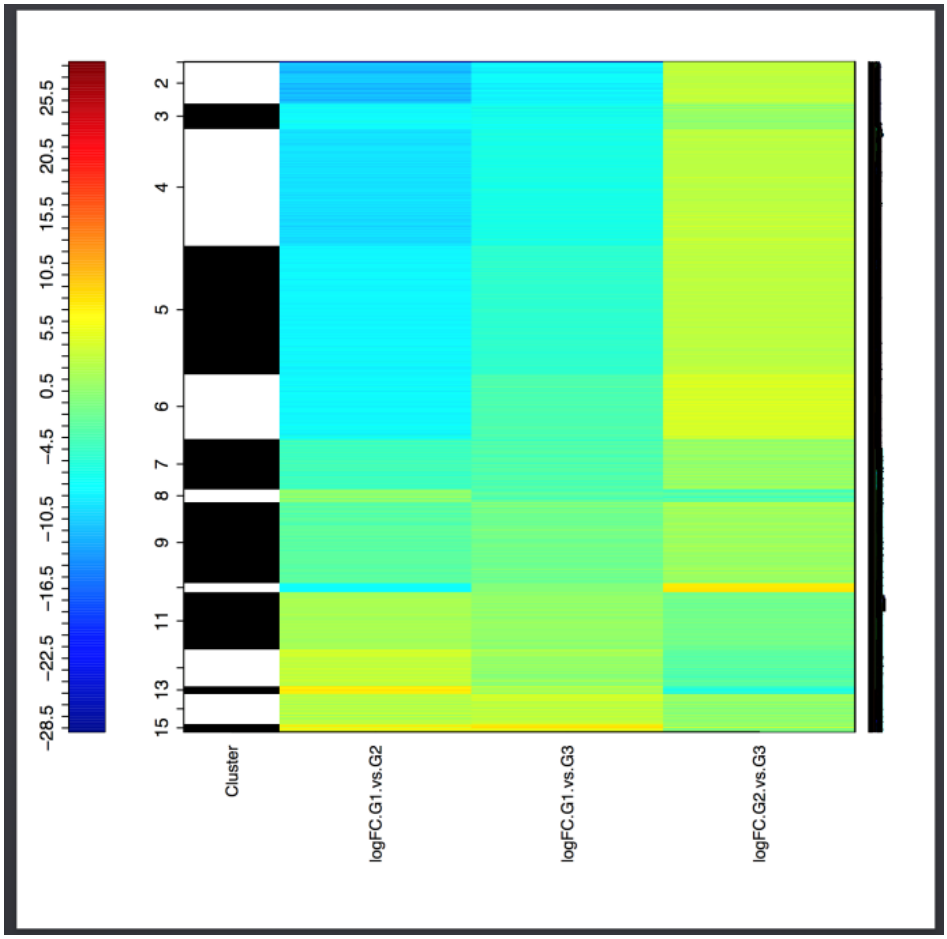


Figure 5. 11: Heatmap showing the expression patterns of the transcripts that formed 15 clusters. Individual cells represent up-regulation (yellow) or down-regulation (blue) of the clusters. logFC G1.vs G2, protocorm vs root samples; logFC G1.vs G3, protocorm vs corm samples; and logFC G2.vs G3, root vs corm samples.

5.3.5.6 Functional annotation

5.3.5.6.1 Genes/ transcripts of interest

The top ten clusters according to total FPKM are listed in Table 5.10 to 5.12 and these are among the top 100 significant differentially expressed genes. These were identified by comparing three groups G1 vs G2 (protocorm group), G1 vs G3 (root group), and G2 vs G3 (corm group). The level of gene expression was determined based on the total value of FPKM and FDR ≤ 0.001 to judge significance of gene expression. For example, a total of 100713 genes DE in group 1 (99634 up regulated / 1079 down regulated) and so on (Full data set for 100 significant gene expression in Appendix 5.1). Furthermore, the number of down regulated exceeded the number of up-regulated in group 1 versus group 2 (82%), while in group 3 versus group 2 there were 77% down-regulated than up-regulated.

Table 5. 10: Top ten sorted genes based on total FPKM value in G1 vs G2.

G1 vs G2					
Gene ID	**Total FPKM**	**P-value**	**FDR-value**	**Protocorm**	**Root**
TRINITY_DN165695_c3_g1	100713	2.00E-27	9.01E-23	down	up
TRINITY_DN151896_c5_g3	96465	6.09891E-14	1.59871E-10	down	up
TRINITY_DN150624_c2_g3	89899	2.47E-16	1.13E-12	down	up
TRINITY_DN151896_c5_g3	45105	6.32E-29	5.69E-24	down	up
TRINITY_DN157662_c0_g4	41607	1.18E-28	6.52E-24	down	up
TRINITY_DN133304_c8_g7	38085	9.52E-14	2.34E-10	down	up
TRINITY_DN145321_c1_g1	31601	8.13E-15	2.68E-11	down	up
TRINITY_DN151896_c5_g3	22139	1.60E-34	4.33E-29	down	up
TRINITY_DN151896_c5_g3	20176	9.82E-16	3.84E-12	down	up
TRINITY_DN133620_c0_g2	10291	3.02E-17	1.85E-13	up	down

Table 5. 11: Top ten sorted genes based on total FPKM value in G1 vs G3.

G1 vs G3					
Gene ID	**Total FPKM**	**P-value**	**FDR-value**	**Protocorm**	**Corm**
TRINITY_DN151896_c5_g3	138984	2.78E-15	1.93E-11	down	up
TRINITY_DN165695_c3_g1	133221	1.83E-28	9.88E-24	down	up
TRINITY_DN151896_c5_g3	60676	7.07E-30	4.77E-25	down	up
TRINITY_DN168031_c5_g4	54795	2.00E-20	3.01E-16	down	up
TRINITY_DN165695_c3_g1	47604	3.03E-24	6.81E-20	down	up
TRINITY_DN133304_c8_g7	46344	6.07E-13	2.52E-09	down	up
TRINITY_DN151896_c5_g3	30906	4.49E-36	1.21E-30	down	up
TRINITY_DN151896_c5_g3	30417	1.57E-17	1.46E-13	down	up
TRINITY_DN145321_c1_g1	29712	3.55E-12	1.27E-08	down	up
TRINITY_DN165695_c3_g1	19401	2.36E-31	3.18E-26	down	up

Table 5. 12: Top ten sorted genes based on total FPKM value in G2 vs G3.

G2 vs G3					
Gene ID	**Total FPKM**	**P-value**	**FDR-value**	**Root**	**Corm**
TRINITY_DN146889_c2_g4	4432	1.00E-14	1.36E-09	down	up
TRINITY_DN155776_c6_g1	516	1.03E-12	4.62E-08	down	up
TRINITY_DN154464_c0_g2	2237	2.67E-11	6.01E-07	down	up
TRINITY_DN154464_c0_g2	1724	1.11E-09	1.19E-05	down	up
TRINITY_DN145590_c2_g6	218	4.84E-06	0.004	down	up
TRINITY_DN142295_c6_g3	564	0.0004	0.05	down	up
TRINITY_DN133620_c0_g2	5588	4.67E-07	0.001	down	up
TRINITY_DN161687_c0_g3	370	1.24E-06	0.002	down	up
TRINITY_DN158184_c4_g2	1496	1.83E-13	1.65E-08	down	up
TRINITY_DN146889_c2_g4	4432	1.00E-14	1.36E-09	down	up

The top ten among top 100 significant genes sorted based on the log2FC ≥ 1 are shown in Tables 5.13 and 5.14. Genes where log2FC is greater than 1 show significant differences. Therefore, G2 vs G3 is not listed here as it is not significant at log2FC <0.

Table 5.13: Top ten sorted genes based on log2FC value in G1 vs G2.

G1 vs G2				
Gene ID	**Log2Fold Change**	**P-value**	**Protocorm**	**Root**
TRINITY_DN154464_c0_g2	9.727883138	3.10E-21	up	down
TRINITY_DN154464_c0_g2	8.504160547	3.42E-19	up	down
TRINITY_DN146889_c2_g4	7.829420116	3.04E-22	up	down
TRINITY_DN145238_c3_g2	6.994905451	9.09E-13	up	down
TRINITY_DN142295_c6_g3	6.533672166	2.46E-17	up	down
TRINITY_DN161687_c0_g3	5.642921239	9.16E-17	up	down
TRINITY_DN155776_c6_g1	5.394135114	3.94E-22	up	down
TRINITY_DN145590_c2_g6	4.619088398	2.71E-17	up	down
TRINITY_DN133620_c0_g2	3.461436153	3.02E-17	up	down
TRINITY_DN158184_c4_g2	3.355553855	2.00E-12	up	down

Table 5. 14: Top ten sorted genes based on log2FC value in G1 vs G3.

G1 vs G3				
Gene ID	**Log2Fold Change**	**P-value**	**Protocorm**	**Corm**
TRINITY_DN145238_c3_g2	5.592602093	5.24E-13	up	down
TRINITY_DN142295_c6_g3	3.789173938	4.80E-07	up	down
TRINITY_DN154464_c0_g2	2.872894868	0.003022955	up	down
TRINITY_DN154464_c0_g2	2.714821274	0.002838451	up	down
TRINITY_DN161687_c0_g3	2.337354197	0.000197248	up	down
TRINITY_DN145590_c2_g6	2.091748179	1.32E-05	up	down
TRINITY_DN146889_c2_g4	1.581209821	0.041930745	up	down
TRINITY_DN155776_c6_g1	1.407205088	0.003448288	up	down
TRINITY_DN133620_c0_g2	1.394709446	0.000620097	up	down

5.3.5.6.2 Protein Domains based on Pfam

Among the total of 18,143 differentially expressed clusters, there were matches of 44,045, 47,811, and 29,178 clusters to data in Pfam, Go Ontology Blast, and Go Ontology Pfam database, respectively. A selection of these proteins have been listed in Table 5.15 which are upregulated in corm and root, while the Table 5.16 shows proteins that up regulated in protocorm with p-value $\leq$ 0.05. The Hsp20 (Hsp20/alpha crystallin family), AAA (ATPase family associated with various cellular activities), and Clp-N (Clp amino terminal domain, pathogenicity island component) were the most highly represented proteins in the root and corm samples. Retrotran_gag_2 (gag-polypeptide of LTR copia-type) was highly represented in the protocorm and corm. Many of these proteins have binding (e.g. NAD, fatty acid and protein) or transporter (e.g. lipid, ion, *O*-methyl and phosphoribosyl transferase) functions, and other proteins that are implicated in the stress response including Hsp20 and a protease inhibitor/seed storage protein (Tryp_alpha_amyl). Proteins that were found up-regulated in protocorm and corm may play important roles in the growth of plant through domains such as Pkinase, Pkinase_Try, and ATP binding. Interestingly, genes related to photosynthesis were down regulated in protocorm tissue and up regulated in the root and corm tissues.

Table 5.15: Summary of Pfam domains for genes significantly up-regulated in root and corm.

Accession	ID	Description	E-value
PF04969.15	CS	CS domain	E:0.0027
PF00011.20	HSP20	Hsp20/alpha crystallin family	E:3.4e-28
PF07724.13	AAA_2	AAA domain (Cdc48 subfamily)	E:2.4e-15
PF00004.28	AAA	ATPase family associated with various cellular activities (AAA)	E:0.00014
PF02861.19	Clp_N	Clp amino terminal domain, pathogenicity island component	E:5.1e-06
PF00158.25	Sigma54_activat	Sigma-54 interaction domain	E:4.5e-06
PF07728.13	AAA_5	AAA domain (dynein-related subfamily)	E:3.2e-11
F10431.8	ClpB_D2-small	C-terminal, D2-small domain, of ClpB protein	E:3.9e-20
PF08100.10	Dimerisation	Dimerisation domain	E:1.6e-13
PF00891.17	Methyltransf_2	O-methyltransferase	E:1.2e-05
PF14223.5	Retrotran_gag_2	gag-polypeptide of LTR copia-type	E:7.2e-25
PF16561.4	AMPK1_CBM	Glycogen recognition site of AMP-activated protein kinase	E:5.2e-18
PF13450.5	NAD_binding_8	NAD(P)-binding Rossmann-like domain	E:5.6e-11
PF01593.23	Amino_oxidase	Flavin containing amine oxidoreductase	E:3.6e-74
PF00078.26	RVT_1	Reverse transcriptase (RNA-dependent DNA polymerase)	E:1.2e-26
PF02866.17	Ldh_1_C	lactate/malate dehydrogenase, alpha/beta C-terminal domain	E:3.5e-41
PF01554.17	MatE	MatE	E:5.4e-25
PF03098.14	An_peroxidase	Animal haem peroxidase	E:2.3e-10
PF00067.21	p450	Cytochrome P450	E:2.8e-15

Table 5. 16: Summary of Pfam domains for genes significantly up-regulated in protocorm.

Accession	ID	Description	E-value
PF14368.5	LTP_2	Probable lipid transfer	E:9.1e-11
PF00234.21	Tryp_alpha_amyl	Protease inhibitor/seed storage/LTP family	E:3e-05
PF14223.5	Retrotran_gag_2	gag-polypeptide of LTR copia-type	E:7.2e-25
PF08212.11	Lipocalin_2	Lipocalin-like domain	E:2.5e-37
PF00061.22	Lipocalin	cytosolic fatty-acid binding protein family	E:0.00018
PF00520.30	Ion_trans	Ion transport protein	E:2.7e-10
PF05703.10	Auxin_canalis	Auxin canalisation	E:4e-79
PF00504.20	Chloroa_b-bind	Chlorophyll A-B binding protein	E:2.5e-46
PF00156.26	Pribosyltran	Phosphoribosyl transferase domain	E:9.7e-17

5.3.5.6.3 Functional annotation and Gene Ontology classification:

The Gene Ontology assignment of 100 transcripts significantly differentially expressed, based on the Pfam ontology, is shown in Figure 5.12. The GO classification involved three main categories: biological process, cellular component and molecular function. The largest number of transcripts was annotated by molecular function such as ATP binding and protein binding, followed by protein kinase activity and catalytic activity, then cellular component, namely photosystem 1 and integral component of membrane, and then photosynthesis in biological process.

The GO for all other clusters, based on their Blast annotation is shown in Tables 5.17 to 5.22. Many clusters in molecular function categories were assigned to 'binding' and up-regulated in all tissues such as for ATP, RNA, lipid, protein, metal ion, zinc, calmodulin, cAMP, cGMP, iron ion, and DNA (Table 5.21 and 5.22). For biological process, the most numerous categories were 'response' (e.g. cold, heat, high light intensity, auxin, water deprivation, paraquat, and freezing) and 'cellular process' (Table 5.17, 5.18 and Figure 5.13, 5.14). Within cellular component was 'cell membrane' e.g. plasma and chloroplast membranes and 'organelle' such as mitochondrion and nucleus were most abundant in corm and root (Table 5.19). Some other gene functions in Table 5.17 that are up regulated in protocorm included, in terms of biological process, the term hyperosmotic salinity response (GO:0042538; to any change in the concentration of salts, sodium, and chloride ions) and seed maturation (GO:0010431; responsible for seed development after embryogenesis).

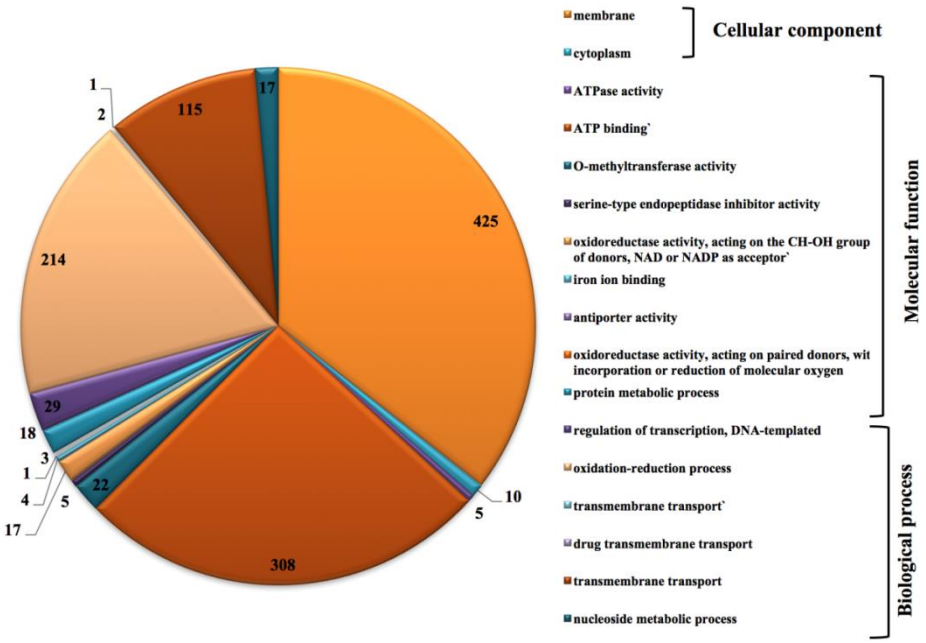


Figure 5.12: GO functional classification of top 100 DE genes for three tissues (protocorm, corm and root) based on Pfam. It represents the three categories biological process, cellular component, and molecular function. The number in each sector indicates frequency of each GO assignment.

