



**Uniwersytet
Gdański**

Michał May

*Dziedzina nauk ścisłych i przyrodniczych
Dyscyplina nauki biologiczne*

**PLASTYCZNOŚĆ STRATEGII ODŻYWIANIA
STORCZYKÓW MIKSOTROFICZNYCH JAKO
STADIUM POŚREDNIEGO W EWOLUCJI
MYKOHETEROTROFII**

**PLASTICITY OF NUTRITION STRATEGIES IN MIXOTROPHIC
ORCHIDS AS AN INTERMEDIATE STAGE IN EVOLUTION OF
MYCOHETEROTROPHY**

Rozprawa doktorska
wykonana
w Katedrze Taksonomii Roślin i Ochrony Przyrody /
/ Pracowni Symbioz Roślinnych
pod kierunkiem
prof. Marc-André Selosse
dr. Marcin Jąkowski

Gdańsk 2022

Table of contents

Table of contents	1
List of figures.....	4
List of tables.....	4
Abbreviation list	5
1 Abstract / Streszczenie	6
2 Preface.....	10
2.1 Background and motivation - personal and scientific - of presented dissertation	10
2.2 Scope of this dissertation	11
2.3 Acknowledgements	13
3 Introduction.....	14
3.1 Plants exhibit multiple carbon acquisition strategies	14
3.2 Orchidaceae family as a model for plant nutrition study	15
3.3 Autotrophy – a strategy for well-lit habitats.....	16
3.4 Mycoheterotrophy - an alternative carbon source for low light habitats.....	17
3.5 Mycorrhiza	17
3.5.1 Ectomycorrhiza	19
3.5.2 Endomycorrhiza.....	19
3.5.3 Orchid mycorrhiza	19
3.5.4 Mycoheterotrophy in adulthood.....	20
3.5.5 Mixotrophy	21
3.5.6 Mycobiont selection and specificity	22
4 Research questions and aims of this study	24
4.1 General goals.....	24
4.2 Specific goals	25
5 Publications and research articles	27
5.1 Article 1 The complete chloroplast genome sequence of <i>Dactylorhiza majalis</i> (Rchb.) P.F. Hunt et Summerh. (Orchidaceae).....	28
5.2 Article 2 The complete chloroplast genome sequence of <i>Platanthera chlorantha</i> (Orchidaceae).....	37
5.3 Article 3 Thirteen new plastid genomes from mixotrophic and autotrophic species provide insights into heterotrophy evolution in <i>Neottieae</i> orchids.....	44

0. Table of contents

5.4	Article 4 Three-year pot culture of <i>Epipactis helleborine</i> reveals autotrophic survival, without mycorrhizal networks, in a mixotrophic species	66
5.5	Article 5 The Genomic Impact of Mycoheterotrophy in Orchids	87
6	Methodology	110
6.1	DNA sequencing	110
6.2	RNA sequencing.....	110
6.3	Stable isotope analyses.....	111
6.4	In silico data processing	111
6.5	Plant tissue sampling procedure.....	112
7	Discussion.....	113
7.1	Framework of presented work	113
7.2	Plastomes of <i>P. chlorantha</i> and <i>D. majalis</i> present a solid genetic image of autotrophy	114
7.3	Description of 13 plastidial genomes of AT and MX Neottieae explains the possibility of elasticity in nutrition	115
7.4	<i>E. helleborine</i> without access to mycorrhizal network further proves high plasticity between AT and MX	118
7.5	MH species emerge by expression reprogramming and function losses, rather than metabolic innovations	119
7.6	Effective <i>de novo</i> assembly of plant transcriptomes requires tailored approach and refined tools – and can be improved upon.....	122
8	Practices for reliable hybrid <i>de novo</i> assembly of plant transcriptomes	124
9	Final discussion.....	152
9.1	Obstacles and opportunities in research process	152
9.2	Extending scope of research and further research perspectives.....	155
9.3	Conclusion.....	157
10	Supplementary materials	159
10.1	List of species involved in presented work	159
10.2	Photosynthesis – extended description.....	160
10.3	Non-photosynthetic functionality of chloroplasts	164
10.4	Practices for reliable hybrid <i>de novo</i> assembly – supplementary materials...	165

0. Table of contents

10.5	Methodology – extended description	184
10.5.1	Genomics	184
10.5.2	Three generations of DNA sequencing	185
10.5.2.A	Illumina SBS	185
10.5.2.B	Roche 454.....	186
10.5.2.C	Ion Torrent.....	186
10.5.2.D	SMRT (PacBio).....	187
10.5.2.E	ONT Nanopore.....	187
10.6	Transcriptomics	189
10.6.1	RNA sequencing	189
10.6.2	RNA-seq experiment design.....	189
10.6.3	Choice of samples and experimental groups	189
10.6.4	rRNA depletion	190
10.6.5	Sequencing library construction	190
10.6.6	Sequencing platform	191
10.6.7	Raw read processing	191
10.6.8	Transcriptome assembly	192
10.6.9	Transcriptome quality evaluation	193
10.6.10	Annotation and downstream application.....	193
10.6.11	Expression quantification and differential gene expression.....	194
10.7	Plastome sequencing	195
10.7.1	DNA extraction for plastome sequencing.....	195
10.7.2	Plastidial DNA enrichment	196
10.7.3	PCR amplification of plastomes	196
10.7.4	Direct plastome sequencing	197
10.7.5	Plastome assembly	197
10.8	Fungal identification by ITS barcoding.....	199
10.9	Phylogenetic inference	201
10.10	Isotopic analysis methods.....	203
10.10.1	Stable isotope natural abundance analysis	203
10.10.2	Isotope ratio mass spectrometry.....	205
11	Bibliography	206

List of figures

- Fig. 1** Phylogenetic tree of mycoheterotrophic lineages of Orchidaceae. Marked with red bars are loss of photosynthesis events, and mycoheterotrophic lineages names in red font
- Fig. 2** Structure of a circular plastidial genome on the example of first sequenced plastome, of tobacco plant. Kelvin Ma, after Wakasugi et al., 1998; CC BY-SA 3.0
- Fig. 3** A diagram illustrating the process of plastome reduction during loss of photosynthetic capability – after Graham et al. 2017
- Fig. 4** Field sampling team at work under the protective tent
- Fig. 5** Application of 3D printing for production of custom low cost laboratory equipment
- Fig. 6** Ribosomal RNA gene cluster structure
- Fig. 7** Commonly used fungal ITS primers and their annealing sites
- Fig. 8.1** Hybrid assembly mapping stats plot
- Fig. 8.2** Comparison between IDPdenovo output assemblies and their corresponding pre-assemblies

List of tables

- Tab. 1** Summary of particular types of mycorrhizae (after van der Heijden et al., 2015)
- Tab. 2** List of Orchidaceae described in this work, their phylogenetic belonging and nutrition strategy
- Tab. 3** Illumina sequencing platforms parameters, according to manufacturer's specifications
- Tab. 4** ONT sequencing platforms parameters, according to manufacturer's specifications
- Tab. 8.1.** SeqStat based general statistics from consecutive steps of Nanopore long read correction
- Tab. 8.2.** Mapping statistics of long reads from each step of correction.
- Tab. 8.3.** Basic statistics of the final hybrid assemblies.
- Tab. 8.4.** Basic statistics of *de novo* pre-assemblies based on Illumina short reads. Overall highest scores are underlined.
- Tab. 8.5.** Comparison of hybrid assembly mapping statistics.
- Tab. 8.6.** Timing and memory usage records.

Abbreviation list

(In order of occurrence)

AT	autotrophy	PEP	plastid encoded polymerase
MH	mycoheterotrophy	NEP	nuclear encoded polymerase
MX	mixotrophy	A-T	adenine-thymine
DNA	deoxyribonucleic acid	G-C	guanine-cytosine
RNA	ribonucleic acid	NGS	next generation sequencing
h _v	light energy	RNAseq	RNA sequencing
NADPH	nicotinamide adenine dinucleotide phosphate	PCR	polymerase chain reaction
ATP	adenosine triphosphate	GB	gigabytes
LSC	large single copy region	cDNA	complementary DNA
SSC	small single copy region	rRNA	ribosomal RNA
cpDNA	chloroplast DNA	mRNA	messenger RNA
EcM	ectomycorrhiza	DIY	do-it-yourself
AM	arbuscular mycorrhiza	FDM	fused deposition modeling
OM	orchid mycorrhiza	PAR	photosynthetic active radiation
PVPP	polyvinyl polypyrrolidone	LHC	light harvesting complex
kbp	kilobasepair (1000 basepairs)	RuBP	ribulose 1,5-bisphosphate
bp	basepair	ROS	reactive oxygen species
LR	long reads (Nanopore)	SBS	sequencing by synthesis
SR	short reads (Illumina)	DEG	differentially expressed gene
LN2	liquid nitrogen	scRNAseq	single cell RNAseq
IR	inverse repeats	RIN	RNA integrity number
NADH	nicotinamide adenine dinucleotide	ITS	internal transcribed spacer
NNA	<i>Neottia nidus-avis</i>	RFLP	restrictive fragment length polymorphism
EA	<i>Epipogium aphyllum</i>	IRMS	isotope ratio mass spectrometry

1 Abstract / Streszczenie

Members of Orchidaceae family exhibit three diverse nutrition strategies: autotrophy (AT) based on photosynthesis, mycoheterotrophy (MH), in which sustenance is obtained from a mycorrhizal partner – a soil fungus, and mixotrophy (MX), which involves both photosynthetic activity and reversed mycorrhizal carbon flow simultaneously. All orchid species demonstrate germination strategy that relies on fungal carbon and are therefore mycoheterotrophic in their juvenile state. After developing photosynthetic organs most of the species transition to AT and revert the direction of carbon flow to the functioning expected of common mycorrhizal relationship. Some species, however, do not undergo the switch and remain MH in adulthood as an evolutionary adaptation to low light conditions. The transition from AT to MH occurred independently multiple times in evolutionary history in multiple separate lineages of Orchidaceae. MX nutrition can be considered an intermediate state between those two strategies and is an evolutionary pre-requisite to transition into MH. Multiple MX species exhibit a high adaptation potential to revert to AT nutrition in conditions that favor photosynthesis over dependency on fungal carbon. The presented thesis encloses several studies, involved in a consistent sequence of cause-and-effect research questions and hypotheses that emerged from gathered, analyzed and discussed results. Sequencing of plastidial genomes of two AT orchid species, *P. chlorantha* and *D. majalis*, done to set a baseline template for a genetic makeup of an autotroph, revealed that both species possess intact plastomes and carry functional photosynthetic machinery. Further description of 13 new plastomes sequenced from orchid species, 9 of which were previously described as mixotrophic and 4 as autotrophic, revealed 10 cases of intact plastomes and 3 cases of reductions in plastome compositions that do not affect their photosynthetic capabilities. No correlation between degree of dependency on fungal carbon and gene loss was detected. MX plants retain genetic capability of reverting to AT nutrition, which is further observed on mixotrophic species *E. helleborine* grown in a greenhouse. The absence of trees, that contribute photosynthetic carbon to a common mycorrhizal partner and act as a primary carbon source in mycorrhizal network facilitates reverting to autotrophic nutrition, confirmed by isotopic content analysis. The analysis of MH orchid transcriptomes indicates far reaching degradation in their plastidial genomes but detects no genes unique to mycoheterotrophs; and a comparison of expression profiles between their above- and below ground organs suggests that physiology of MH plants is a result of expression reprogramming rather than novel genetic constructs. The assembly of transcriptomes for

1. Abstract / Streszczenie

non-model species is challenging, especially among plants – without fully sequenced genome to act as a reference for guided assembly, a less accurate *de novo* assembly approach is required. In an attempt to improve its quality, a novel approach combining two sequencing strategies was evaluated. This hybrid assembly technique involves employing short reads from Illumina sequencing and long reads from MinION platform to overcome drawbacks of both technologies and deliver a transcriptome of higher quality. In addition to the exploratory study of plasticity in plant nutrition strategies, this work delivered a unified, optimized toolset for future investigation of trophic modes in Orchidaceae.

Przedstawiciele rodziny Orchidaceae (storzycowate) wykazują trzy odmienne strategie odżywiania: opartą na fotosyntezie **autotrofię** (AT), **mykoheterotrofię** (MH), w przypadku której całość substancji odżywczych pozyskiwana jest od partnera mykoryzowego – grzyba zasiedlającego glebę, oraz **miksotrofię** (MX), która łączy w sobie jednocześnie wykorzystanie aktywności fotosyntetycznej i odwrócony przepływ mykoryzowy węgla. Wszystkie gatunki storczykowatych cechują się strategią kiełkowania uzależnioną od pozyskiwania węgla od partnera mykoryzowego, są więc mykoheterotrofami w stadium młodocianym. Większość gatunków po wytworzeniu organów fotosyntetycznych przechodzi na odżywianie AT i przywraca typowy dla mykoryzy kierunek przepływu węgla organicznego. Niektóre gatunki pozostają jednak MH w dorosłości, jako ewolucyjne przystosowanie do zasiedlania nisz ekologicznych o niskiej dostępności światła. Przejście z AT do MH zachodziło wielokrotnie, w niezależnych zdarzeniach ewolucyjnych, w oddzielnych liniach ewolucyjnych storczykowatych. Odżywianie MX można uznać za stadium pośrednie pomiędzy tymi strategiami. Stanowi ono ewolucyjny punkt wyjściowy do przejścia na ściśle odżywianie MH. Wiele gatunków MX wykazuje wysoki potencjał adaptacyjny do powrotu do odżywiania AT w warunkach promujących fotosyntezę względem uzależnienia od węgla pochodzenia grzybowego. Sekwencjonowanie genomów plastydowych dwóch autotroficznych gatunków storczyków, *P. chlorantha* i *D. majalis*, w celu wyznaczenia wzorcowego stanu plastomu rośliny samożywej, wykazało, że oba gatunki posiadają nienaruszony, kompletny plastom i w pełni funkcjonalną maszynę fotosyntetyczną. Opis dalszych 13 plastomów zsekwencjonowanych dla gatunków storczykowatych, z których 9 opisano przedtem w literaturze jako miksotroficzne a 4 jako autotroficzne, wykazał 10 przypadków kompletnych plastomów i 3 przypadki zajęcia redukcji składu genomu plastydowego, które jednak nie upośledzają funkcjonalności fotosyntetycznej tych gatunków. Nie odnotowano korelacji pomiędzy stopniem zależności od węgla pochodzenia grzybowego a degradacją genomów plastydowych. Rośliny MX zachowały genetyczne podstawy do przywrócenia ściśle AT strategii odżywiania – zjawiska, które zaobserwowano w przypadku miksotroficznego storczyka z gatunku *E. helleborine* uprawianego w szklarni. Brak obecności drzew, które stanowią główne źródło węgla fotosyntetycznego w sieciach mykoryzowych, sprzyjał przywróceniu odżywiania AT, co potwierdziła analiza składu izotopowego. Analiza transkryptomów storczyków MH wykazała daleko posuniętą degradację ich genomów plastydowych, ale nie wykryła obecności genów unikalnych dla mykoheterotrofów. Porównanie profili ekspresji pomiędzy częściami nadziemnymi i podziemnymi sugeruje, że fizjologia roślin MH wynika nie

1. Abstract / Streszczenie

z innowacji genetycznej, lecz z reprogramowania ekspresji istniejących genów. Składanie transkryptomów dla niemodelowych gatunków roślin jest utrudnione, ze względu na brak dostępności pełnych sekwencji genomowych, które mogłyby posłużyć do kierowanego procesu składania odczytów z sekwencjonowania RNA. W takiej sytuacji konieczne jest zastosowanie mniej dokładnego podejścia *de novo*. W celu poprawy jego jakości przeprowadzono testy strategii opartej o zastosowanie dwóch odmiennych platform sekwencjonowania. Podejście hybrydowe używa krótkich odczytów sekwencji z platformy Illumina i długich odczytów generowanych przez platformę MinION w celu przezwyciężenia ograniczeń technicznych obu tych metod i wygenerowania transkryptomu wysokiej jakości bez wsparcia genomu referencyjnego. Dodatkowo, na podstawie doświadczeń z badań przeprowadzonych nad plastycznością strategii odżywiania roślin, udało się skonstruować zoptymalizowany i ujednolicony zestaw narzędzi i procedur, które wykorzystane zostaną w dalszych badaniach.

2 Preface

2.1 Background and motivation - personal and scientific - of presented dissertation

Through the course of my academic education, I was involved with multiple diverse fields of biological sciences – beginning from protein biochemistry, through microbiology, molecular biology and evolutionary biology, that together led me to my current research interest. My first “serious” laboratory work revolved around *Arthrospira*. Being a cyanobacterium, a primitive photosynthesizing organism and a descendant of evolutionary pioneers of photosynthesis, it sparked my fascination with complex evolutionary mechanisms that stood behind development of modern-day higher plant autotrophic nutrition. Consecutively, the subject of my MSc thesis research, concerning the evolutionary arms race between plants and their microbial pathogens, led to my interest in the tricky and rarely unambiguous nature of coexistence and interaction between plants and microorganisms. However, the subject of interactions between plant and bacterial domains, and its molecular background, is not new and has been relatively thoroughly investigated and explained. Therefore, I was captivated by delving into a much more mysterious relationship that plants create with another kingdom of eukaryotes - fungi.

Despite the mycorrhizal relationship between plants and fungi being extremely common and prevalent in nature (about 90% of known terrestrial plants have been reported to engage in mycorrhizas), many of its molecular mechanisms still have not been sufficiently described to provide a clear explanation. Naturally, this topic poses a significant research challenge and has much left to be explored, investigated, and theorized upon. Furthermore, a broad spectrum of rapidly developing, novel and challenging methods needs to be employed to effectively investigate this subject. Commonly applied research methodic and commercial solutions are often not suited to be readily applied to our subject and goals, and the knowledge base of methods optimized for application in this field is scarce and often narrowly specified. Therefore, new variations of available techniques are required to effectively investigate this challenging pairing of organisms, especially in *in situ* conditions. Existing methods need redesigning and adjustment to be effective in this field. As an exciting perspective, both from a point of view of scientific community and a single molecular biology enthusiast, various existing methods can be applied as means to a novel goal they

were not originally designed for, fine-tuned for specific conditions and materials, or combined to outperform conventional approaches.

To build a connection between the topic of evolution of plant nutrition and mycorrhizal interactions, an outstanding family of plants is brought in. Orchidaceae is a massive and diverse group of organisms, exhibiting some unusual properties that make them especially valuable for the research of evolutionary origin of various plant nutrition modes. The group exhibits impressive biodiversity and extremely broad area of occurrence, thus making it a relevant model for diversified research applications. Their uniqueness lies in their disparate nutrition modes and their common unusual germination strategy - we can witness all stages of evolutionary history of plant trophism within a single, closely related group of organisms, and in multiple cases, even within a single species. Despite this variability, all these species exhibit a common trait of heterotrophic nutrition in early seedling development, and most of them undergo a complete or partial switch to autotrophy. Therefore, deeper investigation of their metabolism and genetic makeup will bring invaluable insight into our understanding of evolutionary history of plant nutrition.

Orchid species of large economic and horticultural significance are relatively well-described. However, more “niche” families still lack descriptions and genetic reference needed for bioinformatic and phylogenetic analysis. Obtaining such resources is crucial for construction of detailed descriptions of nutritional gradient within the family. Many terrestrial species of Orchidaceae that occur in Poland are barely known to the general public and their biology is often misunderstood, which makes them particularly attractive as a subject in both investigatory research and nature conservation.

2.2 Scope of this dissertation

Presented thesis is a selection of excerpts from wide research on orchid transcriptome, metabolome, plastome and phylogeny. It consists of a combination of reference data gathering, practical research methods optimisation and exploratory and investigatory work that ties together into a comprehensive study of plasticity of plant nutrition strategies. The first part of conducted work concentrated on data accumulation – large genetic datasets were constructed, later to be subjected to extensive comparative analyses. Only then an attempt to derive meaningful conclusions, and finally deliver reasonably well-grounded statements on the researched subject could be accomplished. Due to the multi-threaded nature of this investigation, numerous new research questions and opportunities emerged during the entire

2. Preface

study and were thoroughly addressed and investigated. In addition, in progress of this work a novel, optimized, and unified methodical toolbox was designed that addresses the needs and difficulties that emerged in progress of our investigation. The toolset consists of in situ sampling procedures, laboratory analysis and sequencing workflow and bioinformatic pipeline for efficient and reliable data processing. These novel methods were applied to refine the outcome of presented research and will be used to extend their scope in further investigative and accumulative research.

The introduction chapter (3) of this thesis contains contextual explanation of core concepts within this work. It establishes diverse nutrition strategies in Plantae kingdom, delivers description of Orchidaceae family in context of its application as a model organism in this research, and describes the phenomenon of mycorrhiza, its types and influence it has on plant trophic strategies.

Chapter 4 outlines and summarizes main research questions and goals that were undertaken in presented work.

Chapter 5 consists of five research articles, that were published in the course of my studies and are the main outcome of presented research.

Chapter 6 provides a brief list of methods used in this work, which are explained in-depth inside the articles and in supplementary materials section.

The discussion chapter (7) provides my interpretation of findings from presented publications and ties them together in a common context of this dissertation.

Chapter 8 consists of yet unpublished results, formatted in a layout of a scientific article. It addresses the issues encountered during the assembly of transcriptomes performed in chapter 5.5, suggests a novel solution and evaluates its effectivity.

Chapter 9 delivers a final discussion of presented results and summarizes findings of this work. It presents the tangible side of efforts and problem-solving approach that enabled me to formulate the conclusions of this dissertation and presents the perspectives for further research that can expand and further build upon this work. The thesis is concluded with a final summary.

Chapter 10 contains supplementary information that explain the extended context and technicalities behind presented work.

2.3 Acknowledgements

I would like to express my gratitude to my PhD supervisor, prof. Marc-André Selosse, for his guidance, mentoring, and for infecting me with his passion for investigation and popularizing knowledge. Without his influence, presenting this work would have been utterly impossible.

I would like to thank dr. Julita Minasiewicz for her guidance and overseeing my work, both in laboratory work and in exploring the data.

Dr. Marcin Jąkowski, for introducing me to the world of modern bioinformatics, for his patience in aiding me in understanding and performing the *in silico* research, and for his invaluable insights that allowed me to write and redact this dissertation.

Etienne Delannoy and personnel of INRA, for introducing me to the next-generation sequencing technologies in both theory and practice.

University of Gdańsk, for providing me with opportunity to conduct my research and for assembling an outstanding group of researchers to share knowledge with.

National Science Centre, for funding the MAESTRO 7 grant that sponsored my research.

University of Cologne for inviting me into a cooperation to describe a unique research opportunity, and for teaching me the ropes of RNA handling I could later expand upon.

3 Introduction

3.1 Plants exhibit multiple carbon acquisition strategies

All life forms known on Earth exhibit a requirement for several basic nutritional elements that are necessary for their survival and development. The main element considered a basis of all life – carbon (C) – is a backbone of organic matter building nearly all the biologically crucial compounds, such as proteins, lipids, carbohydrates, or deoxyribonucleic acids. Through the past 4 billion years evolutionary processes have led to the development of a number of diverse strategies responsible for obtaining carbon, both organic and inorganic. However, of them all, photosynthesis-based autotrophy is considered the most significant for Earth's entire biosphere (Whitmarsh and Govindjee, 1999).

Within the plant kingdom, the evolutionary adaptation processes led to formation of a specific gradient of nutritional strategies. On one end, the most common strategy employed by plants – autotrophy (AT) – allows them to utilize solar energy as a mean of binding atmospheric or water-dissolved carbon dioxide into biologically useful compounds. On the opposite end, an evolutionary adaptative regression resulted in secondary formation of a particular type of heterotrophy – mycoheterotrophy (MH), in which plants derive their entire organic carbon pool from a fungal partner bound in a mycorrhizal relationship. This unusual strategy emerged as a culmination of a complex avalanche of evolutionary processes, and coexists with multiple transient, intermediate stages that were necessary for its development, commonly referred to as mixotrophy (MX) (Selosse and Roy, 2009; Buchanan et al., 2015).

3.2 Orchidaceae family as a model for plant nutrition study

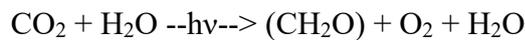
Orchidaceae are one of the largest families in the kingdom of Plantae and the largest family of flowering plants (Chase et al., 2015; Fay and Chase, 2009). It counts almost 28 000 described species divided into 5 main subfamilies: Apostasioideae, Cyripedioideae, Orchidoideae, Epidendroideae and Vanilloideae, and approximately 800 genera (The Plant List 1.1, 2020 – www.theplantlist.org). The family exhibits a great diversity of adaptive specializations, variability in morphology and habitat selection. Though occurring on all continents excluding Antarctica, the representatives of this family are the most numerous as epiphytes inhabiting tropical rainforests. In temperate climates, most Orchidaceae are terrestrial plants, inhabiting all sorts of habitats – from dry meadows, through damp marshes, to dense forest understory. Their wide range of occurrence and diversity make them a relevant model for general plant studies and evolutionary research.

Orchidaceae exhibit an interesting set of alterations to the most common plant nutrition model, and the entire family employs a common unusual reproductive strategy. Orchidaceae propagate by dust seeds – minuscule seeds devoid of nutrition reserves and surrounded by hard cover, produced in great quantities (even up to several millions seeds per seed pod), yet unable to germinate independently. For successful germination, formation of mycorrhizal relationship with a soil fungus is required. The fungus associates with an imbibed seed, aids in perforating its seed cover and redirects its own carbohydrates to the germinating plant, sustaining it until its own photosynthetic organs are developed. Therefore, all orchid species are mycoheterotrophic at one stage of their life. After maturation, most orchid species transition to independent carbohydrate synthesis and become autotrophic. Multiple species, however, remain fully mycoheterotrophic in adulthood or exhibit various degrees of mixotrophy in their nutrition, as an adaptation to certain ecological niches, usually characterized by reduced availability of sunlight. Example species involved in this work, with their respective nutrition strategies, are listed in Table 3 (chapter 10.1).

Dependent on symbiosis at germination, and tightly cooperating – or benefiting from – fungi in adult forms, the richness of diverse strategies between closely related species creates an opportunity for a comprehensive study of evolutionary emergence, adaptation process and plasticity of nutrition on evolutionary path to mycoheterotrophy. Therefore, Orchidaceae are a valuable and well-suited model for a study of modification of plant physiology to increase suitability for inhabiting dark environments (Lallemand et al., 2019c), in both physiological and evolutionary aspects.

3.3 Autotrophy – a strategy for well-lit habitats

Autotrophic mode of nutrition is an adaptation to deriving sustenance by fixation of inorganic carbon into complex organic compounds through a physicochemical process powered by energy harvested from sunlight - photosynthesis. In plants, algae and certain photosynthetic bacteria, photosynthetic activity occurs through absorption of atmospheric or water-dissolved carbon dioxide (CO₂) to produce carbohydrates, and results in release of oxygen as a byproduct – therefore it is classified as oxygenic photosynthesis. Predominant in complex organisms, it can be summarized with a reaction:



The solar energy is utilized for stripping electrons from molecules of water and biosynthesis of reduced nicotinamide adenine dinucleotide phosphate (NADPH) and adenosine triphosphate (ATP), which are considered volatile, basic energy carrier inside living cells. This reaction can be considered a mean of conversion of light into chemical energy, which – in the case of plants – is later used for synthesis of the main product of the process, namely glucose. Though seemingly simple, photosynthesis consists of several distinct processes and requires nearly 30 specialized proteins building an intricate network of membrane-bound structures (Buchanan et al., 2015; Whitmarsh and Govindjee, 1999).

Complex photosynthetic organisms have compartmentalized their photosynthetic activity inside dedicated organelle – chloroplasts. Contemporary chloroplast is a construct specific to and occurring only in *Plantae* cells. As a chlorophyll-containing variant of highly versatile group of organelles – plastids (gr. *plastikos* – molded), it is characterized by its basic functionality including photosynthesis, biosynthesis and storage of starch, synthesis of lipids and amino acids, and involvement in multiple metabolic pathways directly unrelated to photosynthetic activity (Hölzl and Dörmann, 2019; Lancien et al., 2007, Rottet et al., 2015, Buchanan et al., 2015; Maréchal, 2018). A system of internal thylakoid membranes houses the 3 functional photosystems and acts as a site of photosynthetic activity. Chloroplasts exhibit a sustained semi-autonomy from the 'host' cell they are contained within – they retain their own, highly conserved genome (plastome), plastid-specific genetic machinery and transcriptional apparatus, and are capable of synthesizing their own proteins and of multiplication by division (Chen et al., 2018). However, most of their proteins require complementation with imported nuclear-encoded proteins to form functional structures. The

3. Introduction

number of chloroplasts present in a single plant cell is variable and changes in different developmental stages (Buchanan et al., 2015; Maréchal, 2018). Most angiosperms exhibit inheritance of plastids in maternal lineage, while gymnosperms usually inherit them from paternal pollen (Salinas, 2018). The chloroplast genomes of vascular plants exhibit high conservation of their coding regions, structures and general linear layout of their content. The contents of the plastidial chromosome can be divided into 3 functional categories: protein-coding sequences, introns and intergenic spacers. Following the example of *Nicotiana tabacum* as a model plastidial genome (Shinozaki et al., 1986), 43% of LSC and SSC are non-coding sequences; 15 present introns comprise 10.6% of single copy cpDNA, and 92 spacers make up 32%. Most land plants exhibit highly similar content of plastomes. Most commonly occurring genes divided by their function include photosynthesis-related genes and various genetic machinery-related genes. However, genetic content of plastidial DNA has been observed to be a subject of alterations, usually paired with transition to non-photosynthetic nutrition strategies (Graham et al., 2017). The degree of modification varies between stages of evolutionary transition to heterotrophy and differs between lineages. The principles of this differentiation still require detailed description and investigation (Barrett et al., 2014; Barrett et al., 2019; Jacquemyn and Merckx, 2019; Wickett et al., 2011).

3.4 Mycoheterotrophy - an alternative carbon source for low light habitats

Developing alternative – backup – strategies of obtaining carbon may create a possibility to abandon photosynthesis altogether in favor of more cost-efficient strategies, or strategies that are better adjusted to current environmental conditions (light availability). The development of mycorrhiza - a nutrient trade relationship formed between plant roots and fungal hyphae – was such an opportunity for approximately 50 lineages from different clades of terrestrial plants (Hynson et al., 2013; Lallemand et al., 2019c). In a specific variant of mycorrhiza, plants can alter their nutrition from autotrophy (AT) to mycoheterotrophy (MH), in which they manage to obtain their entire organic carbon from a fungal partner (Hynson et al., 2013; Selosse et al., 2016).

3.5 Mycorrhiza

Mycorrhiza is a predominantly symbiotic or mutualistic relationship between plant roots and fungal hyphae (from Greek: *mykes* – fungus, *rhiza* – root) (Frank, 2005). The relationship occurs extremely common in nature – with exception of a few families

3. Introduction

(Brassicaceae, Caryophyllaceae, Proteaceae, Cyperaceae) it is utilized by over 80% of living plant species and is considered one of the most important adaptations of terrestrial plants (Wang and Qiu, 2006). The first examples of mycorrhiza are believed to have occurred in Ordovician and enabled plants to effectively colonize dry land (Feijen et al., 2018; Pirozynski and Malloch, 1975; Selosse et al., 2015).

In the most basic model, the interaction resembles a trade, where the fungal partner provides water with dissolved mineral nutrition, mainly nitrogen (N), phosphorus (P) and potassium (K), while phototroph exchanges it for its products of photosynthesis – organic carbon in a form of carbohydrates. Additionally, both mycorrhizal partners derive benefits in form of protection against biotic (pathogens) and abiotic (drought, salinity, toxic compounds) stress factors. Such relationship is, therefore, mutually beneficial and symbiotic.

Two major mycorrhizal types have been described based on the spatial structure of co-habiting symbiotes: **ectomycorrhiza (EcM)** and **endomycorrhiza**, further divided into arbuscular, ericoid, and orchid mycorrhiza, based on morphology and taxonomy of partners.

Tab. 1 Summary of particular types of mycorrhizae (after van der Heijden et al., 2015)

Type	Plant groups ¹	No. of species ¹	Fungal taxa	No. of fungal species	Characteristic structures
Ectomycorrhiza	Pinaceae, angiosperms (mostly temperate shrubs and trees) ⁵ , some liverworts ⁶	6000	Basidiomycota, Ascomycota, few Zygomycota	20 000 ⁷	Hyphal sheath, Hartig net
Arbuscular mycorrhiza	most herbs, grasses, trees; multiple hornworts ² and liverworts ³	200 000	Glomeromycota	300-1600 ⁴	Arbuscules, vesicles
Orchid mycorrhiza	Orchidaceae	20 – 35 000	Basidiomycota	25 000	Pelotons
Ericoid mycorrhiza	Ericaceae, some liverworts ⁶	3900	mainly Ascomycota, some Basidiomycota	>150 ⁸	-
Nonmycorrhizal plants ⁹	Brassicaceae, Crassulaceae, Orobanchaceae, Proteaceae	51 500	-	-	-

¹ after Brundrett, 2009

² after Desirò et al., 2013

³ after Ligrone et al., 2007

⁴ after Kivlin et al., 2011; Öpik et al., 2013

⁵ after Brundrett, 2009

⁶ after Read et al., 2000

⁷ after Rinaldi et al., 2008; Tedersoo et al., 2010

⁸ after Walker et al., 2011

⁹ after Lambers and Teste, 2013

3.5.1 Ectomycorrhiza

Ectomycorrhiza occurs in 3% of seed plant species, but still can be considered common among shrubs and trees of temperate forests and is widespread in almost all climates, from tropical to boreal forests. As the name indicates, in this type of mycorrhiza, fungal hyphae do not penetrate through the cell walls of root cells, instead creating an intercellular net (Hartig net) and forming characteristic sheath (mantle) around the root and spreading into surrounding soil (Smith and Read, 2002). This type of mycorrhiza is usually formed between fungal phylae Basidiomycota, Ascomycota, rarely Zygomycota, and plants (mostly) from families Pinaceae, Betulaceae, Fagaceae, Salicaceae, Rosaceae, Myrtaceae and Dipterocarpaceae (Tedersoo et al., 2010).

3.5.2 Endomycorrhiza

As opposed to ectomycorrhiza, endomycorrhiza – indicating its internalized character in host roots – is developed by both intercellular growth of fungal network, and formation of hyphal structures inside the cytoplasm of cortical cells of a root. Over 90% of land plants have been observed to form endomycorrhizal relationships with partner soil fungi (Smith and Read, 2002).

3.5.3 Orchid mycorrhiza

Orchid mycorrhiza, a type of endomycorrhiza specific to Orchidaceae family, involves formation of specialized intracellular structures of hyphae inside plant cells. These formations, resembling a ball of yarn, are called pelotons (Dearnaley et al., 2012). They act as a dynamic interface of nutrient exchange, both by active and passive membrane transport and by total digestion of entire structure. Due to fungus-dependent nature of dust seed germination, juvenile seedlings (protocorms) are mycoheterotrophic (MH). The relationship involves at least three parties – the fungus is incapable of carbohydrate production on its own, and receives it from another, green mycorrhizal partner before passing it to the orchid, its final recipient (Bidartondo et al., 2004; Hynson et al., 2013; Selosse et al., 2006). Such mycorrhizal network formed by many photosynthetic plants interacting with multiple soil fungi, jokingly called a "wood wide web" allows for nutrition and mineral transfer and signaling between many cohabitating organisms. After formation of leaves a switch occurs and nutrient trade continues along the 'default' model.

3.5.4 Mycoheterotrophy in adulthood

Multiple orchid species never develop photosynthesis and remain mycoheterotrophic into adulthood. They exhibit achlorophyllous morphology, and their leaves are rudimentary. Such nutrition strategy emerged in at least 50 known separate evolutionary events in 17 separate lineages (Jacquemyn and Merckx, 2019; Barrett et al., 2019; Ogura-Tsujita et al., 2021) (Fig. 1). Loss of photosynthesis is an adaptation to low light conditions and allows to reduce energetic cost while colonizing dark forest understory.

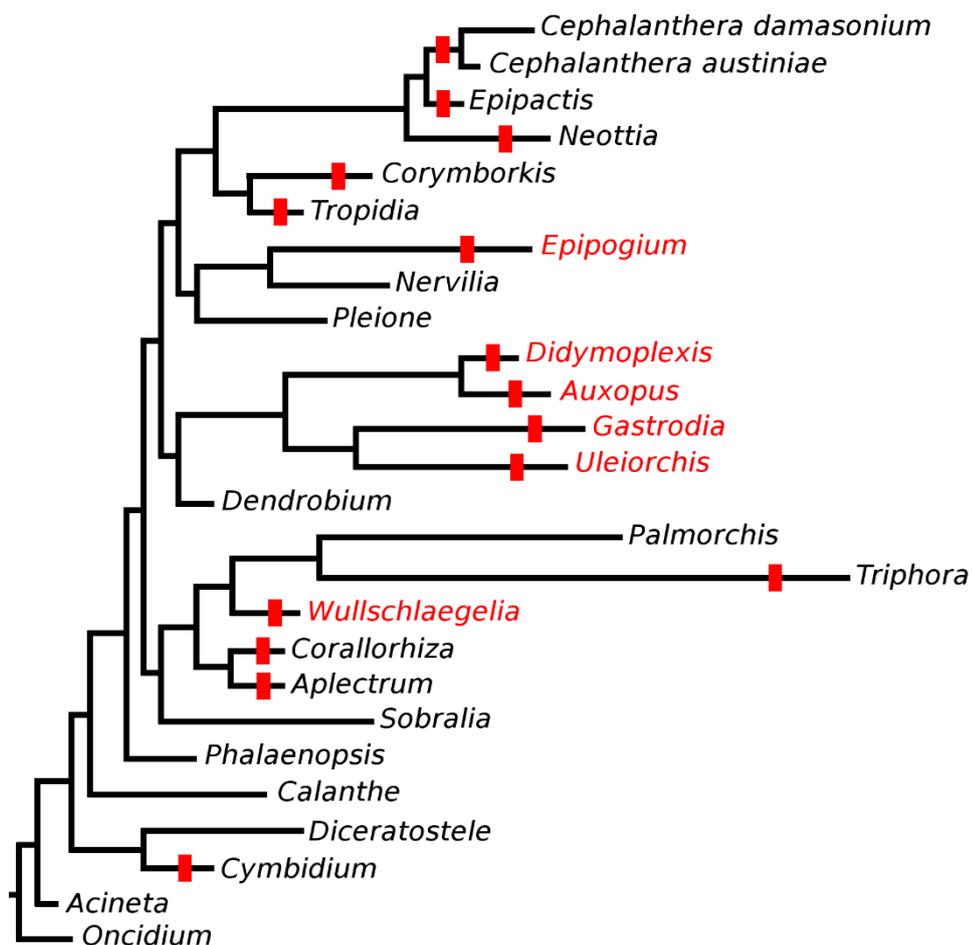


Fig. 1 Phylogenetic tree of mycoheterotrophic lineages of Orchidaceae. Marked with red bars are loss-of-photosynthesis events, and mycoheterotrophic lineages names in red font.

3.5.5 Mixotrophy

The development of strict mycoheterotrophy, including a full loss of photosynthetic capabilities, is a conclusion of multiple coinciding evolutionary events. Multiple lineages of Orchidaceae exhibit sets of traits that are pre-requisites to a successful complete trophism shift. Their current trophic mode can be considered an intermediate stage in evolution of fully MH lineage – **mixotrophy (MX)**, in this case also referred to as **partial mycoheterotrophy**. A similar stage has been observed in multiple plant families during transition to heterotrophy, most notably parasitism (Press and Graves, 1995; Schulze et al., 1991; Těšitel et al., 2010). In such state, the plant is still capable of photosynthesis, but obtains part of its carbon from a secondary nutrition source – in case of Orchidaceae, their mycorrhizal partner. This strategy exhibits a high potential for regulation and reversal (Bidartondo et al., 2004). The phenomenon of mixotrophy has received more researchers' attention in planktonic algal taxa (Kamjunke and Tittel, 2009), but still remains relatively poorly known in terrestrial plants (Moore, 2013; Schmidt et al., 2013). Remarkably, these two groups do not share the same origin of MX nutrition – in algae, mixotrophy developed as an extension of pure heterotrophy (uptake of dissolved organic carbon or phagotrophy), by gaining additional photosynthetic functionality. Vascular plants, as photosynthetic plastid-bearing organisms, have somewhat "reverted" to it, by secondary evolution of previously lost heterotrophy, enabling them to draw part of their nutrition from host plants sap (parasitism) or by fungal association (partial mycoheterotrophy/mixotrophy) (Leake, 1994). The genetic basis of this evolutionary process is predominantly based on gene loss events, mainly in photosynthesis-related areas (Delannoy et al. 2011; Braukmann et al. 2013; Wicke et al. 2013). Various mixotrophs and heterotrophs exhibit different degrees of plastome and nuclear genome reduction, which yet appear to be largely convergent in different lineages (Feng et al., 2016; Graham et al., 2017). Several models of gene reduction were proposed, but the exact mechanism governing this process still remains largely unexplained and requires deeper investigation (Graham et al., 2017). The capability of reversal of this process and its effect on plasticity of nutrition is also a subject of discussion.

Whereas fully MH plants are easily identified by their lack of photosynthesis, distinction of mixotrophs is less obvious. Several species have been identified by observation of viable albinotic specimens *in natura*. Such observations were made mostly for species from Neottieae tribe, such as *Epipactis* (Salmia, 1986; Selosse et al., 2004) (e.g., *helleborine*, *microphylla*) and *Cephalanthera* (Julou et al., 2005; Stöckel et al., 2011) (e.g., *damasonium*,

3. Introduction

longifolia) genus, which are closely related to known strict mycoheterotrophs. However, more distant clades also exhibit such tendencies (e.g., *Platanthera* (Yagame et al., 2012) and *Cymbidium* (Motomura et al., 2010)). To unambiguously quantify plant's nutrition mode, analysis of isotopic content is required. Such experiments have proven that not only is mixotrophy prevalent in multiple clades of Orchidaceae in association with multiple fungal families, but also that the strategy is subject to fine-tuning on individual level in response to habitat lighting conditions (Gonneau et al., 2014; Preiss et al., 2010).

Mycorrhizal partnership of mixotrophic orchids is also not restricted to a single clade of fungal partners – not only (as previously suspected) EcM species, but also numerous rhizoctonias are exploited in such relationships (Gebauer et al., 2016; Selosse and Martos, 2014). The exact dynamics and mechanisms involved in the process still require deeper investigation and will be discussed in this work.

3.5.6 Mycobiont selection and specificity

Until recently, Orchidaceae were commonly divided into ‘ectomycorrhiza-associated’ species, exhibiting partial or full mycoheterotrophy, and ‘rhizoctonia-associated’ species, considered autotrophic as adults (Bidartondo et al., 2004; Schiebold et al., 2018). However, recent evidence (Gebauer et al., 2016; Schiebold et al., 2018; Schweiger et al., 2018) questions autotrophy of rhizoctonia-associated orchid species by delivering evidence of multiple sources of nutrition being utilized. Multiple species from both groups were also revealed to form associations with both rhizoctonias and EcM fungi (Jacquemyn et al., 2017, 2015, 2014). These findings undermine the previously agreed upon division between EcM- and rhizoctonia-associated species.

The multi-species nature of mycorrhizas seems to be common in both autotrophic and mycoheterotrophic plant species. Selecting multiple partners may be beneficial by complementation of multiple needs fulfilled by different partners, or by increasing the chance for association with the most beneficial partner (Batstone et al., 2018). Particular partners can be also exploited in different life stages, which is supported by observations of gradual accumulation of additional mycobiont species over plant's lifetime (Bidartondo and Read, 2008). Various partners also exhibit different adaptation to stress conditions, therefore multi-sided mycorrhiza allows choice of the most optimal fungus for a particular stress factor (McCormick et al., 2006).

3. Introduction

The turnover of mycobionts appears not to be caused by simple availability factor. *Neottia ovata* was observed to alter the proportions of associated fungal partners, while the surrounding soil communities did not reflect these changes (Oja et al., 2015), indicating on it being a controlled shift.

4 Research questions and aims of this study

4.1 General goals

We embarked on the presented research as a multi-threaded exploratory work, guided by the main goal of providing a deeper understanding of the diverse range of trophic types exhibited by the Orchidaceae family and establish it in an evolutionary framework. The primary approach aims to perform a comprehensive survey of evolutionary and genetic aspects of plant adaptation to low-light conditions. A wide set of modern research methods was applied to an orchid model, aiming to increase the understanding of photosynthetic plasticity in plant evolution and of the phenomenon of mixotrophy itself.

During the progress of investigation, as new data were gathered and analyzed, a number of new questions were formulated to thoroughly exploit the research opportunities that arose, and multiple new approaches had to be introduced.

In order to investigate the genome reduction process on the way from autotrophy towards mycoheterotrophy, we faced a need to construct a clear baseline “definition” of an autotroph as a point of origin and mixotroph as an intermediate stage between the ends of the spectrum, in a genomic and transcriptomic aspect. From this point, we realized a need for construction of strict and unified guidelines for further exploration and comparative description of AT, MX and MH species in data accumulation process. The existing methods used in this field turned out to be unsatisfying and insufficient to produce high-throughput datasets of high quality, especially due to the challenging character of research material (scarcity, difficulties in processing and analysis). Therefore, we set out to refine and optimize current procedures to build a modern toolset for performing large scale comparisons between orchid species from the lineages of interest.

Thanks to application of refined methods, we could finally obtain data that would allow us to perform further comparative analyses and track the environmental impact on functioning of MX species – concentrating on light conditions and their surroundings, to investigate the plasticity of their nutrition.

4.2 Specific goals

Article 1. (Chapter 5.1) The complete chloroplast genome sequence of *Platanthera chlorantha* (Orchidaceae): The aim of this research was to provide a complete plastome sequence of *P. chlorantha*, thus producing a valuable resource for phylogenetic and evolutionary studies. Such resources are especially needed for Orchidoideae tribe, as only 14 plastomes of its members are currently available, and only one from genus *Platanthera*. In a wider scope of this work, constructed plastome provides a necessary reference dataset describing a strictly autotrophic orchid, grounding further analysis for mixo- and autotrophic species.

Article 2. (Chapter 5.2) The complete chloroplast genome sequence of *Dactylorhiza majalis* (Rchb.) P.F. Hunt et Summerh. (Orchidaceae): The aim of this study was to provide a complete plastidial DNA sequence for *D. majalis* – the first fully sequenced plastome for its genus *Dactylorhiza*.

Article 3. (Chapter 5.3) Three-year pot culture of *Epipactis helleborine* reveals autotrophic survival, without mycorrhizal networks, in a mixotrophic species: The goal of this study was to investigate mycorrhizal and nutritional status of an unusual specimen of *E. helleborine*, usually described as a MX species, growing in pot culture without access to mycorrhizal network with photosynthetic carbon source. The study aims to shed new light on plasticity of mixotrophic nutrition and assess its potential for reversal to autotrophy.

Article 4. (Chapter 5.4) Thirteen new plastid genomes from mixotrophic and autotrophic species provide insights into heterotrophy evolution in *Neottieae* orchids: This study investigates selective pressure acting on photosynthesis-related gene loss in *Neottieae* tribe at various stages of evolutionary progression from autotrophy to mycoheterotrophy.

Article 5. (Chapter 5.5) The genomic impact of mycoheterotrophy in orchids: The goal of this study was to survey the transcriptomic profiles of *E. aphyllum* and *N. nidus-avis*, two MH orchids, and perform a comparison on organ level and between species to derive

4. Research questions and aims of this study

conclusions about gene loss and metabolic changes that plants undergo during their transition to MH nutrition.

Unpublished work (Chapter 8): Practices for reliable hybrid de novo assembly of plant transcriptomes: This work investigates multiple novel methods of transcriptomic assembly, aiming to combine two state-of-the-art sequencing platforms to overcome limitations of *de novo* assembly processes. A comparison of multiple tools and approaches is carried out to achieve the most optimal results on a well-described model species (*A. thaliana*) transcriptome assembly before extending it to non-model research subjects.

5 Publications and research articles

This chapter consists of 5 published and peer-reviewed research articles. Each article is prefaced by a collection of statements delivered by co-authors on their involvement in presented work.

5.1 Article 1

The complete chloroplast genome sequence of *Dactylorhiza majalis* (Rchb.) P.F. Hunt et Summerh. (Orchidaceae)

Michał May; Alžběta Novotná; Julita Minasiewicz; Marc-André Selosse; Marcin Jąkałski

MITOCHONDRIAL DNA PART B

2019, VOL. 4, NO. 2, 2821–2823

<https://doi.org/10.1080/23802359.2019.1660282>

in this work cited as (May et al., 2019)

My contribution in presented work involved:

- Assembling and annotation of plastidial genome
- Analysis and description of plastidial genome
- Manuscript preparation

Uniwersytet Gdański
Wydział Biologii
Ul. Wita Stwosza 59
80-308, Gdańsk

data

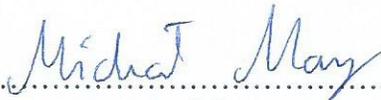
OŚWIADCZENIE AUTORA O UDZIALE W PUBLIKACJI

Tytuł publikacji: *"The complete chloroplast genome sequence of Dactylorhiza majalis (Rchb.) P.F. Hunt et Summerh. (Orchidaceae)"*

Imię autora: Michał May

Oświadczam, że w wymienionym wyżej artykule, opublikowanym w *Mitochondrial DNA Part B* 4:2, pages 2821-2823; doi.org/10.1080/23802359.2019.1660282 mój udział obejmował:

- Składanie i anotację genomu plastydowego
- Analiza i opis genomu
- Przygotowanie manuskryptu


.....
podpis autora

University of South Bohemia
Faculty of Science
Branišovská 31, 370 05
České Budějovice, Czech Republic

date 10/2/2022

AUTHOR'S CONTRIBUTION STATEMENT

Work title: "The complete chloroplast genome sequence of *Dactylorhiza majalis* (Rchb.) P.F. Hunt et Summerh. (Orchidaceae)"

Author: Alžběta Novotná

I declare that my contribution in article mentioned above and published in Mitochondrial DNA Part B 4:2, pages 2821-2823; doi.org/10.1080/23802359.2019.1660282 included:

- DNA isolation

Novotná Alžběta

Author's signature

Uniwersytet Gdański
Wydział Biologii
Ul. Wita Stwosza 59
80-308, Gdańsk

data M.03.2022

OŚWIADCZENIE AUTORA O UDZIALE W PUBLIKACJI

Tytuł publikacji: "The complete chloroplast genome sequence of *Dactylorhiza majalis* (Rchb.) P.F. Hunt et Summerh. (Orchidaceae)"

Imię autora: Julita Minasiewicz

Oświadczam, że w wymienionym wyżej artykule, opublikowanym w Mitochondrial DNA Part B 4:2, pages 2821-2823; doi.org/10.1080/23802359.2019.1660282 mój udział obejmował:

- Analizy statystyczne i filogenetyczne


.....
podpis autora

Institut de Systématique, Evolution, Biodiversité (ISYEB)
Muséum national d'Histoire naturelle
CNRS, Sorbonne Université, EPHE
Paris, France

date

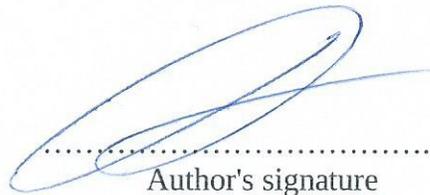
AUTHOR'S CONTRIBUTION STATEMENT

Work title: "The complete chloroplast genome sequence of *Dactylorhiza majalis* (Rchb.) P.F. Hunt et Summerh. (Orchidaceae)"

Author: Marc-André Selosse

I declare that my contribution in article mentioned above and published in Mitochondrial DNA Part B 4:2, pages 2821-2823; doi.org/10.1080/23802359.2019.1660282 included:

- Experimental concept
- Obtaining funding and resources
- Research supervision



.....
Author's signature

M-A. Selosse

The complete chloroplast genome sequence of *Dactylorhiza majalis* (Rchb.) P.F. Hunt et Summerh. (*Orchidaceae*)

Michał May^a , Alžběta Novotná^b , Julita Minasiewicz^a , Marc-Andre Selosse^{a,c}  and Marcin Jąkowski^a 

^aDepartment of Plant Taxonomy and Nature Conservation, Faculty of Biology, University of Gdańsk, Gdańsk, Poland; ^bFaculty of Science, University of South Bohemia, České Budějovice, Czech Republic; ^cInstitut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Paris, France

ABSTRACT

The complete chloroplast genome of *Dactylorhiza majalis* (Rchb.) P.F. Hunt et Summerh. (*Orchidaceae:Orchidoideae*) was assembled and characterized using next-generation sequencing data. The plastome (154,108 bp) possesses the typical circular structure consisting of a large single-copy region (LSC; 83,196 bp), a small single-copy region (SSC; 26,580 bp), and two copies of inverted repeats (17,752 bp each). Its overall GC content is 36.99% and the plastome encodes 134 genes. Reconstruction of phylogenetic relationships using complete plastome sequences of *Orchidaceae* representatives showed that *D. majalis* was nested within the *Orchidoideae* tribe *Orchideae*. The complete plastome comprises a valuable tool in elucidating taxonomic uncertainties within the genus *Dactylorhiza*.

ARTICLE HISTORY

Received 19 July 2019
Accepted 3 August 2019

KEYWORDS

Chloroplast genome;
Dactylorhiza majalis; orchid;
Orchidaceae; next-
generation sequencing

Dactylorhiza Necker ex Nevski is a temperate orchid genus known from its complex evolutionary relationships between species driven by high frequency of hybridization, introgression, and polyploidization. Reticulate evolution pattern along with relatively great morphological variability within a species is considered a major challenge in taxonomy of the genus (Hedrén 1996; Pillon et al. 2007, and references therein). *Dactylorhiza majalis* (Rchb.) P.F. Hunt et Summerh. can be found in western and central Europe, Baltic region, and northern Russia (Hultén and Fries 1986; Balao et al. 2016). It is an allotetraploid species belonging to a polyploid complex formed iteratively by crosses between *Dactylorhiza incarnata* s.l. and *Dactylorhiza maculata* s.l. with the last species being always maternal parent (Hedrén et al. 2008 and references therein). Complete, annotated plastidial genome, upon which new molecular markers can be described, would be a valuable tool in untangling evolutionary history within the genus. Chloroplast genomes provide researchers with data invaluable for resolving major phylogenetic relationships between orchid subfamilies (Givnish et al. 2015). Currently, only 15 complete chloroplast genomes are available within the subfamily *Orchidoideae* (Delannoy et al. 2011; Lin et al. 2015; Yu et al. 2015; Zhu et al. 2016; Roma et al. 2018; Lallemand et al. 2019; Oh et al. 2019). This makes this subfamily largely underrepresented among other orchids. Species from the *Dactylorhiza* genus were so far only subject to phylogenetic studies employing ITS, microsatellite loci, selected marker genes, or morphology, and results of these still often remain incongruent (Bateman et al. 2003; Shipunov et al. 2004; Balao et al. 2016; Jin et al. 2017).

Fresh leaves were collected from an individual growing in Psary, Poland (N50°22'07.4" E19°04'53.3"). Leaves dried in silica gel (voucher SG-13237, Herbarium of University of Gdansk, UGDA) were used for extraction of the total genomic DNA with Dneasy Plant Mini Kit (Qiagen, Hilden, Germany). Sequencing library was generated with Accel-NGS[®] 1S Plus DNA Library Kit (Swift Biosciences Inc., Ann Arbor, MI). Next-generation paired-end sequencing was performed with Illumina HiSeq 4000 (San Diego, CA). The obtained reads were used for genome assembly with the Geneious software version 10.2.4 (<https://www.geneious.com>) with medium-low sensitivity parameters and a subset of 25% of the reads, followed by mapping to the closest reference plastome (*Platanthera japonica*, NC_037440.1), and reassembly with medium sensitivity parameters to increase the assembly quality. Annotation was performed within Geneious as well as using GeSeq (Tillich et al. 2017), and manually corrected afterwards. Phylogenetic relationships of *D. majalis* with other orchids were inferred from maximum-likelihood analyses with RAXML-NG (Kozlov et al. 2019) using selected available complete orchid plastomes aligned with MAFFT (Katoh and Standley 2013).

The chloroplast DNA of *D. majalis* is 154,108 bp in length, presenting the overall GC content (the proportion of guanine and cytosine bases) of 36.99%. Consistent with other known orchid plastomes it is comprised two inverted repeats (IRa and IRb) with 26,580 bp in length, an 83,196 bp large single-copy region (LSC), and a 17,752 bp long small single-copy region (SSC). A total of 134 genes were annotated, of which 113 are unique. These are 4 rRNA genes, 30 tRNA genes, and

CONTACT Marcin Jąkowski  marcin.jakalski@biol.ug.edu.pl  Department of Plant Taxonomy and Nature Conservation, Faculty of Biology, University of Gdańsk, ul. Wita Stwosza 59, Gdańsk 80-308, Poland

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

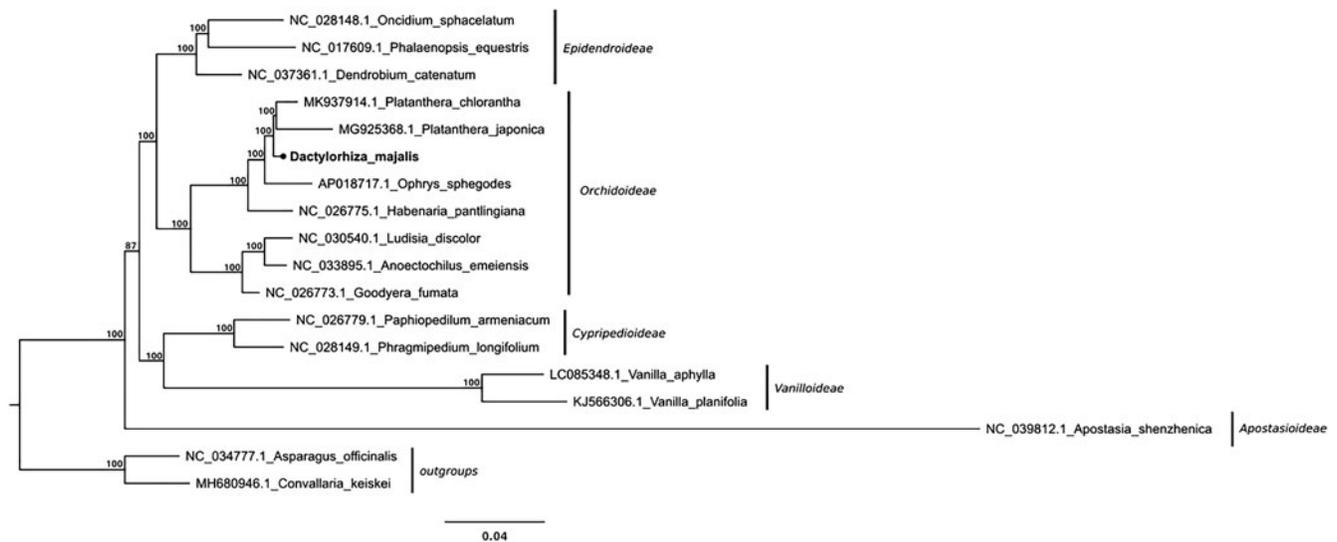


Figure 1. Phylogenetic relationships inferred from maximum-likelihood analyses with full length plastome sequences of *Orchidaceae* representatives, including the newly assembled *D. majalis* plastome. Node support values are derived from RAXML assessment with 1000 bootstraps replicates. Non-orchid monocots were used for tree rooting.

79 protein-coding genes. Twenty genes are duplicated in the IR region. Additionally, 12 protein-coding genes and 6 tRNA genes contain introns. The annotated sequence was deposited at GenBank with the accession number MK984209. Results of the phylogenetic relationships investigation between *D. majalis* and other members of the *Orchidaceae* showed its clustering together with representatives of the subtribe *Orchidinae* (Figure 1). The complete plastome sequence we provided here constitutes a valuable aid for addressing the taxonomic uncertainties within the genus *Dactylorhiza*, as well as analysing the genetic diversity of the *Orchidaceae* family.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The presented research was supported by the grant funded by Narodowe Centrum Nauki (National Science Center, Poland), grant No. [2015/18/A/NZ8/00149].

ORCID

Michał May <http://orcid.org/0000-0001-8027-1369>
 Alžběta Novotná <http://orcid.org/0000-0002-9835-0855>
 Julita Minasiewicz <http://orcid.org/0000-0002-0330-7011>
 Marc-Andre Selosse <http://orcid.org/0000-0003-3471-9067>
 Marcin Jąkowski <http://orcid.org/0000-0002-5481-9148>

References

Balao F, Tannhäuser M, Lorenzo MT, Hedrén M, Paun O. 2016. Genetic differentiation and admixture between sibling allopolyploids in the *Dactylorhiza majalis* complex. *Heredity*. 116:351–361.

Bateman RM, Hollingsworth PM, Preston J, Yi-Bo L, Pridgeon AM, Chase MW. 2003. Molecular phylogenetics and evolution of Orchidinae and selected Habenariinae (*Orchidaceae*). *Bot J Linn Soc*. 142:1–40.

Delannoy E, Fujii S, Colas Des Francs-Small C, Brundrett M, Small I. 2011. Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Mol Biol Evol*. 28:2077–2086.

Givnish TJ, Spalink D, Ames M, Lyon SP, Hunter SJ, Zuluaga A, Iles WJD, Clements MA, Arroyo MTK, Leebens-Mack J, et al. 2015. Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proc R Soc B Biol B*. 282:20151553.

Hedrén M. 1996. Genetic differentiation, polyploidization and hybridization in northern European *Dactylorhiza* (*Orchidaceae*): evidence from allozyme markers. *Plant Syst Evol*. 201:31–55.

Hedrén M, Nordström S, Ståhlberg D. 2008. Polyploid evolution and plastid DNA variation in the *Dactylorhiza incarnata/maculata* complex (*Orchidaceae*) in Scandinavia. *Mol Ecol*. 17:5075–5091.

Hultén E, Fries M. 1986. Atlas of north European vascular plants: north of the tropic of cancer. Germany: Koeltz Scientific Books.

Jin WT, Schuitman A, Chase MW, Li JW, Chung SW, Hsu TC, Jin XH. 2017. Phylogenetics of subtribe *Orchidinae* s.l. (*Orchidaceae*; *Orchidoideae*) based on seven markers (plastid matK, psaB, rbcL, trnL-F, trnH-psbA, and nuclear nrITS, Xdh): implications for generic delimitation. *BMC Plant Biol*. 17:222.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30:772–780.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*. btz305, doi:10.1093/bioinformatics/btz305

Lallemant F, May M, Ichnatowicz A, Jąkowski M. 2019. The complete chloroplast genome sequence of *Platanthera chlorantha* (*Orchidaceae*). *Mitochondrial DNA B*. 4:2683–2684.

Lin CS, Chen JJW, Huang YT, Chan MT, Daniell H, Chang WJ, Hsu CT, Liao DC, Wu FH, Lin SY, et al. 2015. The location and translocation of *ndh* genes of chloroplast origin in the *Orchidaceae* family. *Sci Rep*. 5:9040.

Oh SH, Suh HJ, Park J, Kim Y, Kim S. 2019. The complete chloroplast genome sequence of a morphotype of *Goodyera schlechtendalia* (*Orchidaceae*) with the column appendages. *Mitochondrial DNA B*. 4: 626–627.

Pillon Y, Fay MF, Hedrén M, Bateman RM, Devey DS, Shipunov AB, van der Bank M, Chase MW. 2007. Evolution and temporal diversification

- of western European polyploid species complexes in *Dactylorhiza* (*Orchidaceae*). *Taxon*. 56:1185–1208.
- Roma L, Cozzolino S, Schlüter PM, Scopece G, Cafasso D. 2018. The complete plastid genomes of *Ophrys iricolor* and *O. sphegodes* (*Orchidaceae*) and comparative analyses with other orchids. *PLoS One*. 13:e0204174.
- Shipunov AB, Fay MF, Pillon Y, Bateman RM, Chase MW. 2004. *Dactylorhiza* (*Orchidaceae*) in European Russia: combined molecular and morphological analysis. *Am J Bot*. 91:1419–1426.
- Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq – versatile and accurate annotation of organellar genomes. *Nucleic Acids Res*. 45:W6–W11.
- Yu C, Lian Q, Wu K, Yu S, Xie L, Wu Z. 2015. The complete chloroplast genome sequence of *Anoectochilus roxburghii*. *Mitochondrial DNA*. 27:1–2.
- Zhu S, Niu Z, Yan W, Xue Q, Ding X. 2016. The complete chloroplast genome sequence of *Anoectochilus emeiensis*. *Mitochondrial DNA A*. 27:3565–3566.

5.2 Article 2

The complete chloroplast genome sequence of *Platanthera chlorantha* (Orchidaceae)

Félix Lallemand; **Michał May**; Anna Ihnatowicz; Marcin Jąkałski

MITOCHONDRIAL DNA PART B

2019, VOL. 4, NO. 2, 2683–2684

<https://doi.org/10.1080/23802359.2019.1644551>

In this work cited as (Lallemand et al., 2019d)

My contribution in presented work involved:

- Assembling and annotation of plastidial genome
- Analysis and description of plastidial genome
- Manuscript preparation

Institut de Systématique, Evolution, Biodiversité (ISYEB)
Muséum national d'Histoire naturelle
CNRS, Sorbonne Université, EPHE
Paris, France

date

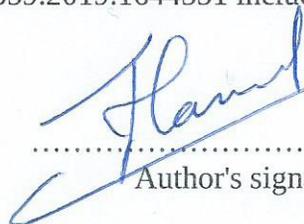
AUTHOR'S CONTRIBUTION STATEMENT

Work title: "The complete chloroplast genome sequence of *Platanthera chlorantha* (Orchidaceae)"

Author: Félix Lallemand

I declare that my contribution in article mentioned above and published in Mitochondrial DNA Part B 5:1, pages 63-64; doi.org/10.1080/23802359.2019.1644551 included:

- Laboratory analysis
- Phylogenetic analysis and statistics


.....
Author's signature

Uniwersytet Gdański
Wydział Biologii
Ul. Wita Stwosza 59
80-308, Gdańsk

data

OŚWIADCZENIE AUTORA O UDZIALE W PUBLIKACJI

Tytuł publikacji: *"The complete chloroplast genome sequence of Platanthera chlorantha (Orchidaceae)"*

Imię autora: Michał May

Oświadczam, że w wymienionym wyżej artykule, opublikowanym w Mitochondrial DNA Part B 5:1, pages 63-64; doi.org/10.1080/23802359.2019.1644551mój udział obejmował:

- Składanie i anotację genomu plastydowego
- Analiza i opis genomu
- Przygotowanie manuskryptu


.....
podpis autora

Międzyuczelniany Wydział Biotechnologii
MWB i GUMed
ul. Abrahama 58
80-307 Gdańsk

data

OŚWIADCZENIE AUTORA O UDZIALE W PUBLIKACJI

Tytuł publikacji: "The complete chloroplast genome sequence of *Platanthera chlorantha* (Orchidaceae)"

Imię autora: Anna Ihnatowicz

Oświadczam, że w wymienionym wyżej artykule, opublikowanym w Mitochondrial DNA Part B 5:1, pages 63-64; doi.org/10.1080/23802359.2019.1644551 mój udział obejmował:

- Optymalizacja procedury izolacji i preparacji materiału DNA

.....

.....
podpis autora

Uniwersytet Gdański
Wydział Biologii
Ul. Wita Stwosza 59
80-308, Gdańsk

data

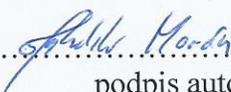
OŚWIADCZENIE AUTORA O UDZIALE W PUBLIKACJI

Tytuł publikacji: "*The complete chloroplast genome sequence of Platanthera chlorantha (Orchidaceae)*"

Imię i nazwisko autora: Marcin Jąkowski

Oświadczam, że w wymienionym wyżej artykule, opublikowanym w Mitochondrial DNA Part B 5:1, pages 63-64; doi.org/10.1080/23802359.2019.1644551 mój udział obejmował:

- analizę filogenetyczną
- redagowanie manuskryptu
- nadzór merytoryczny

.....

.....
podpis autora

The complete chloroplast genome sequence of *Platanthera chlorantha* (Orchidaceae)

Félix Lallemand^a, Michał May^b, Anna Ihnatowicz^c and Marcin Jąkowski^b 

^aInstitut de Systématique, Evolution, Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, Paris, France; ^bDepartment of Plant Taxonomy and Nature Conservation, Faculty of Biology, University of Gdańsk, Gdańsk, Poland; ^cIntercollegiate Faculty of Biotechnology, University of Gdansk and Medical University of Gdansk, Gdańsk, Poland

ABSTRACT

Here, we report the first complete chloroplast genome of *Platanthera chlorantha* (Orchidaceae: Orchidoideae). The circular genome with the length of 154,260 bp possesses the typical structure consisting of a large single copy region (LSC) of 83,279 bp and a small single copy region (SSC) of 17,759 bp, separated from each other by two copies of inverted repeats (IRs) of 26,611 bp. The plastome encodes 134 genes, of which 88 were protein-coding, eight encoded ribosomal RNA, and 38 transfer RNAs. The overall GC content was 36.74%. The plastome sequence provided here constitutes a valuable resource for analyzing genetic diversity of the *Orchidoideae* family.

