Available online at www.sciencedirect.com
SCIENCE DIRECT
Plant Science 170 (2006) 426–432
PLANT SCIENCE
www.elsevier.com/locate/plantsci

Expression analysis of the ESTs derived from the flower buds of *Phalaenopsis equestris*

Wen-Chieh Tsai [a,1], Yu-Yun Hsiao [a,1], Shu-Hua Lee [b], Chun-Wei Tung [a], Dan-Ping Wang [a], Hei-Chia Wang [c], Wen-Huei Chen [d], Hong-Hwa Chen [a,b,*]

[a] *Department of Life Sciences National Cheng Kung University, Tainan 701, Taiwan*
[b] *Institute of Biotechnology National Cheng Kung University, Tainan 701, Taiwan*
[c] *Institute of Information Management, National Cheng Kung University, Tainan 701, Taiwan*
[d] *Department of Life Sciences, National University of Kaohsiung, Kaohsiung 811, Taiwan*

Received 3 June 2005; accepted 31 August 2005
Available online 2 November 2005

Abstract

Orchids have profound diversity of specialized pollination and ecological strategies and provide a rich setting for studying evolutionary relationships and molecular biology. The sophisticated orchid flower morphology includes two whorls of perianth segments, three sepals and three petals, one of which is modified as labellum (or lip). In addition, the male and female reproductive organs are fused to form a gynostemium. In order to study gene expression in orchid reproductive organs, a cDNA library of mature flower buds of *Phalaenopsis equestris*, a native diploid species of *Phalaenopsis* in Taiwan, was constructed. A total of 5593 expressed sequence tags (ESTs) from randomly selected clones were identified and characterized. Cluster analysis allowed the identification of a unigene set of 3688 sequences. This abundance of transcripts with predicted cellular roles was functionally characterized by the BLASTX matches to known proteins. Comparison of the relative EST frequencies based on functional categories among floral tissues of five species including *P. equestris*, *Acorus Americanus*, *Asparagus officinalis*, *Oryza sativa* and *Arabidopsis thaliana* was performed. The most highly transcribed genes in *Phalaenopsis* floral buds are those coding for RNA-dependent RNA polymerase of *Cymbidium* mosaic virus, followed by heat shock protein genes. A total 217 putative transcription factor related ESTs were identified. C3H and trihelix families occupied 25% of transcribed transcription factor genes, indicating that the profile of the transcription factors in orchid flower buds is polarized. The extensive analysis of the genes in floral organs adds to the growing repertoire of known plant genes and may also reveal unique features of the reproductive organs of orchids.

Keywords: Expressed sequence tags; Flower bud; Unigene; Gene expression; *Phalaenopsis equestris*; Transcription factor

1. Introduction

The family of Orchidaceae has an estimated 17,000 to 35,000 species, making it among the largest families of the flowering plants [1]. They are known for their diversity of both specialized pollination and ecological strategies and provide a rich subject in which to investigate evolutionary relationships and molecular biology. The versatility and specialization in orchid floral morphology, structure and physiological properties have fascinated botanists and collectors for centuries. Their most astonishing evolution is seen in their reproductive biology. Like most angiosperm flowers, orchids have two whorls of perianth segments. The outer whorl consists of three sepals. Three petals make up the inner whorl, one of them (the labellum or lip) being highly evolved with different size, form, color, and general appearance from the other two. In some genera, such as *Paphiopedilum* and *Phragmipedium*, the labellum has even evolved into a pouch. In all genera, the male and female reproductive organs are fused to form a gynostemium. Modifications of the perianth, androecium, and gynoecium can represent a basis for a variety of floral morphology. The highly sophisticated flower organization offers the opportunity to discover new variant genes and different levels of complexity in the morphogenetic networks [2].

* Corresponding author. Tel.: +886 6 275 7575x65521; fax: +886 6 235 6211.
E-mail address: hhchen@mail.ncku.edu.tw (H.-H. Chen).
[1] These two authors have equal contributions.

0168-9452/$ – see front matter © 2005 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.plantsci.2005.08.029

Based on morphological characteristics, the genus *Phalaenopsis* comprises approximately 45 species that are grouped into nine sections [3]. They have broad geographical distribution and commercial value as floricultural commodities. Two native species of *Phalaenopsis* have been reported in Taiwan, *P. equestris* and *P. aphrodite* var. formosa [4]. Few molecular studies of *Phalaenopsis* orchids' reproductive biology exist because of their long life cycles and inefficient transformation systems. By use of flow cytometry, nuclear DNA contents of *P. equestris* were estimated to be 1.6×10^9 bp ($2n = 2x = 38$), about four-fold that of the rice genome [5,6]. The expansion of plant genomes has mainly been the result of multiplications of retrotransposon repeat sequences [7]. Thus, expressed sequence tags (ESTs) avoid the highly repetitive DNA that makes up the bulk of most plant genomes and offer both a reasonable cost and rapid generation of data that can be exploited for gene discovery and comparative genomics.

