Identification of Floral ESTs from *Calamus manan* Inflorescence Library

KALAIVANI NADARAJAH[1]*, LEONG SHANG JYE[2], CHOONG CHEE YEN and WICKNESWARI RATNAM[2]

[1]School of Biosciences and Biotechnology, [2]School of Environment and Natural Resources Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia - 436 00 (Malaysia).

(Received: January 03, 2009; Accepted: March 05, 2009)

ABSTRACT

Calamus manan is a rattan species that plays an important role in the furniture industry of Malaysia. To assure the availability of cane supplies, rattan plantations and seed nurseries have been initiated. Sex determination of rattan at a young stage is very important for a successful seed nursery to produce more female trees. Expressed sequence tags (ESTs) were generated in order to identify genes found in the inflorescence tissues of *C. manan* especially floral-related genes that may be used in sex determination. A total of 2304 cDNA clones were partially sequenced from 5' end to produce 1179 good quality ESTs with an average size of 400 nucleotides. The sequences were analysed using BLASTX and the EST data set had revealed six floral-related genes, i.e. stigma/stylar cysteine-rich adhesin precursor, flower-specific gamma-thionine precursor (defensin), Men-7, FRIGIDA, proline rich anther-specific protein and Early Flowering 5. *Calamus manan* ESTs that showed a match to floral-related genes were compared with corresponding floral gene sequences from other plants using ClustalW in order to obtain more comprehensive information about sequence similarity. In general, the alignment scores obtained were between 21 to 91% identity. The ESTs that are floral-related genes will be used in developing molecular markers such as CAPS (cleaved amplified polymorphic sequence) that may be used for sex determination in *C. manan*.

Key words: *Calamus manan,* furniture industry, ESTs, flowering genes.

INTRODUCTION

Calamus manan is a hard, compact and single trunk, large diameter rattan that can grow to a height of 100 m (Witono 1995). It can be found on the slopes of the dipterocarpaceae forest though it is able to grow equally well on plains. This particular breed of rattan is used extensively in the furniture industry (Lim 1992) and therefore a continuous supply is needed to support the industry. The cultivation of *Calamus manan* has always been via seeds. The rattan has also been cultivated via wildings, suckers or tissue culture. Vegetative propagation, micropropagation and tissue culture techniques have been used to generate rattan plantlets in large scale (Aziah 1992; Dransfield *et al*. 1992). However although the tissue culture technique is able to generate many plantlets, this technique is not widely used by the cultivators due to the high cost.

In order to generate sufficient rattan plants to support the industry, it is essential to have a good male to female ratio of plants in the plantation (Aminuddin 1995; Jaya Kumar *et al*. 2005). This however is a bit of a problem as the male and female plants can not be distinguished without the floral organs and the inflorescence is only produced when the plant is about 5 years old (Darus & Ab. Rasip 1989; Manokaran 1985). Therefore it is necessary for a method to be devised for early identification of male and female species to assist in the planning of a suitable planting ratio.

The expression of flowering specific genes are important in the study of the development of floral organ. These genes may be used as candidate genes in the breeding program to generate genetic markers. These genetic markers may be used in differentiating between the male and female plants. The study of floral gene expression has been conducted in *Arabidopsis thaliana* (Bowman *et al.* 1989) and *Antirrhinum majus* (Schwarz-Sommer *et al*. 1990).

These studies have shown that the genetic mechanism that determines organ identity is conserved amongst angiosperms. This mechanism has been studied extensively in *Arabidopsis* and *Antirrhinum*, as the ABC model in flower development (Coen & Meyerowitz 1991). According to this model there are 3 classes of homeotic genes (A, B and C) that act individually or in combination to differentiate the 4 organs; stamen, petal, carpel and sepal.

To date, there have been no genetic studies on *C.manan* especially with regards to sex determination. There have also been no reports on marker development for sex determination in *C.manan*. Here we have developed a total of eight cDNA libraries, four for each sex covering the four developmental stages of flowering in *C.manan* with the hope of identifying flowering related genes that may be developed as markers for early detection of sex in *C.manan*. In this paper we report some of the ESTs that have shown identity to flowering genes.

MATERIAL AND METHODS

cDNA library

RNA was extracted from the four different flowering stages. The RNA was used in the synthesis of the cDNA using the Creator™ SMART cDNA Library Construction Kit (Clontech). The synthesised cDNA was later inserted into pSPORT1 and transformed into *E. coli ElectroMAX DH10B* (Invitrogen, USA) via electroporation. This culture was grown and colonies carrying the cDNA were cultured and duplicated into 96 well plates. These 96 well plate libraries were made to represent different stages of flowering in the male and female flowers separately.

