

Comparison of Expressed Sequence Tags from Male and Female Sexual Organs of *Marchantia polymorpha*

Rie NISHIYAMA,[†] Katsuyuki T. YAMATO, Kenji MIURA, Megumi SAKAIDA, Sachiko OKADA, Kaoru KONO, Masayoshi TAKAHAMA, Takefumi SONE,[‡] Mizuki TAKENAKA, Hideya FUKUZAWA, and Kanji OHYAMA*

Laboratory of Plant Molecular Biology, Division of Integrated Life Science, Graduate School of Biostudies, Kyoto University, Kyoto 606-8502, Japan

(Received 3 April 2000; revised 10 May 2000)

Abstract

A total of 935 expressed sequence tags (ESTs) from male immature sexual organ were determined, of which 600 ESTs were assembled into 110 non-redundant groups, resulting in 445 unique EST sequences. Of these, 244 sequences shared significant similarities to known nucleotide or amino acid sequences in other organisms. The remaining 201 unique sequences showed no significant matches and thus are likely to be novel transcripts. ESTs from male and female immature sexual organs of a liverwort, *Marchantia polymorpha*, were compared to characterize gene expression patterns during sex differentiation. Ninety-nine male ESTs turned out to be common genes found also in the female library. Interestingly, one of the ESTs found only in male shows a significant similarity to the *transformer-2* gene involved in sex determination in *Drosophila*. In female, several unique lectin ESTs were found that are not present in the male library.

Key words: *Marchantia polymorpha*; EST; male sexual organ; sex differentiation; transformer-2

1. Introduction

Systematic cataloging of cDNA clones has been facilitating characterization of gene expression patterns during the morphological development of various organisms. Expressed sequence tags (ESTs) have also been used for gene identification and construction of gene-based maps in conjunction with genome sequencing projects.¹ To yield information concerning the mechanism of sexual differentiation, EST analyses from human adult testis² and mouse prepubertal testis³ have been performed. ESTs from reproductive organs of plants have been reported for *Arabidopsis thaliana*,^{4,5} *Oryza sativa*⁶ and *Brassica campestris*,⁷ but in these plant species the reproductive organs are bisexual. The present EST analysis from immature female sexual organs of a liverwort, *Marchantia*

polymorpha, is to our knowledge the first investigation for sexually heteromorphic plant.⁸

To better understand the sexual differentiation programs in plants, the dioecious plant *M. polymorpha* was chosen for study. *M. polymorpha* is haploid throughout its life cycle except for the time immediately after fertilization. For sexual reproduction, individual gametophytes develop male or female sexual organs, which are morphologically quite simple yet distinct.⁹ Both of them completely lack any rudiments of the opposite sexual organ primordia. The morphological distinctions between sexual organs suggest that expression of the regulatory genes causing the differences between male and female individuals are confined to the respective sexual organ. Furthermore, *M. polymorpha* has heteromorphic sex chromosomes, a Y chromosome for males and an X chromosome for females; these are the smallest sex chromosomes among plants.¹⁰ Consequently, the Y chromosome should code dominant male factors and/or female-suppressing factors, while the X chromosome likely encodes genes for feminization and/or inhibition of masculinization. The advantages of this system should allow the isolation of key regulatory genes involved in sex differentiation and determination of their functions in *M. polymorpha*.

Here we report an EST analysis of immature male sexual organs from *M. polymorpha* and a comparative analysis to the 717 female ESTs described previously.⁸

Communicated by Satoshi Tabata

* To whom correspondence should be addressed. Tel. +81-75-753-6389, Fax. +81-75-753-6127, E-mail: kohyama@lif.kyoto-u.ac.jp

† Present address: Laboratory of Plant Molecular Breeding, Research and Education Center of Genetic Information, Nara Institute of Science and Technology, Ikoma 630-0101, Japan

‡ Present address: Laboratory of Cell Physiology, Department of Biotechnology, Graduate School of Engineering, Osaka University, Suita 565-0871, Japan

§ Abbreviations; EST: expressed sequence tag

¶ The nucleotide sequence data reported in this paper appear in the NCBI, EMBL and DDBJ nucleotide sequence databases with the accession numbers, AU081613-AU082057.

This comparative EST analysis concerning sexual organ-specific gene expression of non-flowering plants will aid in identifying genes characteristic to sexual organs possibly also in other plant species.

2. Materials and Methods

2.1. Plant materials and cDNA library

The male *M. polymorpha* used in this study was originally derived from a single gemma of a thallus in a local wild population and has been maintained on M51C medium.¹¹ Immature male sexual organs of approximately 2 mm in height were harvested from male thalli propagated on rockwool. Isolation of poly(A)⁺ RNA and construction of a cDNA library were performed as described previously.⁸ Insert size of each cDNA clone was estimated by agarose gel electrophoresis after digestion with *Pvu* II.

