Transcriptome de novo assembly was performed using the short reads assembly program Trinity (Grabherr et al. 2011) (parameter: min\_glue = 2, V = 10, edge-thr = 0.05, min\_kmer\_cov = 2, path\_reinforcement\_distance = 75, group\_pairs\_distance = 250). It first links reads with a certain length of overlap to shape longer fragments without any gap that are called contigs. The reads are subsequently used to align back to contigs. Using paired-end reads, contigs from the same transcript and the distances between are detected. Next, the contigs to scaffolds are linked using N to represent unknown sequences between two contigs. Paired-end reads are used for gap filling of scaffolds. Such sequences are defined as Unigenes. When multiple samples from the same species are sequenced, the Unigenes from the assembly of each sample are further processed through sequence splicing and redundancy removing using sequence clustering software to acquire non-redundant Unigenes of long lengths. In the final step, blastx alignments (e-value < 0.00001) between Unigenes and protein databases, such as the NCBI non-redundant database (NR), Swiss-Prot (www.uniprot.org), KEGG (www.genome.jp/kegg/) and COG (http://www.geneontology.org), were performed, and the best aligning results were used to determine the sequence direction of the Unigenes.

Species and the unigene file names are as follows:

**Species**  **code**

Phalaenopsis equestris PHAlgyTARAAPE

Hemipilia forrestii PHAlgyTBRAAPE

Cymbidium sinense PHAlgyTCRAAPE

Vanilla shenzhenica PHAlgyTDRAAPE

Gastrochilus calceolaris PHAlgyTERAAPE

Holcoglossum amesianum PHAlgyTFRAAPE

Paphiopedilum armeniacum PHAlgyTGRAAPE

Neuwiedia malipoensis PHAlgyTHRAAPE

Cypripedium singchii PHAlgyTIRAAPE

Gastrodia elata PHAlgyTJRAAPE