Selection of cDNA for sequencing

Two libraries were selected for sequencing purposes ECM04 (male) and ECF03 (female). In each library, 1152 clones were selected randomly for sequencing. These clones were cultured in 96 well plates containing LB-ampicilin (100µg/ml) overnight at 37°C. The plasmids from these clones were extracted and sent for sequencing at the Genomics Laboratory at UKM-MTDC.

Processing of sequences

The sequences were processed via *PHRED* (http://www.phrap.com/phred/) to eliminate bad sequences and to generate sequences in *FASTA* format. The sequences were then imported into *stackPACK*™ v2.1 (http://www.sanbi.ac.za) through *stack_Import*. The parameters for analysis were set at default.

Homology search

The EST sequences that were processed were then compared to other protein sequences in the NCBI site (http://www.ncbi.nlm.nih.gov) through BLASTX (Altschul *et al.* 1990), at default settings. The BLASTX results are given in a table through *parse-blast*.

Classification into functionality groups

The ESTs were classified into four categories based on sequence identity with other protein sequences. The default setting of the EST in BLASTX was 10. Therefore an EST is taken as showing low level homology if the E value is $e^{-10} < E \geq^{-10}$ and score < 100. When E > 10, the EST is considered to have no sequence homology to any proteins in NCBI database.

Classification into protein functionality groups was done using the MIPS (http://mips.gsf.de/proj/funcatDB/search_main_frame.html) database. The MIPS site contained 29 functional categories. We later modified this to only include seven categories; metabolism, synthesis and protein processing, organisation and cell function, defence and stress management, flowering, growth and unclassified proteins.

Comparative studies of flowering genes

ESTs of *C. manan* that were comparable to flowering genes were identified and compared

against flowering genes from other plant species. The EST sequences were converted into protein sequences in their six open reading frames through the use of *Transeq* (http://www.ebi.ac.uk/emboss/transeq/index.html). Once converted the protein sequences are then compared against proteins of other flowering genes through the use of *ClustalW Multiple Sequence Alignment* (http://www.ebi.ac.uk/clustalw). The parameters for the comparison studies in *ClustalW* were at default settings.

From the 6 open reading frames that are generated, the most suitable protein sequence is selected for the comparison studies where the conserved regions and degree of identity will be determined. The selected ORF should contain the methionine codon at the 5" end and the stop codon at the 3' terminal.

RESULTS

The BLASTX results identified a few genes that were related to flowering. They are the stigma/stylar cysteine-rich adhesin precursor, flower specific gamma thionine precursor, Men-7, FRIGIDA, proline rich anther-specific protein dan Early Flowering 5. Most of these ESTs are significantly identical to flowering related genes with an E < e^{-10}.

Comparison of flowering Genes

The *ClustalW* software was used to compare the translated sequences of *C. manan* ESTs with known floral genes in plants from the NCBI database. Two contigs, 25 and 39, showed identity to stigma/stylar cysteine-rich adhesin precursor. The longer contig, contig 39 was translated into a protein with 118 amino acids and aligned with the stigma/stilar precursor from *Lilium longiflorum* that is a 113 amino acids long. The comparison showed 91 identical amino acids with a score of 80 (Fig. 1). Motifs for lipid transport, seed storage and trypsin-α amilase inhibitor were found between (IPR003612) amino acids 30-114 of the sequence from contig 39. The level of identity between the sequences compared was 80%.

Table 1: *C. manan* ESTs with identity to flowering related genes

Flowering Related Genes	EST ID
stigma/stylar cysteine-rich adhesin precursor	contig 25 (566 bp), contig 39 (689 bp)
Flower specific gamma thionine precursor	contig 70 (780 bp), contig 71 (737 bp)
Men-7	ECF03_GS001A.G10 (577 bp)
FRIGIDA	ECM04_GS002C.A12 (663 bp)
Proline rich anther-specific protein	ECF03_GS002D.A03 (681 bp)
Early Flowering 5	ECM04_GS001A.C07 (193 bp)

Table 2: The alignment score for flower specific gamma thionine gene in *C.manan* (contig 70- third frame) to *S. lycopersicum*, *C. chinense*, *N. tabacum*, *N. alata*, *C. annuum*, *H. annuus* and *P. hybrida*