ARTICLE HISTORY

Received 24 May 2019
Accepted 10 July 2019

KEYWORDS

Chloroplast genome;
Platanthera chlorantha;
orchids; Orchidoideae

Due to their extensive variability and biotic interactions, orchids present a compelling subject for evolutionary studies (Selosse 2014). Chloroplast genomes provide researchers with data invaluable for studying genetic history and phylogeny and thus were already used to resolve major phylogenetic relationships between orchid subfamilies (Givnish et al. 2015). In this study, we sequenced for the first time the plastome of *Platanthera chlorantha* (Custer) Rchb. This species inhabits mostly woods, meadows, and damp heaths and is widely spread through Europe, excluding eastern, northern, and south-western edges (Hultén and Fries 1986). It belongs to the *Orchidoideae* subfamily represented by almost 3800 species. With the main focus of researchers being put on the *Epidendroideae* tribes (142 of 175 orchid species with complete plastomes; NCBI GenBank, as of April 2019), currently only 14 complete chloroplast genomes are available within *Orchidoideae* tribes (Yu et al. 2015; Zhu et al. 2016; Roma et al. 2018; Oh et al. 2019) and only one of those belongs to the member of the same genus *Platanthera*, namely *P. japonica* (Dong et al. 2018).

Fresh plant leaves were collected from Kalina Lisiniec, Poland (50°21'43.7"N 20°09'37.4"E) and dried in silica-gel (voucher SG-13236, Herbarium of University of Gdansk, UGDA). Extraction of the total genomic DNA was performed using DNeasy Plant Mini Kit (Qiagen), followed by sequencing library generation with Accel-NGS[®] 1S Plus DNA Library Kit (Swift Biosciences Inc., USA) and paired-end sequencing on Illumina HiSeq 4000. Geneious (version 10.2.4, <https://www.geneious.com>), with the Geneious assembling algorithm (medium-low sensitivity parameters) and a subset of 25% of

the reads was used for plastome assembly. Annotations were obtained with Geneious annotations transfer procedure from *P. japonica* (NC_037440.1) with a 77% similarity threshold, additionally aided using GeSeq (Tillich et al. 2017) and followed by manual verification. Inferring of phylogenetic position of *P. chlorantha* was based on ML analysis performed with RAxML-NG (Kozlov et al. 2019) and multiple sequence alignment of full chloroplast sequences conducted with MAFFT (Katoh and Standley 2013).

The complete circular chloroplast genome of *P. chlorantha* (GenBank MK937914) is a 154,260 bp long sequence, with 36.74% GC content and a typical structure consisting of two inverted repeats (IRa and IRb) of 26,611 bp each, separated by a large single-copy region (LSC) of 83,279 bp and a small single-copy region (SSC) of 17,759 bp in length. The plastome contains 134 (113 unique) genes, of which eight are rRNA genes, 38 encode tRNAs, and 88 are protein-coding. One of the *ycf1* genes located at the IRb/SSC boundary is truncated but with complete open reading frame and one copy of the *rpl22* gene is a pseudogene. Introns were annotated for 12 protein-coding genes and six tRNA genes. Phylogenetic analyses (Figure 1) showed that *P. chlorantha* grouped together with *P. japonica* and other representatives of *Orchidoideae*. Despite floral similarities, it is phylogenetically distant from the genus *Habenaria* as already noted before (Bateman et al. 2003). In the genus *Platanthera*, some species display a trend to recover biomass from the symbiotic fungi colonizing their roots (i.e. a partial heterotrophy/mixotrophy, based on their mycorrhizal fungi; see Selosse and Roy 2009; Yagame et al. 2012; references therein). Our result show plastids with

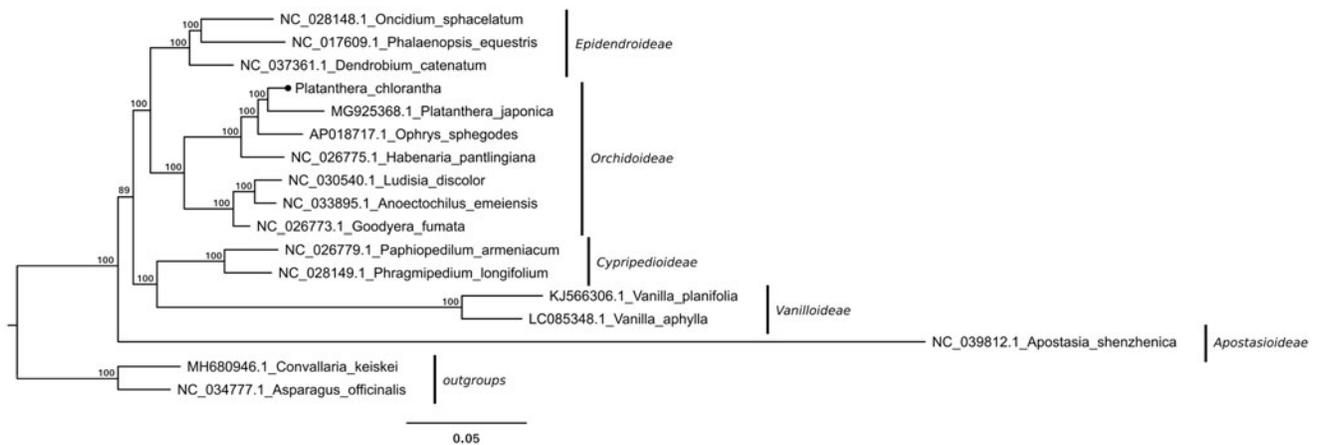


Figure 1. Maximum-likelihood-based phylogenetic tree of *Orchidaceae* representatives constructed using full length plastome sequences. The numbers on each node are RAxML bootstrap values (based on 1000 replicates). The tree is rooted with non-orchid monocot representatives.

apparently intact photosynthetic abilities in *P. chlorantha*, which is in agreement with indirect evidence for autotrophy published by Bidartondo et al. (2004) for this species. The plastome sequence we provided here constitutes a valuable aid for analyzing the genetic diversity of the *Orchidaceae* family.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The presented research was supported by the 2015/18/A/NZ8/00149 grant funded by National Science Center (Poland).

ORCID

Marcin Jąkowski  <http://orcid.org/0000-0002-5481-9148>

References

- Bateman RM, Hollingsworth PM, Preston J, Yi-Bo L, Pridgeon AM, Chase MW. 2003. Molecular phylogenetics and evolution of Orchidinae and selected Habenariinae (Orchidaceae). *Bot J Linn Soc.* 142:1–40.
- Bidartondo MI, Burghardt B, Gebauer G, Bruns TD, Read DJ. 2004. Changing partners in the dark: isotopic and molecular evidence of ectomycorrhizal liaisons between forest orchids and trees. *Proc R Soc London Ser B Biol Sci.* 271:1799–1806.
- Dong W-L, Wang R-N, Zhang N-Y, Fan W-B, Fang M-F, Li Z-H. 2018. Molecular evolution of chloroplast genomes of orchid species: insights into phylogenetic relationship and adaptive evolution. *IJMS.* 19:716.
- Givnish TJ, Spalink D, Ames M, Lyon SP, Hunter SJ, Zuluaga A, Iles WJD, Clements MA, Arroyo MTK, Leebens-Mack J, et al. 2015. Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proc R Soc B Biol Sci.* 282:20151553.
- Hultén E, Fries M. 1986. Atlas of North European vascular plants: north of the Tropic of Cancer. Königstein (Federal Republic of Germany): Koeltz Scientific Books.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30:772–780.
- Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics.* <https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btz305/5487384>
- Oh S-H, Suh HJ, Park J, Kim Y, Kim S. 2019. The complete chloroplast genome sequence of a morphotype of *Goodyera schlechtendaliana* (Orchidaceae) with the column appendages. *Mitochondrial DNA B.* 4: 626–627.
- Roma L, Cozzolino S, Schlüter PM, Scopece G, Cafasso D. 2018. The complete plastid genomes of *Ophrys iricolor* and *O. sphegodes* (Orchidaceae) and comparative analyses with other orchids. *PLoS One.* 13:e0204174.
- Selosse M-A. 2014. The latest news from biological interactions in orchids: in love, head to toe. *New Phytol.* 202:337–340.
- Selosse M-A, Roy M. 2009. Green plants that feed on fungi: facts and questions about mixotrophy. *Trends Plant Sci.* 14:64–70.
- Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45:W6–W11.
- Yagame T, Orihara T, Selosse M-A, Yamato M, Iwase K. 2012. Mixotrophy of *Platanthera minor*, an orchid associated with ectomycorrhiza-forming Ceratobasidiaceae fungi. *New Phytol.* 193:178–187.
- Yu C, Lian Q, Wu K, Yu S, Xie L, Wu Z. 2015. The complete chloroplast genome sequence of *Anoectochilus roxburghii*. *Mitochondrial DNA.* 27: 1–2.
- Zhu S, Niu Z, Yan W, Xue Q, Ding X. 2016. The complete chloroplast genome sequence of *Anoectochilus emeiensis*. *Mitochondrial DNA A.* 27: 3565–3566.

5.3 Article 3

Thirteen new plastid genomes from mixotrophic and autotrophic species provide insights into heterotrophy evolution in *Neottieae* orchids

Félix Lallemand; Maria Logacheva; Isabelle Le Clainche; Aurélie Bérard; Ekaterina Zheleznaia; **Michał May**; Marcin Jąkowski; Étienne Delannoy; Marie-Christine Le Paslier; Marc-André Selosse

GENOME BIOLOGY AND EVOLUTION

Volume 11, Issue 9, September 2019, Pages 2457–2467

<https://doi.org/10.1093/gbe/evz170>

In this work cited as (Lallemand et al., 2019b)

My contribution in presented work involved:

- Research material collection
- Investigation process
- Data processing and analysis

Institut de Systématique, Evolution, Biodiversité (ISYEB)
Muséum national d'Histoire naturelle
CNRS, Sorbonne Université, EPHE
Paris, France

date

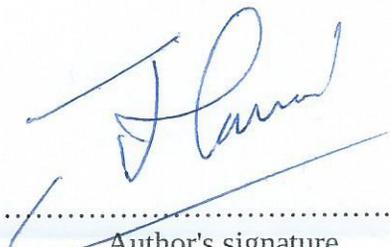
AUTHOR'S CONTRIBUTION STATEMENT

Work title: "Thirteen New Plastid Genomes from Mixotrophic and Autotrophic Species Provide Insights into Heterotrophy Evolution in Neottieae Orchids"

Author: Félix Lallemand

I declare that my contribution in article mentioned above and published in *Genome Biology and Evolution*, Volume 11, Issue 9, September 2019, Pages 2457–2467; doi.org/10.1093/gbe/evz170 included:

- Conceptualization
- Formal analysis
- Investigation
- Data curation
- Writing the original draft
- Review and editing the manuscript



.....
Author's signature

AUTHOR'S CONTRIBUTION STATEMENT

Work title: "Thirteen New Plastid Genomes from Mixotrophic and Autotrophic Species Provide Insights into Heterotrophy Evolution in Neottieae Orchids"

Author: Maria Logacheva

I declare that my contribution in article mentioned above and published in *Genome Biology and Evolution*, Volume 11, Issue 9, September 2019, Pages 2457–2467; doi.org/10.1093/gbe/evz170 included:

- Investigation
- Data curation
- Obtaining and providing resources
- Review and editing the manuscript

Sincerely yours,
Maria Logacheva, assistant professor
Skolkovo Institute of Science and Technology,
Center of Life Sciences
e-mail m.logacheva@skoltech.ru



Etude du Polymorphisme des Génomes Végétaux (EPGV),
INRA, Université Paris-Saclay,
Evry, France

date 10. Février 2022

AUTHOR'S CONTRIBUTION STATEMENT

Work title: "Thirteen New Plastid Genomes from Mixotrophic and Autotrophic Species Provide Insights into Heterotrophy Evolution in Neottieae Orchids"

Author: Isabelle Le Clainche

I declare that my contribution in article mentioned above and published in *Genome Biology and Evolution*, Volume 11, Issue 9, September 2019, Pages 2457–2467; doi.org/10.1093/gbe/evz170 included:

- Investigation process
- Obtaining and providing resources

LE CLAINCHE.....

Author's signature



Etude du Polymorphisme des Génomes Végétaux (EPGV),
INRA, Université Paris-Saclay,
Evry, France

date 10/02/2022

AUTHOR'S CONTRIBUTION STATEMENT

Work title: "Thirteen New Plastid Genomes from Mixotrophic and Autotrophic Species Provide Insights into Heterotrophy Evolution in Neottieae Orchids"

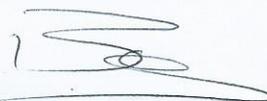
Author: Aurélie Bérard

I declare that my contribution in article mentioned above and published in *Genome Biology and Evolution*, Volume 11, Issue 9, September 2019, Pages 2457–2467; doi.org/10.1093/gbe/evz170 included:

- Investigation process
- Data curation
- Obtaining and providing resources

Aurélie Bérard.....

Author's signature



Uniwersytet Gdański
Wydział Biologii
Ul. Wita Stwosza 59
80-308, Gdańsk

data

OŚWIADCZENIE AUTORA O UDZIALE W PUBLIKACJI

Tytuł publikacji: *"Thirteen New Plastid Genomes from Mixotrophic and Autotrophic Species Provide Insights into Heterotrophy Evolution in Neottieae Orchids"*

Imię autora: Michał May

Oświadczam, że w wymienionym wyżej artykule, opublikowanym w *Genome Biology and Evolution*, Volume 11, Issue 9, September 2019, Pages 2457–2467; doi.org/10.1093/gbe/evz170 mój udział obejmował:

- Proces dochodzeniowy
- Obróbka, przetwarzanie i analiza danych
- Pozyskanie i dostarczenie materiału badawczego


.....

podpis autora

Peoples' Friendship University of Russia
Timiryazev State Biological Museum
Moscow, Russia

date 12.11.2021

AUTHOR'S CONTRIBUTION STATEMENT

Work title: "Thirteen New Plastid Genomes from Mixotrophic and Autotrophic Species Provide Insights into Heterotrophy Evolution in Neottieae Orchids"

Author: Ekaterina Zheleznaia

I declare that my contribution in article mentioned above and published in *Genome Biology and Evolution*, Volume 11, Issue 9, September 2019, Pages 2457–2467; doi.org/10.1093/gbe/evz170 included:

- Obtaining and providing resources
- Review and editing the manuscript



.....
Author's signature

Uniwersytet Gdański
Wydział Biologii
Ul. Wita Stwosza 59
80-308, Gdańsk

data

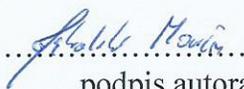
OŚWIADCZENIE AUTORA O UDZIALE W PUBLIKACJI

Tytuł publikacji: *"Thirteen New Plastid Genomes from Mixotrophic and Autotrophic Species Provide Insights into Heterotrophy Evolution in Neottieae Orchids"*

Imię i nazwisko autora: Marcin Jąkowski

Oświadczam, że w wymienionym wyżej artykule, opublikowanym w *Genome Biology and Evolution*, Volume 11, Issue 9, September 2019, Pages 2457–2467; doi.org/10.1093/gbe/evz170 mój udział obejmował:

- Obróbka, przetwarzanie i analiza danych
- Korekta i redagowanie manuskryptu

.....

.....
podpis autora

Institute of Plant Sciences Paris-Saclay (IPS2)
CNRS, INRA, Université Paris-Sud
Orsay, France
Université Evry, Université Paris-Saclay
Orsay, France

date

AUTHOR'S CONTRIBUTION STATEMENT

Work title: "Thirteen New Plastid Genomes from Mixotrophic and Autotrophic Species Provide Insights into Heterotrophy Evolution in Neottieae Orchids"

Author: Étienne Delannoy

I declare that my contribution in article mentioned above and published in *Genome Biology and Evolution*, Volume 11, Issue 9, September 2019, Pages 2457–2467; doi.org/10.1093/gbe/evz170 included:

- Review and editing the manuscript



.....
Author's signature

Etude du Polymorphisme des Génomes Végétaux (EPGV)
INRA, Université Paris-Saclay
Evry, France

date 2022-02-07

AUTHOR'S CONTRIBUTION STATEMENT

Work title: "Thirteen New Plastid Genomes from Mixotrophic and Autotrophic Species Provide Insights into Heterotrophy Evolution in Neottieae Orchids"

Author: Marie-Christine Le Paslier

I declare that my contribution in article mentioned above and published in *Genome Biology and Evolution*, Volume 11, Issue 9, September 2019, Pages 2457–2467;
doi.org/10.1093/gbe/evz170

included:

- Investigation
- Obtaining and providing resources
- Data curation
- Review and editing the manuscript
- Supervision



M-C LE PASLIER

.....
Author's signature

Institut de Systématique, Evolution, Biodiversité (ISYEB)
Muséum national d'Histoire naturelle
CNRS, Sorbonne Université, EPHE
Paris, France

date

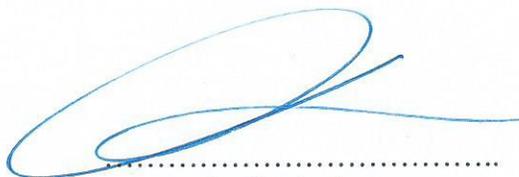
AUTHOR'S CONTRIBUTION STATEMENT

Work title: "Thirteen New Plastid Genomes from Mixotrophic and Autotrophic Species Provide Insights into Heterotrophy Evolution in Neottieae Orchids"

Author: Marc-André Selosse

I declare that my contribution in article mentioned above and published in *Genome Biology and Evolution*, Volume 11, Issue 9, September 2019, Pages 2457–2467; doi.org/10.1093/gbe/evz170 included:

- Conceptualization
- Review and editing the manuscript
- Supervision
- Funding acquisition



Author's signature

A.-A. Selosse

Thirteen New Plastid Genomes from Mixotrophic and Autotrophic Species Provide Insights into Heterotrophy Evolution in Neottieae Orchids

Félix Lallemand^{1,*}, Maria Logacheva^{2,3}, Isabelle Le Clainche⁴, Aurélie Bérard⁴, Ekaterina Zheleznaia⁵, Michał May⁶, Marcin Jakalski⁶, Étienne Delannoy^{7,8}, Marie-Christine Le Paslier⁴, and Marc-André Selosse^{1,6}

¹Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Paris, France

²Laboratory of Plant Genomics, Institute for Information Transmission Problems, Moscow, Russia

³Skolkovo Institute of Science and Technology, Moscow, Russia

⁴Etude du Polymorphisme des Génomes Végétaux (EPGV), INRA, Université Paris-Saclay, Evry, France

⁵Peoples' Friendship University of Russia, Timiryazev State Biological Museum, Moscow, Russia

⁶Faculty of Biology, Department of Plant Taxonomy and Nature Conservation, University of Gdańsk, Poland

⁷Institute of Plant Sciences Paris-Saclay (IPS2), CNRS, INRA, Université Paris-Sud, Orsay, France

⁸Université Evry, Université Paris-Saclay, Orsay, France

*Corresponding author: E-mail: felix.lallemand@protonmail.com.

Accepted: August 7, 2019

Data deposition: This project has been deposited at GenBank under the accession MH590345-MH590357; at NCBI BioProject under the accession PRJNA484137; at NCBI SRA under the accession SRP160956; at TreeBASE under the accession 24919; at the GitHub repository https://github.com/flallema/Neottieae_plastomes.

Abstract

Mixotrophic species use both organic and mineral carbon sources. Some mixotrophic plants combine photosynthesis and a nutrition called mycoheterotrophy, where carbon is obtained from fungi forming mycorrhizal symbiosis with their roots. These species can lose photosynthetic abilities and evolve full mycoheterotrophy. Besides morphological changes, the latter transition is associated with a deep alteration of the plastid genome. Photosynthesis-related genes are lost first, followed by housekeeping genes, eventually resulting in a highly reduced genome. Whether relaxation of selective constraints already occurs for the plastid genome of mixotrophic species, which remain photosynthetic, is unclear. This is partly due to the difficulty of comparing plastid genomes of autotrophic, mixotrophic, and mycoheterotrophic species in a narrow phylogenetic framework. We address this question in the orchid tribe Neottieae, where this large assortment of nutrition types occurs. We sequenced 13 new plastid genomes, including 9 mixotrophic species and covering all 6 Neottieae genera. We investigated selective pressure on plastid genes in each nutrition type and conducted a phylogenetic inference of the group. Surprisingly, photosynthesis-related genes did not experience selection relaxation in mixotrophic species compared with autotrophic relatives. Conversely, we observed evidence for selection intensification for some plastid genes. Photosynthesis is thus still under purifying selection, maybe because of its role in fruit formation and thus reproductive success. Phylogenetic analysis resolved most relationships, but short branches at the base of the tree suggest an evolutionary radiation at the beginning of Neottieae history, which, we hypothesize, may be linked to mixotrophy emergence.

Key words: Mycorrhiza, mycoheterotrophy, mixotrophy, phylogeny, Neottieae, plastome.

Introduction

Mixotrophic organisms grow by using both inorganic and organic carbon resources, either simultaneously or successively,

and therefore represent intermediates between autotrophy and heterotrophy (Selosse et al. 2017; Těšitel et al. 2018). Land plants widely associate with fungi through mycorrhizal

© The Author(s) 2019. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Table 1
Length, GC Content, and Genes Lost of the 13 Neottieae Plastomes Sequenced

Species ^a	Accession	Length				GC Content				Gene Loss ^b
		Total	LSC	SSC	IR	Total	LSC	SSC	IR	<i>Pseudogene/Undetected</i>
<i>P. pabstii</i>	MH590357	163,909	90,710	18,823	27,188	37.3	35	31	43.3	
<i>Ce. damasonium</i>	MH590345	161,699	88,720	19,085	26,947	37.3	35.1	30.7	43.2	
<i>Ce. longibracteata</i>	MH590346	161,986	88,888	19,138	26,980	37.2	35	30.6	43.1	
<i>Ce. rubra</i>	MH590347	162,277	88,814	19,199	27,132	37.2	35	30.6	43.1	
<i>E. albensis</i>	MH590348	159,763	87,237	18,782	26,872	37.3	35.1	30.7	43.2	
<i>E. atrorubens</i>	MH590349	159,790	87,237	18,803	26,875	37.3	35.1	30.7	43.3	
<i>E. gigantea</i>	MH590350	158,977	87,101	18,664	26,606	37.3	35.2	30.8	43.2	
<i>E. helleborine</i>	MH590351	159,822	87,313	18,785	26,862	37.3	35.1	30.7	43.2	
<i>E. microphylla</i>	MH590352	159,236	86,706	18,784	26,873	37.4	35.2	30.7	43.3	<i>ndhD, ndhE, ndhF, ndhH</i>
<i>E. palustris</i>	MH590353	159,134	87,114	18,702	26,659	37.4	35.2	30.8	43.2	
<i>E. purpurata</i>	MH590354	159,864	87,246	18,786	26,916	37.3	35.1	30.7	43.2	
<i>L. abortivum</i>	MH590355	128,822	85,544	15,099	27,102	36.3	35.1	30.4	43.1	<i>ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK, cemaA</i>
<i>N. cordata</i>	MH590356	147,034	82,416	12812	25903	37.5	35.1	29.3	43.4	<i>ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>

^aComplete genera names are *Palmorchis*, *Cephalanthera*, *Epipactis*, *Limodorum*, and *Neottia*.

^bA gene was presumed pseudogenized if a frameshift indel resulted in a reduction in protein length higher than 25% of its original length.

symbiosis and usually exchange their photosynthates for mineral nutrients (Smith and Read 2008). Repeatedly, plants evolved the ability to gain organic compounds also from their mycorrhizal fungi, a mixotrophic nutrition also called partial mycoheterotrophy (Selosse and Roy 2009; Hynson et al. 2013; Těšitel et al. 2018). We hereafter use the word mixotrophy for these plants that combine photosynthetic and mycorrhizal carbon sources. In many mixotrophic lineages from different families, species eventually lost photosynthesis and became fully mycoheterotrophic (Leake 1994; Merckx, Freudenstein, et al. 2013). Traits of mycoheterotrophy such as chlorophyll loss, reduction in leaf surface area, development of short clumpy roots, or production of reserve-less “dust” seeds have been first studied (Merckx, Mennes, et al. 2013). An increasing amount of molecular data now clarifies the genetic traits linked with the evolution to mycoheterotrophy.

The plastid genomes (or plastomes) of 24 fully mycoheterotrophic species have been sequenced to date (Graham et al. 2017; Petersen et al. 2018). Their comparative analysis has resulted in a model of sequential plastid genome degradation associated with mycoheterotrophy, which parallels what has been observed for heterotrophic parasitic plants (Funk et al. 2007; Barrett and Davis 2012; Barrett et al. 2014; Wicke et al. 2016; Graham et al. 2017). Genes encoding subunits of the nicotinamide adenine dinucleotide H dehydrogenase (NDH)-like complex, which regulates excessive electron flow in the plastidial electron transfer chain, are the first to be lost, followed by most photosynthesis-related genes. Rubisco large subunit (*rbcl*), adenosine triphosphate (ATP)-synthase, and plastid-encoded RNA polymerase (PEP) genes are often

retained in the early stage of photosynthesis loss. Eventually, degradation encompasses housekeeping genes involved in plastid translation and other specific functions. This model summarizes well how the plastid genome evolves in mycoheterotrophic and parasitic plant as they become more and more dependent on organic carbon uptake (Wicke et al. 2016; reviewed in Graham et al. [2017] and Wicke and Naumann [2018]).

Due to the few data available, still little is known about what occurs in the evolutionary steps prior to mycoheterotrophy, that is, in mixotrophic species that retain photosynthetic abilities. Whether these species already experience relaxed selective pressure on photosynthesis remains unclear in the general case. This is a challenging question because of the scarcity of lineages that have retained together autotrophic, mixotrophic, and mycoheterotrophic species. Known examples include the genus *Burmannia* L. (Burmanniaceae; Bolin et al. 2017), the tribe Pyroleae (Ericaceae; Lallemand et al. 2016), and the tribe Neottieae (Orchidaceae; Selosse and Roy 2009; Gonneau et al. 2014). In the leafless genus *Corallorhiza* Gagnebin (Orchidaceae), where no full autotroph occurs, comparison of plastomes from mixotrophic (green) and mycoheterotrophic (nongreen) species showed very limited gene loss in the former and relaxed selective constraints on photosynthesis and ATP-synthase genes in the latter (Barrett et al. 2014). Consistently, the selective regime does not differ between mixotrophic *Corallorhiza* and autotrophic outgroups (Barrett et al. 2014). The absence of autotrophic *Corallorhiza* as reference yet prevents firm conclusions regarding selective regime in mixotrophs. In the orchid genus *Cymbidium* Sw., *C. macrorhizon* is leafless but still holds

chlorophyll in the stem and fruits and the importance of its photosynthesis for fruiting has recently been demonstrated (Suetsugu et al. 2018). The plastome of this mixotrophic species shows no gene loss except for a few NDH genes and is mostly under purifying selection (Kim et al. 2018). The two genera *Corallorhiza* and *Cymbidium* where photosynthesis already undergoes relaxed selection in mixotrophic species (Barrett et al. 2014) call for extensive analyses in other lineages.

We sequenced plastomes of mixotrophs in Neottieae, an orchid tribe mostly found in northern hemisphere forests. Neottieae include 6 genera and about 200 species (WCSP 2018) with various nutrition types: *Palmorchis* Barb. Rodr. (all putative autotrophs), *Neottia* Guett. (many mycoheterotrophs and some probable mixotrophs; Těšitelová et al. 2012; Yagame et al. 2016; Schiebold et al. 2018), *Epipactis* Zinn (numerous mixotrophs and some autotrophs; Lallemand et al. 2018), *Cephalanthera* Rich. (some mycoheterotrophs and many mixotrophs), *Limodorum* Boehm. (all mixotrophs; Girlanda et al. 2006), and *Aphyllorchis* Blume (all mycoheterotrophs; Roy et al. 2009). Early studies on mycoheterotrophic and mixotrophic Neottieae contributed to our understanding of the ecophysiology and evolution of plants feeding on fungal carbon (Selosse et al. 2002; Gebauer and Meyer 2003; Julou et al. 2005). The transition to mixotrophy in this group is consistently linked to a shift in type of mycorrhizal fungi (supplementary table S1, Supplementary Material online). Autotrophic orchids associate with a polyphyletic group of saprotrophic and endophytic fungi, the rhizoctonias (Dearnaley et al. 2012), whereas mixotrophic Neottieae use ectomycorrhizal fungi from different genera, which simultaneously form mycorrhizas with surrounding trees (Selosse and Martos 2014). For the purpose of this study, we had to assign a given nutrition type to each species investigated. Even if there is a continuum between autotrophy and mixotrophy (Jacquemyn et al. 2017), we decided to consider a species as mixotrophic when it dominantly associated with ectomycorrhizal fungi and showed an enrichment in ^{13}C compared with surrounding autotrophs, a usual marker of fungal carbon gain (Hynson et al. 2013; see supplementary table S1, Supplementary Material online, for references regarding species nutrition).

Plastome sequencing of some Neottieae species by Feng et al. (2016) has yielded important results: 1) new hypotheses regarding the controversial phylogenetic relationships between Neottieae genera (Bateman et al. 2005; Roy et al. 2009; Górniak et al. 2010; Xiang et al. 2012), 2) patterns of gene loss in mycoheterotrophs matching the model described above, and 3) no relationship between the chlorophyll level (presence or absence based on visual estimation) and the plastome degradation in the leafless *Neottia* clade. Their sampling, however, included only four out of six Neottieae genera and one mixotrophic species (*Cephalanthera longifolia* (L.) Fritsch), limiting the phylogenetic resolution and the analysis

of plastome evolution during the emergence of mixotrophy and mycoheterotrophy. Here, we sequenced 13 additional Neottieae plastomes, including representatives of the 2 remaining genera, as well as 9 mixotrophic species (table 1). We investigated how gene content and selective pressure are related to nutrition type and carried out a tentative plastome-based phylogenetic reconstruction of Neottieae.

Plastome Structure, Gene Content, and Selective Regime

None of the plastomes of the 13 photosynthetic species sequenced displayed loss of large chromosome regions (>5 kb) or changes in synteny compared with related autotrophic Neottieae (Feng et al. 2016). Their sizes and GC contents were similar, with the exception of *Neottia cordata* and *Limodorum abortivum* which have smaller genomes as a consequence of NDH gene loss (table 1). A given gene was presumed pseudogenized if a frameshift indel resulted in a reduction in protein length higher than 25% of its original length. It was considered physically lost when it failed to pass the 77% similarity threshold (compared with closest available reference species) used for gene annotation and was thus undetected. Gene loss (pseudogenization or undetected) was restricted to the NDH complex in a few species (table 1): *N. cordata* (10 genes lost, *ndhI* with 148 amino acids vs. 182 in autotrophic *Neottia* species may also be nonfunctional), *L. abortivum* (all genes) and *Epipactis microphylla* (4 genes). An exception was *cemA* (chloroplast envelope membrane protein), which is lost in *L. abortivum* similarly to some mixotrophic *Corallorhiza* (Barrett et al. 2014), which may indicate a nonessential photosynthetic function. A 9-amino-acid deletion and a 28-amino-acid insertion occurred respectively in *rpoC1* for *Cephalanthera rubra* and *accD* for *N. cordata*. However, we did not find any stop codons and/or frameshift indels in their sequence, and *accD* is known to be highly variable in length (Kim and Lee 2004; Gurdon and Maliga 2014; Wicke and Naumann 2018), suggesting that both proteins may still be functional in these species.

The selective regime was analyzed for five main gene subsets (as in Graham et al. [2017]): photosynthesis-related, NDH, ATP, PEP, and housekeeping (table 2). We used two different methods, both based on ω , the ratio of nonsynonymous substitutions per nonsynonymous site (d_N) to synonymous substitutions per synonymous site (d_S). ω provides insights into selection intensity but is also sensitive to population size and other parameters such as time of divergence (Ohta and Ina 1995). The first analysis was done with the PAML program, which allows to compute different ω values for different groups according to their nutrition type. Although this method provides useful information, it is not sufficient to confidently conclude about intensification or relaxation of selective pressure. A higher ω value can especially be a consequence of either increased positive selection or

Table 2

Analysis of Selective Pressure for Different Sets of Genes Conserved among Autotrophic and Mixotrophic Neottieae

Genes and Models	PAML Analysis				RELAX Analysis	
	np ^a	ln L ^b	Λ ^c	P Value	k ^d	P Value
Photosynthesis^e						
M0: ω ₀ = 0.11	43	-33,266.07				
M1: ω ₀ = 0.12, ω _N = 0.11	44	-33,265.49	1.17	0.28	1.03	0.67
M2: ω ₀ = 0.12, ω _p = 0.15, ω _{N*} = 0.10	45	-33,263.31	4.35	0.04	1.13	0.44
M3: ω ₀ = 0.12, ω _p = 0.15, ω _{Na} = 0.11, ω _{Nx} = 0.09	46	-33,263	0.62	0.43	0.91	1
rbcl						
M0: ω ₀ = 0.12	43	-2,687.57				
M1: ω ₀ = 0.21, ω _N = 0.09	44	-2,685.94	3.26	0.07	0.57	0.03
M2: ω ₀ = 0.20, ω _p = 0.52, ω _{N*} = 0.04	45	-2,676.02	19.85	10 ⁻⁵	5.24	10 ⁻⁴
M3: ω ₀ = 0.20, ω _p = 0.52, ω _{Na} = 0.05, ω _{Nx} = 0.01	46	-2,675.11	1.82	0.18	10.97	0.04
NDH^f						
M0: ω ₀ = 0.21	29	-18,772.43				
M1: ω ₀ = 0.18, ω _N = 0.22	30	-18,771.82	1.22	0.27	1.63	0.22
M2: ω ₀ = 0.18, ω _p = 0.16, ω _{N*} = 0.24	31	-18,770.10	3.44	0.06	1.61	0.10
M3: ω ₀ = 0.18, ω _p = 0.16, ω _{Na} = 0.24, ω _{Nx} = 0.23	32	-18,770.09	0.02	0.90	0.57	0.77
ATP^g						
M0: ω ₀ = 0.13	43	-10,149.71				
M1: ω ₀ = 0.15, ω _N = 0.13	44	-10,149.43	0.56	0.45	1.08	0.49
M2: ω ₀ = 0.15, ω _p = 0.12, ω _{N*} = 0.13	45	-10,149.4	0.05	0.82	0.72	0.16
M3: ω ₀ = 0.15, ω _p = 0.12, ω _{Na} = 0.13, ω _{Nx} = 0.12	46	-10,149.36	0.09	0.76	1.01	0.93
PEP^h						
M0: ω ₀ = 0.26	43	-22,109.79				
M1: ω ₀ = 0.23, ω _N = 0.26	44	-22,109.33	0.92	0.34	0.93	0.83
M2: ω ₀ = 0.23, ω _p = 0.19, ω _{N*} = 0.28	45	-22,107.93	2.8	0.09	1.98	0.04
M3: ω ₀ = 0.23, ω _p = 0.19, ω _{Na} = 0.30, ω _{Nx} = 0.25	46	-22,107.14	1.59	0.21	0.69	0.39
Housekeepingⁱ						
M0: ω ₀ = 0.26	43	-25,635.43				
M1: ω ₀ = 0.23, ω _N = 0.27	44	-25,634.46	1.94	0.16	1.16	0.56
M2: ω ₀ = 0.22, ω _p = 0.30, ω _{N*} = 0.26	45	-25,634.21	0.49	0.48	1.1	0.75
M3: ω ₀ = 0.22, ω _p = 0.30, ω _{Na} = 0.27, ω _{Nx} = 0.25	46	-25,633.98	0.46	0.5	1.18	0.02
matK						
M0: ω ₀ = 0.35	43	-4,121.08				
M1: ω ₀ = 0.39, ω _N = 0.34	44	-4,120.92	0.33	0.56	1.11	0.7
M2: ω ₀ = 0.39, ω _p = 0.37, ω _{N*} = 0.34	45	-4,120.9	0.05	0.83	1.34	1
M3: ω ₀ = 0.39, ω _p = 0.37, ω _{Na} = 0.45, ω _{Nx} = 0.23	46	-4,118.08	5.63	0.02	0.48	0.52
ycf1 + ycf2						
M0: ω ₀ = 0.68	43	-27,092.79				
M1: ω ₀ = 0.78, ω _N = 0.65	44	-27,091.67	2.23	0.14	1.12	0.46
M2: ω ₀ = 0.78, ω _p = 0.57, ω _{N*} = 0.66	45	-27,091.38	0.59	0.44	1.1	1
M3: ω ₀ = 0.78, ω _p = 0.57, ω _{Na} = 0.75, ω _{Nx} = 0.55	46	-27,088.8	5.16	0.02	1.13	0.13

^aNumber of parameters for the model.

^bLog-likelihood of the data for the model.

^cLog-likelihood ratio test statistic used to compute P value.

^dSelection intensity parameter.

^eccsA, petA, petB, petD, petG, petL, petN, psaA, psbA, psbC, psal, psaj, psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbl, psbj, psbK, psbL, psbM, psbN, psbT, psbZ, rbcl, ycf3, and ycf4.

^fndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, and ndhK; species analyzed were *Cephalanthera damasonium*, *Ce. longibracteata*, *Ce. longifolia*, *Ce. rubra*, *Epipactis albensis*, *E. atrorubens*, *E. gigantea*, *E. helleborine*, *E. palustris*, *E. purpurata*, *Neottia fugongensis*, *N. ovata*, *Palmeria pabstii*, *Sobralia callosa*, and *Calanthe triplicata*.

^gatpA, atpB, atpE, atpF, atpH, and atpI.

^hrpoA, rpoB, rpoC1, and rpoC2.

ⁱaccD, clpP, infA, matK, rpl2, rpl14, rpl16, rpl20, rpl22, rpl23, rpl32, rpl33, rpl36, rps2, rps3, rps4, rps7, rps8, rps11, rps12, rps14, rps15, rps16, rps18, and rps19.

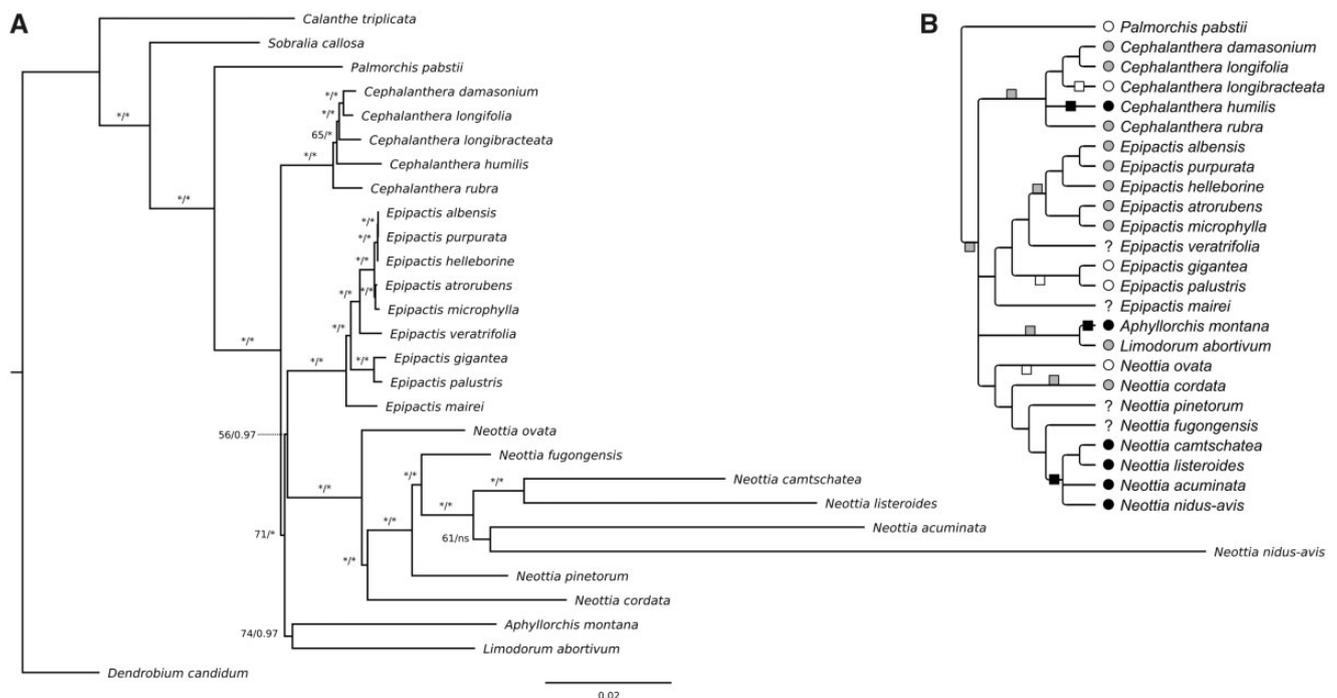


Fig. 1.—(A) Phylogeny of Neottieae based on whole plastome analysis. Numbers above or to the left of branches represent bootstrap values (1,000 replicates) from ML analysis (left, stars indicate numbers above 95) and posterior probabilities from Bayesian inference (right, stars indicate probabilities above 0.98), ns means that the branch was not recovered in the Bayesian analysis. Scale bar: number of substitutions per site. (B) Two scenarios of mixotrophy and mycoheterotrophy evolution among Neottieae, that is, autotrophy (changes above the line) versus mixotrophy (changes below the line) of the Neottieae (excl. *Palmorchis*) ancestor. The tree is a consensus of the three analyses carried out on different sets of species and plastid DNA data (see [supplementary fig. S1, Supplementary Material](#) online, for other analyses). Nutrition type evolution is mapped considering the most parsimonious scenario in each case. Circles in front of species names indicate autotrophy (white), mixotrophy (gray), or mycoheterotrophy (black) and squares on the branches indicate shifts to these nutrition types in each scenario. Nutrition type evolution is mapped considering the most parsimonious scenario in each case. Boxes on the branches show changes occurring in both scenarios.

relaxation (Wertheim et al. 2015). We used the program RELAX as a second method to refine our analysis. RELAX allows more subtle estimation of ω variations by considering different categories of sequence sites that can evolve under different selective constraints (Wertheim et al. 2015).

For PAML analysis, four nested models allowing the calculation of different ω values for some subsets of branches were defined: M0, the null model with one ω for the whole tree; M1, which allowed ω to differ between outgroups (ω_0) and Neottieae (ω_N); M2, same as M1 but allowing a different ω for *Palmorchis* (ω_p) and the remaining Neottieae (ω_{N^*} ; below noted as Neottieae* = Neottieae without *Palmorchis*); M3, same as M2 but allowing ω to differ between autotrophic (ω_{Na}) and mixotrophic Neottieae* (ω_{Nx} , see [fig. 1B](#) for species considered as mixotrophic). Models were compared by likelihood ratio tests following this scheme: M1 versus M0, M2 versus M1, and M3 versus M2. For all subsets, successive comparisons of the four models of ω variation among branches did not show a better fit for one model against another one, except for the photosynthesis genes for M2 versus M1 ([table 2](#)). M2 allowed different ω for outgroups ($\omega_0 = 0.12$), *Palmorchis pabstii* ($\omega_p = 0.15$), and the

remaining Neottieae ($\omega_{N^*} = 0.10$; [table 2](#)). These results were not supported by the RELAX analysis, which indicated instead intensification of selective constraints for PEP genes comparing Neottieae* and *Palmorchis* ($k = 1.98$; $P = 0.04$) and for housekeeping genes comparing mixotrophic and autotrophic Neottieae* ($k = 1.18$; $P = 0.02$; [table 2](#)). But one should consider that these results are not adjusted for multiple comparisons.

The same analysis carried out for individual photosynthesis-related genes showed a variation in selective regime only for *rbcl* ([table 2](#) for *rbcl*, other genes shown in [supplementary table S2, Supplementary Material](#) online). *rbcl* data fitted models M2 much better than M1. In M2, ω was lower for Neottieae* ($\omega_{N^*} = 0.04$) compared with outgroups ($\omega_0 = 0.20$) and *P. pabstii* ($\omega_p = 0.52$; [table 2](#)). This was supported by the RELAX analysis, which showed an intensification of selective constraints in Neottieae* compared with *Palmorchis* on the one hand ($k = 5.24$; $P = 10^{-4}$), and in mixotrophic compared with autotrophic Neottieae* on the other hand ($k = 10.97$; $P = 0.04$). Thus, *rbcl* experiences stronger purifying selection in Neottieae*, and more particularly in mixotrophic species. For individual nonphotosynthesis genes,

PAML analysis showed that both *matK* and *ycf1 + ycf2* fitted M3 better than M2, with higher ω for autotrophic ($\omega_{Na} = 0.45$, resp. 0.75) compared with mixotrophic Neottieae* ($\omega_{Nx} = 0.23$, resp. 0.55; [table 2](#)). This was, however, not supported by the RELAX analysis ([table 2](#)).

To summarize, we did not find evidence for relaxed selection in plastomes of mixotrophic species. By contrast, both PAML and RELAX analyses showed stronger selective pressure in mixotrophic Neottieae* compared with *Palmorchis* for *rbcl*.

Plastome Evolution in Mixotrophic Neottieae

The observed loss of NDH genes in three plastomes confirms that this gene complex is sensitive to evolutionary decay in Neottieae (Feng et al. 2016) and more generally in orchids (Barrett et al. 2014; Luo et al. 2014; Kim et al. 2015; Lin et al. 2017; Kim et al. 2018) and some other photosynthetic lineages (Wicke and Naumann 2018). Interestingly, it does not correlate with species nutrition. For example, the 11 genes were lost in *L. abortivum* and *N. cordata* (if *ndhI* is indeed nonfunctional). The first is mixotrophic and highly dependent on fungal carbon because its photosynthesis does not even compensate for its respiration (Girlanda et al. 2006); the second displays a more versatile use of fungal resources (Schiebold et al. 2018), which may even be optional (Těšitelová et al. 2015). Similarly, *E. microphylla* is the only species of the mixotrophic *Epipactis* section (*E. albensis*, *E. atrorubens*, *E. helleborine*, and *E. purpurata*) (Jin et al. 2014) for which NDH gene loss occurred, whereas some losses have been reported in *E. veratrifolia* and *E. mairei* (Feng et al. 2016) that are supposed autotrophic (pending more investigations). This suggests a species-specific decay of NDH genes in orchids, independently of their autotrophic or mixotrophic habit. In addition, we did not observe any evidence for intensification or relaxation of selective pressure on these genes ([table 2](#)). We suppose that relaxation on NDH genes occurs in the whole orchid clade, making any difference among orchid species hardly detectable. Changes in nutrition type do not seem to alter this pattern.

Photosynthesis-related genes neither are lost nor experience relaxed selective constraints in mixotrophic species but, in Neottieae*, they showed a slightly lower ω ; *rbcl* was even under stronger purifying selection compared with *P. pabstii* ([table 2](#)). Some parts of the plastomes unrelated to photosynthesis also experienced higher selective constraints in mixotrophs, as shown by 1) lower ω for *matK* and *ycf1 + ycf2* in mixotrophic than in autotrophic Neottieae* and 2) the RELAX analyses of PEP and housekeeping genes (displaying $k > 1$). These results are concordant with what has been reported for mixotrophic *Corallorhiza*, which 1) retain all photosynthesis-related genes except *cemA* and *psbM*, 2) display apparently unimpaired photosynthesis, and 3) show no difference in plastome selective regime compared with autotrophic references

(Barrett et al. 2014). This also agrees with the retention of purifying selection on plastidial genes observed in the mixotrophic *C. macrorhizon* (Kim et al. 2018). However, our results contrast with other kinds of mixotrophy where relaxed purifying selection has been observed for photosynthesis, PEP and ATP-synthase genes, namely 1) in carnivorous plants (*psa*, *atp*, and *rpo* genes in Wicke et al. [2014]) and 2) in obligate hemiparasites (*psa*, *psb*, *pet*, *atp*, and *rbcl* in Petersen et al. [2015]; photosynthesis and *atp* genes in Wicke et al. [2016]). For this latter trophic strategy, retention of selective constraints on photosynthesis genes has nevertheless been reported for photosynthesis genes in *Cuscuta* (McNeal et al. 2007) and also for some housekeeping genes in *Viscum* (*accD*, *cemA*, *clpP*, and *ycf2*, Petersen et al. 2015). However, these comparisons between studies have to be taken with caution given the differences in scales among them or in a given study (individual genes vs. gene subsets).

In Neottieae, depending on the plastid genes considered, mixotrophic species thus feature either no change or a lowly supported increase in selective pressure compared with other mixotrophic groups such as *Corallorhiza*, carnivorous or parasitic plants (Barrett et al. 2014; Wicke et al. 2014, 2016; Petersen et al. 2015). Photosynthesis therefore seems to remain advantageous in mixotrophic Neottieae, and this is congruent with physiological evidence for the importance of photosynthetic carbon in fruit development and reproduction in these species (Roy et al. 2013; Bellino et al. 2014; Gonneau et al. 2014; Lallemand et al. 2019). We hypothesize that the retention of genes and purifying selection in the plastome of *C. macrorhizon* (Kim et al. 2018) indicates a similar crucial role of photosynthesis in reproduction, as illustrated by Suetsugu et al. (2018).

Neottieae Phylogeny and Mycoheterotrophy Evolution

Phylogenetic analysis of the 28 available plastomes confirmed Neottieae monophyly with *Palmorchis* in sister position to the remaining species ([fig. 1A](#)). *Neottia*, *Epipactis*, and *Cephalanthera* genera were each monophyletic and most intrageneric relationships were well defined. *Aphyllorchis* and *Limodorum* were sister genera but with mild support from maximum likelihood (ML) analysis (bootstrap = 74, [fig. 1A](#)). However, very short branches with low support prevented strong conclusions about intergeneric relationships ([fig. 1A](#)). Because mycoheterotrophic *Neottia* species display accelerated evolutionary rates (see long branches in [fig. 1A](#)), we removed them to increase alignment quality: We then obtained a congruent tree with better support for most but not all clades ([supplementary fig. S1A](#), [Supplementary Material](#) online). In particular, *Neottia*, *Epipactis*, *Aphyllorchis*, and *Limodorum* clustered together in a monophyletic group sister to *Cephalanthera* with better support from ML analysis (bootstrap = 87 vs. 71). When restricting

the analysis to the plastidial coding sequences (CDS) of non-mycoheterotrophic species, that is, excluding noncoding regions to allow better alignment, *Limodorum* was sister to *Epipactis* but *Neottia* no longer clustered with these genera in a monophyletic group (supplementary fig. S1B, Supplementary Material online). Combining these results suggests that at least *Cephalanthera* is sister clade to the rest of Neottieae. Analysis based on ITS (internal transcribed spacer of nuclear ribosomal DNA, including ITS1-5.8S rRNA-ITS2) confirmed *Aphyllorchis* and *Limodorum* as sister genera but failed to clarify further intergeneric relationships (supplementary fig. S1C, Supplementary Material online).

Phylogenetic relationships between the Neottieae genera have long been controversial and studies using different species and marker combinations yielded contradictory results, without any clear intergeneric relationships (ITS and the plastid *trnL* intron in Bateman et al. [2005]; ITS, *trnS-G* spacer, and *rbcl* in Roy et al. [2009]; ITS, *Xdh*, *rbcl*, *matK*, *psaB*, and *trnL-F* spacer in Xiang et al. [2012]; ITS, *rbcl*, and *matK* in Zhou and Jin [2018]). More confusingly, the phylogeny of Feng et al. (2016), based on a subset of the plastomes we analyzed, disagrees with our results, displaying *Cephalanthera* and *Aphyllorchis* as sister genera, and *Epipactis* and *Neottia* as successive sisters. This inconsistency may result from the difference in the number and diversity of species included in the analysis. We added 13 species and 2 new genera (*Palmorchis* and *Limodorum*) but failed to reproduce their results when considering only their subset of species, suggesting that this is rather due to different phylogenetic methods and/or to evolutionary signals conveyed by the plastome regions used. Feng et al. (2016) selected conserved plastome regions, whereas we made a dual analysis of whole plastome alignment and CDS-restricted alignment. Moreover, we chose not to filter or manually edit alignments following Tan et al. (2015) for accurate phylogenetic inferences (see also Hallas et al. 2017).

Neottieae intergeneric relationships thus remain difficult to estimate confidently. Our plastome-based analysis showed that the branches separating these genera are very short ($<10^{-3}$ substitutions per site). Thus, the common ancestor of Neottieae* (i.e., Neottieae with exclusion of *Palmorchis*) likely experienced rapid speciation, with little time for informative mutations to accumulate (Glor 2010). Such a fast evolution of ancestral lineage makes it prone to hybridization and incomplete lineage sorting (Glor 2010; Wang et al. 2014), so that independent loci (e.g., nuclear and plastidial markers) may encapsulate different phylogenetic stories and make any combined analysis unreliable. Another possibility is heterotachy (Fitch and Markowitz 1970), a variation (here a decrease) in evolutionary rate in the ancestral lineage of Neottieae*. Arbitrating between fast speciation and heterotachy would be particularly challenging given the very short branch lengths. Confident reconstruction of early speciation in Neottieae will be complicated, but different gene trees

could provide interesting insights into early Neottieae evolution and possible hybridization events.

We also wanted to compare two possible evolution of nutrition in Neottieae*. Based on our plastid phylogeny, we display two hypotheses, namely 1) an autotrophic ancestor and 2) a mixotrophic ancestor for Neottieae* (see Materials and Methods for reconstruction of intermediate characters). Assuming an autotrophic common ancestor, mixotrophy evolved at least four times (fig. 1B, gray boxes above the branches). Conversely, assuming a mixotrophic common ancestor would give three reversions to autotrophy (fig. 1B; white boxes below the branches). Each scenario implies two emergences of mycoheterotrophy, but unfortunately, the lack of information about the trophic status of some taxa (such as *E. mairei* or *Neottia fugongensis*) and the absence of several mycoheterotrophic clades limit our analysis. For example, green *Neottia* other than *N. cordata* or *Epipactis* species outside the *Epipactis* section may actually show mixotrophic nutrition. Assuming a mixotrophic common ancestor implies one change less than an autotrophic one but includes reversion to autotrophy. Although unlikely at first glance, reversion is allowed because mixotrophy is flexible and displays a continuum from autotrophy to mycoheterotrophy (Jacquelyn et al. 2017). Most importantly, the selective pressure maintaining photosynthesis genes reported here allows a reversion from mixotrophy to autotrophy, because potential for autotrophy remains intact in mixotrophic Neottieae. The question whether mixotrophy is a plesiomorphy of Neottieae thus remains open, pending for analyses of more species.

A final speculation comes at this point: If mixotrophy turns out to be indeed the ancestral state of Neottieae, could it have been a key factor in the fast early diversification speculated above? Mixotrophy allows plants to grow in light-limited forest understory and so to occupy a broad new ecological niche (Těšitel et al. 2018). Orchids show important dispersion abilities and most mixotrophic Neottieae do not seem very selective regarding the fungi they associate with (Těšitelová et al. 2012). One can therefore speculate that an early mixotrophic Neottieae lineage could have rapidly colonized distant and heterogeneous forest patches, with ability to keep or lose photosynthetic abilities, thus favoring radiation.

Materials and Methods

DNA Extraction and Sequencing

Total DNA was extracted from the leaves or fruits of the following Neottieae species (table 1 and supplementary table S1, Supplementary Material online): *Palmorchis pabstii* Veyret, *N. cordata* (L.) Rich., *E. albensis* Nováková & Rydlo, *E. atrorubens* (Hoffm.) Besser, *E. gigantea* Douglas ex Hook., *E. helleborine* (L.) Crantz, *E. microphylla* (Ehrh.) Sw., *E. palustris* (L.) Crantz, *E. purpurata* Sm., *Cephalanthera damasonium* (Mill.) Druce, *Ce. longibracteata* Blume, *Ce. rubra* (L.)

Rich., and *L. abortivum* (L.) Sw. The following extraction kits and protocols were used (supplementary table S3, Supplementary Material online): DNeasy Plant Mini Kit (Qiagen) with final elution in distilled water, NucleoSpin Plant II kit (Macherey-Nagel), CTAB extraction as in Porebski et al. (1997).

Given low yield due to DNA degradation in some old samples, the Accel-NGS 1S Plus DNA Library Kit for Illumina with the 1S Plus Indexing kit (Swift Biosciences Inc., Ann Arbor, MI) was used for preparation of libraries, following the manufacturer's instructions. We did so for the set of samples handled in EPGV laboratory (supplementary table S3, Supplementary Material online). For other samples, handled in Moscow laboratory, the TruSeq DNA sample preparation kit (Illumina, USA) and the NEBNext Ultra DNA kit (New England Biolabs, USA) were used for preparation of libraries (supplementary table S3, Supplementary Material online). Paired-end sequencing on multiplexed libraries was performed either on Illumina HiSeq 2000, Illumina HiSeq 2500 using the rapid run mode, and Illumina HiSeq 4000 or Illumina MiSeq, depending on the samples (different material available in the different teams involved in the study; supplementary table S3, Supplementary Material online). Raw sequences for *Ce. longibracteata*, *Ce. rubra*, *E. albensis*, *E. atrorubens*, *E. gigantea*, *E. microphylla*, *E. palustris*, *E. purpurata*, and *P. pabstii* were submitted to the NCBI database: BioProject PRJNA484137, SRA accession number ongoing.

Read Cleaning, Plastome Assembly, and Annotation

Redundant reads were removed with a homemade C++ script. The redundancies were sought by pairs read1–read2, the redundancy with the best Phred score was kept. Trimming of the low-quality 3' end was done using a homemade C++ script, one base after the other until reaching a base with a Phred score > 30 or until the mean Phred score of the read was above 30. Pairs with at least one read shorter than 30 bp after this step or with N were removed. Read number before and after cleaning are indicated in supplementary table S4, Supplementary Material online.

For samples that were prepared with the Accel-NGS 1S Plus DNA Library Kit, adapter trimming was conducted using BBDuk Trimmer plugin version 37.64 in Geneious version 11.1.3 (<http://www.geneious.com>; last accessed September 2018; Kearsse et al. 2012) with the following parameters: adapters = Illumina Truseq DNA adapters, ktrim = r, k = 27, hdist = 1, edist = 0, mink = 6, minlength = 10, trimbyoverlap = t, and minoverlap = 6. Addition of a low complexity tail to the 3' end of fragments was part of the library preparation kit and had to be removed from the beginning of reads 2. Following the manufacturer's recommendations, trimming an additional 10 bp at the 5' end was done for all reads.

De novo assembly was done differently depending on the species (supplementary table S3, Supplementary Material online). 1) A subset of 25% of the reads was assembled using

the Geneious algorithm with medium-low sensitivity parameters. Resulting contigs were mapped on the plastome of *Neottia ovata* (L.) Bluff & Fingerh. (KU551271), *Cephalanthera longifolia* (L.) Fritsch (KU551263), or *E. veratrifolia* Boiss. & Hohen (KU551267), depending on the species. Contigs corresponding to the plastome were then dissolved and reassembled with medium sensitivity parameters to increase assembly quality. The resulting contigs were again mapped to the reference genome and large single copy (LSC), inverted repeat B (IRB), and small single copy (SSC) regions were isolated. IRB was duplicated and reverse complemented to obtain inverted repeat A (IRA). Concatenation of LSC, IRB, SSC, and IRA led to the final plastome sequence. 2) The CLC Genomics Workbench de novo assembly procedure was used. In this latter case, reads were first trimmed with the following settings: quality limit = 0.01 (corresponds to $Q \geq 20$), remove adapters = yes (Illumina Truseq adapters), discard short reads = yes (limit 50 nt for MiSeq data, 25 nt for HiSeq data). Then de novo assembly was performed with the following settings: automatic bubble size = yes, minimum contig length = 1,000 bp, automatic word size = yes, perform scaffolding = yes, and auto-detect paired distances = yes. Contigs of plastid origin were selected based on the results of BLAST search with *Phalaenopsis aphrodite* plastome used as query. Contigs corresponding to LSC, IR, and SSC were concatenated accordingly and resulting sequences were verified and corrected by performing the gap closure method using back mapping of the reads. IRs were duplicated and reverse complemented to obtain the final parts (IRA) of the plastomes.

Coverage values of the contigs used for assembly are indicated in supplementary table S4, Supplementary Material online. Depending on the species, the phylogenetically closest reference plastome was used for annotations transfer in Geneious with a 77% similarity threshold that turned out, after preliminary attempts, to be low enough to detect similar gene sequences and sufficiently high to avoid false positives. Manual correction and addition of missing information was then carried out for each plastome. Annotated plastome sequences were submitted to GenBank under the accession numbers MH590345–MH590357 (table 1).

Phylogenetic Inferences

Including the 12 already published sequences, we ended up with a set of 25 Neottieae plastomes and added 3 outgroup species from the same subfamily Epidendroideae (supplementary table S1, Supplementary Material online). The four mycoheterotrophic *Neottia* were either kept or removed to test whether their rapidly evolving plastomes (much higher than the two other mycoheterotrophs; fig. 1A) and the increased length and quality of the alignment obtained without them alter phylogeny reconstruction. The second copy of the inverted repeat region (IRA) was removed from all plastomes

and the resulting sequences were aligned a first time using the Mauve progressive algorithm, as implemented in Geneious (parameters set to automatically calculate seed weight and minimum locally collinear block (LCB) score, compute LCBs, full alignment), to detect potential rearrangements (Darling et al. 2004). The LCBs obtained were individually aligned a second time using the MAFFT online service (Kato et al. 2017) with automatic selection of alignment strategy, gap opening penalty set to 3.0, offset value set to 0.1 and other parameters left as default. The resulting alignments were then concatenated. Model selection was done with the ModelFinder algorithm (Kalyaanamoorthy et al. 2017) implemented in the IQ-TREE program (Nguyen et al. 2015). The model with lowest Bayesian information criterion score was used for ML analysis with IQ-TREE (Nguyen et al. 2015). The first model with the lowest Bayesian information criterion compatible with MrBayes v3.2 parameters (Ronquist et al. 2012) was chosen for Bayesian inference with this program. A complementary analysis using the nuclear marker ITS, including ITS1-5.8S rRNA-ITS2, was done using available Neottieae sequences (supplementary table S1, Supplementary Material online).

On the plastid phylogeny (fig. 1B), we displayed two possible evolution of nutrition based on the hypotheses of 1) an autotrophic ancestor or 2) a mixotrophic ancestor for Neottieae*. Intermediate characters for other nodes were proposed in order to minimize the number of transitions between trophic types (autotrophic, mixotrophic, or mycoheterotrophic), that is, the most parsimonious scenario under each hypothesis, considering the nutritional status of extant species (supplementary table S1, Supplementary Material online). A given species was considered as mixotrophic when it dominantly associated with ectomycorrhizal fungi and showed an enrichment in ^{13}C compared with surrounding autotrophs, a usual marker of fungal carbon gain (Hynson et al. 2013; see supplementary table S1, Supplementary Material online, for references regarding species nutrition).

Selective Regime Analyses

All CDS were extracted from the same set of plastomes used for phylogeny reconstruction, except the six mycoheterotrophic species. Except for NDH complex, genes that were lost or pseudogenized in at least one of the remaining species were discarded from subsequent analyses (see table 2 for the remaining genes). For NDH, because some species had lost all the genes, we rather only kept for subsequent analysis only the 15 species, which retained the 11 NDH genes (table 2). CDS were aligned based on their amino acid sequences with subsequent backtranslation using the AlignTranslation function (Wright 2015) of the DECIPHER package (Wright 2016) in the R environment for statistical computing (R Development Core Team 2007). Functional gene groups were built by concatenating CDS for photosynthesis, NDH, ATP-synthase,

PEP, and housekeeping genes (table 2). *ycf1* and *ycf2* were combined and analyzed independently because of their specific evolutionary rates (Barrett et al. 2014). In addition, some genes were analyzed individually when long enough for confident parameter estimation: *ccsA*, *psaA*, *psaB*, *psbA*, *psbB*, *psbC*, *psbD*, *rbcl*, *accD*, and *matK* (see supplementary table S2, Supplementary Material online). ω , the ratio of nonsynonymous substitutions per nonsynonymous site (d_N) to synonymous substitutions per synonymous site (d_S), was used to estimate selective pressure for different sets of genes. The codeml program implemented in PAML version 4.9 (Yang 2007) was used to compute ω values, with the branch model and the tree shown in figure 1B. The codon frequency parameter was set to F3X4. Four nested models allowing the calculation of different ω values for some subsets of branches were defined: M0, the null model with a unique ω for the whole tree; M1, which allowed ω to differ between outgroups (ω_O) and Neottieae (ω_N); M2, same as M1 but allowing a different ω for *Palmorchis* (ω_P) and the remaining Neottieae (ω_{N^*}); M3, same as M2 but allowing ω to differ between autotrophic (ω_{Na}) and mixotrophic Neottieae* (ω_{Nx} , see fig. 1B for species considered as mixotrophic). Models were compared by likelihood ratio tests following this scheme: M1 versus M0, M2 versus M1, and M3 versus M2. The statistic of the test (Λ) was calculated as two times the difference in model likelihood and compared with a chi-squared distribution with degrees of freedom equaling the difference in parameters number between the two models. *P* values were obtained using the function *pchisq* of the R stats package.

We completed the PAML analysis with a specific test dedicated to the detection of changes in selective pressure (relaxation or intensification). This was carried out with the program RELAX, available at the Datamonkey webserver <http://datamonkey.org/relax>; last accessed March 6, 2019. RELAX allows more subtle estimation of ω variations by considering different categories of sequence sites that can evolve under different selective constraints (Wertheim et al. 2015). ω distributions are estimated for a test (T) and a reference (R) subset of branches and a selection intensity parameter *k* such that $\omega_T = \omega_R^k$ is introduced. The program then compares the goodness of fit of the data to a branch-site evolutionary model with *k* = 1 (null model) or *k* being a free parameter (alternative model). When the alternative model shows a better fit following a likelihood ratio test, *k* < 1 (resp. *k* > 1) indicates a relaxation (resp. an intensification) of selective constraints in the test group. Three different tests were done with three different designs of test versus reference groups, copying what has been done with the PAML nested analysis (see supplementary fig. S2, Supplementary Material online, for group designs).

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We acknowledge Aurélie Chauveau for preparation of the libraries, Élodie Marquand and Aurélie Canaguier for data processing, two anonymous referees and Dennis Lavrov for their corrections, and David Marsh for English corrections. This work was funded by the Fondation de France (funds from the Fondation Ars Cuttoli & Paul Appell, project name : EVOLUTION ET PHYSIOLOGIE DES PLANTES QUI SE NOURRISSENT DE CHAMPIGNONS (MYCOHETEROTROPHIE)) and the National Science Center (Poland; grant no. 2015/18/A/NZ8/00149). We thank the Commissariat à l'Énergie Atomique et aux Énergies Alternatives – Institut de Génomique/Centre National de Génotypage for conducting the DNA quality control and for giving the INRA-EPGV group access to their Illumina Sequencing Platform. Plastomes of *Ce. damasonium* and *L. abortivum* were sequenced with support of budgetary subsidy to IITP RAS (Laboratory of Plant Genomics). We thank the MSU Herbarium (MW) for providing a sample of *Ce. longibracteata* Blume MW0047886 and “Service de Systématique Moléculaire” (UMS2700 MNHN/CNRS) for granting access to its technical platform.

Author Contributions

Conceptualization: F.L. and M.-A.S.; formal analysis: F.L.; investigation: F.L., M.L., M.M., I.L.C., A.B., and M.-C.L.P.; resources: M.L., I.L.C., A.B., E.Z., M.M., and M.-C.L.P.; data curation: F.L., M.L., M.M., M.J., and A.B.; writing—original draft: F.L.; writing—review and editing: F.L., M.L., E.Z., E.D., M.J., M.-C.L.P., and M.-A.S.; supervision: M.-C.L.P. and M.-A.S.; funding acquisition: M.-A.S.