The GO terms in biological process that were up-regulated in corm and root and most abundant were “metabolic process” such as protein and carbohydrate, followed by “response” (e.g. heat, copper, toxic substance, and hydrogen peroxide) as shown in Table 5.17 and Figure 5.13.

Table 5. 17: GO assignment in terms of biological process of Blast ontology for significantly DE (p-value ≤ 0.05) up-regulated in protocorm. Frequency more than one.

ID	Description	Frequent
GO:0006869	lipid transport	0
GO:0009627	systemic acquired resistance	0
GO:0009862	systemic acquired resistance, salicylic acid mediated signaling pathway	0
GO:0015074	DNA integration	137
GO:0030644	cellular chloride ion homeostasis	0
GO:0006883	cellular sodium ion homeostasis	0
GO:0010286	heat acclimation	0
GO:0042538	hyperosmotic salinity response	0
GO:0006629	lipid metabolic process	150
GO:1902884	positive regulation of response to oxidative stress	0
GO:1901002	positive regulation of response to salt stress	0
GO:0009409	response to cold	0
GO:0009735	response to cytokinin	0
GO:0050826	response to freezing	0
GO:0009408	response to heat	0
GO:0009644	response to high light intensity	0
GO:0009416	response to light stimulus	0
GO:1901562	response to paraquat	0
GO:0009414	response to water deprivation	0
GO:0010431	seed maturation	0
GO:0042391	regulation of membrane potential	0
GO:0009734	auxin-activated signaling pathway	0
GO:0010305	leaf vascular tissue pattern formation	0
GO:0010087	phloem or xylem histogenesis	0
GO:0009733	response to auxin	14
GO:0018298	protein-chromophore linkage	0
GO:0006168	adenine salvage	0
GO:0044209	AMP salvage	0
GO:0006166	purine ribonucleoside salvage	0

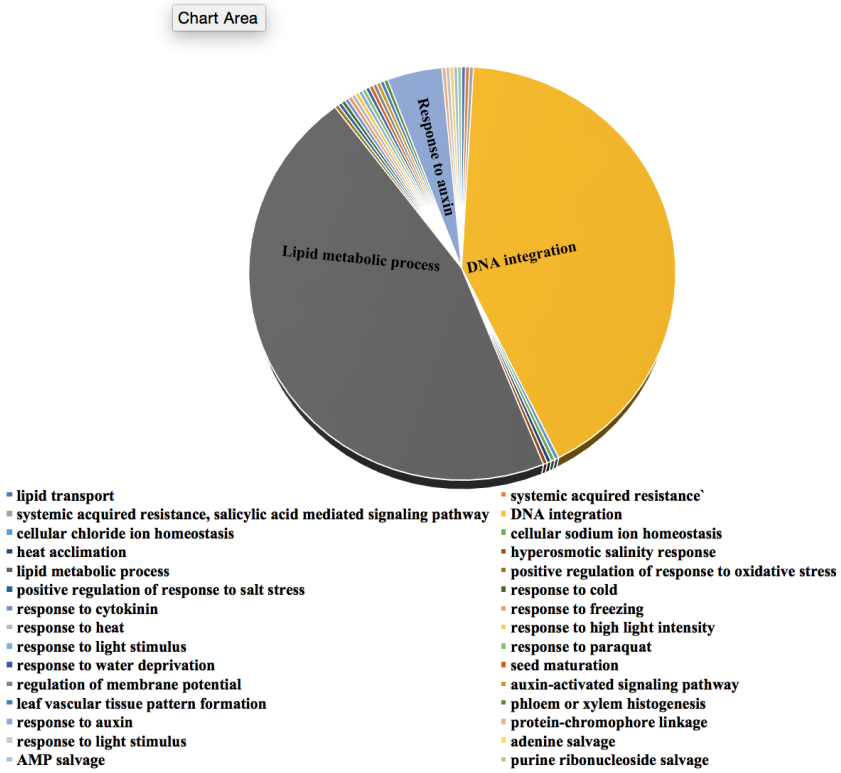


Figure 5.13: GO functional classification of top 100 DE genes up-regulated in protocorm, based on Blast ontology. It represents the biological process according to frequency of each GO assignment.

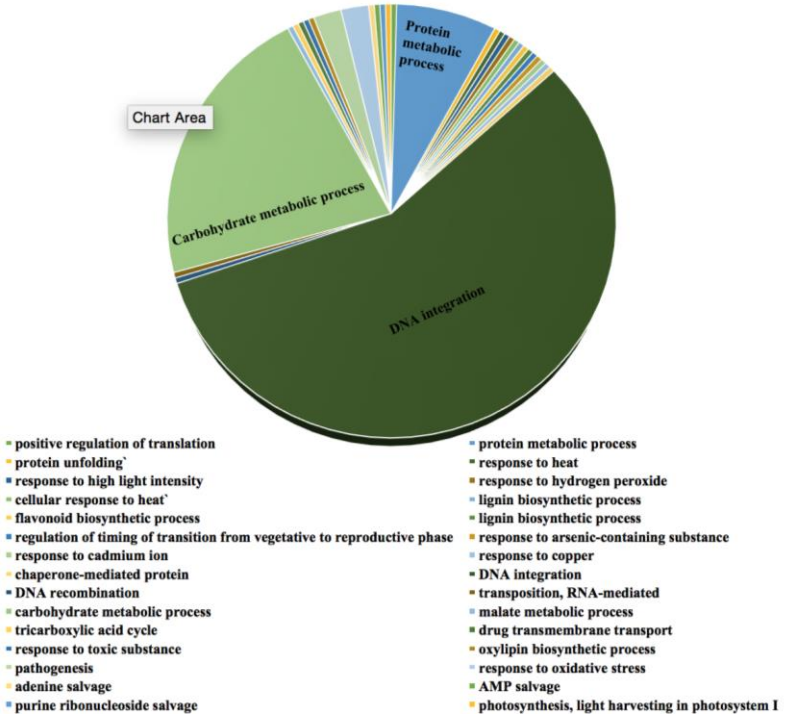


Figure 5.14: GO functional classification of top 100 DE genes up-regulated in corm and root, based on Blast ontology. It represents the biological process according to frequency of each GO assignment.

Table 5.18: GO assignment in terms of biological process of Blast ontology for significantly DE (p-value ≤ 0.05) up-regulated in corm and root. Frequency more than one.

ID	Description	Frequent
GO:0045727	positive regulation of translation	0
GO:0019538	protein metabolic process	18
GO:0043335	protein unfolding	0
GO:0009408	response to heat	0
GO:0009644	response to high light intensity	0
GO:0042542	response to hydrogen peroxide	0
GO:0034605	cellular response to heat	0
GO:0009809	lignin biosynthetic process	0
GO:0009813	flavonoid biosynthetic process	0
GO:0048510	regulation of timing of transition from vegetative to reproductive phase	0
GO:0046685	response to arsenic-containing substance	0
GO:0046686	response to cadmium ion	0
GO:0046688	response to copper	0
GO:0061077	chaperone-mediated protein	0
GO:0016032	viral process	0
GO:0006310	DNA recombination	0
GO:0032197	transposition, RNA-mediated	0
GO:0005975	carbohydrate metabolic process	52
GO:0006108	malate metabolic process	0
GO:0006099	tricarboxylic acid cycle	0
GO:0006855	drug transmembrane transport	0
GO:0009636	response to toxic substance	0
GO:0031408	oxylipin biosynthetic process	0
GO:0009405	pathogenesis	5
GO:0006979	response to oxidative stress	5
GO:0009768	photosynthesis, light harvesting in photosystem I	0

Table 5. 19: GO assignment in terms of cellular component of Blast ontology for significantly DE (p-value ≤ 0.05) up-regulated in corm and root. Frequency more than one.

Go ontology	Cellular component	Frequent
GO:0005737	cytoplasm	10
GO:0009941	chloroplast envelope	1
GO:0009570	chloroplast stroma	1
GO:0005829	cytosol	1
GO:0005634	nucleus	65
GO:0009507	chloroplast	1
GO:0000943	retrotransposon nucleocapsid	1
GO:0005759	mitochondrial matrix	12
GO:0005739	mitochondrion	5

Table 5. 20: GO assignment in terms of cellular component of Blast ontology for significantly DE (p-value ≤ 0.05) up-regulated in protocorm. Frequency more than one.

Go ontology	Cellular component	
GO:0048046	apoplast`	1
GO:0005783	endoplasmic reticulum	6
GO:0009506	plasmodesma	1
GO:0016021	integral component of membrane	333
GO:0009941	chloroplast envelope	1
GO:0031969	chloroplast membrane`	0
GO:0005737	cytoplasm	10
GO:0009898	cytoplasmic side of plasma	0
GO:0005794	Golgi apparatus	0
GO:0005739	mitochondrion	5
GO:0005886	plasma membrane	38
GO:0009506	plasmodesma	0
GO:0005774	vacuolar membrane	0
GO:0005773	vacuole	0
GO:0005887	integral component of plasma membrane	0
GO:0009522	photosystem I	0
GO:0009523	photosystem II	0
GO:0010287	plastoglobule	0

Table 5. 21: GO assignment in terms of molecular function of Blast ontology for significantly DE (p-value ≤ 0.05) up-regulated in root and corm. Frequency more than one.

Go ontology	Molecular function	Frequent
GO:0005524	ATP binding`	308
GO:0047763	caffeate O-methyltransferase activity	0
GO:0046983	protein dimerization activity	163
GO:0008171	O-methyltransferase activity	22
GO:1990534	thermospermine oxidase activity	0
GO:0051082	unfolded protein binding	7
GO:0004190	aspartic-type endopeptidase activity	27
GO:0003677	DNA binding	193
GO:0003887	DNA-directed DNA polymerase activity	0
GO:0008233	peptidase activity	12
GO:0004540	ribonuclease activity	3
GO:0003723	RNA binding	175
GO:0004523	NA-DNA hybrid ribonuclease activity	0
GO:0030060	L-malate dehydrogenase activity	0
GO:0004252	serine-type endopeptidase activity	1
GO:0015297	antiporter activity	0
GO:0015238	drug transmembrane transporter activity	51
GO:0016853	isomerase activity	2
GO:0052878	linoleate 8R-lipoxygenase activity	0
GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	0
GO:0004601	peroxidase activity	73
GO:0020037	heme binding	13
GO:0005506	iron ion binding	4
GO:0016710	trans-cinnamate 4-monooxygenase activity	0
GO:0003999	adenine phosphoribosyltransferase activity	0

Table 5.22: GO assignment in terms of molecular function of Blast ontology for significantly DE (p-value ≤ 0.05) up-regulated in protocorm. Frequency more than one.

Go ontology	Molecular function	Frequent
GO:0005504	fatty acid binding	0
GO:0043621	protein self-association	0
GO:0008270	zinc ion binding	209
GO:0008289	ipid binding	37
GO:0004190	aspartic-type endopeptidase activity	27
GO:0004519	endonuclease activity	1
GO:0003676	nucleic acid binding	412
GO:0003964	RNA-directed DNA polymerase activity	10
GO:0020037	heme binding	13
GO:0005506	iron ion binding	4
GO:0004497	monooxygenase activity	0
GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	0
GO:0045735	nutrient reservoir activity	26
GO:0005215	transporter activity	1
GO:0005516	calmodulin binding	15
GO:0030552	cAMP binding	0
GO:0030553	cGMP binding	0
GO:0005249	voltage-gated potassium channel activity	0
GO:0016168	chlorophyll binding	11
GO:0046872	metal ion binding	234
GO:0031409	pigment binding	0
GO:0003999	adenine phosphoribosyltransferase activity	0

5.4 Discussion:

5.4.1 RNA isolation:

Obtaining RNA of adequate quality and integrity is challenging for many researchers in plant molecular biology, because of contamination with other components such as polysaccharides, phenols, and other secondary metabolites. The use of a rapid CTAB method yielded very low quantity (14-18 ng/µl) and quality. TRIzol® Reagent protocol was successful for extraction of RNA from root tissues of good quantity (~1000-2000 ng/µl) and quality, but was not successful with protocorms and corms, probably indicating a high level of contamination due to starch, protein, phenols and other components. The innuPREP Plant RNA Kit provided a satisfactory amount of RNA from protocorm and corm even though these were considered a difficult tissue from which to extract RNA.

5.4.2 RNA quality and integrity:

After using the TRIzol® Reagent protocol and innuPREP Plant RNA Kit in root, corm and protocorm respectively, followed by DNase treatment, to prepare the RNA samples, they were sent for sequencing. The CGR measured the integrity of the RNA using an Agilent Bioanalyser to determine several RNA quality metrics (Ward *et al.*, 2012). A RIN value of more than 8 indicates high quality, but with 7 also considered sufficient (Ward *et al.,* 2012; Johnson *et al.,* 2012). The ratio of 28S:18S must also be considered, and although this was below 2.0 for some samples, they were taken forward and ultimately yielded sequence data.

5.4.3 rRNA depletion and library preparation:

The result of the rRNA depletion using the Ribo-Zero rRNA removal kit in most samples under study was successful with an absence of 28S and 18S peaks and a ratio 28S/18S of 0.0 (Benes et al., 2011) although the ratios in two samples (Sample_1-P1 and Sample_14-corm) were higher (12.4 and 1.3 respectively). However, these two samples with inadequate rRNA depletion were processed the same as the other samples and yielded adequate sequence data. These results therefore indicate that measurements of RNA quality do not always predict sequencing success.

5.4.4 RNA-seq data analysis

5.4.4.1 Sequence mapping

Initially, the *Ophrys apifera* transcriptome sequences were mapped to the genome data of *Arabidopsis thaliana* and to the orchid *Phalaenopsis equetris* using CyVerse, but the result was poor, 0.6-0.8% and 0.9-1.4% respectively. Also, it was mapped with an *Ophrys* transcriptome assembly (Sedeek et al., 2013), but again the rate of alignment was only between 17%-34% (section 5.3.7.1). Therefore, *de novo* transcriptome assembly using Trinity was performed for reconstructing the transcripts in the samples because of the absence of reference. Trinity was designed to resolved many problems in RNA-seq such as sequencing errors, determine splice isoforms, and gene variants (Grabherr et al., 2011b).

Application of the Trinity method has been reported in the orchid *Cymbidium hybridum* transcriptome as a successful way to identify genes related to plant-fungal interaction (Zhao et al., 2014). In the present study the sequence alignment for transcripts of protocorm, root and corm showed mapping coverage ranging between 31% and 40.8% of total input sequences. A rate of 70% is considered as an accurate differential expression analysis of mapping. The lower values obtained here may be due to the presence of contamination in the samples, or poor read quality (Trapnell et al., 2012).

N50 statistics (1,687 bp) was used to assess the quality of the assembly, with a higher value of N50 indicating better quality assembly (Liu et al., 2013). In comparison with other plant transcriptomes, the N50 value for the transcriptome of *Gastrodia elata* was 1592 (Tsai et al., 2016).

5.4.4.2 Differential gene expression:

This study demonstrated that Illumina platforms have a power for analysis of gene expression by generating a high number of reads (Wang et al., 2009). A total of three combinations were included here (“protocorm vs. root”, “protocorm vs. corm”, and “root vs. corm”). A transcript can be said to be more highly expressed in any tissue when comparing with other tissues, only if the level of expression met the following criteria: firstly, the average of expression level of biological replicates of one tissue was five times or higher than the average of expression level of biological replicates

of the second tissue. Secondly, the p-value must be less than 0.05 (Meng et al., 2016). Finally, the FPKM should be higher than 3 and log2FC ≥ 1(Tsai et al., 2016).

