Evolution and development of orchid flowers and fruits

Dirks, A.

Chapter 5

Transcriptome and ancestral character state analysis of orchid fruits

Anita Dirks-Mulder[1,2], Israa Ahmed[2], Marian Bemer[3], Mark uit het Broek[2], Nemi Dorst[2], Jasmijn Snier[2], Anne van Winzum[2], Martijn van 't Wout[2], Stephen Pieterman[2], Floyd Wittink[2], Jamie Zeegers[2], Bertie Joan van Heuven[1], Jaco Kruizinga[4], Erik F. Smets[1,5,6], Diego Bogarin[1] & Barbara Gravendeel[1,2,6]

[1]Endless Forms group, Naturalis Biodiversity Center, Darwinweg 2, 2333 CR Leiden, The Netherlands

[2]Faculty of Science and Technology, University of Applied Sciences Leiden, Zernikedreef 11, 2333 CK Leiden, The Netherlands

[3]Department of Plant Sciences, Laboratory of Molecular Biology, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

[4]Hortus botanicus, Leiden University, Rapenburg 73, 2311 GJ Leiden, The Netherlands

[5]Ecology, Evolution and Biodiversity Conservation cluster, KU Leuven, Kasteelpark Arenberg 31, 3001 Leuven, Belgium

[6]Institute of Biology Leiden, Leiden University, Sylviusweg 72, 2333 BE Leiden, The Netherlands

Abstract

The orchid family is known for its vast floral diversity. Orchid fruits are highly diverse as well but much less studied. The first aim of this study is to detect genes and gene networks involved in fruit development of *Erycina pusilla* so targeting a much wider set than only the MADS-box genes discussed in the previous chapter. Secondly, we investigated possible correlated evolution of a total of 12 binary traits of orchid fruits from 41 different species in all five subfamilies on a combined nrITS, *mat*K and *rbc*L based phylogeny.

Next-generation sequencing of RNA isolated from entire fruits of *E. pusilla* harvested from different developmental stages, was used to generate a full fruit transcriptome. Fresh and spirit conserved mature fruits were collected, hand cut and stained with phloroglucinol to visualize lignification. A reversible-jump Markov chain Monte Carlo search was employed to search among possible models to describe joint evolution of fruit traits on a phylogeny of the orchid family.

Preliminary results from the orchid fruit transcriptomes analyzed, indicate that genes related to pollen tube formation, seed development and lignification are differentially expressed during development. A striking variation was found in fruit type, direction, ripening period, number of slits and lignification patterns. The latter varied from lignification of the exocarp, various tissue types such as valves or dehiscence zones in the endocarp, to no lignification at all except for the trichomes. The posterior distributions calculated indicate that an epiphytic or lianaceous habit and fruit ripening period of more than four months have clearly coevolved in orchids. There is also strong evidence for the hypothesis that an epiphytic or lianaceous habit co-evolved with a smaller number of opening slits of orchid fruits. Similarly, pendant orchid fruits co-evolved with a preference for growing at intermediate to high temperatures. Strong support was also found for the hypothesis that pendant orchid fruits co-evolved with an epiphytic or lianaceous habit. Lastly, lignification of the valves evolved in both pendant and erect orchid fruits. Future comparative transcriptome analyses including more species may reveal which developmental genes drive the high morphological diversity of orchid fruits.

Key words: Bayesian analyses - Coevolution - Endocarp - Exocarp - Fruit dehiscence - Lignification - OmicsBox - RNAseq

Introduction

Pollination of flowers is a key regulatory event in plant reproduction. When an orchid flower is pollinated, the perianth withers, the stigma closes, and the ovary starts to grow. The ovary of orchids is inferior and made up of three carpels. Unlike most flowering plants, pollination and fertilization are separated in orchids by relatively long periods of time in which the oocyte and embryo sac develop (Dirks-Mulder *et al*., 2019). Once a flower is pollinated, pollen tubes grow from the stigma into the gynostemium towards the ovary, the placenta starts proliferating and ovules are formed. Fertilization of orchid ovules can occur days to months after pollination. Finally, the inferior ovary develops into a fruit, containing up to millions of seeds.

There are two different main types of orchid fruits, fleshy and dry (Dressler, 1993). Most orchids produce dehiscent capsules, which split open along the midline of each carpel when the capsule is fully ripe. Just a few orchids species produce fleshy berries or capsules, which can be found in several genera such as *Neuwiedia* (Apostasioideae), *Palmorchis* (Epidendroideae), *Cyrtosia* and *Vanilla* (Vanilloideae) (Dressler, 1981;1993;Kocyan and Endress, 2001). Indehiscent capsules can be found in *Dictyophyllaria* and *Vanilla* (Vanilloideae) and *Selenipedium* (Cypripedioideae) (Stern *et al*., 2014).

Fleshy orchid fruits are generally indehiscent, except for a few exceptions such as *Vanilla trigonocarpa*, while dry fruits can be either dehiscent or indehiscent. Dehiscent fruits open and release the seeds to be spread by wind, water or animals, while indehiscent fruits have to decay or be eaten and digested by animals in order to spread their seeds. Dehiscence can be defined as the completion of the reproductive cycle, in which the fruit will open and release seeds. This process is achieved by a coordinated cell separation (Mayer *et al*., 2011).

Although orchid fruits are very diverse in size and shape, they all have the same basic structure, which is the result of the differentiation and development of the inferior carpels. The carpels develop into six different valves, three fertile valves, with a placenta, and three sterile valves, not connected to any seed tissue. When a capsule starts to dehisce, the splitting of the valves occurs along the midline of a carpel, between the sterile and the fertile valves. The fruit maximally splits into three wide fertile valves and three narrow sterile ones (Dirks-Mulder *et al*., 2019). The number of dehiscent splits can vary from one, two, three to six splits (Beer, 1863;Horowitz, 1901;Dressler, 1981;1993). In **chapter 4**, we discovered that at the dehiscence zone of orchid fruits either a cuticle layer develops, such as for instance in *Erycina pusilla*, or that valves become lignified during fruit development, such as in *Epipactis helleborine* (both Epidendroideae) and *Cynorkis fastigiate* (Orchidaceae). Lignin is a very common biopolymer found in vegetables, fruits and the secondary cell walls of plants. In Arabidopsis, lignification of the valve margin and endocarp is necessary for creating a mechanical tension in the fruit to open (dehisce) and shatter its seeds (Di Vittori *et al*., 2019).

To gain more insight into the genes regulating these processes, we performed RNA-seq analysis of *E. pusilla* fruits from different stages: 0 and 5 days after pollination (DAP) and 4 and 8 weeks after pollination (WAP). In **chapter 4,**

we divided fruit development of *E. pusilla* in four developmental stages based on morphology: (I) 0 DAP - 1 WAP, (II) 2 WAP - 5 WAP, (III) 6 WAP - 11 WAP and (IV) 12 WAP - 16 WAP. In the first stage, cell division takes place in the fertile and sterile valves and the trichomes start to develop. We therefore expected genes involved in these processes to be up-regulated. In the second stage, when the fruit starts to elongate, six bundles of pollen tubes are formed. We therefore expected an upregulation of expression of genes involved in pollen tube formation and cell elongation in this stage.

During the third stage, when all the trichomes become lignified, the dehiscence zone is developing, and embryonic development takes place. We expected genes involved in all these processes to show a change in expression here.

Secondly, to investigate which fruit traits evolved together and which evolved independently of each other, we expanded our analyses to ripe fruits from additional species belonging to all five subfamilies of the orchids and carried out an ancestral character state study for twelve different fruit characters (Camus *et al*., 1921;Ziegenspeck, 1936). We hypothesized that lignification could be an adaptation for fast fruit dehiscence and therefore expected it to have evolved especially in orchids coping with relatively short seasons suitable for seed dispersal. To test this hypothesis, we studied lignin formation in fruits of orchid species from five subfamilies (Apostasioideae, Cypripedioideae, Epidendroideae, Orchidoideae and Vanilloideae) with different growth forms (either epiphytic, terrestrial or lianaceous). A reversible-jump Markov chain Monte Carlo search was employed to search among possible models to describe joint evolution on a phylogeny. According to this model, transitions among combinations of states result from two binary valuables. Posterior distributions were calculated to estimate statistical support for hypotheses of correlated evolution.

Materials and methods

Plant material

A more than 20-year-old inbred line of *Erycina pusilla* originally collected in Surinam was grown in climate rooms under controlled conditions (7.00h – 19.00h light regime), at a temperature of 22 °C and a relative humidity of 50%. The orchids were cultured in vitro under sterile conditions on Phytamax™ orchid medium with charcoal and banana powder (Sigma-Aldrich) with 4g/L Gelrite™ (Duchefa) culture medium. Pollination was conducted manually by placing the pollinia of flowers on each other's stigma to produce fruits and seeds.

Total RNA extraction and RNA-sequencing

Carpels and fruits were collected from *E. pusilla* at 0 and 5 days after pollination (DAP) and 4 and 8 weeks after pollination (WAP), frozen in liquid nitrogen and stored at -80 °C until they were used for RNA extraction. There were three replicates for each sample. Total RNA was extracted using the RNeasy Plant Mini Kit (QIAGEN) and DNase I (QIAGEN) treated, following the manufacturer's protocol. A maximum of 100 mg plant material was placed in a 2.2 ml micro centrifuge tube with 7 mm

glass beads. The TissueLyser II (QIAGEN) was used to grind the plant material. The amount of RNA was measured using the NanoVue Plus™ (GE Healthcare Life Sciences) and its integrity was assessed on a 2100 Bioanalyzer (Agilent) using the Plant RNA nano protocol. RNA was stored at -80 °C until further use.

Messenger-RNA selection, library preparation and strand-specific sequencing of in total twelve RNA pools (0 DAP, 5 DAP, 4 WAP and 8 WAP x 3 biological replicates) was performed at BGI to generate 150-bp paired-end (PE) reads following the Illumina TruSeq Stranded sample preparation protocol. Summarized, mRNA was isolated using oligo(dT)-attached magnetic beads and fragmented. Then cDNA was synthesized using random hexamer primers, purified, end-repaired, poly-A tailed, adaptor ligated and loaded on an Illumina Hiseq4000™ sequencer (Illumina, United States) for sequencing. The raw reads were filtered with internal software at BGI for low-quality, adaptor-polluted and high content of unknown base(N) reads in order to obtain clean reads. The cleaned reads (in fastq format) were downloaded and stored on a High-End Workstation (**Table S1**) for further analysis.

RNA sequence analysis

Two different methods were used for the RNAseq analysis, both implemented in OmicsBox (v1.0.34) (OmicsBox;Conesa *et al*., 2005). First of all, a d*e novo* assembly for whole-transcriptome construction was carried out using the clean reads from BGI. The reads were assembled with Trinity (Grabherr *et al*., 2011;Haas *et al*., 2013), creating a sequence table containing the assembled transcripts sequences. Using local Blast, the assembled sequences were blasted against a local database consisting of all orchid sequences "txid4747" [Organism] in the NCBI database. BLAST hits were then functionally annotated using Blast2GO (Conesa *et al*., 2005;Gotz *et al*., 2008) and Gene Ontology (GO) terms (Ashburner *et al*., 2000) were assigned to transcripts with BLAST hits to understand gene function in the developing fruits.

RNA-Seq by Expectation-Maximization (RSEM) was used for alignment-based abundance estimation using Bowtie (Li and Dewey, 2011;Langmead and Salzberg, 2012). Since the assembled transcriptome represented all the transcripts found in the fruits of all four time-points, the sequence reads from each sample were aligned to the assembled transcriptome to differentiate transcripts unique to one stage from those present throughout fruit development. The read counts for each transcript were analyzed using edgeR (Robinson *et al*., 2010) for differential expression analysis, and the logarithmic fold change and false detection rate (FDR) were calculated for each transcript. The results of differential expression and GO annotation were combined for Gene Ontology enrichment using GOseq, which identifies GO terms enriched as a result of differential expression between samples.