Literature Cited

- Barrett CF, Davis JI. 2012. The plastid genome of the mycoheterotrophic *Corallorhiza striata* (Orchidaceae) is in the relatively early stages of degradation. *Am J Bot.* 99(9):1513–1523.
- Barrett CF, et al. 2014. Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. *Mol Biol Evol.* 31(12):3095–3112.
- Bateman RM, Hollingsworth PM, Squirrel J, Hollingsworth M. 2005. Tribe Neottieae. In: Pridgeon A, Cribb PJ, Chase MW, Rasmussen FN, editors. *Genera Orchidacearum Volume 4: Epidendroideae, Part 1*. Oxford, New York: Oxford University Press.
- Bellino A, et al. 2014. Nutritional regulation in mixotrophic plants: new insights from *Limodorum abortivum*. *Oecologia* 175(3):875–885.
- Bolin JF, Tennakoon KU, Majid MBA, Cameron DD. 2017. Isotopic evidence of partial mycoheterotrophy in *Burmanniaceae*. *Plant Species Biol.* 32(1):74–80.
- Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14(7):1394–1403.
- Dearnaley JDW, Martos F, Selse M-A. 2012. Orchid mycorrhizas: molecular ecology, physiology, evolution and conservation aspects. In: Hock B, editor. *Fungal Associations. The Mycota (a comprehensive treatise on fungi as experimental systems for basic and applied research)*. Vol. 9. Berlin/Heidelberg (Germany): Springer. p. 207–230.
- Feng Y-L, et al. 2016. Lineage-specific reductions of plastid genomes in an orchid tribe with partially and fully mycoheterotrophic species. *Genome Biol Evol.* 8(7):2164–2175.
- Fitch WM, Markowitz E. 1970. An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. *Biochem Genet.* 4(5):579–593.
- Funk HT, Berg S, Krupinska K, Maier UG, Krause K. 2007. Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. *BMC Plant Biol.* 7:45.
- Gebauer G, Meyer M. 2003. ¹⁵N and ¹³C natural abundance of autotrophic and myco-heterotrophic orchids provides insight into nitrogen and carbon gain from fungal association. *New Phytol.* 160(1):209–223.
- Girlanda M, et al. 2006. Inefficient photosynthesis in the Mediterranean orchid *Limodorum abortivum* is mirrored by specific association to ectomycorrhizal Russulaceae. *Mol Ecol.* 15(2):491–504.
- Glor RE. 2010. Phylogenetic insights on adaptive radiation. *Annu Rev Ecol Syst.* 41(1):251–270.
- Gonneau C, et al. 2014. Photosynthesis in perennial mixotrophic *Epipactis* spp. (Orchidaceae) contributes more to shoot and fruit biomass than to hypogeous survival. *J Ecol.* 102(5):1183–1194.
- Górniak M, Paun O, Chase MW. 2010. Phylogenetic relationships within Orchidaceae based on a low-copy nuclear coding gene, *Xdh*: congruence with organellar and nuclear ribosomal DNA results. *Mol Phylogenet Evol.* 56(2):784–795.
- Graham SW, Lam VKY, Merckx V. 2017. Plastomes on the edge: the evolutionary breakdown of mycoheterotroph plastid genomes. *New Phytol.* 214(1):48–55.
- Gurdon C, Maliga P. 2014. Two distinct plastid genome configurations and unprecedented intraspecific length variation in the *accD* coding region in *Medicago truncatula*. *DNA Res* 21: 417–427.
- Hallas JM, Chichvarkhin A, Gosliner TM. 2017. Aligning evidence: concerns regarding multiple sequence alignments in estimating the phylogeny of the *Nudibranchia* suborder Doridina. *R Soc Open Sci.* 4(10):171095.
- Hynson NA, et al. 2013. The physiological ecology of mycoheterotrophy. In: Merckx VSFT, editor. *Mycoheterotrophy: the biology of plants living on fungi*. New York: Springer. p. 297–342.
- Jacquemyn H, et al. 2017. Mycorrhizal associations and trophic modes in coexisting orchids: an ecological continuum between auto- and mixotrophy. *Front Plant Sci.* 8:1497.
- Jin X-H, et al. 2014. The evolution of floral deception in *Epipactis veratrifolia* (Orchidaceae): from indirect defense to pollination. *BMC Plant Biol.* 14:63.
- Jolou T, et al. 2005. Mixotrophy in orchids: insights from a comparative study of green individuals and nonphotosynthetic individuals of *Cephalanthera damasonium*. *New Phytol.* 166(2):639–653.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler AV, Jermini LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 14(6):587–589.
- Katoh K, Rozewicki J, Yamada KD. 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* bbx 108.
- Kearse M, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647–1649.
- Kim HT, Shin C-H, Sun H, Kim J-H. 2018. Sequencing of the plastome in the leafless green mycoheterotroph *Cymbidium macrorhizon* helps us to understand an early stage of fully mycoheterotrophic plastome structure. *Plant Syst Evol.* 304(2):245–258.
- Kim HT, et al. 2015. Seven new complete plastome sequences reveal rampant independent loss of the *ndh* gene family across orchids

- and associated instability of the inverted repeat/small single-copy region boundaries. *PLoS One* 10(11):e0142215.
- Kim K-J, Lee H-L. 2004. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res.* 11(4):247–261.
- Lallemant F, Robionek A, Courty P-E, Selosse M-A. 2018. The ¹³C content of the orchid *Epipactis palustris* (L.) Crantz responds to light as in autotrophic plants. *Bot Lett.* 165:265–273.
- Lallemant F, et al. 2016. The elusive predisposition to mycoheterotrophy in Ericaceae. *New Phytol.* 212(2):314–319.
- Lallemant F, et al. 2019. Mixotrophic orchids do not use photosynthates for perennial underground organs. *New Phytol.* 221(1):12–17.
- Leake JR. 1994. The biology of myco-heterotrophic ('saprophytic') plants. *New Phytol.* 127(2):171–216.
- Lin C-S, et al. 2017. Concomitant loss of NDH complex-related genes within chloroplast and nuclear genomes in some orchids. *Plant J* 90: 994–1006.
- Luo J, et al. 2014. Comparative chloroplast genomes of photosynthetic orchids: insights into evolution of the Orchidaceae and development of molecular markers for phylogenetic applications. *PLoS One* 9(6):e99016.
- McNeal JR, Kuehl JV, Boore JL, de Pamphilis CW. 2007. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biol.* 7:57.
- Merckx VSFT, Freudenstein JV, et al. 2013. Taxonomy and classification. In: Merckx VSFT, editor. *Mycoheterotrophy: the biology of plants living on fungi*. New York: Springer. p. 19–101.
- Merckx VSFT, Mennes CB, Peay KG, Geml J. 2013. Evolution and diversification. In: Merckx VSFT, editor. *Mycoheterotrophy: the biology of plants living on fungi*. New York: Springer. p. 215–244.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274.
- Ohta T, Ina Y. 1995. Variation in synonymous substitution rates among mammalian genes and the correlation between synonymous and non-synonymous divergences. *J Mol Evol.* 41(6):717–720.
- Petersen G, Cuenca A, Seberg O. 2015. Plastome evolution in hemiparasitic mistletoes. *Genome Biol Evol.* 7(9):2520–2532.
- Petersen G, Zervas A, Pedersen HÆ, Seberg O. 2018. Genome reports: contracted genes and dwarfed plastome in mycoheterotrophic *Sciaphila thaidanica* (Triuridaceae, Pandanales). *Genome Biol Evol.* 10(3):976–981.
- Porebski S, Bailey LG, Baum BR. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol Biol Rep.* 15(1):8–15.
- R Development Core Team. 2007. R: a language and environment for statistical computing.
- Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61(3):539–542.
- Roy M, et al. 2009. Two mycoheterotrophic orchids from Thailand tropical dipterocarpacean forests associate with a broad diversity of ectomycorrhizal fungi. *BMC Biol.* 7:51.
- Roy M, et al. 2013. Why do mixotrophic plants stay green? A comparison between green and achlorophyllous orchid individuals in situ. *Ecol Monogr* 83:95–117.
- Schiebold J-I, Bidartondo MI, Lenhard F, Makiola A, Gebauer G. 2018. Exploiting mycorrhizas in broad daylight: partial mycoheterotrophy is a common nutritional strategy in meadow orchids. *J Ecol.* 106(1):168–178.
- Selosse M-A, Charpin M, Not F. 2017. Mixotrophy everywhere on land and in water: the grand écart hypothesis. *Ecol Lett.* 20(2):246–263.
- Selosse M-A, Martos F. 2014. Do chlorophyllous orchids heterotrophically use mycorrhizal fungal carbon? *Trends Plant Sci.* 19(11):683–685.
- Selosse M-A, Roy M. 2009. Green plants that feed on fungi: facts and questions about mixotrophy. *Trends Plant Sci.* 14(2):64–70.
- Selosse M-A, Weiß M, Jany J-L, Tillier A. 2002. Communities and populations of sebacinoid basidiomycetes associated with the achlorophyllous orchid *Neottia nidus-avis* (L.) L.C.M. Rich. and neighbouring tree ectomycorrhizae. *Mol Ecol.* 11(9):1831–1844.
- Smith SE, Read DJ. 2008. *Mycorrhizal symbiosis*. London: Academic Press.
- Suetsugu K, Ohta T, Tayasu I. 2018. Partial mycoheterotrophy in the leafless orchid *Cymbidium macrorhizon*. *Am J Bot.* 105(9):1595–1596.
- Tan G, et al. 2015. Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Syst Biol* 64: 778–791.
- Těšitel J, Těšitelová T, Minasiewicz J, Selosse M-A. 2018. Mixotrophy in Land Plants: why to Stay Green? *Trends Plant Sci.* 23(8):656–659.
- Těšitelová T, Těšitel J, Jersáková J, Ríhová G, Selosse M-A. 2012. Symbiotic germination capability of four *Epipactis* species (Orchidaceae) is broader than expected from adult ecology. *Am J Bot.* 99(6):1020–1032.
- Těšitelová T, et al. 2015. Two widespread green *Neottia* species (Orchidaceae) show mycorrhizal preference for Sebaciniales in various habitats and ontogenetic stages. *Mol Ecol.* 24(5):1122–1134.
- Wang W, Li H, Chen Z. 2014. Analysis of plastid and nuclear DNA data in plant phylogenetics-evaluation and improvement. *Sci China Life Sci.* 57(3):280–286.
- WCSP. 2018. World checklist of selected plant families. Facilitated by the Royal Botanic Gardens, Kew. Available from: <http://wcsp.science.kew.org/> (accessed February 17, 2018).
- Wertheim JO, et al. 2015. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol.* 32(3):820–832.
- Wicke S, Naumann J. 2018. Molecular evolution of plastid genomes in parasitic flowering plants. In: Chaw S-M, Jansen RK, editors. *Advances in botanical research*. Vol. 85. Cambridge: Academic Press. p. 315–347.
- Wicke S, Schäferhoff B, dePamphilis CW, Müller KF. 2014. Disproportional plastome-wide increase of substitution rates and relaxed purifying selection in genes of carnivorous Lentibulariaceae. *Mol Biol Evol.* 31(3):529–545.
- Wicke S, et al. 2016. Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proc Natl Acad Sci U S A.* 113(32):9045–9050.
- Wright ES. 2015. DECIPHER: harnessing local sequence context to improve protein multiple sequence alignment. *BMC Bioinformatics* 16:322.
- Wright ES. 2016. Using DECIPHER v2.0 to analyze big biological sequence data in R. *R J.* 8(1):352–359.
- Xiang X-G, et al. 2012. Phylogenetic placement of the enigmatic orchid genera *Thaia* and *Tangtsinia*: evidence from molecular and morphological characters. *Taxon* 61(1):45–54.
- Yagame T, et al. 2016. Fungal partner shifts during the evolution of mycoheterotrophy in *Neottia*. *Am J Bot.* 103(9):1630–1641.
- Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol.* 24(8):1586–1591.
- Zhou T, Jin X. 2018. Molecular systematics and the evolution of mycoheterotrophy of tribe Neottieae (Orchidaceae, Epidendroideae). *PhytoKeys* 94:39–49.

Associate editor: Dennis Lavrov

5.4 Article 4

Three-year pot culture of *Epipactis helleborine* reveals autotrophic survival, without mycorrhizal networks, in a mixotrophic species

Michał May; Marcin Jąkałski; Alžběta Novotná; Jennifer Dietel; Manfred Ayasse; Félix Lallemand; Tomáš Figura; Julita Minasiewicz; Marc-André Selosse

MYCORRHIZA

(2020) 30:51–61

<https://doi.org/10.1007/s00572-020-00932-4>

In this work cited as (May et al., 2020)

My contribution in presented work involved:

- Genetic material isolation
- Fungal barcoding
- Investigation process

Uniwersytet Gdański
Wydział Biologii
Ul. Wita Stwosza 59
80-308, Gdańsk

data

OŚWIADCZENIE AUTORA O UDZIALE W PUBLIKACJI

Tytuł publikacji: *"Three-year pot culture of Epipactis helleborine reveals autotrophic survival, without mycorrhizal networks, in a mixotrophic species"*

Imię autora: Michał May

Oświadczam, że w wymienionym wyżej artykule, opublikowanym w *Mycorrhiza* (2020, 30:51–61; doi: 10.1007/s00572-020-00932-4) mój udział obejmował:

- Przygotowanie manuskryptu
- Izolacja DNA z materiału roślinnego
- Analizy statystyczne danych
- Barcoding i analiza danych



.....
podpis autora

Uniwersytet Gdański
Wydział Biologii
Ul. Wita Stwosza 59
80-308, Gdańsk

data

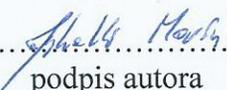
OŚWIADCZENIE AUTORA O UDZIALE W PUBLIKACJI

Tytuł publikacji: "Three-year pot culture of *Epipactis helleborine* reveals autotrophic survival, without mycorrhizal networks, in a mixotrophic species"

Imię i nazwisko autora: Marcin Jąkałski

Oświadczam, że w wymienionym wyżej artykule, opublikowanym w *Mycorrhiza* (2020, 30:51–61; doi: 10.1007/s00572-020-00932-4) mój udział obejmował:

- metabarcoding sekwencji grzybowych
- redagowanie manuskryptu

.....

.....
podpis autora

University of South Bohemia
Faculty of Science
Branišovská 31, 370 05
České Budějovice, Czech Republic

date 20/2/2022

AUTHOR'S CONTRIBUTION STATEMENT

Work title: "Three-year pot culture of *Epipactis helleborine* reveals autotrophic survival, without mycorrhizal networks, in a mixotrophic species"

Author: Alžběta Novotná

I declare that my contribution in article mentioned above and published in *Mycorrhiza* (2020, 30:51–61; doi: 10.1007/s00572-020-00932-4) included:

- DNA isolation

Novotná Alžběta

Author's signature

Institute of Evolutionary Ecology and Conservation Genomics
University of Ulm
Albert-Einstein Allee 11
D-89081 Ulm, Germany

date

AUTHOR'S CONTRIBUTION STATEMENT

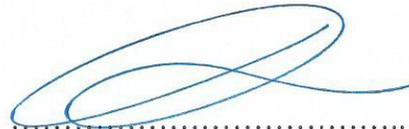
Work title: "Three-year pot culture of *Epipactis helleborine* reveals autotrophic survival, without mycorrhizal networks, in a mixotrophic species"

Author: Jennifer Dietel

Corresponding author: Marc-André Selosse

Due to author's unavailability, as a corresponding author I declare that co-author Jennifer Dietel's contribution to the article mentioned above and published in *Mycorrhiza* (2020, 30:51–61; doi: 10.1007/s00572-020-00932-4) included:

- Research material acquisition
- Plant culture and sampling



.....
Corresponding author's signature

M.A. Selosse .

Institute of Evolutionary Ecology and Conservation Genomics
University of Ulm
Albert-Einstein Allee 11
D-89081 Ulm, Germany

date Ulm, 14.3.2022

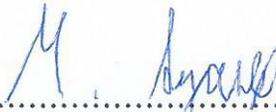
AUTHOR'S CONTRIBUTION STATEMENT

Work title: "Three-year pot culture of *Epipactis helleborine* reveals autotrophic survival, without mycorrhizal networks, in a mixotrophic species"

Author: Manfred Ayasse

I declare that my contribution in article mentioned above and published in *Mycorrhiza* (2020, 30:51–61; doi: 10.1007/s00572-020-00932-4) included:

- Designing the sampling program for *E. helleborine*
- Maintaining the culture facility


.....
Author's signature

Muséum National d'Histoire Naturelle
Institut de Systématique, Evolution, Biodiversité (ISYEB)
Muséum national d'Histoire naturelle
CNRS, Sorbonne Université, EPHE,
CP 39, 57 rue Cuvier
F-75005 Paris, France

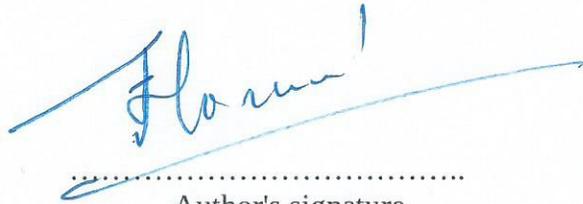
date

AUTHOR'S CONTRIBUTION STATEMENT

Work title: "Three-year pot culture of *Epipactis helleborine* reveals autotrophic survival, without mycorrhizal networks, in a mixotrophic species"

Author: Félix Lallemand

I declare that my contribution in article mentioned above and published in *Mycorrhiza* (2020, 30:51–61; doi: 10.1007/s00572-020-00932-4) included:
- Isotopic content data generation and analysis



.....
Author's signature

Department of Experimental Plant Biology
Faculty of Science, Charles University
Viničná 5, 128 44
Prague, Czech Republic

date 13.9.2021.....

AUTHOR'S CONTRIBUTION STATEMENT

Work title: "Three-year pot culture of *Epipactis helleborine* reveals autotrophic survival, without mycorrhizal networks, in a mixotrophic species"

Author: Tomáš Figura

I declare that my contribution in article mentioned above and published in *Mycorrhiza* (2020, 30:51–61; doi: 10.1007/s00572-020-00932-4) included:

- Isotopic content analysis
- Field sampling
- Elenber index calculations


.....
Author's signature

Uniwersytet Gdański
Wydział Biologii
Ul. Wita Stwosza 59
80-308, Gdańsk

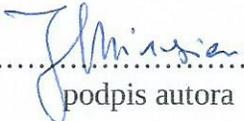
data14.03.22.....

OŚWIADCZENIE AUTORA O UDZIALE W PUBLIKACJI

Tytuł publikacji: "Three-year pot culture of *Epipactis helleborine* reveals autotrophic survival, without mycorrhizal networks, in a mixotrophic species"

Imię autora: Julita Minasiewicz

Oświadczam, że w wymienionym wyżej artykule, opublikowanym w *Mycorrhiza* (2020, 30:51–61; doi: 10.1007/s00572-020-00932-4) mój udział obejmował:
- Barcoding i analiza danych


.....
podpis autora

Muséum National d'Histoire Naturelle
Institut de Systématique, Evolution, Biodiversité (ISYEB)
Muséum national d'Histoire naturelle
CNRS, Sorbonne Université, EPHE,
CP 39, 57 rue Cuvier
F-75005 Paris, France

date

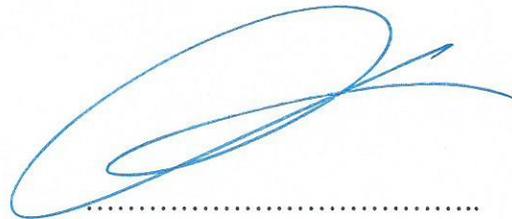
AUTHOR'S CONTRIBUTION STATEMENT

Work title: "Three-year pot culture of *Epipactis helleborine* reveals autotrophic survival, without mycorrhizal networks, in a mixotrophic species"

Author: Marc-André Selosse

I declare that my contribution in article mentioned above and published in *Mycorrhiza* (2020, 30:51–61; doi: 10.1007/s00572-020-00932-4) included:

- Experimental concept and design
- Writing and editing the manuscript
- Supervision of research and publication process
- Acting as corresponding author



.....
Author's signature

M.A. Selosse



Three-year pot culture of *Epipactis helleborine* reveals autotrophic survival, without mycorrhizal networks, in a mixotrophic species

Michał May¹ · Marcin Jąkałski¹ · Alžběta Novotná^{1,2} · Jennifer Dietel³ · Manfred Ayasse³ · Félix Lallemand⁴ · Tomáš Figura^{4,5} · Julita Minasiewicz¹ · Marc-André Selosse^{1,4}

Received: 5 October 2019 / Accepted: 10 January 2020 / Published online: 21 January 2020
© The Author(s) 2020

Abstract

Some mixotrophic plants from temperate forests use the mycorrhizal fungi colonizing their roots as a carbon source to supplement their photosynthesis. These fungi are also mycorrhizal on surrounding trees, from which they transfer carbon to mixotrophic plants. These plants are thus reputed difficult to transplant, even when their protection requires it. Here, we take profit of a successful ex situ pot cultivation over 1 to 3 years of the mixotrophic orchid *Epipactis helleborine* to investigate its mycorrhizal and nutrition status. Firstly, compared with surrounding autotrophic plants, it did not display the higher N content and higher isotopic (¹³C and ¹⁵N) abundance that normally feature mixotrophic orchids because they incorporate N-, ¹³C-, and ¹⁵N-rich fungal biomass. Second, fungal barcoding by next-generation sequencing revealed that the proportion of ectomycorrhizal fungi (expressed as percentage of the total number of either reads or operational taxonomic units) was unusually low compared with *E. helleborine* growing in situ: instead, we found a high percentage of rhizoctonias, the usual mycorrhizal partners of autotrophic orchids. Altogether, this supports autotrophic survival. Added to the recently published evidence that plastid genomes of mixotrophic orchids have intact photosynthetic genes, this suggests that at least some of them have abilities for autotrophy. This adds to the ecological plasticity of mixotrophic plants, and may allow some reversion to autotrophy in their evolution.

Keywords Mycoheterotrophy · Mycorrhizae · Orchid transplantation · Rhizoctonia · Stable isotopes · ¹³C · ¹⁵N

Introduction

Some plants from temperate forests display a mixotrophic nutrition that relies on both their photosynthates and resources extracted from the mycorrhizal fungi colonizing their roots

(for reviews, see (Selosse and Roy 2009; Selosse et al. 2016; Jacquemyn and Merckx 2019). These plants belong to the orchid (e.g., from the Neottieae tribe) and *Ericaceae* families and are also called partial mycoheterotrophs, because the heterotrophic nutrition using carbon from mycorrhizal fungi is called mycoheterotrophy (Hynson et al. 2013). Such mixotrophic plants rely on the ability of mycorrhizal fungi to establish networks between plants of different species, due to a low mycorrhizal specificity: the so-called mycorrhizal networks (Selosse et al. 2006; Simard et al. 2012) allow fungi to gain nutrients on some plants and to deliver part of it to others. Mixotrophic temperate plants rely on the network formed by ectomycorrhizal fungi, which also associate with surrounding trees (Selosse and Roy 2009).

Five lines of evidence demonstrated this mixotrophic status. Firstly, the mixotrophic plants shifted from the usual mycorrhizal fungi of their respective family (e.g., the rhizoctonias in orchids; (Dearnaley et al. 2012)) to ectomycorrhizal taxa, establishing mycorrhizal networks with nearby trees (Hynson et al. 2013). Secondly, the (hitherto elusive) compounds provided by ectomycorrhizal fungi to mixotrophs is naturally

✉ Marc-André Selosse
ma.selosse@wanadoo.fr

¹ Faculty of Biology, University of Gdańsk, ul. Wita Stwosza 59, 80-308 Gdańsk, Poland

² Present address: Faculty of Science, University of South Bohemia, Branišovská 31, 370 05 České Budějovice, Czech Republic

³ Institute of Evolutionary Ecology and Conservation Genomics, University of Ulm, Albert-Einstein Allee 11, D-89081 Ulm, Germany

⁴ Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, CP 39, 57 rue Cuvier, F-75005 Paris, France

⁵ Department of Experimental Plant Biology, Faculty of Science, Charles University, Viničná 5, 128 44 Prague, Czech Republic

enriched in ^{13}C as compared with photosynthates, and the biomass of mixotrophic plants is thus enriched in ^{13}C compared with that of autotrophic plants (Gebauer and Meyer 2003; Bidartondo et al. 2004; Julou et al. 2005; Hynson et al. 2013); note that the biomass acquired by mycoheterotrophy is often also enriched in ^{15}N). Importantly, ^{13}C abundance in mixotrophs can be used to calculate the percentage of mycoheterotrophy in each organ and season (Gebauer and Meyer 2003; Gonneau et al. 2014) or in different light environments (Preiss et al. 2010; Gonneau et al. 2014). Thirdly, beyond isotopic evidence of a raw C flow from the fungus to the orchid, the survival of rare achlorophyllous variants, in mixotrophic *Neottiae* orchids at least, demonstrates a net flow in favor of the plant: these achlorophyllous variants survive well (Selosse et al. 2004; Julou et al. 2005; Lewis 2015; Shefferson et al. 2016), but produce a limited fruit set (Salmia 1989; Roy et al. 2013; Gonneau et al. 2014). Fourthly, green mixotrophic individuals display a limited photosynthesis rate, even under the compensation point (where photosynthesis equals respiration), due to the shade of the tree canopy (Julou et al. 2005) or to limited intrinsic photosynthetic abilities (Girlanda et al. 2006). Fifthly, mixotrophs often contain more nitrogen (N) as compared with autotrophic, non-N fixing plants, probably because (i) plant respiration eliminates C and thus concentrates N from the biomass received from fungi, as described in heterotrophs, and/or (ii) the fungal biomass is likely already N-rich (Abadie et al. 2006; Hynson et al. 2013).

Recently, a split use of the two (photosynthetic and mycoheterotrophic) resources by mixotrophic *Neottiae* orchids was revealed, based on the variations of ^{13}C abundance in various organs and labeling of photosynthates in situ. Photosynthates produced after leaf expansion are mostly used for aerial parts, leaves, and fruits which are green before ripening (Roy et al. 2013; Bellino et al. 2014). Mycoheterotrophic resources are mostly used for underground roots and rhizomes (Gonneau et al. 2014): see Fig. 5 therein), as well as for elaboration of starch reserves (Lallemand et al. 2019a). This explains why achlorophyllous variants produce fewer seeds (Roy et al. 2013), but have good rhizome survival (Shefferson et al. 2016).

Mixotrophic orchids thus strongly depend on mycorrhizal networks for survival, and these results in some difficulties in transplantation attempts, as reported by Sadovsky (1965) for mixotrophic *Neottiae*. This puts constraints on protection and transplantations to save populations menaced by changes in land use. Yet, there are seldom examples of successful transplantation: for example, Delforge (2016, 2017) reports that two *Epipactis helleborine* individuals (*Neottiae* tribe) survived transplantation to a forest edge environment the first year and even flowered. While one individual then disappeared, the other one persisted over 6 years at least (Delforge 2016, 2017). Furthermore, some commercial

nurseries sell mixotrophic *Epipactis* spp. grown in pots, such as *E. helleborine*, although every time we accessed these products ($n = 2$) they turned out to belong to *E. palustris*, a related but autotrophic species (Lallemand et al. 2018). The *Epipactis helleborine* orchid species usually harbors a dominance of ectomycorrhizal fungi in its roots (Bidartondo et al. 2004; Ogura-Tsujita and Yukawa 2008; Těšitelová et al. 2012; Jacquemyn et al. 2016; Jacquemyn and Merckx 2019) and, from many isotopic data, largely relies on mycoheterotrophy for its rhizome survival and growth of young shoots (Gebauer and Meyer 2003; Gonneau et al. 2014; Schiebold et al. 2017; Lallemand et al. 2019a); Xing et al. 2019), so that successful transplantation and pot culture appear unexpected.

Here, we used a common garden growth experiment to investigate the stability of the phenotype of various morphologically distinct subspecies of *E. helleborine* after transplantation (see Delforge 2016, for a review of these subspecies; the taxonomic outcome of this experiment will be reported elsewhere). Transplanted *E. helleborine* were successfully grown in pots placed in a common garden, where the absence of a mycorrhizal network prompted the questions of (1) their mycorrhizal associates after transplantation, and (2) their level of autotrophy. Using, respectively, metabarcoding methods to identify the fungal community in roots and isotopic and N abundance to characterize the autotrophy level, we evidence here autotrophic survival of the mixotrophic *E. helleborine*.

Material and methods

E. helleborine culture ex situ

The investigated plants were harvested with 2 L of undisturbed soil surrounding each plant, in 2013 or 2015 (Table 1). Their forests of origin were mixed but dominated by *Fagus sylvatica*, with a dense canopy as is typical for the ecology of this species in Central Europe (Těšitelová et al. 2012). The plants belonged to three different subspecies of *E. helleborine* (see Table 1 and Delforge 2016, for a review of these subspecies and their debated taxonomic status). After cutting two long roots, the plants were individually placed in square pots (18 × 18 cm, height 20 cm; Fig. 1) filled with the soil collected at the same time as the orchids. After potting, orchids were deposited at a propagation bed of the Botanical Garden of the University of Ulm on a 5-cm layer of sand in order to prevent waterlogging. Pots were moved weekly to limit the ability of local soil fungi to establish permanent links with the plants (indeed, some ectomycorrhizal trees grow at a distance > 5 m in the garden). They were watered daily and put below a large grid covered with an aluminum top in order to provide shadier conditions (distance to the top of highest orchids was 75 cm). Despite shading, plants grew in a luminous conditions based on Ellenberg's indicator value ($7.0 \pm$

Table 1 Origin and sampling of the *E. helleborine* individuals from the Botanical Garden of the University of Ulm investigated in this study (sampling in 2016)

Transplantation to the Botanical Garden Ulm	Sub-species	Sampled individuals*	Forest of origin	Geocodes
June 2015	<i>E. helleborine</i> type	0/1	Ulm	48° 24' 05" N 09° 55' 05" E
July 2013	<i>E. helleborine minor</i>	5/1	Königsbronn	48° 44' 10" N 10° 05' 31" E
July 2013	<i>E. helleborine</i> type	5/1	Königsbronn	48° 44' 10" N 10° 05' 31" E
July 2013	<i>E. helleborine moratoria</i>	5/1	Ulm	48° 24' 05" N 09° 55' 05" E

*Number of individuals sampled for leaf isotopic abundances/number of individuals sacrificed for fungal barcoding of root fungi and mycorrhiza isotopic abundances

0.63; mean \pm confidence interval) calculated on the basis of the plants spontaneously growing at the same place and in the same conditions (*Brassica napus*, *Sinapis arvensis*, *Echinacea purpurea*, *E. pallida*, *Heracleum sphondylium*, *Lythrum salicaria*, *Agrimonia eupatoria*, *Campanula trachelium*, *C. persicifolia*, *C. medium*, *Taraxacum officinale*, *Leontodon autumnalis*, *Malva sylvestris*, *M. moschata*, *Verbascum densiflorum*, *V. phlomoides*, *Geranium sanguineum*, *Plantago major*). Ellenberg's indicator value represents the preference of individual species, based on empirical field observations in Central Europe, ranging from 1 (deep shade) to 9 (full sunlight; Ellenberg et al. 1991).



Fig. 1 Pot cultures of *E. helleborine* individuals investigated in this study (*E. helleborine* sampled in 2015 near Ulm, pictured July 2017)

Survival rate after transplantation in the garden was ca. 60%, but surviving plants flowered in all years after sampling (vegetative and reproductive descriptions will be reported in a separate paper investigating the three different subspecies of *E. helleborine*).

Stable isotope analyses

We sampled about 1 g (fresh weight) of leaves from each *E. helleborine* individual in July 2016, 3 years after their transplantation to pot culture (Table 1). We also added for these analyses mycorrhizal roots of the four individuals sampled for fungal barcoding (see below and Table 1; two root pieces per individual). As a reference for autotrophic plants, we used weeds growing in the same pots, i.e., an unidentified *Arabidopsis* species, *Plantago major*, and *Taraxacum sp.* The two latter species form arbuscular mycorrhizas while the first one is not mycorrhizal but harbors various root endophytes as do other Brassicaceae (Almario et al. 2017). Six replicates from independent plants were sampled for each species. To minimize environmental variations influencing ^{13}C abundance, all leaf samples were collected at a similar light level and at a similar distance from the pot soil, i.e., under the grid covered with aluminum. Samples were ground in 2-mL Eppendorf tubes in a ball mill MM200 (Retsch GmbH, Haan, Germany) and analyzed for total N concentration, as well as $^{13}\text{C}/^{12}\text{C}$ and $^{15}\text{N}/^{14}\text{N}$ ratios using an elemental analyzer (EA) coupled to a ThermoFinnigan DeltaV Advantage Continuous-Flow Isotope-ratio mass spectrometer, and expressed as δ -values (Hynson et al. 2013). Isotope values were calibrated using internal calibrated standards (EDTA and ammonium oxalate). The standard deviations of the replicated standard samples were 0.024‰ for ^{13}C and 0.199‰ for ^{15}N . Statistical analyses were performed using R environment for statistical computing (R Core Team 2015). Analysis of variance (ANOVA) was used to evaluate differences in mean $\delta^{13}\text{C}$, $\delta^{15}\text{N}$, and % N among species from a given site

(function `aov {stats}`). The alpha type I error threshold was set at 0.05.

Identification of *E. helleborine* root fungi

For metabarcoding of root fungi by next-generation sequencing (NGS), roots of four plants (see Table 1; i.e., two plants from each site of origin) were harvested in 2016, i.e., 1 or 3 years after transplantation to pot culture (Table 1). This sampling was intended to limit damage in the common garden growth experiment. Roots were screened for mycorrhizal colonization, and 4 independent colonized root pieces were submitted to DNA extraction for the subsequent assessment of fungal communities as in Schneider-Maunoury et al. (2018, 2019). We amplified the ITS2 region by using the two general primer pairs ITS3/ITS4-OF and ITS86-F/ITS4, as such a choice of primers covers a large fungal set, including orchid mycorrhizal fungi (Waud et al. 2014; Jacquemyn et al. 2016), but not necessarily arbuscular mycorrhizal ones, known to be underrepresented in orchids. The four pooled amplicon libraries, generated in a PCR reaction (see Schneider-Maunoury et al. (2018) for detailed parameters), were sequenced with an Ion Torrent sequencer (Life Technologies, Carlsbad, USA). In brief, the downstream analyses included at first a R-OTU (Reference Operational Taxonomic Unit) database generation utilizing the full-length amplicons only, i.e., reads containing both the ITS86-F and ITS4 primers, in both types of amplicons (the primer pair ITS86-F/ITS4 resides within the stretch amplified by the ITS3/ITS4-OF pair), allowing a 25% error rate for primer recognition and a minimum required length of 200 bp. Trimmed sequences were clustered into R-OTUs with SWARM (Mahé et al. 2014), followed by removal of singletons, as well as chimera removal with UCHIME (Edgar et al. 2011) against the UNITE database (Kõljalg et al. 2013). Next, reads from the original amplicon sets were extracted and trimmed if they contained the ITS86F and ITS4 primers, and assigned to R-OTUs using BLASTN (Altschul et al. 1990) with a 97% similarity threshold. Assigning a taxonomy to each R-OTU was finally accomplished by comparing the representative sequences of each R-OTU to the UNITE reference database using BLASTN with a 90% similarity threshold. The above-described steps were carried out using selected scripts from the QIIME package v1.9.1 (Caporaso et al. 2010), as well as home-made scripts. Representative sequences for each mycorrhizal OTU found in this study were submitted to GenBank under accession numbers MN459665–MN459894. OTUs were manually screened for possible orchid mycorrhizal families based on Dearnaley et al. (2012) and information of previously detected mycorrhizal fungi from the roots, germinating seeds and protocorms of various *Epipactis* species (Bidartondo et al. 2004; Selosse et al. 2004; Ogura-Tsujita and Yukawa 2008; Těšitelová et al. 2012; Jacquemyn et al. 2016; Jacquemyn and

Merckx 2019); we also included all potentially ectomycorrhizal fungi according to (Tedersoo et al. 2010; 2013). Analysis was restricted to these taxa. We tested the null hypothesis of no difference in the proportions of rhizoctonia and non-rhizoctonia fungi sequences and OTUs number among four groups represented by data from the present study, these from Těšitelová et al. (2012) and these from Jacquemyn et al. (2016 and 2019), i.e., whenever format of published data allowed respective comparisons: for this, we performed the chi-square test with Yate's correction followed by a pairwise proportional test with the Bonferroni correction. Proportions instead of raw sequences and OTU number data were applied in the calculation to normalize results between the studies.

Results

Stable isotope analyses

The leaf isotopic abundance in ^{13}C and ^{15}N of the three subspecies of *E. helleborine* (Fig. 2a) did not differ significantly from those of the reference autotrophic *Arabidopsis* sp., *Plantago major*, and *Taraxacum* sp. growing in the same pots and conditions, whatever the sub-species (see statistics in caption of Fig. 2a). Mycorrhizal roots displayed the same ^{13}C abundance as leaves (Fig. 2b). The average total N content of *E. helleborine* was lower than that of the reference autotrophic species, significantly for *Arabidopsis* sp. and *Plantago major* (Fig. 3). Thus, neither N content nor isotopic abundances indicated any contribution of N-rich, $^{13}\text{C}/^{15}\text{N}$ -enriched biomass originating from ectomycorrhizal fungi in the aerial and root biomass of pot-cultivated *E. helleborine* individuals.

Identification of *E. helleborine* root fungi

The quality-filtered pyrosequencing data set comprised 584 OTUs represented by 613,118 sequences. After analysis, 88.9% of the total number of sequences (544,929 sequences, 454 OTUs) could be assigned to Ascomycota and Basidiomycota, and a relatively large representation of Glomeromycota (arbuscular mycorrhizal fungi; 11,646 sequences, 63 OTUs, 10.7% of all fungal OTUs) were recovered. Putatively orchid mycorrhizal according to Dearnaley et al. (2012) and/or ectomycorrhizal taxa covered 167 OTUs (402,272 sequences, 65.6%). Among these, we found the usual fungal associates of autotrophic orchids, the so-called rhizoctonias: from the three rhizoctonia families, Ceratobasidiaceae were ubiquitous and highly abundant (198,383 sequences, 49.3% of all sequences in this category; in all, 49 OTUs; Fig. 4); Serendipitaceae (1 OTU) occurred in 3 plants (but reached high abundance in only one of these); no Tulasnellaceae was found. Several ectomycorrhizal clades potentially mycorrhizal on

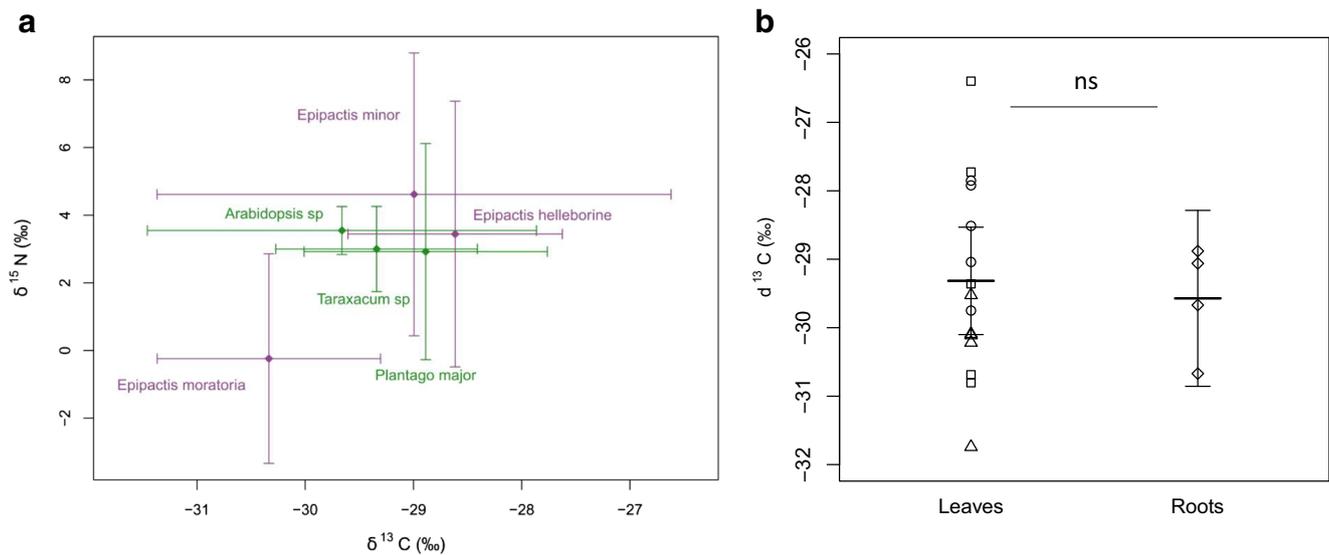


Fig. 2 Isotopic abundance of *E. helleborine* individuals (violet; $n = 5$ replicates for leaves and $n = 4$ for roots) and surrounding autotrophic weeds (green; $n = 6$ replicates). **a** $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values of leaves (distinguishing the 3 sub-species). **b** $\delta^{13}\text{C}$ of roots versus leaves of *E. helleborine* (all sub-species pooled). Bars represent the 95%

confidence intervals of the mean for each species. No significant difference was found between species for $\delta^{13}\text{C}$ nor $\delta^{15}\text{N}$ (ANOVA; $F = 1.50$ and $p = 0.232$ for ^{13}C ; $F = 2.26$ and $p = 0.086$ for ^{15}N), or between *E. helleborine* organs (to a Tukey honest significant difference test; $p = 0.65$)

E. helleborine were recorded. Helotiales (76,164 sequences, 18.9%) and Herpotrichellaceae (33,708 sequences, 8.4%) were ubiquitous, although with different abundances from one plant to another, while other taxa occurred on some plants only (Fig. 4a), namely Tuberaceae (*Tuber anniae*, 14,380 sequences, 3.6%, from 2 plants), Pyronemataceae (67,960 sequences, 16.9%, from 4 plants), Sebacinaceae, *Inocybe* and *Cortinarius* (for these three taxa: 64 sequences, 0.01%, from 3 plants; Fig. 4). The individuals transplanted to pot more recently (1 year of cultivation; column 1 in Fig. 4a) revealed more abundant ectomycorrhizal fungi than the ones cultivated for 3 years (columns 2–4; respectively 70.65% of all sequences versus 48.18% on average): *Pyronemataceae* dominated in its fungal community (12 OTUs and 24,681 sequences).

On average, the proportion of rhizoctonias found in this study was significantly higher than that of other available studies (Table 2) calculated as sequence proportions ($\chi^2(0.05,3) = 161.07$; $p < 0.0001$) or as OTU proportions ($\chi^2(0.05,2) = 39.37$ $p < 0.001$).

Discussion

We observed 1- to 3-year survival in pots for *E. helleborine*, with normal development and flowering (Fig. 1; developmental traits will be reported later in a comparative study of *E. helleborine* subspecies) that correlates with (i) an unusually high abundance of rhizoctonias and (ii) isotopic and N signatures that do not differ from nearby autotrophic plants.

Fig. 3 Total N content of leaves of *E. helleborine* individuals ($n = 5$ replicates) and surrounding autotrophic weeds ($n = 6$). Different letters in brackets after the plant names indicate different content according to a Tukey honest significant difference test

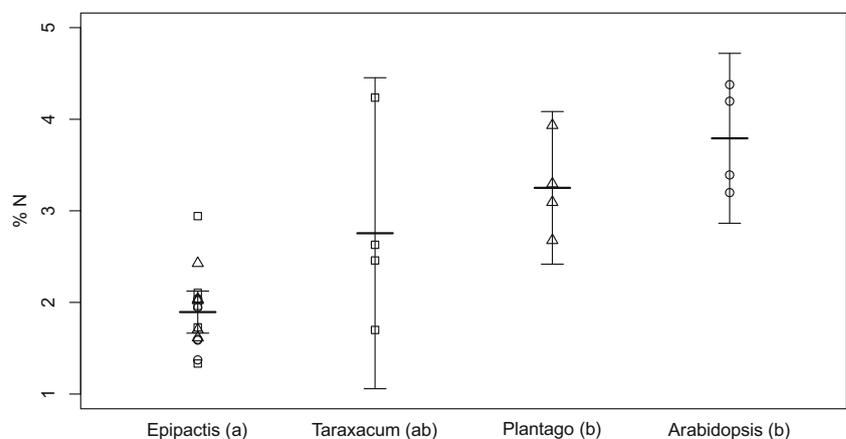
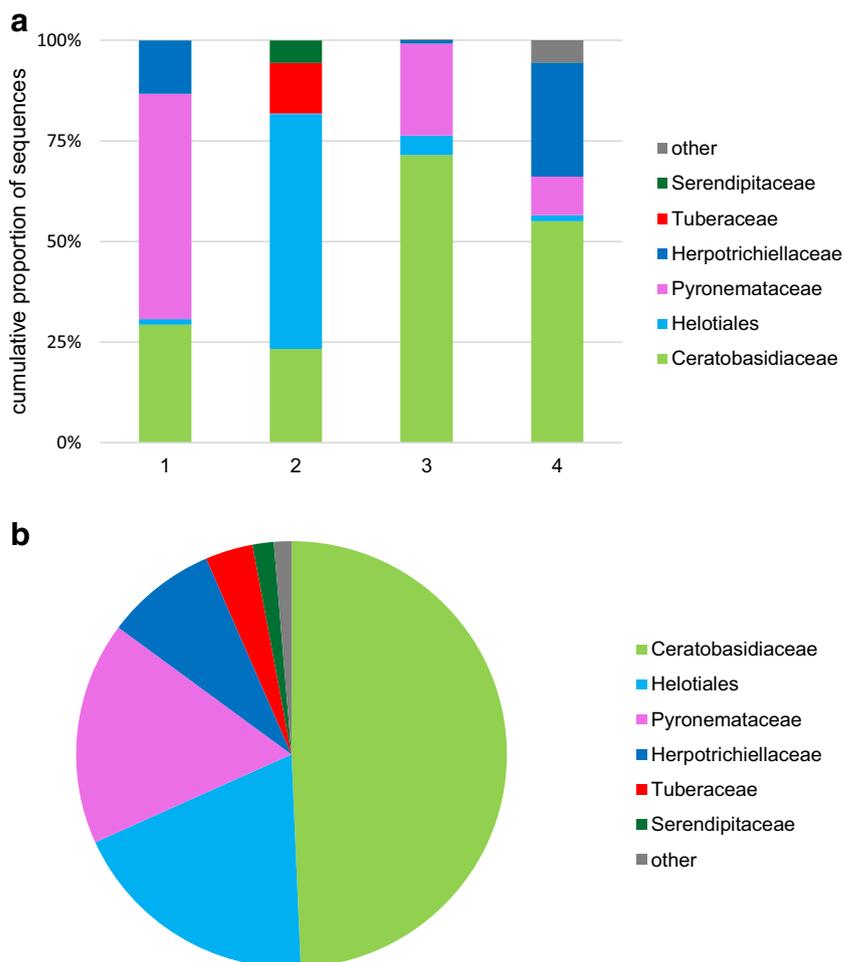


Fig. 4 Putative mycorrhizal fungal families (i.e., orchid mycorrhizal or ectomycorrhizal) detected in *E. helleborine* individuals from pot culture. Relative abundance of each mycorrhizal fungal family (i.e., orchid mycorrhizal or ectomycorrhizal) is calculated as the proportion of DNA sequences. **a** Communities for the four individuals (Table 1), namely: 1, *E. helleborine* (type) from Ulm after 1 year in pots; 2, *E. helleborine minor* from Königsbronn; 3, *E. helleborine* (type) from Königsbronn; *E. helleborine moratoria* from Ulm; the latter three were kept for 3 years in pot culture. **b** Total community found in the four barcoded *E. helleborine* individuals. The category “other” includes Sebacinaceae and unidentified Sebacinales, *Inocybe* and *Cortinarius*



Although this short-term experiment may reflect survival rather than a sustainable niche (i.e., we do not assess average life expectancy), we discuss these observations in terms of mycorrhizal interaction and physiology of *E. helleborine*, in the general framework of evolution and biological conservation of mixotrophic plants.

Table 2 Percentage of rhizoctonias versus non-rhizoctonia ectomycorrhizal taxa including species whose exact status, endophytic or ectomycorrhizal, is debated, such as Helotiales or

Study:	Current study	Xing et al., 2019	Jacquemyn et al. 2016	Těšitelová et al. 2012
Growth conditions:	Pot culture	<i>In natura</i>	<i>In natura</i>	<i>In natura</i>
Barcoding method:	NGS	NGS	NGS	Sanger (cloning)
Rhizoctonia taxa	50.87% [29.9%]	3.30% [10.6%]	0.03%	0% [0%]
Non-rhizoctonia ecto-mycorrhizal taxa	49.13% [49.1%]	96.70% [89.4]	99.97%	100% [100%]
Statistics*	<i>a</i> [a]	<i>b</i> [b]	<i>b</i>	<i>b</i> [b]

*Column with different letters differ significantly according to the chi-square test with Yate's correction ($\chi^2 (0.05,3) = 161.07, p < 0.0001$ for percentage of reads, $\chi^2 (0.05,3) = 39.37, p < 0.001$ for percentage of OTUs), followed by pairwise proportional test with the Bonferroni correction

Mycorrhizal fungi are dominated by rhizoctonias

Pot-cultivated *E. helleborine* revealed a community of root mycorrhizal fungi (including some endophytes sensu Wilson 1995, i.e., fungi colonizing the roots loosely without forming true mycorrhizas) qualitatively (= taxonomically) close to that

Herpotrichiellaceae. The value is calculated as percentage of reads (in italics) or as percentage of OTUs (bold, in brackets), which was not calculable for the work by Jacquemyn et al. (2016)

reported from natural sites, but with striking quantitative differences. The finding of *Glomeromycota*, which normally form arbuscular mycorrhizae and are poorly targeted by the primers we used, was unexpected (but see e.g. Abadie et al. 2006): although their interaction with orchid roots is unexpected and unclear, we cannot exclude an asymptomatic colonization resulting from culture in the company of some arbuscular mycorrhizal weeds such as *Taraxacum* and *Plantago* spp. (a non-mycorrhizal colonization described in non-arbuscular mycorrhizal; Cosme et al. 2018).

On the one hand, the main taxa found here were also reported in situ for *E. helleborine* (see Bidartondo et al. 2004; Ogura-Tsujita and Yukawa 2008; Těšitelová et al. 2012; Jacquemyn et al. 2016; 2019). This includes rhizoctonias, which occur in many autotrophic orchids (Dearnaley et al. 2012): *Ceratobasidiaceae* that abound here are also reported *in natura* in *E. helleborine*. Ectomycorrhizal ascomycetes with possible endophytic abilities are also reported *in natura* in *E. helleborine*: taxa with this dual ecology include *Tuber* spp. (Schneider-Maunoury et al. 2018, 2019), *Pyronemataceae* (Hansen et al. 2013), Helotiales (Wang et al. 2006), and *Herpotrichellaceae* (such as *Exophiala*, which include the so-called dark septate endophytes; Jumpponen 2001). Ectomycorrhizal basidiomycota, finally, are usually much more diverse *in natura* than the few taxa found in this study (e.g., Těšitelová et al. 2012; Jacquemyn et al. 2016; 2019).

On the other hand, not only the diversity but also the abundance of ectomycorrhizal and/or endophytic taxa (asco- or basidiomycetes) is very low in pot-cultivated *E. helleborine*. A comparison with samplings *in natura*, barcoded by NGS or cloning, clearly supports this (Table 2). Of course, differences in methods and choice of primers for NGS may affect this comparison: however, our unpublished data set on *E. helleborine* from four forests sites in Europe (M.-A. Selosse, E. Delannoy & J. Minasiewicz, unpubl; 15 individuals barcoded with exactly the same methods of analysis as in this study) revealed 99.1% of ectomycorrhizal/endophytic fungal sequences (difference with the current data significant, $\chi^2(0.05,1) = 64.9, p < 0.05$), representing 93.27% of the relative number of OTUs ($\chi^2(0.05,1) = 16.4, p < 0.001$). Thus, the abundance of rhizoctonias, especially *Ceratobasidiaceae*, is unusually high in pot cultures (Table 2). Unfortunately, the composition of the root mycorrhizal community of the *E. helleborine* populations of origin (Table 1) at the time of sampling, which may predispose to this composition, remains unknown.

The finding of few potentially ectomycorrhizal taxa is unexpected in these pots where C-providing plant hosts are lacking. Indeed, mixotrophic orchids are unlikely to give them carbon, and instead even exploit them (see discussion below). We do not believe that contamination explains our data, but we consider three non-exclusive possibilities. Firstly,

ectomycorrhizal taxa may be surviving here, perhaps declining over time: indeed, they are more numerous in the plants transplanted 1 year before (especially *Pyronemataceae*), but a firm conclusion cannot be drawn from this single plant. Secondly, ectomycorrhizal taxa may colonize the pot from the soil, since some ectomycorrhizal trees exist at some distance in the surroundings: however, we did not see any direct contact of pots with the soil, which was limited by (i) a layer of sand and (ii) weekly moving of the pots. Yet, we cannot exclude colonization by transient contacts reaching the orchid roots. Thirdly, there is increasing evidence that several ectomycorrhizal fungi, beyond the ascomycetous taxa mentioned above, also have endophytic abilities, i.e., colonize the roots of non-ectomycorrhizal plants in a loose way (for indirect evidence and a review on this, see Schneider-Maunoury et al. 2018, 2019; Selosse et al. 2018). This especially applies to the genus *Tuber* (Gryndler et al. 2014; Schneider-Maunoury et al. 2018, 2019), although this is not demonstrated for *Tuber anniae*, the North American species recently found to occur also in Europe (Wang et al. 2013) that was detected here. This is also demonstrated for *Sebacinaceae* (Selosse et al. 2009; Weiß et al. 2016), and remains pending for other ectomycorrhizal basidiomycetes (including *Inocybe*: Schneider-Maunoury et al. 2018). In this explanation, ectomycorrhizal mycelia may survive in pots by colonizing endophytically the roots of co-occurring weeds (Figs. 1 and 3) and/or *E. helleborine*.

Isotopic and N signatures of autotrophy

Whatever the reason for their presence, these root fungi did not provide detectable contribution to the biomass of the orchid since isotopic ^{13}C and ^{15}N abundances as well as total N content were similar to those of surrounding autotrophic references. A very small flow, which would not entail significant deviations in isotopic abundances, may of course occur, but the N content, lower than that of autotrophic references, argues against this. This lack of apparent mycoheterotrophy is in good agreement with the paucity of ectomycorrhizal fungi in roots and the lack of nutritional resources for the few detected, because links to surrounding ectomycorrhizal hosts are regularly disturbed.

Mixotrophy in *Neottiae* (the orchid tribe encompassing *Epipactis*) may thus display some plasticity. The ratio of aerial biomass acquired by photosynthesis and mycoheterotrophy is reported to vary with the light level in mixotrophs: increasing light positively correlates with higher contribution of photosynthesis and, thus, a lower ^{13}C content (Preiss et al. 2010; Gonneau et al. 2014). Indeed, the site of pot cultivation is rather sunny, as shown by a relatively high Ellenberg's indicator value for light (value of 7 for a maximum of 9), and there is no competition for diffuse light with similarly high plants. A similar situation is sometimes reported for *E. helleborine* in

full light, in open fields quite far away from nearest ectomycorrhizal trees (e.g., Rydlo 2008). We may have here an extreme of the continuum from mycohetero- to autotrophy, leading to undetectable mycoheterotrophy because of light level and lack of fungal resources. The possibility of autotrophic survival in *E. helleborine* is also congruent with the recent report that its plastid genome retains a full set of photosynthetic genes without any evidence of selective relaxation (Lallemand et al. 2019b) and has intact photosynthetic abilities. Moreover, the phylogenetically related mixotrophic *Limodorum abortivum* displayed some photosynthetic compensation (higher chlorophyll content and possibly higher photosynthetic activity) after experimental eradication of its fungal partners in situ (Bellino et al. 2014). Here, the lack of fungal resources in pots may have entailed a similar compensation.

In this framework, the observation of similar ^{13}C abundance in mycorrhizas and leaves is very relevant. As exemplified in the related mixotrophic *Cephalanthera damasonium* (Lallemand et al. 2019a), ^{13}C -enrichment of mycorrhizal fungi and derived plant starch normally induce higher ^{13}C abundance in mycorrhizas than in leaves. In pot-cultivated *E. helleborine*, the two organs displayed similar abundances, which can be explained if (i) the underground carbon is photosynthetic in origin and (ii) there is little biomass of ectomycorrhizal fungi, either as hyphae or delivered to the root cells. Alternatively, but less likely, we cannot exclude that root fungi provide resources from a different nutrition, e.g., saprotrophic or endophytic, which would not entail isotopic differences as compared with photosynthetic resources (see below). Finally, this autotrophic signature for underground parts is somewhat unexpected in the current model of nutrition in mixotrophic *Neottiae*, where photosynthates are mostly used for aboveground parts (Bellino et al. 2014; Gonneau et al. 2014) and migrate poorly underground (Lallemand et al. 2019a). Yet, a small flow to underground parts exists in labeling experiments (Lallemand et al. 2019a) and we speculate that, in pots, the absence (or limited presence) of fungi entails a stronger sink that directs more plant C to roots than when nutrients flow from the mycorrhizal network. This may also mean that autotrophic life in *E. helleborine* requires quite high light levels, as in this study.

Autotrophic survival in *E. helleborine* in evolution of mixotrophy

Pot-cultivated *E. helleborine* displays three features that contrast with those of mixotrophy (as described in Section 1): (i) rhizoctonias, not ectomycorrhizal fungi, dominate in their roots and since an endophytic, non-ectomycorrhizal niche can be proposed for the few existing ones, there is likely no mycorrhizal network with surrounding plants; (ii) neither ^{13}C nor ^{15}N abundances offer significant evidence for gain from

the few available ectomycorrhizal fungi; (iii) their N content is not higher than that of surrounding autotrophs. This supports autotrophy in these specific conditions, for a species otherwise reported as mixotrophic in its natural environments.

There is currently an open question on autotrophy in rhizoctonia-associated orchids: their slight isotopic difference with surrounding autotrophs (Selosse and Martos 2014) as well as their ^2H abundance (Gebauer et al. 2016; Schiebold et al. 2018) suggests that they recover some fungal biomass. Yet, the net flow, i.e., when also considering the potential reverse flow from orchid to fungus, is unknown. One ex situ experiment involving *Ceratobasidiaceae* reveals a net flow in favor of the fungus (Cameron et al. 2008), and no achlorophyllous variants of rhizoctonia-associated orchid survive in nature. Thus, the question of the net contribution of rhizoctonias remains open (see discussion in Lallemand et al. 2017) and, to facilitate reading, we provisionally consider below rhizoctonia-associated orchids as “autotrophic.” Pot-cultivated *E. helleborine* displays all the features of such orchids; even the presence of a few ectomycorrhizal fungi in the roots is reported from rhizoctonia-associated orchids (e.g., Jacquemyn et al. 2017). Our observations have consequences for plant protection and for the evolution of mixotrophy.

In terms of plant protection, ex situ conservation and transplantation is thus possible. The successful *E. helleborine* transplantation by Delforge (2016, 2017), which this author explains by a reconnection to the mycorrhizal network, may have been helped by autotrophic survival, at least transiently. We show here that absence of mycorrhizal network may even not impair survival, at least transiently. The failed transplantations by Sadovsky (1965) may be due to two factors. Firstly, not all plants, and even not all populations, may have a physiological status allowing transplantation, as shown by the 60% survival in our study (see Section 2): more on the physiology and autotrophy level of the population of origin could help, but this was overlooked in this study, which was initially designed as a taxonomic study. Secondly, not all receiving sites may be suitable, e.g., in terms of light (see above): sunnier sites may be targeted to enhance photosynthesis, although a negative trade-off is possible with desiccation in the absence of an artificial cover, as in our garden. We are far from predicting the factors allowing transplantation of mixotrophs, and more data on transplantation of various mixotrophic species (which often stay in the “gray literature”) are required to help save threatened populations.

In terms of evolution, one should remember that similar mycorrhizal, nitrogen, and isotopic features occur in closely related *Epipactis* species, such as *E. palustris* (Lallemand et al. 2018) and *E. gigantea* (Schiebold et al. 2017), which associate with *Ceratobasidiaceae* (Bidartondo et al. 2004; Jacquemyn et al. 2016, 2017). These *Epipactis* species belong to rhizoctonia-associated, putatively autotrophic orchids, and thus, an autotrophy-to-mixotrophy transition (or vice-versa)

occurred in the evolution of the genus *Epipactis*. In the framework of *Neottiae* evolution as a whole, it is still unclear in which direction this transition occurred, and a reversion from mixo- to autotrophy is also possible (Lallemand et al. 2019b, c). Autotrophic survival in *E. helleborine*, combining the intact plastid genomes of mixotrophic orchids (retaining all photosynthetic genes; (Feng et al. 2016; Lallemand et al. 2019b), makes a reversion possible, even if the physiology of mixotrophs is deeply rooted in their dependence on two carbon sources, as mentioned above (Gonneau et al. 2014; Lallemand et al. 2019a).

Our findings add to the reported plasticity of mixotrophs: they are known to survive the disappearance of chlorophyll (achlorophyllous variants; e.g., Julou et al. 2005; Gonneau et al. 2014) and to adapt to light level (Preiss et al. 2010; Gonneau et al. 2014); they now turn out to display nearly autotrophic survival in some environments. Such nutritional plasticity is a major attribute for successional species, such as *Epipactis* species, which colonize early-stage forests with variable access to light and ectomycorrhizal networks. The mycorrhizal interaction with rhizoctonias in autotrophic orchids versus ectomycorrhizal fungi in mixotrophs recently turned out to be a continuum rather than an alternative (Jacquemyn et al. 2017, and references therein); autotrophic versus mixotrophic nutrition also turns out to be a continuum, even within a given species. A quite similar statement was made in the mixotrophic *Pyrola japonica* under different light environments (Matsuda et al. 2012), in the *Ericaceae* family where mixotrophs are considered transplantable (Figura et al. 2019). Thus, autotrophic survival of mixotrophs may be sought in various phylogenetic backgrounds, even beyond orchids.

Conclusion

Some mixotrophic *E. helleborine* can be cultivated in pots, where they behave as autotrophs. Many other mixotrophic orchids may have a high trophic plasticity that was overlooked till now. They display reduced mycorrhizal colonization by the fungi that usually link them to surrounding trees, and from which they indirectly extract part of their carbon resources in forest (mixotrophy). Instead, they associate with rhizoctonia taxa that normally colonize autotrophy orchids. This further suggests that mixotrophy is ecologically, if not evolutionarily, plastic.

Acknowledgments The authors would like to acknowledge Dietrich Bergfeld and Werner Hiller, who helped us to collect and determine subspecies, two anonymous referees that improved the previous version of this manuscript, and the gardeners of the Botanical Garden Ulm. They also thank David Marsh for English editing and the Fondation de France for support to Marc-André Selosse's lab in Paris.