Briefly, among the total 271,338 DE clusters, 18,143 clusters have significant differential expression. A total of 14,272, 1,630, and 2,241 were differentially expressed genes or transcripts from protocorm, root, and corm respectively. Thus, clustering analysis using k-means method was performed to detected DE clusters and it revealed 2,818 transcripts /genes were up-regulated in protocorm, 728 were up-regulated in root, and 1,453 were up-regulated in corm.

Focusing on the top 100 significantly differentially expressed transcripts /genes were detected with FDR-adjusted p-value ≤ 0.05. The transcriptome comparison between the three groups by following the same criteria revealed that transcripts/ genes within G1 vs G2 were up-regulated in 29 clusters out of 100, and 27 clusters were up-regulated in G1 vs G3, while G2 vs G3 shown 33 clusters up-regulated. Interestingly, a total of 1,453 genes were identified in corm to be expressed at a higher level in comparison with the two other tissues.

We found a higher values for the number of genes that were highly expressed in root (728 transcripts/ genes) than the number reported from root of *Dendrobium officinale* (256 transcripts / genes) reported by Meng et al. (2016). In the orchid *Gastrodia elata* it has been found that corm tissue had a higher number of genes up-regulated (405 transcripts/ genes) than juvenile tubers (298 transcripts / genes up-regulated) (Tsai et al., 2016).

5.4.4.3 Functional annotation of differentially expressed genes:

After using Trinity to identify transcripts that belong to differentially expressed clusters, Trinotate was used for functional annotation including NCBI-BLAST, protein domain identification (HMMR / PFAM), and comparison annotation databases such as (EMBL Uniprot eggNOG / GO pathways). *Apostasia shenzhenica* (isolate ASH160606) was added as additional custom annotation (Zhang et al 2017). Thus, a total of 88,467, 45,165, 44,045, 47,811 and 29,178 transcripts were matched against BlastX, BLASTP, Pfam, Gene Ontology Blast, and Gene ontology Pfam respectively. The information about output of annotation can obtained from (https://trinotate.github.io/#OutputReport). The number of transcripts that matched to

a_shenzhenica_BLASTX was 78737 and 39418 a_shenzhenica_BLASTP matching. Many transcripts did not match to the databases; for example, from 100 significant clusters only 35 were matched to a database. This is a common feature as a significant number of non-annotated clusters / transcripts exist in databases of even well-studied plants such as rice and *Arabidopsis.* Unknown contigs may be related to fragmented RNA (e.g. untranslated regions, introns, long non-coding RNAs). Some types of transcripts including contaminated transcripts are from unidentified sources (Su et al., 2011). Moreover, some may be considered as novel transcripts, or are too short to match known sequences, as has been pointed by others (Liang et al., 2008, Mittapalli et al., 2010, Wang et al., 2004).

However, in this particular study, mycorrhizal fungi may be an additional source of transcripts in both tissues from the field (root and corm). These should be absent from the protocorm (from aseptic condition). Further bioinformatics analysis could be undertaken to identify whether such transcripts are indeed present. Some have been located in other systems. Zhao et al (2014) identified putative fungal genes in symbiotic orchid roots of *Cymbidium hybridum* with functions related to plant cell wall degradation, changing the fungal cell wall and nutrient transport in addition to ribosomal proteins and enzymes. The symbiotic root system often contains fungal structures in different developmental stages (contact, penetrating, hyphae, intracellular hyphae, active and degenerating pelotons), and thus genes will be at different levels of expression. It has also been shown in *Spiranthes sinensis* that in symbiotic protocorms, the degeneration of hyphae played a significant role in the transfer of elements to the plant cells, but fungal carbon and nitrogen also transferred from live hyphae to plant cell through interface between the symbionts (Kuga et al., 2014). The situation with regard to fungal gene expression will therefore be complex. Nevertheless, a total of 163 fungal genes could be characterized through combining the annotation information of *de novo* assembled transcripts with the comparative expression level in different symbiotic roots (Zhao et al., 2014).

5.4.4.3.1 Protein domain:

Among 100 clusters significantly expressed, 31 of them were matched in Pfam domain database. Some proteins that found up-regulated in protocorm such as LTPs and Tryp_alpha_amyl are from the LTP family. This contains trypsin-alpha amylase

inhibitors, seed storage proteins and lipid transfer proteins from plants, and play a role as defense against pathogen, as well as adaptation under environmental changes, and growth of mitochondria (Rico et al., 1996).

The phosphoribosyl transferase domain (Chloroa_b-bind) characterises a family of diverse phosphoribosyl transferase enzymes such as Uracil phosphoribosyl transferase, Adenine phosphoribosyl transferase, and Ribose-phosphate pyrophosphokinase. More generally, these proteins, as been found in Arabidopsis, contain an N-terminal domain known as the P-Loop NTPase domain (Islam et al., 2007). Under changing light condition, this protein plays a crucial role in reversible phosphorylation of light harvesting chlorophyll a / b binding protein to act as a system for balancing energy between photosystem 1 and ll (Liu and Shen, 2004).

It is clear that most proteins that have been detected from protocorm are related to transfer materials, defense, adaptation to environmental changes, and storage in seed as these functions help protocorm maintenance and development.

Proteins up-regulated in root and corm included proteins of An-peroxidase, RNA-binding and heme-binding which are involved in response to stress and active defense. These proteins have been also found in *Phalaenopsis aphrodite* (Su et al., 2011), along with other proteins such as glycine-rich domain and glutathione transferase as highly expressed in all vegetative and reproductive tissues of *Phalaenopsis aphrodite*. Protein AMPK1_CBM (Glycogen recognition site of AMP-activated protein kinase) was up-regulated only in root and this protein with a surface carbohydrate-binding pocket is reasonably involved in carbohydrate metabolism (Polekhina et al., 2005). Cytochrome P450 proteins were also identified. These are a very large family of plant proteins and play roles in the oxidation steps in secondary metabolism and also catalyse degradation of environmental toxins (Coon, 2005, Morant et al., 2003).

5.4.4.3.2 GO ontology:

GO ontology was used for functional categorization of the annotated transcripts. The number of terms does not correspond to the number of annotated transcripts because some transcripts are mapped to few GO terms while others transcript may have multiple GO terms (Wang et al., 2013b). This GO annotation showed a diverse

functionality of the transcripts. The most frequent GO terms, in the molecular function category, were the terms “binding” and “catalytic activity”, and in the biological process the term “metabolic process” was most frequent. Functions involving membranes were highly present under the cellular component annotation, and indeed “membrane” was the most frequent term as has been reported previously (Tsai et al., 2016).

5.4.4.3.3 Genes related to transport:

Genes / transcripts were associated to membrane transport, among them Ion transport, AAA-ATPase, chaperone and lipid. Only Ion transporter and lipid transport were found up-regulated in protocorm tissue, providing interesting insight into function in this structure. AAA-ATPase and chaperone are found up-regulated in root and corm tissues. It has been previously reported that some of these genes are co-induced in symbiotic *Cymbidium hybridium* roots including inorganic phosphate transporters and a plasma membrane ATPase, as well as AAA-ATPase and chaperone, among other genes (Zhao et al., 2014). The functional analysis focused here on only the top 100 significant DE genes / transcripts and there were further transporter genes among that data. ATP binding transcripts were up-regulated in corm and root. This gene relates to an ATPase subunit and may be involved in energy metabolism or ion transport as part of the larger class of multisubunit transmembrane ABC transporters (Lennarz and Lane, 2013). An ABC transporter has been directly associated with nutrient transport and is regulated by many abiotic and biotic factors in *Oryza sativa*. It has also been identified as an essential gene for AM symbiosis with involvement in regulation of the symbiosis (Zhang et al., 2010). Following up these ATP binding transcripts could therefore be important for further understanding of the symbiosis in orchid roots.

5.4.4.3.4 Other genes of interest:

An auxin protein (GO:0009734) was detected in our samples as up-regulated in protocorm although not in root. It has been pointed out that an auxin efflux facilitator related to SIPINI of *Solanum lycopersicum* was up-regulated in symbiotic roots of *Cymbidium hybridium* and was a participant of an auxin activated signaling pathway (Zhao et al., 2014). Auxin is known to play a role in many plant-microbe interaction (Sukumar et al. (2013) and it is also known that the auxin/cytokinin ratio determines

the root initiation.

Some of the proteins that were detected have roles to promote the growth of orchids such as carbohydrate, amino acid metabolism, processing and synthesis of protein, in agreement with the findings of other studies (e.g., Zhao et al., 2014). No genes related to pathogenesis response were identified as up-regulated among the top 100 DE in protocorm and this supports previous finding from a study of protocorm in the orchid *Serapias vomeracea*. The study showed that there were no pathogenesis or other stress-related genes significantly up-regulated in mycorrhizal tissues, suggesting that the fungal partner did not influence a strong plant defence response (Perotto et al., 2014).

6 Chapter Six: Conclusions and future work

6.1 Life history of plant:

The monitoring of above ground growth of a small population of *Ophrys apifera* throughout the growing season over three years has determined some of the phenology aspects of its growth for the first time. The data have given a picture of the different stages of vegetative and reproductive development, and performance under fluctuations in the environmental conditions. Now that a baseline has been established, this type of study could be extended to populations growing in more diverse locations to identify the role of the environment in more detail. In addition, it was impossible to study below-ground growth in a small population or to consider the age of the plants that were monitored. A site with a much larger population could allow these features to be measured. In particular, very young plants could be located and recorded to monitor recruitment to the population and also development over years prior of reproduction.

6.2 Seed germination *in vitro:*

This study succeeded in sowing and cultivating seed of *Ophrys apifera in vitro* asymbiotically. The germination of the orchid seed is extremely complex, with the need for a fungus partner or an alternative substrate to support germination (Rasmussen et al., 2015). Based on the results of the present study, several important requirements should be considered carefully for effective orchid seed germination when using artificial culture such as media, temperature, light, time of seed harvest prior to culturing (mature or immature seeds), and aseptic condition to avoid contamination.

Studying seed germination with or without fungi has an important role to provide new information about propagation of orchid species. Knowing the optimal condition during culture for that specific genus or species is helpful for conservative and phylogenetically purposes (Szendrák, 1997b). Moreover, since information about germination requirements is beneficial for advancing orchid conservation and habitat restoration, it is useful for researchers to share their results even if they were unsuccessful (Rasmussen et al., 2015). The work reported in this study clearly has some limitations, such as unsuccessful symbiotic seed germination. However, this

work has proved that it possible to germinate seed asymbiotically *in vitro* to produce material suitable for transcriptomic study.

6.3 Orchid mycorrhizal fungi:

Isolation and culture methods are usually used to identify fungi from a root, but in fact, some fungi cannot be readily cultured in artificial media. These methods are also time-consuming. Therefore, sequencing PCR products obtained for the ITS region was adopted and preferred, particularly for samples from roots. Several sets of primers were used, all of which had been designed for the ITS region and reported as effective by others. This included using the primer set of ITS1-OF and ITS4-OF, specifically recommend to be specific for amplification of *Basidiomycota* and reducing amplification of plant sequence (Taylor and McCormick, 2008).

This project underlined the value of using Illumina, or another high-throughput platform, for DNA amplicon sequencing of soil samples to study complex fungal communities. Although cost is an issue, the size and comparative speed with which the data can be obtained, making it a method for serious consideration as opposed to other strategies such as cloning. In addition, the high throughput method provides greater flexibility for data analysis and a much better view of fungal species richness (Schmidt et al., 2013). One difficulty experienced in this project, obtaining DNA samples suitable for analysis, is common to any DNA sequencing based strategy.

However, this platform has methodological biases, limitations of markers, and bioinformatics challenging for analysis data, as well as a risk of producing artificial results and misleading conclusion (Lindahl et al., 2013). One particular point is the requirement for characterized ITS fungal sequences from databases. The strategy adopted in this study was to use one high quality, curated source (UNITE). This clearly contained some examples from orchid symbiotic fungi but it would be interesting to repeat the data analysis using one or more different sources of ITS data.

The results of this study show that there is clearly a relationship between specific fungi and orchids, which has already been reported in the literature. However, this is the first study on the identification of Rhizoctinia-like fungi with the terrestrial orchid species *Ophrys apifera* and specifically presence from the soil around the orchid plant. Information from the soil samples indicated that genera and species

suitable for an orchid mycorrizal association were present throughout the seasons, locations and years sampled. This addresses the question of whether this association is determined by fungal distribution, or selection by orchid or both (Pellegrino et al., 2016). The data obtained here indicate that appropriate fungi are constantly present but the plant side of the relationship requires further study. The distribution and abundance of orchid mycorrhizal fungi may well impact on the above-ground distribution of the orchid host in the field since presence does not mean that a relationship will be established. This has been clearly shown in a study when the relevant fungi declined with increasing distance between adult plant and successful seedling establishment (Waud et al., 2016). Further, deeper characterization of the relationship could consider whether the fungi shift with the season and how these match with different stages of orchid growth. The insight into the life history of *Ophrys apifera* gained during the early part of this project would be useful in making these connections.

6.4 Transcriptome analysis:

This project has obtained high coverage data from RNA-seq analysis of transcriptomes from protocorm, root, and corm, adopting a strategy that has now been used successfully in many studies of plant, including orchid, function (Vijay et al., 2013). Several different tools were used at various steps in the downstream analysis which influence the outcome of the work (Wolf, 2013). For example, although a GO ontology database is important for comparing gene functions to link gene expression analysis to gene interaction information (Chindelevitch et al., 2012), the data is often based on inbred-strains of model organisms and can give limited insight into function. In addition, although information on plants in databases has increased substantially, it can still be sparse and not provide a full picture of the roles of genes and proteins. The data in this study therefore, provides a significant resource to add to current knowledge.

Previous transcriptomic work with *Ophrys* species has focused on their floral structures. Therefore, this is the first project to concentrate on the underground organs of protocorm, root and corm and characterize differential gene expression for these tissues in *Ophrys apifera*. It has shown that genes are up-regulated in the protocorm related to functions in nutrient transfer, adaptation to environmental

changes and storage in seed, while others such as an-peroxidase, RNA-binding and heme-binding functions which are responsive to stress and active defense were up-regulated in corm and root. Some genes that have been reported in previous work, such as involvement in phosphate transport in *Cymbidium* (Zhao et al., 2014), were not among the top DE genes in this study.

7 References

ABARENKOV, K., TEDERSOO, L., NILSSON, R. H., VELLAK, K., SAAR, I., VELDRE, V., PARMASTO, E., PROUS, M., AAN, A. & OTS, M. 2010. PlutoF—a web based workbench for ecological and taxonomic research, with an online implementation for fungal ITS sequences. *Evolutionary Bioinformatics Online,* 6**,** 189.

AGRIOS, G. 2005. Plant diseases caused by fungi. *Plant pathology,* 4.

AL NAGI, M., ABD-ELSALAM, K. A., BAHKALI, A. H. & MONGY, A.-A. A. 2013. Improved method for DNA isolation from different types of soil infested with three fungal genera. *Romanian Agricultural Research,* 30**,** 329-333.

ALAGNA, F., D'AGOSTINO, N., TORCHIA, L., SERVILI, M., RAO, R., PIETRELLA, M., GIULIANO, G., CHIUSANO, M. L., BALDONI, L. & PERROTTA, G. 2009. Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *Bmc Genomics,* 10**,** 399.

ALEXANDER, C. & ALEXANDER, I. Seasonal changes in populations of the orchid Goodyera repens Br. and in its mycorrhizal development. Transactions of the Botanical Society of Edinburgh, 1984. Taylor & Francis, 219-227.

AMIRI, A., HOLB, I. J. & SCHNABEL, G. 2009. A new selective medium for the recovery and enumeration of Monilinia fructicola, M. fructigena, and M. laxa from stone fruits. *Phytopathology,* 99**,** 1199-1208.

ANDERS, S. & HUBER, W. 2010. Differential expression analysis for sequence count data. *Genome biology,* 11**,** R106.

ANDERS, S., PYL, P. T. & HUBER, W. 2014. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics***,** btu638.

ANDERSON, A. 1991. Symbiotic and asymbiotic germination and growth of Spiranthes magnicamporum (Orchidaceae). *Lindleyana,* 6**,** 183-186.