		1	2	3	4	5	6	7	8
1	*C. annuum* (57 aa)	X							
2	*C. chinense* (58 aa)	56	X						
3	*H. annuus* (56 aa)	68	52	X					
4	*N. alata* (55 aa)	56	72	58	X				
5	*N. tabacum* (55 aa)	56	74	58	98	X			
6	*S. lycopersicum* (56 aa)	57	82	53	85	87	X		
7	*P. hybrida* (57 aa)	56	60	52	79	87	66	X	
8	Contig 70 (56 aa)	73	50	89	40	40	50	50	X

	....|....| ....|....| ....|....| ....|....|
	10 20 30 40
L. longiflorum	MARS--SAVC FLLLLAFLIG T---ASAITC GQVDSDLTSC
Contig 39 ORF 1	**MAR**FGV**SAVC** FA**LLAAFL**V**G** **T**PPM**ASAITC** **GQVDS**Y**LTSC**
	....|....| ....|....| ....|....| ....|....|
	50 60 70 80
L. longiflorum	LGYARKGGVI PPGCCAGVRT LNNLAKTTPD RQTACNCLKS
Contig 39 ORF 1	IA**YAR**R**GG**TV **PPGCC**T**GVR**G **LNN**A**AKTTPD** RR**TAC**T**CLK**N
	....|....| ....|....| ....|....| ....|....|
	90 100 110 120
L. longiflorum	LVNPSLGLNA AIVAGIPGKC GVNIPYPIRM QTDCNKVR
Contig 39 ORF 1	**LVNPSLGLN**P RLI**AGIPGKC** **GVNIPYPI**SA S**TD**CS**KV**K

Fig. 1: Comparison between the stigma/stylar cysteine-rich adhesin precursor protein of *C.manan* (contig 39) and *L. longiflorum*. Sequences in bold are identical

	....|....| ....|....| ....|....| ....|....|
	10 20 30 40
S. lycopersicum	VTYEVEAQQI -CKAPSQTFP GLCFMDSSCR KYCIK-E-KF
C. chinense	VANGVQGQNN ICKTTSKHFK GLCFADSKCR KVCIQ-EDKF
N. tabacum	VAYEVQARE- -CKTESNTFP GICITKPPCR KACIS--EKF
N. alata	VAYEVQARE- -CKTESNTFP GICITKPPCR KACIS--EKF
C. annuum	FATDMMAEAK ICEALSGNFK GLCLSSRDCG NVCRR--EGF
H. annuus	MGGPLVVEAR TCESQSHKFK GTCLSDTNCA NVCHS--ERF
P. hybrida	AAYETEAGT- -CKAECPTWE GICINKAPCV KCCKAQPEKF
Contig 70 ORF-3	MAYEEVVEAR KCESQSHKFK GTCLSQPPCR NVCIS--EGF
	VAYEVVVAAR ICETESHNFK GLCLSSPPCR NVCIS—EKF
	....|....| ....|....| ....|....| ....|....|
	50 60 70 80
S. lycopersicum	TGGHCSKLQR KCLCTKPCV
C. chinense	EDGHCSKLQR KCLCTKNCV
N. tabacum	TDGHCSKLLR RCLCTKPCV
N. alata	TDGHCSKIRR RCLCTKPCV
C. annuum	TDGSCIGFRL QCFCTKPCA
H. annuus	SGGKCRGFRR RCFCTTHC-
P. hybrida	TDGHCSKILR RCLCTKPCA
Contig 70 ORF-3	PDGDCSKIRR RCFCTKTCQ
	TDGHCSKIRR RCLCTKPCV

Fig. 2: The comparison of the flower specific gamma thionine gene of *C.manan* (contig 70) with *S. lycopersicum, C. chinense, N. tabacum, N. alata, C. annuum, H. annuus* and *P. hybrida.* Sequences in bold are the consensus sequence of the gene. Identical sequences are in bold

	10	20	30	40
A. thaliana 1	VPMVSGIVES	SIKRGMHIEA	LEMVYTFGME	DKFSAALVLT
A. thaliana 2	VPVISGIVES	SIKRGMHIEA	LEMVYTFGME	DKFSAALVLT
A. arenosa	VPMISGIVES	SIKRGMHIEA	LEMVYTFGME	DKFSASSVLT
ECM04_GS002C.A12 ORF1	**VP**VR**SGIVES**	**SIKRGMHIEA**	INFAYT**FGME**	**DKFSA**VP**VLT**