Secondly, the *Phalaenopsis equestris* reference genome (Cai *et al*., 2015) together with its annotation were downloaded from Plaza 4.0 (Van Bel *et al*., 2018). The *E. pusilla* Illumina reads were mapped to the *P. equestris* reference genome using STAR (Spliced Transcript Alignment to a Reference) (Dobin *et al*., 2013)/, changing the default parameters to a minimal match of 0.05 in order to obtain >75% of the total reads aligned. The output BAM files (annotated files) were used to create a Count Table for every time point, based on HTSeq package (Anders *et al*.,

2015). A Single Series Time-Course Expression Analysis was performed using default settings (Conesa *et al*., 2006) in order to obtain expression profiles, either clustered or by gene.

Fruit material

A total of 41 ripe fruits from different orchid species and, as outgroup, *Narcissus* and *Hypoxis* fruits were freshly collected in the Hortus botanicus (Leiden, The Netherlands), Lankester Botanical Garden (Cartago, Costa Rica), Utrecht Botanic Gardens (The Netherlands), Bochum Botanical Garden (Germany) and from private collections of orchid breeders throughout Europe. The ripe fruits were fixed and stored in 70% ethanol at room temperature.

Lignin staining and visualization

Lignin formation in fruits was visualized using phloroglucinol staining. Handmade cross sections of fruits were cleared with 100% lactic acid (Merck) for one hour at 60 °C, stained with 1% phloroglucinol (Merck) in 96% ethanol for one hour and 25%-(v/v) hydrochloric acid (HCl) (Merck) for 2-5 minutes until a clear staining was visible and immediately examined under a Binocular microscope (Zeiss SteREO Discovery. V12). Between each step the sections were washed three times with demineralized water.

Alignments and phylogenetic analyses

Representative *mat*K, *nr*ITS and *rbc*L sequences were downloaded from NCBI GenBank (https://www.ncbi.nlm.nih.gov/genbank/) including 117 accessions of 50 species covering every subfamily of the Orchidaceae (**Table S2**). In addition, we included two accessions of *Hypoxis curtissii* Rose (Hypoxidaceae) and two of *Narcissus bulbocodium* L. (Amaryllidaceae) as outgroups due to the close relationship of these families with orchids within the monocots (Givnish *et al*., 2015). We aligned and trimmed the matrices of each marker in Geneious® R9 (Biomatters Ltd., Auckland, New Zealand (Kearse *et al*., 2012) using MAFFT (Multiple Alignment using Fast Fourier Transform). The concatenated dataset (nrITS+*mat*K+*rbc*L) was built with Sequence Matrix v100.0 (Vaidya *et al*., 2011). When sequences for a specific marker were not available, they were included as missing data. When a species analyzed morphologically was not available in NCBI GenBank, it was substituted in the DNA concatenated matrix by a DNA sequence of a close relative.

To obtain ultrametric trees for the ancestral state reconstructions, we estimated the divergence times in BEAST v.1.8.2 using the CIPRES Science Gateway (Miller *et al*., 2010). The following MCMC parameters were set to 20×10^6 generations and a sampling frequency of 1000 yielding 20,001 trees per run. The substitution model selected was GTR, estimated with 4 gamma categories, estimated lognormal relaxed clock (uncorrelated) and the Yule Process (Y) speciation tree model. A normal prior distribution of 105.06 (±2.5 standard deviations) Ma was assigned to the root node of the Orchidaceae and 94.75 (±3.5 standard deviations) Ma to the node containing all Orchidaceae members. These secondary calibrations were obtained from various dating studies of the Orchidaceae (Ramirez *et al*., 2007;Chomicki *et*

al., 2015;Givnish *et al*., 2015;Poinar and Rasmussen, 2017). The convergence of independent runs and the MCMC parameters (ESS values >200) were inspected in Tracer v.1.6. Finally, a maximum clade credibility (MCC) tree was obtained with a 10% burnin using TreeAnnotator v.1.8.2. Resulting trees and the 95% highest posterior density (HPD) estimations were visualized in FigTree v1.4.3 (Rambaut, 2014) and manipulated with R programming language (R Core Team, 2018) under R Studio (Gandrud, 2015) using the packages APE, ggtree and phytools (Paradis *et al*., 2004;Revell, 2012;Yu *et al*., 2016).

Ancestral State Reconstruction

Ancestral state reconstructions (ASRs) were assessed with ML and stochastic character mapping (SCM) using ultrametric trees. For the ML approach we tested several models: equal rates (ER), symmetrical (SYM) and all rates different (ARD) with the re-rooting method of Yang *et al*. (1995) and the function ACE implemented in the R (R Development Core Team, 2018)packages APE, ggtree and phytools (Paradis *et al*., 2004;Revell, 2012;Yu *et al*., 2016). A likelihood ratio test, comparing the log-likelihoods among models, was used to select the best-fitting model. For the SCM analysis we performed 100 replicates on 100 randomly selected trees (10,000 mapped trees) derived from the BEAST analysis. These trees were randomly selected using the R function *samples.trees* (http://coleoguy.blogspot.de/ 2012/09/ randomly-sampling-trees.html). Results of transitions and the proportion of time spent in each state were obtained with the functions *make.simmap* and *describe. simmap* (Bollback, 2006;Revell, 2012). The mean probabilities retrieved at each node were plotted with phytools on the MCC tree for each character analyzed.

Correlated evolution between traits

We tested correlations among traits using the program BayesTraits V3.0.1 (Pagel, 1994;Pagel and Cunningham, 1999;Pagel and Meade, 2006) by performing 120 comparisons between each pair of the 12 characters of fruits assessed. We tested the two models under Bayesian Inference: a dependent model which allows correlation among traits against the independence model with no correlation among traits (correlations set to zero). To account for phylogenetic uncertainty, we used a set of 1000 trees (randomly selected with the R function *samples.trees* as described above) from the post burnin sample of the 20,001 ultrametric trees obtained from the time-calibrated BEAST analysis. The MCMC parameters of each model were set to 1,010,000 iterations, sample period of 1,000, burnin of 10,000, auto tune rate deviation and stepping stones 100 10,000 with reversible jump hyper-prior exponential. The BayesTraits outputs files were analyzed in R with the BayesTraits wrapper (btw) by Randi H Griffin (http://rgriff23.github.io/projects/ btw.html) and other functions from btrtools and BTprocessR (https://github.com/ hferg). The MCMC stationarity of parameters (ESS values >200) and convergence of chains were checked with the R package coda (Plummer *et al*., 2006) and the function *mcmcPlots* of BTprocessR. Using the log marginal likelihoods obtained from the outputs, we estimated a log Bayes factor (logBF) for the dependent model and the independent model with: logBF = 2*(SS dependent model (complex model)

– SS independent model (simple model)). We interpreted the logBF as suggested in the BayesTraitsV3 manual (http://www.evolution.rdg.ac.uk/BayesTraitsV3.0.1/Files/BayesTraitsV3.Manual.pdf): <2: weak evidence; >2: positive evidence; 5-10: strong evidence and >10: very strong evidence.

Results

Transcriptome sequencing of *Erycina pusilla* fruits

To discover more about genes involved in orchid fruit formation, total RNA was extracted of *E. pusilla* fruits from different stages: 0 and 5 days after pollination (DAP) and 4 and 8 weeks after pollination (WAP). The RNA was sequenced at the Beijing Genomics Institute (BGI) using an Illumina Hiseq 4000 high-throughput sequencing platform.

Because there is no *E. pusilla* reference genome available, we performed a *de novo* assembly. With the clean reads obtained from BGI, Trinity (Grabherr *et al*., 2011) produced a set of 501,110 non-redundant putative transcripts and 260,320 unigenes with an average length of 788 bp and an N50 value of 1,357 bp. The length of transcripts ranged from 201 to 47,179 bp with a GC content of 36.9%.

The assembled 260,320 unigenes were annotated by comparison to the orchid proteins in the NCBI's non-redundant (NR) protein database. A Principal Component Analysis (PCA) was conducted, resulting in an MDS Plot, showing that not all three independent biological replicates of each sample had a good reproducibility (**Figure S1A**). The two deviating samples were excluded from further analysis (**Figure S1B**). Through BLASTx, transcripts (43.8% of the transcriptome) could be annotated with a description based on sequence homology. These transcripts were functionally annotated using Gene Ontology (GO) terms to understand the role of the genes in the tissue analyzed. Of the 219,296 transcripts annotated using BLASTx, 196,859 (89.8%) could be assigned to GO terms.

To identify differentially expressed genes (DEGs) between the different fruit time points, we determined relative expression levels using DEseq and found an increasing number of both up- and down-regulated features (**Figure 1**).

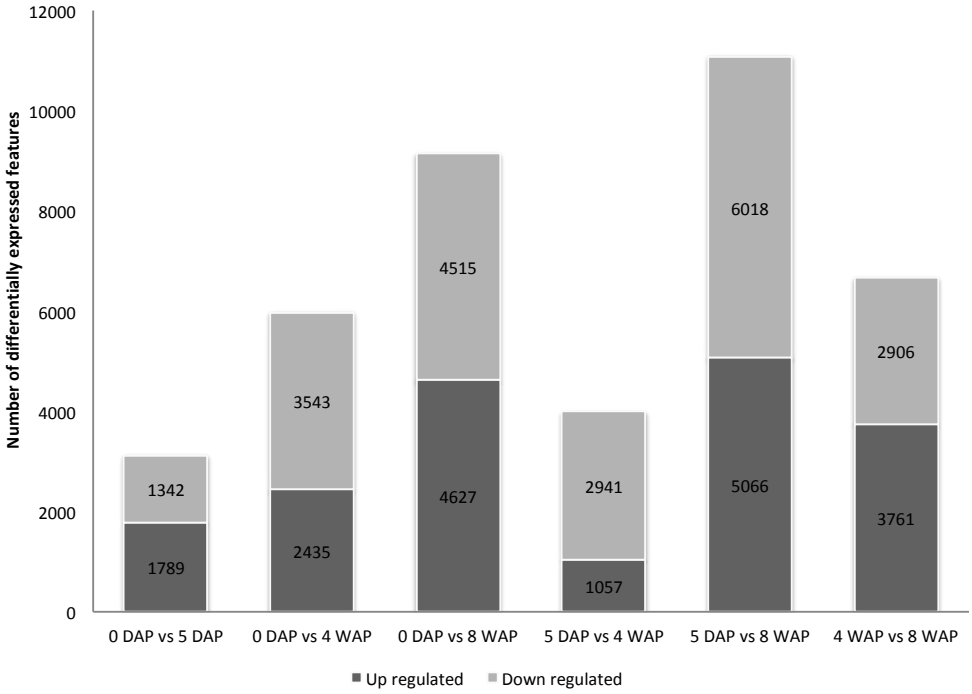


Figure 1: **Changes in gene expression profiles of *E. pusilla* fruits sampled at 0 DAP, 5 DAP, 4 WAP and 8 WAP as expressed in the number of features**. Abbreviations: DAP = days after pollination, WAP = weeks after pollination.

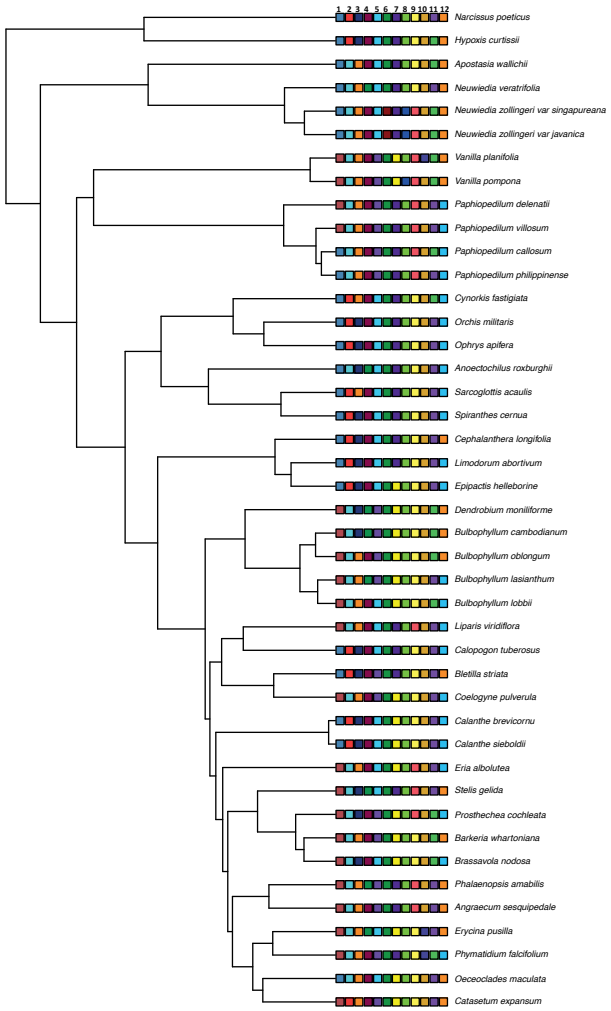
For *E. pusilla* we were not able to detect gene-networks involved in fruit development with a *de novo* assembled transcriptome approach, because of high redundancy in the raw contigs. Using CD-HIT (Li and Godzik, 2006)d to remove redundant contigs from *de novo* contig assemblies didn't improve much.