ARDITTI, J. 1977. Clonal propagation of orchids by means of tissue culture--a manual. *Orchid Biology: Reviews and Perspectives (USA).*

ARDITTI, J. 1992. *Fundamentals of orchid biology*, John Wiley & Sons.

ARDITTI, J. 2009. *Micropropagation of orchids*, John Wiley & Sons.

ARDITTI, J. & ERNST, R. 1993. Micropropagation of orchids. John Willey & Sons. *Inc., NY*.

ARDITTI, J. & GHANI, A. K. A. 2000. Tansley Review No. 110. Numerical and physical properties of orchid seeds and their biological implications. *The New Phytologist,* 145**,** 367-421.

ARIZA, M. C. 2013. *Mycorrhizal Associations, Life History, and Habitat Characteristics of the Endangered Terrestrial Orchid Spiranthes parksii Correll and Sympatric Congener Spiranthes cernua: Implications for Conservation.*

ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., DAVIS, A. P., DOLINSKI, K., DWIGHT, S. S. & EPPIG, J. T. 2000. Gene Ontology: tool for the unification of biology. *Nature genetics,* 25**,** 25-29.

AUER, P. L. & DOERGE, R. 2010. Statistical design and analysis of RNA sequencing data. *Genetics,* 185**,** 405-416.

AXTNER, J. & SOMMER, S. 2009. Validation of internal reference genes for quantitative real-time PCR in a non-model organism, the yellow-necked mouse, Apodemus flavicollis. *BMC research notes,* 2**,** 264.

BAILAROTE, B. C., LIEVENS, B. & JACQUEMYN, H. 2012. Does mycorrhizal specificity affect orchid decline and rarity? *American Journal of Botany,* 99**,** 1655-1665.

BALESTRINI, R. & LUMINI, E. 2017. Focus on mycorrhizal symbioses. *Applied Soil Ecology*.

BÁLINT, M., SCHMIDT, P. A., SHARMA, R., THINES, M. & SCHMITT, I. 2014. An Illumina metabarcoding pipeline for fungi. *Ecology and evolution,* 4**,** 2642-2653.

BALLARD, W. W. 1987. Sterile propagation of Cypripedium reginae from seeds. *American Orchid Society bulletin (USA)*.

BARAKAT, A., DILORETO, D. S., ZHANG, Y., SMITH, C., BAIER, K., POWELL, W. A., WHEELER, N., SEDEROFF, R. & CARLSON, J. E. 2009. Comparison of the transcriptomes of American chestnut (Castanea dentata) and Chinese chestnut (Castanea mollissima) in response to the chestnut blight infection. *BMC Plant Biology,* 9**,** 51.

BATEMAN, R. M., HOLLINGSWORTH, P. M., PRESTON, J., YI-BO, L., PRIDGEON, A. M. & CHASE, M. W. 2003. Molecular phylogenetics and evolution of Orchidinae and selected Habenariinae (Orchidaceae). *Botanical Journal of the Linnean Society,* 142**,** 1-40.

BATTY, A. L., DIXON, K. W., BRUNDRETT, M. C. & SIVASITHAMPARAM, K. 2002. Orchid conservation and mycorrhizal associations. *Microorganisms in plant conservation and biodiversity.* Springer.

BAYMAN, P. & OTERO, J. T. 2006. Microbial endophytes of orchid roots. *Microbial root endophytes.* Springer.

BEARDMORE, J. & PEGG, G. 1981. A technique for the establishment of mycorrhizal infection in orchid tissue grown in aseptic culture. *New Phytologist,* 87**,** 527-535.

BECKER, R. 2011. Genetic approaches for characterizing soil microbial communities.

BELETSKY, A. V., FILYUSHIN, M. A., GRUZDEV, E. V., MAZUR, A. M., PROKHORTCHOUK, E. B., KOCHIEVA, E. Z., MARDANOV, A. V., RAVIN, N. V. & SKRYABIN, K. G. 2017. De novo transcriptome assembly of the mycoheterotrophic plant Monotropa hypopitys. *Genomics Data,* 11**,** 60-61.

BENES, V., BLAKE, J. & DOYLE, K. 2011. Ribo-Zero Gold Kit: improved RNA-seq results after removal of cytoplasmic and mitochondrial ribosomal RNA. *Nature Methods,* 8.

BENJAMINI, Y. & HOCHBERG, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)***,** 289-300.

BERGMAN, F. J. 1996. Disinfecting orchid seed. *Orchids October***,** 1078-9.

BEYRLE, H., SMITH, S., FRANCO, C. & PETERSON, R. 1995. Colonization of Orchis morio protocorms by a mycorrhizal fungus: effects of nitrogen nutrition and glyphosate in modifying the responses. *Canadian Journal of Botany,* 73**,** 1128-1140.

BHADAURIA, V. 2017. *Next-generation Sequencing and Bioinformatics for Plant Science*, Caister Academic Press.

BHARGAVA, V., HEAD, S. R., ORDOUKHANIAN, P., MERCOLA, M. & SUBRAMANIAM, S. 2014. Technical variations in low-input RNA-seq methodologies. *Scientific reports,* 4**,** 3678.

BIDARTONDO, M. I., BRUNS, T. D., WEIß, M., SÉRGIO, C. & READ, D. J. 2003. Specialized cheating of the ectomycorrhizal symbiosis by an epiparasitic liverwort. *Proceedings of the Royal Society of London B: Biological Sciences,* 270**,** 835-842.

BLAUD, A., MENON, M., VAN DER ZAAN, B., LAIR, G. J. & BANWART, S. 2017. Chapter Five-Effects of Dry and Wet Sieving of Soil on Identification and Interpretation of Microbial Community Composition. *Advances in Agronomy,* 142**,** 119-142.

BORRISS, H. & ALBRECHT, L. 1969. Rationelle Samenvermehrung und Anzucht europäischer Erdorchideen. *Gartenwelt,* 69**,** 511-513.

BORSOS, O. 1990. An anatomical-histochemical study of wild orchid tubers in Hungary. *Acta Univ. Wratislaviensis***,** 25-31.

BOUGOURE, J. J., BOUGOURE, D. S., CAIRNEY, J. W. & DEARNALEY, J. D. 2005. ITS-RFLP and sequence analysis of endophytes from Acianthus, Caladenia and Pterostylis (Orchidaceae) in southeastern Queensland. *Mycological research,* 109**,** 452-460.

BOUGOURE, J. J., BRUNDRETT, M. C. & GRIERSON, P. F. 2010. Carbon and nitrogen supply to the underground orchid, Rhizanthella gardneri. *New Phytologist,* 186**,** 947-956.

BRAY, J. R. & CURTIS, J. T. 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecological monographs,* 27**,** 325-349.

BREDDY, N. 1991. Orchid mycorrhiza and symbiotic raising techniques. *American Orchid Society bulletin (USA).*

BREITKOPF, H. 2011. Dynamic speciation processes in the Mediterranean orchid genus Ophrys L.(Orchidaceae).

BROWN, D., GROOM, C., CVITANIK, M., BROWN, M., COOPER, J. & ARDITTI, J. 1981. Effects of fungicides and bactericides on orchid seed germination and shoot tip cultures in vitro. *Plant Cell, Tissue and Organ Culture,* 1**,** 165-180.

BRUNDRETT, M. 2004. Diversity and classification of mycorrhizal associations. *Biological Reviews,* 79**,** 473-495.

BRUNDRETT, M. C. 2006. Understanding the roles of multifunctional mycorrhizal and endophytic fungi. *Microbial root endophytes.* Springer.

BRUNDRETT, M. C. 2007. Scientific approaches to Australian temperate terrestrial orchid conservation. *Australian Journal of Botany,* 55**,** 293-307.

BUSTIN, S. A. & NOLAN, T. 2013. Analysis of mRNA expression by real-time PCR. *Real-Time PCR: Adv Technol Appl,* 51**,** 111-135.

CAI, J., LIU, X., VANNESTE, K., PROOST, S., TSAI, W.-C., LIU, K.-W., CHEN, L.-J., HE, Y., XU, Q. & BIAN, C. 2015. The genome sequence of the orchid Phalaenopsis equestris. *Nature genetics,* 47**,** 65.

CAIRNEY, J. W. 2011. Ectomycorrhizal fungi: the symbiotic route to the root for phosphorus in forest soils. *Plant and Soil,* 344**,** 51-71.

CAPORASO, J. G., KUCZYNSKI, J., STOMBAUGH, J., BITTINGER, K., BUSHMAN, F. D., COSTELLO, E. K., FIERER, N., PEÑA, A. G., GOODRICH, J. K. & GORDON, J. I. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature methods,* 7**,** 335-336.

CHAISIRI, K. 2016. *Molecular ecology of chigger mites (Acari: Trombiculidae) and associated bacteria in Thailand.* University of Liverpool.

CHANG, E., ZHAO, Y., WEI, Q., SHI, S. & JIANG, Z. 2016. Isolation of high-quality RNA from Platycladus orientalis and other Cupressaceae plants. *Electronic Journal of Biotechnology,* 23**,** 21-27.

CHANG, Y.-Y., CHU, Y.-W., CHEN, C.-W., LEU, W.-M., HSU, H.-F. & YANG, C.-H. 2011. Characterization of Oncidium 'Gower Ramsey'transcriptomes using 454 GS-FLX pyrosequencing and their application to the identification of genes associated with flowering time. *Plant and Cell Physiology,* 52**,** 1532-1545.

CHAO, A. 1984. Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of statistics***,** 265-270.

CHINDELEVITCH, L., ZIEMEK, D., ENAYETALLAH, A., RANDHAWA, R., SIDDERS, B., BROCKEL, C. & HUANG, E. S. 2012. Causal reasoning on biological networks: interpreting transcriptional changes. *Bioinformatics,* 28**,** 1114-1121.

CHOMCZYNSKI, P. 1993. A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *Biotechniques,* 15**,** 532-4, 536-7.

CHOMCZYNSKI, P. & SACCHI, N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical biochemistry,* 162**,** 156-159.

CHUTIMA, R. 2012. *Enhancing the propagation successof thai terrestrial orchids with mycorrhizal fungi= การ เพิ่ม ความ สำเร็จ ใน การ ขยาย พันธุ์ กล้วยไม้ ดิน ของ ไทย ด้วย เชื้อรา ไม คอ ร์ ไร ซา.* Chiang mai: Graduate School, Chiang Mai University, 2012.

CLEMENTS, M. & ELLYARD, R. 1979. The symbiotic germination of Australian terrestrial orchids [Pterostylis, Diuris, Thelymitra inocultates with mycorrhizal fungi Tulasnella and Ceratobandium]. *American Orchid Society Bulletin*.

CLEMENTS, M., MUIR, H. & CRIBB, P. 1986. A preliminary report on the symbiotic germination of European terrestrial orchids. *Kew Bulletin***,** 437-445.

COKE, J. 1990. Aseptic germination and growth of some terrestrial orchids. *North American native terrestrial orchid propagation and production***,** 90-91.

COLE, J. R., WANG, Q., CARDENAS, E., FISH, J., CHAI, B., FARRIS, R. J., KULAM-SYED-MOHIDEEN, A. S., MCGARRELL, D. M., MARSH, T., GARRITY, G. M. & TIEDJE, J. M. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research,* 37**,** D141-D145.

CONG, J., YANG, Y., LIU, X., LU, H., LIU, X., ZHOU, J., LI, D., YIN, H., DING, J. & ZHANG, Y. 2015. Analyses of soil microbial community compositions and functional genes reveal potential consequences of natural forest succession. *Scientific reports,* 5**,** 10007.

CONSORTIUM, G. O. 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic acids research,* 32**,** D258-D261.

COON, M. J. 2005. Cytochrome P450: nature's most versatile biological catalyst. *Annu. Rev. Pharmacol. Toxicol.,* 45**,** 1-25.

CULLEN, D. & HIRSCH, P. 1998. Simple and rapid method fordirect extraction of microbial DNA fromsoil for PCR. *Soil Biology and Biochemistry,* 30**,** 983-993.

CURRAH, R., SMRECIU, E. & HAMBLETON, S. 1990. Mycorrhizae and mycorrhizal fungi of boreal species of Platanthera and Coeloglossum (Orchidaceae). *Canadian Journal of Botany,* 68**,** 1171-1181.

CURRAH, R., ZETTLER, L. & MCINNIS, T. 1997. Epulorhiza inquilina sp. nov. from Platanthera (Orchidaceae) and a key to Epulorhiza species.

DAVIDSON, N. M. & OSHLACK, A. 2014. Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome biology,* 15**,** 410.

DAVIES, P. J. 2010. *The plant hormones: their nature, occurrence, and functions*, Springer.

DAVIS, N., IYER, S. & DIENER, U. 1987. Improved method of screening for aflatoxin with a coconut agar medium. *Applied and Environmental Microbiology,* 53**,** 1593-1595.

DEARNALEY, J. D. 2007. Further advances in orchid mycorrhizal research. *Mycorrhiza,* 17**,** 475-486.

DEARNALEY, J. D., MARTOS, F. & SELOSSE, M.-A. 2012. 12 Orchid Mycorrhizas: Molecular Ecology, Physiology, Evolution and Conservation Aspects. *Fungal associations.* Springer.

DEVEY, D. S., BATEMAN, R. M., FAY, M. F. & HAWKINS, J. A. 2009. Genetic structure and systematic relationships within the Ophrys fuciflora aggregate (Orchidaceae: Orchidinae): high diversity in Kent and a wind-induced discontinuity bisecting the Adriatic. *Annals of Botany,* 104**,** 483-495.

DEVISETTY, U. K., KENNEDY, K., SARANDO, P., MERCHANT, N. & LYONS, E. 2016. Bringing your tools to CyVerse Discovery Environment using Docker. *F1000Research,* 5.

DING, R., CHEN, X.-H., ZHANG, L.-J., YU, X.-D., QU, B., DUAN, R. & XU, Y.-F. 2014. Identity and specificity of Rhizoctonia-like fungi from different populations of Liparis japonica (Orchidaceae) in Northeast China. *PloS one,* 9**,** e105573.

DIX, N. J. & WEBSTER, J. 1995. Fungi of extreme environments. *Fungal Ecology.* Springer.

DOBNIK, D., BAEBLER, Š., KOGOVŠEK, P., POMPE-NOVAK, M., ŠTEBIH, D., PANTER, G., JANEŽ, N., MORISSET, D., ŽEL, J. & GRUDEN, K. 2013. β-1, 3-glucanase class III promotes spread of PVYNTN and improves in planta protein production. *Plant biotechnology reports,* 7**,** 547-555.