Funding information This study was financially supported by the Arbeitskreis heimische Orchideen Baden-Württemberg (AHO).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abadie JC, Püttsepp Ü, Gebauer G, Faccio A, Bonfante P, Selosse M-A (2006) *Cephalanthera longifolia* (Neottieae, Orchidaceae) is mixotrophic: a comparative study between green and nonphotosynthetic individuals. *Can J Bot* 84:1462–1477
- Almario J, Jeena G, Wunder J, Langen G, Zuccaro A, Coupland G, Bucher M (2017) Root-associated fungal microbiota of nonmycorrhizal *Arabidopsis thaliana* and its contribution to plant phosphorus nutrition. *Proc Natl Acad Sci U S A* 114:E9403–E9412
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Bellino A, Alfani A, Selosse MA, Guerrieri R, Borghetti M, Baldantoni D (2014) Nutritional regulation in mixotrophic plants: new insights from *Limodorum abortivum*. *Oecologia* 175:875–885
- Bidartondo MI, Burghardt B, Gebauer G, Bruns TD, Read DJ (2004) Changing partners in the dark: isotopic and molecular evidence of ectomycorrhizal liaisons between forest orchids and trees. *Proc R Soc London, Ser B* 271:1799–1806
- Cameron DD, Johnson I, Read DJ, Leake JR (2008) Giving and receiving: measuring the carbon cost of mycorrhizas in the green orchid *Goodyera repens*. *New Phytol* 180:176–184
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336
- Cosme M, Fernández I, Van der Heijden MGA, Pieterse CMJ (2018) Non-mycorrhizal plants: the exceptions that prove the rule. *Trends Plant Sci* 23:577–587
- Dearnaley JDW, Martos F, Selosse M-A (2012) Orchid mycorrhizas: molecular ecology, physiology, evolution and conservation aspects. In: Hock B (ed) *The Mycota IX: fungal associations*, 2nd edn. Springer, Berlin, pp 207–230
- Delforge P (2016) Que devient un individu robuste d'*Epipactis helleborine* (L.) Crantz après une transplantation réussie ? Implications pour le statut d'*E. helleborine* var. *minor* Engel et d'*E. helleborine*. *Nat Belges* 97:89–124
- Delforge P (2017) Que devient en 2017 un individu robuste d'*Epipactis helleborine* (L.) C RANTZ après une transplantation réussie, effectuée en 2011 ? *Nat Belges* 98:62–68
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194–2200

- Ellenberg H, Weber EH, Düll R, Wirth V, Werner W, Paulißen D (1991) Zeigerwerte von Pflanzen in Mitteleuropa. *Scr Geobot* 18:1–248
- Feng YL, Wicke S, Li JW, Han Y, Lin CS, Li DZ, Zhou TT, Huang WC, Huang LQ, Jin XH (2016) Lineage-specific reductions of plastid genomes in an orchid tribe with partially and fully mycoheterotrophic species. *Genome Biol Evol* 8:2164–2175
- Figura T, Tylová E, Šoch J, Selosse MA, Ponert J (2019) In vitro axenic germination and cultivation of mixotrophic Pyroloideae (Ericaceae) and their post-germination ontogenetic development. *Ann Bot* 123:625–639
- Gebauer G, Meyer M (2003) ^{15}N and ^{13}C natural abundance of autotrophic and myco-heterotrophic orchids provides insight into nitrogen and carbon gain from fungal association. *New Phytol* 160:209–223
- Gebauer G, Preiss K, Gebauer AC (2016) Partial mycoheterotrophy is more widespread among orchids than previously assumed. *New Phytol* 211:11–15
- Girlanda M, Selosse MA, Cafasso D, Brilli F, Delfine S, Fabbian R, Ghignone S, Pinelli P, Segreto R, Loreto F, Cozzolino S, Perotto S (2006) Inefficient photosynthesis in the Mediterranean orchid *Limodorum abortivum* is mirrored by specific association to ectomycorrhizal Russulaceae. *Mol Ecol* 15:491–504
- Gonneau C, Jersáková J, de Tredem E, Till-Bottraud I, Saarinen K, Sauve M, Roy M, Hájek T, Selosse M-A (2014) Photosynthesis in perennial mixotrophic *Epipactis* spp. (Orchidaceae) contributes more to shoot and fruit biomass than to hypogeous survival. *J Ecol* 102:1183–1194
- Gryndler M, Černá L, Bukovská P, Hršelová H, Jansa J (2014) *Tuber aestivum* association with non-host roots. *Mycorrhiza* 24:603–610
- Hansen K, Perry BA, Dranginis AW, Pfister DH (2013) A phylogeny of the highly diverse cup-fungus family Pyronemataceae (Pezizomycetes, Ascomycota) clarifies relationships and evolution of selected life history traits. *Mol Phylogenet Evol* 67:311–335
- Hynson NA, Madsen TP, Selosse M-A, Adam IKU, Ogura-Tsujita Y, Roy M, Gebauer G (2013) The physiological ecology of mycoheterotrophy. In: Merckx VSFT (ed) *Mycoheterotrophy*. Springer, New York, pp 297–342
- Jacquemyn H, Merckx VSFT (2019) Mycorrhizal symbioses and the evolution of trophic modes in plants. *J Ecol* 107:1567–1581
- Jacquemyn H, Waud M, Lievens B, Brys R (2016) Differences in mycorrhizal communities between *Epipactis palustris*, *E. helleborine* and its presumed sister species *E. neerlandica*. *Ann Bot* 118:105–114
- Jacquemyn H, Waud M, Brys R, Lallemand F, Courty PE, Robioneck A, Selosse M-A (2017) Mycorrhizal associations and trophic modes in coexisting orchids: an ecological continuum between auto- and mixotrophy. *Front Plant Sci* 8:1–12
- Julou T, Burghardt B, Gebauer G, Berveiller D, Damesin C, Selosse M-A (2005) Mixotrophy in orchids: insights from a comparative study of green individuals and nonphotosynthetic individuals of *Cephalanthera damasonium*. *New Phytol* 166:639–653
- Jumpponen A (2001) Dark septate endophytes - are they mycorrhizal? *Mycorrhiza* 11:207–211
- Köljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M, Bates ST, Bruns TD, Bengtsson-Palme J, Callaghan TM, Douglas B, Drenkhan T, Eberhardt U, Dueñas M, Grebenc T, Griffith GW, Hartmann M, Kirk PM, Kohout P, Larsson E, Lindahl BD, Lücking R, Martín MP, Matheny PB, Nguyen NH, Niskanen T, Oja J, Peay KG, Peintner U, Peterson M, Pöldmaa K, Saag L, Saar I, Schüßler A, Scott JA, Senés C, Smith ME, Suija A, Taylor DL, Telleria MT, Weiss M, Larsson KH (2013) Towards a unified paradigm for sequence-based identification of fungi. *Mol Ecol* 22:5271–5277
- Lallemand F, Puttsepp Ü, Lang M, Luud A, Courty PE, Palancade C, Selosse M-A (2017) Mixotrophy in Pyroleae (Ericaceae) from Estonian boreal forests does not vary with light or tissue age. *Ann Bot* 120:361–371
- Lallemand F, Robioneck A, Courty PE, Selosse M-A (2018) The ^{13}C content of the orchid *Epipactis palustris* (L.) Crantz responds to light as in autotrophic plants. *Bot Lett* 165:265–273
- Lallemand F, Figura T, Damesin C, Fresneau C, Griveau C, Fontaine N, Zeller B, Selosse M-A (2019a) Mixotrophic orchids do not use photosynthates for perennial underground organs. *New Phytol* 221:12–17
- Lallemand F, Logacheva M, Le Clainche I, Bérard A, Zheleznaia E, May M, Jakalski M, Delannoy É, Le Paslier M-C, Selosse M-A (2019b) Thirteen new plastid genomes from mixotrophic and autotrophic species provide insights into heterotrophy evolution in Neottieae orchids. *Genome Biol Evol*:49–66
- Lallemand F, Martin-Magniette ML, Gilard F, Gakière B, Launay-Avon A, Delannoy E, Selosse M-A (2019c) In situ transcriptomic and metabolomic study of the loss of photosynthesis in the leaves of mixotrophic plants exploiting fungi. *Plant J* 98:826–841
- Lewis L (2015) Some observations on the nomenclature of achlorophyllous forms of *Epipactis purpurata*, *E. helleborine* and *E. dumensis*. *J Eur Orchid* 47:111–122
- Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M (2014) Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ* 2014:1–13
- Matsuda Y, Shimizu S, Mori M, Ito S-I, Selosse M-A (2012) Seasonal and environmental changes of mycorrhizal associations and heterotrophy levels in mixotrophic *Pyrola japonica* (Ericaceae) growing under different light environments. *Am J Bot* 99:1177–1188
- Ogura-Tsujita Y, Yukawa T (2008) *Epipactis helleborine* shows strong mycorrhizal preference towards ectomycorrhizal fungi with contrasting geographic distributions in Japan. *Mycorrhiza* 18:331–338
- Preiss K, Adam IKU, Gebauer G (2010) Irradiance governs exploitation of fungi: fine-tuning of carbon gain by two partially mycoheterotrophic orchids. *Proc R Soc London, Ser B* 277:1333–1336
- R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>. Accessed 6 Oct 2019
- Roy M, Gonneau C, Rocheteau A, Berveiller D, Thomas JC, Damesin C, Selosse M-A (2013) Why do mixotrophic plants stay green? A comparison between green and achlorophyllous orchid individuals in situ. *Ecol Monogr* 83:95–117
- Rydlo J (2008) *Epipactis helleborine* jako polni plevel. *Muz a Současnost, Roztoky, ser natur* 23:211–218
- Sadovsky O (1965) *Orchideen im eigenen Garten*. Bayer, Landwirtschaftsverlag
- Salmia A (1989) General morphology and anatomy of chlorophyll-free and green forms of *Epipactis helleborine* (Orchidaceae). *Ann Bot Fenn* 26:95–105
- Schiebold JMI, Bidartondo MI, Karasch P, Gravendeel B, Gebauer G (2017) You are what you get from your fungi: nitrogen stable isotope patterns in *Epipactis* species. *Ann Bot* 119:1085–1095
- Schiebold JMI, Bidartondo MI, Lenhard F, Makiola A, Gebauer G (2018) Exploiting mycorrhizas in broad daylight: partial mycoheterotrophy is a common nutritional strategy in meadow orchids. *J Ecol* 106:168–178
- Schneider-Maunoury L, Leclercq S, Clément C, Covès H, Lambourdière J, Sauve M, Richard F, Selosse MA, Taschen E (2018) Is *Tuber melanosporum* colonizing the roots of herbaceous, non-ectomycorrhizal plants? *Fungal Ecol* 31:59–68
- Schneider-Maunoury L, Deveau A, Moreno M, Todesco F, Belmonto S, Murat C, Courty P-E, Jakalski M, Selosse M-A (2019) Two ectomycorrhizal truffles, *Tuber melanosporum* and *T. aestivum*, endophytically colonise roots of non-ectomycorrhizal plants in natural environments. *New Phytol* in press
- Selosse MA, Martos F (2014) Do chlorophyllous orchids heterotrophically use mycorrhizal fungal carbon? *Trends Plant Sci* 19:683–685
- Selosse MA, Roy M (2009) Green plants that feed on fungi: facts and questions about mixotrophy. *Trends Plant Sci* 14:64–70

- Selosse MA, Faccio A, Scappaticci G, Bonfante P (2004) Chlorophyllous and achlorophyllous specimens of *Epipactis microphylla* (Neottieae, Orchidaceae) are associated with ectomycorrhizal septomycetes, including truffles. *Microb Ecol* 47:416–426
- Selosse MA, Richard F, He X, Simard SW (2006) Mycorrhizal networks: des liaisons dangereuses? *Trends Ecol Evol* 21:621–628
- Selosse MA, Dubois MP, Alvarez N (2009) Do Sebaciales commonly associate with plant roots as endophytes? *Mycol Res* 113:1062–1069
- Selosse MA, Bocayuva MF, Kasuya MCM, Courty P-E (2016) Mixotrophy in mycorrhizal plants: extracting carbon from mycorrhizal networks. In: Martinn F (ed) *Molecular mycorrhizal symbiosis*. Springer, Berlin, pp 451–471
- Selosse MA, Schneider-Maunoury L, Martos F (2018) Time to re-think fungal ecology? Fungal ecological niches are often prejudged. *New Phytol* 217:968–972
- Shefferson RP, Roy M, Püttsepp Ü, Selosse M-A (2016) Demographic shifts related to mycoheterotrophy and their fitness impacts in two *Cephalanthera* species. *Ecology* 97:1452–1462
- Simard SW, Beiler KJ, Bingham MA, Deslippe JR, Philip LJ, Teste FP (2012) Mycorrhizal networks: mechanisms, ecology and modelling. *Fungal Biol Rev* 26:39–60
- Tedersoo L, Smith ME (2013) Lineages of ectomycorrhizal fungi revisited: foraging strategies and novel lineages revealed by sequences from belowground. *Fungal Biol Rev* 27:83–99
- Tedersoo L, May TW, Smith ME (2010) Ectomycorrhizal lifestyle in fungi: global diversity, distribution, and evolution of phylogenetic lineages. *Mycorrhiza* 20:217–263
- Těšitelová T, Tešitel J, Jersáková J, Říhová G, Selosse M-A (2012) Symbiotic germination capability of four *Epipactis* species (Orchidaceae) is broader than expected from adult ecology. *Am J Bot* 99:1020–1032
- Wang Z, Johnston PR, Takamatsu S, Spatafora JW, Hibbett DS (2006) Toward a phylogenetic classification of the Leotiomycetes based on rDNA data. *Mycologia* 98:1065–1075
- Wang XH, Benucci GMN, Xie XD, Bonito G, Leisola M, Liu PG, Shamekh S (2013) Morphological, mycorrhizal and molecular characterization of Finnish truffles belonging to the *Tuber anniae* species-complex. *Fungal Ecol* 6:269–280
- Waud M, Busschaert P, Ruyters S, Jacquemyn H, Lievens B (2014) Impact of primer choice on characterization of orchid mycorrhizal communities using 454 pyrosequencing. *Mol Ecol Resour* 14:679–699
- Weiß M, Waller F, Zuccaro A, Selosse M-A (2016) Sebaciales - one thousand and one interactions with land plants. *New Phytol* 211:20–40
- Wilson D (1995) Endophyte: the evolution of a term, and clarification of its use and definition. *Oikos* 73:274–276. <https://doi.org/10.2307/3545919>
- Xing X, Gao Y, Zhao Z, Waud M, Duffy KJ, Selosse M, Jakalski M, Liu N, Jacquemyn H, Guo S, Merckx V, (2019) Similarity in mycorrhizal communities associating with two widespread terrestrial orchids decays with distance. *Journal of Biogeography*

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

5.5 Article 5

The Genomic Impact of Mycoheterotrophy in Orchids

Marcin Jąkowski, Julita Minasiewicz, José Caius, **Michał May**, Marc-André Selosse,
Etienne Delannoy

FRONTIERS IN PLANT SCIENCE

12:632033

<https://doi.org/10.3389/fpls.2021.632033>

In this work cited as (Jąkowski et al., 2021)

My contribution in presented work involved:

- Research material collection
- Transcriptomic data analysis
- Debugging and testing the data analysis pipeline
- Manuscript preparation

Uniwersytet Gdański
Wydział Biologii
Ul. Wita Stwosza 59
80-308, Gdańsk

data

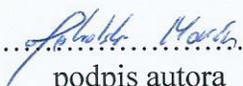
OŚWIADCZENIE AUTORA O UDZIALE W PUBLIKACJI

Tytuł publikacji: "The Genomic Impact of Mycoheterotrophy in Orchids"

Imię i nazwisko autora: Marcin Jąkałski

Oświadczam, że w wymienionym wyżej artykule, opublikowanym w *Frontiers in Plant Science* (doi: 10.3389/fpls.2021.632033) mój udział obejmował:

- Zbiór prób materiału badawczego
- Analiza danych
- Przygotowanie manuskryptu publikacji

.....

.....
podpis autora

Uniwersytet Gdański
Wydział Biologii
Ul. Wita Stwosza 59
80-308, Gdańsk

data ..14.03.2022

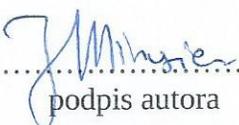
OŚWIADCZENIE AUTORA O UDZIALE W PUBLIKACJI

Tytuł publikacji: "The Genomic Impact of Mycoheterotrophy in Orchids"

Imię autora: Julita Minasiewicz

Oświadczam, że w wymienionym wyżej artykule, opublikowanym w *Frontiers in Plant Science* (doi: 10.3389/fpls.2021.632033) mój udział obejmował:

- Zbiór prób materiału badawczego
- Przygotowanie manuskryptu publikacji

.....

.....
podpis autora

Université Paris-Saclay

date04/02/2022...

.....
CNRS, INRAE, Univ Evry
Institute of Plant Sciences Paris-Saclay (IPS2), F-91405
Orsay, France

AUTHOR'S CONTRIBUTION STATEMENT

Work title: "The Genomic Impact of Mycoheterotrophy in Orchids"

Author: José Caius

I declare that my contribution in article mentioned above and published in *Frontiers in Plant Science* (doi: 10.3389/fpls.2021.632033) included:

- RNA-seq data generation



.....
Author's signature

Uniwersytet Gdański
Wydział Biologii
Ul. Wita Stwosza 59
80-308, Gdańsk

data

OŚWIADCZENIE AUTORA O UDZIALE W PUBLIKACJI

Tytuł publikacji: "The Genomic Impact of Mycoheterotrophy in Orchids"

Imię autora: Michał May

Oświadczam, że w wymienionym wyżej artykule, opublikowanym w *Frontiers in Plant Science* (doi: 10.3389/fpls.2021.632033) mój udział obejmował:

- Zbiór prób materiału badawczego
- Analiza danych transkryptomicznych
- Rozwiązywanie problemów i testowanie procedury analizy danych
- Przygotowanie manuskryptu

.....
podpis autora

Muséum National d'Histoire Naturelle
Institut de Systématique, Evolution, Biodiversité (ISYEB)
Muséum national d'Histoire naturelle
CNRS, Sorbonne Université, EPHE,
CP 39, 57 rue Cuvier
F-75005 Paris, France

date

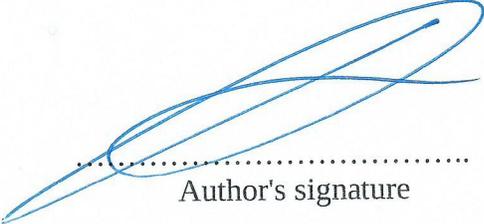
AUTHOR'S CONTRIBUTION STATEMENT

Work title: "The Genomic Impact of Mycoheterotrophy in Orchids"

Author: Marc-André Selosse

I declare that my contribution in article mentioned above and published in *Frontiers in Plant Science* (doi: 10.3389/fpls.2021.632033) included:

- Experimental concept and design
- Plant sample collection
- Supervision of research and publication process


.....
Author's signature

M.A. Selosse .

Université Paris-Saclay
CNRS, INRAE, Univ Evry
Institute of Plant Sciences Paris-Saclay (IPS2), F-91405
Orsay, France

date

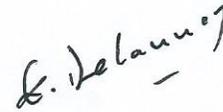
AUTHOR'S CONTRIBUTION STATEMENT

Work title: "The Genomic Impact of Mycoheterotrophy in Orchids"

Author: Etienne Delannoy

I declare that my contribution in article mentioned above and published in *Frontiers in Plant Science* (doi: 10.3389/fpls.2021.632033) included:

- Experimental concept and design
- Data analysis
- Writing and refining the manuscript
- Acting as the corresponding author in publication process



.....

Author's signature



The Genomic Impact of Mycoheterotrophy in Orchids

Marcin Jałkowski¹, Julita Minasiewicz¹, José Caius^{2,3}, Michał May¹,
Marc-André Selosse^{1,4†} and Etienne Delannoy^{2,3*†}

¹ Department of Plant Taxonomy and Nature Conservation, Faculty of Biology, University of Gdańsk, Gdańsk, Poland, ² Institute of Plant Sciences Paris-Saclay, Université Paris-Saclay, CNRS, INRAE, Univ Evry, Orsay, France, ³ Université de Paris, CNRS, INRAE, Institute of Plant Sciences Paris-Saclay, Orsay, France, ⁴ Sorbonne Université, CNRS, EPHE, Muséum National d'Histoire Naturelle, Institut de Systématique, Evolution, Biodiversité, Paris, France

OPEN ACCESS

Edited by:

Susann Wicke,
Humboldt University of Berlin,
Germany

Reviewed by:

Maria D. Logacheva,
Skolkovo Institute of Science
and Technology, Russia
Sean W. Graham,
University of British Columbia,
Canada
Craig Barrett,
West Virginia University, United States

*Correspondence:

Etienne Delannoy
etienne.delannoy@inrae.fr

†These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Systematics and Evolution,
a section of the journal
Frontiers in Plant Science

Received: 21 November 2020

Accepted: 14 May 2021

Published: 09 June 2021

Citation:

Jałkowski M, Minasiewicz J,
Caius J, May M, Selosse M-A and
Delannoy E (2021) The Genomic
Impact of Mycoheterotrophy
in Orchids.
Front. Plant Sci. 12:632033.
doi: 10.3389/fpls.2021.632033

Mycoheterotrophic plants have lost the ability to photosynthesize and obtain essential mineral and organic nutrients from associated soil fungi. Despite involving radical changes in life history traits and ecological requirements, the transition from autotrophy to mycoheterotrophy has occurred independently in many major lineages of land plants, most frequently in Orchidaceae. Yet the molecular mechanisms underlying this shift are still poorly understood. A comparison of the transcriptomes of *Epipogium aphyllum* and *Neottia nidus-avis*, two completely mycoheterotrophic orchids, to other autotrophic and mycoheterotrophic orchids showed the unexpected retention of several genes associated with photosynthetic activities. In addition to these selected retentions, the analysis of their expression profiles showed that many orthologs had inverted underground/aboveground expression ratios compared to autotrophic species. Fatty acid and amino acid biosynthesis as well as primary cell wall metabolism were among the pathways most impacted by this expression reprogramming. Our study suggests that the shift in nutritional mode from autotrophy to mycoheterotrophy remodeled the architecture of the plant metabolism but was associated primarily with function losses rather than metabolic innovations.

Keywords: mycorrhiza, photosynthesis, metabolic evolution, mycoheterotrophy, orchids, transcriptome, *Epipogium aphyllum*, *Neottia nidus-avis*

INTRODUCTION

More than 85% of vascular plants grow in association with soil fungi, forming a mycorrhizal symbiosis (Wang and Qiu, 2006; van der Heijden et al., 2015; Brundrett and Tedersoo, 2018). Thanks to this symbiosis, plant growth, and fitness are substantially improved by better mineral nutrition and increased resistance to biotic and abiotic stresses (see Jung et al., 2012 for a review). In this mutualism, the fungal partner provides mineral nutrients (e.g., water, N, P, and K) in exchange for organic compounds from the photosynthesis of the plant (see Rich et al., 2017 for a review). However, more than 500 plant species, called mycoheterotrophs, have lost their ability to photosynthesize and entirely rely on their fungal partners for mineral and organic nutrients, reversing the usual direction of the net carbon flow. The metabolic evolution associated with this switch, which is still poorly understood, has occurred in parallel at least 50 times in 17 plant families,

including at least 30 independent transitions in Orchidaceae (Merckx et al., 2009; Merckx and Freudenstein, 2010; Těšitel et al., 2018; Barrett et al., 2019). One of the characteristics that may make orchids more prone to this evolutionary change in their nutrition mode lies in their minute seeds devoid of nutritional reserves (Rasmussen, 1995). The seeds fully depend on their mycorrhizal fungi for carbon supply at germination and early developmental stages (protocorms), which are thus always mycoheterotrophic (Merckx, 2013; Dearnaley et al., 2016). Most orchid species later shift to autotrophy once photosynthesis becomes possible and establish more reciprocal exchanges with mycorrhizal fungi at adulthood. However, some species still recover carbon from the fungi at adulthood in addition to their photosynthesis (mixotrophy; Selosse and Roy, 2009), and from this nutrition some even evolved complete loss of photosynthesis and mycoheterotrophy at adulthood (Merckx et al., 2009; Dearnaley et al., 2016). This versatile relation between orchids and their mycorrhizal partners provides a useful framework for understanding the metabolic evolution resulting in mycoheterotrophy (Suetsugu et al., 2017; Lallemand et al., 2019b).

The impact of mycoheterotrophy on plant physiology can be analyzed through the changes in genomes of mycoheterotrophs compared to autotrophic relatives. As mycoheterotrophy is associated with the loss of photosynthesis, sequencing of the plastid genome has been targeted first, thanks to next-generation methods (e.g., Logacheva et al., 2014; Schelkunov et al., 2015; Lim et al., 2016; Ravin et al., 2016; Yudina et al., 2021). A common feature among plastid genomes of mycoheterotrophs is a strong reduction in size and gene content, especially with, as expected, a loss of all photosynthetic genes (Graham et al., 2017; Hadariová et al., 2018). However, the plastid genome contains only a tiny fraction of plant genes and the absence of genes from the plastid genome does not rule out the possibility that some of them were transferred into the nuclear genome, rather than lost (Bock, 2017).

Furthermore, in addition to photosynthesis, the transition to mycoheterotrophy can be expected to affect other metabolic processes, which cannot be assessed without the complete gene repertoire of all three plant genomes. Out of three published full genomes of heterotrophic plants, two belong to obligate plant parasites (Vogel et al., 2018; Yoshida et al., 2019) and one to an east Asian mycoheterotrophic orchid (*Gastrodia elata* Blume; Yuan et al., 2018). When compared to photosynthetic orchids, the genome of *G. elata* is characterized by a reduction of gene content, including the loss of most of the genes associated with photosynthesis, and the reduction of gene families involved in resistance to pathogens. At the same time, it shows an expansion of gene families that are putatively involved in the interaction with fungi (Yuan et al., 2018).

Despite the decrease in sequencing costs, the *de novo* characterization of a complete plant genome is still = expensive and tedious, especially in the case of relatively large genomes of achlorophyllous orchids, [from about 6 Gb for *Corallorhiza trifida* Chatelain to about 16 Gb for *Neottia nidus-avis* (L.) L.C.M. Rich; Pellicer and Leitch, 2020]. Another approach for studying gene content is to analyze transcriptomes. RNA-seq

focuses on the transcribed fraction of the genome, which includes the protein-coding genes. Transcriptomes of five mycoheterotrophic plants are currently available (Schelkunov et al., 2018; Leebens-Mack et al., 2019). The transcriptomes of two orchids, *Epipogium aphyllum* Sw. and *Epipogium roseum* (D. Don) Lindl., and the Ericaceae *Monotropa hypopitys* L. show a loss of the photosynthetic genes (Schelkunov et al., 2018). Surprisingly, but in accordance with results from obligate parasitic plants (Wickett et al., 2011; Chen et al., 2020), the chlorophyll synthesis pathway was mostly conserved in these plants, even if incomplete. However, transcriptome analysis only identifies the genes expressed in the tissue(s) under study, and as the previous studies of mycoheterotrophic species concentrated on the aerial part only, a fraction of the extant genes was likely missed. In addition, the missed genes include all the genes specifically expressed in the roots and mycorrhiza, which are fundamental to understanding of the mechanism of the interaction between a mycoheterotrophic plant and its fungal partners. Finally, it is the most likely that the switch to mycoheterotrophy not only results in gene losses, but also in neofunctionalizations and changes in the expression profiles of some retained genes, which are difficult or impossible to capture in simple analyses of gene repertoires.

Here, we explored the transcriptome and gene expression profiles in the mycorrhiza, stems, and flowers of the MH orchids *N. nidus-avis* and *E. aphyllum* (Figure 1). Both studied species are achlorophyllous and, like *G. elata*, belong to the orchid subfamily Epidendroideae. Despite their rarity, they have a broad Eurasian range (Hulten and Fries, 1986) and, together with *G. elata*, they represent three independent evolutionary origins of mycoheterotrophy in orchids (Merckx and Freudenstein, 2010). Their shoots have minute achlorophyllous scales and produce a few large flowers in *E. aphyllum* (Taylor and Roberts, 2011) and many small flowers in *N. nidus-avis* (Selosse, 2003). Both species are considered allogamous, producing scent and a little amount of nectar (Ziegenspeck, 1936; Claessens, 2011; Jakubska-Busse et al., 2014; Krawczyk et al., 2016) however some level of autogamy is also expected in *N. nidus-avis* (Claessens, 2011).

Considering their underground parts, *N. nidus-avis* has a clump of short and thick mycorrhizal roots growing out of a short and thin rhizome, and *E. aphyllum* forms a fleshy, dichotomously branched and rootless rhizome that hosts the fungus (Roy et al., 2009). Thus, roots of *N. nidus-avis* and the rhizome of *E. aphyllum* are not anatomically equivalent but both correspond to the underground organs where the interactions with their fungal partner are taking place.

We used RNA-seq in flowers, stems, and mycorrhizal parts (roots or rhizomes) sampled in natural forest conditions, and identified expressed gene sets in each case. In combination with published data from the *G. elata* genome, we compared the gene sets of the three mycoheterotrophic orchids to that of three autotrophic orchid species, in order to highlight the gene losses and gains associated with the switch to mycoheterotrophy in orchids. We also identified genes that are differentially expressed between the three investigated tissues. As no equivalent dataset (expression profiles per organ for the same individuals with biological replicates) was available for autotrophic orchids at



FIGURE 1 | Morphology of *Neottia nidus-avis* and *Epipogium aphyllum*. Top left: roots of *N. nidus-avis*. Bottom left: inflorescence of *N. nidus-avis*. Top right: inflorescence of *E. aphyllum* (Courtesy of Emilia Krawczyk). Bottom right: rhizome of *E. aphyllum* (Courtesy of Emilia Krawczyk).

the time of study, we compared these profiles to expression in other autotrophic non-orchid plants. This comparison suggested that, in addition to gene losses, the switch to mycoheterotrophy induced extensive expression reprogramming.

RESULTS

Sequencing, *de novo* Assembly and Functional Annotation

In total, 12 cDNA libraries from flowers, stems, and mycorrhizal roots (two replicates per tissue and species) were sequenced resulting in 304,280,061 reads for *N. nidus-avis* and 178,486,849 reads for *E. aphyllum*. After assembly and filtering of the probable contaminating transcripts, the final set of transcripts comprised 44,451 and 38,488 sequences for *N. nidus-avis* and *E. aphyllum*, respectively (**Table 1**). As expected, the fraction of contaminating contigs was much higher in the mycorrhizal samples (roots in *N. nidus-avis* and rhizome in *E. aphyllum*), and indeed almost all the contaminating transcripts were most probably of fungal origin (**Supplementary Tables 3, 4**). Thanks to the presence of a few contigs corresponding to ITS, the main fungal partners can be identified as *Inocybe cervicolor* (Pers.) Quél. and *Hebeloma incarnatum* A.H. Sm. for *E. aphyllum* and *Sebacina epigaea*

(Berk. & Broome) Bourdot & Galzin for *N. nidus-avis* as expected (McKendrick et al., 2002; Selosse et al., 2002; Roy et al., 2009).

We functionally annotated the transcripts recovered for the two studied species. Roughly 46 and 50% of the transcripts could be attributed to any annotation category in *N. nidus-avis* and *E. aphyllum*, respectively (**Supplementary Table 5 and Supplementary Data 1**).

We also assessed the completeness of the generated transcriptomes using several analyses. The BUSCO (Benchmarking Universal Single-Copy Orthologs) analysis showed 78.5 and 71% of completeness for *N. nidus-avis* and

TABLE 1 | Statistics of the final assemblies.

	<i>Neottia nidus-avis</i>	<i>Epipogium aphyllum</i>
Number of transcripts	43,451	38,488
Number of "genes"	39,731	36,275
Median/mean transcript length (bp)	675/993	555/920
Shortest/longest transcript (bp)	201/13,537	201/15,415
Transcripts over 1/10k bp length	15,391/10	11,980/9
Transcript n50 (bp)	1,506	1,480
GC %	45.56	44.26
Total assembled bases	43,160,528	35,412,792

E. aphyllum, respectively, comparable to or higher than that for the *G. elata* genome (73.1%), and much higher than that for the *E. aphyllum* transcriptome generated by Schelkunov et al. (2018) (53.4%; **Supplementary Table 6**). We also checked the mapping rate of the RNA-seq reads on these transcriptomes (**Supplementary Table 7**). This was higher than 94%, except for mycorrhizal samples as expected due to the presence of the mycorrhizal fungi. Finally, we looked for the orthologs of the plant KEGG pathways and of the Mapman4 bins and compared them to the *G. elata* gene content (**Supplementary Data 2**). An ortholog was counted if at least one transcript or gene was associated with it. Out of the 140 tested KEGG pathways representing 15,150 potential orthologs, none were differentially represented between our transcriptomes and *G. elata*. Similarly, none of the 1,196 tested Mapman4 bins representing 4,966 potential orthologs were differentially represented. Even when relaxing the stringency of the test (raw *p*-value < 0.05), no bin or pathway suggesting missed orthologs in our transcriptomes compared to *G. elata* was statistically significant. Taken together, these results strongly support that our transcriptomes were complete at least for the pathways considered in the comparison and that any missing orthologs from our transcriptomes were probably

lost by the corresponding species. It also suggests that the gene repertoires of *E. aphyllum*, *N. nidus-avis*, and *G. elata* are very similar.

Impact of Mycoheterotrophy on Gene Repertoires in Mycoheterotrophic Orchids

By considering the gene repertoires of three mycoheterotrophic orchids that experienced independent evolutionary origins of mycoheterotrophy, we can also address its impact on their gene sets. A comparison with the genomes of *Phalaenopsis equestris*, *Dendrobium catenatum*, and *Apostasia shenzhenica*, three autotrophic orchids, using the KEGG and Mapman4 pathways described above (**Table 2** and **Supplementary Data 2**), shows that the switches to mycoheterotrophy result in the loss of orthologs exclusively associated with pathways directly related to photosynthesis. Even when relaxing the stringency of the test (raw *p*-value < 0.05), there is no indication that the switch to mycoheterotrophy is associated with any gain of novel genes or that pathways other than those associated with plastid or photosynthesis showed significant gene losses (**Supplementary Data 2**). It should also be noted that none of the genes lost

TABLE 2 | Gene content: pathways impacted by the switch to mycoheterotrophy.

	Code	Size	AS	DC	PE	GE	EA	NNA	MH impact
Mapman4									
Photosynthesis	1	291	161	196	183	42	33	41	Losses
Photophosphorylation	1.1	239	124	157	143	17	13	18	Losses
ATP synthase complex	1.1.9	12	6	12	12	0	0	1	Losses
Chlororespiration	1.1.8	41	22	17	4	5	4	4	Losses
Cytochrome b6/f complex	1.1.2	19	10	19	19	0	0	0	Losses
Linear electron flow	1.1.5	5	4	5	5	1	1	1	Losses
Photosystem I	1.1.4	28	20	26	26	1	1	0	Losses
Photosystem II	1.1.1	74	58	74	73	10	7	12	Losses
Calvin cycle	1.2	30	22	24	25	12	7	10	Losses
RuBisCo activity	1.2.1	14	13	13	14	7	4	4	Losses
Galactolipid and sulfolipid biosynthesis	5.3	7	7	7	7	3	3	4	Losses
Coenzyme metabolism	7	224	145	155	154	129	132	135	Losses
Phylloquinone biosynthesis	7.13	8	8	8	8	2	2	2	Losses
Tetrapyrrole biosynthesis	7.12	58	37	39	38	25	27	29	Losses
Chlorophyll metabolism	7.12.6	23	21	22	22	10	11	12	Losses
Organelle RNA polymerase machinery	15.6	35	21	28	29	7	6	6	Losses
Organelle RNA polymerase activities	15.6.1	27	19	26	27	6	5	5	Losses
Organelle RNA processing	16.12	102	73	79	72	55	49	53	Losses
Organelle RNA editing	16.12.5	42	33	36	28	23	17	19	Losses
Organelle RNA stability	16.12.4	6	6	6	5	3	2	2	Losses
Chloroplast disulfide bond formation	18.11.2	3	3	3	3	0	0	1	Losses
Plastid movement	20.5	10	9	9	9	4	4	4	Losses
Total	total	5963	3945	4211	4185	3790	3813	3891	Losses
KEGG pathways									
Photosynthesis	195	63	54	54	33	10	4	4	Losses
Photosynthesis – antenna proteins	196	42	11	11	11	0	3	1	Losses

Code, code of the Mapman4 bin or KEGG pathway; Size, ortholog content of the bin or pathway; AS, *A. shenzhenica*; DC, *D. catenatum*; PE, *P. equestris*; GE, *G. elata*; EA, *E. aphyllum*; NNA, *N. nidus-avis*; MH impact, impact of the switch to mycoheterotrophy.

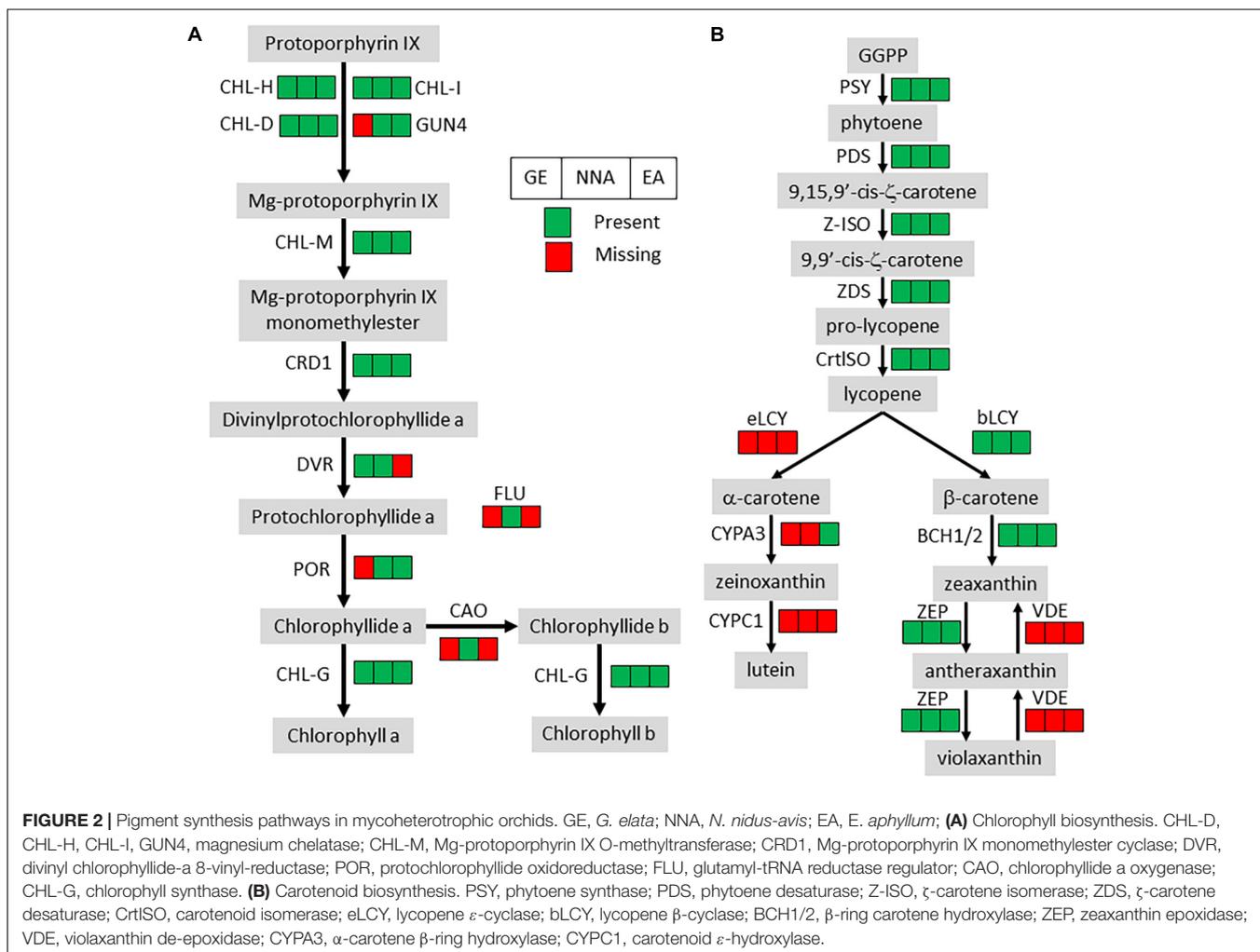
from their plastid genomes were found in the transcriptomes of *E. aphyllum* and *N. nidus-avis*. This suggests that the switch to mycoheterotrophy selectively resulted in gene losses (and not in gene transfers to their nucleus) in pathways associated with photosynthesis. However, most of these pathways were not completely lost.

All the orthologs required for photosystems were not detected, but the losses in the chlorophyll metabolism pathway were almost exclusively restricted to chlorophyll degradation and interconversion. As seen before (Wickett et al., 2011; Schelkunov et al., 2018), the chlorophyll synthesis pathway was mostly conserved but incomplete in *G. elata* and *E. aphyllum* (Figure 2A). By contrast, *N. nidus-avis* expressed the full extent of genes required for the biosynthesis of chlorophyll as well as some chlorophyll *a/b* binding proteins [Light-Harvesting-Complex A3 (LHCA3), LHCB1, LHCB2, stress-enhanced protein 1 (SEP1), SEP3, SEP5, and early light-induced protein (ELIP) genes]. Similarly, the three mycoheterotrophic species were missing the *lycE* and *lut5* genes required for the synthesis of lutein, a photoprotective pigment, but possessed a complete biosynthesis pathway to violaxanthin (Figure 2B). No gene coding for a

violaxanthin de-epoxidase, required for the xanthophyll cycle to occur, was found in any of the three MH species.

Gene losses in MH species mirror the loss of photosynthesis but another characteristic of MH species is the lack of developed leaves which are reduced to small scales. Leaf initiation and expansion is controlled by a network of transcription factors and hormone gradients (Bar and Ori, 2014; Wang et al., 2021). None of the important gene families involved in leaf development and present in the genome of autotrophic orchids (YABBY, KNOX, BOP, PINOID, KNAT, ARF, HD-ZIPIII, and NGATHA) are missing from MH orchids.

An examination of known pathways, mainly deriving from model autotrophic plants, allows identification of gene losses linked to the switch to mycoheterotrophy, but this misses any potential new pathways or genes that may be associated with this transition. To address this, we performed an orthology analysis including the coding genes of the six previous orchid species and three grasses [*Zea mays*, *Brachypodium distachyon*, and *Oryza sativa* (Supplementary Data 3 and Supplementary Tables 8, 9)]. Out of the 18,259 orthogroups identified, only 38 contained exclusively genes from all three MH orchid



species. Twenty-two of these orthogroups contained only unannotated genes and the 16 remaining did not have specific annotations (**Supplementary Data 4**). These results suggest that the switch to mycoheterotrophy in orchids does not involve new pathways or functions.

A Transcriptome Analysis Highlights the Organ-Specific Functions of Mycoheterotrophic Orchids

The pairwise comparisons of the transcriptome profiles of flower, stem, and mycorrhizal root/rhizome of *E. aphyllum* and *N. nidus-avis* identified the genes differentially expressed between these organs as well as organ-specific genes (**Supplementary Data 5**). We identified 18,817 and 12,331 differentially expressed genes as well as 6,351 and 4,520 organ-specific genes in *N. nidus-avis* and *E. aphyllum*, respectively (**Table 3**). The highest numbers of differentially expressed genes were observed between underground and aerial organs. Similarly, most organ-specific genes were identified in the mycorrhizal root/rhizome.

To elucidate which functions are served by the differentially expressed and organ-specific genes, Gene Ontology, Mapman, and KEGG enrichment analyses were performed (**Supplementary Data 6**). While very few enrichments were found in the organ-specific genes, the differentially expressed genes showed that numerous metabolic functions were differentially activated in the three organs, following a strikingly similar pattern in *N. nidus-avis* and *E. aphyllum*. **Figure 3** summarizes the Mapman and KEGG enrichment analyses, which are fully supported by the GO enrichment analyses. The metabolic functions are indicated where their transcriptomic activity appeared to be peaking. The aerial parts shared high levels of amino acid and fatty acid synthesis as well as high level of primary cell wall metabolism. Light signaling pathways were also activated in aerial parts. The flowers specifically showed high cell division and phenolic activities. In *N. nidus-avis*, the chlorophyll synthesis pathway was active mostly in the flowers along with other plastid associated pathways. At the other end of the plant, the mycorrhizal roots of *N. nidus-avis* and the mycorrhizal rhizomes of *E. aphyllum* showed some contrasted metabolic functions. Nevertheless they also showed a remarkable convergence with an increased activity

of pathways related to pathogen and symbiont interactions, as well as of the transportome (e.g., ABC transporters and solute carriers) and degradation capacities (i.e., proteasome and glycosaminoglycan degradations).

Comparison of Expression Profiles in Underground Organ and Stem of Mycoheterotrophic and Autotrophic Species

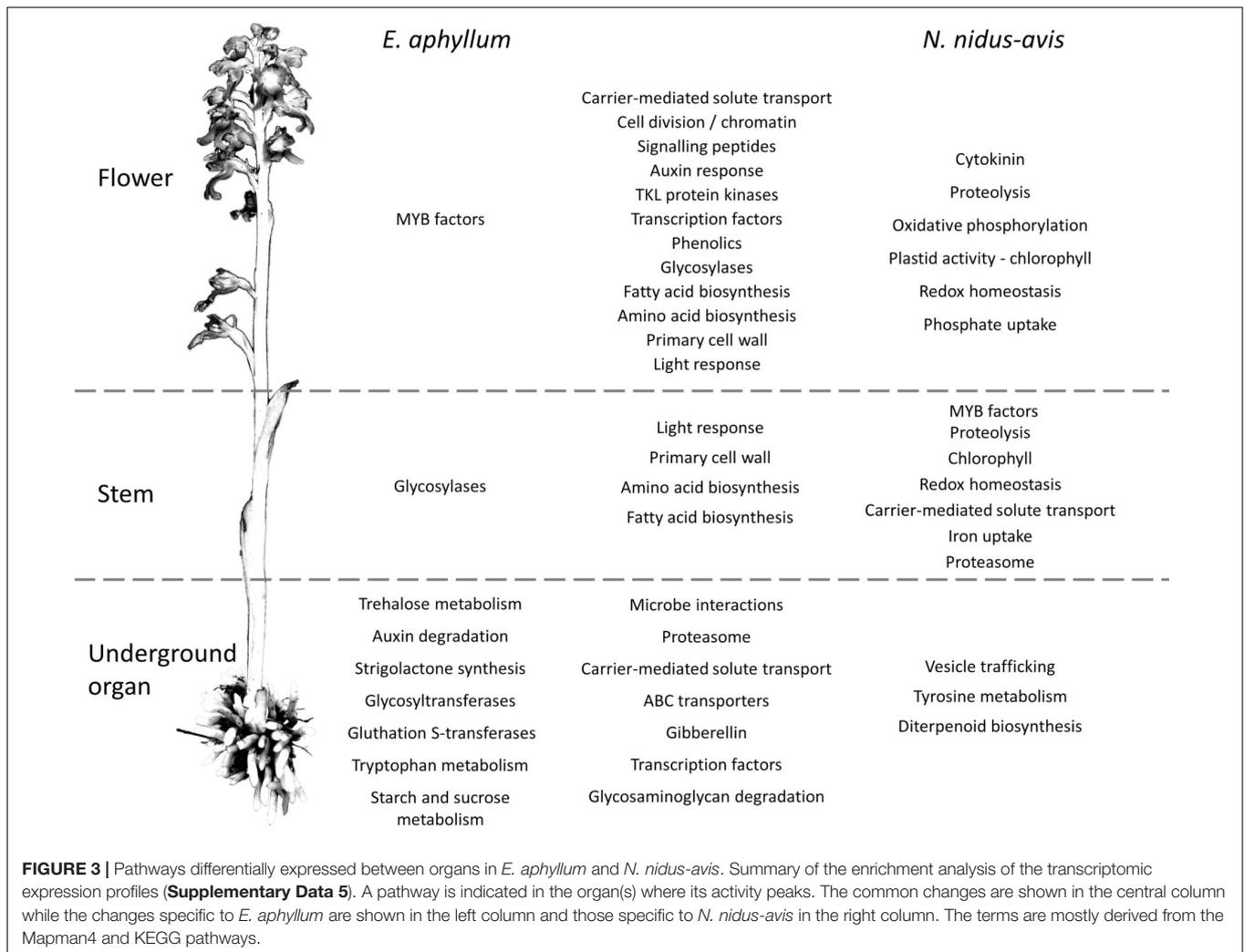
To understand the consequences of mycoheterotrophy for the expression profiles, it would be preferable to compare our mycoheterotrophic orchids to autotrophic relatives species from a transcriptomic point of view. However, no equivalent transcriptomic dataset (pairs of underground and stem organs from the same individuals with biological replicates) for autotrophic orchids or other monocots was publicly available at the time of analysis. We used datasets from two grasses, *B. distachyon* and *Z. mays*. We analyzed only the 8,620 (out of 18,259) orthogroups detected in the roots or stem of all four species. This filter removes most of the orthogroups associated with photosynthesis, but these pathways are an obvious major difference between the two trophic types. While 2,378 and 3,617 orthogroups were differentially expressed between underground organ and stem in autotrophic and mycoheterotrophic species, respectively, 3,359 orthogroups (39% of the analyzed orthogroups) showed a significantly different underground to aboveground ratio between the two trophic types, including 2,536 (30%) with inverted ratios (**Supplementary Data 7**).

The pathway enrichment analysis of the differentially expressed orthogroups in the mycoheterotrophic orchids (**Supplementary Data 8**) showed results similar to the transcriptomic analysis of *E. aphyllum* and *N. nidus-avis* genes, supporting the idea that orthogroup expression patterns are biologically relevant. **Figure 4** summarizes the results of the pathway enrichment analysis of these orthogroups. It is particularly noteworthy that the orthogroups associated with fatty acid biosynthesis, amino acid biosynthesis, primary cell wall metabolism, glycosidases and secondary metabolism are more expressed in the stem of the mycoheterotrophic orchids than in their underground organs and are more expressed in the roots of autotrophic grasses than in their stem. The opposite is true for the orthogroups involved in RNA metabolism and DNA damage response. The orthogroups of some pathways (those involved in solute transport, symbiosis, trehalose degradation, and cytochrome P450) were more expressed in the underground organs than in the stems for both autotrophic and mycoheterotrophic species but differed between the two suggesting that the species of the two trophic types either induced these pathways to different levels or used different orthologs.

The latter can be illustrated for the solute transport pathway. The 192 orthogroups showing a different root/shoot ratio between AT and MH (out of 392 orthogroups belonging to the solute transport pathway) are distributed in most transporter families, and in each family there

TABLE 3 | Summary of differential gene expression analyses among the sampled tissues.

	<i>Neottia nidus-avis</i>	<i>Epipogium aphyllum</i>
Stem vs. flower	9,109/4,644 down, 4,465 up	5,315/2,123 down, 3,192 up
Mycorrhiza vs. flower	13,701/6,465 down, 7,236 up	7,596/3,430 down, 4,166 up
Mycorrhiza vs. stem	11,360/4,866 down, 6,494 up	7,849/3,955 down, 3,894 up
Flower-specific	55	297
Stem-specific	508	175
Mycorrhiza-specific	5,788	4,048
Total	25,168 (57.92%)	16,851 (43.78%)



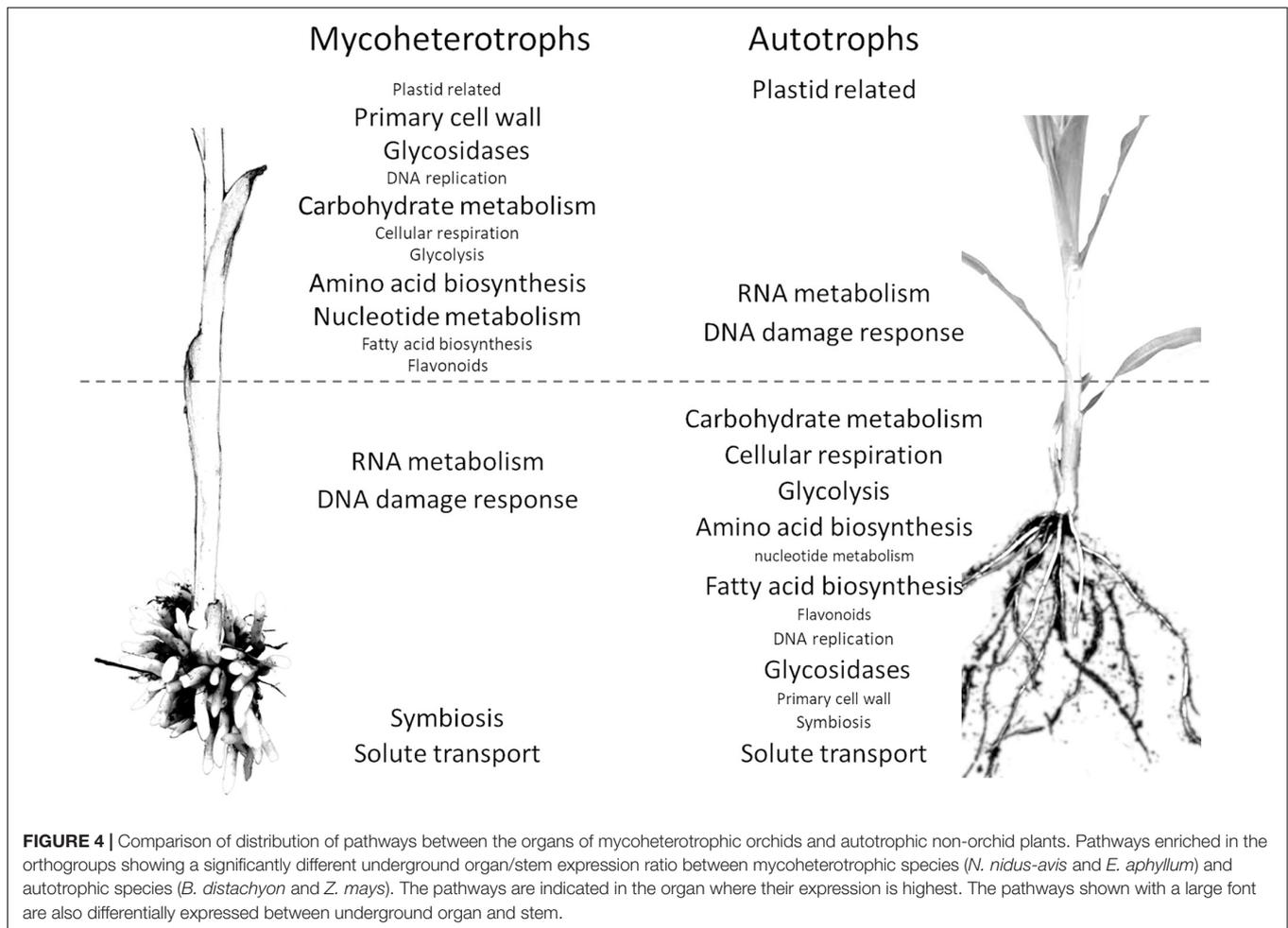
are orthologs showing different behavior in MH and AT (Supplementary Data 7). This collectively illustrates that the consequences of mycoheterotrophy extend well beyond photosynthesis and associated gene losses observed previously. Mycoheterotrophy has remodeled a large fraction of gene expression and metabolism.

DISCUSSION

No Genetic Innovation, but Convergent Gene Loss in Mycoheterotrophic Orchids

Using the RNA-seq data from rhizome/root, stem and flowers of *E. aphyllum* and *N. nidus-avis*, we identified gene sets that are probably almost complete, based on their comparison with the genome of *G. elata* (Supplementary Data 2). This comparison also illustrates that these three independent evolutionary events of mycoheterotrophy have converged to essentially the same gene repertoire. When comparing the molecular functions encoded in these three mycoheterotrophic orchids with those of autotrophic orchids, the switch to mycoheterotrophy entailed

a global reduction of molecular functions, as was previously demonstrated for *G. elata* (Supplementary Data 2; Yuan et al., 2018). In addition, we could not detect any major gain of function associated with mycoheterotrophy. Obviously, it is difficult to identify potentially unknown functions and our transcriptome analysis must have missed some genes, potentially including those specifically expressed during germination. However, all orchids regardless of their nutrition at adulthood, are mycoheterotrophic during germination (Bernard, 1899; Merckx, 2013; Dearnaley et al., 2016). This indirectly indicates that they have all the genes and metabolic pathways required to obtain nutrients through mycoheterotrophic nutrition, suggesting that major gains/innovation are not essential for the transition to mycoheterotrophy. When looking for orthologs present in our three mycoheterotrophic orchid species, but not in the other six autotrophic species, we found only a handful of mycoheterotrophy-specific orthologs. It is highly unlikely that they are the key genes required for the switch to mycoheterotrophy, although more extensive sampling of mycoheterotrophic and autotrophic species should be done to verify this assumption.



In addition to a general reduction of gene content, Yuan et al. (2018) showed that some gene families, mostly associated with interactions with fungi, expanded in the *G. elata* genome. Our transcriptome assemblies include large numbers of contigs putatively coding for enzymes such as mannose-specific lectins or β -glucosidases, indicating the possible expansion of some gene families in *E. aphyllum* and *N. nidus-avis*. However, using transcriptome assemblies (and despite or because of a step of redundancy reduction in our analysis), it is difficult to count the number of genes precisely because it is not possible to distinguish between two transcript isoforms and two copies of a gene. Only high-quality assemblies of the large genome of these species (16.96 Gb for *N. nidus-avis*; Vesely et al., 2012) will allow the confirmation of the expansion of such gene families in these species.

Pigments and Secondary Metabolism: Compensatory Protection and Camouflage?

The gene losses observed in the mycoheterotrophic orchids reflect the evolution of their plastomes: massive gene loss restricted to photosynthetic pathways and functions. The only

genes retained in their plastid genomes have non-photosynthetic functions (Graham et al., 2017; Barrett et al., 2019; Mohanta et al., 2020). By extension to the nuclear genome, we can assume that the orthologs not detected in mycoheterotrophic species are probably exclusively associated with photosynthesis, while the conserved orthologs probably have non-photosynthetic functions. Thus, the comparison of the gene contents of mycoheterotrophic and autotrophic species should provide useful information for the functional analysis of genes even in model plants, as shown by two examples below.

The loss of photosynthesis resulted in gene losses in several pigment synthesis pathways (Table 2). In *N. nidus-avis*, Pfeifhofer (1989) detected high amounts of zeaxanthin but no lutein. In the three MH species, the genes coding for the enzymatic activities of the carotenoid pathway required for the synthesis of zeaxanthin, but not lutein, are conserved (Figure 2). Lutein is associated with the dissipation of excess energy from the photosystems and zeaxanthin is part of the xanthophyll cycle, which has the same function (Niyogi et al., 1997). However, the loss of violaxanthin de-epoxidase shows loss of the xanthophyll cycle in these species. The fact that zeaxanthin is also a precursor of abscisic acid may explain the conservation of a functional synthesis pathway. Thus, the switch to mycoheterotrophy appears to have trimmed the

multifunctional carotenoid synthesis pathway to keep only the enzymes required for its non-photosynthetic functions.

Because of the potential photo-toxicity of chlorophylls and their precursors (Rebeiz et al., 1984), a null expectation might be that mycoheterotrophic species should lose the chlorophyll synthesis pathway. It is nonetheless mostly conserved, even if incomplete, in *E. aphyllum* and *G. elata* (Figure 2). Such conservation has been observed in holoparasitic and mycoheterotrophic plants (Wickett et al., 2011; Barrett et al., 2014) and in coral-infecting apicomplexan (Kwong et al., 2019), and suggests that chlorophylls or their intermediates should have a non-photosynthetic function. It remains unclear what this function is (Ankele et al., 2007). *N. nidus-avis* differs from the two other species in having a complete and functional chlorophyll synthesis pathway. Its activity, along with other plastid activities, was detected in *N. nidus-avis*, mostly in the flowers (Figure 3). This is consistent with the detection of chlorophyll a and b in the inflorescence (Pfeifhofer, 1989). Chlorophyll was also detected in other MH orchids (Barrett et al., 2014) and the authors proposed that it would support a minimal and localized photosynthetic activity providing additional carbon to the production of seeds. This hypothesis is consistent with the demonstration that the products of the photosynthesis of the mixotrophic orchid *Cephalanthera damasonium* are targeted to fruits and seeds (Lallemant et al., 2019a). It is also supported by Menke and Schmid (1976), which reported cyclic photophosphorylation in the flower of *N. nidus-avis*. However this report is incompatible with the absence of most plastid and nuclear genes coding for photosystem I and cytochrome b6/f and deserves further study.

As free chlorophylls are photo-toxic (Rebeiz et al., 1984), the accumulation of chlorophyll requires a photo-protection mechanism. Flowers of *N. nidus-avis* are not green, but they turn green upon heating (Supplementary Figure 1), suggesting that the chlorophyll is stored in a heat-labile complex and that this may limit toxicity. Indeed, Cameron et al. (2009) failed to detect any chlorophyll fluorescence in this species, supporting its lack of photochemical activity. When compared with *G. elata* and *E. aphyllum*, the activity of the chlorophyll synthesis pathway in *N. nidus-avis* is associated with the presence of several SEP and ELIP genes. The SEP1 and ELIP Arabidopsis orthologs are induced in response to high light and are believed to bind chlorophyll (Adamska et al., 1999; Heddad, 2000; Rossini et al., 2006), but their exact molecular functions are unknown. Their conservation in *N. nidus-avis*, but not in *E. aphyllum* or *G. elata*, suggests that they may indeed bind chlorophyll to inactivate its ability to capture light.

Another, non-exclusive possible explanation for conservation of a functional chlorophyll synthesis pathway and the accumulation of zeaxanthin to high levels in *N. nidus-avis* (Pfeifhofer, 1989) may be camouflage. By visually blending the plants into the background of leaf litter, the dull colors of MH species protect them against herbivory (Klooster et al., 2009).

In any case, we show that the switch to mycoheterotrophy is mostly dominated by function losses, and does not require major, massive metabolic innovations. In mixotrophic species (representing an evolutionary transition from autotrophy to mycoheterotrophy; Selosse and Roy, 2009), a metabolomic and

transcriptomic analysis showed that their response to the loss of photosynthesis by mutation was similar to the response of achlorophyllous mutants of autotrophic plants (Lallemant et al., 2019b). This suggests that the ability of achlorophyllous variants of otherwise green mixotrophic species to sustain an almost normal growth without photosynthesis is mostly based on the plasticity of plant metabolism. Furthermore, mycoheterotrophy is not a rare event (it has occurred > 50 times in 17 plant families; Merckx et al., 2009; Těšitel et al., 2018; Barrett et al., 2019), suggesting that it primarily entails functional losses and not complex gene gains.

Another characteristic of mycoheterotrophic orchids is the lack of developed leaves. They are not missing but are reduced to small scales (Figure 1). The genes supposedly involved in leaf initiation but also leaf blade development are not missing, most probably because they function in other developmental processes. So the lack of developed leaves in mycoheterotrophic orchids could be explained by impaired expression profiles of these genes.

An Upside-Down Metabolic Architecture

Photosynthesis is considered to be at the core of plant metabolism and so its loss in normally green plants severely impacts their metabolism (Aluru et al., 2009; Abadie et al., 2016; Lallemant et al., 2019b). We analyzed the physiology of mycoheterotrophic orchids through gene expression in different organs (Figure 3 and Supplementary Data 6). Many genes were differentially expressed, reflecting a partition of metabolic functions between the organs of most plants. The flowers showed a higher activity of cell division, primary cell wall and signaling pathways, which can be attributed to floral development. Similarly, higher phenolic compound synthesis can be associated with pollinator attraction involving flower pigmentation and production of fragrant phenolics (Jakubska-Busse et al., 2014). Conversely, the different underground organs of *N. nidus-avis* (roots) and *E. aphyllum* (rhizome) converged toward a higher activity of pathways likely involved in the interaction with their fungal partners (microbe interactions, proteasome, and transporters). This transcriptomic convergence probably results from the similar function as organs where nutrient exchange at plant-fungus interfaces takes place. This is also evidenced in their anatomical convergence (reduced number of xylem elements) or functional similarities (nutritional independency from the other organs of plant; Rasmussen, 1995). Although *N. nidus-avis* and *E. aphyllum* showed similar pathway enrichments, especially in the aerial organs, there were some idiosyncrasies. These differences are difficult to interpret clearly as they may result from the different phylogenetic backgrounds, the anatomical differences (roots vs. rhizome) but also from different fungal partners. For example, the peak of trehalose, tryptophan, starch, and sucrose metabolism observed in the rhizome of *E. aphyllum* as opposed to a peak of tyrosine metabolism in the roots of *N. nidus-avis* (Figure 3 and Supplementary Data 6) may provide clues to the specificities of the nutrient fluxes in these two pairs of partners.

Comparing symbiotic and asymbiotic protocorms of the orchid *Serapias vomeracea*, Fochi et al. (2017) highlighted the importance of organic N metabolism and especially lysine

histidine transporters (LST) in its interaction with its fungal partner. In our analysis, several LST genes were differentially expressed between the organs for both *N. nidus-avis* and *E. aphyllum*, but some were induced in flowers while others were more transcribed in stems or mycorrhizal parts (**Supplementary Data 7**). In a similar analysis in *G. elata*, the upregulation of clathrin genes in symbiotic protocorms compared to asymbiotic protocorms suggested the involvement of exocytosis in the interaction between the orchid and its fungal partner (Zeng et al., 2017). Our analysis showed no signal specific to N metabolism or exocytosis. The different conditions considered in these studies may help explain such discrepancies, but they may also illustrate different kinds of evolutionary adjustments occurring in different mycorrhiza.

Comparison of expression profiles of the mycoheterotrophic orchids to similar datasets in the autotrophic species: *B. distachyon* and maize provides additional evidence of the impact of mycoheterotrophy on plant metabolism. The interpretation of differences should be done carefully because it is limited by factors such as different phylogenetic backgrounds, possibly different growth conditions (including the possible absence of mycorrhizal fungi in the autotrophic plants considered here), or the restriction of the comparison to orthogroups detected in all four species. Despite these limitations, we can state that almost 40% of the analyzed orthogroups had a significantly different root/stem ratio between mycoheterotrophic and autotrophic species, and that 30% of the orthogroups, from numerous pathways, showed inverted underground organ/stem ratios, suggesting that the metabolism of mycoheterotroph species has been inverted compared to photosynthetic taxa. This inversion of the metabolism architecture likely coincided with the inversion of the usual source/sink relationship: in mycoheterotrophs, underground organs are sources, while they are a sink in photosynthetic species. The sink organs were associated with a higher activity of several major metabolic pathways (carbohydrate and nucleotide metabolism, amino acid and fatty acid biosynthesis, glycolysis, and respiration). In association with a higher DNA replication and primary cell wall activity (which involves glycosidases) and a higher expression of auxin transporters, sink organs likely experience stronger growth than their source counterparts. Mycoheterotrophic roots and rhizomes are generally short, thick and compact to minimize accidental loss of a part of a source organ and nutrient transfer effort (Imhof et al., 2013), stems are ephemeral (<2 months) but fast growing (e.g., 4 cm/day in *E. aphyllum*, J. Minasiewicz personal observations) organs involved in sexual reproduction but without nutritional functions. Conversely, fibrous roots of grasses have high growth rate as nutrient uptake depends largely on the root length (Fitter, 2002). Even with different growth habits, some pathways showed similar overall expression underground organ/stem ratios in mycoheterotrophic orchids and photosynthetic grasses. Plastid-related pathways (chlorophyll synthesis, plastid translation) are more active in stems than in underground organs, while symbiosis and trehalose degradation are more active in underground organs than stems. Trehalose is almost absent from vascular plants, where its 6-phosphaste precursor is

an important growth regulator (Lunn et al., 2014). However, it is an abundant storage carbohydrate in mycorrhizal fungi and it has been suggested that it is transferred to mycoheterotrophic orchids to be cleaved into glucose (Müller and Dulieu, 1998). A comparison between leaves of achlorophyllous mutants (thus with mycoheterotrophic nutrition) and green individuals in mixotrophic orchids showed an upregulation of trehalase, but also of trehalose-6-P phosphatases (TPP) and trehalose-6-P synthase (TPS; Lallemand et al., 2019b). Similarly, the mycoheterotrophic orchids demonstrated a higher underground organ/stem ratio of trehalase and TPP expression (but not TPS) compared to photosynthetic grasses. This result supports the hypothesis that trehalose is transferred from mycorrhizal fungi to mycoheterotrophic orchids. Many other nutrients are exchanged at this interface and our analysis suggests numerous differences between the trophic types: close to half of the orthogroups involved in solute transport showed different underground organ to stem ratios between autotrophic and mycoheterotrophic species. Some SWEET (Sugar Will Eventually be Exported) transporters were induced in the mycorrhiza of achlorophyllous MH mutants of the mixotrophic orchid *Epipactis helleborine* (Suetsugu et al., 2017) and in the protocorms of *Serapias* (Perotto et al., 2014). The three SWEET orthogroups in our analysis behaved differently between autotrophic and mycoheterotrophic species, but showed contrasting differences, indicating that autotrophic and mycoheterotrophic species both used SWEET transporters in underground organs and stems but corresponding to different orthologs. Similarly, 13 out of the 15 ABCG transporter orthogroups or 10 out of the 13 NRT1/PTR transporter orthogroups showed contrasted differences between autotrophic and mycoheterotrophic species. The same could be observed for all transporter families (**Supplementary Data 7**): autotrophic and mycoheterotrophic species use different orthologs for the transport of solutes in stem and roots, demonstrating extensive expression reprogramming. These differences are probably associated with changes in the fluxes of nutrients in autotrophic and mycoheterotrophic species, including in mycorrhizas. Understanding these changes is a central question in the study of mycoheterotrophy. However, the specificity of transporters can vary even within a gene family. For example, transporters of the NRT1/PTR family were identified as nitrate transporters, but some transport other molecules (Corratgé-Faillie and Lacombe, 2017). Further investigations of the changes of nutrient fluxes associated with this reprogramming of transporter expression should be directed at a detailed analysis of each orthogroup (assuming that the substrate specificity is the same for all transporters within an orthogroup). However, such an analysis should not replace direct measurement of these fluxes with labeling experiments, which will also be required to better understand the processes involved.

CONCLUSION

The shift to mycoheterotrophy induces diverse changes in the genome of MH plants. From the analysis of the gene

repertoires, we were not able to identify new functions associated with mycoheterotrophy, and large losses appeared to be restricted to genes exclusively involved in photosynthetic functions. This could superficially suggest that no metabolic innovation is required for mycoheterotrophy. However, our transcriptome analysis shows extensive changes in numerous pathways, probably associated with changes in the plant lifecycle and in the interaction with fungal partners induced by mycoheterotrophy. This reprogramming illustrates the versatility of plant metabolism and can be considered as a metabolic innovation in itself. It may also help explain why the shift to mycoheterotrophic nutrition has occurred so frequently in plant evolution: becoming mycoheterotrophic may be based more on reprogramming of existing metabolism and gene loss than on genetic innovation involving new genes or pathways.

MATERIALS AND METHODS

Sampling Procedures

The specimens of *N. nidus-avis* and *E. aphyllum* were sampled in their natural habitats in southern Poland in 2017 at the peak of their flowering season, at 10.00 AM local time (**Supplementary Table 1**; these two species cannot be cultivated *ex situ*). Two plants per species were selected as biological replicates based on their size and healthy condition, keeping the parameters similar among the replicates. Each plant was excavated with surrounding soil. A fully open flower and associated stem were cut off and processed (see below) right away, while the underground organs were first cleaned thoroughly by gentle scrubbing and rinsing in distilled water to remove all visible soil and foreign material. In-field processing consisted of slicing and dividing material into samples of 150 mg in weight before immediate preservation in liquid nitrogen to inhibit RNA degradation. The presence of mycorrhizal fungus was checked on thin cross-sections of colonized organs adjacent to the sampling and examined later under a light microscope.

RNA Extraction and Purification

The samples collected *in situ* were transferred from liquid nitrogen to a -80°C freezer until the RNA extraction step. Flower samples were homogenized in liquid nitrogen using TissueLyser II (Qiagen) in 2 mL Eppendorf tubes containing ceramic beads. This method has proven ineffective in processing root, rhizome, and stem tissue samples due to their hardness when frozen and so manual grinding in mortars with liquid nitrogen had to be applied instead. Homogenized material was subjected to the NucleoZol (Macherey-Nagel, Dueren, Germany) reagent extraction process following the manufacturer's protocol with the addition of polyvinylpyrrolidone (PVPP) during the grinding of root and rhizome. RNAs were further purified using Agencourt RNAClean XP (Beckman Coulter, Brea, CA, United States) magnetic beads following the manufacturer's instructions.

Digestion with DNase Max (Qiagen, Hilden, Germany) was subsequently conducted to purify RNA isolates from remaining genomic DNA contamination.

Finally, RNA integrity and purity were assessed by Agilent BioAnalyzer 2100 survey using the Plant RNA Nano Chip (Agilent Technology, Santa Clara, CA, United States) and RNA concentration was measured by RiboGreen assay (Thermo Fisher Scientific, Waltham, MA, United States). Samples exhibiting high integrity (RIN > 7) were selected for sequencing.

Sequencing

The sequencing libraries were constructed using TruSeq Stranded Total RNA with the Ribo-Zero Plant kit (Illumina, San Diego, CA, United States), following the manufacturer's instructions. Next, they were sequenced on a NextSeq500 (Illumina, San Diego, CA, United States) platform in a paired-end mode with read length of 150 bases. The sequences obtained were quality-controlled and trimmed using the Trimmomatic software (version 0.36, parameters PE and ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:true LEADING:21 TRAILING:21 MINLEN:30) (Bolger et al., 2014) and the residual ribosomal RNAs were filtered out with SortMeRNA (version 2.1, against the databases silva-bac-16s-id90, silva-bac-23s-id98, silva-euk-18s-id95, and silva-euk-28s-id98 with the parameters `-e 1e-07 -paired_in`).

Transcriptome Assembly

Due to the lack of reference genomes of the sampled plants, their transcriptomes were assembled *de novo* using the Trinity RNA-seq assembler (Haas et al., 2013) version v2.6.6 with all parameters at their default settings except `--SS_lib_type RF`. Taking into account the possible contamination of our samples collected *in situ*, which may vary between collected tissues, in order to avoid mis-assemblies and/or chimeric transcript generation, we assembled the transcriptomes of each plant tissue separately. Subsequently, the assemblies for each species were merged and the redundancy of the isoforms was decreased with the tr2acds pipeline from the EvidentialGene package¹ (version 2017.12.21; Gilbert, 2019). According to the pipeline description, we kept only those contigs that were classified as “main” or “noclass”, i.e., primary transcripts with alternates or with no alternates, respectively, to form the final unigene set.

Identification of Contaminating Contigs

As our samples were collected in the field, the total RNAs contain transcripts from the microbiota associated with our plants, and especially abundant transcripts of the mycorrhizal fungi in underground organs, which means that the previous unigene set is contaminated with sequences that do not belong to the plants.