	50	60	70	80
A. thaliana 1	SFLKMSKESF	ER-AKRKAQS	PLAFKEAATK	QLAVLSSVMQ
A. thaliana 2	SFLKMSKESF	ER-AKRKAQS	PLAFKEAATK	QLAVLSSVMQ
A. arenosa	SFLRMSKESF	ER-AKRKAQS	PLAFKEAAAK	QLAALSSVMQ
ECM04_GS002C.A12 ORF1	**SFL**KD**SKE**AT	PSS**AKRKA**NS	GQ**AFKEA**SR**K**	EQS**A**IR**SVM**K

	90	100	110	120
A. thaliana 1	CMETHKLDPA	KELPGWQIKE	QIVSLEKDTL	QLDKEMEEKA
A. thaliana 2	CMETHKLDPA	KELPGWQIKE	QIVSLEKDTL	QLDKEMEEKA
A. arenosa	CMETHKLDPV	KELPGWQIKE	QIVNLEKDTL	QLDKEMEEKA
ECM04_GS002C.A12 ORF1	**CMET**R**KL**EAE	FP**L**EG--**IKE**	RLES**LEKD**KV	EKK**KEMEE**G**A**

	130	140	150	160
A. thaliana 1	RSLSLMEEAA	LAKRMYNQQI	KRPRLSPMEM	PPVTSSSYSP
A. thaliana 2	RSLSLMEEAA	LAKRMYNQQI	KRPRLSPMEM	PPVASSSYSP
A. arenosa	RSISLMEEAV	LAKRMYNQQM	KRPRLSPMEM	PPVASSSYSP
ECM04_GS002C.A12 ORF1	----------	-**AKRM**T**NQ**NN	G**RP**------**M**	**PPV**KAG**SYS**N

	170	180	190	200
A. thaliana 1	IYRDRSFPSQ	RDDDQDEISA	LVSSYLGPST	SFPHRS--RR
A. thaliana 2	IYRDRSFPSQ	RDDDQDEISA	LVSSYLGPST	SFPHRS--RR
A. arenosa	LYLDRSFPSQ	RDEDRDEISA	LVSSYLGPSS	SFPHRSSLRR
ECM04_GS002C.A12 ORF1	**L**AYVS**SFPSQ**	**RD**F**D**RS--**SA**	**LVSSY**PAAAP	PY**PHRS**----

	210	220	230	240
A. thaliana 1	SPEYMVPLPH	GGLGRSVYAY	EHLAPNSYSP	GHGHRLHRQY
A. thaliana 2	SPEYMVPLPH	GGLGRSVYAY	EHLAPNSYSP	GHGQRLHRQY
A. arenosa	SPEYMVPLPP	GGLGRSVYAY	EHLPPNSYSP	GHGQRLPRQY
ECM04_GS002C.A12 ORF1	------**PL**GH	**G**-------V**Y**	GSRS**PN**A**Y**RD	**GH**AYPAEE**QY**

	250	260	270	280
A. thaliana 1	SPSLVHGQRH	PLQYSPPIHG	QQQLPYGIQR	VYRHSPSEE
A. thaliana 2	SPSLVHGQRH	PLQYSPPIHG	QQQLPYGIQR	VYRHSPSEE
A. arenosa	SPSPVHGQRH	PRQYSPPIHG	QQQIPFGLQR	VYRHSPSEE
ECM04_GS002C.A12 ORF1	**SPS**A**VHG**SYP	AAPM**SPP**PY**G**	------**GLQR**	WFGQVT**SEE**

Fig. 3: The comparison of FRIGIDA gene sequence in *C.manan* (ECM04_GS002C.A12) to *A. thaliana* 1 [AAP49810.1], *A. thaliana* 2 [AAX97727] and *A. arenosa*. Identical sequences are in bold

	10	20	30	40
A. thaliana	APKPAPPPAP	KPGPCPSPPK	PP-APTPKPV	PPHGPPPKPA
B. oleracea	KPPPAPAPSP	KPGPSPPPPK	PP-SPVPKPV	PPPAPSPKPS
B. napus	KPPPAPGPSP	KPGPSPSPPK	PPPSPAPKPV	PPPSPSPKPS
ECF03_GS002D.A03 ORF3	MVV**PAP**ASQV	**KPGP**T**P**GTDR	S**P**--------	---G**P**A**PKP**D