DOYLE, J. 1991. DNA protocols for plants. *Molecular techniques in taxonomy*. Springer.
DRESSLER, R. L. 1993. *Phylogeny and classification of the orchid family*, Cambridge University Press.
DYER, S. K. & MCCAMMON, S. 1994. Detection of toxigenic isolates of Aspergillus flavus and related species on coconut cream agar. *Journal of Applied Microbiology,* 76**,** 75-78.
EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research,* 32**,** 1792-1797.
EDGAR, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics,* 26**,** 2460-2461.
EDGCOMB, V. P., MCDONALD, J. H., DEVEREUX, R. & SMITH, D. W. 1999. Estimation of bacterial cell numbers in humic acid-rich salt marsh sediments with probes directed to 16S ribosomal DNA. *Applied and environmental microbiology,* 65**,** 1516-1523.
EGAN, A. N., SCHLUETER, J. & SPOONER, D. M. 2012. Applications of next-generation sequencing in plant biology. Botanical Soc America.
EKBLOM, R. & GALINDO, J. 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity,* 107**,** 1-15.
EL-TARABILY, K. A. 2008. Promotion of tomato (Lycopersicon esculentum Mill.) plant growth by rhizosphere competent 1-aminocyclopropane-1-carboxylic acid deaminase-producing streptomycete actinomycetes. *Plant and Soil,* 308**,** 161-174.
EPP, L. S., BOESSENKOOL, S., BELLEMAIN, E. P., HAILE, J., ESPOSITO, A., RIAZ, T., ERSEUS, C., GUSAROV, V. I., EDWARDS, M. E. & JOHNSEN, A. 2012. New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Molecular ecology,* 21**,** 1821-1833.
FALGUERAS, J., LARA, A. J., FERNÁNDEZ-POZO, N., CANTÓN, F. R., PÉREZ-TRABADO, G. & CLAROS, M. G. 2010. SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read. *BMC bioinformatics,* 11**,** 38.
FANG, S.-C., CHEN, J.-C. & WEI, M.-J. 2016. Protocorms and protocorm-like bodies are molecularly distinct from zygotic embryonic tissues. *Plant physiology***,** pp. 00841.2016.
FAST, G. 1982. European terrestrial orchids (symbiotic and asymbiotic methods)[Seed germination and culture, methods]. *Orchid Biology: Reviews and Perspectives*.
FELSENSTEIN, J. & FELENSTEIN, J. 2004. *Inferring phylogenies*, Sinauer associates Sunderland, MA.
FLEIGE, S. & PFAFFL, M. W. 2006. RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular aspects of medicine,* 27**,** 126-139.
FOCHI, V., CHITARRA, W., KOHLER, A., VOYRON, S., SINGAN, V. R., LINDQUIST, E. A., BARRY, K. W., GIRLANDA, M., GRIGORIEV, I. V. & MARTIN, F. 2017. Fungal and plant gene expression in the Tulasnella calospora–Serapias vomeracea symbiosis provides clues about nitrogen pathways in orchid mycorrhizas. *New Phytologist,* 213**,** 365-379.
FRANSSEN, S. U., SHRESTHA, R. P., BRÄUTIGAM, A., BORNBERG-BAUER, E. & WEBER, A. P. 2011. Comprehensive transcriptome analysis of the highly complex Pisum sativum genome using next generation sequencing. *BMC genomics,* 12**,** 227.
FRERICKS, J. 2014. The effects of endophytic fungi of NZ terrestrial orchids: developing methods for conservation.
FROESCHKE, G. & VON DER HEYDEN, S. 2014. A review of molecular approaches for investigating patterns of coevolution in marine host-parasite relationships. *Adv. Parasitol,* 84**,** 209-252.
FROSCH, W. 1982. Beseitigung der durch die innere Hulle bedingten Keimhemmung bei europaischen Orchideen. *Orchidee,* 33**,** 145-146.

FROSTEGÅRD, Å., COURTOIS, S., RAMISSE, V., CLERC, S., BERNILLON, D., LE GALL, F., JEANNIN, P., NESME, X. & SIMONET, P. 1999. Quantification of bias related to the extraction of DNA directly from soils. *Applied and environmental microbiology,* 65**,** 5409-5420.

FUCHS, A. & ZIEGENSPECK, H. 1924. Aus der Monographie des Orchis traunsteineri Saut. III. Entwicklungsgeschichte einger deutscher Orchideen. *Bot. Arch,* 5**,** 120-32.

GAMBINO, G., PERRONE, I. & GRIBAUDO, I. 2008. A rapid and effective method for RNA extraction from different tissues of grapevine and other woody plants. *Phytochemical Analysis,* 19**,** 520-525.

GARBER, M., GRABHERR, M. G., GUTTMAN, M. & TRAPNELL, C. 2011. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature methods,* 8**,** 469-477.

GARCÍA, V. G., ONCO, M. P. & SUSAN, V. R. 2006. Review. Biology and systematics of the form genus Rhizoctonia. *Spanish Journal of Agricultural Research,* 4**,** 55-79.

GARDES, M. & BRUNS, T. D. 1993. ITS primers with enhanced specificity for basidiomycetes-application to the identification of mycorrhizae and rusts. *Molecular ecology,* 2**,** 113-118.

GEORGE, E. F., HALL, M. A. & DE KLERK, G.-J. 2008. The components of plant tissue culture media II: organic additions, osmotic and pH effects, and support systems. *Plant propagation by tissue culture.* Springer.

GEZGIN, Y. & ELTEM, R. 2009. Diversity of endophytic fungi from various Aegean and Mediterranean orchids (saleps). *Turkish Journal of Botany,* 33**,** 439-445.

GHOSH, S. & CHAN, C.-K. K. 2016. Analysis of RNA-Seq data using TopHat and Cufflinks. *Plant Bioinformatics: Methods and Protocols***,** 339-361.

GIONGO, A., CRABB, D. B., DAVIS-RICHARDSON, A. G., CHAULIAC, D., MOBBERLEY, J. M., GANO, K. A., MUKHERJEE, N., CASELLA, G., ROESCH, L. F. & WALTS, B. 2010. PANGEA: pipeline for analysis of next generation amplicons. *The ISME journal,* 4**,** 852-861.

GIRLANDA, M., SEGRETO, R., CAFASSO, D., LIEBEL, H. T., RODDA, M., ERCOLE, E., COZZOLINO, S., GEBAUER, G. & PEROTTO, S. 2011. Photosynthetic Mediterranean meadow orchids feature partial mycoheterotrophy and specific mycorrhizal associations. *American Journal of Botany,* 98**,** 1148-1163.

GOFF, L. A., TRAPNELL, C. & KELLEY, D. 2012. CummeRbund: visualization and exploration of Cufflinks high-throughput sequencing data. *R package version,* 2.

GOFF, S. A., VAUGHN, M., MCKAY, S., LYONS, E., STAPLETON, A. E., GESSLER, D., MATASCI, N., WANG, L., HANLON, M. & LENARDS, A. 2011. The iPlant collaborative: cyberinfrastructure for plant biology. *Frontiers in plant science,* 2**,** 34.

GOH, F. J. 2016. *Effects Of Selected Additives For Growth Efficiency On In Vitro Grown Vanilla Planifolia Andrews.* Universiti Sains Malaysia.

GONÇALVES, A. B., PATERSON, R. R. M. & LIMA, N. 2006. Survey and significance of filamentous fungi from tap water. *International Journal of hygiene and environmental health,* 209**,** 257-264.

GONZALEZ, D., CARLING, D. E., KUNINAGA, S., VILGALYS, R. & CUBETA, M. A. 2001. Ribosomal DNA systematics of Ceratobasidium and Thanatephorus with Rhizoctonia anamorphs. *Mycologia***,** 1138-1150.

GOODWIN, S., MCPHERSON, J. D. & MCCOMBIE, W. R. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics,* 17**,** 333-351.

GÖTZ, S., GARCÍA-GÓMEZ, J. M., TEROL, J., WILLIAMS, T. D., NAGARAJ, S. H., NUEDA, M. J., ROBLES, M., TALÓN, M., DOPAZO, J. & CONESA, A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids research,* 36**,** 3420-3435.

GRABHERR, M. G., HAAS, B. J., YASSOUR, M., LEVIN, J. Z., THOMPSON, D. A., AMIT, I., ADICONIS, X., FAN, L., RAYCHOWDHURY, R. & ZENG, Q. 2011a. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology,* 29**,** 644.

GRABHERR, M. G., HAAS, B. J., YASSOUR, M., LEVIN, J. Z., THOMPSON, D. A., AMIT, I., ADICONIS, X., FAN, L., RAYCHOWDHURY, R. & ZENG, Q. 2011b. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature biotechnology,* 29**,** 644.

GRIFFIN, D. W., KELLOGG, C., PEAK, K. & SHINN, E. 2002. A rapid and efficient assay for extracting DNA from fungi. *Letters in applied microbiology,* 34**,** 210-214.

GRIFFITHS, R. I., WHITELEY, A. S., O'DONNELL, A. G. & BAILEY, M. J. 2000. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA-and rRNA-based microbial community composition. *Applied and environmental microbiology,* 66**,** 5488-5491.

GUO, S., ZHENG, Y., JOUNG, J.-G., LIU, S., ZHANG, Z., CRASTA, O. R., SOBRAL, B. W., XU, Y., HUANG, S. & FEI, Z. 2010. Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types. *BMC genomics,* 11**,** 384.

HAAS, B. J., PAPANICOLAOU, A., YASSOUR, M., GRABHERR, M., BLOOD, P. D., BOWDEN, J., COUGER, M. B., ECCLES, D., LI, B. & LIEBER, M. 2013a. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols,* 8**,** 1494-1512.

HAAS, B. J., PAPANICOLAOU, A., YASSOUR, M., GRABHERR, M., BLOOD, P. D., BOWDEN, J., COUGER, M. B., ECCLES, D., LI, B. & LIEBER, M. 2013b. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols,* 8**,** 1494.

HAAS, N. 1977. [Asymbiontic propagation of European terrestrial orchids. 2: Nigritella nigra and N. miniata]. *Orchidee (Germany, FR).*

HADLEY, G. 1982. European terrestrial orchids [Seed germination and culture, methods]. *Orchid Biology: Reviews and Perspectives*.

HADLEY, G. & HARVAIS, G. 1968. The effect of certain growth substances on asymbiotic germination and development of Orchis purpurella. *New Phytologist,* 67**,** 441-445.

HAJIBABAEI, M., SHOKRALLA, S., ZHOU, X., SINGER, G. A. & BAIRD, D. J. 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS one,* 6**,** e17497.

HAMADY, M. & KNIGHT, R. 2009. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome research,* 19**,** 1141-1152.

HAN, R., RAI, A., NAKAMURA, M., SUZUKI, H., TAKAHASHI, H., YAMAZAKI, M. & SAITO, K. 2016. Chapter Two-De Novo Deep Transcriptome Analysis of Medicinal Plants for Gene Discovery in Biosynthesis of Plant Natural Products. *Methods in Enzymology,* 576**,** 19-45.

HARTMAN, A. L., RIDDLE, S., MCPHILLIPS, T., LUDÄSCHER, B. & EISEN, J. A. 2010. Software Introducing WATERS: a Workflow for the Alignment, Taxonomy, and Ecology of Ribosomal Sequences.

HARVAIS, G. 1974. Notes on the biology of some native orchids of Thunder Bay, their endophytes and symbionts. *Canadian Journal of Botany,* 52**,** 451-460.

HARVAIS, G. 1982. An improved culture medium for growing the orchid Cypripedium reginae axenically. *canadian Journal of botany,* 60**,** 2547-2555.

HASELWANDTER, K., PASSLER, V., REITER, S., SCHMID, D., NICHOLSON, G., HENTSCHEL, P., ALBERT, K. & WINKELMANN, G. 2006. Basidiochrome–A Novel Siderophore of the Orchidaceous Mycorrhizal Fungi Ceratobasidium and Rhizoctonia spp. *BioMetals,* 19**,** 335-343.

HAWKSWORTH, D. L. 1991. The fungal dimension of biodiversity: magnitude, significance, and conservation. *Mycological research,* 95**,** 641-655.
HAWKSWORTH, D. L. 2001. The magnitude of fungal diversity: the 1.5 million species estimate revisited** Paper presented at the Asian Mycological Congress 2000 (AMC 2000), incorporating the 2nd Asia-Pacific Mycological Congress on Biodiversity and Biotechnology, and held at the University of Hong Kong on 9-13 July 2000. *Mycological research,* 105**,** 1422-1432.
HAWKSWORTH, D. L. & LUECKING, R. 2017. Fungal Diversity Revisited: 2.2 to 3.8 Million Species. *Microbiology spectrum,* 5.
HEBERT, P. D., CYWINSKA, A. & BALL, S. L. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences,* 270**,** 313-321.
HELLMUTH, M., WIESEKE, N., LECHNER, M., LENHOF, H.-P., MIDDENDORF, M. & STADLER, P. F. 2015. Phylogenomics with paralogs. *Proceedings of the National Academy of Sciences,* 112**,** 2058-2063.
HERRERA, H., VALADARES, R., CONTRERAS, D., BASHAN, Y. & ARRIAGADA, C. 2017. Mycorrhizal compatibility and symbiotic seed germination of orchids from the Coastal Range and Andes in south central Chile. *Mycorrhiza,* 27**,** 175-188.
HIBBETT, D. & THORN, R. 2001. Basidiomycota: homobasidiomycetes. *Systematics and Evolution.* Springer.
HIETALA, A. M., SEN, R. & LILJA, A. 1994. Anamorphic and teleomorphic characteristics of a uninucleate Rhizoctonia sp. isolated from the roots of nursery grown conifer seedlings. *Mycological Research,* 98**,** 1044-1050.
HODKINSON, B. P. & GRICE, E. A. 2015. Next-generation sequencing: a review of technologies and tools for wound microbiome research. *Advances in wound care,* 4**,** 50-58.
HOPPE, E. & HOPPE, H. 1988. Tissue culture of Ophrys apifera. *Lindleyana,* 3**,** 190-194.
HSIAO, Y.-Y., CHEN, Y.-W., HUANG, S.-C., PAN, Z.-J., FU, C.-H., CHEN, W.-H., TSAI, W.-C. & CHEN, H.-H. 2011. Gene discovery using next-generation pyrosequencing to develop ESTs for Phalaenopsis orchids. *BMC genomics,* 12**,** 360.
HUANG, J.-Z., LIN, C.-P., CHENG, T.-C., CHANG, B. C.-H., CHENG, S.-Y., CHEN, Y.-W., LEE, C.-Y., CHIN, S.-W. & CHEN, F.-C. 2015. A de novo floral transcriptome reveals clues into Phalaenopsis orchid flower development. *PloS one,* 10**,** e0123474.
HUH, Y. S., LEE, J. K., NAM, S. Y., HONG, E. Y., PAEK, K. Y. & SON, S. W. 2016. Effects of altering medium strength and sucrose concentration on in vitro germination and seedling growth of Cypripedium macranthos Sw. *Journal of Plant Biotechnology,* 43**,** 132-137.
HUMMON, A. B., LIM, S. R., DIFILIPPANTONIO, M. J. & RIED, T. 2007. Isolation and solubilization of proteins after TRIzol® extraction of RNA and DNA from patient material following prolonged storage. *Biotechniques,* 42**,** 467.
HUTCHINGS, M. J. 2010. The population biology of the early spider orchid Ophrys sphegodes Mill. III. Demography over three decades. *Journal of Ecology,* 98**,** 867-878.
ISLAM, M. R., KIM, H., KANG, S.-W., KIM, J.-S., JEONG, Y.-M., HWANG, H.-J., LEE, S.-Y., WOO, J.-C. & KIM, S.-G. 2007. Functional characterization of a gene encoding a dual domain for uridine kinase and uracil phosphoribosyltransferase in Arabidopsis thaliana. *Plant molecular biology,* 63**,** 465-477.
JAAKOLA, L., PIRTTILÄ, A. M., HALONEN, M. & HOHTOLA, A. 2001. Isolation of high quality RNA from bilberry (Vaccinium myrtillus L.) fruit. *Molecular biotechnology,* 19**,** 201-203.
JANEČKOVÁ, P., WOTAVOVÁ, K., SCHÖDELBAUEROVÁ, I., JERSÁKOVÁ, J. & KINDLMANN, P. 2006. Relative effects of management and environmental