With the availability of an annotated reference genome of the orchid *Phalaenopsis equestris,* we decided to map the reads against this genome (Cai *et al*., 2015;Van Bel *et al*., 2018).

After aligning the *E. pusilla* reads against the *P. equestris* genome, >75% of the reads could be aligned. A Count Table (BAM+GFF) for gene level quantification was created in OmixcBox and the Counts per Category were visualized showing the number of reads of each input file sorted by different categories. For every input file, more than 10,000 read counts could be assigned to a feature (**Figure S2**), of a total number of 29,431 features.

For the Time Course Expression analyses, we performed a Single Series Time Course, with Time factor Group. The results were visualized using K-Means Clustering, resulting in 5,984 significant features divided over 9 clusters (**Table S3**). Genes belonging to Cluster 2 decreased their expression after pollination, for cluster 5 expression increased at 4 WAP and for cluster 7-9 expression increased at 8 WAP. Genes in these clusters belonging to GO terms involved in fruit and seed development, pollen tube development and lignin biosynthesis were further

analyzed in order to identify gene-networks involved in orchid fruit development (**Figure S3** and **Table S4**).



Number	Character	State 0	State 1
1	Life form	Terrestrial	Epiphytic/Liana
2	Strategy	Evergreen	Deciduous
3	Minimum temperature requirement	Cool/Intermediate (<15 °C)	High (> 15 °C)
4	Production	All year round	Seasonal
5	Ripening period (months)	≤ 4	> 4
6	Type	Berry	Capsule
7	Direction	Erect	Pendant
8	Dehiscence type	Indehiscent	Dehiscent
9	Number of slits in mature fruit	<3	≥ 3
10	United valves	United at the base and apex	United at the base
11	Valves	No	Yes
12	Dehiscence zones	No	Yes

Figure 2: **Consensus tree used for the ancestral state analyses of 41 orchid species and two non-orchid species (used as outgroup) plotted together with a character matrix score for 12 characters.** Almost all branches are supported by 90% bootstrap support or higher.

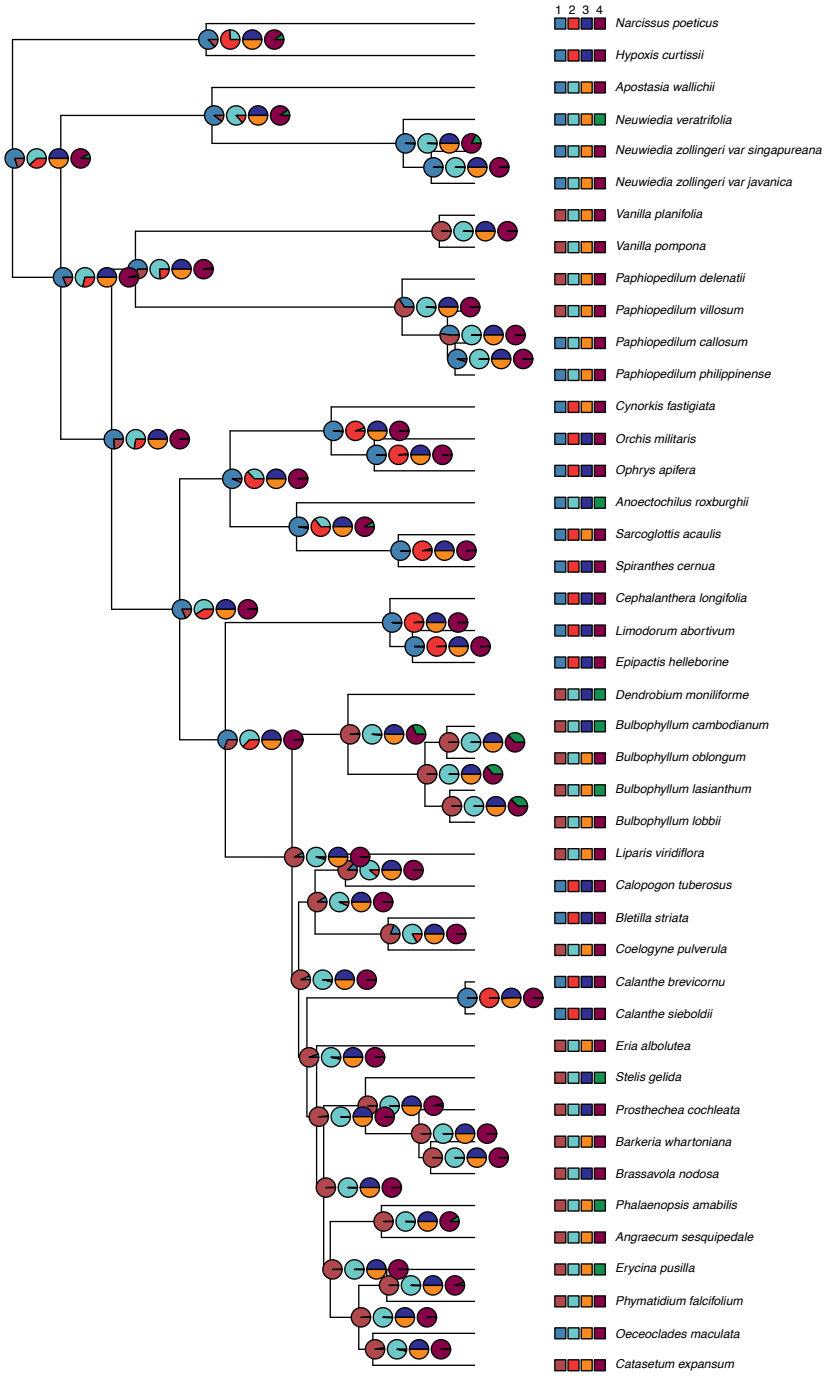


Figure 3: **Ancestral state reconstructions of selected characters from stochastic mapping analysis based on joint sampling (10,000 mapped trees).** Posterior probabilities (pie charts) are mapped in a random stochastic character map. Each tree represents four of the twelve characters represented in **Figure 2**.

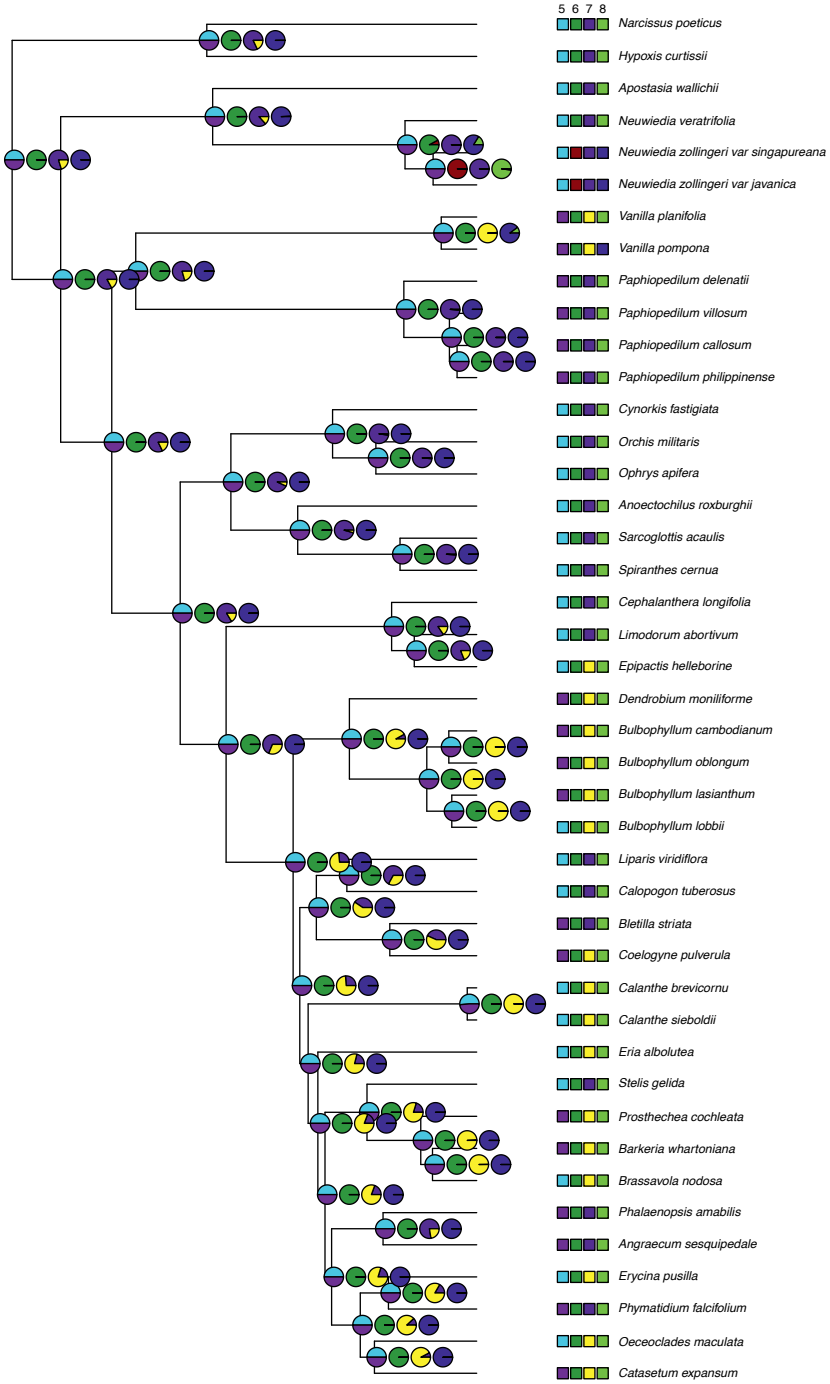


Figure 3 continued: Ancestral state reconstructions of selected characters from stochastic mapping analysis based on joint sampling (10,000 mapped trees).