	50	60	70	80
A. thaliana	PA----PTPA	PSPK--PAPS	PPKPENKTIP	AVFFFGDSVF
B. oleracea	PP---APSPK	PKPS-PPAPL	PPKPENKTIP	AVFFFGDSIF
B. napus	PPKPPAPSPK	PSPPKPPAPS	PPKPQNKTIP	AVFFFGDSIF
ECF03_GS002D.A03 ORF3	LS--------	----------	-----RQQ**IP**	**A**--------

	90	100	110	120
A. thaliana	DTGNNNNLET	KIKSNYRPYG	MDFKFRVATG	R**F**SNGMVASD
B. oleracea	DTGNNNNLKS	KIKSNYRPYG	MDFPSRVATG	R**F**SNGKVASD
B. napus	DTGNNNNLDT	KLKCNYRPYG	MDFPMGVATG	R**F**SNGRVASD
EcF03_GS002D.A03 ORF3	----------	----------	------LGTG	H**FSN**HHR**ASD**

	130	140	150	160
A. thaliana	YLAKYMG**V**KE	IVPAYLDPKI	QPN------D	**LL**TGVSF**A**SG
B. oleracea	YLSTYLG**V**KE	IVPAYLDQKL	QQN-QLQRSD	**LL**TGVSF**A**SG
B. napus	YLSKYLG**V**KE	IVPAYVDKKL	QQNNELQQSD	**LL**TGVSF**A**SG
ECF03_GS002D.A03 ORF3	**YL**GL**Y**I**GV**TH	A**VPA**GAQRIL	LKR-----PC	**LL**GSHH**FA**LV

	170	180	190	200
A. thaliana	GA**G**YNPTTSE	AANVIPMLDQ	LTYFQDYIEK	VNRLVRQHKS
B. oleracea	GA**G**FDPETSE	SVEVIPMLDQ	LSYFQDYIKR	VK--------
B. napus	GA**G**YLPQTSE	SWKVTTMLDQ	LTYFQDYKKR	MK--------
ECF03_GS002D.A03 ORF3	**GAG**TT**P**G**TSE**	ARV**V**IRLAEG	**L**VT**F**G-----	----------

	210	220	230	240
A. thaliana	QYKLVGLEKT	NQLISKGVAI	VVG**G**SNDLII	TYFGSGAQRL
B. oleracea	--KLVGKKEA	KRIVSKGVAI	VVA**G**GTDLIY	TYFGIGAQHL
B. napus	--KLVGKKKT	KKIVSKGAAI	VVA**G**SNDLIY	TYFGNGAQHL
ECF03_GS002D.A03 ORF3	--**KLVG**FHQS	CHALA**KG**RLA	**V**II**G**PQP**L**MA	**TYF**ITAH**QHL**

	250	260	270	280
A. thaliana	KNDI**D**SYTTI	IADSAAS**F**VL	QLY**G**Y**G**AR**R**I	GVIGTPPL**G**C
B. oleracea	KTDI**D**SYTTL	MADSAAS**F**VL	QLY**G**Y**G**AR**R**I	GVIGTPPL**G**C
B. napus	KNDV**D**SFTTM	MADSAAS**F**VL	QLY**G**Y**G**AR**R**I	GVIGTPPI**G**C
ECF03_GS002D.A03 ORF3	TLAA**D**VPDIV	E**A**RGQHH**F**II	EA**YG**H**G**QL**R**G	-------L**G**H

	290	300	310	320
A. thaliana	VPSQ**R**LKKKK	ICNEELNYAS	Q	
B. oleracea	TPSQ**R**VKDKK	ICDEEINYAA	Q	
B. napus	TPSQ**R**VKKKK	ICNEDLNYAA	Q	
ECF03_GS002D.A03 ORF3	VL**S**L**R**D**K**FAN	VVGRTVLFVQ	A	

Fig. 4: The comparison of proline rich anther specific protein in *C.manan* (ECFO3_ GS002D.A03) with *A. thaliana, B. oleraceae* and *B. napus*. Identical sequences are in bold.