conditions on performance and survival of populations of a terrestrial orchid, Dactylorhiza majalis. *Biological Conservation,* 129**,** 40-49.
JEEWON, R., LIEW, E. & HYDE, K. 2004. Phylogenetic evaluation of species nomenclature of Pestalotiopsis in relation to host association. *Fungal Diversity*.
JERVIS-BARDY, J., LEONG, L. E., MARRI, S., SMITH, R. J., CHOO, J. M., SMITH-VAUGHAN, H. C., NOSWORTHY, E., MORRIS, P. S., O'LEARY, S. & ROGERS, G. B. 2015. Deriving accurate microbiota profiles from human samples with low bacterial content through post-sequencing processing of Illumina MiSeq data. *Microbiome,* 3**,** 19.
JEVSNIK, T. & LUTHAR, Z. 2015. Successful disinfection protocol for orchid seeds and influence of gelling agent on germination and growth. *Acta agriculturae Slovenica,* 105**,** 95-102.
JHENG, F.-Y., DO, Y.-Y., LIAUH, Y.-W., CHUNG, J.-P. & HUANG, P.-L. 2006. Enhancement of growth and regeneration efficiency from embryogenic callus cultures of Oncidium 'Gower Ramsey'by adjusting carbohydrate sources. *Plant Science,* 170**,** 1133-1140.
JOHNSON, L. F., CURL, E. A., BOND, J. H. & FRIBOURG, H. A. 1959. Methods for studying soil microflora: plant disease relationships. Minnesota, US: Burgess Publishing.
JOHNSON, M. T., CARPENTER, E. J., TIAN, Z., BRUSKIEWICH, R., BURRIS, J. N., CARRIGAN, C. T., CHASE, M. W., CLARKE, N. D., COVSHOFF, S. & EDGER, P. P. 2012. Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. *PloS one,* 7**,** e50226.
JORDON-THADEN, I. E., CHANDERBALI, A. S., GITZENDANNER, M. A. & SOLTIS, D. E. 2015. Modified CTAB and TRIzol protocols improve RNA extraction from chemically complex Embryophyta. *Applications in plant sciences,* 3**,** 1400105.
JORDON-THADEN, I. E., CHANDERBALI, A. S., GITZENDANNER, M. A. & SOLTIS, D. E. 2015. Modified CTAB and TRIzol protocols improve RNA extraction from chemically complex Embryophyta. *Applications in plant sciences,* 3.
JOSHI, N. & FASS, J. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33)[Software].
JOST, L. 2006. Entropy and diversity. *Oikos,* 113**,** 363-375.
JUKES, T. H., CANTOR, C. R. & MUNRO, H. 1969. Evolution of protein molecules. *Mammalian protein metabolism,* 3**,** 132.
KAUR, J., POFF, K. E. & SHARMA, J. 2018. A rare temperate terrestrial orchid selects similar Tulasnella taxa in ex situ and in situ environments. *Plant Ecology,* 219**,** 45-55.
KENNEDY, K., HALL, M. W., LYNCH, M. D., MORENO-HAGELSIEB, G. & NEUFELD, J. D. 2014. Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles. *Applied and environmental microbiology,* 80**,** 5717-5722.
KÉRY, M. & GREGG, K. B. 2003. Effects of life-state on detectability in a demographic study of the terrestrial orchid Cleistes bifaria. *Journal of Ecology,* 91**,** 265-273.
KÉRY, M. & GREGG, K. B. 2004. Demographic analysis of dormancy and survival in the terrestrial orchid Cypripedium reginae. *Journal of Ecology,* 92**,** 686-695.
KINDLMANN, P. & BALOUNOVÁ, Z. 2001. Irregular flowering patterns in terrestrial orchids: theories vs. empirical data. *Web Ecology,* 2**,** 75-82.
KITSAKI, C., ZYGOURAKI, S., ZIOBORA, M. & KINTZIOS, S. 2004. In vitro germination, protocorm formation and plantlet development of mature versus immature seeds from several Ophrys species (Orchidaceae). *Plant cell reports,* 23**,** 284-290.
KNUDSON, L. 1922. Nonsymbiotic germination of orchid seeds. *Botanical gazette,* 73**,** 1-25.

KNUDSON, L. 1948. Chlorox and calcium hypochlorite as disinfectants for orchid seed. *Am Orchid Soc Bull,* 17**,** 348-353.
KOHOUT, P., TĚŠITELOVÁ, T., ROY, M., VOHNÍK, M. & JERSÁKOVÁ, J. 2013. A diverse fungal community associated with Pseudorchis albida (Orchidaceae) roots. *Fungal Ecology,* 6**,** 50-64.
KÕLJALG, U., NILSSON, R. H., ABARENKOV, K., TEDERSOO, L., TAYLOR, A. F., BAHRAM, M., BATES, S. T., BRUNS, T. D., BENGTSSON-PALME, J. & CALLAGHAN, T. M. 2013. Towards a unified paradigm for sequence-based identification of fungi. *Molecular ecology,* 22**,** 5271-5277.
KOOPOWITZ, H. & LAVARACK, P. 2003. The nature of threats to orchid conservation. In 'Orchid conservation'.(Eds KW Dixon, SP Kell, RL Barrett, PJ Cribb) pp. 25–42. Natural History Publications (Borneo): Kota Kinabalu, Sabah, Malaysia.
KRISTIANSEN, K. A., FREUDENSTEIN, J. V., RASMUSSEN, F. N. & RASMUSSEN, H. N. 2004. Molecular identification of mycorrhizal fungi in Neuwiedia veratrifolia (Orchidaceae). *Molecular Phylogenetics and Evolution,* 33**,** 251-258.
KRISTIANSEN, K. A., TAYLOR, D. L., KJOLLER, R., RASMUSSEN, H. N. & ROSENDAHL, S. 2001. Identification of mycorrhizal fungi from single pelotons of Dactylorhiza majalis (Orchidaceae) using single-strand conformation polymorphism and mitochondrial ribosomal large subunit DNA sequences. *Molecular Ecology,* 10**,** 2089-2093.
KUGA, Y., SAKAMOTO, N. & YURIMOTO, H. 2014. Stable isotope cellular imaging reveals that both live and degenerating fungal pelotons transfer carbon and nitrogen to orchid protocorms. *New Phytologist,* 202**,** 594-605.
KUMAR, S., BANKS, T. W. & CLOUTIER, S. 2016a. SNP discovery through next-generation sequencing and its applications. *Crop Breeding: Bioinformatics and Preparing for Climate Change.* CRC Press.
KUMAR, S., CARLSEN, T., MEVIK, B.-H., ENGER, P., BLAALID, R., SHALCHIAN-TABRIZI, K. & KAUSERUD, H. 2011. C LOTU: An online pipeline for processing and clustering of 454 amplicon reads into OTUs followed by taxonomic annotation. *BMC bioinformatics,* 12**,** 182.
KUMAR, S., STECHER, G. & TAMURA, K. 2016b. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular biology and evolution,* 33**,** 1870-1874.
KUMAR, S., TAMURA, K. & NEI, M. 1994. MEGA: molecular evolutionary genetics analysis software for microcomputers. *Bioinformatics,* 10**,** 189-191.
KURBIDAEVA, A. & NOVOKRESHCHENOVA, M. 2011. Somaclonal variations of Nicotiana tabacum transgenic plants. *Moscow University biological sciences bulletin,* 66**,** 86-90.
LACAP, D., HYDE, K. & LIEW, E. 2003. An evaluation of the fungal'morphotype'concept based on ribosomal DNA sequences. *Fungal Diversity.*
LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology,* 10**,** R25.
LEE, S.-S. 2002. A review of orchid mycorrhizae in Korea. *The Plant Pathology Journal,* 18**,** 169-178.
LEHNEBACH, C. A., ROBERTSON, A. W. & HEDDERLEY, D. 2005. Pollination studies of four New Zealand terrestrial orchids and the implication for their conservation. *New Zealand Journal of Botany,* 43**,** 467-477.
LENNARZ, W. J. & LANE, M. D. 2013. *Encyclopedia of biological chemistry,* Academic Press.
LESICA, P. & STEELE, B. M. 1994. Prolonged dormancy in vascular plants and implications for monitoring studies. *Natural Areas Journal,* 14**,** 209-212.

LEWIS, C., BILKHU, S., ROBERT, V., EBERHARDT, U., SZOKE, S., SEIFERT, K. A. & LÉVESQUE, C. A. 2011. Identification of fungal DNA barcode targets and PCR primers based on Pfam protein families and taxonomic hierarchy. *Open Appl Inform J,* 5**,** 30-44.

LI, D., YIN, H., ZHAO, C., ZHU, G. & LŮ, F. 2014. Transcriptome Analysis of Tessellated and Green Leaves in Paphiopedilum Orchids Using Illumina Paired-End Sequencing and Discovery Simple Sequence Repeat Markers. *J Plant Biochem Physiol,* 2**,** 2.

LI, R., YU, C., LI, Y., LAM, T.-W., YIU, S.-M., KRISTIANSEN, K. & WANG, J. 2009. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics,* 25**,** 1966-1967.

LI, X., LUO, J., YAN, T., XIANG, L., JIN, F., QIN, D., SUN, C. & XIE, M. 2013. Deep sequencing-based analysis of the Cymbidium ensifolium floral transcriptome. *PloS one,* 8**,** e85480.

LIANG, H., CARLSON, J. E., LEEBENS-MACK, J. H., WALL, P. K., MUELLER, L. A., BUZGO, M., LANDHERR, L. L., HU, Y., DILORETO, D. S. & ILUT, D. C. 2008. An EST database for Liriodendron tulipifera L. floral buds: the first EST resource for functional and comparative genomics in Liriodendron. *Tree Genetics & Genomes,* 4**,** 419-433.

LINDAHL, B. D., NILSSON, R. H., TEDERSOO, L., ABARENKOV, K., CARLSEN, T., KJØLLER, R., KÕLJALG, U., PENNANEN, T., ROSENDAHL, S. & STENLID, J. 2013. Fungal community analysis by high-throughput sequencing of amplified markers–a user's guide. *New Phytologist,* 199**,** 288-299.

LIU, S.-H., EDWARDS, C., HOCH, P. C., RAVEN, P. H. & BARBER, J. C. 2016. Complete plastome sequence of Ludwigia octovalvis (Onagraceae), a globally distributed wetland plant. *Genome Announcements,* 4**,** e01274-16.

LIU, S., LI, W., WU, Y., CHEN, C. & LEI, J. 2013. De novo transcriptome assembly in chili pepper (Capsicum frutescens) to identify genes involved in the biosynthesis of capsaicinoids. *PloS one,* 8**,** e48156.

LIU, T.-H. A., LIN, J.-J. & WU, R.-Y. 2006. The effects of using trehalose as a carbon source on the proliferation of Phalaenopsis and Doritaenopsis protocorm-like-bodies. *Plant cell, tissue and organ culture,* 86**,** 125-129.

LIU, X.-D. & SHEN, Y.-G. 2004. NaCl-induced phosphorylation of light harvesting chlorophyll a/b proteins in thylakoid membranes from the halotolerant green alga, Dunaliella salina. *FEBS letters,* 569**,** 337-340.

LUCKE, E. 1971. Zur samenkeimung mediterraner Ophrys. *Orchidee,* 22**,** 62-65.

LUNTER, G. & GOODSON, M. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome research,* 21**,** 936-939.

MACKENZIE, D. J., MCLEAN, M. A., MUKERJI, S. & GREEN, M. 1997. Improved RNA extraction from woody plants for the detection of viral pathogens by reverse transcription-polymerase chain reaction. *Plant Disease,* 81**,** 222-226.

MAHESWARI, T. & RAVI, M. 2016. Next Generation Sequencing–Technical Report. *LIFE SCIENCE RESEARCH,* 1.

MAMIATIS, T., FRITSCH, E., SAMBROOK, J. & ENGEL, J. 1985. Molecular cloning–A laboratory manual. New York: Cold Spring Harbor Laboratory. 1982, 545 S., 42$. Wiley Online Library.

MARTIN, F. 2016. *Molecular Mycorrhizal Symbiosis*, John Wiley & Sons.

MARTIN, J. A. & WANG, Z. 2011. Next-generation transcriptome assembly. *Nature Reviews Genetics,* 12**,** 671-682.

MARTIN, K. J. & RYGIEWICZ, P. T. 2005. Fungal-specific PCR primers developed for analysis of the ITS region of environmental DNA extracts. *BMC microbiology,* 5**,** 28.

MARTIN, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal,* 17**,** pp. 10-12.

MCWILLIAM, H., LI, W., ULUDAG, M., SQUIZZATO, S., PARK, Y. M., BUSO, N., COWLEY, A. P. & LOPEZ, R. 2013. Analysis tool web services from the EMBL-EBI. *Nucleic acids research,* 41**,** W597-W600.

MELO, S., PUNGARTNIK, C., CASCARDO, J. & BRENDEL, M. 2006. Rapid and efficient protocol for DNA extraction and molecular identification of the basidiomycete Crinipellis perniciosa. *Genet. Mol. Res,* 5**,** 851-855.

MENG, Y., YU, D., XUE, J., LU, J., FENG, S., SHEN, C. & WANG, H. 2016. A transcriptome-wide, organ-specific regulatory map of Dendrobium officinale, an important traditional Chinese orchid herb. *Scientific reports,* 6**,** 18864.

METZKER, M. L. 2010. Sequencing technologies-the next generation. *Nature reviews genetics,* 11.

MINH, B. Q., NGUYEN, M. A. T. & VON HAESELER, A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Molecular biology and evolution,* 30**,** 1188-1195.

MITTAPALLI, O., BAI, X., MAMIDALA, P., RAJARAPU, S. P., BONELLO, P. & HERMS, D. A. 2010. Tissue-specific transcriptomics of the exotic invasive insect pest emerald ash borer (Agrilus planipennis). *PLoS one,* 5**,** e13708.

MÖLLER, O. 1987. Die subterrane Innovation und der Wachstumszyklus einiger Erdorchideen. *Die Orchidee,* 38**,** 13-22.

MOORE, D., ROBSON, G. D. & TRINCI, A. P. 2011. *21st century guidebook to fungi with CD*, Cambridge University Press.

MOORE, R. T. 1987. The genera of Rhizoctonia-like fungi: Ascorhizoctonia,

Ceratorhiza gen. nov., Epulorhiza gen. nov., Moniliopsis and Rhizoctonia. *Mycotaxon,* 29**,** 91-99.

MORANT, M., BAK, S., MØLLER, B. L. & WERCK-REICHHART, D. 2003. Plant cytochromes P450: tools for pharmacology, plant protection and phytoremediation. *Current opinion in biotechnology,* 14**,** 151-162.

MORENO, F., FIGUEIRO, G., MANNISE, N., IRIARTE, A., GONZÁLEZ, S., DUARTE, J. M. B. & COSSE, M. 2016. Use of next-generation molecular tools in archaeological neotropical deer sample analysis. *Journal of Archaeological Science: Reports,* 10**,** 403-410.

MORTAZAVI, A., WILLIAMS, B. A., MCCUE, K., SCHAEFFER, L. & WOLD, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods,* 5**,** 621-628.

MRKVICKA, A. C. 1992. Liste der Chromosomenzahlen europaischer Orchideen. *List of chromosome numbers of European orchids.) Mitteilungsbl. Arbeitskr. Heim. Orch. Baden-Wurttemberg,* 24**,** 125-140.

MÜLLER, J., WIEMKEN, A. & AESCHBACHER, R. 1999. Trehalose metabolism in sugar sensing and plant development. *Plant Science,* 147**,** 37-47.

MURASHIGE, T. & SKOOG, F. 1962. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia plantarum,* 15**,** 473-497.

MUTSUURA, O., ITO, I. & NAKAHIRA, R. 1962. Studies on the germination and the development of seedlings of Taeniophyllum aphyllum (Makino) Makino. *Sci. Rep. Kyoto Pref. Univ.(Nat. Sci. & Liv. Sci.), Ser. A,* 3**,** 189-194.

NADARAJAN, J., WOOD, S., MARKS, T., SEATON, P. & PRITCHARD, H. 2011. Nutritional requirements for in vitro seed germination of 12 terrestrial, lithophytic and epiphytic orchids. *Journal of Tropical Forest Science***,** 204-212.

NAGTONG, T., THANONKEO, S., KLANRIT, P. & THANONKEO, P. 2009. Cloning and characterization of 1-aminocyclopropane-1-carboxylate oxidase gene from orchid (Dendrobium spp.). *World Appl. Sci. J,* 7**,** 11-18.

NAURIN, S., HANSSON, B., HASSELQUIST, D., KIM, Y.-H. & BENSCH, S. 2011. The sex-biased brain: sexual dimorphism in gene expression in two species of songbirds. *BMC genomics,* 12**,** 37.