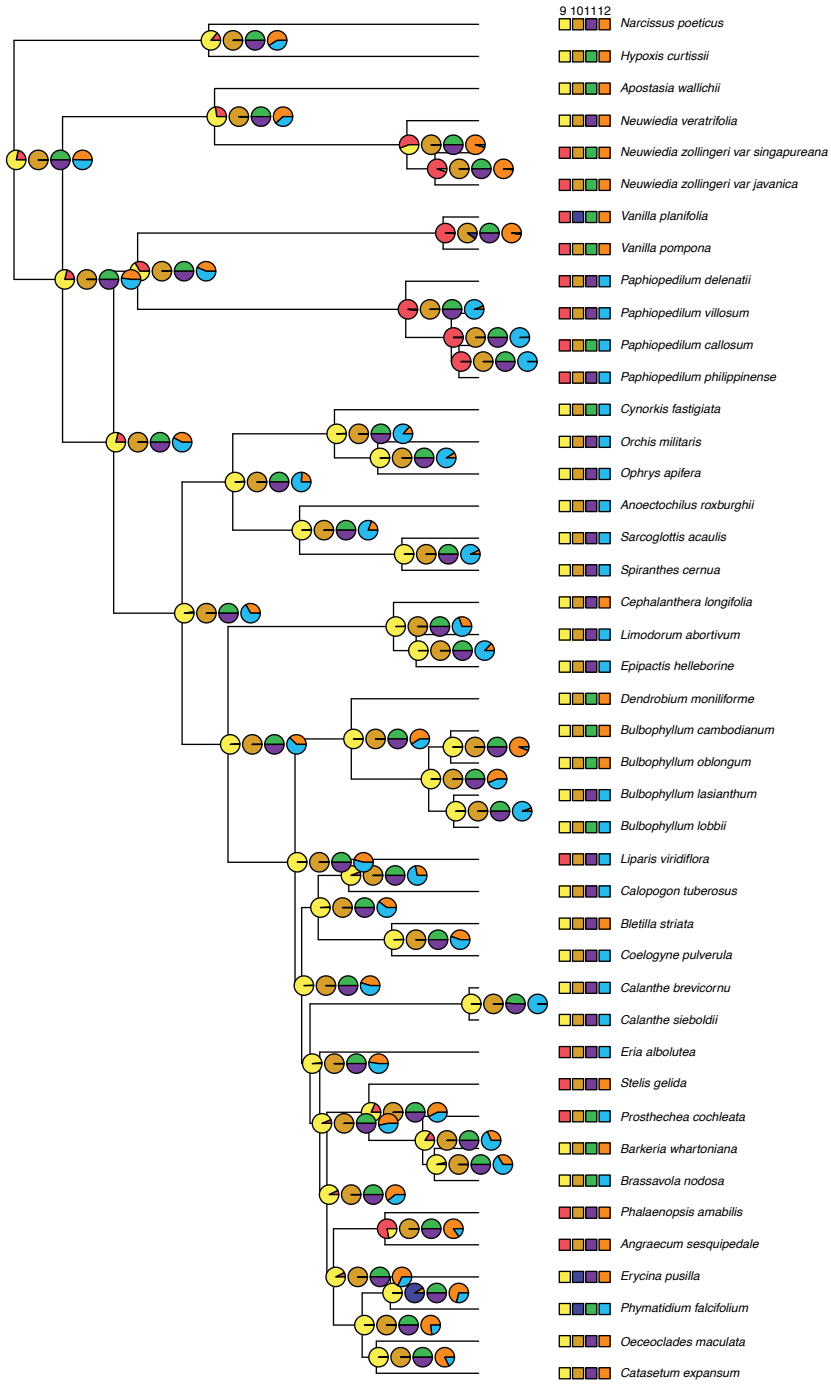


Figure 3 continued: Ancestral state reconstructions of selected characters from stochastic mapping analysis based on joint sampling (10,000 mapped trees).

Ancestral character state analyses

A total of 12 different characters, all binary states, could be scored based on the variation among the species studied as observed from literature, personal observations in the greenhouse, and after staining with phloroglucinol (**Figure S4, S5 and S6**). These twelve characters were plotted on a consensus of the orchid phylogeny (**Figure 2**) and further analyzed with ancestral state analyses.

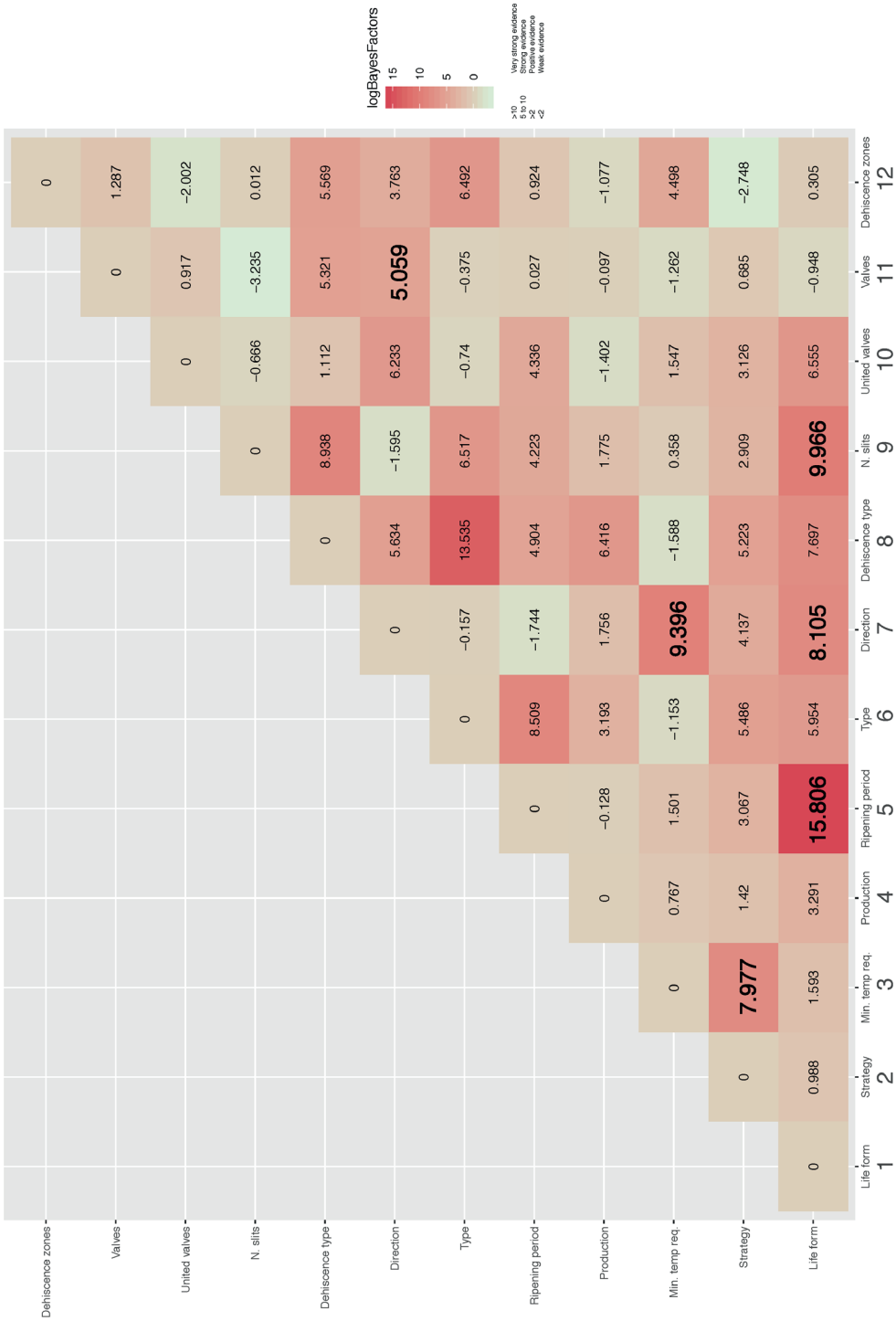
The ASR were based on the ER (equal rate) model, which performed consistently better than the SYM (Symmetrical rate) and the ARD (all rates different) models. The results of the Bayesian analyses are summarized in **Figure 3** showing that for character 1 terrestrial, character 7 erect and for character 9 ≥ 3 slits are the most ancestral states.

Possible concerted evolution among the different traits was further investigated using reversible jump Markov Chain Monte Carlo searches. For several character combinations, models of dependent evolution received high statistical support. Of the 66 possible correlations (**Figure 4**), we found two correlations with very strong support (LogBayes Factor >10), 19 with strong support (LogBayes Factor 5 to 10), 11 with some support (LogBayes Factor >2) and 34 with weak support (LogBayes Factor <2).

Because of the fact that only two indehiscent berries were included (character 6), and almost all included capsules dehisced (character 8), correlations found among these characters were excluded. Correlations found for fruits which are united at the base or base and apex (character 10) were also excluded due to too small sampling size, resulting in five comparisons with either very strong or strong evidence of correlation, as depicted in **Figure 5**.

As can be seen in **Figure 5**, there is strong evidence for the hypothesis that an epiphytic or lianaceous habit co-evolved with a relatively longer orchid fruit ripening time (**Figure 5A**). There is also evidence for the hypothesis that an epiphytic or lianaceous habit co-evolved with a smaller number of opening slits of orchid fruits (**Figure 5B**). Pendant orchid fruits seem to co-evolved with a preference for growing at intermediate to high temperature, although the interpretation of these two characters is ambiguous (**Figure 5C**). Strong support was also found for the hypothesis that pendant orchid fruits co-evolved with an epiphytic or lianaceous habit (**Figure 5D**). Lastly, orchids that have a deciduous life cycle are more common in cool to intermediate temperatures (**Figure 5E**), which makes sense due to the presence of more strictly defined seasons in temperate regions as compared with the tropics. No direction for valve lignification and the direction of orchid fruits was found (**Figure 5F**).

Figure 4 (next page): **Heatmap of the correlations between each character.** LogBayes factor >10 Very strong evidence (warm color), 5 to10 strong evidence, >2 positive evidence and <2 weak evidence (cold color).


logBayesFactors
15
10
5
0
>10
>5 to 10
>2
<2
Very strong evidence
Strong evidence
Positive evidence
Weak evidence
Life form 1
Strategy 2
Min. temp req. 3
Production 4
Ripening period 5
Type 6
Direction 7
Dehiscence type 8
N. slits 9
United valves 10
Valves 11
Dehiscence zones 12
Life form
Strategy
Min. temp req.
Production
Ripening period
Type
Direction
Dehiscence type
N. slits
United valves
Valves
Dehiscence zones
0
0.988
1.593
3.291
15.806
5.954
8.105
7.697
6.555
–0.948
0.305
0
7.977
1.42
3.067
5.486
4.137
5.223
3.126
0.685
–2.748
0
0.767
1.501
–1.153
9.396
–1.588
1.547
–1.262
4.498
0
–0.128
3.193
1.756
6.416
1.775
–1.402
–0.097
–1.077
0
8.509
–1.744
4.904
4.223
4.336
0.027
0.924
0
–0.157
13.535
6.517
–0.74
–0.375
6.492
0
5.634
–1.595
6.233
5.059
3.763
0
8.938
1.112
5.321
5.569
0
–0.666
–3.235
0.012
0
–2.002
0
0.917
1.287
0

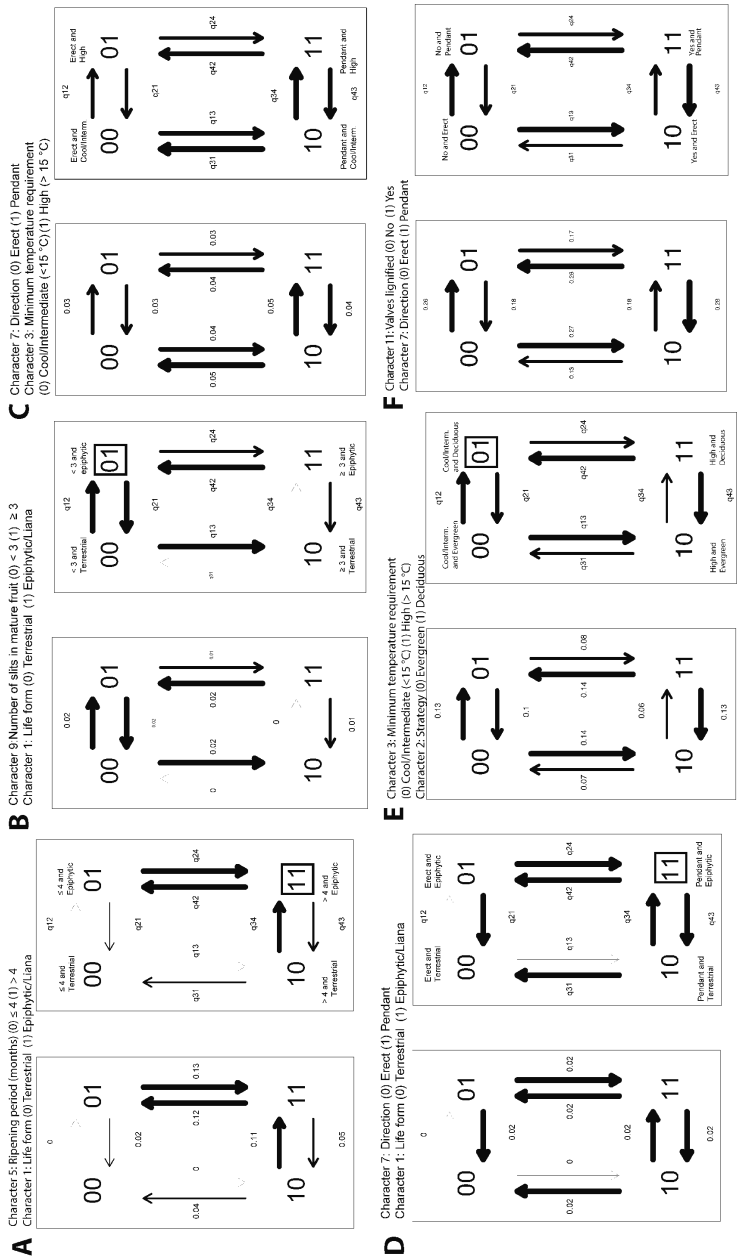


Figure 5: **Ancestral state reconstructions of selected morphological orchid fruit characters from stochastic mapping analyses based on joint sampling (10,000 mapped trees).** Arrows represent transitions between states and numbers represent the estimated number of evolutionary changes with proportion in parenthesis and the time spent in each state. Posterior probabilities (pie charts) are mapped in a random stochastic character map. (A) Character 5 vs. 1, (B) Character 9 vs. 1, (C) Character 7 vs. 3, (D) Character 7 vs. 1, (E) Character 3 vs. 2 and (F) Character 11 vs. 7. Character combinations indicated with a square are interpreted as having co-evolved.