Contig 70 and 71 of *C. manan* showed identity to flower specific gamma thionine precursor (defensin). Contig 70 being the longer of the two (780 bp) was used in the *ClustalW*. The translated sequence of the third frame (56 amino acids) was compared against seven flower specific gamma thionine precursor sequences in plants such as *Solanum lycopersicum*, *Capsicum chinense*, *Nicotiana tabacum*, *Nicotiana alata*, *Capsicum annuum*, *Helianthus annuus* and *Petunia hybrida*. It was observed that the sequence of *C. manan* has the highest level of identity (score 89) with *H. annuus* and the lowest value (score 50) with *C. chinense*, *S. lycopersicum* and *P. hybrida*. In the overall comparison of the sequences, 46 amino acids are identical to the rest of the sequences. This resulted in a 62% identity (in average) when compared against the 59 amino acids (Fig. 2). The gamma thionine motif (IPR008176) was observed between amino acids 9-55 of the third open reading frame of contig 70.

The first open reading frame sequence of EST ECM04_GS002C.A12 (235 amino acid) was translated and compared to the FRIGIDA sequence of *A. thaliana* 1 [AAP49810.1], *A. thaliana* 2 [AAX97727] and *Arabidopsis arenosa* that was 609, 588 and 611 amino acids long respectively. In Table 3 we see that the protein sequence of *C. manan* has sequence identity to the FRIGIDA sequence of *Arabidopsis*. The percentage of sequence identity was 91% (Figure 3). FRIGIDA-like motif (IPR012474) was located between amino acid 6 and 118 of the first open reading frame in ECM04_GS002C.A12.

Table 3: The alignment score for FRIGIDA (ECM04_GS002C. A12 - 1st frame) in comparison to *A. thaliana* 1 [AAP49810.1], *A. thaliana* 2 [AAX97727] and *A. arenosa*

		1	2	3	4
1	*A. arenosa* (278 aa)	X			
2	*A. thaliana* 1 (276 aa)	90	X		
3	*A. thaliana* 2 (276 aa)	91	98	X	
4	ECM04_GS002C.A12 (235 aa)	91	91	91	X

Table 4: The alignment score of proline rich anther specific protein of *C.manan* (ECF03_GS002D.A03 - third frame), with *A. thaliana*, *B. oleracea* and *B. napus*

		1	2	3	4
1	*A. thaliana* (288 aa)	X			
2	*B. napus* (291 aa)	73	X		
3	*B. oleracea* (285 aa)	74	82	X	
4	ECF03_GS002D.A03 (203 aa)	45	52	30	X

Table 5: The alignment score for the Men-7 gene ECF03_GS001A.G10 (ORF3), with *L. henryi* and *S. latifolia*

		1	2	3
1	*L. henryi* (89 aa)	X		
2	*S. latifolia* (129 aa)	74	X	
3	ECF03_GS001A.G10 (174 aa)	29	21	X

The third open reading frame sequence of EST ECF03_GS002D.A03 was translated and the 203 amino acid long protein sequence was compared against the proline rich anther-specific proteins of *A. thaliana* (288 amino acids), *Brassica oleracea* (285 amino acids) and *Brassica napus* (291 amino acids) (Fig. 4). The full length of all these three proteins are 534, 525 and 449 amino acids respectively. The comparison score values for the three species and *C. manan* is provided in Table 4. The *C. manan* protein sequence had low sequence identity with the sequences of *A. thaliana*, *B. oleracea* and *B. napus*. The average identity of these sequences are 36%.

The third frame sequence of *C.manan* EST ECF03_GS001A.G10 was translated and compared against the sequence of Men-7 in *Lilium henryi* and *Silene latifolia*. Table 5 has the scores for the comparison of ECF03_GS001A.G10 with *L. henryi* dan *S. latifolia*. The comparison shows that the sequence of *C.manan* has higher level of identity to *L. henryi*. On the overall, 38 amino acids were identical amongst the sequences compared (Fig. 5). The trypsin-α amylase motif was (IPR003612) observed between amino acids 34-111 of ECF03_GS001A.G10. No comparative studies were conducted on early flowering 5 EST sequence as the sequence length was too short to obtain any significant information.

	10	20	30	40
L. henryi	MAAMKS---L	ATALLVVLLL	RRLPRGLS--	-QNCSAAIGE
S. latifolia	FVAFLS----	---LLVSRLL	LGLDKPMAG-	-PACSDAIPK
ECF03_GS001A.G10 ORF3	MA**A**FR**S**SDRW	LPL**L**V**V**LC**LL**	SAASSDVSSD	AQE**CS**N**A**LVG

	50	60	70	80
L. henryi	LMTCGPYVLP	GGGAAPEQAC	CSGLKAVNHG	--------CL
S. latifolia	LTFCLLYIT-	GGSPAPSDAC	CSGLKTVAST	SKDKNDAVCL
ECF03_GS001A.G10 ORF3	**L**ST**C**LS**Y**VQ-	**GG**AK**AP**TPD**C**	**CSGLK**QLLLK	**S**PK-----**CL**