NEHLS, U. 2008. Mastering ectomycorrhizal symbiosis: the impact of carbohydrates. *Journal of experimental botany,* 59**,** 1097-1108.
NELDER, J. A. & BAKER, R. J. 1972. *Generalized linear models*, Wiley Online Library.
NELSON, M. C., MORRISON, H. G., BENJAMINO, J., GRIM, S. L. & GRAF, J. 2014. Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. *PloS one,* 9**,** e94249.
NGUYEN, L.-T., SCHMIDT, H. A., VON HAESELER, A. & MINH, B. Q. 2014. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution,* 32**,** 268-274.
NILSSON, R. H., VELDRE, V., HARTMANN, M., UNTERSEHER, M., AMEND, A., BERGSTEN, J., KRISTIANSSON, E., RYBERG, M., JUMPPONEN, A. & ABARENKOV, K. 2010. An open source software package for automated extraction of ITS1 and ITS2 from fungal ITS sequences for use in high-throughput community assays and molecular ecology. *Fungal Ecology,* 3**,** 284-287.
NIU, S.-C., XU, Q., ZHANG, G.-Q., ZHANG, Y.-Q., TSAI, W.-C., HSU, J.-L., LIANG, C.-K., LUO, Y.-B. & LIU, Z.-J. 2016. De novo transcriptome assembly databases for the butterfly orchid Phalaenopsis equestris. *Scientific Data,* 3**,** 160083.
NOGUEIRA, R. E., BERG, C. V. D., PEREIRA, O. L. & KASUYA, M. C. M. 2014. Isolation and molecular characterization of Rhizoctonia-like fungi associated with orchid roots in the Quadrilátero Ferrífero and Zona da Mata regions of the state of Minas Gerais, Brazil. *Acta Botanica Brasilica,* 28**,** 298-300.
NONTACHAIYAPOOM, S., SASIRAT, S. & MANOCH, L. 2010. Isolation and identification of Rhizoctonia-like fungi from roots of three orchid genera, Paphiopedilum, Dendrobium, and Cymbidium, collected in Chiang Rai and Chiang Mai provinces of Thailand. *Mycorrhiza,* 20**,** 459-471.
NOTREDAME, C., HIGGINS, D. G. & HERINGA, J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology,* 302**,** 205-217.
NOVAES, E., DROST, D. R., FARMERIE, W. G., PAPPAS, G. J., GRATTAPAGLIA, D., SEDEROFF, R. R. & KIRST, M. 2008. High-throughput gene and SNP discovery in Eucalyptus grandis, an uncharacterized genome. *BMC genomics,* 9**,** 312.
OGURA-TSUJITA, Y., YOKOYAMA, J., MIYOSHI, K. & YUKAWA, T. 2012. Shifts in mycorrhizal fungi during the evolution of autotrophy to mycoheterotrophy in Cymbidium (Orchidaceae). *American Journal of Botany,* 99**,** 1158-1176.
OLIVER, S. L., LENARDS, A. J., BARTHELSON, R. A., MERCHANT, N. & MCKAY, S. J. 2013. Using the iPlant collaborative discovery environment. *Current protocols in bioinformatics***,** 1.22. 1-1.22. 26.
OSHLACK, A., ROBINSON, M. D. & YOUNG, M. D. 2010. From RNA-seq reads to differential expression results. *Genome biology,* 11**,** 220.
OSIADACZ, B. & KRĘCIAŁA, M. 2014. Ophrys apifera Huds.(Orchidaceae), a new orchid species to the flora of Poland. *Biodiversity Research and Conservation,* 36**,** 11-16.
OTERO, J. T., ACKERMAN, J. D. & BAYMAN, P. 2002a. Diversity and host specificity of endophytic Rhizoctonia-like fungi from tropical orchids. *American Journal of Botany,* 89**,** 1852-1858.
OTERO, J. T., ACKERMAN, J. D. & BAYMAN, P. 2002b. Diversity and host specificity of endophytic Rhizoctonia-like fungi from tropical orchids. *American Journal of Botany,* 89**,** 1852-1858.
OTERO, J. T., FLANAGAN, N. S., HERRE, E. A., ACKERMAN, J. D. & BAYMAN, P. 2007. Widespread mycorrhizal specificity correlates to mycorrhizal function in the neotropical, epiphytic orchid Ionopsis utricularioides (Orchidaceae). *American Journal of Botany,* 94**,** 1944-1950.
OTERO OSPINA, J. T. & BAYMAN, P. 2009. Symbiotic vs. asymbiotic seed germination in epiphytic orchids. *Acta Agronómica,* 58**,** 270-276.

PANDEY, R. V., NOLTE, V. & SCHLÖTTERER, C. 2010. CANGS: a user-friendly utility for processing and analyzing 454 GS-FLX data in biodiversity studies. *BMC research notes,* 3**,** 3.

PANT, B. 2013. Medicinal orchids and their uses: tissue culture a potential alternative for conservation. *African Journal of Plant Science,* 7**,** 448-467.

PARK, M. R., WANG, Y.-H. & HASENSTEIN, K. H. 2014. Profiling gene expression in germinating Brassica roots. *Plant molecular biology reporter,* 32**,** 541-548.

PARNISKE, M. 2008. Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nature Reviews Microbiology,* 6**,** 763-775.

PARRY, D. W. 1990. *Plant pathology in agriculture,* CUP Archive.

PASZKIEWICZ, K. & STUDHOLME, D. J. 2010. De novo assembly of short sequence reads. *Briefings in bioinformatics***,** bbq020.

PEASE, J. & SOOKNANAN, R. 2012. A rapid, directional RNA-seq library preparation workflow for Illumina® sequencing. *nature methods,* 9.

PEAY, K. G., BRUNS, T. D., KENNEDY, P. G., BERGEMANN, S. E. & GARBELOTTO, M. 2007. A strong species–area relationship for eukaryotic soil microbes: island size matters for ectomycorrhizal fungi. *Ecology letters,* 10**,** 470-480.

PECORARO, L., GIRLANDA, M., LIU, Z.-J., HUANG, L. & PEROTTO, S. 2015. Molecular analysis of fungi associated with the Mediterranean orchid Ophrys bertolonii Mor. *Annals of microbiology,* 65**,** 2001-2007.

PEDERSEN, H. Æ., FAURHOLDT, N. & MANUEL, R. L. 2007. *Ophrys: the bee orchids of Europe,* Kew Publishing Kew.

PELLEGRINO, G., LUCA, A. & BELLUSCI, F. 2016. Relationships between orchid and fungal biodiversity: Mycorrhizal preferences in Mediterranean orchids. *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology,* 150**,** 180-189.

PEREIRA, O. L., ROLLEMBERG, C. L., BORGES, A. C., KASUYA, M. C. & MATSUOKA, K. 2003. Epulorhiza epiphytica sp. nov. isolated from mycorrhizal roots of epiphytic orchids in Brazil. *Mycoscience,* 44**,** 153-155.

PEROTTO, S., RODDA, M., BENETTI, A., SILLO, F., ERCOLE, E., RODDA, M., GIRLANDA, M., MURAT, C. & BALESTRINI, R. 2014. Gene expression in mycorrhizal orchid protocorms suggests a friendly plant–fungus relationship. *Planta,* 239**,** 1337-1349.

PIERCE, S. & BELOTTI, J. 2011. The conservation of terrestrial orchids. *Kindle eBook edition (English language edition translated from the Italian print edition, La Conservazione delle Orchidee Terrestri: dalle Alpi alla Pianura Padana Lombarda). Galbiate: The Native Flora Centre of the Lombardy Region (CFA)*.

PIERCE, S. & CERABOLINI, B. 2011. Asymbiotic germination of the White Mountain Orchid (Pseudorchis albida) from immature seed on media enriched with complex organics or phytohormones. *Seed Science and Technology,* 39**,** 199-203.

PIERCE, S., CERIANI, R. M., VILLA, M. & CERABOLINI, B. 2006. Quantifying relative extinction risks and targeting intervention for the orchid flora of a natural park in the European prealps. *Conservation biology,* 20**,** 1804-1810.

POLEKHINA, G., GUPTA, A., VAN DENDEREN, B. J., FEIL, S. C., KEMP, B. E., STAPLETON, D. & PARKER, M. W. 2005. Structural basis for glycogen recognition by AMP-activated protein kinase. *Structure,* 13**,** 1453-1462.

PONERT, J. 2009. *Early Stages of development of terrestrial orchids: the effect of saccharides and phytohormones.* M. Sc. thesis, Charles University Prague.[In Czech.].

PONERT, J., FIGURA, T., VOSOLSOBĚ, S., LIPAVSKÁ, H., VOHNÍK, M. & JERSÁKOVÁ, J. 2013. Asymbiotic germination of mature seeds and protocorm development of Pseudorchis albida (Orchidaceae) are inhibited by nitrates even at extremely low concentrations. *Botany,* 91**,** 662-670.

PONERT, J., VOSOLSOBĚ, S., KMECOVÁ, K. & LIPAVSKÁ, H. 2011. European orchid cultivation–from seed to mature plant. *European Journal of Environmental Sciences,* 1.

POP, M. & SALZBERG, S. L. 2008. Bioinformatics challenges of new sequencing technology. *Trends in Genetics,* 24**,** 142-149.

PORRAS-ALFARO, A. & BAYMAN, P. 2007. Mycorrhizal fungi of Vanilla: diversity, specificity and effects on seed germination and plant growth. *Mycologia,* 99**,** 510-525.

POTTER, R. M. 2008. *Constructing Phylogenetic Trees using Multiple Sequence Alignment.* University of Washington, Tacoma.

PRABHUDAS, S. K., RAJU, B., KANNAN THODI, S., PARANI, M. & NATARAJAN, P. 2016. The complete chloroplast genome sequence of Indian mustard (Brassica juncea L.). *Mitochondrial DNA Part A,* 27**,** 4622-4623.

PRINCE, R. C. & GUNSON, D. E. 1994. Just plain vanilla? *Trends in biochemical sciences,* 19**,** 521.

PULMAN, J. 2014. *A transcriptomics approach to understanding polymorphic and transcript level differences linked to isoquinoline alkaloid production in triploid varieties of Narcissus pseudonarcissus.* University of Liverpool.

RADEMACHER, M. (ed.) 2012. *Orchids in quarries and gravel pits*.

RAJA, H. A., MILLER, A. N., PEARCE, C. J. & OBERLIES, N. H. 2017. Fungal identification using molecular tools: a primer for the natural products research community. *Journal of natural products,* 80**,** 756-770.

RAO, P. R. 1970. Studies on soil fungi III. Seasonal variation and distribution of microfungi in some soils of Andhra Pradesh (India). *Mycopathologia,* 40**,** 277-298.

RASMUSSEN, H. N. 1990. Cell differentiation and mycorrhizal infection in Dactylorhiza majalis (Rchb. f.) Hunt & Summerh.(Orchidaceae) during germination in vitro. *New phytologist,* 116**,** 137-147.

RASMUSSEN, H. N. 1992. Seed dormancy patterns in Epipactis palustris (Orchidaceae): Requirements for germination and establishment of mycorrhiza. *Physiologia Plantarum,* 86**,** 161-167.

RASMUSSEN, H. N. 1995. *Terrestrial orchids: from seed to mycotrophic plant*, Cambridge University Press.

RASMUSSEN, H. N. 2002. Recent developments in the study of orchid mycorrhiza. *Plant and soil,* 244**,** 149-163.

RASMUSSEN, H. N., DIXON, K. W., JERSÁKOVÁ, J. & TĚŠITELOVÁ, T. 2015. Germination and seedling establishment in orchids: a complex of requirements. *Annals of botany,* 116**,** 391-402.

RASMUSSEN, H. N. & RASMUSSEN, F. N. 2009. Orchid mycorrhiza: implications of a mycophagous life style. *Oikos,* 118**,** 334-345.

RICO, M., BRUIX, M., GONZÁLEZ, C., MONSALVE, R. I. & RODRÍGUEZ, R. 1996. 1H NMR assignment and global fold of napin BnIb, a representative 2S albumin seed protein. *Biochemistry,* 35**,** 15672-15682.

ROBERT, K. C. & JONATHAN, A. C. 1994. Estimating terrestrial biodiversity through extrapolation. Nov.

ROBERTS, P. 1999. *Rhizoctonia-forming fungi*, Herbarium, Royal Botanic Gardens.

RONDON, M. R., AUGUST, P. R., BETTERMANN, A. D., BRADY, S. F., GROSSMAN, T. H., LILES, M. R., LOIACONO, K. A., LYNCH, B. A., MACNEIL, I. A. & MINOR, C. 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied and environmental microbiology,* 66**,** 2541-2547.

RUBIN, B. E., SANDERS, J. G., HAMPTON-MARCELL, J., OWENS, S. M., GILBERT, J. A. & MOREAU, C. S. 2014. DNA extraction protocols cause differences in 16S

rRNA amplicon sequencing efficiency but not in community profile composition or structure. *Microbiologyopen,* 3**,** 910-921.
RUIBAL, M. P., PEAKALL, R., SMITH, L. M. & LINDE, C. C. 2013. Phylogenetic and microsatellite markers for Tulasnella (Tulasnellaceae) mycorrhizal fungi associated with Australian orchids. *Applications in plant sciences,* 1**,** 1200394.
SAITOU, N. & NEI, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution,* 4**,** 406-425.
SALTER, S. J., COX, M. J., TUREK, E. M., CALUS, S. T., COOKSON, W. O., MOFFATT, M. F., TURNER, P., PARKHILL, J., LOMAN, N. J. & WALKER, A. W. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC biology,* 12**,** 87.
SCHLOSS, P. D., WESTCOTT, S. L., RYABIN, T., HALL, J. R., HARTMANN, M., HOLLISTER, E. B., LESNIEWSKI, R. A., OAKLEY, B. B., PARKS, D. H. & ROBINSON, C. J. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology,* 75**,** 7537-7541.
SCHMIDT, P.-A., BÁLINT, M., GRESHAKE, B., BANDOW, C., RÖMBKE, J. & SCHMITT, I. 2013. Illumina metabarcoding of a soil fungal community. *Soil Biology and Biochemistry,* 65**,** 128-132.
SCHMIEDER, R., LIM, Y. W., ROHWER, F. & EDWARDS, R. 2010. TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. *BMC bioinformatics,* 11**,** 341.
SCHOCH, C. L., SEIFERT, K. A., HUHNDORF, S., ROBERT, V., SPOUGE, J. L., LEVESQUE, C. A., CHEN, W., BOLCHACOVA, E., VOIGT, K. & CROUS, P. W. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences,* 109**,** 6241-6246.
SEDEEK, K. E., QI, W., SCHAUER, M. A., GUPTA, A. K., POVEDA, L., XU, S., LIU, Z.-J., GROSSNIKLAUS, U., SCHIESTL, F. P. & SCHLÜTER, P. M. 2013. Transcriptome and proteome data reveal candidate genes for pollinator attraction in sexually deceptive orchids. *PLoS One,* 8**,** e64621.
SELOSSE, M.-A. & ROY, M. 2009. Green plants that feed on fungi: facts and questions about mixotrophy. *Trends in plant science,* 14**,** 64-70.
SHARMA, J., ZETTLER, L. W. & VAN SAMBEEK, J. 2003. A Survey of Mycobionts of Federally Threatened.
SHARON, M., KUNINAGA, S., HYAKUMACHI, M., NAITO, S. & SNEH, B. 2008. Classification of Rhizoctonia spp. using rDNA-ITS sequence analysis supports the genetic basis of the classical anastomosis grouping. *Mycoscience,* 49**,** 93-114.
SHEFFERSON, R. P. 2009. The evolutionary ecology of vegetative dormancy in mature herbaceous perennial plants. *Journal of Ecology,* 97**,** 1000-1009.
SHEFFERSON, R. P., KULL, T., TALI, K. & KELLETT, K. M. 2012. Linking vegetative dormancy to fitness in two long-lived herbaceous perennials. *Ecosphere,* 3**,** 1-19.
SHIMURA, H., SADAMOTO, M., MATSUURA, M., KAWAHARA, T., NAITO, S. & KODA, Y. 2009. Characterization of mycorrhizal fungi isolated from the threatened Cypripedium macranthos in a northern island of Japan: two phylogenetically distinct fungi associated with the orchid. *Mycorrhiza,* 19**,** 525-534.
SHOKRALLA, S., SPALL, J. L., GIBSON, J. F. & HAJIBABAEI, M. 2012. Next-generation sequencing technologies for environmental DNA research. *Molecular ecology,* 21**,** 1794-1805.
SKIDMORE, E., KIM, S.-J., KUCHIMANCHI, S., SINGARAM, S., MERCHANT, N. & STANZIONE, D. iPlant atmosphere: a gateway to cloud infrastructure for the plant sciences. Proceedings of the 2011 ACM workshop on Gateway computing environments, 2011. ACM, 59-64.

SMIT, E., LEEFLANG, P., GLANDORF, B., VAN ELSAS, J. D. & WERNARS, K. 1999. Analysis of fungal diversity in the wheat rhizosphere by sequencing of cloned PCR-amplified genes encoding 18S rRNA and temperature gradient gel electrophoresis. *Applied and Environmental Microbiology,* 65**,** 2614-2621.

SMITH, J. E. & NAIR, R. 2005. The architecture of virtual machines. *Computer,* 38**,** 32-38.

SMITH, S. & READ, D. 2008. Colonization of roots and anatomy of arbuscular mycorrhiza. *Mycorrhizal Symbiosis. Academic Press: London***,** 42-90.