To identify and filter out these contigs, each reduced transcriptome was searched with the blastx algorithm against the NCBI NR database using Diamond software (Buchfink et al., 2015). Local Diamond version 0.9.16 installation was run with the following set of parameters: `-sensitive, -index-chunks 2, -block-size 20, -max-target-seqs 50, -no-self-hits, -evalue 0.001, -taxonmap prot.accession2taxid.gz`, with the latest parameter specifying the taxonomic information obtained from

¹<http://arthropods.eugenics.org/EvidentialGene/trassembly.html>

the NCBI ftp pages². Both the NCBI NR database and the taxonomy information were current as of December 2018. All contigs with best hits inside the *Streptophyta* clade of plants were considered as *bona fide* orchid contigs.

However, this analysis may miss genes highly conserved across kingdoms. Hence, we performed an orthology analysis against several orchid and monocotyledon species. The analysis included proteomes of *N. nidus-avis* and *E. aphyllum* generated here, as well as published reference sets of *B. distachyon* (L.) P.Beauv., *Z. mays* L., *O. sativa*, and the orchids *G. elata*, *D. catenatum* Lindley, *A. shenzhenica* Z.J.Liu & L.J.Chen, and *P. equestris* (Schauer) Rchb.f. (see **Supplementary Table 2**). We identified orthologous groups using the OrthoFinder software (version 2.2.7, default parameters, except -S diamond) (Emms and Kelly, 2019).

Contigs sharing the same orthogroup as any protein of these seven species were considered as *bona fide* orchid contigs. For contigs with no hit at all we applied a further filtering criterion based on the expression pattern, i.e., we required such transcripts to be expressed in at least two out of our six samples. Expression analyses were performed with Seal from the BBTools package³ (version 38.22).

Identification of the Fungal Partners

The contig sets were searched for ITS sequences using ITSx software (version 1.1.2; Bengtsson-Palme et al., 2013) and the identified contigs were queried against the UNITE database version 8.2 (Nilsson et al., 2019).

Annotations

Annotation of transcripts was performed with the Trinotate suite⁴ (version 3.1.1; Bryant et al., 2017). Trinotate was fed with results of several independent analyses. To annotate protein domains, hmmscan from the HMMER 3.1b2 package (Eddy, 2011) was run against the Pfam-A entries from the PFAM database (Finn et al., 2016). The UniProt/SwissProt protein database was searched with blastp (Altschul et al., 1997) from the blast+ 2.7.1 package to retrieve gene ontology (GO), KEGG (Kyoto Encyclopedia of Genes and Genomes), and eggNOG annotations. The presence of signal peptides was assessed with signalP (Petersen et al., 2011) software.

Additionally, the transcripts were assigned to KEGG orthologs and pathways using the KAAS server (Moriya et al., 2007) with BLAST and the BBH (bi-directional best hit) method. They were also assigned to the Mapman4 pathways using the Mercator4 v2.0 online tool (Schwacke et al., 2019).

In all the above analyses, transcripts were represented by either their nucleotide sequences derived directly from the assembly or by their amino acid sequences, as derived from the open reading frames (ORFs) determined by the tr2aacds pipeline. To reduce technical bias when comparing species, the gene sets of all species were re-annotated with the same tools and parameters. The annotation of the orthogroups was derived from the annotations of their genes independently of the origin of these

genes. Orthogroups were annotated with terms representing at least 25% of their genes.

Comparison of Gene Sets

The quality and completeness of the final transcriptomes (unigene sets) for *E. aphyllum* and *N. nidus-avis* were benchmarked with BUSCO v3.0.2 (Seppey et al., 2019) against the *Liliopsida:odb10* plant-specific reference database and compared with the abovementioned species. We also compared the representation of the KEGG pathways and Mapman4 bins in each species. The unigene sets of *E. aphyllum* and *N. nidus-avis* were first completed with their plastid gene lists extracted from the NCBI accessions NC_026449.1 and NC_016471.1, respectively. We counted whether a KEGG ortholog or its Mapman equivalent was detected independently of the number of genes associated with it. Fisher's exact test was performed to compare *E. aphyllum* and *N. nidus-avis* to *G. elata* and to compare these three mycoheterotrophic orchids to the three autotrophic orchids in each pathway or bin. Pathways or bins with an adjusted *p*-value (Bonferroni adjustment) below 0.05 were considered as differentially represented.

Gene Expression Analyses

Sequencing read libraries were mapped separately to their corresponding final transcriptome (unigene set) using BBmap (see text footnote 3). The software was run with the additional "rpkm" parameter, which yields per-contig raw counts directly along the standard SAM/BAM output files. Next, a raw count matrix was generated for each species' unigene set and fed into edgeR (Robinson et al., 2010) for differential expression testing by fitting a negative binomial generalized log-linear model (GLM) including a tissue factor and a replicate factor to the TMM-normalized read counts for each unigene. Unigenes detected in less than three of the six samples were considered as poorly expressed and filtered out from the analysis. We performed pairwise comparisons of tissues, i.e., flower vs. underground organ (FL vs. MR), flower vs. stem (FL vs. ST), and underground organ vs. stem (MR vs. ST). The distribution of the resulting *p*-values followed the quality criterion described by Rigai et al. (2018). Genes with an adjusted *p*-value [FDR, Benjamini and Hochberg (1995)] below 0.05 were considered to be differentially expressed.

Given the sets of up- and down-regulated genes for each species from pairwise tissue comparisons, we performed enrichment analysis for GO terms, KEGG, and Mapman4 pathways using hypergeometric tests. Terms with an adjusted *p*-value (Bonferroni adjustment) below 0.05 were considered to be enriched.

Comparison of Underground Organ/Stem Expression Profiles Between Autotrophs and Mycoheterotrophs

Biological replicates are required to perform a statistical analysis and identify differentially expressed genes. Another constraint of this analysis was the comparison of the transcriptomes from

²<ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/>

³<https://jgi.doe.gov/data-and-tools/bbtools/>

⁴<https://trinotate.github.io/>

different species. One option is to perform the same analysis as previously for each of the four species and compare the results of the enrichment analyses. However, this would lead only to very broad results at the level of pathways. The other option is to directly compare the four transcriptomes of the four species but this introduces various challenges and biases (Dunn et al., 2013). The first one is to identify the quadruplets of orthologous genes. In this study, we used the expression of the 18,259 orthogroups identified above as a proxy of the expression of the various molecular functions present in the stem and underground organs. This approximation should be taken into account when interpreting the results but is similar to the approach of McWhite et al. (2020). The second one is that the absolute read counts of each species for a given orthogroup cannot be directly compared because the number and length of the genes in each orthogroup can differ from one species to another. To remove this bias, we instead considered the underground organ/stem expression ratios.

As no equivalent dataset is available for autotrophic orchids, we used datasets from *Z. mays* and *B. distachyon* as autotrophic species for comparison. We focused on the underground and stem tissues using roots and internodes as the corresponding tissues for autotrophic monocotyledons. Expression values for *Z. mays* were extracted from the SRA project PRJNA217053. The samples SRR957475 and SRR957476 correspond to internodes, SRR957460 and SRR957461 to roots. Expression values for *B. distachyon* were extracted from the SRA project PRJNA419776. The samples SRR6322422 and SRR6322429 correspond to internodes, SRR6322386 and SRR6322417 to roots. Counts were calculated after mapping of the reads to their corresponding reference transcriptome (*Zea_mays.B73_RefGen_v4.cdna.all.fa* and *Brachypodium_distachyon.Brachypodium_distachyon_v3.0.cdna.all.fa*) using BBmap with the same parameters as previously.

Any orthogroup whose expression was not detected in at least one sample of all four species was filtered out from further analysis. As an orthogroup can group different numbers of genes from each species, the absolute counts cannot be compared directly. However, as the stem and underground organ samples are paired, it is possible to compare the underground organ/stem ratios. After normalization with the TMM method (Robinson et al., 2010) to correct the library size effect, the counts were transformed with the vst method of the cseq package v1.2 (Rau and Maugis-Rabusseau, 2018). The log₂ root/shoot ratios calculated from the transformed counts were analyzed using the lmFit contrasts.fit and eBayes functions of the limma package v3.34.9 (Smyth, 2004). In our model, the log₂ ratio was expressed as a linear combination of a species effect and the *p*-values corresponding to the difference between the average of the two mycoheterotrophic species and the average of the two autotrophic species were calculated. The distribution of the resulting *p*-values followed the quality criterion described by Rigai et al. (2018). The Benjamini–Hochberg correction was used to control false discovery rate. We considered orthogroups with an adjusted *p*-value < 0.05 to have a different underground organ/stem/ ratio between the mycoheterotrophic orchids and the photosynthetic grasses.

Enrichment analyses were performed as described previously with orthogroups being annotated with terms representing at least 25% of their genes.

DATA AVAILABILITY STATEMENT

The reads are available at the NCBI database under Bioproject PRJNA633477. The GFF file and annotation of the unigene sets for *E. aphyllum* and *N. nidus-avis* as well as the raw count matrices are available at <https://doi.org/10.15454/HR9KUX>.

AUTHOR CONTRIBUTIONS

M-AS and ED designed the study. M-AS supervised the project. ED, MM, and MJ analyzed the data. ED, JM, and MJ wrote the manuscript. JC generated the RNA-seq data. JM, MJ, MM, and M-AS collected the samples. ED agreed to serve as the author responsible for contact and ensures communication. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported by grants from the National Science Center, Poland (project No: 2015/18/A/NZ8/00149) to M-AS. The IPS2 benefited from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007).

ACKNOWLEDGMENTS

We thank Emilia Krawczyk for the photos of *E. aphyllum*.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.632033/full#supplementary-material>

Supplementary Figure 1 | The effect of heat on the flowers of *N. nidus-avis*.

Supplementary Table 1 | Details of sampling location and dates for the studied orchids.

Supplementary Table 2 | Genomic datasets used in this study.

Supplementary Table 3 | Comparison of the intermediate and final assemblies generated.

Supplementary Table 4 | Composition of contamination sources among sampled tissues.

Supplementary Table 5 | Annotation statistics of the generated transcriptome assemblies.

Supplementary Table 6 | Summary statistics of the BUSCO analysis of completeness for the generated transcriptomes in comparison to the *E. aphyllum* transcriptome from Schelkunov et al. (2018) and another mycoheterotrophic orchid *G. elata* with a sequenced genome.

Supplementary Table 7 | Statistics of per-tissue read mapping to the intermediate and final assemblies.

Supplementary Table 8 | Per-species statistics among the generated orthologous groups.

Supplementary Table 9 | Species overlaps among orthologous groups.

Supplementary Data 1 | Distribution of GO terms in the three mycoheterotrophic orchids. Only the 20 most abundant terms for each species and each ontology are shown.

Supplementary Data 2 | Comparison of ortholog numbers in Mapman and KEGG pathways for the three mycoheterotrophic orchids and three autotrophic orchids. This excel file includes one sheet for each annotation plus a legend sheet.

Supplementary Data 3 | Output of the Orthofinder analysis. This a tabulated file where each line corresponds to an orthogroup and each column provides the list of proteins (comma separated, using their standard ID) included in the orthogroup for a given species.

Supplementary Data 4 | Composition and annotation of the mycoheterotroph-specific orthogroups. This excel file is composed of one sheet and a legend sheet.

Supplementary Data 5 | Differential expression analysis of *N. nidus-avis* and *E. aphyllum* organs. This excel file includes four sheets plus a legend sheet.

Supplementary Data 6 | Mapman, KEGG, and GO enrichment analysis of *N. nidus-avis* and *E. aphyllum* expression. This excel file contains six sheets and a legend sheet.

Supplementary Data 7 | Differential analysis of the root/shoot expression ratios. This excel file contains two sheets and a legend sheet.

Supplementary Data 8 | Mapman, KEGG, and GO enrichment analysis of the expression ratios.

REFERENCES

- Abadie, C., Lamothe-Sibold, M., Gilard, F., and Tcherkez, G. (2016). Isotopic evidence for nitrogen exchange between autotrophic and heterotrophic tissues in variegated leaves. *Funct. Plant Biol.* 43:298. doi: 10.1071/FP15187
- Adamska, I., Roobol-Bóza, M., Lindahl, M., and Andersson, B. (1999). Isolation of pigment-binding early light-inducible proteins from pea. *Eur. J. Biochem.* 260, 453–460. doi: 10.1046/j.1432-1327.1999.00178.x
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
- Aluru, M. R., Zola, J., Foudree, A., and Rodermel, S. R. (2009). Chloroplast photooxidation-induced transcriptome reprogramming in *Arabidopsis* imtmutans white leaf sectors 1[w][OA]. *Plant Physiol.* 150, 904–923. doi: 10.1104/pp.109.135780
- Ankele, E., Kindgren, P., Pesquet, E., and Strand, A. (2007). In vivo visualization of Mg-protoporphyrin IX, a coordinator of photosynthetic gene expression in the nucleus and the chloroplast. *Plant Cell* 19, 1964–1979. doi: 10.1105/tpc.106.048744
- Bar, M., and Ori, N. (2014). Leaf development and morphogenesis. *Development* 141, 4219–4230. doi: 10.1242/dev.106195
- Barrett, C. F., Freudenstein, J. V., Li, J., Mayfield-Jones, D. R., Perez, L., Pires, J. C., et al. (2014). Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. *Mol. Biol. Evol.* 31, 3095–3112. doi: 10.1093/molbev/msu252
- Barrett, C. F., Sinn, B. T., Kennedy, A. H., and Pupko, T. (2019). Unprecedented parallel photosynthetic losses in a heterotrophic orchid genus. *Mol. Biol. Evol.* 36, 1884–1901. doi: 10.1093/molbev/msz111
- Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., et al. (2013). Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods Ecol. Evol.* 4, 914–919. doi: 10.1111/2041-210X.12073
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300. doi: 10.2307/2346101
- Bernard, N. (1899). Sur la germination du *Neottia nidus-avis*. *C. R. Hebd. Seances Acad. Sci.* 128, 1253–1255.
- Bock, R. (2017). Witnessing genome evolution: experimental reconstruction of endosymbiotic and horizontal gene transfer. *Annu. Rev. Genet.* 51, 1–22. doi: 10.1146/annurev-genet-120215-035329
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Brundrett, M. C., and Tedersoo, L. (2018). Evolutionary history of mycorrhizal symbioses and global host plant diversity. *New Phytol.* 220, 1108–1115. doi: 10.1111/nph.14976
- Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzinger, D., et al. (2017). A tissue-mapped *Axolotl* De Novo transcriptome enables identification of limb regeneration factors. *Cell Rep.* 18, 762–776. doi: 10.1016/j.celrep.2016.12.063
- Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176
- Cameron, D. D., Preiss, K., Gebauer, G., and Read, D. J. (2009). The chlorophyll-containing orchid *Corallorhiza trifida* derives little carbon through photosynthesis. *New Phytol.* 183, 358–364. doi: 10.1111/j.1469-8137.2009.02853.x
- Chen, J., Yu, R., Dai, J., Liu, Y., and Zhou, R. (2020). The loss of photosynthesis pathway and genomic locations of the lost plastid genes in a holoparasitic plant *Aeginetia indica*. *BMC Plant Biol.* 20:199. doi: 10.1186/s12870-020-02415-2
- Claessens, J. (2011). *The Flower of the European Orchid: Form and Function*. Geulle: Jean Claessens & Jaques Kleynen.
- Corratgé-Faillie, C., and Lacombe, B. (2017). Substrate (un)specificity of *Arabidopsis* NRT1/PTR FAMILY (NPF) proteins. *J. Exp. Bot.* 68, 3107–3113. doi: 10.1093/jxb/erw499
- Dearnaley, J., Perotto, S., and Selosse, M. A. (2016). “Structure and development of orchid mycorrhizas,” in *Molecular Mycorrhizal Symbiosis*, ed. F. Martin (Hoboken, NJ: John Wiley & Sons, Inc.), 63–86. doi: 10.1002/9781118951446.ch5
- Dunn, C. W., Luo, X., and Wu, Z. (2013). Phylogenetic analysis of gene expression. *Integr. Comp. Biol.* 53, 847–856. doi: 10.1093/icb/ict068
- Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7:e1002195. doi: 10.1371/journal.pcbi.1002195
- Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238. doi: 10.1186/s13059-019-1832-y
- Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344
- Fitter, A. (2002). “Characteristics and functions of root systems,” in *Plant Roots—the Hidden Half*, eds Y. Waisel, A. Eshel, and U. Kafkafi (New York, NY: Marcel Dekker, Inc.), 21–50.
- Fochi, V., Chitarra, W., Kohler, A., Voyron, S., Singan, V. R., Lindquist, E. A., et al. (2017). Fungal and plant gene expression in the *Tulasnella calospora*–*Serapias vomeracea* symbiosis provides clues about nitrogen pathways in orchid mycorrhizas. *New Phytol.* 213, 365–379. doi: 10.1111/nph.14279
- Gilbert, D. G. (2019). Longest protein, longest transcript or most expression, for accurate gene reconstruction of transcriptomes? *bioRxiv* [Preprint] doi: 10.1101/829184
- Graham, S. W., Lam, V. K. Y., and Merckx, V. S. F. T. (2017). Plastomes on the edge: the evolutionary breakdown of mycoheterotroph plastid genomes. *New Phytol.* 214, 48–55. doi: 10.1111/nph.14398
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084
- Hadariová, L., Vesteg, M., Hampl, V., and Krajčovič, J. (2018). Reductive evolution of chloroplasts in non-photosynthetic plants, algae and protists. *Curr. Genet.* 64, 365–387. doi: 10.1007/s00294-017-0761-0

- Heddad, M. (2000). Light stress-regulated two-helix proteins in *Arabidopsis thaliana* related to the chlorophyll a/b-binding gene family. *Proc. Natl. Acad. Sci. U.S.A.* 97, 3741–3746. doi: 10.1073/pnas.050391397
- Hulten, E., and Fries, M. (1986). *Atlas of North European Vascular Plant, North of the Tropic of Cancer*. Königstein: Koeltz.
- Imhof, S., Massicotte, H. B., Melville, L. H., and Peterson, R. L. (2013). “Subterranean morphology and mycorrhizal structures,” in *Mycoheterotrophy: The Biology of Plants Living on Fungi*, ed. V. Merckx (New York, NY: Springer), 157–214. doi: 10.1007/978-1-4614-5209-6_4
- Jakubka-Busse, A., Jasicka-Misiak, I., Poliwoda, A., Świeczkowska, E., and Kafarski, P. (2014). The chemical composition of the floral extract of *Epipogium aphyllum* Sw. (Orchidaceae): a clue for their pollination biology. *Arch. Biol. Sci.* 66, 989–998. doi: 10.2298/ABS1403989B
- Jung, S. C., Martinez-Medina, A., Lopez-Raez, J. A., and Pozo, M. J. (2012). Mycorrhiza-induced resistance and priming of plant defenses. *J. Chem. Ecol.* 38, 651–664. doi: 10.1007/s10886-012-0134-6
- Klooster, M. R., Clark, D. L., and Culley, T. M. (2009). Cryptic bracts facilitate herbivore avoidance in the mycoheterotrophic plant *Monotropa odorata* (Ericaceae). *Am. J. Bot.* 96, 2197–2205. doi: 10.3732/ajb.0900124
- Krawczyk, E., Rojek, J., Kowalkowska, A. K., Kapusta, M., Znaniecka, J., and Minasiewicz, J. (2016). Evidence for mixed sexual and asexual reproduction in the rare European mycoheterotrophic orchid *Epipogium aphyllum*, Orchidaceae (ghost orchid). *Ann. Bot.* 118, 159–172. doi: 10.1093/aob/mcw084
- Kwong, W. K., del Campo, J., Mathur, V., Vermeij, M. J. A., and Keeling, P. J. (2019). A widespread coral-infecting apicomplexan with chlorophyll biosynthesis genes. *Nature* 568, 103–107. doi: 10.1038/s41586-019-1072-z
- Lallemand, F., Figura, T., Damesin, C., Fresneau, C., Griveau, C., Fontaine, N., et al. (2019a). Mixotrophic orchids do not use photosynthates for perennial underground organs. *New Phytol.* 221, 12–17. doi: 10.1111/nph.15443
- Lallemand, F., Martin-Magniette, M.-L., Gilard, F., Gakière, B., Launay-Avon, A., Delannoy, É, et al. (2019b). In situ transcriptomic and metabolomic study of the loss of photosynthesis in the leaves of mixotrophic plants exploiting fungi. *Plant J.* 98, 826–841. doi: 10.1111/tj.14276
- Leebens-Mack, J. H., Barker, M. S., Carpenter, E. J., Deyholos, M. K., Gitzendanner, M. A., Graham, S. W., et al. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574, 679–685. doi: 10.1038/s41586-019-1693-2
- Lim, G. S., Barrett, C. F., Pang, C. C., and Davis, J. I. (2016). Drastic reduction of plastome size in the mycoheterotrophic *Thismia tentaculata* relative to that of its autotrophic relative *Tacca chantrieri*. *Am. J. Bot.* 103, 1129–1137. doi: 10.3732/ajb.1600042
- Logacheva, M. D., Schelkunov, M. I., Nuraliev, M. S., Samigullin, T. H., and Penin, A. A. (2014). The plastid genome of mycoheterotrophic monocot *Petrosavia stellaris* exhibits both gene losses and multiple rearrangements. *Genome Biol. Evol.* 6, 238–246. doi: 10.1093/gbe/evu001
- Lunn, J. E., Delorge, I., Figueroa, C. M., Van Dijck, P., and Stitt, M. (2014). Trehalose metabolism in plants. *Plant J.* 79, 544–567. doi: 10.1111/tj.12509
- McKendrick, S. L., Leake, J. R., Taylor, D. L., and Read, D. J. (2002). Symbiotic germination and development of the myco-heterotrophic orchid *Neottia nidus-avis* in nature and its requirement for locally distributed *Sebacina* spp. *New Phytol.* 154, 233–247.
- McWhite, C. D., Papoulas, O., Drew, K., Cox, R. M., June, V., Dong, O. X., et al. (2020). A pan-plant protein complex map reveals deep conservation and novel assemblies. *Cell* 181, 460.e–474.e. doi: 10.1016/j.cell.2020.02.049
- Menke, W., and Schmid, G. H. (1976). Cyclic photophosphorylation in the mykorrhizal orchid *Neottia nidus-avis*. *Plant Physiol.* 57, 716–719. doi: 10.1104/pp.57.5.716
- Merckx, V., Bidartondo, M. I., and Hynson, N. A. (2009). Myco-heterotrophy: when fungi host plants. *Ann. Bot.* 104, 1255–1261. doi: 10.1093/aob/mcp235
- Merckx, V., and Freudenstein, J. V. (2010). Evolution of mycoheterotrophy in plants: a phylogenetic perspective. *New Phytol.* 185, 605–609. doi: 10.1111/j.1469-8137.2009.03155.x
- Merckx, V. S. F. T. (2013). *Mycoheterotrophy: The Biology of Plants Living on Fungi*. New York, NY: Springer-Verlag.
- Mohanta, T. K., Mishra, A. K., Khan, A., Hashem, A., Abd-Allah, E. F., and Al-Harrasi, A. (2020). Gene loss and evolution of the plastome. *Genes (Base)*. 11:1133. doi: 10.3390/genes11011133
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, W182–W185. doi: 10.1093/nar/gkm321
- Müller, J., and Dulieu, H. (1998). Enhanced growth of non-photosynthesizing tobacco mutants in the presence of a mycorrhizal inoculum. *J. Exp. Bot.* 49, 707–711. doi: 10.1093/jxb/49.321.707
- Nilsson, R., Larsson, K., Taylor, A., Bengtsson-Palme, J., Jeppesen, T., Schigel, D., et al. (2019). The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.* 47, D259–D264. doi: 10.1093/nar/gky1022
- Niyogi, K. K., Björkman, O., and Grossman, A. R. (1997). The roles of specific xanthophylls in photoprotection. *Proc. Natl. Acad. Sci. U.S.A.* 94, 14162–14167. doi: 10.1073/pnas.94.25.14162
- Pellicer, J., and Leitch, I. J. (2020). The plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comparative studies. *New Phytol.* 226, 301–305. doi: 10.1111/nph.16261
- Perotto, S., Rodda, M., Benetti, A., Sillo, F., Ercole, E., Rodda, M., et al. (2014). Gene expression in mycorrhizal orchid protocorms suggests a friendly plant-fungus relationship. *Planta* 239, 1337–1349. doi: 10.1007/s00425-014-2062-x
- Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701
- Pfeifhofer, H. W. (1989). Evidence for chlorophyll b and lack of lutein in *Neottia nidus-avis* plastids. *Biochem. Physiol. Pflanz.* 184, 55–61. doi: 10.1016/s0015-3796(89)80120-9
- Rasmussen, H. N. (1995). *Terrestrial Orchids: From Seed to Mycotrophic Plant*. Cambridge: Cambridge University Press.
- Rau, A., and Maudis-Rabusseau, C. (2018). Transformation and model choice for RNA-seq co-expression analysis. *Brief. Bioinform.* 19, 425–436. doi: 10.1093/bib/bbw128
- Ravin, N. V., Gruzdev, E. V., Beletsky, A. V., Mazur, A. M., Prokhortchouk, E. B., Filyushin, M. A., et al. (2016). The loss of photosynthetic pathways in the plastid and nuclear genomes of the non-photosynthetic mycoheterotrophic eudicot *Monotropa hypopitys*. *BMC Plant Biol.* 16 (Suppl 3):153–161. doi: 10.1186/s12870-016-0929-7
- Rebeiz, C. A., Montazer-Zouhoor, A., Hopen, H. J., and Wu, S. M. (1984). Photodynamic herbicides: I. concept and phenomenology. *Enzyme Microb. Technol.* 6, 390–396. doi: 10.1016/0141-0229(84)90012-7
- Rich, M. K., Nouri, E., Courty, P. E., and Reinhardt, D. (2017). Diet of arbuscular mycorrhizal fungi: bread and butter? *Trends Plant Sci.* 22, 652–660. doi: 10.1016/j.tplants.2017.05.008
- Rigaill, G., Balzergue, S., Brunaud, V., Blondet, E., Rau, A., Rogier, O., et al. (2018). Synthetic data sets for the identification of key ingredients for RNA-seq differential analysis. *Brief. Bioinform.* 19, 65–76. doi: 10.1093/bib/bbw092
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Rossini, S., Casazza, A. P., Engelmann, E. C. M., Havaux, M., Jennings, R. C., and Soave, C. (2006). Suppression of both ELIP1 and ELIP2 in *Arabidopsis* does not affect tolerance to photoinhibition and photooxidative stress. *Plant Physiol.* 141, 1264–1273. doi: 10.1104/pp.106.083055
- Roy, M., Yagame, T., Yamato, M., Koji, I., Heinz, C., Faccio, A., et al. (2009). Ectomycorrhizal *Inocybe* species associate with the mycoheterotrophic orchid *Epipogium aphyllum* but not its asexual propagules. *Ann. Bot.* 104, 595–610. doi: 10.1093/aob/mcn269
- Schelkunov, M. I., Penin, A. A., and Logacheva, M. D. (2018). RNA-seq highlights parallel and contrasting patterns in the evolution of the nuclear genome of fully mycoheterotrophic plants. *BMC Genomics* 19:602. doi: 10.1186/s12864-018-4968-3
- Schelkunov, M. I., Shtratnikova, V., Nuraliev, M., Selosse, M., Penin, A. A., and Logacheva, M. D. (2015). Exploring the limits for reduction of plastid genomes: a case study of the mycoheterotrophic orchids *Epipogium aphyllum* and *Epipogium roseum*. *Genome Biol. Evol.* 7, 1179–1191.
- Schwacke, R., Ponce-Soto, G. Y., Krause, K., Bolger, A. M., Arsova, B., Hallab, A., et al. (2019). MapMan4: a refined protein classification and annotation framework applicable to multi-omics data analysis. *Mol. Plant* 12, 879–892. doi: 10.1016/j.molp.2019.01.003

- Selosse, M.-A., and Roy, M. (2009). Green plants that feed on fungi: facts and questions about mixotrophy. *Trends Plant Sci.* 14, 64–70. doi: 10.1016/j.tplants.2008.11.004
- Selosse, M. (2003). La Néottie, une “mangeuse” d’arbres. *L’Orchidophile* 155, 21–31.
- Selosse, M. A., Weiß, M., Jany, J. L., and Tillier, A. (2002). Communities and populations of sebacinoïd basidiomycetes associated with the achlorophyllous orchid *Neottia nidus-avis* (L.) L.C.M. Rich. and neighbouring tree ectomycorrhizae. *Mol. Ecol.* 11, 1831–1844. doi: 10.1046/j.1365-294X.2002.01553.x
- Seppey, M., Manni, M., and Zdobnov, E. M. (2019). “BUSCO: assessing genome assembly and annotation completeness,” in *Methods in Molecular Biology*, ed. M. Kollmar (New York, NY: Humana Press Inc.), 227–245. doi: 10.1007/978-1-4939-9173-0_14
- Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3:3. doi: 10.2202/1544-6115.1027
- Suetsugu, K., Yamato, M., Miura, C., Yamaguchi, K., Takahashi, K., Ida, Y., et al. (2017). Comparison of green and albino individuals of the partially mycoheterotrophic orchid *Epipactis helleborine* on molecular identities of mycorrhizal fungi, nutritional modes and gene expression in mycorrhizal roots. *Mol. Ecol.* 26, 1652–1669. doi: 10.1111/mec.14021
- Taylor, L., and Roberts, D. L. (2011). Biological flora of the British Isles: *Epipogium aphyllum* Sw. *J. Ecol.* 99, 878–890. doi: 10.1111/j.1365-2745.2011.01839.x
- Těšitel, J., Tišitelová, T., Minasiwicz, J., and Selosse, M. A. (2018). Mixotrophy in land plants: why to stay green? *Trends Plant Sci.* 23, 656–659. doi: 10.1016/j.tplants.2018.05.010
- van der Heijden, M. G. A., Martin, F. M., Selosse, M. A., and Sanders, I. R. (2015). Mycorrhizal ecology and evolution: the past, the present, and the future. *New Phytol.* 205, 1406–1423. doi: 10.1111/nph.13288
- Vesely, P., Bures, P., Smarda, P., and Pavlicek, T. (2012). Genome size and DNA base composition of geophytes: the mirror of phenology and ecology? *Ann. Bot.* 109, 65–75. doi: 10.1093/aob/mcr267
- Vogel, A., Schwacke, R., Denton, A. K., Usadel, B., Hollmann, J., Fischer, K., et al. (2018). Footprints of parasitism in the genome of the parasitic flowering plant *Cuscuta campestris*. *Nat. Commun.* 9:2515. doi: 10.1038/s41467-018-04344-z
- Wang, B., and Qiu, Y. L. (2006). Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* 16, 299–363. doi: 10.1007/s00572-005-0033-6
- Wang, H., Kong, F., and Zhou, C. (2021). From genes to networks: the genetic control of leaf development. *J. Integr. Plant Biol.* doi: 10.1111/jipb.13084
- Wickett, N. J., Honaas, L. A., Wafula, E. K., Das, M., Huang, K., Wu, B., et al. (2011). Transcriptomes of the parasitic plant family orobanchaceae reveal surprising conservation of chlorophyll synthesis. *Curr. Biol.* 21, 2098–2104. doi: 10.1016/j.cub.2011.11.011
- Yoshida, S., Kim, S., Wafula, E. K., Tanskanen, J., Kim, Y. M., Honaas, L., et al. (2019). Genome sequence of *Striga asiatica* provides insight into the evolution of plant parasitism. *Curr. Biol.* 29, 3041.e–3052.e. doi: 10.1016/j.cub.2019.07.086
- Yuan, Y., Jin, X., Liu, J., Zhao, X., Zhou, J., Wang, X., et al. (2018). The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. *Nat. Commun.* 9:1615. doi: 10.1038/s41467-018-03423-5
- Yudina, S. V., Schelkunov, M. I., Nauheimer, L., Crayn, D., Chantanaorrapint, S., Hroneš, M., et al. (2021). Comparative analysis of plastid genomes in the non-photosynthetic genus *Thismia* reveals ongoing gene set reduction. *Front. Plant Sci.* 12:602598. doi: 10.3389/fpls.2021.602598
- Zeng, X., Li, Y., Ling, H., Liu, S., Liu, M., Chen, J., et al. (2017). Transcriptomic analyses reveal clathrin-mediated endocytosis involved in symbiotic seed germination of *Gastrodia elata*. *Bot. Stud.* 58:31. doi: 10.1186/s40529-017-0185-7
- Ziegenspeck, H. (1936). *Orchidaceae: Lebensgeschichte der Blütenpflanzen Mitteleuropas*. Stuttgart: Eugen Ulmer.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Jałkalski, Minasiwicz, Caius, May, Selosse and Delannoy. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

6 Methodology

Within this work, a number of diverse strategies and techniques was employed to generate and analyze data. This chapter contains brief listing and summary of the most mention-worthy methodological approaches that contributed to the presented work – for detailed methodologies please refer to individual research articles (Chapter 5) and relevant chapters of supplementary material (Chapter 10).

6.1 DNA sequencing

In presented works, several specifically optimized approaches were used for generating DNA sequences for various applications.

For generating plastidial sequences in May et al. 2019, Lallemand et al. 2019b and Lallemand 2019d (Chapters 5.1, 5.3, 5.2, respectively), DNA material was extracted using DNeasy Plant Mini kit (Qiagen) as a method of choice. Selected samples in Lallemand et al. 2019b were extracted using NucleoSpin Plant II kit (Macherey-Nagel) and CTAB isolation protocol after Porebski et al. 1997. The differences were caused by having multiple teams performing experiments in different locations using available resources. Sequencing libraries were constructed using Accel-NGS 1s plus DNA library kit (or TruSeq (Illumina) and NEBNext Ultra DNA kit (New England Biolabs) applied by laboratory in Moscow in Lallemand et al. 2019b). Sequencing was performed using Illumina HiSeq 2000, 2500, 4000 and MiSeq. The subject of DNA sequencing is further extended in Chapter 10.5.2 supplementary data.

For fungal barcoding in May et al. 2020 (Chapter 5.4), DNA extraction and amplification was performed using Qiagen DNeasy Plant mini kit and REDExtract-N-Amp kit (Sigma Aldrich). Sequencing was carried out by Ion Torrent (Thermo Fisher Scientific). The subject of fungal barcoding is further extended in Chapter 10.8 supplementary data.

6.2 RNA sequencing

In Jąkowski et al., 2021 (Chapter 5.5) and in unpublished research presented in this work (Chapter 8), a common protocol for handling RNA material was established and optimized for specific types of plant material. RNA extraction was performed using NucleoZOL (Macherey-Nagel) protocol with addition of PVPP. Isolates were purified using Agencourt RNAClean XP kit (Beckman Coulter) and subjected to DNase digestion. Integrity

6. Methodology

(RIN) control was performed by Agilent BioAnalyzer 2100 assay on Plant RNA nano chips. Libraries were constructed using TruSeq Stranded Total RNA with Ribo-Zero Plant kit (Illumina) and sequenced with NextSeq 500 (Illumina) for Jąkalski et al., 2021 (chapter 5.5), and HiSeq 2500 (Illumina) for unpublished research described in this work (chapter 8). The subject of RNA sequencing is further extended in Chapter 10.6 supplementary data.

Long read sequences for unpublished study on hybrid *de novo* assembly (chapter 8) were generated by MinIon (Oxford Nanopore Technologies) sequencing of libraries constructed by SQK-PCB109 (Oxford Nanopore Technologies) kit. The subject of LR sequencing is further extended in Chapter 10.5.2 supplementary data.

6.3 Stable isotope analyses

In May et al. 2020 (Chapter 5.4), isotopic content analysis was performed to measure N concentration and $^{13}\text{C}/^{12}\text{C}$ and $^{15}\text{N}/^{14}\text{N}$ ratios. The measurements were carried out by an environmental analyzer coupled to ThermoFinnigan Delta V Advantage continuous flow isotope ratio mass spectrometer. The subject of isotopic content analysis is further extended in Chapter 10.10 supplementary data.

6.4 In silico data processing

Multiple bioinformatic data handling techniques were applied in the process of data generation, processing and analysis.

For the plastid sequence assembly, depending on the processed species, Geneious suite software (www.geneious.com) and CLC Genomics Workbench (digitalinsights.qiagen.com) *de novo* assembly procedure were employed. Alignments for the purpose of annotation of sequences of plastidial origin and phylogenetic inference, were conducted by, respectively, MAFFT (Kazutaka et al., 2002) and BLAST (Altschul et al., 1990) tools. For selective regime analyses, we used PAML (<http://abacus.gene.ucl.ac.uk/software/paml.html>) tools and RELAX (<http://datamonkey.org/relax>). This subject is further extended in Chapter 10.7.5 supplementary data.

In transcriptomic experiments, for sequence assembly we employed Trinity RNAseq assembler (Grabherr et al., 2013), followed by refining by tr2aacds pipeline from EvidentialGene package (<http://arthropods.eugenesis.org/EvidentialGene/trassembly.html>). Contaminant sequences were identified by blastx alignment to NCBI NR database performed

with Diamond software (Buchfink et al., 2021). Multiple tools and databases were used to annotate generated contigs and perform comparative analysis of differential gene expression, described in-detail in Jąkalski et al., 2021 (Chapter 5.5). The subject of transcriptome assembly is further extended in Chapter 10.6.8 supplementary data.

6.5 Plant tissue sampling procedure

Within the presented work, a unified procedure for *in situ* plant sampling for transcriptomic procedures was constructed. Multiple researched species were impossible to cultivate in controlled conditions, therefore sampling of wild specimens was necessary. The sampling was carried out during a narrow time window around 10:00 AM. Double biological replicates were selected for statistical relevancy of results. Plants were excavated with surrounding soil. Fully developed inflorescences were collected, along with stems. Root sections were mechanically cleaned of soil remnants and rinsed. Material was divided into samples of 150mg and cryopreserved in LN₂ to prevent decay of RNA. Additional verification of presence of symbiotic mycobionts was conducted by optical microscope observation of selected root cross-sections. This procedure was most notably applied in Jąkalski et al., 2021 (Chapter 5.5) and is further discussed in Chapter 9.1.

7 Discussion

7.1 Framework of presented work

The main reason and goal of this research was our intent to better understand evolutionary context and genetic basis of transition between plant trophic modes in Orchidaceae, beginning from their autotrophic common ancestors, through an intermediate stage of mixotrophy, to strict achlorophyllous mycoheterotrophs. The presented work was initiated with an exploratory approach, without clearly stated hypothesis. As the guiding milestones in my investigation, a set of broadly outlined goals was appointed: first, to build a “baseline” for further research into modified trophism, we needed to survey the condition of plastomes of AT species, expecting to describe their fully functional photosynthetic machinery. The sequences delivered for *P. chlorantha* and *D. majalis* confirmed these expectations (Chapters 5.1 and 5.2). Next stage involved similar analysis performed on a group of species reported as MX. A comparison of these would reveal possible evolutionary tendencies involving preservation or degeneration of photosynthetic apparatus in presence of alternative nutrient source as an early prerequisite for further progress towards MH and possibly indicate whether reversal of this transition would be likely if facilitated by specific environmental condition. Among most of the researched species, no reduction in photosynthetic functionality was observed, and the gene losses detected in 3 MX species did not affect the capability of reversal to full autotrophy (Chapter 5.3). Observed tendency did not reflect the degree of relying on a fungal partner for nutrition, but rather appeared to be species-specific. An additional opportunity to study nutrition of a mixotroph in an environment lacking primary suppliers for mycorrhizal nutrition was taken advantage of and provided further indication that reversal to autotrophy is a viable strategy in conditions that would facilitate this choice for a mixotroph (Chapter 5.4). Lastly, MH species needed to be introduced into the survey as a final evolutionary gradient stage. Their highly degraded plastomes have been previously described but were not sufficient to be projected onto the overall condition of plant’s gene pool. Without access to genomic data, we settled on a transcriptomic approach to detect possible genetic novelties required for achieving MH and to observe how the metabolic transition affected general state of plant’s transcriptome, possibly determining genetic basis of occurring changes. Furthermore, the expression profile comparison between MH plant’s organs were expected to bring insight into the alterations in plant’s functioning. The investigation provided no evidence of novel mechanisms and

found loss-of-function events only, while simultaneously indicating a high degree of reprogramming in gene expression, most notably in transport-related mechanisms (Chapter 5.5). The challenging nature of generating transcriptomes for non-model species without genomic reference created a need for improvement in *de novo* assembly procedure, which was addressed by evaluating a multitude of bioinformatic tools in a hybrid approach, combining short and long sequencing reads to overcome limitations of both technologies, and indicating potential for further improvement (Chapter 8).

7.2 Plastomes of *P. chlorantha* and *D. majalis* present a solid genetic image of autotrophy

The sequencing and assembly of two plastidial genomes (Chapter 5.1 and 5.2) delivered a valuable dataset for future works involving phylogeny and genetic diversity in the Orchidaceae family. As previously described (Bidartondo et al. 2004) both species were observed to be strictly autotrophic in adulthood, therefore we expected their plastidial genomes' structure to be consistent with those of other strictly photosynthetic plants. The chloroplast genome of vascular plants usually exhibits high conservation of its general linear layout, structural organization, and its coding regions, while exhibiting more variability in its overall length. Chloroplast DNA (cpDNA) is organized into a single, circular chromosome, ranging from 100 to over 200 kbp (*Pelargonium*) in length, and containing on average around 120-160 genes involved primarily in photosynthesis, transcription, and translation (Pyke, 2009). Both newly sequenced plastomes exhibited the expected, average size (154 260 and 154 108 bp respectively), and we were able to annotate 134 genes (113 unique) in both cases.

With seldom exceptions, cpDNA of fully functional plastid of an autotroph contains two regions of inverse repeats (IR). In higher plants, both IRs consist of approximately 20 – 25 kbp long sequences and are a mirror image of each other in terms of gene complement. They are separated, from one side, by Small Single Copy region (SSC) and Large Single Copy region (LSC) from the other (Daniell et al., 2016) – as presented in Fig.2 on example of well researched model species, *Nicotiana tabacum*. We observed this exact organization of structure in both newly sequenced plastomes, as expected of a plant carrying fully functional autotrophic nutrition machinery. This state of plastome content and organization

7. Discussion

will be further considered a baseline for autotrophic members of Orchidaceae for comparative purposes in the scope of this compilatory work.

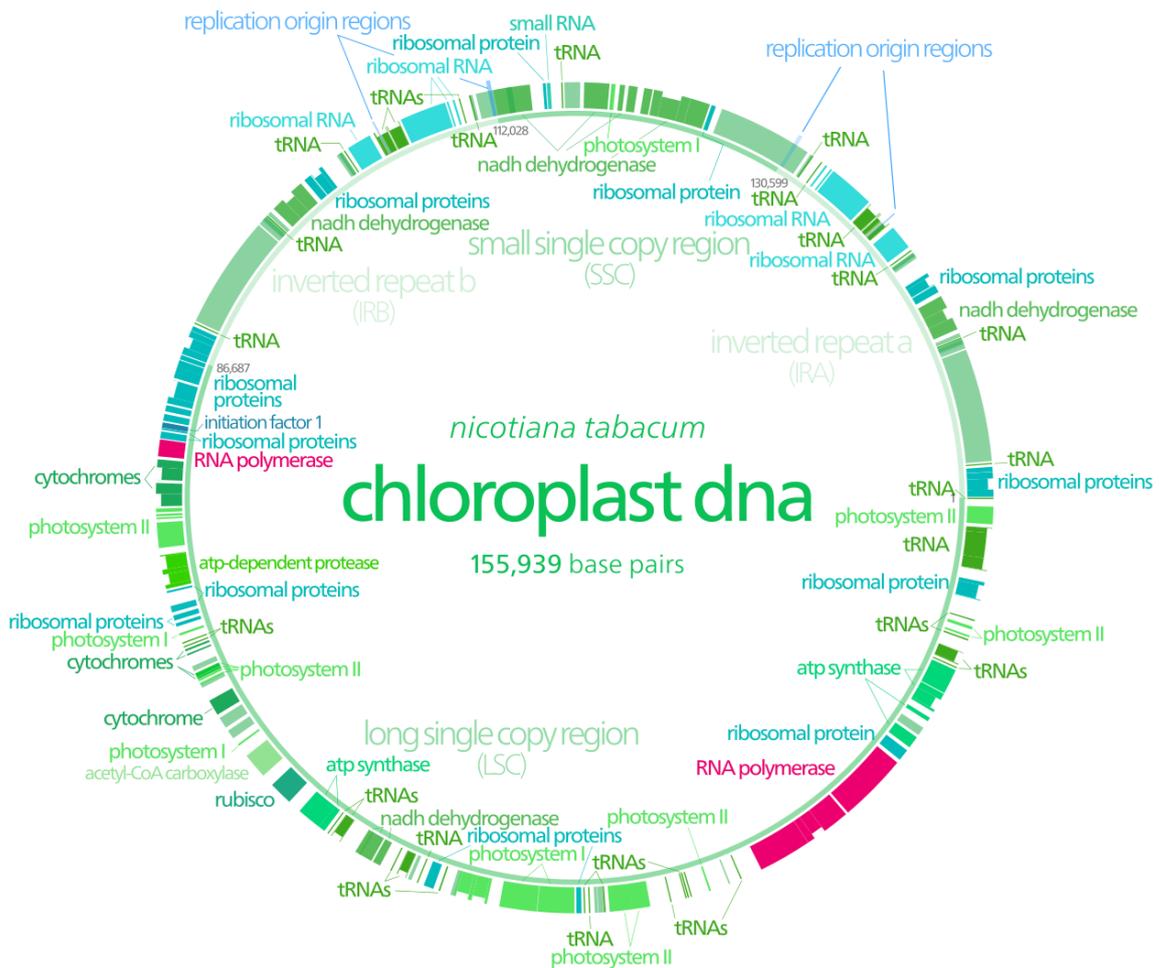


Fig. 2 Structure of a circular plastidial genome on the example of first sequenced plastome, of tobacco plant. Kelvin Ma, after Wakasugi et al., 1998; CC BY-SA 3.0

7.3 Description of 13 plastidial genomes of AT and MX Neottieae explains the possibility of elasticity in nutrition

As our plastome sequencing enterprise was expanded by a set of 13 new Orchidaceae species, the conducted experiment allowed us to generate valuable datasets and draw conclusions on changes that appear in plastomes of mixotrophs in the course of evolutionary process of heading towards MH nutrition. Plastomes of most described species followed the structure characteristic for previously described autotrophs – with the exception of *E. microphylla*, *N. cordata* (highly adaptable mixotroph (Tesitelova et al. 2015)), and *L. abortivum* (mixotroph with high tendency towards MH and strongly reduced AT), which

7. Discussion

exhibited limited gene loss. However, no strict correlation was detected between the degree of fungal dependency of mixotrophs and degradation of their plastomes, as observed on the example of NDH genes.

The evolutionary process of plastome reduction has been previously described for multiple mixotrophic and mycoheterotrophic species along the gradient of increasing fungal dependency. Due to the relaxation of selective pressure by removing the need for photosynthesis and by redundancy caused by nuclear gene transfer, mycoheterotrophic orchids exhibit far-fetched reductions in their plastidial genomes. A common model of plastome loss in MH orchid species has been described (Graham et al., 2017) (Figure 3).

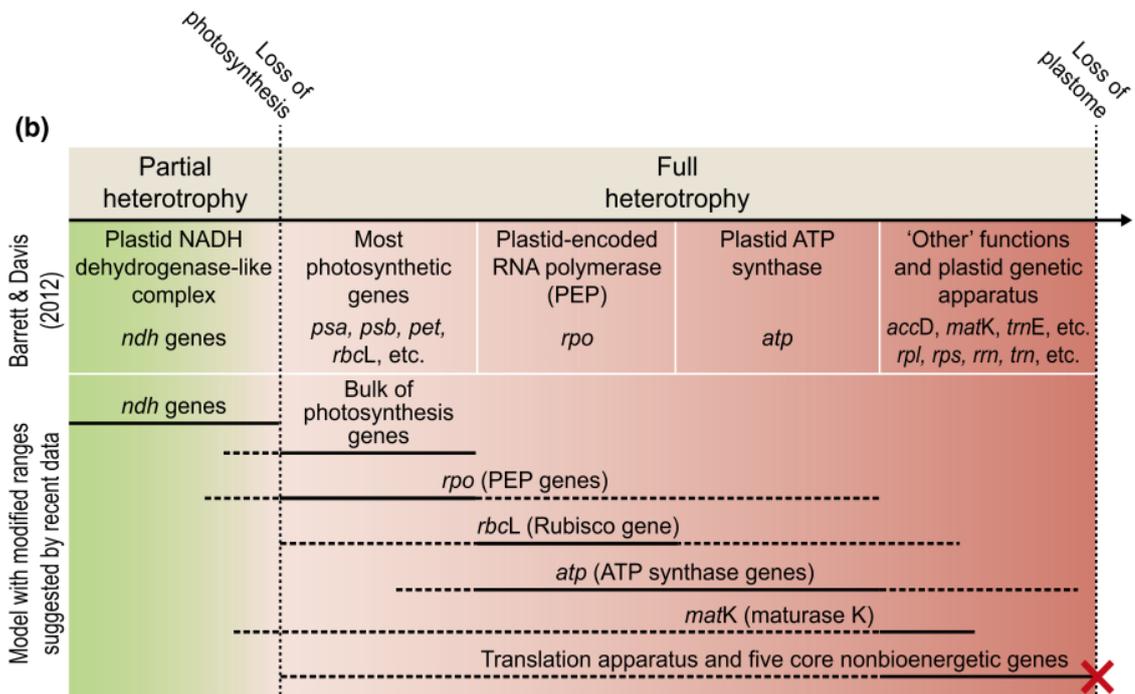


Fig. 3 A diagram illustrating the process of plastome reduction during loss of photosynthetic capability – after Graham et al. 2017

The first plastidial genes that commonly undergo reduction or pseudogenization, often before the actual loss of photosynthesis occurs, are 11 NADH dehydrogenase-like complex NDH1 coding *ndh* genes (Hadariová et al., 2018), as observed here in *E. microphylla*, *N. cordata* and *L. abortivum*. Being a ferredoxin-plastoquinon reductase, NDH1 is believed to act as an electron flow regulator, oxidative stress moderator and a mean for fine-tuning of photosynthesis efficiency. Its functions can be easily substituted by nuclear-encoded genes and therefore its loss was previously reported in multiple non-MH green lineages. We cannot detect such nuclear complementation events solely based on the

7. Discussion

plastomic data - a total transcriptome analysis is required. The total loss of *ndh* genes may be an adaptation to low light conditions in forest understory (related to low oxidative stress). Being irreversible, however, it results in difficulty in re-adaptation to bright conditions, and pushes the plant further towards the evolutionary switch to MH nutrition. In most cases, the process of plastome reduction during transition to mycoheterotrophy follows a similar tendency, despite exhibiting minor differences in order of events. The differences seem to be strongly lineage-specific (Graham et al., 2017).

Along with previously described evidence (Feng et al., 2016; Jin et al., 2014), the degradation process appears to be species-specific in Neottieae family. No significant changes in selective pressure on these genes were detected, suggesting that their loss is facilitated by selective relaxation present in an entire Orchidaceae family without correlation with their nutrition strategy. This observation seems to contradict a principle observed previously in other plant families that developed mixotrophy (carnivorous and hemiparasitic plants) (Petersen et al., 2015b; Wicke et al., 2016, 2014), that indicates relaxation of selective pressure on retaining intact plastomic sequences in mixotrophs. Compared to other families that developed alternative sources of carbon nutrients, no losses of genes or selective pressure relaxation on the photosynthesis-related genes were observed in Neottieae. The advantage of upholding photosynthetic apparatus is potentially related to their reproductive strategy and beneficial role of photosynthetic carbon in successful fruit formation.

Unfortunately, no unambiguous explanation could be delivered on Neottieae evolutionary history based on our current, limited dataset. Two possible explanations were formulated. If we assume that contemporary Neottieae originate from an autotrophic common ancestor, at least 4 separate events of emergence of mixotrophy would have had to occur in progress of formation of current phylogenetic tree. On the other hand, if a single mixotrophic common ancestor was involved, occurrence of 3 events of regression to AT would be needed to explain the current state. Due to selective pressure that facilitates retaining of photosynthetic capability in mixotrophic Neottieae and high plasticity widely exhibited by mixotrophic species, occurrence of such events is not unlikely and both hypotheses are plausible. The model basing on common MX ancestor is more parsimonious, i.e., provides a simpler explanation requiring fewer major changes, but further work and expanding the scope of this research by new species with clearly indicated trophism is required to resolve this dilemma and deliver a strong conclusion.

In a narrowed scope of this dissertation, this part of my work commits towards explaining the basis of high nutritional plasticity exhibited by mixotrophic orchids. A tendency to retain fully functional plastomes on the course of evolutionary progression towards mycoheterotrophy is necessary for MX plants to preserve their ability to regress back to AT in conditions that would favor such transition, be it an evolutionary change or fine-tuning of plant-fungus interaction and internal nutrition dynamics.

7.4 *E. helleborine* without access to mycorrhizal network further proves high plasticity between AT and MX

In this part of my research an unusual collection of specimens of a mixotrophic species, *Epipactis helleborine*, was investigated to survey the species' dependency on its fungal partner and ability to survive without connection to a mycorrhizal network. The specimens were explanted from their habitat and cultivated in pots inside a greenhouse. The conditions of their growth exclude the possibility of integration with mycorrhizal network involving sources of photosynthetic carbon. Despite this, the subject plants exhibited 1-3 years survival without inhibition of development and flowering. In their mycorrhizas we detected a high abundance (compared to *in natura*) of rhizoctonia (common mycorrhizal partner of AT orchids) over EcM fungi, that are considered main mycorrhizal partners of mixotrophic *E. helleborine* (Ogura-Tsujita and Yukawa 2008) in nature. In further research of such cases, it would be beneficial to collect root samples before explantation from *in situ* conditions to observe changes in composition of mycorrhizal species on a specimen basis.

This experiment did not compare the viability between *in natura* and explanted specimens in quantitative terms. Compared to previously described explantation cases, which concentrated on ability to reconnect to a new mycorrhizal network, we deliver proof that *E. helleborine* is able to survive without this connection whatsoever, and not suffer any critical developmental impairments.

The isotopic signatures measured in pot-cultured *E. helleborine* did not differ significantly from surrounding autotrophic species, and no enrichment was detected in ^{13}C between *Epipactis*' leaves and roots. Furthermore, *Epipactis* exhibited lower average nitrogen content than reference autotrophs. The measured abundance of ^{13}C and ^{15}N and total N content did not provide significant evidence to support influx of fungal carbon from EcM fungi normally described in MX plants. These findings, together with low presence of EcM species, seemingly indicate that AT nutrition was performed by this otherwise MX orchid.

However, considering the possibility of a bi-directional nature of carbon transfer between mycorrhizal partners, in this experiment only the raw flow of nutrients can be observed - the net balance of transfer remains unknown. Rhizoctonia, which contribute to majority of detected mycobionts, are saprotrophs and utilize carbon of plant origin. Therefore, the influx of carbon delivered by these species would not be detected alongside fungal carbon presence in isotopic ratio measurement. Currently it is unclear and a matter of discussion whether rhizoctonia engage in delivering carbon to associated AT *Orchidaceae*. To derive a full confirmation of a switch from MX to AT, additional experiments are required. However, along with plastomic data assembled in previous research, indicating an intact and fully functional plastidial genome in *E. helleborine*, we observe a state that highly resembles that of closely related AT orchid species. Considering these findings in the context of evolutionary history of *Neottieae* described earlier, the possibility of performing a reversion from MX to AT nutrition as an adjustment to specific environmental conditions is probable. We can assume, that under high abundance of light and lack of alternative carbon source from mycorrhizal partners, mixotrophic species can adjust by exhibiting behavior very similar to autotrophic plants. This strongly contributes to a theory of high adaptive flexibility of mixotrophy.

7.5 MH species emerge by expression reprogramming and function losses, rather than metabolic innovations

To represent a final stage on the evolutionary gradient of plant nutrition, transcriptomes of two strictly mycoheterotrophic species were sequenced and analyzed: *Epipogium aphyllum* (EA) and *Neottia nidus-avis* (NNA). Both being achlorophyllous, they indisputably exhibit a purely MH strategy of carbon acquisition. A comparison to a reference genome from *Gastrodia elata*, another MH orchid species, uncovered a highly convergent gene set shared by MH species originating from separate loss-of-autotrophy evolutionary events. Based on the newly generated transcriptomic data, we can state that no novel genetic mechanisms were detected. It is possible that novel genes were missed in our assemblies due to assembling and cleanup error caused by the drawbacks of *de novo* sequencing approach without genomic reference available for our species, but considering all *Orchidaceae* being mycoheterotrophic during germination, we can expect all the genes necessary for this trophism to be commonly present in the family, in AT, MX and MH species alike. Under

7. Discussion

such assumption, no genetic novelties are required to perform a trophic shift to MH. We speculate that most of these common traits are based on regulated use of transporters, but to describe this mechanism in detail further research is required. Previous studies on *G. elata* genome indicated expansions of multiple MH-related gene groups. We are, unfortunately, unable to confirm such changes basing on transcriptomic data only – a high quality genomic sequencing would be required to perform such observation. The general tendency widely exhibited by investigated transcriptomes is reduction of molecular functions compared to AT species.

According to Graham's et al. 2017 model of gene loss during transition to MH nutrition (Figure 3), loss of previously described *ndh* genes (completely lost from EA and NNA) is usually followed by reduction of bulk photosynthesis genes. A knockout at any step of photosynthetic reaction chain disables the entire functionality, causing total loss of photosynthesis, removing selective pressure for its maintenance (Barrett et al., 2014; Barrett and Davis, 2012). Pseudogenized genes rapidly accumulate point mutations, and selection for lower energy cost promotes simplification of plastome by their gradual removal. In the cases investigated in this study, no orthologs responsible for photosystems were detected, but chlorophyll-related pathways remained largely unaffected – gene loss events were mostly isolated to chlorophyll interconversion and degradation pathways. In both researched mycoheterotrophs, chlorophyll synthesis pathway remained (at least partially) conserved, congruently to observations on MH plants from other families (Barrett et al., 2014; Wickett et al., 2011). It is unclear which non-photosynthetic function of chlorophyll intermediates facilitates their conservation. *N. nidus-avis* retains a fully functional chlorophyll synthesis pathway, activity of which was detected mainly in inflorescence, together with several other remains of plastome-related activity. It is speculated that minimal level photosynthesis may be conducted within these organs – a number of MX species use photosynthesis predominantly in production of inflorescence and fruit (Lallemand et al., 2019a), suggesting a possible evolutionary remnant of this function in form of trace photosynthetic activity. It is, however, unlikely, as most major genes responsible for formation of photosystems are missing from MH transcriptomes. Another theory speculates that in *N. nidus-avis* chlorophyll might be employed purely as a dye to provide camouflage. The metabolic switch from MX to MH in investigated species bases on gene losses. Their general response to the loss of photosynthetic functionality proceeds similarly to what was observed in albinotic MX specimens (Lallemand et al. 2019c). The possibility of survival without photosynthetic

7. Discussion

carbon source appears to be a result of high plasticity of plant metabolism, and as the emergence of MH nutrition is not an isolated occurrence, it is likely an effect of simple loss of function event, rather than an incredibly coherent, convergent gain of complex mechanism and functionality.

The next group of genes to be lost in Graham's et al. 2017 model are usually *rbcL* (present in NNA, lost in EA), *atp* (mostly lost in both EA and NNA) and *rpo* (PEP genes – completely lost in EA and NNA). Because of their dual functionality they are more resistant to genome reduction – Rubisco, for example, is retained in multiple mycoheterotrophic lineages, most likely for its role in lipid biosynthesis. Similarly, ATP synthase can be alternatively employed for proton gradient creation through ATP hydrolysis, to enable transport of protein through thylakoid membrane. PEP knockout is commonly observed in most transitions to MH. The last stage of plastome reduction involves loss of plastid genetic apparatus and five core non-bioenergetic genes. In all currently known MH plastomes, even those highly degraded, at least one of these genes is retained (Delannoy et al., 2011; Kim et al., 2015).

Leaf reduction to a form of scales is a notable simplification in MH orchids, logically applied for reduction of expenses on a non-functional organ. However, a complete set of genes responsible for initiation and formation of leaf is present in the generated transcriptomes. Retaining these growth-regulating factors occurs due to their multifunctional role in multiple developmental regulatory processes. As no gene loss occurred, reduction of leaves is therefore likely caused by an expression reprogramming event.

Gene loss observed in *E. helleborine* and *N. nidus-avis* transcriptomes proceeded according to previously described models – a massive gene loss was detected mostly in photosynthetic pathways. It can be assumed that most orthologs missing from MH transcriptomes are associated exclusively with photosynthesis, and conserved elements are retained thanks to performing other functions in unrelated pathways. Observed trimming of functions from within complex biosynthesis pathways appears precise and acts under strict selective pressure, affecting only elements of no external significance outside photosynthesis.

Complementation of knocked-out or inactive genes by their nuclear homologues, originating from either convergent evolution or transferring the plastidial sequence into nuclear genome, is often considered a major factor facilitating their loss from plastidial genomes by removal of purifying selective pressure. Such occurrence would be possible to be observed in a conducted transcriptome investigation. However, in this case, we did not

detect any complementation by transcripts of nuclear origin, meaning that loss of these genes was an actual loss-of-function event.

In comparison of differential gene expression between hypogeal and epigeal parts, high convergence between *E. aphyllum* and *N. nidus-avis* was observed, presumably due to functional similarities between the species. Multiple differences affecting metabolic functions may be related to differences in species' interactions with their fungal partners. Compared to AT species, we most notably observed a kind of topographic functionality inversion, as roots and above-ground parts switched their function as a source and recipient of nutrition.

A similar observation was made as previously described in albino MX-turned-MH orchids (Lallemand et al. 2019c) compared to green MX specimens. An increase of trehalose metabolism in carbon-starved achlorophyllous leaves was detected, drawing a parallel to its upregulation in below-ground structures of mycoheterotrophs compared to AT plants. This observation is compliant with a theory that trehalose acts in these species as a fungal carbon carrier. Alongside the similarities, we observed numerous differences in how AT and MH plants employ different transporters for different compounds, due to inverted fluxes of metabolites between organs. These changes indicate that large-scale reprogramming of expression patterns occurred to adjust plants physiology to altered nutrition mode. Describing and explaining these discrepancies is crucial to understanding dynamics of mycoheterotrophy and requires further investigation – both on transcriptomic level and by direct tracking of nutrient flow in labelling experiments. The findings of large expression pattern reprogramming rather than genetic novelty in transition from AT to MH nutrition further attest to the observation of commonness of such events in plant evolution, proving tremendous plasticity of plant nutrition modes.

7.6 Effective *de novo* assembly of plant transcriptomes requires tailored approach and refined tools – and can be improved upon

In transcriptomic experiments concerning model species of plants or those of high agricultural importance, researchers have the beneficial advantage of using high quality genomic references that were sequenced for many species. Their availability makes guided

7. Discussion

approach to transcriptome assembly much easier, delivering final transcriptomes of much higher quality than un-guided *de novo* approach (or even completely removes the need for transcriptome-based approach, by enabling a direct analysis on the reference genome and its gene annotations). Multiple orchid genomes were sequenced up to date, mainly of epiphytic orchids such as *Phalaenopsis* or *Cymbidium*, which are of large economic significance for floristic industry. None of the fully sequenced species are, however, sufficiently closely related to our species of interest in this study. Therefore, we were forced to resort to *de novo* transcriptome assembly. As the sequencing experiments were conducted mostly on Illumina platform, our sequences could be classified as short reads, up to – optimistically – 150bp. Despite their very low error rate, their length poses an obstacle in assembling repeat-rich areas, long homopolymer stretches, genes exhibiting complex isoforms or of low expression (Li et al., 2016). An alternative approach involves sequencing long reads using e.g., MinION (ONT) platform, which generates significantly longer reads (up to 10-15kbp), at the cost of higher error rate (up to 15% misattribution). To address this obstacle, we investigated a possibility of a combined “hybrid” approach, utilizing both sequencing strategies to obtain the best possible *de novo* transcriptomes without the need for (extremely costly and laborious) full genome sequencing (Chapter 10). The hybrid assembly is widely considered a promising novel approach and a wide array of tools is constantly being developed to optimize and streamline the process. We present currently unpublished results of the tests that were conducted on model species *Arabidopsis thaliana*, to evaluate the potential benefits of applying hybrid *de novo* transcriptome assembly to our research.

8 Practices for reliable hybrid *de novo* assembly of plant transcriptomes

Michał May, Étienne Delannoy, Marcin Jąkowski

UNPUBLISHED DATA

Task division:

Michał May:

Bioinformatic processing, data handling, manuscript preparation

Étienne Delannoy:

Experimental concept, providing plant material, RNA sequencing

Marcin Jąkowski:

Bioinformatic procedures design, data processing, manuscript editing

Uniwersytet Gdański
Wydział Biologii
Ul. Wita Stwosza 59
80-308, Gdańsk

data

OŚWIADCZENIE AUTORA O UDZIALE W PUBLIKACJI

Tytuł publikacji: *"Practices for reliable hybrid de novo assembly of plant transcriptomes"*

Imię i nazwisko autora: Michał May

Oświadczam, że w przygotowaniu wymienionego wyżej niepublikowanego artykułu mój udział obejmował:

- Przeprowadzenie procedur bioinformatycznych
- Przetwarzanie danych
- Sporządzenie manuskryptu

.....
podpis autora

Université Paris-Saclay
CNRS, INRAE, Univ Evry
Institute of Plant Sciences Paris-Saclay (IPS2), F-91405
Orsay, France

date

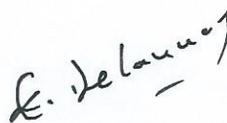
AUTHOR'S CONTRIBUTION STATEMENT

Work title: *"Practices for reliable hybrid de novo assembly of plant transcriptomes"*

Author: Etienne Delannoy

I declare that my contribution in the unpublished article mentioned above included:

- Experimental concept
- Providing plant material
- RNA sequencing



.....
Author's signature

Uniwersytet Gdański
Wydział Biologii
Ul. Wita Stwosza 59
80-308, Gdańsk

data

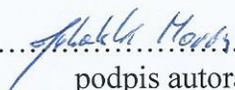
OŚWIADCZENIE AUTORA O UDZIALE W PUBLIKACJI

Tytuł publikacji: *"Practices for reliable hybrid de novo assembly of plant transcriptomes"*

Imię i nazwisko autora: Marcin Jąkałski

Oświadczam, że w przygotowaniu wymienionego wyżej niepublikowanego artykułu mój udział obejmował:

- Zaplanowanie procedur bioinformatycznych
- Przetwarzanie danych
- Redagowanie manuskryptu

.....

.....
podpis autora

Practices for reliable hybrid *de novo* assembly of plant transcriptomes

Michał May¹, Étienne Delannoy^{2,3}, Marcin Jąkowski¹

¹ Faculty of Biology, Department of Plant Taxonomy and Nature Conservation, University of Gdańsk, Poland

² Institute of Plant Sciences Paris-Saclay (IPS2), CNRS, INRA, Université Paris-Sud, Orsay, France

³ Université Evry, Université Paris-Saclay, Orsay, France

Keywords: transcriptome assembly, *de novo* assembly, RNA-seq, long reads

ABSTRACT

Progress in RNA sequencing has allowed researchers to construct high quality transcriptomes without access to genomic data. Those *de novo* methods assemble transcriptomic sequences from RNA-seq Illumina read sets. Due to their low length, however, this process introduces numerous biases, mainly in regions of high repeatability, and does not allow for reliable alternative splice site detection and isoform prediction. A hybrid approach has been developed, that involves employing long reads generated by Oxford Nanopore MinIon or PacBio SMRT sequencing, to improve the quality of short read-based assemblies. A wide

range of software has been developed that performs hybrid assemblies of transcriptomes. Here, we assess and evaluate multiple methodical approaches to develop a set of practices for effective, reliable, and efficient assembly of plant transcriptomes in high-throughput pipelines. As a testing model we use Illumina and Oxford Nanopore MinIon sequencing data from a widely used model plant *Arabidopsis thaliana*.

INTRODUCTION

In the last decade a tremendous progress has been made in the development of next generation sequencing technology (NGS). Researchers were bestowed with accessible, powerful and ever-growing set of tools for thorough analysis of data stored within nucleic acids, both genomic and transcriptomic. Among these, one of the most substantial breakthroughs that would allow for tying genetic and biochemical research together on a whole new level, was the possibility to successfully construct high quality transcriptomes (Lowe et al., 2017; Zhang, 2019). In contrast to stable and unchanging genome, transcriptomes exhibit high plasticity and are dynamically adjusted to the internal and external conditions influencing the researched organism. Their content also varies between developmental stages and nutritional mode of the subject (Kumar et al., 2016). Therefore, a study of transcriptome reflects on the dynamic profile of expressed genes and isoforms, which when coupled with high quality annotation, allow researchers to derive meaningful and well-grounded biological conclusions.