Discussion

Transcriptome analyses of *Erycina pusilla* fruits

In contrast to orchid floral transcriptomes, not many orchid fruit transcriptomes have been published yet. We therefore generated transcriptomes of fruits of *E. pusilla* sampled at different developmental stages in order to discover more about gene networks involved in orchid fruit development. We used two different strategies for data analysis.

First of all, the candidate unigenes that most likely had the longest ORFs (Open Reading Frames) or same Trinity cluster (e.g. DNXXXX_c0_g1) were identified and then filtered by their fragments per kilobase per million mapped base pairs of sequenced (FPKM) values. The high number of unigenes retrieved could not be analyzed with the computer facilities available though. After assembly, we were therefore not able to select the longest transcript from each locus as a unigene for subsequent annotation and thus could not reduce redundancy and potential assembly errors for all the unigenes. The *de novo* assembled unigenes and related groups therefore did not provide very specific information about possible gene networks involved in orchid fruit formation.

Aligning the Ilumina reads to an annotated reference genome, in our case *P. equestris*, downloaded from Plaza4.0 (Van Bel *et al*., 2018) turned out to be more effective as more than 75% of the *E. pusilla* reads could be aligned to this reference. With a Time-Course Expression Analysis, we could detect genes differentially expressed during development of the fruits of *E. pusilla*. In particular the genes involved in the lignin pathway (**Figure S3C** and **Table S4**), which probably also include genes involved in suberin and cutin biosynthesis (Taylor-Teeples *et al*., 2015), are interesting for further analysis. Expression differences in these genes may explain the variation observed in character state 12 (**Figure 2** and **3**), including the presence of a peculiar waxy layer in *E. pusilla* fruits (**Chapter 4**), compared to different lignification patterns in other fruits.

Aligning reads against a reference genome turned out to be more promising and a first analysis revealed putative candidate genes involved in seed formation, pollen tube development and lignification of orchid fruits.

Ancestral character state analysis of orchid fruits

A first important finding of our study was that orchid fruit orientation (erect or pendant), opening (dehiscent or indehiscent) and lignification of the valves are phylogenetically informative within the orchid family. The pendant fruits in this study evolved from an erect one and indehiscent fruits evolved from dehiscent fruits multiple times. Non-lignified fruits seem the most ancestral, from this state lignified endocarp evolved first, followed by lignified exocarp.

A second important finding was that the orchid fruits that we investigated thus far showed that lignification in the dehiscence zone had a positive correlation with temperature. It is not clear yet if lignification of the dehiscence zone is most prominent at a high temperature or intermediate to cool temperature.

Thirdly, the majority of the species with erect orchid fruits displayed lignification

in the dehiscence zone. The hypothesis that lignification could be an adaptation to fast fruit development implies this character state to be present in orchids coping with relatively short seasons suitable for seed dispersal, cannot yet be answered with the current dataset. Expanding the species sampling might shed more light on this.

For the ancestral character state analysis, we did not discriminate between valve lignification in either the endocarp or exocarp layer, but we clearly saw that only in *Bletilla striata*, *Bulbophyllum lasianthum* and *Coelogyne dayana*, all species belonging to subfamily Epidendroideae, the exocarp layer becomes lignified during fruit development. Lignification of the endocarp occurred in a single species of the Apostasioideae, none of the species of the Vanilloideae, the majority of the species sampled of the Cypripedioideae and Epidendroideae, and in all species sampled from the Orchidoideae (**Table S5)**. It remains to be seen if these trends remain standing once the sampling is expanded. Based on the Genera Orchidacearum series edited by Pridgeon *et al*., of the two genera in Apostasioideae, both were sampled (100%), of the 15 genera in Vanilloideae, one was sampled (6.6%), of the 5 genera in Cypripedioideae, one was sampled (20%), of the 200 genera in the Orchidoideae, 6 were sampled (3%) and of the 650 genera in the Epidendroideae, 21 were sampled (3.2%). Increase in sampling size should be particularly focused on genera with both terrestrial and epiphytic species such as *Cymbidium* and *Malaxis*, and fruits of the enigmatic underground flowering genus *Rhizanthella*.

Another observation is that some orchid fruits do dehisce and shatter their seeds, while not showing any sign of lignification in either the dehiscence zone or valves (e.g. *Apostasia wallichii*, *Barkeria scandens*, *Bulbophyllum phalaenopsis*, *Dendrobium* sp. and *Vanilla planifolia*). Whether these fruits develop a dehiscence zone consisting of small cells similar to *E. pusilla* and *Oncidium flexuosum* (Mayer *et al*., 2011;Dirks-Mulder *et al*., 2019) remains to be investigated. Staining tissue sections of fruits of these orchid species with toluidine blue O at different developmental stages, as has been done for *E. pusilla*, *Epipactis hellborine* and *Cynorkis fastigiata*, and staining entire fruits with phosphotungtic acid (PTA) together with X-ray computed microtomography (micro-CT) is needed to visualize the possible presence of cuticle-like layers.

For the ancestral character state analysis of orchid fruits we can thus far conclude that an epiphytic or lianaceous habit and fruit ripening period of more than four months clearly coevolved in orchids, similar to an epiphytic or lianaceous habit with a smaller number of opening slits of orchid fruits, a pendant orientation of orchid fruits with a preference for growing at intermediate to high temperatures, a pendant orientation of orchid fruits with an epiphytic or lianaceous habit and lignification of the valves in both pendant and erect orchid fruits. Future comparative transcriptome analyses of orchid fruits may reveal which developmental genes drive this morphological diversity.

Supplementary figures and tables

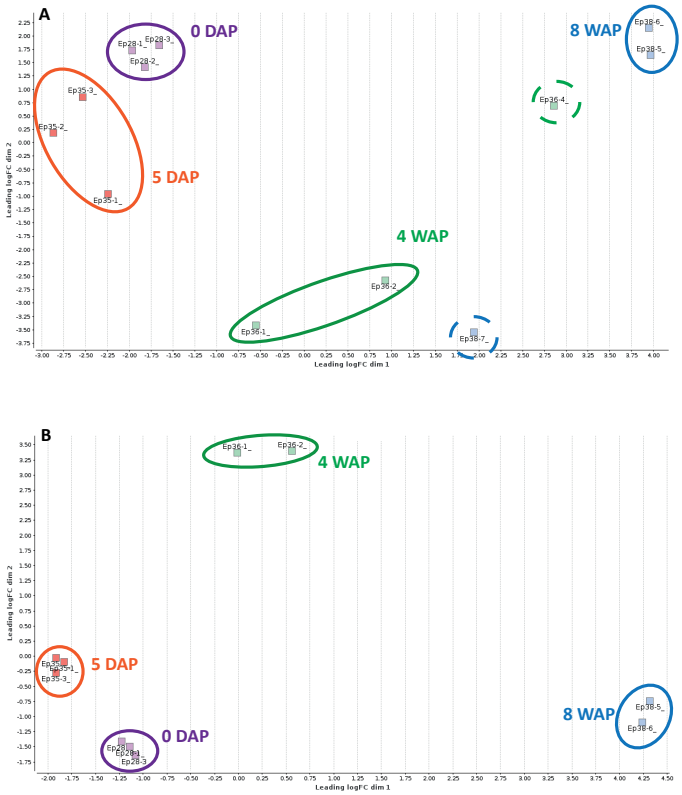


Figure S1: MDS plot resulting from the RNAseq assembly. (A) All samples included. (B) Two deviating samples excluded. DAP, days after pollination; WAP, weeks after pollination.

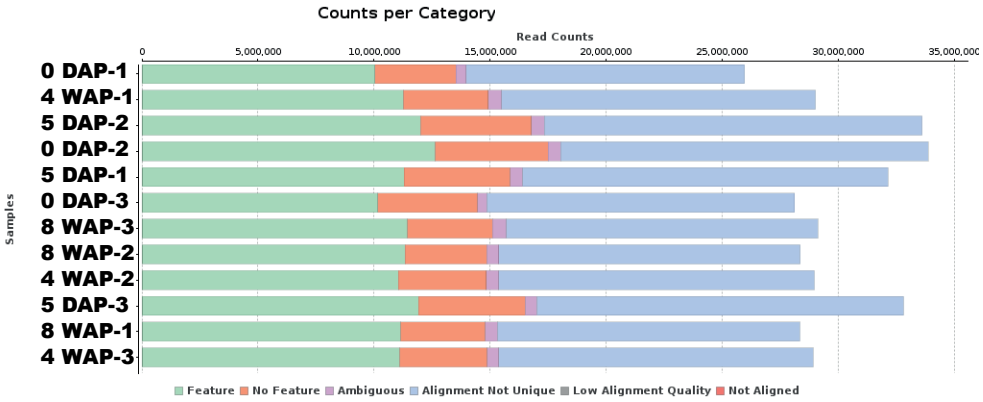


Figure S2: **Read counts per category split up over the different fruits samples analyzed resulting from the RNAseq alignment against the *P. equestris* genome.**

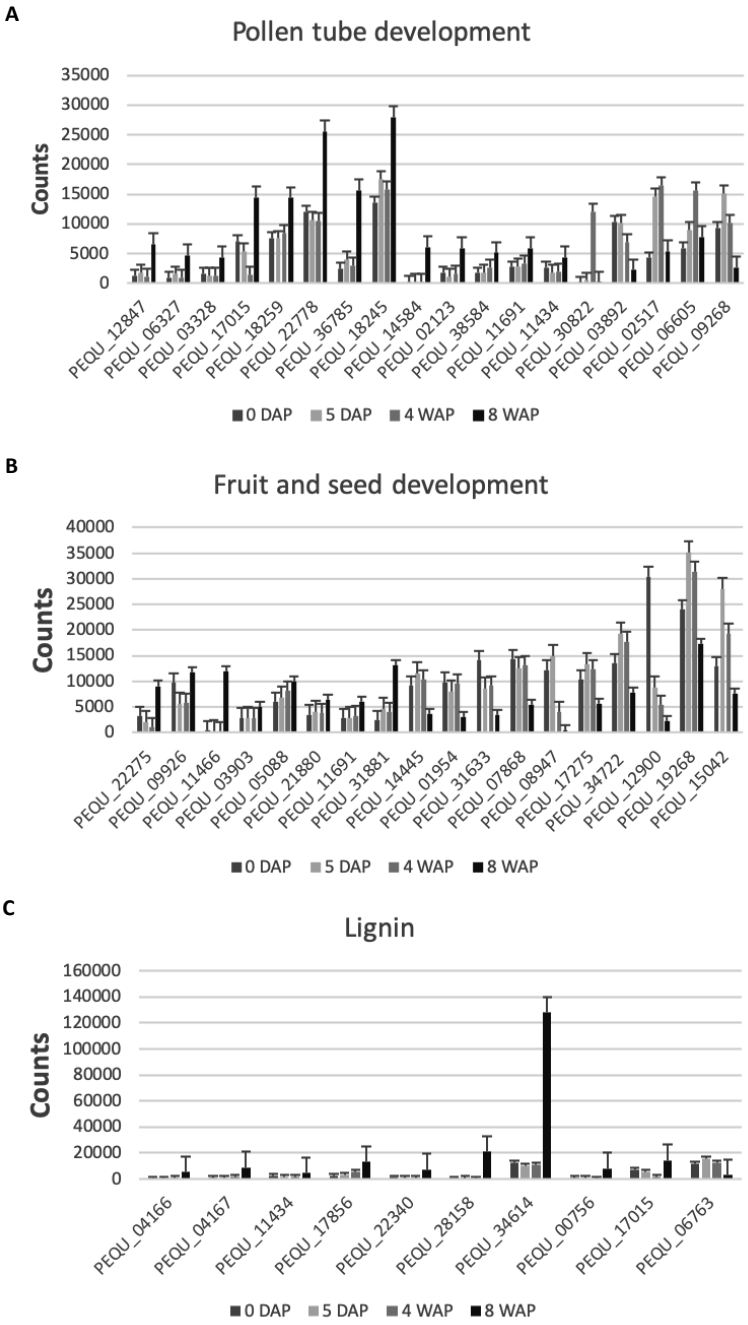


Figure S3: **Fruit specific expression patterns of genes involved in (A) Pollen tube development, (B) Fruit and seed development and (C) Lignification.** Each graph shows the counts of the *E. pusilla* reads aligned against the homologous *P. equestris* gene. DAP, days after pollination; WAP, weeks after pollination. The error bars represent the standard deviation.