So far, ESTs have been analyzed from many plant species [8–16], and the number of reported ESTs is increasing rapidly year by year. In order to analyze gene expression in specific samples from EST data, the numbers of transcripts need to be quantified from the sampled ESTs. In an unbiased cDNA library, the number of ESTs matching a particular gene should reflect the abundance of the corresponding cDNA in the library and the level of its mRNA in the tissue from which the library was derived [17]. EST databases are a valid and reliable source of gene expression data [17–19]. Although more than 26 million ESTs from many species have been deposited in the dbEST division of GenBank, no EST resource has been reported for any orchid species (http://www.ncbi.nlm.nih.gov/dbEST, dbEST 15 April 2005). In this report, a total of 5593 randomly selected cDNA clones corresponding to 3688 unigenes from *P. equestris* flower buds were sequenced and functionally characterized. Results of these studies may lead to a better understanding of the gene expression in the reproductive organs of orchids.

2. Materials and methods

2.1. Plant materials and cDNA library construction

Native species *P. equestris* with red petal and orange lip were grown in greenhouses at Taiwan Sugar Research Institute under natural light and controlled temperature ranging from 23 to 27 °C. Stage IV flower buds (5–10 mm) were collected, frozen immediately in liquid nitrogen and stored at −80 °C until used.

Total RNA samples were extracted from these flower buds using the guanidium thiocyanate method [2]. Poly(A)$^+$ mRNA was prepared using the Poly(A) Quick RNA Isolation kit (Stratagene, La Jolla, CA, USA). The cDNA library was constructed with a cDNA library construction kit (Stratagene) according to the manufacturer's instructions. Synthesized cDNAs were directionally cloned into the λZAPII vector. The cDNA phage clones were excised with the EX Assistant helper phage system (Stratagene), and pBluescript SK+ plasmids were recovered.

2.2. DNA sequencing and sequence processing

Plasmid DNAs were purified from overnight cultures with the miniprep kit (Geneaid, Taipei, Taiwan). More than 6000 sequencing reactions were carried out from the 5′-end using the BigDye^{TM} terminator cycle sequencing kit (Applied Biosystems, Foster city, CA) with a T3 primer 5′-AATTAACCCTCACTAAAGGG-3′. The sequencing analyses were separated on an ABI 377 (Applied Biosystems). Sequence data were analyzed with Sequencher V. 4.1.2 (Gene Codes Corp, Ann Arbor, MI) to remove vector, polyA, adaptor and ambiguous sequences. Residual vector and adaptor sequences were trimmed manually. All sequences were screened for homology to *E. coli* and ribosomal sequences by BLASTN searched and if necessary removed [13]. Fully 5593 flower bud EST sequences have passed these criteria and were submitted to the GenBank databases (GenBank accession no. CB031751–CB035289; CK855526–CK851519). The edited EST sequences were first pre-clustered, and these clusters were then assembled using the PHRAP program (http://www.phrap.org) with at least 40 bases of overlap in which with at least 90% identity.

2.3. Sequence analysis and computer comparison

All sequences were searched for their similarities to known sequences through the use of a BLAST algorithm [20]. Sequences showing BLASTX scores over 50 with known amino acid sequences were taken and the "best hits" annotation was used to represent proteins similar to those encoded by the clones. The BLAST score (bits) used the BLOSUM 62 matrix and Existence 11, Extension 1 Gap costs for BLASTX. Functional assignment of the ESTs was performed against the Munich Information Center for Protein Sequences (MIPS) Functional Catalogue (Funcat) of functionally annotated proteins [21] based on highest scoring of BLASTX results. For comparisons to transcriptomes of other plant floral tissues, ESTs from similar floral development stages of Arabidopsis (5827 ESTs), *Oryza sativa* (6502 ESTs), the basal monocot *Acorus americanus* (7484 ESTs) and the monocot *Asparagus officinalis* (7362 ESTs) were downloaded from public databases. ESTs of Arabidopsis and rice were drawn from http://www.ncbi.nlm.nih.gov, and ESTs of *Acorus* and *Asparagus* from http://www.floralgenome.org [22]. These ESTs were also assembled into unigene sets individually and then functionally characterized as described above.

2.4. Identification of putative transcription factor related ESTs

The protein sequences of Arabidopsis predicted transcription factor were downloaded from AtTFDB (*Arabidopsis thaliana* transcription factor database) of AGRIS (http://arabidopsis.med.ohio-state.edu). AtTFDB contains information on 1583 transcription factors grouped into 42 families [23]. For identification of transcription factor-related ESTs from *P. equestris* flower bud dbEST, the protein sequence set of each predicted Arabidopsis transcription factor family was BLAST