RNA-seq, or sequencing reverse-transcribed RNA, since its first official mention in 2008 (Nagalakshmi et al., 2008) has become a common tool for conducting high-throughput indirect study of transcriptomes, mainly with a goal of identifying differentially expressed genes. Based e.g., on Illumina sequencing, it allows for spatial and temporal quantification of transcripts' expression levels, and thanks to the amplification step it can be conducted virtually on any RNA sample of sufficient purity and integrity. Despite providing high depth of coverage, RNA-seq with Illumina suffers from limitations of the technology and from biases introduced by the PCR amplification step. The reads it provides are of high fidelity, but read length is relatively low – 100-150bp, which is much lower than length of a typical transcript. Therefore, assembly of such data often leads to obtaining incomplete or improperly spliced transcripts, particularly in homologous regions and areas of high repeatability (Byrne et al.,

8. Practices for reliable hybrid de novo assembly of plant transcriptomes

2019). Transcripts of genes exhibiting complex isoform expression and those of low overall expression are particularly difficult to assemble with high degree of confidence. Furthermore, amplification step often introduces a bias against G/C-rich sequences and long homopolymer stretches. This together causes either missing isoforms and lowered completeness of transcriptome produced, or assembly of false transcripts and low quality of annotation.

A different approach to transcriptomic sequence investigation employs a third generation sequencing solution: by applying single molecule sequencing, the analysis can be conducted not only on cDNA, but also directly on mRNA (Branton et al., 2008; Lima et al., 2018). These third generation sequencing technologies, introduced by platforms such as Pacific Biosciences (PacBio) SMRT (Eid et al., 2009; Rhoads and Au, 2015) and Oxford Nanopore Technologies (ONT) MinION (Jain et al., 2016; Laver et al., 2015), excel at producing long reads (LRs) reaching 10-15kbp in length, often covering an entire length of a transcript and greatly improving the overall quality of assembly completeness and isoform identification (Quail et al., 2012; Weirather et al., 2017). A trade-off, however, needs to be made for the price of fidelity – the error rate of long reads reaches up to 15%. Compared to below 1% error rate provided by Illumina sequencing (Pfeiffer et al., 2018), it is a serious shortcoming that creates a requirement for additional correction process for long reads, either self-correction by creating a consensus sequence from multiple reads or correction based on short read alignment (Hackl et al., 2014; Lima et al., 2018; Weirather et al., 2017). Additionally, low throughput and sequencing bias greatly limit usefulness of long reads for isoform abundance estimation.

The *in silico* processing of transcriptomic data presents multiple challenges not encountered in genomic assembly, and for effective application it requires new features, such as alternative splicing detection, multiple isoform recognition and transcript reconstruction. The process of assembling a transcriptome from raw sequencing reads, provided no reference genome is available, can be conducted based on the mutual sequence overlap, or by constructing a de Bruijn graph and resolving the most optimal paths using heuristics or customized criteria (e.g., coverage) (Fu et al., 2015; Grabherr et al., 2011; Pevzner et al., 2001), and converting them into the final contigs. Graph construction involves decomposing the sequence into overlapping strings of set length k , which are commonly referred to as k -mers. The length of k -mers strongly affects the assembler performance and requires fine-tuning to deliver best result for a particular transcriptome. Sequence overlap-based approach generally

8. Practices for reliable hybrid *de novo* assembly of plant transcriptomes

delivers high quality contigs, but conducting complex alignments requires tremendous computational power and time to process large sets of reads. Methods employing De Bruijn graph are much more efficient in terms of calculations and can be conducted within the budget of most research projects (Salmela and Rivals, 2014; Voshall and Moriyama, 2018). However, working on sets of short reads frequently results in a large number of redundant or miss-assembled contigs, especially for highly expressed transcripts. In general, low length of reads used for *de novo* process often results in assembly of shortened or malformed contigs, or missing co-occurring isoforms and alternatively spliced transcripts altogether. The assembly quality can be significantly improved by using a genome-guided approach (Visser et al., 2015; Voshall and Moriyama, 2018). In such case, the reads are assembled by first aligning them to genomic sequences used as a reference scaffold. Resulting assembly is free of ambiguity introduced by decomposing reads into k-mers in de Bruijn graph generation, and provides reliable data on post-translational modifications, splicing and transcription levels of the annotated genes. However, extremely high costs and computational workload required for sequencing a full genome restrict the availability of well-established reference data to common model organisms and species of high economic significance only. Currently, complete genomes have been sequenced for less than 400 plant species (https://plabipd.de/plant_genomes_pa.ep). For all other cases, the research community needs to rely on transcriptomes constructed using a more accessible but less robust *de novo* assembly process (Grabherr et al., 2011; Robertson et al., 2010; Schulz et al., 2012; Weber, 2015).

With the increase of availability and demand for downstream application of third generation sequencing, a requirement has arisen for matching software that could effectively utilize the benefits of both short and long reads at once. Such hybrid approach allows for large improvement in overall performance and accuracy of transcriptomes generated from complementing short reads (SRs) and long reads (LRs). Hybrid correction of long reads is a commonly applied process that involves correcting low accuracy areas of LR data with sequences derived from SRs. A hybrid assembly can be constructed by combining short and long reads into a pre-assembly before performing the proper sequence assembly (Koren et al., 2012). Such construct is usually built by aligning short reads and applying long reads for determination of their proper order and identification of potential missing fragments (Bankevich et al., 2012; Prjibelski et al., 2014). Alternatively, long reads can be used instead of a scaffold (gapped assembly of contigs) for guiding the assembly process of short reads.

8. Practices for reliable hybrid de novo assembly of plant transcriptomes

In the presented research we describe multiple methodical approaches for assembling a high-quality transcriptome from a collection of short reads (employing the Illumina technology) and long reads derived from the Oxford Nanopore MinIon sequencer. We compare and benchmark the data processing procedures and evaluate assemblies obtained through the process. Our main goal is to select a set of tools and practices best suited for performing high-throughput transcriptome assembly for other, non-model plant species.

MATERIALS AND METHODS

Subject and RNA extraction

As a model species, whose transcriptome would be assembled for the evaluation purposes, we have selected the thale cress - *Arabidopsis thaliana*, which is a widely considered and the most commonly used model plant in genetic research. Though its genome is relatively small compared to other plants (approx. 135Mb, while plant genomes range from 80 to 20 000Mb (Michael, 2014)), it has proven to be reliable as a model genome. Its size allows for easier and faster studies, and its thorough description and high quality of mapping and annotation makes *Arabidopsis* a well-suited and highly representative model for our preliminary research.

Most of the published comparisons of transcriptome assembly methods use artificially generated (simulated) read sets. This approach removes a factor of technical biases introduced by manual handling, library construction and sequencing, allowing for the most unbiased and objective comparison of software. However, as our main goal was to optimize workflow for *de novo* transcriptome assembly from *in vivo* samples, we decided to use genuine sequencing reads. *Arabidopsis thaliana* plants were grown *in vitro* at Institut des Sciences des Plantes - Paris-Saclay (IPS2). 7 days old seedlings grown on half strength MS medium containing 1% sucrose were collected and frozen in liquid nitrogen. RNA samples were isolated using NucleoZOL reagent protocol (Macherey-Nagel). Isolates were purified by DNase treatment and Agencourt RNAClean XP magnetic bead rinsing. Ribosomal RNA depletion was conducted using Ribo-Zero rRNA Removal Kit (Illumina).

Library construction and sequencing

Prepared RNA isolates were used for construction of Illumina sequencing libraries. For this goal, TRUseq mRNA stranded kit (Illumina) was used. Sample and library quality was assessed at each step by Agilent 2100 BioAnalyzer. Short reads were produced by Illumina HiSeq 2500 platform.

The libraries for long read sequencing were constructed from 50ng of total RNA using the SQK-PCS109 kit (Oxford Nanopore Technologies) and multiplexed (2 libraries per FC, 75fmols each) on a MinIon R9.4 FC during 48h. the MinIon was controlled by MinKNOW software (v. 19.06.8.)

Read pre-processing

The basecalling was conducted with Guppy (v. 3.0.7+427b5a11). The config file used was *dna_r9.4.1_450bps_fast.cfg* and the model file - *template_r9.4.1_450bps_fast.jsn*.

The identification of the sense and anti-sense reads was done by identification of the VSN (3') and SSP (5') adapters with LAST.

The demultiplexing and trimming of adapters was done with Guppy barcoding software (v. 3.2.2+9fe0a78), and arrangement file *barcode_arrs_lwb.cfg*.

Hardware setup

All *in silico* processing was performed on a DELL PowerEdge R430 computational server of the Laboratory of Plant Symbiosis, University of Gdańsk with the following technical details: 2x 8 core (16 threads) Intel Xeon E5-2620V4 processor and 256 GB of RAM. Peak memory usage and computation times were measured individually for each software used. All software used in this research was given the same allocation of CPU (24 threads) and RAM (120GB).

Raw read error correction

Before incorporating the long reads into our analyses, raw long reads underwent pre-processing, namely sequence correction and read set reduction, with the latter aiming at reducing the computational time of the analyses, however without sacrificing the quality of the expected results. We subjected the forward and reverse read sets to correction protocol using an arbitrarily chosen hybrid error correction program – **LoRDEC** (Long Read DBG Error Correction, version v0.9 using GATB v1.4.1) (<http://www.atgc-montpellier.fr/lordec>)(Salmela and Rivals, 2014). SRs are not aligned directly to the long sequences but are instead used to construct a de Bruijn graph. Replacements for faulty regions of long reads are generated by resolving appropriate paths of the graph. LoRDEC performs two passes over each long read – during the first pass, new robust k-mers are generated. They are subsequently used as starting nodes in the second pass, which is performed in opposite direction.

The LRs were passed through a 4-step iterative correction process using Illumina SRs as a high fidelity reference set. We used the lordec-correct program with increasing k-mer length: 19, 21 and 23, with an additional final pass with k=19, as advised in the manual. Forward and reverse reads were processed separately and merged at the end of the correction phase. Reads were trimmed (lordec-trim command) to remove terminal parts of sequences that could not be corrected. Subsequently, we decided to discard all the sequences below 200bp in length as not suitable for assembly process. Redundant sequences were clustered using CD-HIT-EST (<http://weizhongli-lab.org/cd-hit/>)(Li and Godzik, 2006) based on sequence similarity (identity threshold = 0.97) and inclusion. At each step the reads were mapped to the reference *Arabidopsis* genome using deSALT (as specified in commands list) - a long-read alignment software for transcriptomic applications (<https://github.com/ydLiu-HIT/deSALT>)(Liu et al., 2019), to detect potential loss of meaningful reads and ensure that each step actually increases sequence quality, and to investigate the differences in performance and output quality of *de novo* and hybrid assembly software. The reference genome was obtained from The Arabidopsis Information Resource - TAIR10 (<https://www.arabidopsis.org/download/index.jsp>)(Lamesch et al., 2012).

Transcriptome assembly

Our primary choice for a hybrid assembler was **IDP-denovo** (Version v.2, by Au's lab, 2016/12/08) (<http://www.healthcare.uiowa.edu/labs/au/IDP-denovo>), a tool developed by Fu et al., (2018) with an exact purpose of substituting the benefits of aligning to reference genome with coverage of long reads and fidelity of short reads combined together. Preceding the actual assembly process, a scaffold based on short reads assembly is constructed (hereafter pre-assembly). Long reads are then aligned to the pre-assembly and used to extend and refine the delivered SR scaffolds; the LRs that could not be aligned are clustered by k-mers and later used to generate consensus structures. A pseudo-reference of the exonic regions is generated based on multiple sequence alignment and is used to determine alternative usage of splice sites and exons, and thus annotate corresponding isoform structures of each transcript. Finally, SRs are used to estimate abundance of isoforms.

rnaSPAdes (v3.14.0) (<http://cab.spbu.ru/software/rnaspades/>)(Bushmanova et al., 2019) is a *de novo* transcriptome assembler built upon SPAdes (Bankevich et al., 2012) genome assembly tool with mainly Illumina read processing in mind, however it also supports Ion Torrent reads and long reads generated by PacBio and Oxford Nanopore. SPAdes applies an alternative approach using k-bimers only for initial De Bruijn graph construction. All further graph-theoretical operations are based on graph topology, coverage, and sequence lengths, omitting the sequences themselves (A-Bruijn graph), until restoring the consensus sequence at the final processing step. As a modification for transcriptome data, tip-trimming procedures were modified to be less stringent and preserve transcript ends that would be otherwise lost. Graph simplification procedure has been significantly altered to avoid creation of chimeric structures, non-informative sequences and mis-assemblies. Complex algorithms were also introduced to improve isoform reconstruction.

We have performed 6 separate IDP-denovo hybrid assemblies based on 5 different pre-assemblies selected arbitrarily from 24 short read assemblies described in the following paragraph, using corrected long reads and the Illumina reads. Assembly settings and commands used are specified in commands list. In addition, we performed hybrid assemblies using Trinity, rnaSPAdes and Velvet-Oases pipeline (as specified in commands list).

Pre-assembly construction

For pre-assembly construction we have selected a range of *de novo* assemblers developed for processing the short transcriptomic reads, namely Trinity, TransABYSS, SOAPdenovo-Trans and Velvet-Oases.

Trinity (v. 2.6.6) (<https://github.com/trinityrnaseq/trinityrnaseq/wiki>) (Grabherr et al., 2013)) suite assembly is conducted in a three-step process. First module, *Inchworm*, builds a k-mer dictionary and constructs graphs from its words. A greedy algorithm then searches for paths in the graph, so that k-mers are present only once in a single path and converts them to linear sequences. Contigs containing common k-mers are pooled together by *Chrysalis* and are used to build individual de Bruijn graphs for each pool. Graphs are then cleaned and simplified by *Butterfly*. Finally, plausible paths are derived from the graphs according to their consistency with reads they were based upon and are delivered as final linear sequences. A feature allowing for use of long reads for isoform resolving can also be applied, making it a hybrid assembly. However, as Trinity needs to perform a BLAST alignment during the assembly, processing large sets of long reads is exponentially more time consuming. To avoid very long processing times, that would preclude its use in a high-throughput pipeline, a reduction of LR set is necessary (here conducted as a part of LR correction step). It is expected that future versions of Trinity may introduce faster alternatives to BLAST alignment.

We have performed Trinity assembly twice: once as a pre-assembly using only Illumina reads, and once as a hybrid assembly including Nanopore long sequences. Used commands and parameters are specified in commands list.

TransABYSS (v. 2.0.1) (<https://www.bcgsc.ca/resources/software/trans-abyss>) (Robertson et al., 2010), a modification of ABySS assembler optimized for processing RNA-seq data, is based on de Bruijn graph construction and refining. In addition to removing dead-end branches and merging unambiguously connected nodes, during graph simplification TransABYSS detects specific bubbles of overlapping contigs, isolates them and adds them to the final assembly to reconstruct alternative splicing variants of sequences. The workflow usually includes generating multiple assemblies with different k-mer length. This approach concludes in merging highly sensitive assemblies based on short k-mers and highly specific based on long ones, into a meta-assembly. Such construct exhibits significant improvement

8. Practices for reliable hybrid de novo assembly of plant transcriptomes

in both specificity, detecting less misassembled contigs, and sensitivity, resulting in higher number of assembled contigs than both short and long single k-mer assembly.

We have performed a set of 9 assemblies of Illumina reads with k-mers from 21 to 91, rising in increments of 10, with addition of $k = 25$, and merged them into a single meta-assembly using TransABySS-merge. All commands and parameters used in performing this assembly are specified in commands list.

SOAPdenovo-Trans (v1.04) (<https://github.com/aquaskyline/SOAPdenovo-Trans>) (Xie et al., 2014) is a transcriptome-oriented modification of SOAPdenovo (Li 2009), a de Bruijn graph based assembler for Illumina reads. It implements a global and local correction of k-mers before contig generation and scaffold graph construction. Contigs are generated from linear, fork and bubble paths. During graph simplification short contigs below 100bp are removed from the final sequences, reducing structure ambiguity but introducing gaps in sequences. Therefore, final contigs undergo gap filling correction based on paired-end information.

We used SOAPdenovo-Trans to perform two assemblies of Illumina reads at $k = 25$ and $k = 31$ (as specified in commands list).

Velvet (v. 1.2.10) (<https://www.ebi.ac.uk/~zerbino/velvet/>) (Zerbino and Birney, 2008) was originally designed to build genomic contigs and scaffolds from SRs. It is used in a two-step process - hashing (Velveth) and graph building (Velvetg). Hashing creates a dictionary of all k-mers, which are then used by Velvetg to build a de Bruijn graph, which is later simplified and reduced to its final form. Multiple assemblies of different k values are then merged into a meta-assembly using **Oases** (v. 0.2.09) (<https://www.ebi.ac.uk/~zerbino/oases/>) (Schulz et al., 2012), a transcriptomic assembly tool, to refine the assembly by including both highly specific long k-mer graphs and more sensitive low k-values.

We have tested and analyzed assemblies performed at 9 different k-mer levels between 21 and 91; those were also used for creating meta-assemblies at $k = 27$ and 61. Additionally, a feature of adding long reads to the Oases assembly was utilized to build a hybrid assembly with $k = 71$ (as suggested by Velvet Advisor), and a hybrid meta-assembly at $k = 27$ with LR input (as specified in commands list).

8. Practices for reliable hybrid de novo assembly of plant transcriptomes

In total, 24 pre-assemblies were produced and 5 were chosen for further application in IDPdenovo. Pre-assemblies were chosen from each applied de-novo assembler to cover a wide range of diverse software: Trinity, Velvet-Oases, TransABySS and SOAPdenovo-Trans. For Velvet-Oases and TransABySS we have selected the meta-assemblies that assembled the most sequences and provided representation for a wide range of k-mers. Both contigs and predicted transcript sequences from Velvet-Oases were tested as pre-assembly sets. For SOAPdenovo-Trans we have selected assembly at $k = 31$ which contains less sequences than $k = 25$ but has performed better in terms of assembled contig length and N50 and scored a higher percentage of mapped bases in reads.

Transcriptome evaluation

To compare and evaluate the several implemented assembly methodologies a set of statistics and analyses was performed. For extracting basic quantitative statistics, such as number of contigs in alignment, smallest and largest sequence size, N50, average and median length, we used **TrinityStats** (a utility of Trinity suite) and **SeqStat** (v. SQUID 1.9g) (a part of SQUID library and HMMer package (<http://hmmmer.org/>)).

To assess the quality of built transcriptomes, the assemblies were mapped to a reference *Arabidopsis* genome. **GMAP** (v. 2019-09-12) (<http://research-pub.gene.com/gmap/>) (Wu and Watanabe, 2005), a splice-aware genomic mapping program for mRNA sequences was used for this operation, as described in commands list. Subsequently, mapping statistics were calculated. For this purpose, we employed **SeqStat**, **RNAseqEval** (<https://github.com/kkri-zanovic/RNAseqEval>) (Križanović et al., 2018), **featureCounts** (v.1.6.3) (a part of Subread package (<http://subread.sourceforge.net/>)) and **Bedtools** (v2.25.0) (<https://bedtools.readthedocs.io/en/latest/>) utilities (parameters specified in commands list). Additional statistics were performed using R with in-house developed scripts.

RESULTS

Long read error correction

The raw set of ONT sequencing data consisted of 6 638 367 sequences (Table 8.1). Progressing through increasing k-mer iterations resulted in only minor, insignificant increase in average sequence length (Table 8.1). After performing a sequence trimming, 15 848 sequences were lost, and average sequence length decreased by 14%. A number of sequences that could be mapped to the genome also dropped by 2%. Of 33610 genes in TAIR10 genome, 65.36% were covered after read correction (22 313), and trimming caused a loss of 345 (1%) of these mappings.

Discarding very short sequences below 200bp cutoff threshold reduced the total number of sequences by approximately 10% (from 6 638 367 to 5 909 322) (Table 8.1). It consequently resulted in significant increase in average sequence length (over 10%) and over 4.5% increase in mapped sequence percentage (by decreasing a number of non-mapping sequences) (Table 8.2). However, further 677 genes from the reference genome (2%) were lost. Compared to TAIR10 cDNA, it is over twice the number of contigs below 200bp observed in the reference data.

Clustering redundant sequences further reduced the set to 1 689 883 unique reads (25% of original read set size), while reducing genome coverage only by 0.5%, and decreasing percentage of mappable sequences by 0.06% (Table 8.1).

Overall, during post-correction processing, 2,15% of representation of *Arabidopsis* genome was lost, while reducing the set size by approximately 75% (Table 8.1 and Table 8.2). Such reduction will significantly simplify and speed up further downstream processing, with affordable loss of representation of genome coding part.

Pre-assembly statistics

Optimally, the most desired assembly should contain the largest summed lengths of contigs and avoid over-assembly of reads and constructing of high similarity overlapping contigs from duplicated data containing high degree of sequencing errors. The obtained mean contig

8. Practices for reliable hybrid de novo assembly of plant transcriptomes

length and length variance should be close to a value obtained for reference transcriptome (maximum contig length - 16593bp, average - 1799bp, median - 1602bp).

Trinity pre-assembly consists of 61773 contigs between 201bp and 15571bp in length - 29% more than are present in reference cDNA. It exhibited the highest mean, median and N50 (a weighted median length of sequence in bp, above which, if summed, the sequences cover at least half of the entire assembly length) contig lengths of all constructed pre-assemblies (Table 8.4). The largest assembled read length equals 94% of the longest reference cDNA contig length.

TransABySS generated 85782 contigs at its lowest k-mer setting ($k = 21$), which is the highest number for its single-k assemblies (179% of reference cDNA sequences) and 28383 at the highest ($k = 91$) (59% of reference number) (Table 8.4). All k-mer settings between 25 and 61 resulted in assembly of the longest sequence of 15563bp (94% of longest reference contig). Extreme k-mer values delivered shorter maximum length, with the lowest result at $k = 91$ (9053bp, 55% of longest reference contig). Average and mean length and N50 peaked at $k = 81$. The merged meta-assembly generated the highest number of contigs (105062, over 220% of reference set) of all TransABySS assemblies, and also managed to assemble the longest sequence of 15563bp (Table 8.4). It delivered the highest average length and only slightly below-the-top median and N50 values.

SOAPdenovo-Trans in general delivered less assembled contigs over 200bp in length than the other de novo assemblers tested in this study, being, however, the closest to the number of contigs present in reference cDNA (Table 8.4). As expected, $k = 31$ delivered slightly less sequences than $k = 25$ but exhibited higher average sequence lengths and generated a longer largest contig (94% vs. 84% of largest reference contig).

The number of valid sequences (over 200bp) delivered by simple **Velvet-Oases** assemblies peaked at $k = 41$ with 78337 contigs, while the best average, median and N50 values were obtained at $k = 81$ (Table 8.4). Maximum contig length peaked at $k = 71$ with 13090bp, while the assemblies below $k = 61$ did not manage to assemble any contigs reaching 50% of longest reference contig and delivered the lowest average contig length scores. Meta-assemblies constructed from 9 single k-mer assemblies merged at $k = 27$ and $k = 61$ performed quite closely to each other, delivering over 4 times more valid sequences than the highest number reached by single k-mer assembly (respectively 662% and 655% of reference set).

8. Practices for reliable hybrid de novo assembly of plant transcriptomes

Both managed to assemble the longest contig of 13090bp (79% of longest reference contig) (Table 8.4). Shorter of the two tested k-mers in meta-assemblies performed slightly better in terms of average, median and N50 lengths.

In general, Trinity SR *de novo* assembly performed the best in terms of contig length structure, and the poorest performance in this category was exhibited by Velvet assemblies. Numbers of sequences assembled in multi-k assemblies exceeded single-k usually by an entire order of magnitude.

We have selected a set of 5 pre-assemblies to be applied in further IDP-denovo processing. Our choice was based on selecting the richest and most promising assembly from diverse programs. We have chosen Trinity SR assembly, Velvet meta-assembly performed at $k = 27$ (both transcript and contig sequences were tested), SOAPdenovo contig sequences derived from $k = 31$ assembly and Trans-ABYSS meta-assembly.

Assembly statistics

The results delivered by IDPdenovo assemblies based on different pre-assemblies exhibited highly variable results. The highest abundance of contigs (1 242 992) was generated based on Velvet contigs meta-assembly and was almost twice as high as the second-to-highest value achieved by single-k Velvet assembly at $k = 91$ and exceeded the reference cDNA set 25 times (Table 8.3, Fig. 8.2). However, removal of contigs shorter than 200bp resulted in loss of only 40 (0,005%) of its sequences, while the meta-assembly lost almost 275 000 sequences (over 18%). Assembly based on transcripts predicted in the meta-assembly was approximately half as big and its longest sequence was almost twice as long, despite exhibiting lower average and median length. A reference TAIR10 cDNA base contains 48 037 contigs above 200bp, which is exceeded at least 4 times and over 25 times at most by IDPdenovo assemblies. SOAPdenovo and transABYSS-based IDP assemblies performed the best in maximum read length - their longest sequence was 94% of the longest reference cDNA contig and exhibited the highest average sequence lengths. Utilizing Trinity pre-assembly resulted in the lowest number of assembled contigs, the shortest maximum contig length and lowest average and median length. In general, IDPdenovo output sets were much

more numerous than their corresponding pre-assemblies but had overall worse contig length statistics (Fig. 8.2).

Velvet-Oases hybrid meta-assembly constructed 63549 contigs and predicted 248245 transcripts (Table 8.3). The longest assembled transcript was 15676bp long and surpassed all the other sequences obtained in this paper in length, reaching 94% of the highest contig length in reference cDNA. To reduce number of transcripts and tighten the set to only the most biologically useful sequences, we have performed a "filtration" by EvidentialGene tr2aacds, reducing the number of transcripts to 27515 and greatly improving statistics of sequence lengths. Overall, this set exhibited the most promising basic statistical parameters of all hybrid assemblies built in this paper.

rnaSPAdes assembly delivered the lowest number of sequences of all hybrid assemblies described in this work, being, however, the closest to the reference cDNA in terms of both sequence and residue number. It has scored the best of all tested hybrid assemblies in terms of average and median contig length. Its longest assembled contig is 15571bp long (94% of longest reference contig) (Table 8.3).

Final assembly mapping

A hallmark of a high quality transcriptome is its ability to represent and match the actual - or, in this case, for the sake of evaluative approach, a reference genomic data set previously established for the target species. Therefore, mapping the assembled sequences to the genome can be used as an assessment measure.

Final assemblies were mapped to the TAIR10 reference genome as a mean of quality assessment. As the most meaningful alignment parameters for creating a benchmark, we consider: 1) fraction of bases of assembled sequences that were mapped to the genome (a measure of abundance and correctness), 2) fraction of annotated genes in the reference genome with a hit from transcriptome assembly (measure of completeness) and 3) with 100% coverage, and 4) percent of non-zero mappings (measure of uniqueness). In addition, we compare mean breadth of coverage (percentage of genome covered at certain depth) and depth of coverage (average number of times a base of a reference genome is covered by aligned assembly data).

8. Practices for reliable hybrid de novo assembly of plant transcriptomes

1) The highest fraction of assembly bases (99.48%) was mapped to the reference at a tr2aacds processed Velvet-Oases LR contigs meta-assembly, followed by rnaSPAdes hybrid assembly at 99.19%, rnaSPAdes hybrid assembly with parameter ss-rf (98.98%), followed by *ex aequo* non-filtered Velvet-Oases LR contigs meta-assembly set, and Trinity LR assembly, both at 98.88% (Table 8.5, Fig. 8.1). In contrast, the worst result (expressing high content of abundant or misassembled structures in a transcriptome) was achieved by IDP-denovo assembly based on Velvet meta-assembly contigs set and IDP-denovo assembly running on Trinity pre-assembly (89%).

2) IDP-denovo assembly based on Velvet meta-assembly exhibited partial hits to 99.998% structures present in the genome, which together with low fraction of assembly bases that were mapped to the reference indicates generation of large number of sequences of low specificity. Trinity hybrid assembly scored low in this gauge (83%), and so did rnaSPAdes (80%) (Table 8.5, Fig. 8.1). As both have scored high in the previous criterion, we assume that many sequences of reference remained unaddressed, due to likely being lost during assembly.

3) The highest percentage of genes covered in 100% of their length is present in IDP-denovo hybrid assembly based on velvet meta-assembly contig set (99.98%) (Table 8.5, Fig. 8.1). On average, IDPdenovo assemblies scored the highest in this scope, followed by the set of transcripts predicted by a hybrid velvet-oases assembly. SPAdes, Trinity and rnaSPAdes hybrid assemblies scored much lower, completely covering respectively 67, 58 and 46%.

4) Average percentage of non-zero mappings equals approximately 60%. IDP-denovo assembly based on Velvet meta-assembly exhibited the highest percentage of non-zero mappings (71%), while Velvet-Oases LR contigs and transcripts meta-assembly scored below 50% after processing with tr2aacds (Table 8.5, Fig. 8.1). IDP assembly based on Trinity short read pre-assembly delivered only 6.41% of non-zero mappings.

Timing and memory consumption

For each assembly process, timing and memory consumption data were recorded. The most optimal assembler should exhibit possibly short computation time - this factor is especially important in designing a high-throughput pipeline, as time differences may be insignificant for a single assembly but will accumulate for larger numbers of read sets. Efficiency in memory utilization is especially important while working on low-end computational servers and workstations.

In case of Velvet-Oases pipeline, both Velveth indexing and Velvetg graph processing time were directly related to processed k-mer length, being the slowest for low k values and the fastest for high k. The differences in time were larger for Velvetg process. Oases meta-assembly creation process consisted of Velveth, Velvetg and Oases-merge processes. However, as the data were already processed prior to merging step, the pipeline: hashing, graph processing and merge process took 6 minutes, of which Oases merging time was below 1 minute.

An entire process of creation of Velvet preassembly with single k = 91 took 53 minutes (Table 8.6). In comparison, creation of meta-pre-assembly consumed almost 19 hours. Adding long reads to the assembly significantly increased the calculation time for both Velveth and Velvetg, reaching over 72h for the entire meta-assembly pipeline at k = 27. The almost 4x increase in time does not reflect on the number of assembled sequences - we have observed a 1,5x decrease instead, but the improvement in contig length is significant.

We have performed all assembly processes using both FASTA and FASTQ.gz sequence input files, expecting to observe that processing compressed files would be more time consuming. However, we have noticed no discernible difference in processing time between compressed and uncompressed input. It is therefore beneficial for the sake of disk space to use compressed sequence files, as they require approximately 1/2 of space taken by uncompressed FASTA.

IDP-denovo assemblies exhibited high variability in processing time depending on used pre-assembly file, ranging from e.g., below 1h when working on pre-assembly constructed by Trinity, to over 20h on Velvet-Oases meta-assembly. There is no direct correlation between processing time and number of sequences in pre-assembly or number of contigs assembled

8. Practices for reliable hybrid de novo assembly of plant transcriptomes

by IDPdenovo. Assemblies based on SOAPdenovo, TransABySS and Velvet meta-assembly transcripts set deviate from output-to-time scaling which seems to be linear for assemblies based on Trinity, Velvet $k = 91$ and Velvet meta-assembly contig set. No distinct tendencies were observed in input-to-time scaling.

Processing time of SPAdes and rnaSPAdes was mostly below 30 minutes (Table 8.6), making it the fastest of all tested hybrid assemblers. However, it reflects on its number of assembled sequences.

We have noticed discrepancies between run time of several processes, which had to be repeated several times. For instance, Velvet calculations were conducted twice for $k = 25$ and was completed in respectively 133min and 32min. A possible reason of such a variance is disk access speed, which was not measured during the calculations, and might have been affected by other processes running simultaneously, being calculations or, potentially, RAID matrix maintenance task.

Most of tested assemblers exceeded the amount of memory available on a common workstation (which, for 2020, is assumed at 16GB) - only TransABySS and IDPdenovo did not exceed this amount. SPAdes and rnaSPAdes required 17-20GB RAM, which is available on high-end workstations. The most demanding assembler turned out to be Trinity, peaking at over 67GB RAM used. It was followed by Velvet, which required almost 30GB RAM for de novo and almost 40GB for hybrid assembly (Table 8.6). None of the tested applications came close to our upper memory usage limit set at 120GB.

We did not record complete data required to perform an analysis of effect of number of CPU threads on processing time. All of the assemblers we have tested support multi-threaded calculations. Based on our observations, however, the main bottleneck in computation time was related to processes that could not be divided between multiple threads. Also, calculation time exhibited significant improvements with expanding thread count only up to a certain number of cores used, being the most notable below 8 threads.

DISCUSSION

The research problem of transcriptome construction can be approached using multiple distinct strategies. Decision to use a hybrid assembly process requires to be taken early in the research preparation, as it will require different methodical approach to sample processing and sequencing than relying only on SR data. A choice between various assembly tools can be made later, however, it is still not an easy one. Despite evaluation of transcriptome assemblers having been done numerous times in the past, still no clear conclusion on this subject has been reached and no superior method has been pointed out at the time of this study. The comparisons mostly use synthetic, artificially generated data, and mostly focus on a single tool and a single step within a transcriptome construction workflow. Moreover, what makes such comparison harder to perform and derive conclusions, is lack of unified set of metrics to estimate the quality of a sequence set.

In our work, we have used a set of diverse metrics for transcriptome evaluation, both statistical and reference-based, to come closer to a clear conclusion. The statistics were performed at every point of the whole research process, from sequencing to the final result. Our analysis is also supposed to shed some light on practical implications of used measures to the entirety of the process and aid in creation of improved protocols in the future.

In our approach, we were able to use a superior reference-based quality assessment. However, such solution is not available for most non-model species work, what indicates a large need for non-reference-based evaluation tools based purely on statistics or on a common reference for groups of species.

Although all of the generated assemblies were based on the same sequencing data, every approach generated an individual and different set of results. The simple single-k *de novo* assemblies exhibited less variability in results, but the ensemble approaches exhibited a raise in ambiguity. In theory, the ensemble approaches are supposed to combine the benefits of multiple assembly tools. Our analyses show, that while delivering generally better results than simple single-step assembly, the drawbacks of these tools also do pile up, and may introduce severe errors into the process and result in decreased final assembly quality, and affect further downstream applications, e.g., differential gene expression analysis or annotation.

8. Practices for reliable hybrid de novo assembly of plant transcriptomes

Our assessment, even though performed on a widely accepted model organism, might not be relevant to the intended research application, due to differences in genome structure and size between our model and the non-model species. Even within the species, fine tuning of assembly parameters, such as the k-mer length, resulted in obtaining greatly different results. Our research suggests a general direction to follow while deciding on a tool-set and approach but does not deliver an unambiguous ready solution. We expect that the pipelines we have established will require fine tuning to adjust them to their final subjects.

The conclusion we have reached does not vote strongly in favor of a particular method, but nevertheless it does present the hybrid assembly approach in a less "glorified" manner. Despite being advertised as a solution for all the problems sprouting from the limitations of SR and LR sequencing, it still requires a lot of improvement to effectively solve these issues. Until then, it is understandable that using long reads as a mean of assembly improvement will often not be considered in research projects, in order to divert the funding and resources to more proven SR-based solutions, or to simply save on computation time.

CONCLUSIONS

Our descriptions of various hybrid transcriptome assembly tools allowed us to create a comparison. IDPdenovo generated large sets of contigs but delivered mediocre average contig length stats. The result depended strongly on quality of input pre-assembly. Velvet-Oases generated small contig sets and low contig length statistics, except for its multi-k meta-assemblies, which were, however, very slow to build. rnaSPAdes turned out to be very fast while delivering small sets of contigs of satisfying length. Trinity delivered contig sets of decent length, but not very numerous, while the processing time was slow.

Generally, hybrid assemblies performed better in terms of genome fit and recovery. However, they required much longer computation times.

Multi-k meta-assemblies were a decent improvement of single-k sets, delivering much larger sets, better contig length statistics and scoring better in genome mapping assessment. However, their construction is very time-consuming.

ABBREVIATIONS

LR: long reads (Nanopore)

SR: short reads (Illumina)

ONT: Oxford Nanopore Technologies

Mb: mega basepairs

kb: kilo basepairs

GB: gigabytes

DATA ACCESSIBILITY / SUPPLEMENTARY MATERIALS

All figures, tables and full list of commands used in presented research are attached in supplementary material chapter of this work. For clarity, figures and tables belonging to this chapter are labelled under double digit reference numbers, first digit marking the number of this chapter (Table 8.X, Fig. 8.X). The References section is independent of the main body of this dissertation and is not included in the final citations section.

REFERENCES

- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Branton, D., Deamer, D.W., Marziali, A., Bayley, H., Benner, S.A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., Jovanovich, S.B., Krstic, P.S., Lindsay, S., Ling, X.S., Mastrangelo, C.H., Meller, A., Oliver, J.S., Pershin, Y. V., Ramsey, J.M., Riehn, R., Soni, G. V., Tabard-Cossa, V., Wanunu, M., Wiggin, M., Schloss, J.A., 2008. The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* 26, 1146–1153. <https://doi.org/10.1038/nbt.1495>
- Bushmanova, E., Antipov, D., Lapidus, A., Prjibelski, A.D., 2019. RnaSPAdes: A de novo transcriptome assembler and its application to RNA-Seq data. *Gigascience* 8, 1–13. <https://doi.org/10.1093/gigascience/giz100>
- Byrne, A., Cole, C., Volden, R., Vollmers, C., 2019. Realizing the potential of full-length transcriptome sequencing. *Philos. Trans. R. Soc. B Biol. Sci.* 374. <https://doi.org/10.1098/rstb.2019.0097>
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., DeWinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J., Turner, S., 2009. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* (80-.). 323, 133–138. <https://doi.org/10.1126/science.1162986>
- Fu, S., Ma, Y., Yao, H., Xu, Z., Chen, S., Song, J., Au, K.F., 2018. IDP-denovo: De novo transcriptome assembly and isoform annotation by hybrid sequencing. *Bioinformatics* 34, 2168–2176. <https://doi.org/10.1093/bioinformatics/bty098>
- Fu, S., Tarone, A.M., Sze, S.H., 2015. Heuristic pairwise alignment of de Bruijn graphs to facilitate simultaneous transcript discovery in related organisms from RNA-Seq data. *BMC Genomics* 16, S5. <https://doi.org/10.1186/1471-2164-16-S11-S5>
- Grabherr, M.G., Brian J. Haas, Moran Yassour Joshua Z. Levin, Dawn A. Thompson, Ido Amit, Xian Adiconis, Lin Fan, Raktima Raychowdhury, Qiandong Zeng, Zehua Chen, Evan Mauceli, Nir Hacohen, Andreas Gnirke, Nicholas Rhind, Federica di Palma, Bruce W., N., Friedman, and A.R., 2013. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* 29, 644–652. <https://doi.org/10.1038/nbt.1883>
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. <https://doi.org/10.1038/nbt.1883>
- Hackl, T., Hedrich, R., Schultz, J., Förster, F., 2014. Proovread: Large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics* 30, 3004–3011. <https://doi.org/10.1093/bioinformatics/btu392>
- Jain, M., Olsen, H.E., Paten, B., Akeson, M., 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* 17, 1–11. <https://doi.org/10.1186/s13059-016-1103-0>

8. Practices for reliable hybrid de novo assembly of plant transcriptomes

- Koren, S., Schatz, M.C., Walenz, B.P., Martin, J., Howard, J.T., Ganapathy, G., Wang, Z., Rasko, D.A., McCombie, W.R., Jarvis, E.D., Phillippy, A.M., 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat. Biotechnol.* 30, 693–700. <https://doi.org/10.1038/nbt.2280>
- Križanović, K., Echchiki, A., Roux, J., Šikić, M., 2018. Evaluation of tools for long read RNA-seq splice-aware alignment. *Bioinformatics* 34, 748–754. <https://doi.org/10.1093/bioinformatics/btx668>
- Kumar, D., Bansal, G., Narang, A., Basak, T., Abbas, T., Dash, D., 2016. Integrating transcriptome and proteome profiling: Strategies and applications. *Proteomics* 16, 2533–2544. <https://doi.org/10.1002/pmic.201600140>
- Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., Karthikeyan, A.S., Lee, C.H., Nelson, W.D., Ploetz, L., Singh, S., Wensel, A., Huala, E., 2012. The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Res.* 40, 1202–1210. <https://doi.org/10.1093/nar/gkr1090>
- Laver, T., Harrison, J., O'Neill, P.A., Moore, K., Farbos, A., Paszkiewicz, K., Studholme, D.J., 2015. Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol. Detect. Quantif.* 3, 1–8. <https://doi.org/10.1016/j.bdq.2015.02.001>
- Li, W., Godzik, A., 2006. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>
- Lima, L., Marchet, C., Caboche, S., Da Silva, C., Istace, B., Aury, J.-M., Touzet, H., Chikhi, R., 2018. Comparative assessment of long-read error-correction software applied to RNA-sequencing data. *Brief. Bioinform.* 1–14. <https://doi.org/10.1101/476622>
- Liu, B., Liu, Y., Li, J., Guo, H., Zang, T., Wang, Y., 2019. DeSALT: Fast and accurate long transcriptomic read alignment with De Bruijn graph-based index. *Genome Biol.* 20, 1–14. <https://doi.org/10.1186/s13059-019-1895-9>
- Lowe, R., Shirley, N., Bleackley, M., Dolan, S., Shafee, T., 2017. Transcriptomics technologies. *PLOS Comput. Biol.* 13, e1005457. <https://doi.org/10.1371/journal.pcbi.1005457.t005>
- Michael, T.P., 2014. Plant genome size variation: Bloating and purging DNA. *Briefings Funct. Genomics Proteomics* 13, 308–317. <https://doi.org/10.1093/bfpg/elu005>
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., Snyder, M., 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* (80-.). 320, 1344–1349. <https://doi.org/10.1126/science.1158441>
- Pevzner, P.A., Tang, H., Waterman, M.S., 2001. An Eulerian path approach to DNA fragment assembly. *Proc. Natl. Acad. Sci. U. S. A.* 98, 9748–9753. <https://doi.org/10.1073/pnas.171285098>
- Pfeiffer, F., Gröber, C., Blank, M., Händler, K., Beyer, M., Schultze, J.L., Mayer, G., 2018. Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Sci. Rep.* 8, 1–14. <https://doi.org/10.1038/s41598-018-29325-6>
- Prjibelski, A.D., Vasilinetc, I., Bankevich, A., Gurevich, A., Krivosheeva, T., Nurk, S., Pham, S., Korobeynikov, A., Lapidus, A., Pevzner, P.A., 2014. ExSPAnDer: A universal repeat resolver for DNA fragment assembly. *Bioinformatics* 30, 293–301. <https://doi.org/10.1093/bioinformatics/btu266>
- Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., Gu, Y., 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13, 1. <https://doi.org/10.1186/1471-2164-13-341>

8. Practices for reliable hybrid de novo assembly of plant transcriptomes

- Rhoads, A., Au, K.F., 2015. PacBio Sequencing and Its Applications. *Genomics, Proteomics Bioinforma.* 13, 278–289. <https://doi.org/10.1016/j.gpb.2015.08.002>
- Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M., Qian, J.Q., Griffith, M., Raymond, A., Thiessen, N., Cezard, T., Butterfield, Y.S., Newsome, R., Chan, S.K., She, R., Varhol, R., Kamoh, B., Prabhu, A.L., Tam, A., Zhao, Y., Moore, R.A., Hirst, M., Marra, M.A., Jones, S.J.M., Hoodless, P.A., Birol, I., 2010. De novo assembly and analysis of RNA-seq data. *Nat. Methods* 7, 909–912. <https://doi.org/10.1038/nmeth.1517>
- Salmela, L., Rivals, E., 2014. LoRDEC: Accurate and efficient long read error correction. *Bioinformatics* 30, 3506–3514. <https://doi.org/10.1093/bioinformatics/btu538>
- Schulz, M.H., Zerbino, D.R., Vingron, M., Birney, E., 2012. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086–1092. <https://doi.org/10.1093/bioinformatics/bts094>
- Visser, E.A., Wegrzyn, J.L., Steenkmap, E.T., Myburg, A.A., Naidoo, S., 2015. Combined de novo and genome guided assembly and annotation of the *Pinus patula* juvenile shoot transcriptome. *BMC Genomics* 16, 1–13. <https://doi.org/10.1186/s12864-015-2277-7>
- Voshall, A., Moriyama, E.N., 2018. Next-Generation Transcriptome Assembly: Strategies and Performance Analysis. *Bioinforma. Era Post Genomics Big Data*. <https://doi.org/10.5772/intechopen.73497>
- Weber, A.P.M., 2015. Discovering new biology through sequencing of RNA. *Plant Physiol.* 169, 1524–1531. <https://doi.org/10.1104/pp.15.01081>
- Weirather, J.L., de Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X.-J., Buck, D., Au, K.F., 2017. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research* 6, 100. <https://doi.org/10.12688/f1000research.10571.1>
- Wu, T.D., Watanabe, C.K., 2005. GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21, 1859–1875. <https://doi.org/10.1093/bioinformatics/bti310>
- Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T.W., Li, Y., Xu, X., Wong, G.K.S., Wang, J., 2014. SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* 30, 1660–1666. <https://doi.org/10.1093/bioinformatics/btu077>
- Zerbino, D.R., Birney, E., 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. <https://doi.org/10.1101/gr.074492.107>
- Zhang, H., 2019. The review of transcriptome sequencing: Principles, history and advances. *IOP Conf. Ser. Earth Environ. Sci.* 332. <https://doi.org/10.1088/1755-1315/332/4/042003>

9 Final discussion

9.1 Obstacles and opportunities in research process

The main factor limiting the scope of our research was scarcity of research material available. All of the researched species that naturally occur in Poland are under partial or strict protection (articles 5.1, 5.2, 5.3, 5.5), and the *E. helleborine* specimens cultivated in pots (article 5.4) were of limited number, low mass and provided low yield of extracted nucleic acids. Due to the size and morphology, an amount of dry mass that can be obtained by sampling a single plant is very limited and, in many cases, barely allowed us to deliver presented results, not to mention introducing more biological replicates or perform additional experiments. In natura sampling process turned out to be extremely challenging – to avoid introduction of biases, as most plants were collected with highly sensitive transcriptomic and metabolomic research in mind, collection of plant material had to be performed within a certain time window in uniform weather conditions. Furthermore, the explants (as collected for articles 5.1, 5.2, 5.3, 5.5), had to be immediately processed by dividing into samples of equal mass suitable for extraction of nucleic acids and cryopreserved in liquid nitrogen to prevent rapid degradation of RNA material. Without access to sterile laboratory facilities in the field, a number of aseptic precautions and homebrew protective measures needed to be introduced and optimized to prevent contamination of samples with foreign agents, such as pollen or insect secretions, while remaining portable enough to be carried into the field by a minimal research team, and without drawing unnecessary attention of onlookers, that could endanger the biodiversity of habitats (Fig.5). This optimized fieldwork pipeline can be applied to further research to deliver reproducible and accurate results.



Fig. 4 Field sampling team at work under the protective tent

9. Final discussion

No readily available unified procedures were published for processing our type of samples. The nature of investigated material, mainly obtained from roots, proved to cause difficulties in processing, DNA and RNA extraction and sequencing. The main difficulty came from the presence of various soil-related compounds, e.g., fulvic and humic acid and multiple aromatic compounds, that impeded RNA extraction, enzymatic purification, and PCR amplification in sequencing library construction (Alm et al. 2000) (article 5.5). Introduction of polyvinyl polypyrrolidone (PVPP) as a purifying adsorbent during isolation resulted in significantly improved quality of final isolates, and though not being the most efficient method available, it was the only one feasible in our high-throughput application involving very large number of samples. The conventional protocol delivered by NucleoZOL reagent (our procedure of choice) resulted in unsatisfying yield of extracted RNA (finally resolved by fine tuning timing of extraction steps), that would be further reduced by complex cleaning procedure involving DNase digestions of remaining DNA contaminants and cleaning the isolate with Agencourt RNAClean XP (Beckman Coulter) magnetic beads suspension. That last procedure additionally involved use of laboratory equipment that was unavailable and unobtainable under tight deadlines of the research project. The problem required a DIY approach based on FDM 3D printing and use of repurposed materials, that ultimately replaced a costly commercial tool for magnetic separation and was used in multiple other applications (Fig.6). The 3D model files and assembly instructions have also been released as a freely available open source resource for 3D printing.

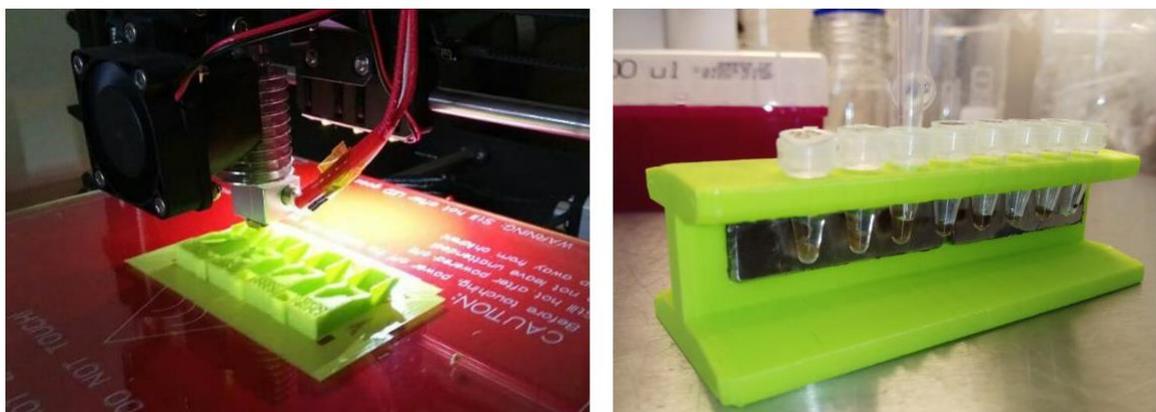


Fig. 5 Application of 3D printing for production of custom low cost laboratory equipment

Similarly, physical properties of processed samples made them extremely difficult to be macerated while deeply frozen. Most conventional equipment does not withstand extremely low temperatures or becomes a severe handling hazard while submerged in liquid nitrogen. A number of diverse methods needed to be developed for crushing different plant

9. Final discussion

tissues into fine dust, considering their hardness and high electrostatic attraction in their frozen state. The overall optimization process was largely based on trial and error and turned out to be particularly interesting. A high throughput approach involves use of a TissueLyser II (Qiagen) homogenizer. The device uses beads to pulverize the material inside test tubes by rapid shaking. After comparing a range of steel, tungsten carbide and ceramic beads, utilizing 6mm ceramic beads resulted in the fastest and most thorough grinding of leaf tissue. The temperatures optimal for grinding (safe for RNA and making tissues brittle) were well below a point of freezing of phenol-based NucleoZOL reagent, immobilizing the beads inside the microtubes - dry grinding was therefore favorable, delaying introducing the liquid reagent to after the tissue was ground. Low temperature made the tubes very brittle, increasing the risk of invisible cracks and fractures from bead impacts. These cracks, paired with low pressure occurring inside the tubes due to rapid contraction of cold air, led to liquid nitrogen seeping into the tubes during brief periods of submersion between grinding cycles. While it evaporated during the cycle, the pressure inside tubes grew, often to the point of explosion. Use of high quality microtubes and distributing them evenly inside the homogenizer's adapters (racks) (pressure of a rubber-coated adapter lid partially absorbed impacts on the tube lids, which were particularly prone to cracking in their central part) reduced this risk, but high pressure often ejected the fine material from the tube upon opening the lid – sometimes together with the lid, tearing it off its hinge and launching it as a painful projectile. These difficulties could be solved only by practice in careful lifting the lid before opening to release the pressure buildup. Electrostatic interaction between plastic tubes and frozen leaf samples often caused the tissue to behave erratically in handling. The timing of grind cycles and nitrogen submersions also required fine tuning, to minimize thawing of samples.

The automated method of grinding with TissueLyser was not effective against harder tissues, such as stem and roots. These samples required manual grinding with mortar and pestle – also flooded with liquid nitrogen. Seemingly simple in nature, even this process required long and pain-staking optimization – its problems including lengthy sterilization of ceramic mortars, pre-cooling them to reduce boil-off of LN₂ poured over the sample, and rather difficult recovery of sample dust from the mortar bowl and its transfer to Eppendorf tubes with NucleoZOL reagent (use of a steel spatula to scrape the ground sample “mud” while it is no longer suspended in boiling LN₂, but before it dries and returns to a form of loose dust; alternative method involving resuspending the dust in NucleoZOL inside the

mortar and collecting it with a pipette after it thaws – leading to more troublesome cleaning of used glassware). The overall process of refining this pipeline was challenging but turned out to be a valuable (and satisfying) lesson in general laboratory work and problem solving.

The rapid development in field of bioinformatics proved to be a challenge while processing data for article 5.5 and unpublished research (chapter 8). The ever-growing base of available software for sequence assembly, processing, and analysis, required much consideration and extensive testing on reduced datasets to commit to a single unified pipeline that would deliver the highest possible quality of results in the shortest computational time, and keep up with the standard in the field. Sharing of server time and resources required coordination of multiple computation tasks from parallel projects. Working with non-typical species meant less reference data available from public databases and less readily optimized methods or complete pipelines. In the end, however, I was able to collect a complete set of reliable procedures and methods that can be applied in further research on investigated species or can act as a basis for further modification and adjustments to new research materials or desired application. Due to rapid development in the field, however, the pipeline is not ultimate and will be constantly adjusted, modified or extended with new tools available in near future.

9.2 Extending scope of research and further research perspectives

As previously mentioned, the main limiting factor of my research was low amount of research material available. This reason alone has barred us from performing multiple additional analyses, such as stable oxygen and hydrogen content measurement. To build upon presented data, new species need to be brought into analysis – species of various tropisms to fill the gaps along the nutrition gradient, and from closely related groups of plants, to identify family-specific traits. Such addition will not only add new subjects for a more extensive comparative survey, but also will result in creating plastomic and transcriptomic references for phylogenetic inference applications. To avoid ambiguity, further isotopic analyses on mixotrophic plants should also include H isotopes, to detect NET carbon flow from rhizoctonia and other saprotrophic mycobionts.

A promising topic for further research branching from presented work concentrates on delivering comparative descriptions not only between species and their organs, but also along multiple trait gradients. A nutritional gradient study should involve assessment of

9. Final discussion

multiple closely related orchid species representing different stages along the progression from AT to MH nutrition. Such research could also involve observation of effect of light intensity in different study sites – to observe plasticity and changes occurring within a single species, but in conditions that would require or facilitate regulation of mixotrophy towards AT or MH. A complete methodology for reliable estimation of light conditions was designed, basing on Ellenberg index, momentary and averaged photosynthetically active radiation (PAR) measurement with solid state radiometers, and complex gap light analysis conducted on hemispherical photography of the sky (or forest canopy) above the researched specimens. We are currently conducting such analyses for multiple selected orchid species.

As all orchids display mycoheterotrophy during germination, a family-unique chance is present to perform analysis along the developmental gradient. By studying protocorms, we can observe orchid species, that exhibit different ultimate modes of trophism, in their juvenile state of obligatory mycoheterotrophy. Observation of two strategies within a single species, similarly to experiments conducted by Suetsugu et al. (2017) on albino and green specimens of mixotrophic *E. helleborine*, and by Lallemand et al. (2019c) on albino *C. damasonium*, *E. helleborine* and *E. purpurata.*, will generate a stronger comparative analysis than between two different species that may exhibit differences originating from lineage specific traits. Such approach may deliver valuable insight into nutrition plasticity and will help understand genesis and function of MH-related genes. The experiment is already underway to introduce protocorms into our comparative analysis together with the previous results. A “seed baiting” technique was used to promote symbiotic germination of orchid seeds *in situ* for multiple species (*C. damasonium*, *D. majalis*, *E. atrorubens*, *E. helleborine*, *E. palustris*, *N. nidus-avis*, *N. ovata*, *P. chlorantha*). Protocorms were successfully collected, and the material is currently undergoing analysis.

Another valuable, but cost- and workload-heavy expansion of existing research would involve complementing available or newly generated transcriptomic data with fitting metabolomic profiles from analyzed specimens and interpolation of transcriptome’s influence on plant’s metabolite makeup and metabolic pathways. This approach will also bring more insight into the previously mentioned concept of complex gene expression reprogramming events shaping the nutrition plasticity of plants.

Construction of high-quality genomic data for any of the researched (or closely related) species would greatly benefit to the depth of our comparative surveys and improve quality of transcriptome assembly, but due to high workload and costs, and low economic

significance of our subject species, this perspective is currently not feasible. The results of hybrid *de novo* transcriptome assembly procedures seem highly promising, and while currently still young and underdeveloped, it is a field for improvements in terms of completeness and accuracy of assembled transcriptomes. While it still requires further research to refine its performance, we have already successfully applied a combined SR and LR assembly to generate transcriptomes for multiple orchid species (results will be published soon).

9.3 Conclusion

In my presented work I have performed a broad investigative study of genetic basis of nutrition strategies in Orchidaceae family, including AT, MX and MH stages of trophic evolution. This dissertation concentrated on how the adjustment to particular trophism was reflected in changes in transcriptomes and plastidial genomes, therefore affecting – or resulting from – high nutritional plasticity of plants. I described research into basic state of plant autotrophy, analyzed tendencies of plastome reduction and retaining in evolution of mixotrophic species, investigated a possibility of adaptative reversal of mixotrophy and studied changes in genome that lead to emergence of mycoheterotrophy and that appeared as its consequences. My performed observations contributed to painting an image of versatile and adjustable genome that effectively undergoes changes to effectively provide nutrition as a response to variable conditions. During presented work I collected, unified, and optimized a toolset of in situ procedures, laboratory handling of samples, sequencing and bioinformatic analysis pipeline, that proved effective and reliable, and can be readily applied in future research. Additionally, my investigation into the potential of combining multiple sequencing approaches in a hybrid transcriptome assembly process will provide a basis for further improvements in consecutive research. This work delivered both potential subject and perspectives for further investigations, created basis and resources for them and developed optimized methods that will ensure consistent results.

In addition to being beneficial towards state of knowledge in the field and in general evolutionary sciences, presented research was a priceless personal learning opportunity and a lesson in non-linear exploratory approach. I learned to conduct and keep track of a complex study, which, rather than concentrate on a single pre-defined goal, dynamically adapts, and is expanded by new questions and hypotheses based upon gradually gathered data. I had a chance to cooperate with experienced specialists, and to benefit from their knowledge and

9. Final discussion

experience to improve my investigative methods and understanding of their results. Furthermore, I had an opportunity to use exciting, top-of-the-line technologies developed for biological research and utilize them to their maximum potential to provide answers to my research questions. The branching nature of this work required me to develop ability to apply both conventional, well established methods derived from literature, and -through their thorough understanding- to modify and combine them to create novel solutions adapted to my “non-model” needs. Personally, I found this study a tough but rewarding lesson in both patience and meticulousness, and in working under pressure of tight time constraints, unforgiving scarcity of irreplaceable research material and brutal reality of limited resources.

10 Supplementary materials

10.1 List of species involved in presented work

Out of 17 orchid species described in this thesis, 2 belong to Orchidoideae, and 15 to Epidendroideae subfamily (Table 3).

Tab. 2 List of Orchidaceae described in this work, their phylogenetic belonging and nutrition strategy

Species	Tribe	Subfamily	Nutrition	Article
<i>Cephalanthera damasonium</i>	Neottieae	Epidendroideae	MX	Lallemand et al., 2019b; chapter 6.3
<i>Cephalanthera longibracteata</i>	Neottieae	Epidendroideae	AT	Lallemand et al., 2019b; chapter 6.3
<i>Cephalanthera rubra</i>	Neottieae	Epidendroideae	MX	Lallemand et al., 2019b; chapter 6.3
<i>Epipactis helleborine</i>	Neottieae	Epidendroideae	MX	Lallemand et al., 2019b; chapter 6.3 May et al., 2020; chapter 5.4
<i>Epipactis albensis</i>	Neottieae	Epidendroideae	MX	Lallemand et al., 2019b; chapter 6.3
<i>Epipactis atrorubens</i>	Neottieae	Epidendroideae	MX	Lallemand et al., 2019b; chapter 6.3
<i>Epipactis gigantea</i>	Neottieae	Epidendroideae	AT	Lallemand et al., 2019b; chapter 6.3
<i>Epipactis microphylla</i>	Neottieae	Epidendroideae	MX	Lallemand et al., 2019b; chapter 6.3
<i>Epipactis palustris</i>	Neottieae	Epidendroideae	AT	Lallemand et al., 2019b; chapter 6.3
<i>Epipactis purpurata</i>	Neottieae	Epidendroideae	MX	Lallemand et al., 2019b; chapter 6.3
<i>Epipogium aphyllum</i>	Gastrodieae	Epidendroideae	MH	Jąkański et al., 2021; chapter 6.5
<i>Limodorum abortivum</i>	Neottieae	Epidendroideae	MX	Lallemand et al., 2019b; chapter 6.3
<i>Neottia cordata</i>	Neottieae	Epidendroideae	MX	Lallemand et al., 2019b; chapter 6.3
<i>Neottia nidus-avis</i>	Neottieae	Epidendroideae	MH	Jąkański et al., 2021; chapter 6.5
<i>Palmarichis pabstii</i>	Neottieae	Epidendroideae	AT	Lallemand et al., 2019b; chapter 6.3
<i>Platanthera chlorantha</i>	Orchideae	Orchidoideae	AT	Lallemand et al., 2019d, chapter 6.2
<i>Dactylohriza majalis</i>	Orchideae	Orchidoideae	AT	May et al., 2019; chapter 6.1

10.2 Photosynthesis – extended description

As important as it is for the functioning of the organisms that use it for nutrition, photosynthesis is a basis of most food chains of Earth's ecosystems. Being an introduction point of biologically useful forms of reduced carbon and energy, it supplies an entire biosphere, simultaneously acting as a major climate-shaping force. It is estimated that photosynthetic organisms on Earth deliver almost 223 petagrams of organic carbon a year, while also acting as a sink for atmospheric CO₂. Each year over 10% of total atmospheric CO₂ undergoes assimilation by photosynthetic organisms and is returned into circulation by metabolic processes of living organisms and bio-matter combustion. The byproduct of oxygen is also necessary for survival of oxygen-consuming organisms (Field et al., 1998).

In current biosphere, oxygenic photosynthesis is observed, most notably, in almost all lineages of plantae, algae and cyanobacteria. The first organisms to develop photosynthetic apparatus were relatively simple, bacterial in nature and similar to the present-day cyanobacteriaceae (Jensen and Leister, 2014). Their appearance is dated to early Proterozoic, 2.5-2.4 billion years ago (Bekker et al., 2004; Buick, 2008). Approximately a billion years later photosynthesis became available to eukaryotes by the means of endosymbiosis.

More sophisticated autotrophs have compartmentalized photosynthetic activity inside chloroplasts. Their origin has been proven to be a particularly tight form of symbiosis. Multiple eukaryotic organisms have developed symbiotic relationships with green prokaryota, but at some point in the history the photosynthesizing bacterium has been internalized by an eukaryotic predator and did not undergo digestion, instead becoming a cellular organelle. This theory of endosymbiosis, raised in 1905 by Mereschkowsky (Martin and Kowallik, 1999; Mereschkowsky, 1905), has since been proven by electron microscopy, biochemical and genetic studies. Chloroplasts were discovered to possess a double-layered membrane (specific to gram-negative bacteria), their own genome and transcriptional machinery, and conduct protein synthesis using a mechanism closely resembling the one utilized by bacteria. Modern phylogenetic studies indicate a close genetic relationship between plastids and cyanobacteria – in particular, *Gloeomargarita* (Hohmann-Marriott and Blankenship, 2011; Moore et al., 2019).

10. Supplementary materials

Photosynthesis consists of four main functions: light absorption, conversion, electron transfer and a multi-step enzymatic pathway of converting CO₂ into carbohydrates. The entire process can be divided into a light-dependent and light-independent (also called carbon reactions) stage. Light dependent reactions are means of generating and storing energy for performing the carbon reactions – Calvin-Benson cycle, carbon reduction cycle and related processes. The processes are spatially separated: thylakoids contain photosystem PS I and PS II membrane complexes and reaction centers responsible for energy harvesting and conversion and acting as the first step of the electron transfer chain. Further, transmembrane cytochrome complexes, plastocyanin proteins and plastoquinone move electrons from aqueous interior of lumen to the soluble redox agents residing in stroma. The process of trans-locating electrons creates an electrochemical gradient – the proton motive force, with potential difference across the thylakoid membrane. The energy of proton flux powers the chloroplast's ATP synthase, allowing it to phosphorylate ADP molecules (Nevo et al., 2009; Pyke, 2009).

All known phototrophs contain a variant of chlorophyll pigment inside and around their photosystems. It absorbs mostly blue and red light (430nm and 680nm wavelength), reflecting the green part of spectrum. Chlorophyll occurs in multiple variants differing in composition of their minor side chains. In vascular plants, variants a (blue-green) and b (yellow-green) are the most prevalent. Multiple other dyes are present in photosynthetic cells, of which the most notable are carotenoids, absorbing blue-green light (400-500nm) (and thus appearing yellow-orange). However, they act only as an accessory light harvesting pigment of minor significance, as their main function is concentrated on managing and alleviating the oxidative stress inside photosynthetic organs and structural role in light harvesting complex (LHC) formation. Of this family of pigments, β -carotene is the most commonly occurring orange-red dye, present in nearly all chloroplasts (Pyke, 2009).

Photosystems are functional and structural units of membrane-bound pigment-protein complexes and consist of reaction center (RC) and light harvesting complexes, also called antenna complexes. Light energy is absorbed by the chlorophyll a – containing antenna structures and transferred to the reaction center, where it causes excitation of chlorophyll molecule and release of electron, that is moved onto acceptor and into the electron transfer chain. Photosystems are differentiated based on their structure and differ in peak wavelength of absorbed light. PS I is based around an iron-sulfur cluster and contains a pair of chlorophyll a P700 dimers as a primary electron donor, while PS II contains quinone

and a pair of chlorophyll b P680 dimers. The photosystems also differ largely on their amino acid sequence level but share the same general structure of their trans-membrane domains and position of electron carriers, suggesting their common ancestry. (Pyke, 2009)

The electrons released from the reaction center are transferred onto electron carriers, one of which is plastoquinol (PQH₂), formed by reduction of plastoquinone in PS II. The electrons are delivered to cytochrome b6f, which transfers them further to plastocyanin (PC), while simultaneously resulting in transfer of protons from stroma into lumen. PS I transfers the electrons from PC to stroma via ferredoxin carrier, that undergoes reduction by specific reductase enzyme (FNR). The created gradient of protons generates a proton-motive force, which is ultimately used to power the ATP synthase. A flow of protons to stroma from lumen enables the enzyme to phosphorylate ADP into ATP – basic cellular energy carrier (Buchanan et al., 2015; Pyke, 2009).

The main mean of photosynthetic CO₂ assimilation is through Calvin-Benson cycle. The process is a cycle of 13 reactions in 3 phases: **carboxylation** of primary CO₂ acceptor – Ribulose 1,5-bisphosphate (RuBP), **reduction** of 3-phosphoglycerate (3-PGA) by employing photosynthetic ATP and NADPH, and **regeneration** of RuBP acceptor by consuming ATP. A noteworthy enzyme, which is crucial for CO₂ assimilation, is RuBisCo, or RuBP carboxylase. Its secondary function is involved in photorespiration, where it acts as an oxygenase. The enzyme consists of eight big and eight small subunits and requires a set of chaperon proteins to be folded correctly (Pyke, 2009; Wise and Hooper, 2006).

Some lineages of plants evolved a modification of this 'default' carbon fixation process (also called C₃ photosynthesis, as CO₂ is primarily bound into 3-PGA). The variant called C₄ photosynthesis involves formation of oxaloacetic acid (OAA) as an intermediate CO₂ carrier. This modification increases the carboxylation rate of Rubisco, while simultaneously decreasing oxygenase activity and inhibitory effect of photorespiration.

Another modification, the crassulacean acid metabolism (CAM) photosynthesis, allows for temporary division between CO₂ uptake and fixation, by storing it inside vacuoles as an intermediate medium – malic acid. This solution is beneficial for plants in dry and hot climate, e.g. *Crassulaceae*, but has also been observed in some *Orchidaceae* (Bone et al., 2015; Luo et al., 2014).