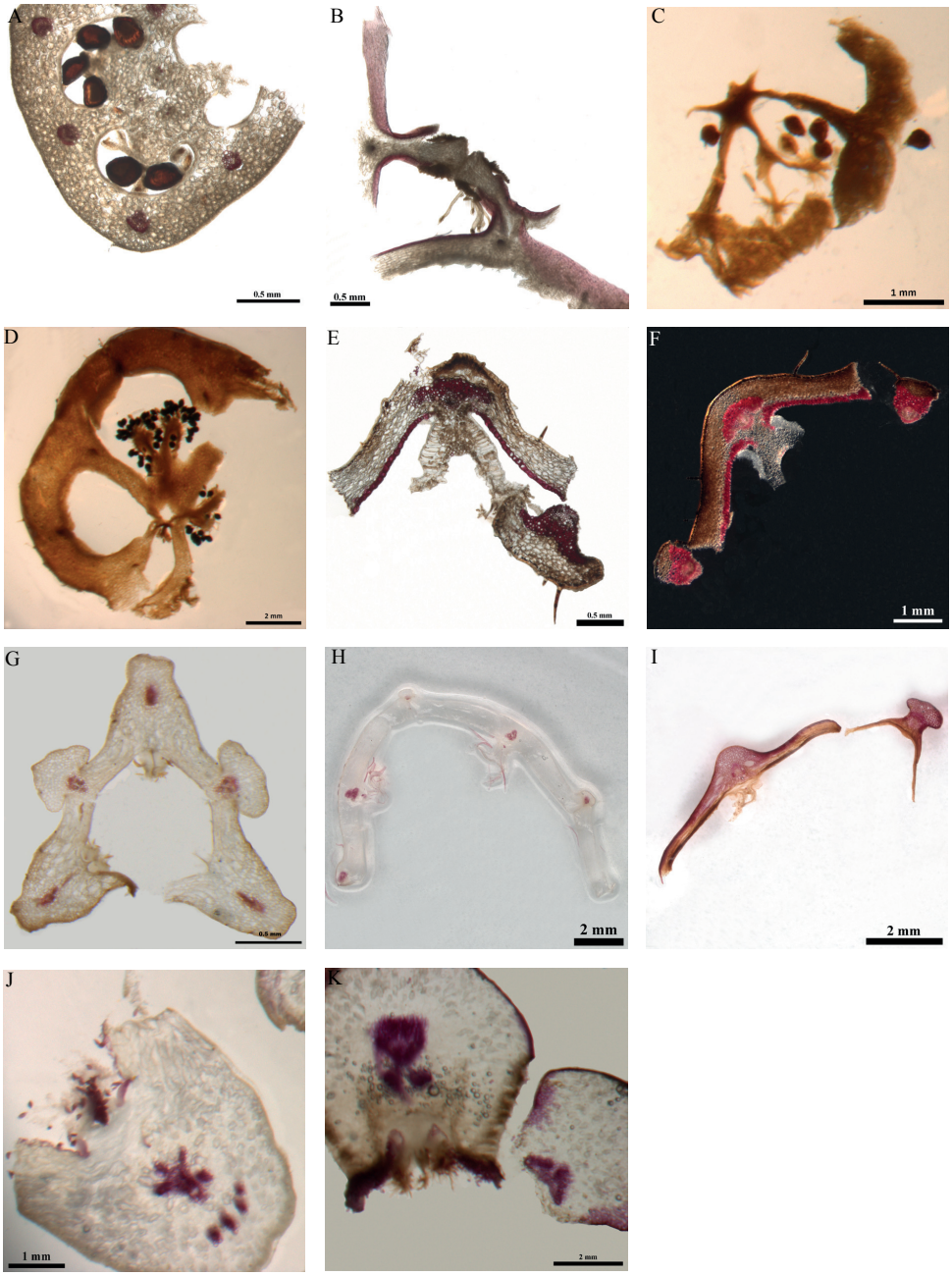


Figure S4: **Phloroglucinol staining of fruit cross-sections. (A)** *Apostasia wallichii*, **(B)** *Neuwiedia veratrifolia*, **(C)** *Neuwiedia zollingeri* var. *javanica*, **(D)** *Neuwiedia zollingeri* var. *singapureana*, **(E)** *Paphiopedilum delenatii*, **(F)** *Paphiopedilum philippinensis*, **(G)** *Angraceum*, **(H)** *Barkeria scandens*, **(I)** *Bletilla striata*, **(J)** *Bulbophylum grandiflorum*, **(K)** *Bulbophylum lasianthu*m. Scale bars: 0.5 mm (A, B, E, G), 1 mm (C, F, J) and 2 mm (D, H, I, K).

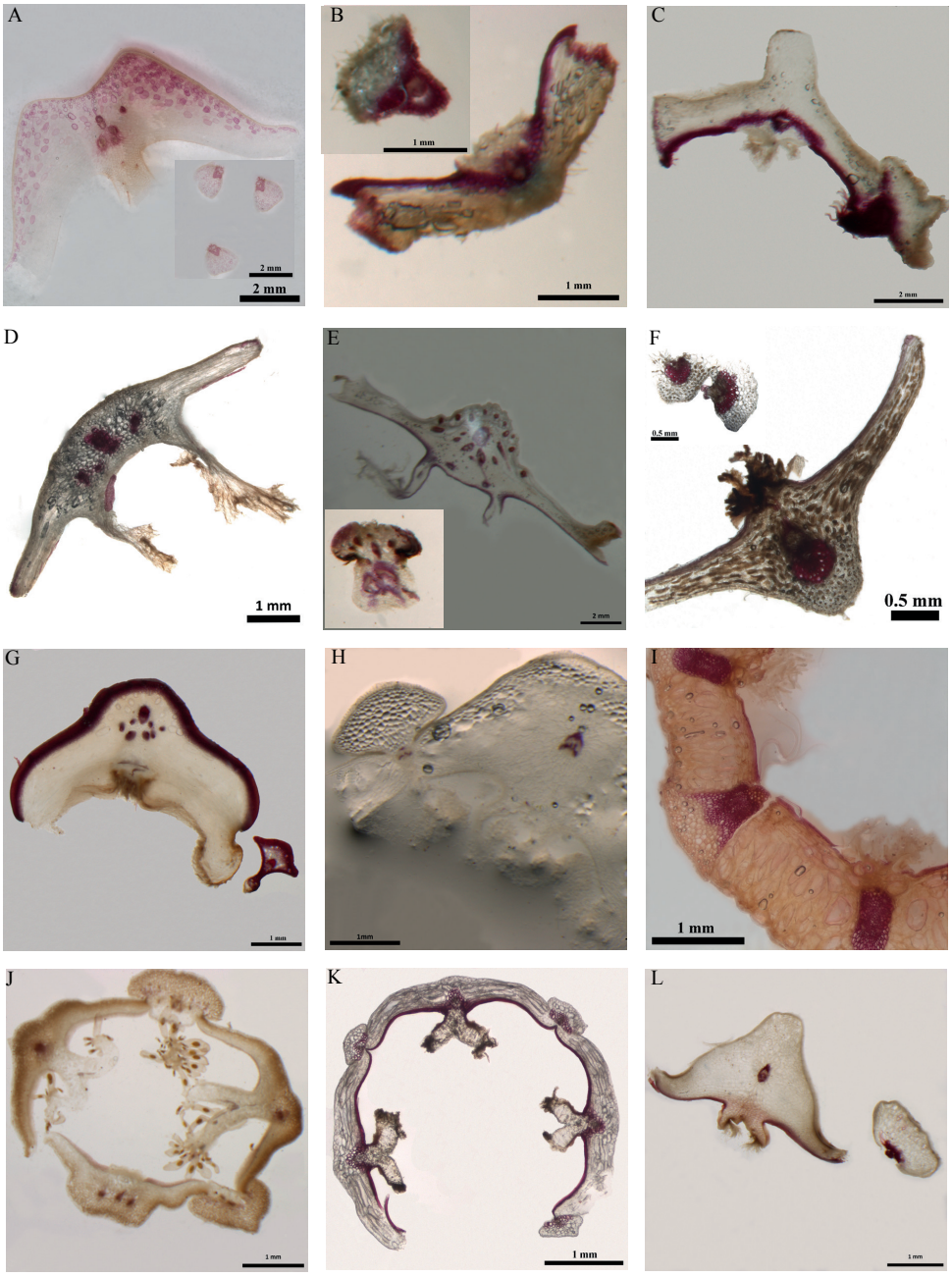


Figure S5: **Phloroglucinol staining of fruit cross-sections. (A)** *Brassavola nodosa*, **(B)** *Calanthe brevicornu*, **(C)** *Calanthe sieboldii*, **(D)** *Calopogon tuberosus*, **(E)** *Catasetum planiceps*, **(F)** *Cephalanthera longifolia*, **(G)** *Coelogyne dayana*, **(H)** *Dendrobium*, **(I)** *Eria albutea*, **(J)** *Limodorum abortivum*, **(K)** *Liparis viridiflora*, **(L)** *Oeceoclades maculat*a. Scale bars: 0.5 mm (F), 1 mm (B, D, G-L) and 2 mm (A, C, E).

Figure S6: **Phloroglucinol staining of fruit cross-sections**. **(A)** *Phalaenopsis*, **(B)** *Phymatidium delicatulum*, **(C)** *Pleurothallis*, **(D)** *Encyclia*, **(E)** *Anoectochilus*, **(F)** *Ophrys*, **(G)** *Orchis militaris*, **(H-I)** *Sarcoglottis*, **(J)** *Spiranthes cernua*, **(K)** *Vanilla planifolia*, **(L)** *Vanilla pompon*a. Scale bars: 0.5 mm (D-G, J), 1 mm (B, C) and 2 mm (A, H, I, K, L). Arrows (B): lignified dehiscence zone.

Table S1. Recommended computer specifications to run OmicsBox.

Product	Description
Memory (RAM)	64 GB Kingston DDR4 2666MHz ECC-registered (2 x 32 GB)
Drives	Storage drive: 2 TB WD Blue™ 3D NAND 2,5" SSD, (max 560 MB/sR | 530 MB/sW) SSD-station: 2 TB SAMSUNG 970 EVO M.2, PCIe NVMe (max 3500 MB/R, 2500 MB/W)
Processor (CPU)	Intel® Xeon® W-2155 10 processor cores (3,3 GHz, 4,5 GHz Turbo, 13,75M Cache)
Motherboard	ASUS® WS C422 PRO/SE (DDR4 RDIMM, 6 Gb/s, CrossFireX/SLI)
Graphics Card	24 GB NVIDIA TITAN RTX - HDMI, 3x DP - RTX VR Ready
Processor cooling	Corsair H55 Hydro-serie power CPU-cooling

Table S2: Accession numbers of *rbc*L, *mat*K and nrITS sequences used in the phylogenetic analysis for the character state analysis.