	90	100	110	120
L. henryi	CETIKDR---	---------I	ISSLPDHCSL	PA--VNCAK-
S. latifolia	CNCLKDR---	LVDLQYQPSL	IASLPDKCSV	SAKLPAISKA
ECF03_GS001A.G10 ORF3	**C**VLI**KDR**DDP	TVGIKFNVTL	AF**SLP**SV**CS**A	P**A**NISECP**K**L

	130	140	150	160
L. henryi	----------	----------	----------	----------
S. latifolia	TDCSKVNPPY	FMMSTNKAIL	KN--------	----------
ECF03_GS001A.G10 ORF3	LNLPPNSTEA	QIFEQGAITT	QVKSNSTSSG	GSTGRSAQAS

	170	180	190	200
L. henryi	----------	----------		
S. latifolia	----------	----------		
ECF03_GS001A.G10 ORF3	SSGNTHHLGG	QRWMGVERVV		

Fig. 5: The comparison of *Men-7* gene from ECFO3_GS001A.G10 with *L. henryi* and *S. latifolia.* Sequence in bold is identical.

DISCUSSION

EST analysis is essential in identifying expressed genes in a specific tissue. Therefore we generated ESTs to study the floral tissues of *C. manan* and to identify genes that are expressed by these tissues. The identification of floral genes are important for us to understand the process of flowering in *C.manan* and at the same time to develop markers that may be used in sex determination in *C. manan.* The ability to identify the sexes is important in designing seed plots for *C. manan.*

From the EST dataset generated, 6 genes related to flowering were identified. They are the stigma/stylar cysteine-rich adhesin precursor, flower specific gamma thionine precursor, Men-7, FRIGIDA, proline rich anther-specific protein and Early Flowering 5. We believe if we had sequenced a larger number of genes, we may have identified a larger number of flowering and floral genes. Three of these gene, stigma/stylar cysteine-rich adhesin precursor, flower specific gamma thionine precursor and FRIGIDA, showed high level of sequence identity to the other plants that were used in the comparison (i.e between 62-91% sequence identity). FRIGIDA and proline rich anther-specific protein however showed low level sequence identity (36 and 21% respectively) though the sequence length obtained for these genes was fairly long and of high quality. There were however regions of sequence identity noted in the sequences compared.

In the analysis conducted in the bud EST of *A. thaliana* (Höfte *et al*. 1993), no flower specific genes were identified from the 234 EST sequenced; 5 flower specific genes were identified from the 1216 ESTs sequenced from the *B. campestris* bud library (Lim et al. 1996). From the comparison between both these reports we can clearly see that the number of ESTs will largely influence the number of floral genes identified. Going by these figures, identifying 6 genes from the 1179 ESTs generated in our studies is reasonable.

According to Park and Lord (2003), the stigma/stylar cysteine-rich adhesin precursor of *L. longiflorum* resembled a lipid transfer protein that was responsible in the binding of the pollen tube. The flower-specific gamma-thionine gene (*defensin*) isolated in this study has been shown to be involved in the defense mechanism in plants. While the FRIGIDA and Early Flowering 5 gene were reported to be involved in the regulation of flowering time. The Men-7 gene is expressed by the tapetum of the anther and it codes for a lipid transfer protein that is responsible we believe in the biogenesis of the anther's membrane (Wirtz 1991). The function of the proline rich anther-specific gene has not yet been clearly defined. However there has been some speculation on its involvement in regulation of development and morphogenesis in plants (Sato *et al*. 2001).

CONCLUSION

These 6 floral ESTs will be used in designing the primers to amplify the genes from the genomic DNA of *C. manan*. These genes will then be used in developing markers that may be used in studying the polymorphism between the sexes in *C. manan*.

ACKNOWLEDGMENTS

We would like to thank the Ministry of Science Technology and Innovation for the Intensified Research Priority Grant (09-02-02-0036-EA132-Sex determination in Calamus manan) and Malaysian Toray Science Foundation Grant (D/39/2000- Establishment of cDNA library for sex determination in Calamus manan) for the grants awarded to supported this research

REFERENCES

1. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D., Basic local alignment search tool. *Journal of Molecular Biology.* **215**: 403-410 (1990).
2. Aminuddin, M., Rattan in Malaysia: conservation status, biodiversity base and strategic programme. In Rao, V.R. & Rao, A.N. (pnyt.). *Bamboo and rattan genetic*

resources and use, p. 45-48. Singapore: IPGRI/INBAR (1995).