The genomes of modern chloroplasts are strongly reduced compared to their free-living ancestors. The cyanobacteria *Gloeomargarita*'s genome contains approximately 3000

10. Supplementary materials

genes, indicating loss of 95% genes since transition from a free-living bacterium. Initial gene loss involved a large set of genes from free-living bacteria, which turned out to be redundant after the endosymbiosis event. The process of reduction continued even after what is considered a common ancestor of all modern plastidial genomes, independently in separate lineages, but convergently under common evolutionary constraints, guided by constant need for carbon fixation (Martin et al., 1998; Race et al., 1999, Kwon et al., 2020). A horizontal transfer of prokaryotic genes from plastome to nuclear genome has been observed. Based on detection of genes of plastidial (prokaryotic) origin, several organisms have been proven to once had possessed chloroplasts, but have since entirely lost them – e.g., some species of diatoms, *Plasmodium* (Wilson et al., 1996, Yu et al., 2018). In terrestrial plants, even 10-15% of nuclear genome may originate from early photosynthetic endosymbionts – even up to estimated 18% of protein-coding genes in *Arabidopsis thaliana* (Martin et al., 2002; Cannon, 2018). It is theorized that plastome-nucleus gene transfer is facilitated by degradation of paternal line chloroplasts in male gametogenesis and uptake of their genetic material (Eckardt, 2006; Bock, 2015). Simple introduction of the sequences via non-homologous end joining in illegitimate repair process is not sufficient to create a retained, stable “nuclear integrant of plastid DNA” (NUPT) (Leister, 2005). Their prokaryotic-type regulatory sequences necessary for initiating transcription are incompatible with nuclear transcription mechanisms. The process of obtaining promoter sequences and compartment-targeting sequences is rather complex (Huang et al., 2004; Matsuo et al., 2005) and usually requires multiple rearrangements to form a functional, transcribable gene that can be retained. If such transfer process is successful, complementation by nuclear expression causes removal of selective pressure on its retaining within plastome, and gene dosage effect may even actively promote its loss. Nuclear genome integration of plastidial genes may also be promoted over retaining a plastome thanks to providing tighter integration into cell’s expression regulatory mechanisms.

A number of nuclear genes compensates for essential functions removed from contemporary plastidial genomes. Rubisco, consisting of two subunits, cannot be constructed without importing externally synthesized protein, as only the large subunit is plastid encoded. HCF136, a crucial factor for PSII reaction center assembly, is encoded by a nuclear gene of clearly prokaryotic origin (Meurer et al., 1998). Genes coding for all 4 subunits of plastidial polymerase PEP (plastid encoded polymerase) core proteins are contained in cpDNA, but most of its cofactors are translated from nuclear genes (Börner et

al., 2015). Second plastidial polymerase, NEP (nuclear encoded polymerase), needs to be imported into the chloroplast from cytoplasm, where it is expressed from its nuclear gene. Due to the division in translational machinery, most of the crucial photosynthetic structures are specific nuclear-plastidial hybrids and require tight cooperation between nuclear and plastidial genome for efficient expression (Sloan et al., 2018).

10.3 Non-photosynthetic functionality of chloroplasts

Outside of their main function of conducting photosynthesis, chloroplasts have been recorded to perform or contribute to multitude of unrelated activities. They are responsible for delivery of multiple crucial organic compounds: they produce all the fatty acids in a plant cell, synthesize and store starch and contain machinery responsible for production of all 20 amino acids (Hözl and Dörmann, 2019; Lancien et al., 2007). They also house galactolipid-synthesizing protein complexes (Rottet et al., 2015) and plastoglobuli – thylakoid-bound lipoprotein bodies, which are a site of lipid synthesis and storage (Buchanan et al., 2015; Maréchal, 2018).

In addition to providing a set of biosynthetic machinery, chloroplasts take part in mediation and performing immune response to pathogen infection. Two pathways of pathogen counteraction have been discovered – by forced apoptosis of infected cells, or by systemic response. Chloroplasts are involved in regulation and initiation of both pathways by decoupling the photosynthetic machinery and production of large concentration of reactive oxygen species (ROS) (Sowden et al., 2018). Their presence directly destroys invading cells, while simultaneously initiating the systemic response by inducing synthesis of anti-pathogenic substances in the rest of the plant. Other stress-related signalization media released by the chloroplasts are jasmonic acid, salicylic acid, and nitric oxide (Ali and Baek, 2020; Fragnire et al., 2011; Jasid et al., 2006).

10.4 Practices for reliable hybrid de novo assembly (...) – supplementary materials

FIGURES AND TABLES

Table 8.1. SeqStat based general statistics from consecutive steps of Nanopore long read correction.

MERGED fwd+rev	Lordec k19	Lordec k19+21	Lordec k19+21+23	Lordec k19+21+23+19	Lordec trim	Min length cutoff: 200 bp	CD-HIT clustering
Number of sequences	6 638 367	6 638 367	6 638 367	6 638 367	6 622 519	5 909 322	1 689 883
Total # residues	6 170 381 157	6 172 098 905	6 172 781 872	6 172 841 105	5 291 700 530	5 224 287 029	1 400 375 461
Smallest	75	75	75	75	19	200	200
Largest	7727	7727	7727	7727	7605	7605	7605
Average length	929.95	930.2	930.3	930.35	799.45	884.1	828.7

Table 8.2. Mapping statistics of long reads from each step of correction.

	non-zero mappings	% of reference [#]	mapped sequences	% mapped
k19+21+23+19	22 313	66.39%	6 490 273	97.77%
k19+21+23+19 trimmed	21 968	65.36%	6 324 395	95.50%
k19+21+23+19 trimmed, over 200 bp	21 291	63.35%	5 908 022	99.98%
k19+21+23+19 trimmed, over 200 bp, clustered	21 078	62.71%	1 688 516	99.92%

[#] Reference annotation of *A. thaliana* with 33610 annotated genes.

Table 8.3. Basic statistics of the final hybrid assemblies.

	assembly	over 200bp									
		ALL sequences	sequences	residues	largest	average	N50	median			
preassembly											
Trinity		208 662	208 662	157 406 482	6870	754,4	1082	507			
Soap 31 contigs		468 134	413 701	347 207 617	15 563	839,27	1042	726			
Velvet91		788 715	788 675	618 371 473	8216	784,06	966	662			
Velvet_meta_27_contigs	IDP	1 517 851	1 242 992	1 026 490 979	7605	825,82	984	757			
Velvet_meta_27_transcripts		751 492	647 389	467 827 967	13 090	722,64	955	532			
trans/Abyss merged		424 809	424 809	382 410 848	15 563	900,19	1180	682			
oases_longMerged.transcripts		269 368	248 245	200 242 432	15 676	806,63	1046	594			
oases_longMerged.contigs	oases	267 316	63 549	25 982 288	5495	408,85	437	323			
oases_longMerged.contigs.tr2aacds		34 003	30 898	16 756 293	5495	542,31	581	450			
	Trinity LONG	60 584	60 584	75 087 243	15 572	1239,39	1732	1012			
	SPAdes	37 759	28 967	38 772 149	15 571	1338,49	1976	1092			
	rnaSPAdes	51 973	47 201	78 973 962	15 571	1673,14	2191	1459			
	rnaSPAdes -ss rf	51 942	47 143	76 050 997	15 571	1613,2	2083	1419			
Reference cDNA		48 359	48 037	86 423 674	16 593	1799,11	2067	1602			

10. Supplementary materials

Table 8.4. Basic statistics of *de novo* pre-assemblies based on Illumina short reads. Overall highest scores are underlined.

Preassembly	k-mer	total sequences	% below 200bp	Above 200bp					
				sequences	residues	largest	average	N50	median
Trinity	25	61773	0%	61 773	88 238 898	<u>15571</u>	<u>1428</u>	<u>1924</u>	<u>1219</u>
Velvet	meta 27	422 045	25%	<u>317 942</u>	<u>193 995 105</u>	13 090	610	748	435
	meta 61	419 030	25%	314 787	186 027 964	13 090	591	711	420
	21	416 111	86%	56 583	17 606 068	2632	311	303	267
	25	350 725	82%	62 712	20 238 396	2845	323	315	272
	31	288 494	76%	69 060	23 179 805	3855	336	330	278
	41	190 270	59%	78 337	30 065 564	4890	384	395	300
	51	112 998	38%	70 237	36 622 671	7526	521	641	358
	61	71 146	18%	58 520	37 513 925	12 304	641	925	407
	71	46 779	2%	45 640	35 409 549	13 090	776	1213	488
	81	34 766	1%	34 460	30 750 842	8505	892	1313	621
	91	28 890	0%	28 850	21 958 355	8216	761	999	545
trans/abyss	21	85 782	0%	85 782	61 387 828	12 013	716	993	469
	25	81 961	0%	81 961	67 587 572	15 563	825	1210	539
	31	82 502	0%	82 502	71 401 437	15 563	865	1287	570
	41	77 569	0%	77 569	72 279 189	15 563	932	1394	630
	51	66 325	0%	66 325	66 521 112	15 563	1003	1504	702
	61	56 582	0%	56 582	59 820 953	15 563	1057	1574	762
	71	47 309	0%	47 309	51 832 055	13 056	1096	1613	826
	81	37 302	0%	37 302	41 340 295	13 055	1108	1619	854
	91	28 383	0%	28 383	28 384 564	9053	1000	1469	770
	meta	105 062	0%	105 062	117 187 700	15 563	1115	1598	817
SOAP contigs	25	116 684	60%	46 399	37 049 120	13 966	798	1230	500
SOAP scaffold	25	70 724	48%	36 474	41 864 600	15 564	1148	1765	825
SOAP contigs	31	96 863	56%	42 430	37 606 096	15 563	886	1401	564
SOAP scaffold	31	70 682	50%	35 511	42 142 156	15 563	1187	1803	886

Table 8.5. Comparison of hybrid assembly mapping statistics.

	% bases in reads of assembly that map to reference	% genes of genome hit by assembly	% genes of reference covered by assembly	% non-zero mappings of reference	mean breadth of coverage	mean depth of coverage
IDP_soap31contigs	90.99%	99.98%	99.91%	66.67%	1556	304.8
IDP_transabyss_merged	91.55%	99.98%	99.92%	62.89%	1557	184.76
IDP_Trinity	88.91%	99.32%	99.22%	6.41%	1543	69.65
IDP_velvet_meta_27_contigs	88.52%	100%	99.98%	71.06%	1557	1911
IDP_velvet_meta_27_transcripts	92.60%	99.99%	99.94%	70.36%	1557	211.9
oases_longMerged.contigs	98.88%	89.69%	75.89%	64.64%	1353	2.489
oases_longMerged.contigs.tr2aacds	99.48%	58.70%	8.38%	47.10%	578.4	0.334
oases_longMerged.transcripts	89.48%	99.54%	99.43%	63.27%	1547	43.6
rnaSPAdes hybrid	99.19%	80.42%	46.40%		1238	2.035
rnaSPAdes hybrid ss rf	98.98%	80.06%	45.94%		1227	1.965
SPAdes	97.69%	84.14%	66.80%		1288	1.554
Trinity_AT_fq_LONG	98.88%	83.25%	57.89%	57.76%	1282	2.179

Table 8.6. Timing and memory usage records.

Process	Runtime (minutes)			Memory usage (GB)		
	Min	Max	Median	Min	Max	Median
IDPdenovo	44	1227	67	4.45	5.60	4.48
Trinity	164			67.13		
Trinity -long	636			67.13		
Oases merge	6 min (in pipeline)			1.01	1.24	1.13
Velvet	22	45	48	23.62	23.62	23.62
Velvetg	28	137	77.5	23.62	28.61	23.62
Velvet -long	51			39.35		
Velvetg -long	263			31.81		
Velvet meta long pipeline	4352			61.69		
SPAdes	22			17		
maSPAdes	30	33	31.5	20		
trans/abyss	55	152	136	5.20	9.09	7.92
trans/abyss merge	314			0.84		
SOAPdenovo/trans	24	56				

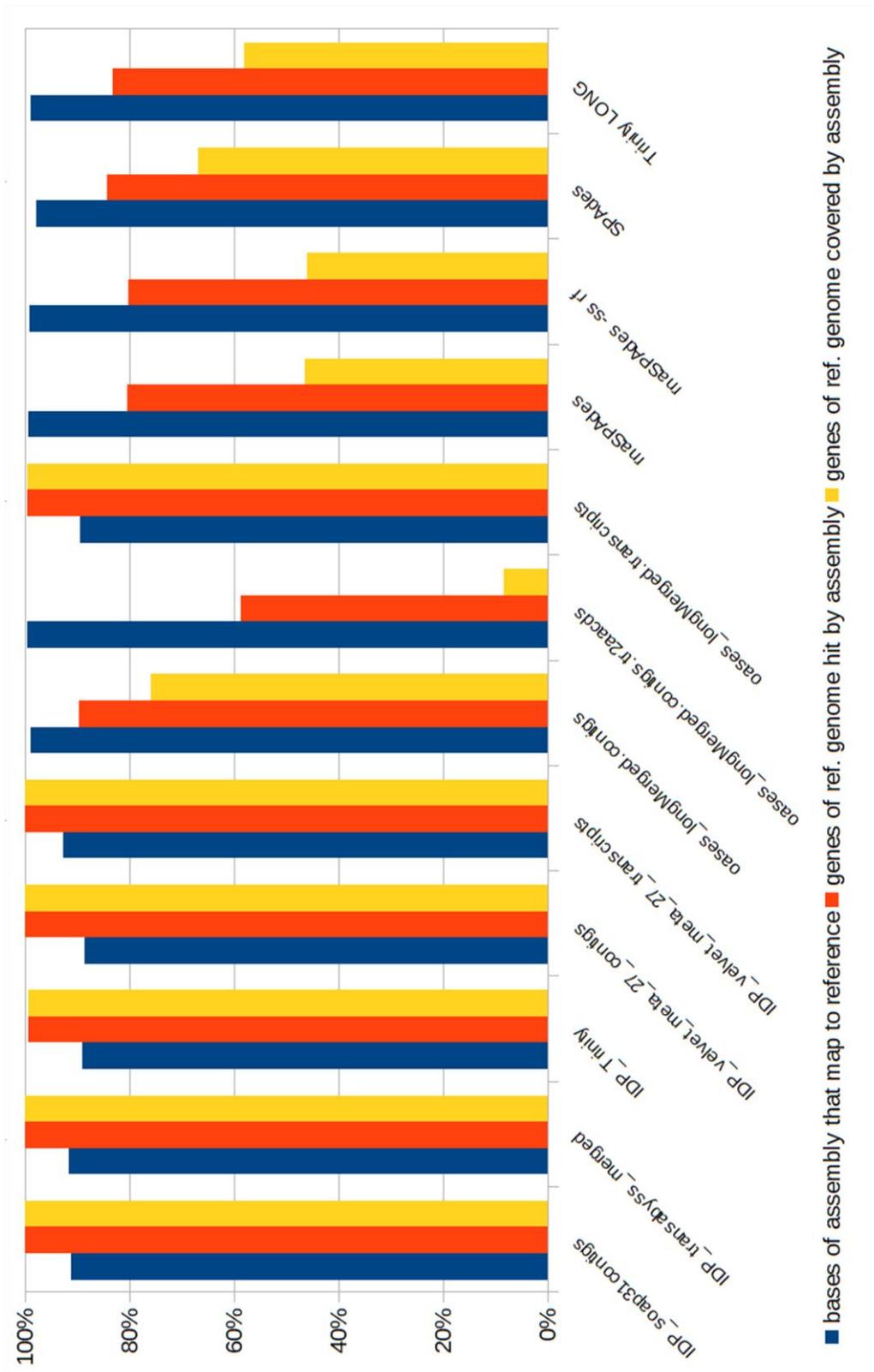


Fig. 8.1 Hybrid assembly mapping stats plot

10. Supplementary materials

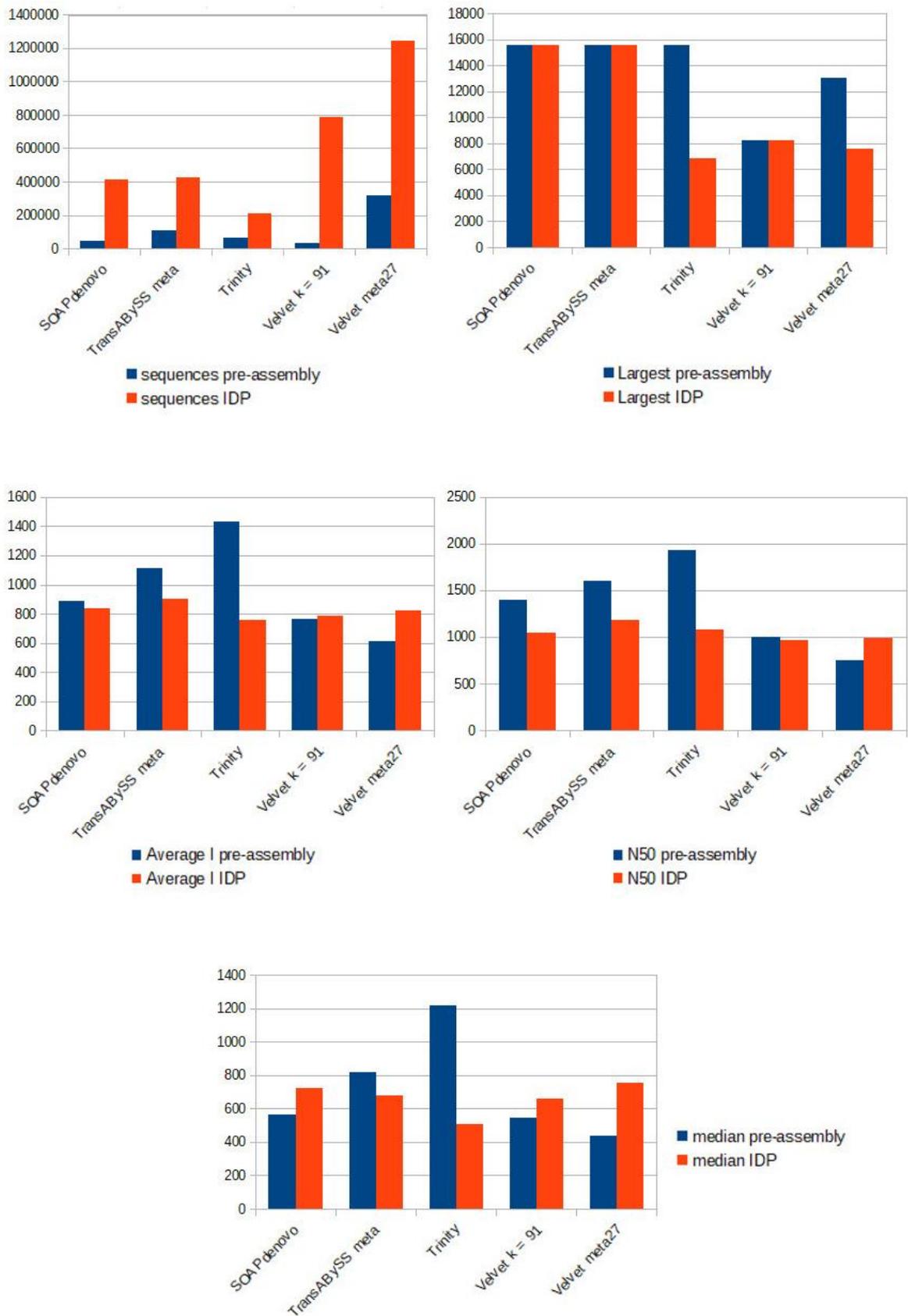


Fig. 8.2 Comparison between IDPdenovo output assemblies and their corresponding pre-assemblies

LIST OF COMMANDS

A) LONG READ CORRECTION:

- LoRDEC (<http://atgc.lirmm.fr/lordec/>)

```
lordec-correct --threads 24 -2 "${FWD}" "${REV}" -k "${KMER}" -s "${ABUNDANCE}" -i "${LONG}" -o "${OUTPUT}"
```

FWD - forward Illumina reads

REV - reverse Illumina reads

KMER - k-mer length - used k values in sequence: 19, 21, 23, 19

ABUNDANCE - Abundance threshold = 3

LONG - Nanopore long read file or corrected long read file (during following correction iterations)

OUTPUT - corrected sequence file

```
lordec-trim -i "${INPUT}" -o "${OUTPUT}"
```

INPUT - corrected sequences for trimming

OUTPUT - trimmed sequence file

- seqtk (<https://github.com/lh3/seqtk>)

```
seqtk seq -L "${CUTOFF}" "${INPUT}" > "${OUTPUT}"
```

INPUT - trimmed sequence file

CUTOFF - length cutoff threshold = 200 (basepairs)

OUTPUT - sequence file above 200bp

- cdhitest (<https://github.com/weizhongli/cdhit>)

```
cd-hit-est -i "${INPUT}" -o "${OUTPUT}" -c "${CUTOFF}" -c "${IDENTITY}" -M "${RAM}" -T "${CPU}" -n "${N-LENGTH}" -d "${DESC}"
```

INPUT - trimmed sequence file

OUTPUT - clustered sequence file

IDENTITY - sequence identity threshold = 0.97

RAM - memory limit = 80000 (MB)

CPU - CPU thread limit = 20

N-LENGTH - word length = 10

DESC - length of description in .clstr file = 150

B) PRE-ASSEMBLY AND DE NOVO ASSEMBLY**TRANSABYSS:**

```
transabyss --pe "${FWD}" "${REV}" --name "${ASSEMBLYNAME}" --outdir "${OUTDIR}" --kmer "${KMER}" --threads  
"${CPU}" --length "${LENGTH}"
```

FWD - forward Illumina reads

REV - reverse Illumina reads

ASSEMBLYNAME - output assembly name

OUTDIR - output directory name

KMER - k-mer length, following values were used: 21, 25, 31, 41, 51, 61, 71, 81, 91

CPU - CPU thread limit = 24

LENGTH - minimum contig length = 200 (basepairs)

TRINITY (<https://github.com/trinityrnaseq/trinityrnaseq>)

```
Trinity --seqType fq --SS_lib_type RF --left "${FWD}" --right "${REV}" --max_memory "${RAM}" --CPU "${CPU}" --  
output "${OUTPUT}"
```

FWD - forward Illumina reads
REV - reverse Illumina reads
RAM - memory limit = 120G (gigabytes)
CPU - CPU thread limit = 24
OUTPUT - output assembly file

SOAPDENOVO:

SOAPdenovo-Trans-31mer all -s soap.config -K "\${KMER}" -p "\${CPU}" -o "\${OUTPUT}"

CPU - CPU thread limit = 24
KMER - k-mer length
OUTPUT - output assembly filename

soap.config configuration file:

```
[LIB]
q1=LEFT_SR.fq
q2=RIGHT_SR.fq
```

C) HYBRID ASSEMBLY

TRINITY (<https://github.com/trinityrnaseq/trinityrnaseq>)

```
Trinity --seqType fq --SS_lib_type RF --left "${FWD}" --right "${REV}" --long_reads "${LONG}" --max_memory "${RAM}" --CPU  
"${CPU}" --output "${OUTPUT}" --full_cleanup
```

FWD - forward Illumina reads

REV - reverse Illumina reads

LONG - long reads file

RAM - memory limit = 120G (gigabytes)

CPU - CPU thread limit = 24

OUTPUT - output assembly filename

VELVET-OASES:

-velveth "hashing"

```
velveth "${NAME}" "${KMER}" -fastq -shortPaired -separate "${FWD}" "${REV}" -strand_specific
```

NAME - assembly name
KMER - k-mer length (values used: 21, 25, 31, 41, 51, 61, 71, 81, 91)
FWD - forward Illumina reads
REV - reverse Illumina reads

-velveth with long reads

```
velveth "${NAME}" "${KMER}" -fastq -shortPaired -separate "${FWD}" "${REV}" -strand_specific -fasta -longPaired -interleaved  
"${LONG}"
```

NAME - assembly name
KMER - k-mer length
FWD - forward Illumina reads
REV - reverse Illumina reads
LONG - long reads file

-velvetg graph processing

```
velvetg "${NAME}"
```

NAME - processed assembly path/name

-OASES hybrid multi-k meta-assembly

```
export OMP_NUM_THREADS=${CPU}
```

```
oases_pipeline.py -m "${K-MIN}" -M "${K-MAX}" -s "${STEP}" -g "${K-MERGE}" -o "${OUTDIR}" -d " -fastq -shortPaired -separate  
"${FWD}" "${REV}" -strand_specific -fasta -longPaired -interleaved "${LONG}"
```

CPU - CPU thread limit = 24

KMIN - minimum k-mer length = 21

KMAX - maximum merged k-mer length = 91

STEP - k-mer length increase increment = 10

KMERGE - k-mer length used for merging process = 27

OUTDIR - output directory name

FWD - forward Illumina reads

REV - reverse Illumina reads

LONG - long reads file

- tr2aacds "filtration"

tr2aacds.pl -mnrnaseq "\${CONTIGS}" -NCPU "\${CPU}" -MAXMEM "\${RAM}"

CONTIGS - processed assembly filename

CPU - CPU thread limit = 24

RAM - memory limit = 120G (gigabytes)

IDPDENOVO:

idpdenovo.py "\${PREASSEMBLY}" "\${LONG}" "\${FWD}" "\${REV}" -o "\${OUTDIR}" --specific_tempdir "\${TEMPDIR}" --threads "\${CPU}"

PREASSEMBLY - pre-assembly sequence filename

LONG - long reads file

FWD - forward Illumina reads

REV - reverse Illumina reads

OUTDIR - output directory name

TEMPDIR - temporary directory filename

CPU - CPU thread limit = 24

SPADES:

-SPAdes assembly

```
spades.py -1 "${FWD}" -2 "${REV}" --nanopore "${LONG}" -t "${CPU}" -o "${OUTPUT}"
```

FWD - forward Illumina reads

REV - reverse Illumina reads

LONG - long reads file

CPU - CPU thread limit = 24

OUTPUT - output assembly filename

-rnaSPAdes assembly

```
spades.py --rna -1 "${FWD}" -2 "${REV}" --nanopore "${LONG}" -t "${CPU}" -o "${OUTPUT}"
```

FWD - forward Illumina reads

REV - reverse Illumina reads

LONG - long reads file

CPU - CPU thread limit = 24

OUTPUT - output assembly filename

-rnaSPAdes assembly with rev-fwd ordering

```
spades.py --rna --ss rf -1 "${FWD}" -2 "${REV}" --nanopore "${LONG}" -t "${CPU}" -o "${OUTPUT}"
```

FWD - forward Illumina reads

REV - reverse Illumina reads

LONG - long reads file

CPU - CPU thread limit = 24

OUTPUT - output assembly filename

D) MAPPING TO REFERENCE GENOME**deSALT (<https://github.com/ydLiu-HIT/deSALT>)**

```
deSALT aln ${TAIR10_chr_all} "${INPUT}" -t "${CPU}" -o "${OUTPUT}"
```

TAIR10_chr_all - TAIR10 chromosomes *Arabidopsis* genome reference file
INPUT - pre-assembly for genome alignment
CPU - CPU thread limit = 24
OUTPUT - .sam alignment file

GMAP (<https://github.com/juliangehring/GMAP-GSNAP>)

```
gmap -D "${gmap_index_dir}" -d "${gmap_genome}" -t "${CPU}" -f samse "${CONTIGS}" 1> "${contigs}"-  
toATgenome.gmap.samse"
```

gmap_index_dir - reference genome GMAP index directory
gmap_genome - reference genome GMAP index filename
CONTIGS - input assembly name
CPU - CPU thread limit = 24
GFF - reference genome gene structure file = TAIR10_GFF3_genes.gff
FASTA - reference genome sequence file = TAIR10_chr_all.fas

E) ASSESSMENT OF MAPPING TO THE REFERENCE GENOME

```
RNAseqEval.py eval-mapping "${FASTA}" "${CONTIGS}.samse" -a "${GFF}" --leave_chrom_names -o
"${CONTIGS}.rnaseqeval.txt"
featureCounts -T "${CPU}" -a "${GFF}" -o "${CONTIGS}.samse-featurecounts.txt" "${CONTIGS}.samse" -I
samtools view -@ "${CPU}" -bs "${CONTIGS}.samse" | samtools sort - -@ "${CPU}" -o "${CONTIGS}.sorted.bam"
bedtools coverage -a "${GFF}" -b "${CONTIGS}.sorted.bam" > "${CONTIGS}.sorted.bam-bedtools_coverage.txt"

bedtools coverage -a "${GFF}" -b "${CONTIGS}.sorted.bam" -mean > "${CONTIGS}.sorted.bam-
bedtools_coverage_mean.txt"

./parse_bedtools.R "${CONTIGS}.sorted.bam-bedtools_coverage.txt" "${CONTIGS}.sorted.bam-
bedtools_coverage_mean.txt"

CONTIGS - input assembly name
CPU - CPU thread limit = 24
GFF - reference genome gene structure file = TAIR10_GFF3_genes.gff
FASTA - reference genome sequence file = TAIR10_chr_all.fas
```

10.5 Methodology – extended description

10.5.1 Genomics

The broad term of genomics involves application of recombinant DNA technology, sequencing methods and bioinformatic analysis for assembling genomes and annotating their functionality, adaptability, and multiple factor response, researching evolutionary processes that shaped them, and for tracing and resolving their phylogeny. A wide range of research methods from the genomic toolset has been used to perform extensive research on a number of species of particular economical or scientific significance, designating them as model organisms for plant sciences. Before the rise of modern sequencing era, many model genomes were sequenced using Sanger sequencing – most notably, *Arabidopsis thaliana* genome, which was sequenced completely in 2000 (The *Arabidopsis* Genome Initiative, 2000). However, due to high cost and workload required for such procedure, it has not been economically feasible for most non-model species. Plants that exhibit long life cycle before delivering offspring, are difficult to grow or propagate outside of their natural habitat and do not play a significant role in horticulture, simply did not make the cut, and are still missing reference sequencing data. The emergence of NGS technologies enabled a new, previously impossible branch of non-model plants investigation (Bräutigam and Gowik, 2010, Unamba et al., 2015).

A high quality, well annotated and well described reference genome is the main basis for understanding complex concepts of gene functionality and regulation (Feuillet et al., 2011). Modern Next Generation Sequencing approach has made these accessible and affordable, but still is not omnipotent. Several traits can pose a significant obstacle in genome construction – most notably, duplication of genome, heterozygosity, ploidy level and presence of highly repetitive sequences are severe problems for genomics (especially in experiments basing on the short-read technologies, such as Illumina – which are particularly convenient, thanks to their cost efficiency) and call for ingenious combined approaches to overcome (Hirsch and Buell, 2013).

10.5.2 Three generations of DNA sequencing

DNA sequencing as a method of investigating genomes revolutionized evolutionary science. First generation methods, biochemical approach described by Sanger and Coulson in 1975 (Sanger and Coulson, 1975) and chemical approach by Maxam and Gilbert in 1977 (Maxam and Gilbert, 1977), were revolutionary, but required extensive workloads, handling radioactive chemicals, were extremely expensive and did not provide the throughput required for extensive research projects. Replacing radioisotopes with fluorescent dyes and early computerization allowed for daily throughput up to a few kbp per day per machine, with read lengths reaching 1200 bp. The introduction of NGS, the Next Generation Sequencing, brought the much-needed throughput and ease of access, with commercially available equipment able to sequence millions of base pairs a day, while simultaneously reducing workload and error rate. All library preparation can be performed *in vitro*, and thanks to introduction of robotic workstations, automated flow cells, advanced digital reading systems, and multiplexing on a single flow cell, the throughput of NGS greatly outperforms its predecessors. An issue that remained to be addressed was read length, which is crucial for resolving complex and repetitive genome sequences. The current, third generation of sequencing technologies concentrates on a single molecule and long read approaches to simplify assembly of structures, that are extremely challenging to short read based NGS technologies (Unamba et al., 2015).

10.5.2.A Illumina SBS

The most prevalent NGS technology for generating short reads is currently Illumina sequencing by synthesis (SBS) (Goodwin et al., 2016). In this platform, a sequencing DNA library has to be first equipped with adapters (consisting of an oligonucleotide for complementary attachment, barcode sequence and primer binding site). Such constructs are bound onto the surface of a flow cell coated with oligonucleotide attachment sites and are amplified in a bridge amplification process. Reverse sequences are removed from the cell. Every time a nucleotide-dye complex mixture is washed over the flow cell, a single complementary nucleotide is attached to the strands, resulting in fluorescence on certain wavelength, which can be detected and used for automated base calling. A single run, depending on used platform and chemistry kit, can generate up to even 750 gigabases of sequence in reads of up to 300 bp in length, sequenced in both directions (paired ends) (Kim

10. Supplementary materials

et al., 2021; Lahens et al., 2017). This technology (Illumina HiSeq 2000, 2500, 4000) was employed in this work to sequence plastidial genomes in May et al., 2019 - chapter 6.1, Lallemand et al., 2019d - chapter 6.2 and Lallemand et al., 2019c - chapter 6.3.

Tab. 3 Illumina sequencing platforms parameters, according to manufacturer's specifications

Platform	Run time	Max. output	Av. read output	Max. reads per run	Max. read length
iSeq 100	9,5-19h	1,2Gb	4 mln	4 mln	2 x 150bp
MiniSeq	4-24h	7,5Gb	14-16 mln	25 mln	2 x 150bp
MiSeq	4-55h	15Gb	22-25 mln	25 mln	2 x 300bp
NextSeq550	12-30h	120Gb	260 mln	400 mln	2 x 150bp
NextSeq 1000	11-48h	360Gb	400 mln	1,2 bln	2 x 150bp
HiSeq 3000/4000	24-84h	750Gb	250-400 mln	2,5 bln	2 x 150bp

* <https://emea.illumina.com/systems/sequencing-platforms.html>

10.5.2.B Roche 454

Roche 454 sequencing technology, currently considered obsolete and discontinued since 2016, is a noteworthy legacy variant of pyrosequencing (King and Scott-Horton, 2008), also bases on sequencing by synthesis. During library construction, DNA is fragmented and ligated with adapters. Clonal amplification is then performed by emulsion PCR on the surface of beads, to which the DNA strands are attached. Actual sequencing process takes place inside fiber optic chip wells, each housing a single bead with copies of a single DNA fragment, together with buffers and enzymes. A microfluidic system sequentially delivers dyed nucleotides for strand synthesis, and incorporation events are detected by light emission. Roche 454 sequencing outperforms Illumina SBS platforms in terms of read length (~700bp) but provides much lower number of reads (700 000 reads, limited by a number of available wells) (Kulski, 2016; Luo et al., 2012).

10.5.2.C Ion Torrent

Ion Torrent sequencing, a third-generation technology, is based on detection of protons released as the nucleotides are incorporated during strand synthesis (Guerrero-Sanchez et al., 2019; Kulski, 2016; Lahens et al., 2017). DNA fragments are labeled with adapters and linked to Ion Sphere Particles (beads), and clonally amplified on their surface by emulsion PCR. Coated beads are loaded on a silicon chip and into wells equipped with

10. Supplementary materials

electronic proton detectors. Sequencing is primed from a specific point on the adapters and bases are introduced sequentially. If the incorporation of currently present base occurs, positive charge of a proton release is detected, signal strength being proportional to a number of bases being attached. The approach of electric potential measurement instead of optical detection of base incorporation removes the need for using modified bases (base-fluorophore complexes), greatly reducing the exploitation cost of chemicals. A single well on a chip can generate a read of up to 400bp. This method was applied to perform fungal barcoding in May et al., 2020, chapter 5.4.

10.5.2.D SMRT (PacBio)

PacBio SMRT, or Single Molecule Real Time sequencing, runs a strand synthesis reaction with a single polymerase molecule on a single DNA strand in each well of the chip. The bases are labeled with fluorescent dyes, which are cleaved during incorporation into the strand. The volume of the well (called ZMW – zero mode waveguide) is small enough to restrict fluorescence to the single released fluorophore molecule and permit its quenching by diffusion outside the well fast enough to observe the synthesis process in real time. The chip's construction allows for sequencing strands of up to 15 kbp length with very high accuracy and uniform coverage (no G-C bias). However, the throughput of SMRT is limited by the number of wells on a chip – typically, a single SMRT cell can deliver 0,5-1 Gbp in 47 000 reads. In addition, it requires larger amount of template material than its counterparts (Kulski, 2016; Rhoads and Au, 2015; Wang et al., 2019).

10.5.2.E ONT Nanopore

ONT Nanopore MinION and PromethION involve use of a transmembrane pore complex (either a biological membrane, or artificial solid-state structure) to sequence particularly long continuous reads. A difference of electrochemical potentials is generated across the membrane, causing translocation of sequenced molecules through the pores. As the structural element currently residing inside the pore affects its ion conductivity, a measurement of current flow allows for identification of shape and dimensions of channel "obstruction" and determine the base sequence (or other structures). The size and portability of this platform (MinION is a USB "dongle" that fits on an open hand and can work with a laptop computer), the remarkable simplicity of its application protocol (can be used for direct

10. Supplementary materials

detection of DNA modification, without any modifications or library construction), and low price of the device itself, distinguish ONT platforms from other technologies. The read error rate of this procedure is relatively high but is compensated by impressive read length – on average 23,8 kb and maximally up to 2,3 Mb (Wang et al., 2021) - and throughput of up to 30Gb on a chip (Amarasinghe et al., 2020; Payne et al., 2019).

Tab. 4 ONT sequencing platforms parameters, according to manufacturer's specifications

Platform	Run time	Max. output per cell	Av. read output	Channels per cell
Flongle	16h	1,8Gb	1Gb	126
MinION Mk1B	72h	42Gb	30Gb	512
MinION Mk1C	72h	42Gb	30Gb	512
GridION Mk1	72h	42Gb	30Gb	512
PromethION 24	64h	245Gb	200Gb	2675
PromethION 48	64h	245Gb	200Gb	2675

* <https://nanoporetech.com/products/comparison>

10.6 Transcriptomics

10.6.1 RNA sequencing

NGS and third generation sequencing has also found its way into RNA applications. While a study of the genome provides us with data about organism's genetic potential, hardly does it reflect on the transient state of proteome, which is strongly affected by various internal and external factors. The key to making the connection is transcriptome – a complete set of all RNA molecules transcribed by the organism (mRNA, rRNA, tRNA and various small coding and non-coding RNAs). Knowledge of full transcriptome sequence allows for novel gene discovery, helps in understanding complex isoforms and alternative gene splicing products, and can be used for quantification of gene expression and detection of differentially expressed genes (DEGs). Before NGS, studies of gene expression were conducted mostly using Sanger-based SAGE (serial analysis of gene expression) and microarrays. **RNA sequencing**, or **RNA-seq**, revolutionized the transcriptomic workshop. It is a powerful and cost-efficient method of obtaining, constructing and characterizing entire transcriptomes (Garg and Jain, 2013; Kulaeva et al., 2017).

10.6.2 RNA-seq experiment design

The methodological design of transcriptomic experiments must be tailored with specific research goal in mind and adjusted to the species being investigated. Knowledge of related species' genomes and transcriptomes may bring valuable insights – expected genome size, complexity, level of ploidy and repeat content can help select efficient and effective strategies for library construction, sequencing, transcriptome assembly and further downstream analysis. A well-thought-out plan is especially crucial to prevent technical biases in experiments involving large numbers of samples that need to be processed in multiple batches (Chung et al., 2021; Conesa et al., 2016).

10.6.3 Choice of samples and experimental groups

Tissue samples will be collected differently for different research goals. To construct a complete transcriptome, multiple organ samples from different stages of life and different conditions will be collected and pooled into the bulk RNA pool (as opposed to single-cell RNA sequencing, scRNA-Seq, which relies on material isolated from a selected cell instead). On the other hand, to observe and compare expression in response to various internal and

external stimuli, or to highlight differences in expression profile between organs, life stages or various aspects of plant's physiological condition, samples will need to be carefully selected to accurately reflect on a particular indicator. A number of biological replicates will be a trade-off between the practical aspect (availability) of sample collection vs. representativeness and statistical power of an experimental group (minimizing the technical and biological variability) (Conesa et al., 2016).

10.6.4 rRNA depletion

Ribosomal RNA (rRNA) constitutes over 80-90% of total cellular RNA. In most transcriptomic experiments its content is of no scientific interest, and its abundance over messenger RNA (mRNA) (1-2%) would result in majority of costly sequencing throughput being wasted on reads that are discarded in subsequent data filtering. To address this issue, the sample for library preparation undergoes either mRNA enrichment or rRNA depletion. Desired mRNA can be enriched using poly(A) selection (capturing polyadenylated 3' ends of transcripts). The process, however, requires input samples with high proportion of non-degraded RNA (quantified as RIN – RNA integrity number), which is notably difficult to isolate, especially with limited amount of sample material. In such cases (and for bacterial rRNA that does not exhibit polyadenylation) the only viable method is to directly deplete rRNA (Herbert et al., 2018). This can be achieved by capturing rRNA by binding to complementary oligonucleotides and separating them by magnetic bead precipitation (Illumina's RiboZero, Qiagen GeneRead rRNA depletion, Lexogen RiboCop), or by hybridizing the rRNA to DNA oligonucleotides and degrading the complexes using RNase H (NEBNext rRNA depletion, Kapa RiboErase, Takara/Clontech's RiboGone). In presented work, such method of rRNA depletion was performed in Jąkalski et al., 2020, using RiboZero Plant kit (Illumina, San Diego, CA, United States).

10.6.5 Sequencing library construction

Library preparation step involves removal of residual DNA contamination, mRNA fragmentation, reverse transcription, and introduction of sequencing adapters. The repeatability and convenience of this process can be greatly improved by the use of automated robotic workstations. Libraries can be multiplexed for the economic utilization of available flowcell lanes and sequencing reads. The procedure itself is specific to a chosen sequencing platform.

10.6.6 Sequencing platform

Choice of sequencing platform will depend on expected parameters of sequenced transcriptome. Currently the most dominant technology on the market is Illumina SBS, which was also our method of choice used in experiments presented in this work – Illumina NextSeq500 was used to perform RNA sequencing in Jąkowski et al., 2020 (chapter 6.5). Choice of sequencing depth (average number of reads covering a single base of complete assembly) is also a trade-off between cost and quality. Increased sequencing depth will result in higher transcript detection, more accurate quantification and easier higher chance of detection of rare transcripts, which are usually expressed at very low levels. Above certain threshold, quality increase from increasing the depth will be minimal, while greatly increasing transcriptional noise, prolonging processing time, and complicating the calculations. Therefore, the usual go-to values are ~10x for reference-based assembly, and over 30x for *de novo* process (Garg and Jain, 2013). A realistic approach has to take into consideration available computing power for completing the assembly in reasonable time (Sims et al., 2014).

10.6.7 Raw read processing

Raw sequencing data require pre-processing. Technical remnants of sequencing process, such as adapter sequences and primer sites are removed, and low-quality terminal parts of reads are trimmed. Then, the reads undergo error correction (especially relevant for LRs, that contain more errors than SRs), using either self-correction or additional sequences to correct mis-sequenced bases. If read cannot be corrected with reasonable degree of certainty, they are discarded from the read pool. The entire pre-processing is usually performed using readily available software, but still requires to be fine-tuned to optimize the outcome.

10.6.8 Transcriptome assembly

The assembly process aligns and merges short sequence segments to reconstruct the original long progenitors. Transcriptome assembly can take two approaches:

- **Reference-based assembly** – a generally preferred method, basing on a reference genome for the studied species or its close phylogenetic relative. Reads from RNA-seq are aligned to the reference. Overlapping sequences are identified, clustered, and used to construct a graph of isoforms. The most frequently applied method is de Bruijn graph, in which sequences are recomposed into a “dictionary” of k-mers, arranged into a graph basing on the overlaps, and resolved for optimal path along the graph (Compeau et al., 2011; Lin et al., 2016). The formulated sequence is used for construction of a final transcriptome. This approach is very sensitive and easily detects transcripts expressed at very low levels. It is also capable of detecting novel transcripts and alternative isoforms and requires relatively low computing power. Unfortunately, for reference-based assembly to be possible, a reference genome of the same species needs to be readily available, which is rarely the case in studies on non-model species. In some cases, a reference from a closely related species may bear sufficient similarity (Conesa et al., 2016; Love et al., 2016).

- **De novo assembly** – applied for non-model species without genomic reference present. The transcriptome construction is performed using only RNA-seq reads with no external reference. The process is much more challenging and requires significantly larger datasets (higher sequencing depth), resulting in higher demand for computational power and time. Many bioinformatic tools are being constantly developed for *de novo* assembly, the most commonly used being Oases (Schulz et al., 2012), Trans-ABYSS (Robertson et al., 2010), SOAPdenovo-Trans (Xie et al., 2014) and Trinity (Grabherr et al., 2013). This approach can detect novel transcripts, even from areas of genome that were previously not sequenced and is capable of assembling multiple alternative isoforms and trans-spliced constructs. In addition to usual short-read-based datasets, the assembly can be refined by using long reads from third generation sequencing platforms in a hybrid assembly.

A combination of these approaches is also possible. To refine a reference-based assembly, a *de novo* process is performed on reads that were not mapped to the genome; alternatively, a *de novo* assembly is ran on both unmapped fragments and contigs previously obtained from a reference-based process.

10.6.9 Transcriptome quality evaluation

The outcome of assembly process, a finished transcriptome, is evaluated based on several criteria. A comparison to a reference transcript reliably informs us about accuracy, completeness, contiguity, chimerism and variant resolution provided by the assembly. In a *de novo* case, the quality has to be estimated indirectly, by assessing the assembly statistics, such as: number of assembled contigs (which should be as close as the number of expected genes and their isoforms), their average length, total length of the assembly, the N50 length parameter (length of the shortest contig in a set of the longest contigs that -summed together- cover 50% of assembly's total length), average sequencing depth, percentage of un-assembled reads that could not be mapped to transcriptome (indicating non-optimal assembly) or percentage of reads that can be mapped to multiple sites in the transcriptome (indicating redundancy or chimerism). Optimization of various assembly parameters may, however, negatively affect other aspects of the outcome – e.g., by increasing sensitivity, detection rate can be improved, but raising algorithm's "greediness" will result in increase in numbers of redundant or mis-assembled contigs. The trade-offs must be managed on a study-specific case, with particular research goal in mind (Honaas et al., 2016; Wang and Gribskov, 2016).

10.6.10 Annotation and downstream application

The finished transcriptome is annotated by assigning functions to detected transcripts. The annotation can be conducted on several criteria:

- A direct sequence similarity search – BLAST (Altschul et al., 1990) alignment against a well annotated nucleotide or protein sequence database (e.g., Nt or Nr databases of NCBI) for closely related species; also, against known protein domains database (alignment performed optimally with HMMER (hmmr.org))
- Gene ontology terms (GO) assignment, basing on homology to proteins from GO databases (geneontology.org) to assign transcripts to 3 categories: biological process, molecular function, and cellular component (Ashburner et al., 2000; Carbon et al., 2021)
- KEGG database of metabolic pathways (<https://www.genome.jp/kegg/>), by similarity to gene families (Kanehisa and Sato, 2020)
- Identification of transcription factors and regulatory mechanisms, can indicate relation to a certain process or pathway

10.6.11 Expression quantification and differential gene expression

Basing on read representation in the assembly, a measure of gene expression level can be derived by calculating reads mapped to each transcript sequence. A basic strategy involves mapping the reads, e.g., using STAR (Dobin et al., 2013), HISAT (Kim et al., 2019), BBSMap (Bushnell 2014) or Salmon (Patro et al., 2017), and converting generated mapping into a matrix of raw mapped reads with tools such as HTSeq-count (Anders et al., 2015) or featureCounts (Liao et al., 2014). Extracting expression data for reliable comparison between multiple samples requires data normalization that will compensate for read length, library size and sequencing biases, e.g., by processing data using TPM (transcripts per million), FPKM (fragments per kilobase of transcript per million fragments mapped) or TMM (trimmed mean of M values) normalization (Robinson and Oshlack, 2010; Zhao et al., 2021). To calculate expression on transcript-level, complex compensating algorithms, such as HISAT (Kim et al., 2019), StringTie2 (Kovaka et al., 2019), Ballgown (Pertea et al., 2016), RSEM (Li and Dewey, 2011) and Sailfish (Patro et al., 2014) have been developed. These methods deliver normalized values corrected for sequencing biases, optimized for comparison between samples.

10.7 Plastome sequencing

Since the construction of the first complete plastidial sequence in 1986 (*Nicotiana tabacum*, Shinozaki et al., 1986), the plastome (plastidial genome) analysis workshop has gone a long way and was equipped with numerous invaluable tools. The original approach involving restriction cleaving and sequencing of overlapping clones was later replaced by Sanger sequencing of PCR-amplified sequences (Taberlet et al., 1991). With the rise of NGS technologies, researchers were provided with multiple cost- and workload-efficient solutions for complete plastome assembly. In general principle the process of plastome construction involves genomic DNA isolation, optional enrichment of sequences that are plastidial in origin, sequencing library construction, sequencing and bioinformatic data processing: dataset filtering, assembly, and post-processing.

10.7.1 DNA extraction for plastome sequencing

Total DNA extraction for plastomic application can be conducted using routine genomic procedures – column-based kits or CTAB isolation process. Oftentimes, however, the process needs to be optimized, to adjust for species-specific difficulties (e.g., high content of phenolic compounds) or degradation of sampled material (e.g., herbarium samples). In presented work, a number of different methods was applied for plastidial DNA extraction: DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) in May et al., 2019 - chapter 6.1 and Lallemand et al., 2019d - chapter 6.2; DNeasy Plant Mini Kit (Qiagen), NucleoSpin Plant II kit (Macherey-Nagel) and CTAB extraction as in Porebski et al. (1997) in Lallemand et al., 2019b - chapter 6.3.

Extracted total DNA isolate contains both desired DNA originating from plastids and nuclear and mitochondrial DNA. Due to high copy number of organelle genomes, highly abundant plastidial DNA (making up to 5% of entire cell's DNA) provides much higher sequencing coverage. Therefore, synthesizing a sequencing library directly from total DNA isolate and using it for plastome assembly is viable and often practiced. Such approach, referred to as "genome skimming", offers significantly lowered workload and high cost efficiency compared to enrichment-based protocols (Twyford and Ness, 2017). Within this work, such approach was performed in May et al., 2019 - chapter 6.1, Lallemand et al., 2019d - chapter 6.2 and Lallemand et al., 2019b - chapter 6.3. However, in species-specific

cases, the abundance of plastidial DNA is not sufficient and the approach has to include an enrichment step.

10.7.2 Plastidial DNA enrichment

Plastidial DNA can be separated from nuclear and mitochondrial contamination by isolation of entire plastids. This can be achieved via several methods: centrifugation in sucrose density gradient, either commercial (e.g., Sigma Chloroplast Isolation Kit) or "home-brew" (Miflin and Beevers, 1974); by precipitation in high salt concentration or by DNase I treatment (Shi et al., 2012). While resulting in high quality *de novo* assemblies, methods based on plastid extraction usually require large amount of tissue material, deliver low yield DNA isolates, and require PCR amplification; furthermore, presence of mitochondrial DNA contamination must be considered. Additional steps result in significantly smaller throughput and increased costs, and the need to fine-tune and tailor the methods specifically to a single species affects the credibility of their application to multi-species surveys.

Another approach involves extraction of whole cellular DNA, and selective capture of plastidial sequences. This can be achieved based on structural differences between nuclear and plastidial DNA. A property that allows for differentiation is strand methylation level. Nuclear DNA, being a perfect example of eukaryotic DNA, displays significantly higher degree of methylation than plastidial DNA, prokaryotic in origin. Degradation of DNA strongly affects efficiency of this method (Yigit et al., 2014).

Hybrid bait capture employs specifically designed oligonucleotide probes ("baits"), that bind known plastidial sequences and can be separated on magnetic beads (Stull et al., 2013). As this method strongly relies on conservation of plastidial DNA, it poses a high likelihood of introducing assembly errors in abnormal plastomes – it will be affected by plastome rearrangements and changes. It is also incapable of detecting nuclear transfer events.

10.7.3 PCR amplification of plastomes

Thanks to relatively small size and high conservation, it is feasible to perform a PCR amplification and sequencing of an entire plastome. Short range PCR paired with Sanger sequencing is an accessible research approach, requiring less advanced facilities than NGS

and simplified bioinformatic processing. However, the approach involves amplification with numerous primers (commonly used set by Dong et al., 2013 consists of 138 primer pairs generating amplicons of 800-1500bp), which creates a large amount of manual labor and prevents high-throughput application; the need for specifically designed primers also requires knowledge of a closely related reference plastomes, which greatly reduces value of this method in evolutionary research. Application of long-range PCR (Bethune et al., 2019) (6-12kbp in length) anchored in highly conserved areas of plastome resolves part of these issues – it requires significantly less manual processing (usually uses around 16 primer pairs), which greatly increases throughput of the process, and transcript tagging and multiplexing can be applied to reduce costs of NGS. This method, however, still carries drawbacks of PCR-based approach – a plastome of closely related species is still required to accurately select a set of primers and avoid generating gaps in coverage; PCR amplification of long fragments from highly degraded samples is also likely to fail.

10.7.4 Direct plastome sequencing

Modern sequencing platforms are capable of generating reads of lengths exceeding most common repeat fragments present in plastidial genomes, allowing for *de novo* assembly of high quality plastomes. For genome skimming approach, a large collection of reads is required to provide necessary depth of coverage – therefore high output platforms like Illumina HiSeq are a go-to tool for this method. For isolated plastids or enriched sequencing libraries, lower output platforms like MiniSeq, MiSeq or Ion Torrent are sufficient. Long read sequencing methods (SMRT, Oxford Nanopore) can be used to sequence entire plastome in a single read, providing a valuable tool for guided assembly if paired with short reads, or on their own for phylogenetic analyses (Scheunert et al., 2020).

10.7.5 Plastome assembly

Plastome assembly for model plant species and well described families is relatively straightforward, thanks to availability of reference plastomes, that can be used to guide the assembly process. In non-model species, highly divergent groups or in species with significant alterations to their plastome (such as parasites or mycoheterotrophs), researchers must resort to *de novo* approach. To improve assembly quality, after standard quality filtering and preliminary read assembly, sequence libraries are filtered to selectively extract plastidial

10. Supplementary materials

sequences and remove nuclear transcripts. Filtering can be based on several factors: differences in read coverage (by measuring rate of k-mer occurrence in library), statistical G-C pair content (Li and Du, 2014) or on similarity to known plastidial sequences by mapping to the closest available reference plastome. Structure-based approach allows for identification and purging of plastidial sequences translocated to nuclear DNA and does not discriminate highly rearranged or altered plastidial sequences, that may be lost in sequence-based filtering. Therefore, a well-balanced combination of filtering techniques is crucial for high quality plastome assembly. Reducing complexity of the library and normalizing the read coverage (by removal of low coverage nuclear fraction artifacts) significantly simplifies (and shortens) the assembly process and improves quality of the outcome. Multiple assembly tools do not rely on preliminary read filtration and include filtering algorithms that are enacted during assembly step (Freudenthal et al., 2020).

The plastome assembly step can be performed with a large variety of available assembler software, often derived from programs designed for bacteria genome sequencing. Commonly used *de novo* tools include ABySS (Simpson et al., 2009), chloroExtractor (Ankenbrand et al., 2018), QUIAGEN CLC Genomic Workbench (digitalinsights.qiagen.com), Edena (Hernandez et al., 2008), Euler-sr (Chaisson and Pevzner, 2008), Fast-Plast (github.com/mrmckain/Fast-Plast), GetOrganelle (github.com/Kinggerm/GetOrganelle), Geneious (www.geneious.com), NOVOPlasty (Dierckxsens et al., 2016), ORG.Asm (pythonhosted.org/ORG.asm), SOAPdenovo (R. Luo et al., 2012), SPAdes (Bankevich et al., 2012), SSAKE (Warren et al., 2007) and Velvet (Zerbino and Birney, 2008) (Ekblom and Wolf, 2014; Freudenthal et al., 2020). The resulting assemblies usually produce several long contigs corresponding to LSC, SSC and IR regions, which can be finally integrated into a full plastome sequence. Naturally, this step is followed by annotation of the plastome-encoded genes, which can be done automatically using a dedicated software, and later curated with the help of an expert.

10.8 Fungal identification by ITS barcoding

An effective tool for accurate identification of fungal species is crucial for investigation of mycorrhizal dynamics. Fungi kingdom is estimated to contain over 3,8 mln distinctive species (Hawksworth and Lücking, 2017) which exhibit highly variable morphology. Identification by morphological features is limited to species developing macroscopic structures, which are often temporary or rare (Slepecky and Starmer, 2009), and often display cryptic features (Bickford 2006, Badotti 2017). Over the years, multiple targets for molecular identification have been suggested, including LSU - large ribosomal subunit, and SSU - small ribosomal subunit sequences and various protein coding genes (*tefl* coding for EF1-a elongation factor, *tub1* and *tub2* beta-tubulin genes, actin-coding *act1* and RNA polymerase II - *rbp1*, *rbp2*). Current consensus was reached in 2012 (Schoch et al., 2012), establishing ITS region polymorphism as a golden standard in fungal barcoding.

ITS, an internal transcribed spacer, is a region inside ribosomal tandem repeat gene cluster, between small and large rRNA subunits (Fig.6). The entire region is transcribed as a single cistron, and is post-translationally removed in splicing, therefore exhibiting high variability as a non-coding sequence. In fungi, ITS spans approximately 600bp, consisting of two variable spacers: ITS1 and ITS2, separated by highly conserved 5.8s rRNA gene and flanked by SSU - 18S rRNA at 5' terminus of ITS1 and LSU - 28S rRNA at the 3' terminus of ITS2. Their location, nested between highly conserved regions, is very convenient for research and allows it to be amplified in most fungal species using universal sets of primers. Furthermore, a haploid genome contains multiple copies in tandem repeats of rRNA gene cluster, increasing yield of template from low amount of material.

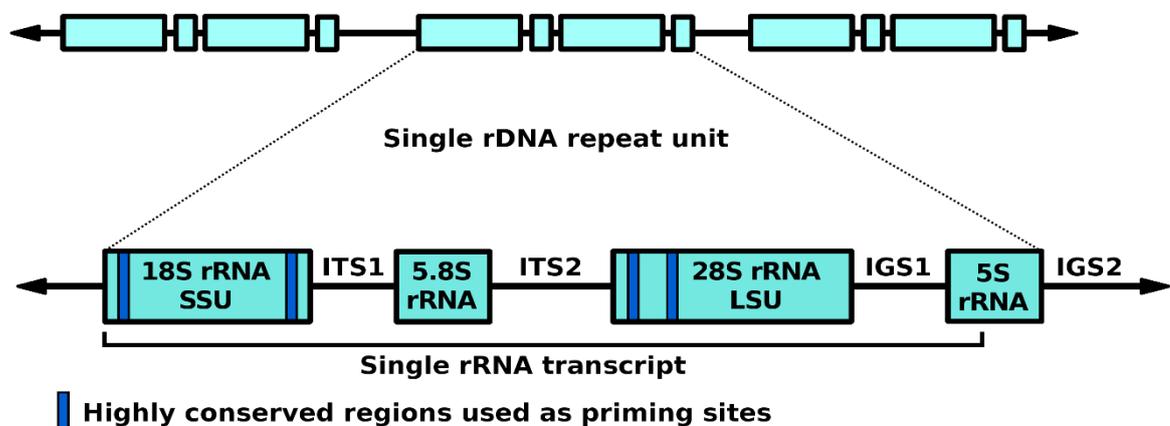


Fig.6 Ribosomal RNA gene cluster structure

10. Supplementary materials

A basic workflow in ITS barcoding involves isolation of DNA, selective amplification of ITS region and its analysis basing on length or sequence data. A universal set of primers has been developed specifically for the task of fungal identification. Their choice is critical for effective species identification and should be the best fit to most fungal taxa to ensure success of PCR and minimize bias towards particular groups, while still providing high specificity to avoid amplification of non-fungal sequences (especially troublesome for soil samples containing diverse fungal communities and non-fungal contaminants, e.g., bacteria, plants, nematodes). An optimal strategy is to select primers closely matching the expected taxa (Toju et al., 2012).

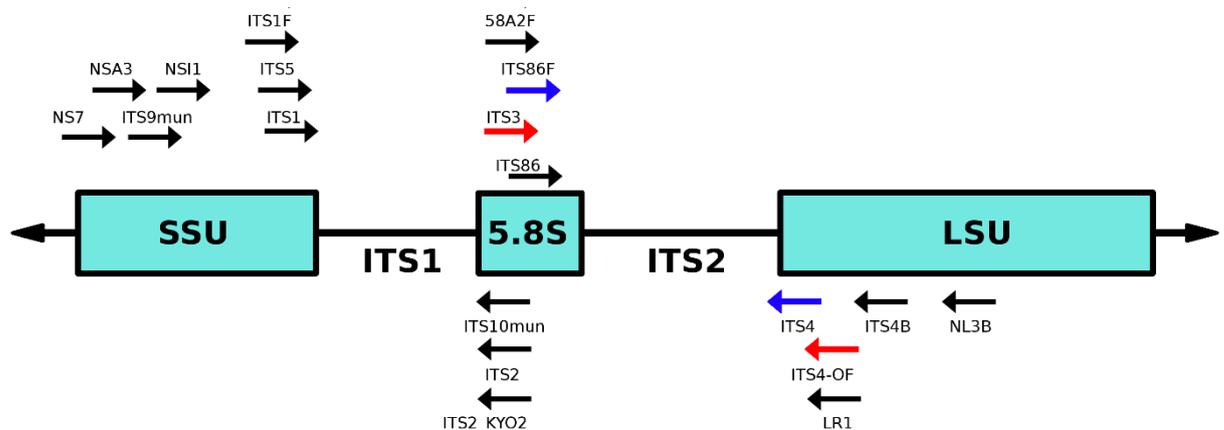


Fig.7 Commonly used fungal ITS primers and their annealing sites. Primer pairs used in this work (May et al., 2019 – chapter 5.4) marked in red and blue

The most common universal ITS barcoding primers (Fig.7) were developed to be used with LH-PCR, RFLP and T-RFLP (White et al., 1990), but are well suited for sequencing with NGS methods and their tagged variants are readily available. Thanks to increasing length of NGS reads, most modern sequencing platforms can generate high quality ITS sequences requiring little to no assembly, that can be used for identification down to species level by alignment to reference sequences (fingerprinting). Multiple databases collect fungal barcode sequences for species identification, including BOLD (www.boldsystems.org), GlobalFungi (globalfungi.com), INSDC (www.insdc.org), ISHAM ITS base (its.mycologylab.org), EPPO-Q-bank (qbank.eppo.int), RDP (rdp.cme.msu.edu) and UNITE (unite.ut.ee).

10.9 Phylogenetic inference

Phylogenetics, or the study of relationships between species, have gone a long way since the Darwinian approach. Modern methods of phylogenetic inference are no longer based on phenotypic and morphological description for determination of evolutionary relationships, turning to highly quantitative and reliable molecular phylogeny instead. The progress of computational phylogenetics fed by NGS data allows for accurate tracking of evolutionary processes that shaped contemporary species and inferring on their time of divergence. Furthermore, extensive amount of data can be obtained from minor amounts of preserved material.

Selection of proper **phylogenetic markers**, being gene or protein sequences under investigation, is a primary factor in experimental design. A marker usually must provide resolving power on the level of species or subspecies in researched taxa. Preferable markers are single copy genes (to avoid multiple variants being present), that do not cause ambiguities in alignment process due to complex structure, large scale insertions, deletions or rearrangements, and exhibit relatively large degree of conservation, and therefore are slow to accumulate mutations. Alternatively, SSRs (simple sequence repeats) and indel (insertion/deletion) sites can be themselves used as markers. An important factor is also availability of specific primers for amplification of a selected marker. Some of the most used markers in plant phylogeny are nucleus-residing ribosomal RNA genes: 18S rRNA from SSU and 5S rRNA, 5.8S rRNA and 25S rRNA from LSU. Internal Transcribed Spacers (ITS) are also frequently used. Another group includes mitochondrial DNA: mitochondrial 12S rRNA. Currently the most promising source of plant phylogeny data are, however, plastomes. Their high degree of conservation, rare occurrence of large structural rearrangements, low length and complexity, are highly useful in resolving species relationships and inferring on evolutionary history. Some of the most used plastidial markers are *atpB* (ATP synthase beta subunit), *matK* (maturase K), *ndhF* (F subunit of NADP dehydrogenase), *rbcL* (rubisco) genes and *rpl16* (non-coding intron). In Orchidaceae, phylogeny surveys based on cpDNA are relatively young due to limited collection of plastome sequences but appear to deliver consistent lineages that correlate well with phylogenies inferred using other datasets (Freudenstein and Chase, 2015; Górnjak et al., 2010; Kim et al., 2020).

After a desired marker set is selected and sequenced, multiple sequence alignment is constructed using aligner software, such as ClustalW (Thompson et al., 1994), Mauve

10. Supplementary materials

(Darling, 2004), MAFFT (Kazutaka et al., 2002). Next, the most fitting substitution model is selected to explain the processes and events that led to creation of current state of relationships between aligned sequences. The most basic models include **Jukes-Cantor model**, that assumes equal rate of substitution of purines and pyrimidines; and **Kimura two parameter model**, that favors transition events over transversion. Model selection can be guided by automated statistical algorithms, such as IQ-TREE ModelFinder (www.iqtree.org/ModelFinder), CIPRES Science Gateway (www.phylo.org) or jModeltest (Posada, 2008).

According to the selected model, a **phylogenetic tree** can be constructed using several methods, basing either solely on evolutionary distance between sequences (a **distance matrix** approach, calculating a mean number of substitutions per site to measure a distance between two sequences descending from a common ancestors, and using it as a measure of relatedness), or taking into consideration the process of change accumulation that generated the differences observed in the alignment (**character-based** approach). These include **maximum parsimony** trees, constructed by comparing all possible trees containing investigated sequences and striving for the lowest number of evolutionary changes that could have led to a current state; and **maximum likelihood** trees, in which all possible trees are evaluated by probability of appearance of a particular sequence as a result of hypothesized evolutionary processes. The robustness of constructed trees can be evaluated using statistical methods such as, for example, **Bootstrapping** or **Jackknifing**. Bootstrapping process involves repeatedly reconstructing the same tree using altered datasets (by dropping sequences and replacing them with duplicates of others) and counting how often does certain branch appear. Jackknifing, instead of replacing sequences, drops them from the database altogether and reconstructs the trees on reduced datasets.

In functional genomes not all sequences exhibit equal rate of evolution. Localized deviations from assumed substitution rate derived from a model can be used as a measure of selective regime acting upon a particular gene, and selective pressure may be quantified by comparing the ratio of synonymous and non-synonymous mutations. Such analyses can be conducted using PAML codeml (<http://abacus.gene.ucl.ac.uk/software/paml.html>) and RELAX algorithm (datamonkey.org/relax).

10.10 Isotopic analysis methods

Radioisotope labeling was first described in 1950s (Badin and Calvin, 1950), and quickly became a crucial tool for molecular biology, that allowed researchers to investigate, trace and explain numerous metabolic pathways and cycles, with photosynthetic carbon circulation as its first achievement. The main principle behind isotopic labeling are nearly identical chemical properties and reactivity shared by isotopes of the same element that compose of the same number of protons and electrons but differ in neutron number. A basic tracer (isotope labeling) experiment involves introduction of a certain amount of isotopically "marked" chemical into the source pool of a biological process, having it pass through the pathway being investigated, and its recovery from one or multiple sink pools. Usage of radioactive isotope as a marker allows for easy quantitative detection by scintillation or visualization by autoradiography (Schimel, 1993). Tracking of radioactive carbon ^{14}C and phosphate ^{32}P was used in 1960 (Bjorkman, 1960) to speculate on the existence of hyphal connections between trees and certain plants of groundcover, which was proven using very similar approach by McKendrick et al. in 2000 (McKendrick et al., 2000).

10.10.1 Stable isotope natural abundance analysis

Stable isotopes may also be used as a molecular tracer. As they do not emit ionizing radiation, their detection is more complicated and requires advanced techniques (IRMS – isotope-ratio mass spectrometry). They can, however, be applied to tracking elements that have no natural radioisotopes (e.g., oxygen and nitrogen), and do not pose a radiation hazard to researchers and environment.

Despite having nearly identical properties, naturally occurring isotopes do exhibit slightly altered kinetics in biological reactions. Therefore, isotopic composition of different ecosystem compartments is predictable and can be employed in **stable isotope natural abundance analysis** experiments for studying the flow of matter and interpreting the outcome of naturally running processes as *in vivo* experiments (Sulzman, 2007). Two main elements in study of plant nutrition, carbon and nitrogen, both occur as two stable isotopes: ^{12}C , ^{13}C and ^{14}N , ^{15}N . The light isotopes, ^{12}C and ^{14}N are much more prevalent in nature than their heavier counterparts, constituting up to, respectively, 98,9% and 99,6% of the element in the environment. These ratios, on global average, have been constant since formation of

Earth. However, they can be altered on a compartment scale by passing through biological processes that favor a particular isotope, e.g., due to differences in isotope fractionation by Rubisco and PEP carboxylase, plants running a C₃ or C₄ cycle exhibit distinct stable carbon isotope signatures and can be distinguished (Craine et al., 2009; Garten et al., 2007; Staddon, 2004). Similar analysis can be used to compare nitrogen content originating from soil compounds and atmospheric uptake by symbiotic bacteria. Therefore, analysis of isotopic saturation ratios is a reliable method to determine the biological source of elements.