	NCBI GenBank accession number		
Species/Phylogenetic marker	***rbc*L**	***mat*K**	**nrITS**
Acianthera_ochreata	AF518038	AY008458	AF366934
Angraecum_sesquipedale	AF074106	AF263621	KX669265
Anoectochilus_roxburghii		KY966708	KY966417
Anoectochilus roxburghii			KR815829
Apostasia_wallichii	HM640552	KC172547	AY557228
Barkeria_whartoniana		FJ238568	AF260170
Bletilla_striata	AF074114	EU490679	AF273334
Brassavola_nodosa		JQ771572	AF260219
Bulbophyllum_cambodianum	KM924495	KM924448	KM924472
Bulbophyllum_lasianthum	JF428026	JF305818	JF428116
Bulbophyllum_lobbii	AF074115	AY368395	AF521074
Bulbophyllum_oblongum	KM924498	KM924475	KM924452
Calanthe_brevicornu	KF852738		
Calanthe_sieboldii	KF296674		AY882613
Calanthe brevicornu		KF852693	
Calopogon_tuberosus	AF264161	AF263635	AF273372
Catasetum_expansum	AF074121	KF660300	KU295260
Cephalanthera_longifolia	FJ454875	KF262096	AY146447
Coelogyne_dayana	matK	AY003879	AF281126
Coelogyne_pulverula	KU877826	KU877846	
Cynorkis_fastigiata	AY381117	MF350008	MF944264
Dendrobium_catenatum		AB847714	KJ881390

Species/Phylogenetic marker	NCBI GenBank accession number		
	*rbc*L	*mat*K	nrITS
Dendrobium_moniliforme	KJ187385	AB847816	
Epipactis_helleborine	Z73707	EU490692	AY154383
Erycina_pusilla	NC_018114: 54886-56328	JN598952	AF350538
Hypoxis_curtissii	KJ773578	KJ772842	
Limodorum_abortivum	JX051376		AY351378
Liparis_viridiflora	matK	AY907174	AY907107
Narcissus bulbocodium	KY992380	JX464569	
Neuwiedia_veratrifolia	AF074200	KC172553	AY557227
Neuwiedia_zollingeri_var._javanica	matK	KC172554	AY557226
Neuwiedia_zollingeri_var._singapureana	LC199503: 57921-59384	LC086542	KY966622
Oeceoclades_maculata	JQ593044	LN831626	KF318917
Ophrys_apifera	AJ542396	HE858501	
Ophrys_sitiaca			AM711736
Ophrys_sphegodes	AP018717: 52206-53639		AY699974
Orchis_militaris	KF997273	KF997352	AY699977
Paphiopedilum_callosum	KP311864	KC692129	JQ929308
Paphiopedilum_delenatii	KX264996	AY368379	JX088548
Paphiopedilum_philippinense	KX755546	KX755566	JQ929341
Paphiopedilum_villosum	KX755529	KX755549	GU993851
Phalaenopsis_amabilis	matK	EU256323	AB217571
Phymatidium delicatulum	nrITS	KR709309	KT709688
Prosthechea_cochleata	KJ773788	KJ773041	AY008545
Sarcoglottis_acaulis	AJ542424	AJ543928	AJ539500
Spiranthes_cernua	AF074229	KM213805	AF301444
Vanilla_planifolia	AF074242	AF263687	AF391786
Vanilla_pompona	NC_036809		EU498164

Table S3: Summary of time course expression results.

Data overview	Features
Total number	29,431
Number after filtering	27,171
Identified differentially expressed	12,410
Significant	5,984
Cluster 1	214
Cluster 2	8
Cluster 3	4,484
Cluster 4	1
Cluster 5	35
Cluster 6	1,118
Cluster 7	4
Cluster 8	100
Cluster 9	20

Table S4: **Heatmap representation of expression profiles of genes involved in pollen tube development, fruit and seed development, and lignin biosynthesis.** The scale for each gene was set to the highest green value. DAP = days after pollination; WAP = weeks after pollination. Colors: No color: PEQU from cluster 7-8-9, orange: PEQU from cluster 2, green: PEQU from cluster 5.

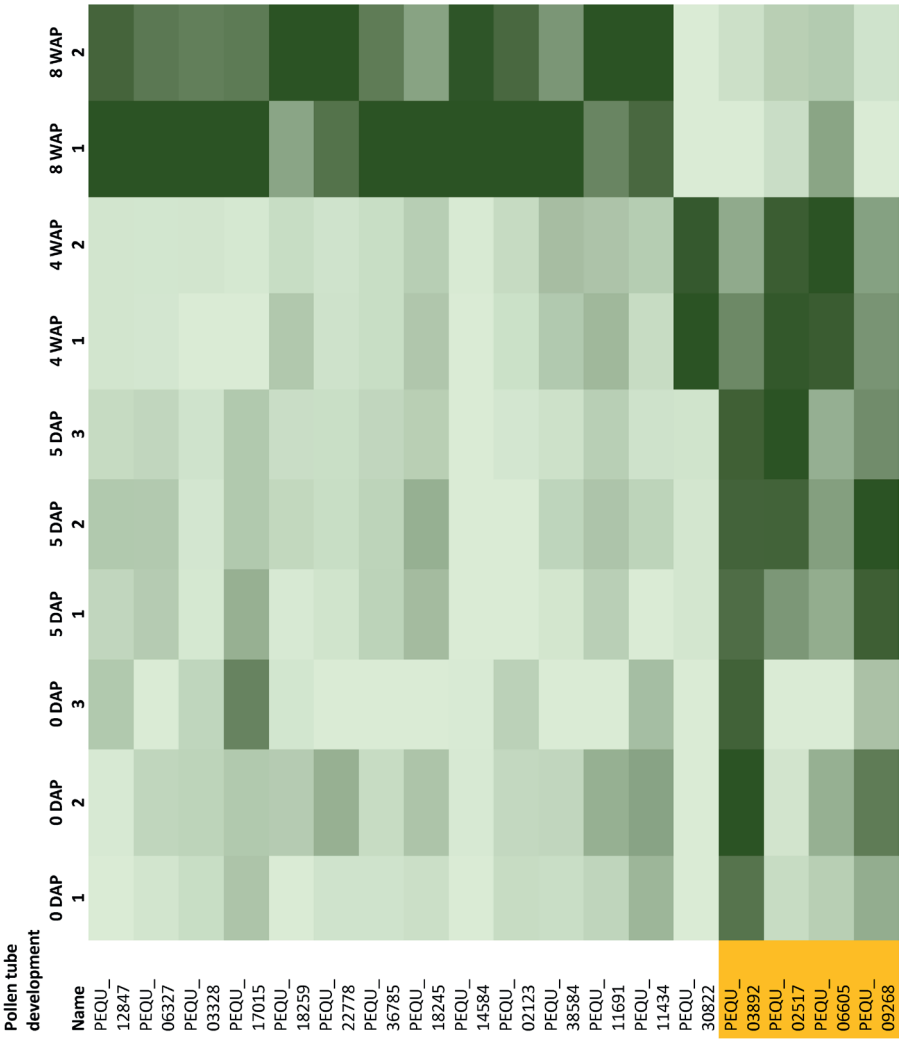


Table S4 continued: Heatmap representation of expression profiles of genes involved in pollen tube development, fruit and seed development, and lignin biosynthesis.

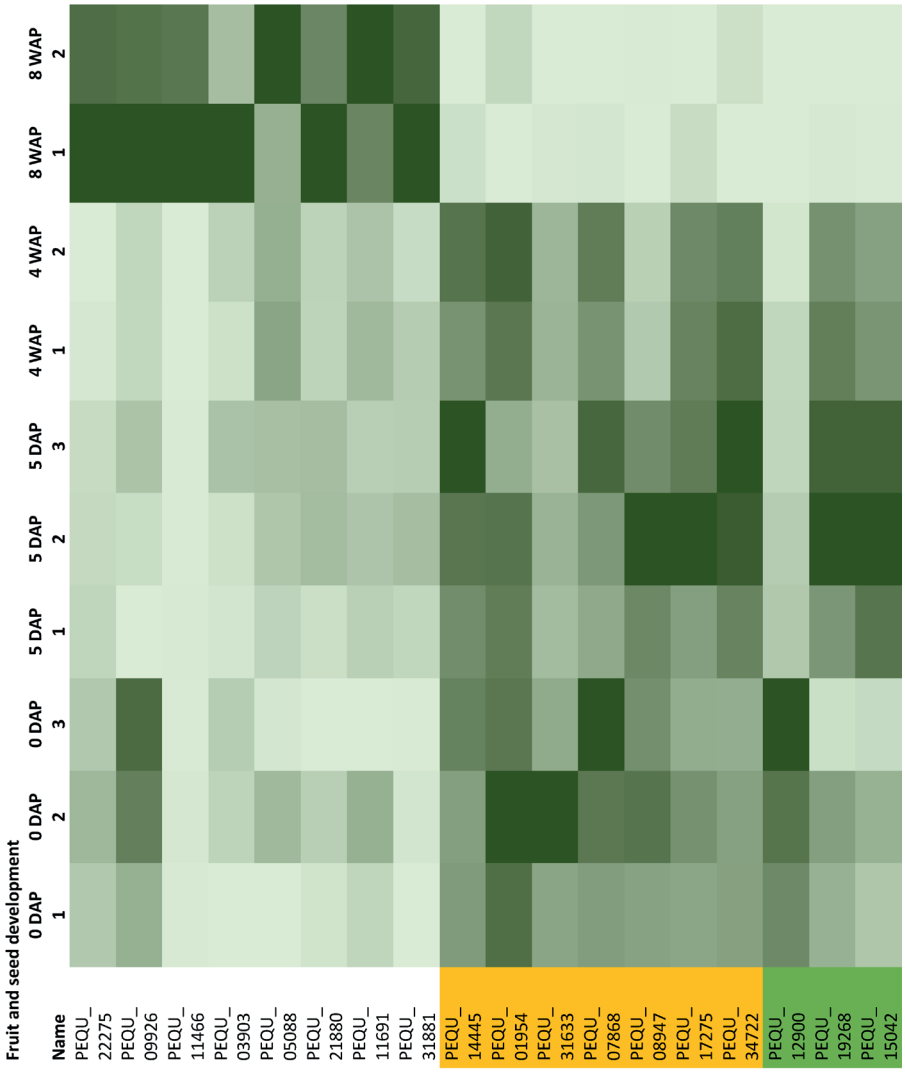


Table S4 continued: Heatmap representation of expression profiles of genes involved in pollen tube development, fruit and seed development, and lignin biosynthesis.

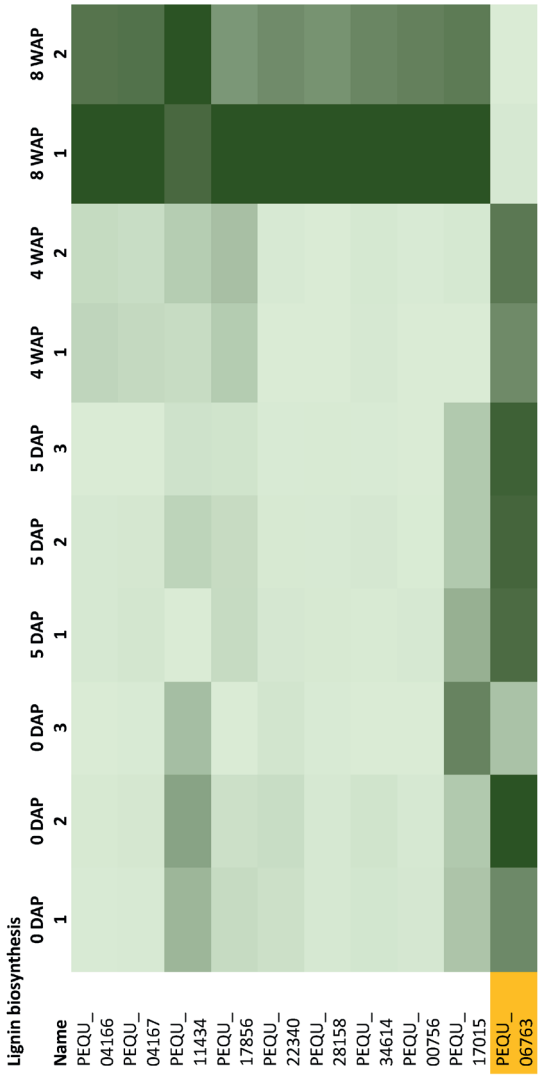


Table S5: **Lignification of the dehiscence zone (DZ) and the valves of 41 orchid species and two non-orchid species.**