SMITH, S. E. & READ, D. J. 1997. *Mycorrhizal symbiosis,* San Diego, Academic Press, Inc.

SMITH, S. E. & READ, D. J. 2010. *Mycorrhizal symbiosis*, Academic press.

SNEH, B., JABAJI-HARE, S., NEATE, S. & DIJST, G. 2013. *Rhizoctonia species: taxonomy, molecular biology, ecology, pathology and disease control*, Springer Science & Business Media.

STEELE, W. K. Large scale seedling production of North American Cypripedium species. North American native terrestrial orchids propagation and production. North American Native Terrestrial Orchid Conference, Germantown, Maryland, 1996. 11-26.

STEWART, S. L. & KANE, M. E. 2006. Asymbiotic seed germination and in vitro seedling development of Habenaria macroceratitis (Orchidaceae), a rare Florida terrestrial orchid. *Plant Cell, Tissue and Organ Culture,* 86**,** 147-158.

STÖCKEL, M., MEYER, C. & GEBAUER, G. 2011. The degree of mycoheterotrophic carbon gain in green, variegated and vegetative albino individuals of Cephalanthera damasonium is related to leaf chlorophyll concentrations. *New Phytologist,* 189**,** 790-796.

STRICKLER, S. R., BOMBARELY, A. & MUELLER, L. A. 2012a. Designing a transcriptome next-generation sequencing project for a nonmodel plant species1. *American Journal of Botany,* 99**,** 257-266.

STRICKLER, S. R., BOMBARELY, A. & MUELLER, L. A. 2012b. Designing a transcriptome next-generation sequencing project for a nonmodel plant species1. *American Journal of Botany,* 99**,** 257-266.

SU, C.-L., CHAO, Y.-T., ALEX CHANG, Y.-C., CHEN, W.-C., CHEN, C.-Y., LEE, A.-Y., HWA, K. T. & SHIH, M.-C. 2011. De novo assembly of expressed transcripts and global analysis of the Phalaenopsis aphrodite transcriptome. *Plant and cell physiology,* 52**,** 1501-1514.

SUKUMAR, P., LEGUE, V., VAYSSIERES, A., MARTIN, F., TUSKAN, G. A. & KALLURI, U. C. 2013. Involvement of auxin pathways in modulating root architecture during beneficial plant–microorganism interactions. *Plant, cell & environment,* 36**,** 909-919.

SWARTS, N. D. & DIXON, K. W. 2009. Terrestrial orchid conservation in the age of extinction. *Annals of Botany,* 104**,** 543-556.

SWEET, H. C. & BOLTON, W. E. 1979. The surface decontamination of seeds to produce axenic seedlings. *American journal of botany***,** 692-698.

SZENDRÁK, E. 1997a. Asymbiotic in vitro seed germination, micropropagation and scanning electron microscopy of several temperate terrestrial orchids (Orchidaceae). *Theses, Dissertations, and Student Research in Agronomy and Horticulture***,** 1.

SZENDRÁK, E. 1997b. Asymbiotic in vitro seed germination, micropropagation and scanning electron microscopy of several temperate terrestrial orchids (Orchidaceae).

TALI, K. 2002. Dynamics of Orchis ustulata populations in Estonia. *Trends and fluctuations and underlying mechanisms in terrestrial orchid populations. Backhuys Publishers, Leiden***,** 33-42.

TAMURA, K., STECHER, G., PETERSON, D., FILIPSKI, A. & KUMAR, S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution,* 30**,** 2725-2729.

TAYLOR, D. & BRUNS, T. 1999. Community structure of ectomycorrhizal fungi in a Pinus muricata forest: minimal overlap between the mature forest and resistant propagule communities. *Molecular Ecology,* 8**,** 1837-1850.

TAYLOR, D., BRUNS, T., LEAKE, J. & READ, D. 2002. Mycorrhizal specificity and function in myco-heterotrophic plants. *Mycorrhizal ecology.* Springer.

TAYLOR, D. L. & BRUNS, T. D. 1997. Independent, specialized invasions of ectomycorrhizal mutualism by two nonphotosynthetic orchids. *Proceedings of the National Academy of Sciences,* 94**,** 4510-4515.

TAYLOR, D. L., BRUNS, T. D., SZARO, T. M. & HODGES, S. A. 2003. Divergence in mycorrhizal specialization within Hexalectris spicata (Orchidaceae), a nonphotosynthetic desert orchid. *American Journal of Botany,* 90**,** 1168-1179.

TAYLOR, D. L. & MCCORMICK, M. K. 2008. Internal transcribed spacer primers and sequences for improved characterization of basidiomycetous orchid mycorrhizas. *New Phytologist,* 177**,** 1020-1033.

TEWOLDEMEDHIN, Y. T. 2005. *Characterisation of Rhizoctonia in cropping systems in the Western Cape Province.* Stellenbosch: University of Stellenbosch.

TOJU, H., TANABE, A. S., YAMAMOTO, S. & SATO, H. 2012. High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. *PloS one,* 7**,** e40863.

TOMASO-PETERSON, M. & TREVATHAN, L. 2007. Characterization of Rhizoctonia-like fungi isolated from agronomic crops and turfgrasses in Mississippi. *Plant disease,* 91**,** 260-265.

TOVAR, P. A. 2015. *Spatial and temporal variation in mycorrhizal associations in a rare North American orchid.*

TRAPNELL, C., PACHTER, L. & SALZBERG, S. L. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics,* 25**,** 1105-1111.

TRAPNELL, C., ROBERTS, A., GOFF, L., PERTEA, G., KIM, D., KELLEY, D. R., PIMENTEL, H., SALZBERG, S. L., RINN, J. L. & PACHTER, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols,* 7**,** 562.

TREANGEN, T. J. & SALZBERG, S. L. 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Reviews Genetics,* 13**,** 36-46.

TSAI, C.-C., WU, K.-M., CHIANG, T.-Y., HUANG, C.-Y., CHOU, C.-H., LI, S.-J. & CHIANG, Y.-C. 2016. Comparative transcriptome analysis of Gastrodia elata (Orchidaceae) in response to fungus symbiosis to identify gastrodin biosynthesis-related genes. *BMC genomics,* 17**,** 212.

TSAI, W.-C., FU, C.-H., HSIAO, Y.-Y., HUANG, Y.-M., CHEN, L.-J., WANG, M., LIU, Z.-J. & CHEN, H.-H. 2013. OrchidBase 2.0: comprehensive collection of Orchidaceae floral transcriptomes. *Plant and Cell Physiology,* 54**,** e7-e7.

TURNBAUGH, P. J., LEY, R. E., HAMADY, M., FRASER-LIGGETT, C., KNIGHT, R. & GORDON, J. I. 2007. The human microbiome project: exploring the microbial part of ourselves in a changing world. *Nature,* 449**,** 804.

VAN WAES, J. 1984. *In vitro studie van de kiemingsfysiologie van Westeuropese orchideeën.* Rijksuniversiteit Gent. Faculteit van de Landbouwwetenschappen.

VEYRET, Y. 1969. 3. La structure des semences des Orchidaceae et leur aptitude à la germination in vitro en cultures pures.

VIJAY, N., POELSTRA, J. W., KÜNSTNER, A. & WOLF, J. B. 2013. Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. *Molecular ecology,* 22**,** 620-634.

VOYRON, S., ERCOLE, E., GHIGNONE, S., PEROTTO, S. & GIRLANDA, M. 2017. Fine-scale spatial distribution of orchid mycorrhizal fungi in the soil of host-rich grasslands. *New Phytologist,* 213**,** 1428-1439.

WAES, J. & DEBERGH, P. 1986a. Adaptation of the tetrazolium method for testing the seed viability, and scanning electron microscopy study of some Western European orchids. *Physiologia Plantarum,* 66**,** 435-442.

WAES, J. V. & DEBERGH, P. 1986b. In vitro germination of some Western European orchids. *Physiologia Plantarum,* 67**,** 253-261.

WALLER, J. M., LENNÉ, J. M. & WALLER, S. J. 2002. *Plant pathologist's pocketbook,* CABI.

WANG, J.-P. Z., LINDSAY, B. G., LEEBENS-MACK, J., CUI, L., WALL, K., MILLER, W. C. & DEPAMPHILIS, C. W. 2004. EST clustering error evaluation and correction. *Bioinformatics,* 20**,** 2973-2984.

WANG, R., XU, S., JIANG, Y., JIANG, J., LI, X., LIANG, L., HE, J., PENG, F. & XIA, B. 2013a. De novo sequence assembly and characterization of Lycoris aurea transcriptome using GS FLX titanium platform of 454 pyrosequencing. *PloS one,* 8**,** e60449.

WANG, Y., PAN, Y., LIU, Z., ZHU, X., ZHAI, L., XU, L., YU, R., GONG, Y. & LIU, L. 2013b. De novo transcriptome sequencing of radish (Raphanus sativus L.) and analysis of major genes involved in glucosinolate metabolism. *BMC genomics,* 14**,** 836.

WANG, Z., GERSTEIN, M. & SNYDER, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics,* 10**,** 57.

WARCUP, J. 1950. The soil-plate method for isolation of fungi from soil. *Nature,* 166**,** 117-118.

WARCUP, J. & TALBOT, P. 1967. Perfect states of Rhizoctonias associated with orchids. *New Phytologist,* 66**,** 631-641.

WARD, J. A., PONNALA, L. & WEBER, C. A. 2012. Strategies for transcriptome analysis in nonmodel plants. *American Journal of Botany,* 99**,** 267-276.

WATERMAN, R. J., BIDARTONDO, M. I., STOFBERG, J., COMBS, J. K., GEBAUER, G., SAVOLAINEN, V., BARRACLOUGH, T. G. & PAUW, A. 2011. The effects of above-and belowground mutualisms on orchid speciation and coexistence. *The American Naturalist,* 177**,** E54-E68.

WATKINSON, J. I. 2002. *Characterization of two genes, trehalose-6-phosphate synthase/phosphatase and nucleotide binding protein, shown to be differentially regulated in roots of Cypripedium parviflorum var. pubescens grown with a mycorrhizal fungus Thanatephorus pennatus.*

WAUD, M., WIEGAND, T., BRYS, R., LIEVENS, B. & JACQUEMYN, H. 2016. Nonrandom seedling establishment corresponds with distance-dependent decline in mycorrhizal abundance in two terrestrial orchids. *New Phytologist,* 211**,** 255-264.

WEATHERHEAD, M., ZEE, S. & BARRETTO, G. 1986. Some observations on the early stages of development of Eulophia yushuiana. Mem. Hong Kong Nat. *Hist. Soc,* 17**,** 85-90.

WEISS, M., SELOSSE, M.-A., REXER, K.-H., URBAN, A. & OBERWINKLER, F. 2004. Sebacinales: a hitherto overlooked cosm of heterobasidiomycetes with a broad mycorrhizal potential. *Mycological research,* 108**,** 1003-1010.

WELLS, T. & COX, R. 1989. HW Pritchard. Modern Methods in Orchid Conservation: The Role of Physiology, Ecology and Management. In. Predicting the probability of the bee orchid (Ophrys apifera) flowering or remaining vegetative from the size and number of leaves. Cambridge University Press. Cambridge.

WELLS, T. & COX, R. 1991. *Demographic and biological studies on Ophrys apifera: some results from a 10 year study*.

WHITE, T. J., BRUNS, T., LEE, S. & TAYLOR, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and applications,* 18**,** 315-322.

WHITLOCK, R., HIPPERSON, H., MANNARELLI, M. & BURKE, T. 2008. A high-throughput protocol for extracting high-purity genomic DNA from plants and animals. *Molecular Ecology Resources,* 8**,** 736-741.

WHITTAKER, R. H. 1972. Evolution and measurement of species diversity. *Taxon***,** 213-251.

WOLF, J. B. 2013. Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Molecular ecology resources,* 13**,** 559-572.

XIAO, M., ZHANG, Y., CHEN, X., LEE, E.-J., BARBER, C. J., CHAKRABARTY, R., DESGAGNÉ-PENIX, I., HASLAM, T. M., KIM, Y.-B. & LIU, E. 2013. Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *Journal of biotechnology,* 166**,** 122-134.

YAM, T. W. & ARDITTI, J. 2009. History of orchid propagation: a mirror of the history of biotechnology. *Plant Biotechnology Reports,* 3**,** 1-56.

YAMATO, M., YAGAME, T., SUZUKI, A. & IWASE, K. 2005. Isolation and identification of mycorrhizal fungi associating with an achlorophyllous plant, Epipogium roseum (Orchidaceae). *Mycoscience,* 46**,** 73-77.

YANAGAWA, T., NAGAI, M., OGINO, T. & MAEGUCHI, R. 1995. Application of disinfectants to orchid seeds, plantlets and media as a means to prevent in vitro contamination. *Lindleyana,* 10**,** 33-36.

ZAMAN, A., GOGOI, M., BORCHETIA, S., KALITA, M. C., YADAV, R. & BANDYOPADHYAY, T. 2016. Comparative assessment of different protocols for isolation of total RNA from various organs of the tea plant (Camellia sinensis).

ZELMER, C. D. & CURRAH, R. 1995. Ceratorhiza pernacatena and Epulorhiza calendulina spp. nov.: mycorrhizal fungi of terrestrial orchids. *Canadian Journal of Botany,* 73**,** 1981-1985.

ZELMER, C. D., CUTHBERTSON, L. & CURRAH, R. S. 1996. Fungi associated with terrestrial orchid mycorrhizas, seeds and protocorms. *Mycoscience,* 37**,** 439-448.

ZHANG, G.-Q., LIU, K.-W., LI, Z., LOHAUS, R., HSIAO, Y.-Y., NIU, S.-C., WANG, J.-Y., LIN, Y.-C., XU, Q. & CHEN, L.-J. 2017. The Apostasia genome and the evolution of orchids. *Nature,* 549**,** 379.

ZHANG, Q., BLAYLOCK, L. A. & HARRISON, M. J. 2010. Two Medicago truncatula half-ABC transporters are essential for arbuscule development in arbuscular mycorrhizal symbiosis. *The Plant Cell,* 22**,** 1483-1497.

ZHAO, X., ZHANG, J., CHEN, C., YANG, J., ZHU, H., LIU, M. & LV, F. 2014. Deep sequencing–based comparative transcriptional profiles of Cymbidium hybridum roots in response to mycorrhizal and non-mycorrhizal beneficial fungi. *BMC genomics,* 15**,** 747.

ZHOU, X., JIANG, X., YANG, C., MA, B., LEI, C., XU, C., ZHANG, A., YANG, X., XIONG, Q. & ZHANG, P. 2016. Cecal microbiota of Tibetan Chickens from five geographic regions were determined by 16S rRNA sequencing. *MicrobiologyOpen,* 5**,** 753-762.

ZHU, G., YU, Z., GUI, Y. & LIU, Z. 2008. A novel technique for isolating orchid mycorrhizal fungi. *Fungal Divers,* 33**,** 123-137.

ZIELIŃSKA, S., RADKOWSKI, P., BLENDOWSKA, A., LUDWIG-GAŁĘZOWSKA, A., ŁOŚ, J. M. & ŁOŚ, M. 2017. The choice of the DNA extraction method may influence the outcome of the soil microbial community structure analysis. *MicrobiologyOpen,* 6.

ZNANIECKA, J., KROLICKA, A., SIDWA-GORYCKA, M., RYBCZYNSKI, J., SZLACHETKO, D. L. & LOJKOWSKA, E. 2005. Asymbiotic germination,

seedling development and plantlet propagation of Encyclia aff. Oncidioides-an endangered orchid. *Acta Societatis Botanicorum Poloniae*, 74.

http://www.metoffice.gov.uk
https://www.metoffice.gov.uk/public/weather/climate/gctb66ydw.

http://www.catalogueoflife.org/col/details/database/id/12

https://www.metoffice.gov.uk/public/weather/climate/gctb66ydw

www.megasoftware.net

http://scata.mykopat.slu.se/)

http://qiime.org

http://qiime.org/scripts/alpha_diversity.html)

www.hardyorchidsociety.org.uk/HOS%201012/Cultivation.html

http://www.ncbi.nlm.nih.gov/genbank/).

http://biom-format.org).

Appendix http://datacat.liverpool.ac.uk/id/eprint/479