Considering possible implications to be derived from such experiment, a notable discovery for study of mycorrhiza was proving, that fungal fruit bodies exhibit significant enrichment in heavy N and C isotopes compared to surrounding autotrophic plants (Gebauer and Dietrich, 1993; Gleixner et al., 1993). Similar enrichment was later described in non-photosynthesizing mycoheterotrophic Orchidaceae and Ericaceae (Gebauer and Meyer, 2003; Trudell et al., 2003). This aberration indicated that MH plants derive their nutrition from a different source than their photosynthesizing neighbors, and the most obvious enriched source were fungi. This dependency turned out to be common in plant-fungus relationships, and the degrees of enrichment are specific to different groups of partnering plants and fungi: EM fungi tend to exhibit higher saturation in heavy C and N isotopes than saprotrophs (Kohzu et al., 1999; Mayor et al., 2009; Taylor et al., 2003); species forming arbuscular mycorrhizas (AM) are more depleted in ¹⁵N than EcM species, and do not exhibit ¹³C enrichment compared to their partnering plants (Courty et al., 2011). Therefore, basing on the isotope food chain theory, these enrichment/depletion patterns can be expected to manifest in associated mycoheterotrophic plants, and can be used to derive scientific conclusions in both directions.

Not only can the analysis of C and N isotopic ratios be used to pinpoint a nutrition source in a binary manner. In cases of mixotrophic plants, it can be used to estimate the degree of dependency on fungal partners' nutritional support, or to track the destination organs of foreign nutrient transfer (Lallemand et al., 2019a). Carbon isotope ratio reflects on the equilibrium between photosynthetic and fungal carbon and is a clear indicator of plant's trophism; nitrogen isotope saturation is less obvious and cannot be directly used to conclude on plant's trophic mode, but nevertheless can indicate on fungal/non-fungal nitrogen contribution, if calibrated to a reference measurement of neighboring autotrophs.

10.10.2 Isotope ratio mass spectrometry

The differences in isotopic content between ecosystem compartments caused by thermodynamic and kinetic effects requires highly sensitive measurement methods to generate meaningful data – the variability of isotopic saturation ranges at a maximum by ~ 1‰ for C and by 0.2‰ for N. Detecting such minuscule differences with sufficient resolution is performed by **IRMS – isotope-ratio mass spectrometry**, considered a current golden standard in biological research (Benson et al., 2006). The process offers reduced flexibility compared to scanning mass spectrometry, not allowing for a wide mass range detection, but delivers a highly precise measurement of ratios for isotopes of expected molecular mass. The sample needs to be converted into simple gas – H₂ for ²H/¹H, N₂ for ¹⁵N/¹⁴N, CO₂ for ¹³C/¹²C, CO for ¹⁸O/¹⁶O and SO₂ for ³⁴S/³²S, usually achieved by high temperature conversion and separation by isothermal gas chromatography. The technique allows for a precise measurement of differences between isotope ratios rather than absolute ratios – therefore, a standard or reference gas is used to compensate for mass discrimination effects and negate the errors that may appear between measurements or devices. Two main approaches differ in a manner of feeding the sample and reference onto the mass spectrometer. In **continuous flow** measurement system, the measured sample is prepared and processed in the elemental analyzer directly before introduction into IRMS, which allows for a single measurement only, and calibration standard is measured before or after the sample. This approach offers high throughput and requires no off-the-line sample processing, at the cost of relatively poor internal precision. In a **dual inlet** measurement system, the gases are prepared, processed and purified before the measurement, and are alternately fed into the IRMS from their separate reservoirs ("bellows"). Though the internal and external precision of this approach is approximately 10x higher, the pre-processing requires extensive sample preparation and large sample volumes, greatly reducing its throughput and applicability for biological research on material collected *in situ* (Carter and Barwick, 2011; Meier-Augenstein, 1999). In presented work, a continuous flow analyzer (ThermoFinnigan DeltaV Advantage Continuous-Flow Isotope-ratio mass spectrometer coupled with elemental analyzer) was used to perform total N concentration and ¹³C/¹²C and ¹⁵N/¹⁴N ratio measurement.

11 Bibliography

This section does not include sources cited inside published articles (chapters 5.1-5.5) and unpublished data section (Chapter 8). To access these, please refer to reference sections of individual articles.

- Ali, M.S., Baek, K.H., 2020. Jasmonic acid signaling pathway in response to abiotic stresses in plants. *Int. J. Mol. Sci.* 21. <https://doi.org/10.3390/ijms21020621>
- Alm, E.W., Zheng, D., Raskin, L., 2000. The Presence of Humic Substances and DNA in RNA Extracts Affects Hybridization Results. *Appl. Environ. Microbiol.* 66, 4547–4554. <https://doi.org/10.1128/AEM.66.10.4547-4554.2000>
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., J., L.D., 1990. Altschul-1990-Basic Local Alignment.pdf. *J. Mol. Biol.* 215, 403–410.
- Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie, M.E., Gouil, Q., 2020. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* 21, 30. <https://doi.org/10.1186/s13059-020-1935-5>
- Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. <https://doi.org/10.1093/bioinformatics/btu638>
- Ankenbrand, M.J., Pfaff, S., Terhoeven, N., Qureischi, M., Gündel, M., L. Weiß, C., Hackl, T., Förster, F., 2018. chloroExtractor: extraction and assembly of the chloroplast genome from whole genome shotgun data. *J. Open Source Softw.* 3, 464. <https://doi.org/10.21105/joss.00464>
- 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., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. <https://doi.org/10.1038/75556>
- Badin, E.J., Calvin, M., 1950. The Path of Carbon in Photosynthesis. IX. Photosynthesis, Photoreduction and the Hydrogen—Oxygen—Carbon Dioxide Dark Reaction 1. *J. Am. Chem. Soc.* 72, 5266–5270. <https://doi.org/10.1021/ja01167a134>
- Badotti, F., de Oliveira, F.S., Garcia, C.F., Vaz, A.B.M., Fonseca, P.L.C., Nahum, L.A., Oliveira, G., Góes-Neto, A., 2017. Effectiveness of ITS and sub-regions as DNA barcode markers for the identification of Basidiomycota (Fungi). *BMC Microbiol.* 17, 42. <https://doi.org/10.1186/s12866-017-0958-x>
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. <https://doi.org/10.1089/cmb.2012.0021>

11. Bibliography

- Barrett, C.F., Davis, J.I., 2012. The plastid genome of the mycoheterotrophic *Corallorhiza striata* (Orchidaceae) is in the relatively early stages of degradation. *Am. J. Bot.* 99, 1513–1523. <https://doi.org/10.3732/ajb.1200256>
- Barrett, C.F., Freudenstein, J. V., Li, J., Mayfield-Jones, D.R., Perez, L., Pires, J.C., Santos, C., 2014. Investigating the Path of Plastid Genome Degradation in an Early-Transitional Clade of Heterotrophic Orchids, and Implications for Heterotrophic Angiosperms. *Mol. Biol. Evol.* 31, 3095–3112. <https://doi.org/10.1093/molbev/msu252>
- Barrett, C.F., Sinn, B.T., Kennedy, A.H., 2019. Unprecedented Parallel Photosynthetic Losses in a Heterotrophic Orchid Genus. *Mol. Biol. Evol.* 36, 1884–1901. <https://doi.org/10.1093/molbev/msz111>
- Batstone, R.T., Carscadden, K.A., Afkhami, M.E., Frederickson, M.E., 2018. Using niche breadth theory to explain generalization in mutualisms. *Ecology* 99, 1039–1050. <https://doi.org/10.1002/ecy.2188>
- Bekker, A., Holland, H.D., Wang, P.-L., Rumble, D., Stein, H.J., Hannah, J.L., Coetzee, L.L., Beukes, N.J., 2004. Dating the rise of atmospheric oxygen. *Nature* 427, 117–120. <https://doi.org/10.1038/nature02260>
- Benson, S., Lennard, C., Maynard, P., Roux, C., 2006. Forensic applications of isotope ratio mass spectrometry—A review. *Forensic Sci. Int.* 157, 1–22. <https://doi.org/10.1016/j.forsciint.2005.03.012>
- Bethune, K., Mariac, C., Couderc, M., Scarcelli, N., Santoni, S., Ardisson, M., Martin, J., Montúfar, R., Klein, V., Sabot, F., Vigouroux, Y., Couvreur, T.L.P., 2019. Long-fragment targeted capture for long-read sequencing of plastomes. *Appl. Plant Sci.* 7, e1243. <https://doi.org/10.1002/aps3.1243>
- Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K., Das, I., 2007. Cryptic species as a window on diversity and conservation. *Trends Ecol. Evol.* 22, 148–155. <https://doi.org/10.1016/j.tree.2006.11.004>
- Bidartondo, M.I., Burghardt, B., Gebauer, G., Bruns, T.D., Read, D.J., 2004. Changing partners in the dark: isotopic and molecular evidence of ectomycorrhizal liaisons between forest orchids and trees. *Proc. R. Soc. London. Ser. B Biol. Sci.* 271, 1799–1806. <https://doi.org/10.1098/rspb.2004.2807>
- Bidartondo, M.I., Read, D.J., 2008. Fungal specificity bottlenecks during orchid germination and development. *Mol. Ecol.* <https://doi.org/10.1111/j.1365-294X.2008.03848.x>
- Bjorkman, E., 1960. *Monotropa Hypopitys* L. - an Epiparasite on Tree Roots. *Physiol. Plant.* 13, 308–327. <https://doi.org/10.1111/j.1399-3054.1960.tb08034.x>
- Bock, R., 2017. Witnessing Genome Evolution: Experimental Reconstruction of Endosymbiotic and Horizontal Gene Transfer. *Annu. Rev. Genet.* 51, 1–22. <https://doi.org/10.1146/annurev-genet-120215-035329>
- Börner, T., Zhelyazkova, P., Legen, J., Schmitz-Linneweber, C., 2014. Chloroplast Gene Expression—RNA Synthesis and Processing, in: *Plastid Biology*. Springer New York, New York, NY, pp. 3–47. https://doi.org/10.1007/978-1-4939-1136-3_1

11. Bibliography

- Braukmann, T., Kuzmina, M., Stefanović, S., 2013. Plastid genome evolution across the genus *Cuscuta* (Convolvulaceae): Two clades within subgenus *Grammica* exhibit extensive gene loss. *J. Exp. Bot.* 64, 977–989. <https://doi.org/10.1093/jxb/ers391>
- Bräutigam, A., Gowik, U., 2010. What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. *Plant Biol.* 12, 831–841. <https://doi.org/10.1111/j.1438-8677.2010.00373.x>
- Brundrett, M.C., 2009. Mycorrhizal associations and other means of nutrition of vascular plants: understanding the global diversity of host plants by resolving conflicting information and developing reliable means of diagnosis. *Plant Soil* 320, 37–77. <https://doi.org/10.1007/s11104-008-9877-9>
- Buchanan, B.B., Gruissem, W., Jones, R.L., 2015. *Biochemistry and Molecular Biology of Plants*, 2nd Edition.
- Buchfink, B., Reuter, K., Drost, H.G., 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* 18, 366–368. <https://doi.org/10.1038/s41592-021-01101-x>
- Buick, R., 2008. When did oxygenic photosynthesis evolve? *Philos. Trans. R. Soc. B Biol. Sci.* 363, 2731–2743. <https://doi.org/10.1098/rstb.2008.0041>
- Bushnell, B., 2014. BBMap: A Fast, Accurate, Splice-Aware Aligner, in: 9th Annual Genomics of Energy & Environment Meeting, Walnut Creek, CA, March 17-20, 2014.
- Cannon, T., 2018. *Genetics and Genetic Engineering*. Scientific e-Resources.
- Carbon, S., Douglass, E., Good, B.M., Unni, D.R., Harris, N.L., Mungall, C.J., Basu, S., Chisholm, R.L., Dodson, R.J., Hartline, E., Fey, P., Thomas, P.D., Albou, L.-P., Ebert, D., Kesling, M.J., Mi, H., Muruganujan, A., Huang, X., Mushayahama, T., LaBonte, S.A., Siegele, D.A., Antonazzo, G., Attrill, H., Brown, N.H., Garapati, P., Marygold, S.J., Trovisco, V., dos Santos, G., Falls, K., Tabone, C., Zhou, P., Goodman, J.L., Strelets, V.B., Thurmond, J., Garmiri, P., Ishtiaq, R., Rodríguez-López, M., Acencio, M.L., Kuiper, M., Lægreid, A., Logie, C., Lovering, R.C., Kramarz, B., Saverimuttu, S.C.C., Pinheiro, S.M., Gunn, H., Su, R., Thurlow, K.E., Chibucos, M., Giglio, M., Nadendla, S., Munro, J., Jackson, R., Duesbury, M.J., Del-Toro, N., Meldal, B.H.M., Paneerselvam, K., Perfetto, L., Porras, P., Orchard, S., Shrivastava, A., Chang, H.-Y., Finn, R.D., Mitchell, A.L., Rawlings, N.D., Richardson, L., Sangrador-Vegas, A., Blake, J.A., Christie, K.R., Dolan, M.E., Drabkin, H.J., Hill, D.P., Ni, L., Sitnikov, D.M., Harris, M.A., Oliver, S.G., Rutherford, K., Wood, V., Hayles, J., Bähler, J., Bolton, E.R., De Pons, J.L., Dwinell, M.R., Hayman, G.T., Kaldunski, M.L., Kwitek, A.E., Laulederkind, S.J.F., Plasterer, C., Tutaj, M.A., Vedi, M., Wang, S.-J., D’Eustachio, P., Matthews, L., Balhoff, J.P., Aleksander, S.A., Alexander, M.J., Cherry, J.M., Engel, S.R., Gondwe, F., Karra, K., Miyasato, S.R., Nash, R.S., Simison, M., Skrzypek, M.S., Weng, S., Wong, E.D., Feuermann, M., Gaudet, P., Morgat, A., Bakker, E., Berardini, T.Z., Reiser, L., Subramaniam, S., Huala, E., Arighi, C.N., Auchincloss, A., Axelsen, K., Argoud-Puy, G., Bateman, A., Blatter, M.-C., Boutet, E., Bowler, E., Breuza, L., Bridge, A., Britto, R., Bye-A-Jee, H., Casas, C.C., Coudert, E., Denny, P., Estreicher, A., Famiglietti, M.L., Georghiou, G., Gos, A., Gruaz-Gumowski, N., Hatton-Ellis, E., Hulo, C., Ignatchenko, A., Jungo, F., Laiho, K., Le Mercier, P., Lieberherr, D., Lock, A., Lussi, Y., MacDougall, A., Magrane, M., Martin, M.J., Masson, P., Natale, D.A., Hyka-Nouspikel, N., Orchard, S., Pedruzzi, I., Pourcel, L., Poux, S., Pundir, S., Rivoire, C., Speretta, E., Sundaram, S., Tyagi, N., Warner, K., Zaru, R., Wu, C.H., Diehl, A.D., Chan, J.N., Grove,

11. Bibliography

- C., Lee, R.Y.N., Muller, H.-M., Raciti, D., Van Auken, K., Sternberg, P.W., Berriman, M., Paulini, M., Howe, K., Gao, S., Wright, A., Stein, L., Howe, D.G., Toro, S., Westerfield, M., Jaiswal, P., Cooper, L., Elser, J., 2021. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* 49, D325–D334. <https://doi.org/10.1093/nar/gkaa1113>
- Carter, J., Barwick, V., 2011. Good practice guide for isotope ratio Mass Spectrometry, Firms.
- Chaisson, M.J., Pevzner, P.A., 2008. Short read fragment assembly of bacterial genomes. *Genome Res.* 18, 324–330. <https://doi.org/10.1101/gr.7088808>
- Chase, M.W., Cameron, K.M., Freudenstein, J. V., Pridgeon, A.M., Salazar, G., van den Berg, C., Schuiteman, A., 2015. An updated classification of Orchidaceae. *Bot. J. Linn. Soc.* 177, 151–174. <https://doi.org/10.1111/boj.12234>
- Chen, C., MacCready, J.S., Ducat, D.C., Osteryoung, K.W., 2018. The Molecular Machinery of Chloroplast Division. *Plant Physiol.* 176, 138–151. <https://doi.org/10.1104/pp.17.01272>
- Chung, M., Bruno, V.M., Rasko, D.A., Cuomo, C.A., Muñoz, J.F., Livny, J., Shetty, A.C., Mahurkar, A., Dunning Hotopp, J.C., 2021. Best practices on the differential expression analysis of multi-species RNA-seq. *Genome Biol.* 22, 121. <https://doi.org/10.1186/s13059-021-02337-8>
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., Mortazavi, A., 2016. A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17, 1–19. <https://doi.org/10.1186/s13059-016-0881-8>
- Courty, P., Walder, F., Boller, T., Ineichen, K., Wiemken, A., 2011. Carbon and Nitrogen Metabolism in Mycorrhizal Networks and Mycoheterotrophic Plants of Tropical Forests : A Stable Isotope Analysis 1 [W] 156, 952–961. <https://doi.org/10.1104/pp.111.177618>
- Craine, J.M., Elmore, A.J., Aidar, M.P.M., Bustamante, M., Dawson, T.E., Hobbie, E.A., Kahmen, A., Mack, M.C., McLauchlan, K.K., Michelsen, A., Nardoto, G.B., Pardo, L.H., Peñuelas, J., Reich, P.B., Schuur, E.A.G., Stock, W.D., Templer, P.H., Virginia, R.A., Welker, J.M., Wright, I.J., 2009. Global patterns of foliar nitrogen isotopes and their relationships with climate, mycorrhizal fungi, foliar nutrient concentrations, and nitrogen availability. *New Phytol.* 183, 980–992. <https://doi.org/10.1111/j.1469-8137.2009.02917.x>
- Daniell, H., Lin, C.-S., Yu, M., Chang, W.-J., 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17, 134. <https://doi.org/10.1186/s13059-016-1004-2>
- Darling, A.C.E., 2004. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Res.* 14, 1394–1403. <https://doi.org/10.1101/gr.2289704>
- Dearnaley, J.D.W., Martos, F., Selosse, M.-A., 2012. 12 Orchid Mycorrhizas: Molecular Ecology, Physiology, Evolution and Conservation Aspects, in: *Fungal Associations*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 207–230. https://doi.org/10.1007/978-3-642-30826-0_12
- Delannoy, E., Fujii, S., Colas des Francs-Small, C., Brundrett, M., Small, I., 2011. Rampant Gene loss in the underground orchid *Rhizanthella Gardneri* highlights evolutionary constraints on plastid genomes. *Mol. Biol. Evol.* 28, 2077–2086. <https://doi.org/10.1093/molbev/msr028>

11. Bibliography

- Desirò, A., Duckett, J.G., Pressel, S., Villarreal, J.C., Bidartondo, M.I., 2013. Fungal symbioses in hornworts: a chequered history. *Proc. R. Soc. B Biol. Sci.* 280, 20130207. <https://doi.org/10.1098/rspb.2013.0207>
- Dierckxsens, N., Mardulyn, P., Smits, G., 2016. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* gkw955. <https://doi.org/10.1093/nar/gkw955>
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Dong, W., Xu, C., Cheng, T., Lin, K., Zhou, S., 2013. Sequencing Angiosperm Plastid Genomes Made Easy: A Complete Set of Universal Primers and a Case Study on the Phylogeny of Saxifragales. *Genome Biol. Evol.* 5, 989–997. <https://doi.org/10.1093/gbe/evt063>
- Eckardt, N.A., 2006. Genomic hopscotch: Gene transfer from plastid to nucleus. *Plant Cell* 18, 2865–2867. <https://doi.org/10.1105/tpc.106.049031>
- Ekblom, R., Wolf, J.B.W., 2014. A field guide to whole-genome sequencing, assembly and annotation. *Evol. Appl.* 7, 1026–1042. <https://doi.org/10.1111/eva.12178>
- Fay, M.F., Chase, M.W., 2009. Orchid biology: From Linnaeus via Darwin to the 21st century. *Ann. Bot.* 104, 359–364. <https://doi.org/10.1093/aob/mcp190>
- Feijen, F.A.A., Vos, R.A., Nuytinck, J., Merckx, V.S.F.T., 2018. Evolutionary dynamics of mycorrhizal symbiosis in land plant diversification. *Sci. Rep.* 8, 1–7. <https://doi.org/10.1038/s41598-018-28920-x>
- Feng, Y.-L., Wicke, S., Li, J.-W., Han, Y., Lin, C.-S., Li, D.-Z., Zhou, T.-T., Huang, W.-C., Huang, L.-Q., Jin, X.-H., 2016. Lineage-Specific Reductions of Plastid Genomes in an Orchid Tribe with Partially and Fully Mycoheterotrophic Species. *Genome Biol. Evol.* 8, 2164–2175. <https://doi.org/10.1093/gbe/evw144>
- Feuillet, C., Leach, J.E., Rogers, J., Schnable, P.S., Eversole, K., 2011. Crop genome sequencing: lessons and rationales. *Trends Plant Sci.* 16, 77–88. <https://doi.org/10.1016/j.tplants.2010.10.005>
- Field, C.B., Behrenfeld, M.J., Randerson, J.T., Falkowski, P., 1998. Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* (80-.). 281, 237–240. <https://doi.org/10.1126/science.281.5374.237>
- Fragnire, C., Serrano, M., Abou-Mansour, E., Métraux, J.P., L'Haridon, F., 2011. Salicylic acid and its location in response to biotic and abiotic stress. *FEBS Lett.* 585, 1847–1852. <https://doi.org/10.1016/j.febslet.2011.04.039>
- Frank, B., 2005. On the nutritional dependence of certain trees on root symbiosis with belowground fungi (an English translation of A.B. Frank's classic paper of 1885). *Mycorrhiza* 15, 267–275. <https://doi.org/10.1007/s00572-004-0329-y>
- Freudenstein, J. V., Chase, M.W., 2015. Phylogenetic relationships in Epidendroideae (Orchidaceae), one of the great flowering plant radiations: progressive specialization and diversification. *Ann. Bot.* 115, 665–681. <https://doi.org/10.1093/aob/mcu253>

11. Bibliography

- Freudenthal, J.A., Pfaff, S., Terhoeven, N., Korte, A., Ankenbrand, M.J., Förster, F., 2020. A systematic comparison of chloroplast genome assembly tools. *Genome Biol.* 21, 1–21. <https://doi.org/10.1186/s13059-020-02153-6>
- Garg, R., Jain, M., 2013. RNA-Seq for transcriptome analysis in non-model plants. *Methods Mol. Biol.* 1069, 43–58. https://doi.org/10.1007/978-1-62703-613-9_4
- Garten, C.T., Hanson, P.J., Todd, D.E., Lu, B.B., Brice, D.J., 2007. Natural¹⁵N- and¹³C-Abundance as Indicators of Forest Nitrogen Status and Soil Carbon Dynamics, in: Michener, R., Lajtha, K. (Eds.), *Stable Isotopes in Ecology and Environmental Science*. Blackwell Publishing Ltd, Oxford, UK, pp. 61–82. <https://doi.org/10.1002/9780470691854.ch3>
- Gebauer, G., Dietrich, P., 1993. Nitrogen Isotope Ratios in Different Compartments of a Mixed Stand of Spruce, Larch and Beech Trees and of Understorey Vegetation Including Fungi. *Isot. Environ. Heal. Stud.* 29, 35–44. <https://doi.org/10.1080/10256019308046133>
- Gebauer, G., Meyer, M., 2003. ¹⁵N and ¹³C natural abundance of autotrophic and myco-heterotrophic orchids provides insight into nitrogen and carbon gain from fungal association. *New Phytol.* 160, 209–223. <https://doi.org/10.1046/j.1469-8137.2003.00872.x>
- Gebauer, G., Preiss, K., Gebauer, A.C., 2016. Partial mycoheterotrophy is more widespread among orchids than previously assumed. *New Phytol.* 211, 11–15. <https://doi.org/10.1111/nph.13865>
- Gleixner, G., Danier, H.J., Werner, R.A., Schmidt, H.L., 1993. Correlations between the ¹³C Content of Primary and Secondary Plant Products in Different Cell Compartments and That in Decomposing Basidiomycetes. *Plant Physiol.* 102, 1287–1290. <https://doi.org/10.1104/pp.102.4.1287>
- Gonneau, C., Jersáková, J., de Tredern, E., Till-Bottraud, I., Saarinen, K., Sauve, M., Roy, M., Hájek, T., Selosse, M.-A., 2014. Photosynthesis in perennial mixotrophic *Epipactis* spp. (Orchidaceae) contributes more to shoot and fruit biomass than to hypogeous survival. *J. Ecol.* 102, 1183–1194. <https://doi.org/10.1111/1365-2745.12274>
- Goodwin, S., McPherson, J.D., McCombie, W.R., 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17, 333–351. <https://doi.org/10.1038/nrg.2016.49>
- Górnaiak, M., Paun, O., Chase, M.W., 2010. Phylogenetic relationships within Orchidaceae based on a low-copy nuclear coding gene, *Xdh*: Congruence with organellar and nuclear ribosomal DNA results. *Mol. Phylogenet. Evol.* 56, 784–795. <https://doi.org/10.1016/j.ympev.2010.03.003>
- Grabherr, M.G., Brian J. Haas, Moran Yassour Joshua Z. Levin, Dawn A. Thompson, Ido Amit, Xian Adiconis, Lin Fan, Raktima Raychowdhury, Qiandong Zeng, Zehua Chen, Evan Mauceli, Nir Hacohen, Andreas Gnirke, Nicholas Rhind, Federica di Palma, Bruce W. N., Friedman, and A.R., 2013. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* 29, 644–652. <https://doi.org/10.1038/nbt.1883>
- Graham, S.W., Lam, V.K.Y., Merckx, V.S.F.T., 2017. Plastomes on the edge: the evolutionary breakdown of mycoheterotroph plastid genomes. *New Phytol.* 214, 48–55. <https://doi.org/10.1111/nph.14398>
- Guerrero-Sanchez, V.M., Maldonado-Alconada, A.M., Amil-Ruiz, F., Verardi, A., Jorrín-Novo, J. V., Rey, M.-D., 2019. Ion Torrent and Illumina, two complementary RNA-seq platforms for

11. Bibliography

- constructing the holm oak (*Quercus ilex*) transcriptome. *PLoS One* 14, e0210356. <https://doi.org/10.1371/journal.pone.0210356>
- Hadariová, L., Vesteg, M., Hampl, V., Krajčovič, J., 2018. Reductive evolution of chloroplasts in non-photosynthetic plants, algae and protists. *Curr. Genet.* 64, 365–387. <https://doi.org/10.1007/s00294-017-0761-0>
- Hawksworth, D.L., Lücking, R., 2017. Fungal Diversity Revisited: 2.2 to 3.8 Million Species. *Microbiol. Spectr.* 5. <https://doi.org/10.1128/microbiolspec.FUNK-0052-2016>
- Herbert, Z.T., Kershner, J.P., Butty, V.L., Thimmapuram, J., Choudhari, S., Alekseyev, Y.O., Fan, J., Podnar, J.W., Wilcox, E., Gipson, J., Gillaspay, A., Jepsen, K., BonDurant, S.S., Morris, K., Berkeley, M., LeClerc, A., Simpson, S.D., Sommerville, G., Grimmett, L., Adams, M., Levine, S.S., 2018. Cross-site comparison of ribosomal depletion kits for Illumina RNAseq library construction. *BMC Genomics* 19, 1–10. <https://doi.org/10.1186/s12864-018-4585-1>
- Hernandez, D., Francois, P., Farinelli, L., Osteras, M., Schrenzel, J., 2008. De novo bacterial genome sequencing: Millions of very short reads assembled on a desktop computer. *Genome Res.* 18, 802–809. <https://doi.org/10.1101/gr.072033.107>
- Hirsch, C.N., Robin Buell, C., 2013. Tapping the Promise of Genomics in Species with Complex, Nonmodel Genomes. *Annu. Rev. Plant Biol.* 64, 89–110. <https://doi.org/10.1146/annurev-arplant-050312-120237>
- Hohmann-Marriott, M.F., Blankenship, R.E., 2011. Evolution of Photosynthesis. *Annu. Rev. Plant Biol.* 62, 515–548. <https://doi.org/10.1146/annurev-arplant-042110-103811>
- Hölzl, G., Dörmann, P., 2019. Chloroplast Lipids and Their Biosynthesis. *Annu. Rev. Plant Biol.* 70, 51–81. <https://doi.org/10.1146/annurev-arplant-050718-100202>
- Honaas, L.A., Wafula, E.K., Wickett, N.J., Der, J.P., Zhang, Y., Edger, P.P., Altman, N.S., Pires, J.C., Leebens-Mack, J.H., DePamphilis, C.W., 2016. Selecting Superior De Novo Transcriptome Assemblies: Lessons Learned by Leveraging the Best Plant Genome. *PLoS One* 11, e0146062. <https://doi.org/10.1371/journal.pone.0146062>
- Huang, C.Y., Ayliffe, M.A., Timmis, J.N., 2004. Simple and complex nuclear loci created by newly transferred chloroplast DNA in tobacco. *Proc. Natl. Acad. Sci.* 101, 9710–9715. <https://doi.org/10.1073/pnas.0400853101>
- Hynson, N.A., Madsen, T.P., Selosse, M.-A., Adam, I.K.U., Ogura-Tsujita, Y., Roy, M., Gebauer, G., 2013. The Physiological Ecology of Mycoheterotrophy, in: *Mycoheterotrophy*. Springer New York, New York, NY, pp. 297–342. https://doi.org/10.1007/978-1-4614-5209-6_8
- Jacquemyn, H., Brys, R., Merckx, V.S.F.T., Waud, M., Lievens, B., Wiegand, T., 2014. Coexisting orchid species have distinct mycorrhizal communities and display strong spatial segregation. *New Phytol.* 202, 616–627. <https://doi.org/10.1111/nph.12640>
- Jacquemyn, H., Brys, R., Waud, M., Busschaert, P., Lievens, B., 2015. Mycorrhizal networks and coexistence in species-rich orchid communities. *New Phytol.* 206, 1127–1134. <https://doi.org/10.1111/nph.13281>
- Jacquemyn, H., Merckx, V.S.F.T., 2019. Mycorrhizal symbioses and the evolution of trophic modes in plants. *J. Ecol.* 107, 1567–1581. <https://doi.org/10.1111/1365-2745.13165>

11. Bibliography

- Jacquemyn, H., Waud, M., Brys, R., Lallemand, F., Courty, P.-E., Robioneck, A., Selosse, M.-A., 2017. Mycorrhizal Associations and Trophic Modes in Coexisting Orchids: An Ecological Continuum between Auto- and Mixotrophy. *Front. Plant Sci.* 8. <https://doi.org/10.3389/fpls.2017.01497>
- Jąkowski, M., Minasiewicz, J., Caius, J., May, M., Selosse, M.A., Delannoy, E., 2021. **The Genomic Impact of Mycoheterotrophy in Orchids.** *Front. Plant Sci.* 12, 1–16. <https://doi.org/10.3389/fpls.2021.632033>
- Jasid, S., Simontacchi, M., Bartoli, C.G., Puntarulo, S., 2006. Chloroplasts as a nitric oxide cellular source. Effect of reactive nitrogen species on chloroplastic lipids and proteins. *Plant Physiol.* 142, 1246–1255. <https://doi.org/10.1104/pp.106.086918>
- Jensen, P.E., Leister, D., 2014. Chloroplast evolution, structure and functions. *F1000Prime Rep.* 6, 1–14. <https://doi.org/10.12703/P6-40>
- Jin, X.-H., Ren, Z.-X., Xu, S.-Z., Wang, H., Li, D.-Z., Li, Z.-Y., 2014. The evolution of floral deception in *Epipactis veratrifolia* (Orchidaceae): from indirect defense to pollination. *BMC Plant Biol.* 14, 63. <https://doi.org/10.1186/1471-2229-14-63>
- Julou, T., Burghardt, B., Gebauer, G., Berveiller, D., Damesin, C., Selosse, M.A., 2005. Mixotrophy in orchids: Insights from a comparative study of green individuals and nonphotosynthetic individuals of *Cephalanthera damasonium*. *New Phytol.* 166, 639–653. <https://doi.org/10.1111/j.1469-8137.2005.01364.x>
- Kamjunke, N., Tittel, J., 2009. MIXOTROPHIC ALGAE CONSTRAIN THE LOSS OF ORGANIC CARBON BY EXUDATION1. *J. Phycol.* 45, 807–811. <https://doi.org/10.1111/j.1529-8817.2009.00707.x>
- Kanehisa, M., Sato, Y., 2020. KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci.* 29, 28–35. <https://doi.org/10.1002/pro.3711>
- Kazutaka, K., Kazuharu, M., Kuma, K., Takashi, M., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. <https://doi.org/10.1093/nar/gkf436>
- Kim, D., Paggi, J.M., Park, C., Bennett, C., Salzberg, S.L., 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. <https://doi.org/10.1038/s41587-019-0201-4>
- Kim, H.-M., Jeon, S., Chung, O., Jun, J.H., Kim, H.-S., Blazyte, A., Lee, H.-Y., Yu, Y., Cho, Y.S., Bolser, D.M., Bhak, J., 2021. Comparative analysis of 7 short-read sequencing platforms using the Korean Reference Genome: MGI and Illumina sequencing benchmark for whole-genome sequencing. *Gigascience* 10. <https://doi.org/10.1093/gigascience/giab014>
- Kim, H.T., Kim, J.S., Moore, M.J., Neubig, K.M., Williams, N.H., Whitten, W.M., Kim, J.-H., 2015. Seven New Complete Plastome Sequences Reveal Rampant Independent Loss of the *ndh* Gene Family across Orchids and Associated Instability of the Inverted Repeat/Small Single-Copy Region Boundaries. *PLoS One* 10, e0142215. <https://doi.org/10.1371/journal.pone.0142215>
- Kim, Y.-K., Jo, S., Cheon, S.-H., Joo, M.-J., Hong, J.-R., Kwak, M., Kim, K.-J., 2020. Plastome Evolution and Phylogeny of Orchidaceae, With 24 New Sequences. *Front. Plant Sci.* 11, 1–27. <https://doi.org/10.3389/fpls.2020.00022>

11. Bibliography

- King, C., Scott-Horton, T., 2008. Pyrosequencing: A Simple Method for Accurate Genotyping. *J. Vis. Exp.* <https://doi.org/10.3791/630>
- Kivlin, S.N., Hawkes, C. V., Treseder, K.K., 2011. Global diversity and distribution of arbuscular mycorrhizal fungi. *Soil Biol. Biochem.* 43, 2294–2303. <https://doi.org/10.1016/j.soilbio.2011.07.012>
- Kohzu, A., Yohsioka, T., Ando, T., Takahashi, M., Koba, K., Wada, E., 1999. Natural ¹³C and ¹⁵N abundance of field-collected fungi and their ecological implications. *New Phytol.* 144, 323–330. <https://doi.org/10.1046/j.1469-8137.1999.00508.x>
- Kovaka, S., Zimin, A. V., Perteza, G.M., Razaghi, R., Salzberg, S.L., Perteza, M., 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *bioRxiv* 1–13. <https://doi.org/10.1101/694554>
- Kulaeva, O.A., Afonin, A.M., Zhernakov, A.I., Tikhonovich, I.A., Zhukov, V.A., 2017. Transcriptomic Studies in Non-Model Plants: Case of *Pisum sativum* L. and *Medicago lupulina* L. *Appl. RNA-Seq Omi. Strateg. - From Microorg. to Hum. Heal.* <https://doi.org/10.5772/intechopen.69057>
- Kulski, J.K., 2016. Next-Generation Sequencing — An Overview of the History, Tools, and “Omic” Applications, in: *Next Generation Sequencing - Advances, Applications and Challenges*. InTech. <https://doi.org/10.5772/61964>
- Kwon, E.-C., Kim, J.-H., Kim, N.-S., 2020. Comprehensive genomic analyses with 115 plastomes from algae to seed plants: structure, gene contents, GC contents, and introns. *Genes Genomics* 42, 553–570. <https://doi.org/10.1007/s13258-020-00923-x>
- Lahens, N.F., Ricciotti, E., Smirnova, O., Toorens, E., Kim, E.J., Baruzzo, G., Hayer, K.E., Ganguly, T., Schug, J., Grant, G.R., 2017. A comparison of Illumina and Ion Torrent sequencing platforms in the context of differential gene expression. *BMC Genomics* 18, 602. <https://doi.org/10.1186/s12864-017-4011-0>
- Lallemand, F., Figura, T., Damesin, C., Fresneau, C., Griveau, C., Fontaine, N., Zeller, B., Selosse, M.A., 2019a. Mixotrophic orchids do not use photosynthates for perennial underground organs. *New Phytol.* 221, 12–17. <https://doi.org/10.1111/nph.15443>
- Lallemand, F., Logacheva, M., Le Clainche, I., Bérard, A., Zheleznaia, E., May, M., Jakalski, M., Delannoy, É., Le Paslier, M.-C., Selosse, M.-A., 2019b. **Thirteen New Plastid Genomes from Mixotrophic and Autotrophic Species Provide Insights into Heterotrophy Evolution in Neottiae Orchids.** *Genome Biol. Evol.* 11, 2457–2467. <https://doi.org/10.1093/gbe/evz170>
- Lallemand, F., Martin-Magniette, M., Gilard, F., Gakière, B., Launay-Avon, A., Delannoy, É., Selosse, M., 2019c. In situ transcriptomic and metabolomic study of the loss of photosynthesis in the leaves of mixotrophic plants exploiting fungi. *Plant J.* 98, 826–841. <https://doi.org/10.1111/tpj.14276>
- Lallemand, F., May, M., Ichnatowicz, A., Jakalski, M., 2019d. **The complete chloroplast genome sequence of *Platanthera chlorantha* (Orchidaceae).** *Mitochondrial DNA Part B* 4, 2683–2684. <https://doi.org/10.1080/23802359.2019.1644551>

11. Bibliography

- Lambers, H., Teste, F.P., 2013. Interactions between arbuscular mycorrhizal and non-mycorrhizal plants: do non-mycorrhizal species at both extremes of nutrient availability play the same game? *Plant. Cell Environ.* n/a-n/a. <https://doi.org/10.1111/pce.12117>
- Lancien, M., Lea, P.J., Azevedo, R.A., 2007. Amino Acid Synthesis in Plastids. pp. 355–385. https://doi.org/10.1007/978-1-4020-4061-0_18
- Leake, J.R., 1994. The biology of myco - heterotrophic (‘saprophytic’) plants. *New Phytol.* 127, 171 – 216. <https://doi.org/10.1111/j.1469-8137.1994.tb04272.x>
- Leister, D., 2005. Origin, evolution and genetic effects of nuclear insertions of organelle DNA. *Trends Genet.* 21, 655–663. <https://doi.org/10.1016/j.tig.2005.09.004>
- Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323. <https://doi.org/10.1186/1471-2105-12-323>
- Li, X.-Q., Du, D., 2014. Variation, Evolution, and Correlation Analysis of C+G Content and Genome or Chromosome Size in Different Kingdoms and Phyla. *PLoS One* 9, e88339. <https://doi.org/10.1371/journal.pone.0088339>
- Li, X., Kong, Y., Zhao, Q.Y., Li, Y.Y., Hao, P., 2016. De novo assembly of transcriptome from next-generation sequencing data. *Quant. Biol.* 4, 94–105. <https://doi.org/10.1007/s40484-016-0069-y>
- Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. <https://doi.org/10.1093/bioinformatics/btt656>
- Ligrone, R., Carafa, A., Lumini, E., Bianciotto, V., Bonfante, P., Duckett, J.G., 2007. Glomeromycotean associations in liverworts: a molecular, cellular, and taxonomic analysis. *Am. J. Bot.* 94, 1756–1777. <https://doi.org/10.3732/ajb.94.11.1756>
- Love, M.I., Anders, S., Kim, V., Huber, W., 2016. RNA-Seq workflow: Gene-level exploratory analysis and differential expression. *F1000Research* 4. <https://doi.org/10.12688/f1000research.7035.2>
- Luo, C., Tsementzi, D., Kyrpides, N., Read, T., Konstantinidis, K.T., 2012. Direct Comparisons of Illumina vs. Roche 454 Sequencing Technologies on the Same Microbial Community DNA Sample. *PLoS One* 7, e30087. <https://doi.org/10.1371/journal.pone.0030087>
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Yunjie, Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Yong, Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.-M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, Jian, Lam, T.-W., Wang, Jun, 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1, 18. <https://doi.org/10.1186/2047-217X-1-18>
- Maréchal, E., 2018. Primary Endosymbiosis: Emergence of the Primary Chloroplast and the Chromatophore, Two Independent Events. pp. 3–16. https://doi.org/10.1007/978-1-4939-8654-5_1
- Martin, W., Goremykin, V., Hansmann, S., 1998. Gene transfer to the nucleus and the evolution of chloroplasts. *Curr. Opin. Plant Biol.* 1, 276. [https://doi.org/10.1016/1369-5266\(88\)80011-6](https://doi.org/10.1016/1369-5266(88)80011-6)

11. Bibliography

- Martin, W., Kowallik, K., 1999. Annotated english translation of mereschkowsky's 1905 paper 'Über natur und ursprung der chromatophoren impflanzenreiche.' Eur. J. Phycol. 34, 287–295. <https://doi.org/10.1080/09670269910001736342>
- Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., Penny, D., 2002. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc. Natl. Acad. Sci. 99, 12246–12251. <https://doi.org/10.1073/pnas.182432999>
- Matsuo, M., Ito, Y., Yamauchi, R., Obokata, J., 2005. The Rice Nuclear Genome Continuously Integrates, Shuffles, and Eliminates the Chloroplast Genome to Cause Chloroplast–Nuclear DNA Flux. Plant Cell 17, 665–675. <https://doi.org/10.1105/tpc.104.027706>
- Maxam, A.M., Gilbert, W., 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. 74, 560–564. <https://doi.org/10.1073/pnas.74.2.560>
- May, M., Jąkowski, M., Novotná, A., Dietel, J., Ayasse, M., Lallemand, F., Figura, T., Minasiewicz, J., Selosse, M.-A., 2020. **Three-year pot culture of *Epipactis helleborine* reveals autotrophic survival, without mycorrhizal networks, in a mixotrophic species.** Mycorrhiza 30, 51–61. <https://doi.org/10.1007/s00572-020-00932-4>
- May, M., Novotná, A., Minasiewicz, J., Selosse, M.-A., Jąkowski, M., 2019. **The complete chloroplast genome sequence of *Dactylorhiza majalis* (Rchb.) P.F. Hunt et Summerh. (Orchidaceae).** Mitochondrial DNA Part B 4, 2821–2823. <https://doi.org/10.1080/23802359.2019.1660282>
- Mayor, J.R., Schuur, E.A.G., Henkel, T.W., 2009. Elucidating the nutritional dynamics of fungi using stable isotopes. Ecol. Lett. 12, 171–183. <https://doi.org/10.1111/j.1461-0248.2008.01265.x>
- McCormick, M.K., Whigham, D.F., Sloan, D., O'Malley, K., Hodkinson, B., 2006. Orchid-fungus fidelity: A marriage meant to last? Ecology 87, 903–911. [https://doi.org/10.1890/0012-9658\(2006\)87](https://doi.org/10.1890/0012-9658(2006)87)
- McKendrick, S.L., Leake, J.R., Read, D.J., 2000. Symbiotic germination and development of myco-heterotrophic plants in nature: transfer of carbon from ectomycorrhizal *Salix repens* and *Betula pendula* to the orchid *Corallorhiza trifida* through shared hyphal connections. New Phytol. 145, 539–548. <https://doi.org/10.1046/j.1469-8137.2000.00592.x>
- Meier-Augenstein, W., 1999. Applied gas chromatography coupled to isotope ratio mass spectrometry. J. Chromatogr. A 842, 351–371. [https://doi.org/10.1016/S0021-9673\(98\)01057-7](https://doi.org/10.1016/S0021-9673(98)01057-7)
- Mereschkowsky, C., 1905. Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. Biol. Cent. 593–604.
- Meurer, J., Plücker, H., Kowallik, K. V., Westhoff, P., 1998. A nuclear-encoded protein of prokaryotic origin is essential for the stability of photosystem II in *Arabidopsis thaliana*. EMBO J. 17, 5286–5297. <https://doi.org/10.1093/emboj/17.18.5286>
- Miflin, B.J., Beevers, H., 1974. Isolation of Intact Plastids from a Range of Plant Tissues. Plant Physiol. 53, 870–874. <https://doi.org/10.1104/pp.53.6.870>

11. Bibliography

- Moore, K.R., Magnabosco, C., Momper, L., Gold, D.A., Bosak, T., Fournier, G.P., 2019. An expanded ribosomal phylogeny of cyanobacteria supports a deep placement of plastids. *Front. Microbiol.* 10, 1–14. <https://doi.org/10.3389/fmicb.2019.01612>
- Moore, L.R., 2013. More mixotrophy in the marine microbial mix. *Proc. Natl. Acad. Sci.* 110, 8323–8324. <https://doi.org/10.1073/pnas.1305998110>
- Motomura, H., Selosse, M.-A., Martos, F., Kagawa, A., Yukawa, T., 2010. Mycoheterotrophy evolved from mixotrophic ancestors: evidence in *Cymbidium* (Orchidaceae). *Ann. Bot.* 106, 573–581. <https://doi.org/10.1093/aob/mcq156>
- Nevo, R., Chuartzman, S.G., Tsabari, O., Reich, Z., Charuvi, D., Shimoni, E., 2009. Architecture of Thylakoid Membrane Networks. pp. 295–328. https://doi.org/10.1007/978-90-481-2863-1_14
- Ogura-Tsujita, Y., Yukawa, T., 2008. *Epipactis helleborine* shows strong mycorrhizal preference towards ectomycorrhizal fungi with contrasting geographic distributions in Japan. *Mycorrhiza* 18, 331–338. <https://doi.org/10.1007/s00572-008-0187-0>
- Ogura-Tsujita, Y., Yukawa, T., Kinoshita, A., 2021. Evolutionary histories and mycorrhizal associations of mycoheterotrophic plants dependent on saprotrophic fungi. *J. Plant Res.* 134, 19–41. <https://doi.org/10.1007/s10265-020-01244-6>
- Oja, J., Kohout, P., Tedersoo, L., Kull, T., Kõljalg, U., 2015. Temporal patterns of orchid mycorrhizal fungi in meadows and forests as revealed by 454 pyrosequencing. *New Phytol.* 205, 1608–1618. <https://doi.org/10.1111/nph.13223>
- Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C., 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. <https://doi.org/10.1038/nmeth.4197>
- Patro, R., Mount, S.M., Kingsford, C., 2014. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat. Biotechnol.* 32, 462–464. <https://doi.org/10.1038/nbt.2862>
- Payne, A., Holmes, N., Rakyar, V., Loose, M., 2019. BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics* 35, 2193–2198. <https://doi.org/10.1093/bioinformatics/bty841>
- Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., Salzberg, S.L., 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* 11, 1650–1667. <https://doi.org/10.1038/nprot.2016.095>
- Petersen, G., Cuenca, A., Seberg, O., 2015a. Plastome evolution in hemiparasitic mistletoes. *Genome Biol. Evol.* 7, 2520–2532. <https://doi.org/10.1093/gbe/evv165>
- Petersen, G., Cuenca, A., Seberg, O., 2015b. Plastome Evolution in Hemiparasitic Mistletoes. *Genome Biol. Evol.* 7, 2520–2532. <https://doi.org/10.1093/gbe/evv165>
- Pirozynski, K.A., Malloch, D.W., 1975. The origin of land plants: A matter of mycotrophism. *Biosystems* 6, 153–164. [https://doi.org/10.1016/0303-2647\(75\)90023-4](https://doi.org/10.1016/0303-2647(75)90023-4)
- Posada, D., 2008. jModelTest: Phylogenetic Model Averaging. *Mol. Biol. Evol.* 25, 1253–1256. <https://doi.org/10.1093/molbev/msn083>

11. Bibliography

- Preiss, K., Adam, I.K.U., Gebauer, G., 2010. Irradiance governs exploitation of fungi: fine-tuning of carbon gain by two partially myco-heterotrophic orchids. *Proc. R. Soc. B Biol. Sci.* 277, 1333–1336. <https://doi.org/10.1098/rspb.2009.1966>
- Press, M., Graves, J. (Eds.), 1995. *Parasitic Plants*, 1st ed. Springer Netherlands.
- Pyke, K., 2009. *Plastid Biology*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511626715>
- Race, H.L., Herrmann, R.G., Martin, W., 1999. Why have organelles retained genomes? *Trends Genet.* 15, 364–370. [https://doi.org/10.1016/S0168-9525\(99\)01766-7](https://doi.org/10.1016/S0168-9525(99)01766-7)
- Read, D.J., Duckett, J.G., Francis, R., Ligrone, R., Russell, A., 2000. Symbiotic fungal associations in ‘lower’ land plants. *Philos. Trans. R. Soc. London. Ser. B Biol. Sci.* 355, 815–831. <https://doi.org/10.1098/rstb.2000.0617>
- Rhoads, A., Au, K.F., 2015. PacBio Sequencing and Its Applications. *Genomics. Proteomics Bioinformatics* 13, 278–289. <https://doi.org/10.1016/j.gpb.2015.08.002>
- Rinaldi, A.C., Comandini, O., Kuyper, T.W., 2008. Ectomycorrhizal fungal diversity: Separating the wheat from the chaff. *Fungal Divers.* 33, 1–45.
- Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M., Qian, J.Q., Griffith, M., Raymond, A., Thiessen, N., Cezard, T., Butterfield, Y.S., Newsome, R., Chan, S.K., She, R., Varhol, R., Kamoh, B., Prabhu, A.L., Tam, A., Zhao, Y., Moore, R.A., Hirst, M., Marra, M.A., Jones, S.J.M., Hoodless, P.A., Birol, I., 2010. De novo assembly and analysis of RNA-seq data. *Nat. Methods* 7, 909–912. <https://doi.org/10.1038/nmeth.1517>
- Robinson, M.D., Oshlack, A., 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11, R25. <https://doi.org/10.1186/gb-2010-11-3-r25>
- Rottet, S., Besagni, C., Kessler, F., 2015. The role of plastoglobules in thylakoid lipid remodeling during plant development. *Biochim. Biophys. Acta - Bioenerg.* 1847, 889–899. <https://doi.org/10.1016/j.bbabi.2015.02.002>
- Salinas, B., 2018. *Plant Cytogenetics, Breeding and Evolution*. Scientific e-Resources.
- Salmia, A., 1986. Chlorophyll-free form of *Epipactis helleborine* (Orchidaceae) in SE Finland. *Ann. Bot. Fenn.* 49–57.
- Sanger, F., Coulson, A.R., 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* 94, 441–448. [https://doi.org/10.1016/0022-2836\(75\)90213-2](https://doi.org/10.1016/0022-2836(75)90213-2)
- Scheunert, A., Dorfner, M., Lingl, T., Oberprieler, C., 2020. Can we use it? On the utility of de novo and reference-based assembly of Nanopore data for plant plastome sequencing, *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0226234>
- Schiebold, J.M.I., Bidartondo, M.I., Lenhard, F., Makiola, A., Gebauer, G., 2018. Exploiting mycorrhizas in broad daylight: Partial mycoheterotrophy is a common nutritional strategy in meadow orchids. *J. Ecol.* 106, 168–178. <https://doi.org/10.1111/1365-2745.12831>
- Schimel, D.S., 1993. *Theory and Application of Tracers*. Academic Press.

11. Bibliography

- Schmidt, S., Raven, J.A., Paungfoo-Lonhienne, C., 2013. The mixotrophic nature of photosynthetic plants. *Funct. Plant Biol.* 40, 425. <https://doi.org/10.1071/FP13061>
- 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., Miller, A.N., Wingfield, M.J., Aime, M.C., An, K.-D., Bai, F.-Y., Barreto, R.W., Begerow, D., Bergeron, M.-J., Blackwell, M., Boekhout, T., Bogale, M., Boonyuen, N., Burgaz, A.R., Buyck, B., Cai, L., Cai, Q., Cardinali, G., Chaverri, P., Coppins, B.J., Crespo, A., Cubas, P., Cummings, C., Damm, U., de Beer, Z.W., de Hoog, G.S., Del-Prado, R., Dentinger, B., Dieguez-Uribeondo, J., Divakar, P.K., Douglas, B., Duenas, M., Duong, T.A., Eberhardt, U., Edwards, J.E., Elshahed, M.S., Fliegerova, K., Furtado, M., Garcia, M.A., Ge, Z.-W., Griffith, G.W., Griffiths, K., Groenewald, J.Z., Groenewald, M., Grube, M., Gryzenhout, M., Guo, L.-D., Hagen, F., Hambleton, S., Hamelin, R.C., Hansen, K., Harrold, P., Heller, G., Herrera, C., Hirayama, K., Hirooka, Y., Ho, H.-M., Hoffmann, K., Hofstetter, V., Hognabba, F., Hollingsworth, P.M., Hong, S.-B., Hosaka, K., Houbraken, J., Hughes, K., Huhtinen, S., Hyde, K.D., James, T., Johnson, E.M., Johnson, J.E., Johnston, P.R., Jones, E.B.G., Kelly, L.J., Kirk, P.M., Knapp, D.G., Koljalg, U., Kovacs, G.M., Kurtzman, C.P., Landvik, S., Leavitt, S.D., Liggenstoffer, A.S., Liimatainen, K., Lombard, L., Luangsa-ard, J.J., Lumbsch, H.T., Maganti, H., Maharachchikumbura, S.S.N., Martin, M.P., May, T.W., McTaggart, A.R., Methven, A.S., Meyer, W., Moncalvo, J.-M., Mongkolsamrit, S., Nagy, L.G., Nilsson, R.H., Niskanen, T., Nyilasi, I., Okada, G., Okane, I., Olariaga, I., Otte, J., Papp, T., Park, D., Petkovits, T., Pino-Bodas, R., Quaedvlieg, W., Raja, H.A., Redecker, D., Rintoul, T.L., Ruibal, C., Sarmiento-Ramirez, J.M., Schmitt, I., Schussler, A., Shearer, C., Sotome, K., Stefani, F.O.P., Stenroos, S., Stielow, B., Stockinger, H., Suetrong, S., Suh, S.-O., Sung, G.-H., Suzuki, M., Tanaka, K., Tedersoo, L., Telleria, M.T., Tretter, E., Untereiner, W.A., Urbina, H., Vagvolgyi, C., Vialle, A., Vu, T.D., Walther, G., Wang, Q.-M., Wang, Y., Weir, B.S., Weiss, M., White, M.M., Xu, J., Yahr, R., Yang, Z.L., Yurkov, A., Zamora, J.-C., Zhang, N., Zhuang, W.-Y., Schindel, D., 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci.* 109, 6241–6246. <https://doi.org/10.1073/pnas.1117018109>
- Schulz, M.H., Zerbino, D.R., Vingron, M., Birney, E., 2012. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086–1092. <https://doi.org/10.1093/bioinformatics/bts094>
- Schulze, E.D., Lange, O.L., Ziegler, H., Gebauer, G., 1991. Carbon and nitrogen isotope ratios of mistletoes growing on nitrogen and non-nitrogen fixing hosts and on CAM plants in the Namib desert confirm partial heterotrophy. *Oecologia* 88, 457–462. <https://doi.org/10.1007/BF00317706>
- Schweiger, J.M. -I., Bidartondo, M.I., Gebauer, G., 2018. Stable isotope signatures of underground seedlings reveal the organic matter gained by adult orchids from mycorrhizal fungi. *Funct. Ecol.* 32, 870–881. <https://doi.org/10.1111/1365-2435.13042>
- Selosse, M.-A., Martos, F., 2014. Do chlorophyllous orchids heterotrophically use mycorrhizal fungal carbon? *Trends Plant Sci.* 19, 683–685. <https://doi.org/10.1016/j.tplants.2014.09.005>
- Selosse, M.A., Bocayuva, M.F., Kasuya, M.C.M., Courty, P.E., 2016. Mixotrophy in mycorrhizal plants: Extracting carbon from mycorrhizal networks. *Mol. Mycorrhizal Symbiosis* 451–471. <https://doi.org/10.1002/9781118951446.ch25>

11. Bibliography

- Selosse, M.A., Faccio, A., Scappaticci, G., Bonfante, P., 2004. Chlorophyllous and achlorophyllous specimens of *Epipactis microphylla* (Neottieae, Orchidaceae) are associated with ectomycorrhizal septomycetes, including truffles. *Microb. Ecol.* 47, 416–426. <https://doi.org/10.1007/s00248-003-2034-3>
- Selosse, M.A., Richard, F., He, X., Simard, S.W., 2006. Mycorrhizal networks: des liaisons dangereuses? *Trends Ecol. Evol.* 21, 621–628. <https://doi.org/10.1016/j.tree.2006.07.003>
- Selosse, M.A., Roy, M., 2009. Green plants that feed on fungi: facts and questions about mixotrophy. *Trends Plant Sci.* 14, 64–70. <https://doi.org/10.1016/j.tplants.2008.11.004>
- Selosse, M.A., Strullu-Derrien, C., Martin, F.M., Kamoun, S., Kenrick, P., 2015. Plants, fungi and oomycetes: A 400-million year affair that shapes the biosphere. *New Phytol.* 206, 501–506. <https://doi.org/10.1111/nph.13371>
- Shi, C., Hu, N., Huang, H., Gao, J., Zhao, Y.-J., Gao, L.-Z., 2012. An Improved Chloroplast DNA Extraction Procedure for Whole Plastid Genome Sequencing. *PLoS One* 7, e31468. <https://doi.org/10.1371/journal.pone.0031468>
- Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H., Sugiura, M., 1986. The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. *EMBO J.* 5, 2043–2049. <https://doi.org/10.1002/j.1460-2075.1986.tb04464.x>
- Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J.M., Birol, I., 2009. ABySS: A parallel assembler for short read sequence data. *Genome Res.* 19, 1117–1123. <https://doi.org/10.1101/gr.089532.108>
- Sims, D., Sudbery, I., Illott, N.E., Heger, A., Ponting, C.P., 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* 15, 121–132. <https://doi.org/10.1038/nrg3642>
- Singh, B.P., 2020. CpGDB: A Comprehensive Database of Chloroplast Genomes. *Bioinformatics* 16, 171–175. <https://doi.org/10.6026/97320630016171>
- Slepecky, R.A., Starmer, W.T., 2009. Phenotypic plasticity in fungi: a review with observations on *Aureobasidium pullulans*. *Mycologia* 101, 823–832. <https://doi.org/10.3852/08-197>
- Sloan, D.B., Warren, J.M., Williams, A.M., Wu, Z., Abdel-Ghany, S.E., Chicco, A.J., Havird, J.C., 2018. Cytonuclear integration and co-evolution. *Nat. Rev. Genet.* 19, 635–648. <https://doi.org/10.1038/s41576-018-0035-9>
- Smith, S.E., Read, D.J., 2002. *Mycorrhizal Symbiosis*. Elsevier. <https://doi.org/10.1016/B978-0-12-652840-4.X5000-1>
- Sowden, R.G., Watson, S.J., Jarvis, P., 2018. The role of chloroplasts in plant pathology. *Essays Biochem.* 62, 21–39. <https://doi.org/10.1042/EBC20170020>
- Staddon, P.L., 2004. Carbon isotopes in functional soil ecology. *Trends Ecol. Evol.* 19, 148–154. <https://doi.org/10.1016/j.tree.2003.12.003>

11. Bibliography

- 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 Phytol.* 189, 790–796. <https://doi.org/10.1111/j.1469-8137.2010.03510.x>
- Stull, G.W., Moore, M.J., Mandala, V.S., Douglas, N.A., Kates, H.-R., Qi, X., Brockington, S.F., Soltis, P.S., Soltis, D.E., Gitzendanner, M.A., 2013. A Targeted Enrichment Strategy for Massively Parallel Sequencing of Angiosperm Plastid Genomes. *Appl. Plant Sci.* 1, 1200497. <https://doi.org/10.3732/apps.1200497>
- Suetsugu, K., Yamato M., Miura C., Yamaguchi K., Takahashi K., Ida Y., Shigenobu S., Kaminaka H., 2017. Comparison of green and albino individuals of the partially mycoheterotrophic orchid *Epipactis helleborine* on molecular identities of mycorrhizal fungi, nutritional modes and gene expression in mycorrhizal roots. *Mol. Ecol.* 26, <https://doi.org/10.1111/mec.14021>
- Sulzman, E.W., 2007. Stable Isotope Chemistry and Measurement: A Primer, in: Michener, R., Lajtha, K. (Eds.), *Stable Isotopes in Ecology and Environmental Science*. Blackwell Publishing Ltd, Oxford, UK, pp. 1–21. <https://doi.org/10.1002/9780470691854.ch1>
- Taberlet, P., Gielly, L., Pautou, G., Bouvet, J., 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* 17, 1105–1109. <https://doi.org/10.1007/BF00037152>
- Taylor, A.F.S., Fransson, P.M., Högborg, P., Högborg, M.N., Plamboeck, A.H., 2003. Species level patterns in ¹³C and ¹⁵N abundance of ectomycorrhizal and saprotrophic fungal sporocarps. *New Phytol.* 159, 757–774. <https://doi.org/10.1046/j.1469-8137.2003.00838.x>
- Tedersoo, L., May, T.W., Smith, M.E., 2010. Ectomycorrhizal lifestyle in fungi: global diversity, distribution, and evolution of phylogenetic lineages. *Mycorrhiza* 20, 217–263. <https://doi.org/10.1007/s00572-009-0274-x>
- Těšitel, J., Plavcová, L., Cameron, D.D., 2010. Interactions between hemiparasitic plants and their hosts: The importance of organic carbon transfer. *Plant Signal. Behav.* 5, 1072–1076. <https://doi.org/10.4161/psb.5.9.12563>
- The *Arabidopsis* Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. <https://doi.org/10.1038/35048692>
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. <https://doi.org/10.1093/nar/22.22.4673>
- 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. <https://doi.org/10.1371/journal.pone.0040863>
- Trudell, S.A., Rygielwicz, P.T., Edmonds, R.L., 2003. Nitrogen and carbon stable isotope abundances support the myco-heterotrophic nature and host-specificity of certain achlorophyllous plants. *New Phytol.* 160, 391–401. <https://doi.org/10.1046/j.1469-8137.2003.00876.x>
- Twyford, A.D., Ness, R.W., 2017. Strategies for complete plastid genome sequencing. *Mol. Ecol. Resour.* 17, 858–868. <https://doi.org/10.1111/1755-0998.12626>

11. Bibliography

- Unamba, C.I.N., Nag, A., Sharma, R.K., 2015. Next generation sequencing technologies: The doorway to the unexplored genomics of non-model plants. *Front. Plant Sci.* 6. <https://doi.org/10.3389/fpls.2015.01074>
- van der Heijden, M.G.A., Martin, F.M., Selosse, M.A., Sanders, I.R., 2015. Mycorrhizal ecology and evolution: The past, the present, and the future. *New Phytol.* 205, 1406–1423. <https://doi.org/10.1111/nph.13288>
- Wakasugi, T., Sugita, M., Tsudzuki, T., Sugiura, M., 1998. Updated Gene Map of Tobacco Chloroplast DNA. *Plant Mol. Biol. Report.* 16, 231–241. <https://doi.org/10.1023/A:1007564209282>
- Walker, J.F., Aldrich-Wolfe, L., Riffel, A., Barbare, H., Simpson, N.B., Trowbridge, J., Jumpponen, A., 2011. Diverse Helotiales associated with the roots of three species of Arctic Ericaceae provide no evidence for host specificity. *New Phytol.* 191, 515–527. <https://doi.org/10.1111/j.1469-8137.2011.03703.x>
- Wang, B., Kumar, V., Olson, A., Ware, D., 2019. Reviving the Transcriptome Studies: An Insight Into the Emergence of Single-Molecule Transcriptome Sequencing. *Front. Genet.* 10. <https://doi.org/10.3389/fgene.2019.00384>
- Wang, B., Qiu, Y.L., 2006. Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* 16, 299–363. <https://doi.org/10.1007/s00572-005-0033-6>
- Wang, D., Jacquemyn, H., Gomes, S.I.F., Vos, R.A., Merckx, V.S.F.T., 2021. Symbiont switching and trophic mode shifts in Orchidaceae. *New Phytol.* 231, 791–800. <https://doi.org/10.1111/nph.17414>
- Wang, S., Gribskov, M., 2016. Comprehensive evaluation of de novo transcriptome assembly programs and their effects on differential gene expression analysis. *Bioinformatics* 33, btw625. <https://doi.org/10.1093/bioinformatics/btw625>
- Warren, R.L., Sutton, G.G., Jones, S.J.M., Holt, R.A., 2007. Assembling millions of short DNA sequences using SSAKE. *Bioinformatics* 23, 500–501. <https://doi.org/10.1093/bioinformatics/btl629>
- White, T.J., Bruns, T.D., Lee, S.B., Taylor, J.W., 1990. Amplification and direct sequencing of fungal ribosomal RNA Genes for phylogenetics, in: *PCR Protocols, a Guide to Methods and Applications*. Academic Press, pp. 315–322.
- Whitmarsh, J., Govindjee, 1999. The Photosynthetic Process. *Concepts Photobiol.* 11–51. https://doi.org/10.1007/978-94-011-4832-0_2
- Wicke, S., Müller, K.F., DePamphilis, C.W., Quandt, D., Bellot, S., Schneeweiss, G.M., 2016. Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proc. Natl. Acad. Sci.* 113, 9045–9050. <https://doi.org/10.1073/pnas.1607576113>
- Wicke, S., Schäferhoff, B., DePamphilis, C.W., Müller, K.F., 2014. Disproportional Plastome-Wide Increase of Substitution Rates and Relaxed Purifying Selection in Genes of Carnivorous Lentibulariaceae. *Mol. Biol. Evol.* 31, 529–545. <https://doi.org/10.1093/molbev/mst261>
- Wickett, N.J., Honaas, L.A., Wafula, E.K., Das, M., Huang, K., Wu, B., Landherr, L., Timko, M.P., Yoder, J., Westwood, J.H., DePamphilis, C.W., 2011. Transcriptomes of the Parasitic Plant

11. Bibliography

- Family Orobanchaceae Reveal Surprising Conservation of Chlorophyll Synthesis. *Curr. Biol.* 21, 2098–2104. <https://doi.org/10.1016/j.cub.2011.11.011>
- Wilson, R.J.M. (Iain.), Denny, P.W., Preiser, P.R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D.J., Moore, P.W., Williamson, D.H., 1996. Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J. Mol. Biol.* 261, 155–172. <https://doi.org/10.1006/jmbi.1996.0449>
- Wise, R.R., Hooper, J.K. (Eds.), 2006. *The Structure and Function of Plastids, Advances in Photosynthesis and Respiration*. Springer Netherlands, Dordrecht. <https://doi.org/10.1007/978-1-4020-4061-0>
- Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T.W., Li, Y., Xu, X., Wong, G.K.S., Wang, J., 2014. SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* 30, 1660–1666. <https://doi.org/10.1093/bioinformatics/btu077>
- Yagame, T., Orihara, T., Selosse, M.-A., Yamato, M., Iwase, K., 2012. Mixotrophy of *Platanthera minor*, an orchid associated with ectomycorrhiza-forming Ceratobasidiaceae fungi. *New Phytol.* 193, 178–187. <https://doi.org/10.1111/j.1469-8137.2011.03896.x>
- Yigit, E., Hernandez, D.I., Trujillo, J.T., Dimalanta, E., Bailey, C.D., 2014. Genome and Metagenome Sequencing: Using the Human Methyl-Binding Domain to Partition Genomic DNA Derived from Plant Tissues. *Appl. Plant Sci.* 2, 1400064. <https://doi.org/10.3732/apps.1400064>
- Yu, M., Ashworth, M.P., Hajrah, N.H., Khiyami, M.A., Sabir, M.J., Alhebshi, A.M., Al-Malki, A.L., Sabir, J.S.M., Theriot, E.C., Jansen, R.K., 2018. Evolution of the Plastid Genomes in Diatoms, in: *Advances in Botanical Research*. Elsevier Ltd., pp. 129–155. <https://doi.org/10.1016/bs.abr.2017.11.009>
- Zerbino, D.R., Birney, E., 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. <https://doi.org/10.1101/gr.074492.107>
- Zhao, Y., Li, M.C., Konaté, M.M., Chen, L., Das, B., Karlovich, C., Williams, P.M., Evrard, Y.A., Doroshov, J.H., McShane, L.M., 2021. TPM, FPKM, or Normalized Counts? A Comparative Study of Quantification Measures for the Analysis of RNA-seq Data from the NCI Patient-Derived Models Repository. *J. Transl. Med.* 19, 1–15. <https://doi.org/10.1186/s12967-021-02936-w>