Subfamily	Name	Lignification of dehisence zone	Lignification of valves in endocarp/exocarp
Apostasioideae	*Apostasia wallichii*	No	No
Apostasioideae	*Neuwiedia veratrifolia*	No	Endocarp
Apostasioideae	*Neuwiedia zollingeri var. javanica*	No	No
Apostasioideae	*Neuwiedia zollingeri var. singapureana*	No	No
Cypripedioideae	*Paphiopedilum callosum*	Yes	No
Cypripedioideae	*Paphiopedilum delenatii*	Yes	Endocarp
Cypripedioideae	*Paphiopedilum philippinense*	Yes	Endocarp
Cypripedioideae	*Paphiopedilum villosum*	Yes	Endocarp
Epidendroideae	*Angraecum sp.*	No	Endocarp
Epidendroideae	*Barkeria scandens*	No	No
Epidendroideae	*Bletilla striata*	No	Exocarp
Epidendroideae	*Bulbophyllum grandiflorum*	No	No
Epidendroideae	*Bulbophyllum lasianthum*	Yes	Exocarp
Epidendroideae	*Bulbophyllum oreonastes*	Yes	No
Epidendroideae	*Bulbophyllum phalaenopsis*	No	No
Epidendroideae	*Brassavola nodosa*	Yes	No
Epidendroideae	*Calanthe brevicornu*	Yes	Endocarp
Epidendroideae	*Calanthe sieboldii*	Yes	Endocarp
Epidendroideae	*Calopogon tuberosus*	Yes	Endocarp
Epidendroideae	*Catasetum planiceps*	No	Endocarp
Epidendroideae	*Cephalanthera longifolia*	No	Endocarp
Epidendroideae	*Coelogyne dayana (C. pulverula)*	Yes	Exocarp
Epidendroideae	*Dendrobium sp. section Crinivera*	No	No
Epidendroideae	*Epipactis helleborine*	Yes	Endocarp
Epidendroideae	*Eria albolutea*	Yes	Endocarp
Epidendroideae	*Erycina pusilla*	No	Endocarp
Epidendroideae	*Limodorum abortivum*	Yes	Endocarp
Epidendroideae	*Liparis viridiflora*	Yes	Endocarp
Epidendroideae	*Oeceoclades maculata*	No	Endocarp
Epidendroideae	*Phaleanopsis hybrid*	No	Endocarp
Epidendroideae	*Phymatidium delicatulum*	Yes	No
Epidendroideae	*Pleurothallis sp.*	No	Endocarp
Epidendroideae	*Prosthechea cochleata (= Encyclia)*	Yes	No
Orchidoideae	*Anoectochilus papuanus*	Yes	Endocarp
Orchidoideae	*Cynorkis fastigiata*	Yes	No
Orchidoideae	*Ophrys sp.*	Yes	Endocarp
Orchidoideae	*Sarcoglottis sp.*	Yes	Endocarp
Orchidoideae	*Orchis militaris*	Yes	Endocarp
Orchidoideae	*Spiranthes cernua*	Yes	Endocarp
Vanilloideae	*Vanilla planifolia*	No	No
Vanilloideae	*Vanilla pompona*	No	No
Amaryllidoideae	*Narcissus*	No	Endocarp
Hypoxidaceae	*Hypoxis angustifolia*	No	No

Acknowledgements

We thank Alexander Kocyan, Elaine Lopes Pereira Nunes, Rogier van Vugt, Gert-Jan de Waard, and Anneke Wagner for providing us with fruit material.

References

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31**,** 166-169.

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., and Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* 25**,** 25-29.

Beer, J.G. (1863). *Beiträge zur morphologie und biologie der familie der orchideen.* Wien: Carl Gerold's Sohn.

Bollback, J.P. (2006). SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* 7**,** 88.

Cai, J., Liu, X., Vanneste, K., Proost, S., Tsai, W.-C., Liu, K.-W., Chen, L.-J., He, Y., Xu, Q., Bian, C., Zheng, Z., Sun, F., Liu, W., Hsiao, Y.-Y., Pan, Z.-J., Hsu, C.-C., Yang, Y.-P., Hsu, Y.-C., Chuang, Y.-C., Dievart, A., Dufayard, J.-F., Xu, X., Wang, J.-Y., Wang, J., Xiao, X.-J., Zhao, X.-M., Du, R., Zhang, G.-Q., Wang, M., Su, Y.-Y., Xie, G.-C., Liu, G.-H., Li, L.-Q., Huang, L.-Q., Luo, Y.-B., Chen, H.-H., De Peer, Y.V., and Liu, Z.-J. (2015). The genome sequence of the orchid Phalaenopsis equestris. *Nature Genetics* 47**,** 65-72.

Camus, E.G., Lecomte, H., and Camus, A.E. (1921). *Iconographie des orchidées d'Europe et du bassin Méditerranéen.* Paris: P. Lechavalier.

Chomicki, G., Bidel, L.P., Ming, F., Coiro, M., Zhang, X., Wang, Y., Baissac, Y., Jay-Allemand, C., and Renner, S.S. (2015). The velamen protects photosynthetic orchid roots against UV-B damage, and a large dated phylogeny implies multiple gains and losses of this function during the Cenozoic. *New Phytol* 205**,** 1330-1341.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21**,** 3674-3676.

Conesa, A., Nueda, M.J., Ferrer, A., and Talon, M. (2006). maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinformatics* 22**,** 1096-1102.

Di Vittori, V., Gioia, T., Rodriguez, M., Bellucci, E., Bitocchi, E., Nanni, L., Attene, G., Rau, D., and Papa, R. (2019). Convergent Evolution of the Seed Shattering Trait. *Genes (Basel)* 10**,** 68.

Dirks-Mulder, A., Ahmed, I., Uit Het Broek, M., Krol, L., Menger, N., Snier, J., Van Winzum, A., De Wolf, A., Van't Wout, M., Zeegers, J.J., Butot, R., Heijungs, R., Van Heuven, B.J., Kruizinga, J., Langelaan, R., Smets, E.F., Star, W., Bemer, M., and Gravendeel, B. (2019). Morphological and Molecular Characterization of Orchid Fruit Development. *Front Plant Sci* 10**,** 137.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29**,** 15-21.

Dressler, R.L. (1981). *The orchids: natural history and classification.* Cambridge: Harvard University Press.

Dressler, R.L. (1993). *Phylogeny and Classification of the Orchid Family.* Cambridge, UK: Cambridge University Press.

Gandrud, C. (2015). *Reproducible Research with R and R Studio.* New York: Chapman and Hall/CRC.

Givnish, T.J., Spalink, D., Ames, M., Lyon, S.P., Hunter, S.J., Zuluaga, A., Iles, W.J., Clements, M.A., Arroyo, M.T., Leebens-Mack, J., Endara, L., Kriebel, R., Neubig, K.M., Whitten, W.M., Williams, N.H.,

and Cameron, K.M. (2015). Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proc Biol Sci* 282.

Gotz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talon, M., Dopazo, J., and Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 36**,** 3420-3435.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* 29**,** 644-652.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., and Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* 8**,** 1494-1512.

Horowitz, A. (1901). Ueber den anatomischen Bau und das Aufspringen der Orchideen frűchte. *Beihefte zum Botanischen Centralblatt* 11**,** 486-521.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28**,** 1647-1649.

Kocyan, A., and Endress, P. (2001). Floral Structure And Development of Apostasia and Neuwiedia (Apostasioideae) and their Relationships to other Orchidaceae. *International Journal of Plant Sciences* 162**,** 847-867.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9**,** 357-359.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12**,** 323.

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22**,** 1658-1659.

Mayer, J.L.S., Carmello-Guerreiro, S.M., and Appezzato-Da-Glória, B. (2011). Anatomical development of the pericarp and seed of Oncidium flexuosum Sims (ORCHIDACEAE). *Flora - Morphology, Distribution, Functional Ecology of Plants* 206**,** 601-609.

Miller, M.A., Pfeiffer, W., and Schwartz, T. (Year). "Creating the CIPRES Science Gateway for inference of large phylogenetic trees", in: *2010 Gateway Computing Environments Workshop (GCE)*), 1-8.

Omicsbox Bioinformatics made easy. BioBam Bioinformatics.

Pagel, M. (1994). Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters. *Proceedings: Biological Sciences* 255**,** 37-45.

Pagel, M., and Cunningham, C. (1999). The Maximum Likelihood Approach to Reconstructing Ancestral Character States of Discrete Characters on Phylogenies. *Systematic Biology* 48**,** 612-622.

Pagel, M., and Meade, A. (2006). Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *Am Nat* 167**,** 808-825.

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20**,** 289-290.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R news* 6**,** 7-11.

Poinar, G., and Rasmussen, F.N. (2017). Orchids from the past, with a new species in Baltic amber. *Botanical Journal of the Linnean Society* 183**,** 327-333.

Pridgeon, A., Cribb, P., Chase, M.W., and Rasmussen, F.N. (1999a). *Genera Orchidacearum Volume 3: Orchidoideae (Part 2) Vanilloideae.* Oxford: Oxford Universitty Press.

Pridgeon, A.M., Cribb, P., Chase, M.W., and Rasmussen, F.N. (2005). *Genera Orchidacearum Volume 4: Epidendroideae.* OUP Oxford.

Pridgeon, A.M., Cribb, P., Chase, M.W., and Rasmussen, F.N. (2009). *Genera Orchidacearum Volume 5: Epidendroideae.* OUP Oxford.

Pridgeon, A.M., Cribb, P.J., Chase, M.W., and Rasmussen, F. (1999b). *Genera Orchidacearum: Volume 1: Apostasioideae and Cypripedioideae.* OUP Oxford.

Pridgeon, A.M., Cribb, P.J., Chase, M.W., and Rasmussen, F.N. (2001). *Genera Orchidacearum: Volume 2. Orchidoideae.* OUP Oxford.

Pridgeon, A.M., Cribb, P.J., Chase, M.W., and Rasmussen, F.N. (2014). *Genera Orchidacearum Volume 6: Epidendroideae.* OUP Oxford.

R Development Core Team (2018). R: A Language and Environment for Statistical Computing}.

Rambaut, A. (2014). *FigTree. v. 1.4. 2: Tree drawing tool.* [Online]. Available: http://tree.bio.ed.ac.uk/software/figtree/ [Accessed].

Ramirez, S.R., Gravendeel, B., Singer, R.B., Marshall, C.R., and Pierce, N.E. (2007). Dating the origin of the Orchidaceae from a fossil orchid with its pollinator. *Nature* 448**,** 1042-1045.

Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3**,** 217-223.

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 (Oxford, England)* 26**,** 139-140.

Stern, W.L., Gregory, M., and Cutler, D.F. (2014). *Anatomy of the Monocotyledons Volume X: Orchidaceae.* OUP Oxford.

Taylor-Teeples, M., Lin, L., De Lucas, M., Turco, G., Toal, T.W., Gaudinier, A., Young, N.F., Trabucco, G.M., Veling, M.T., Lamothe, R., Handakumbura, P.P., Xiong, G., Wang, C., Corwin, J., Tsoukalas, A., Zhang, L., Ware, D., Pauly, M., Kliebenstein, D.J., Dehesh, K., Tagkopoulos, I., Breton, G., Pruneda-Paz, J.L., Ahnert, S.E., Kay, S.A., Hazen, S.P., and Brady, S.M. (2015). An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature* 517**,** 571-575.

Vaidya, G., Lohman, D.J., and Meier, R. (2011). SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* 27**,** 171-180.

Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van De Peer, Y., Coppens, F., and Vandepoele, K. (2018). PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res* 46**,** D1190-D1196.

Yang, Z., Kumar, S., and Nei, M. (1995). A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* 141**,** 1641-1650.

Yu, G., K. Smith, D., Zhu, H., Guan, Y., and Tsan-Yuk Lam, T. (2016). *ggtree : an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data.*

Ziegenspeck, H. (1936). *In: von Kirchner O, Loew E, Schroter C, eds. Lebensgeschichte der Blutenpflanzen Mitteleuropas.* Stuttgart, Germany: Eugen Ulmer Verlag.