3. Aziah, M.Y., Tissue culture of rattans. In Dransfield, J., Wan Razali, W.M. & Manokaran, N. *A Guide to the Cultivation of Rattan*, 149-161 (1992). Kuala Lumpur: FRIM.
4. Bowman, J.L., Smyth, D.R. & Meyerowitz, E.M., Genes directing flower development in *Arabidopsis*. *The Plant Cell.* **1**: 37-52 (1989).
5. Coen, E.S. & Meyerowitz, E.M., The war of the whorls: genetic interactions controlling flower development. *Nature.* **353**: 31-37 (1991).
6. Darus, H.A. & Ab. Rasip, A.G., Flowering and fruiting in *Calamus manan*. *RIC Bulletin* **8**(1/4): 2 (1989).
7. Dransfield, J., Morphological considerations: the structure of rattans. In Dransfield, J., Wan Razali, W.M. & Manokaran, N. *A Guide to the Cultivation of Rattan*, 11-26 (1992a). Kuala Lumpur: FRIM.
8. Dransfield, J., *The rattans of Sarawak*. UK: Kew Royal Botanic Gardens (1992b).
9. Höfte, H., Desprez, T., Amselem, J., Chiapello, H., Caboche, M., Moisan, A., Jourjon, M. F., Charpenteau, J. L., Berthomieu, P., Guerrier, D., Giraudat, J., Quigley, F., Thomas, F., Yu, D.Y., Mache, R., Raynal, M., Cooke, R., Grellet, F., Delseny, M., Parmentier, Y., Marcillac, G., Gigot, C., Fleck, J., Phillips, G., Axelos, M., Bardet, C., Tremousaygue, D. & Lescure, B., An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. *The Plant Journal.* **4**: 1051-1061 (1993).
10. Jaya Kumar, K., Choong, C.Y., Wickneswari, R. & Raja Barizan, R.S., A sex ratio survey of *Calamus manan* planted in a secondary forest. *Proceedings of the Second Regional Symposium on Environment and Natural Resources.* **2**: 59-61 (2005).
11. Lim, C.O., Kim, H.Y., Kim, M.G., Lee, S.I., Chung, W.S., Park, S.H., Hwang, I. & Cho, M.J., Expressed sequence tags of Chinese cabbage flower bud cDNA. *Plant Physiology.* **111**: 577-588 (1996).
12. Manokaran, N.**,** Biological and ecological considerations pertinent to the silviculture of rattans. Dlm. Wong, K.M. & Manokaran, N. *Proceedings of the rattan seminar*. 95-105 (1985). Kuala Lumpur: FRIM.
13. Park, S.Y. & Lord, E. M., Expression studies of SCA in lily and confirmation of its role in pollen tube adhesion. *Plant Molecular Biology* **51**: 183-189 (2003).
14. Raja Barizan, R.S., Phenology of rattans. In Dransfield, J., Wan Razali, W.M. & Manokaran, N. *A Guide to the Cultivation of Rattan*, Kuala Lumpur: FRIM. 39-46 (1992).
15. Saito, A., Yano, M., Kishimoto, N., Nakagahra, M., Yoshimura, A., Saito, K., Kuhara, S., Ukai, Y., Kawase, M., Nagamine, T., Yoshimura, S., Ideta, O., Ohsawa, R., Hayano, Y., Iwata, N. & Sugiura, M., Linkage map of restriction fragment length polymorphism loci in rice. *Jpn. J. Breed.* **41**: 665-670 (1991).
16. Sato, S., Kaneko, T., Nakamura, Y., Asamizu, E., Kato, T. & Tabata, S., Structural Analysis of a *Lotus japonicus* Genome. I. Sequence Features and Mapping of Fifty-six TAC Clones Which Cover the 5.4 Mb Regions of the Genome. *DNA Research.* **8**: 311-318 (2001).
17. Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H. & Sommer, H., Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science.* **250**: 931-936 (1990).
18. Wirtz, K.W.A., Phospholipid transfer proteins. *Annual Review of Biochemistry.* **60**: 73-99 (1991).
19. Witono, J. R., *Calamus manan* Miquel (Rotan manau). *Eksplorasi.* **2**.**3**: 11 (1995).