2.2. Nucleotide sequencing and sequence analysis

Nucleotide sequences of cDNA clones were obtained as described by Nagai et al.⁸ Redundancy of the male ESTs was detected by a contig assembly program, Sequencher™ 3.0 (Gene Codes Co.), with an assembly parameter of 85% identity in 20 bp. Resulting contigs were manually edited. Individual contigs and remaining unique ESTs were searched against the non-redundant nucleotide and amino acid sequence databases of GenomeNet, Japan, using the BLAST algorithm.¹² Sequence similarities were considered significant when the optimized similarity score was 100 or greater at the amino acid sequence level. When no significant similarity at the amino acid sequence level was detected, the optimized similarity score of 200 or greater at the nucleotide sequence level was considered significant. Male and female ESTs tagging the same genes were scored by more than 97% identity in their untruncated nucleotide sequences.

3. Results and Discussion

3.1. Construction of the cDNA library and generation of ESTs from immature male sexual organ

Complementary DNA from poly(A)⁺ RNA derived from immature male sexual organ of *M. polymorpha* was cloned into pBluescript II SK+ (Stratagene) in one orientation so as to obtain cDNA sequences from their 5'-ends. Over 1000 clones with inserts greater than 100 bp were randomly isolated from the cDNA library and were sequenced in a single run from the 5'-end. Removal of sequences derived from ribosomal RNAs and organellar DNAs left 935 ESTs for further analyses.

3.2. Database search

A total of 935 sequences represented 445 unique ESTs because 600 redundant sequences were assembled into

110 independent groups. Redundancy of individual clones in the library turned out to be higher than that of the previously analyzed female library, partly due to a lower cloning efficiency or biased proliferation of clones during incubation just prior to plating. The 445 ESTs with an average length of 505 bp were registered as ESTs through DNA Data Bank of Japan (accession numbers AU081613 through AU082057) and were used for further analyses. Database search was performed for unique cDNA sequences to identify ESTs that are homologous to previously reported genes. Two hundreds and forty-four (54.8%) of the 445 ESTs showed significant amino acid sequence similarities to entries in the public databases, while the remaining 201 ESTs (45.2%) showed no significant matches and appear to represent novel genes. These ESTs sharing similarity with known genes were further classified into two groups, one with similarities to plant sequences (Table 1, A) and the other with matches in non-plant organisms (Table 1, B). One hundred and seventy nine ESTs (73.4%) of the 244 ESTs described above showed sequence similarity to known plant genes. Approximately half (87 of 179) of these ESTs showed more than 70% identity within overlaps longer than 200 bp. The source organism most frequently encountered is *Arabidopsis thaliana*, to which 63 of the *M. polymorpha* ESTs had similarities, while sequences from lower plants such as algae and bryophytes showed similarity to 12 of the *M. polymorpha* ESTs. This bias is obviously due to the few sequences available from the lower plants in comparison to *A. thaliana*.

3.3. Putative gene functions

The 244 male ESTs that showed significant similarity to known sequences were classified by putative gene function as in the female EST analysis, and according to the criteria proposed by Adams et al.,¹³ 85 unique ESTs are involved in gene and protein expression (especially ribosomal proteins) and comprise the largest functional group in the male library (Fig. 1). The ESTs transcribed for metabolism rank next in abundance. Forty-seven of the 244 unique ESTs are similar to hypothetical protein genes or mRNA sequences, which were classified into the "unclassified" or "EST" categories. They are similar to sequences derived from various organisms, the functions of which are yet to be determined.

3.4. Gene families

As ESTs are useful not only in gene identification but also in genetic mapping, it is essential to distinguish closely related members of a multigene family.¹⁴ Here 9 groups of ESTs were found to share over 70% of their amino acid identities, but are highly divergent in their individual nucleotide sequences of the non-coding regions (Table 2). Members in four of the groups (H3.3-like protein, ribosomal proteins L10a, L17 and S18) are well con-

Table 1. Putatively identified ESTs of immature male sexual organ from *Marchantia polymorpha*.

A. ESTs similar to plant sequences

M-EST*	L*	OL*	ID%*	Putative Identification	Source*	Organism*	F-EST*	
M01S009	449	163	70.6	21723 Lambda-PRL1 <i>Arabidopsis thaliana</i> cDNA clone G4C5T7, mRNA sequence	gb	N96665	<i>Arabidopsis thaliana</i>	F01F064
M01-059	651	140	65.0	2352 Lambda-PRL2 <i>Arabidopsis thaliana</i> cDNA clone 49B6T7, mRNA sequence	gb	T14187	<i>Arabidopsis thaliana</i>	
M01-042	629	278	74.5	29536 Lambda-PRL2 <i>Arabidopsis thaliana</i> cDNA clone 202N19T7, mRNA sequence	gb	AA597628	<i>Arabidopsis thaliana</i>	
M01-076	588	240	64.2	30834 Lambda-PRL2 <i>Arabidopsis thaliana</i> cDNA clone 280D8T7, mRNA sequence	gb	AA650720	<i>Arabidopsis thaliana</i>	
M01B079	520	212	66.0	31720 Lambda-PRL2 <i>Arabidopsis thaliana</i> cDNA clone 175H12T7, mRNA sequence	gb	AA712587	<i>Arabidopsis thaliana</i>	
M01F083	517	432	61.8	70 kD peptidylprolyl isomerase	sp	FKB7_WHEAT	<i>Triticum aestivum</i>	
M01D016	631	104	66.3	96GS0863 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa cDNA clone 96GS0863, mRNA sequence	gb	AA752358	<i>Oryza sativa</i>	
M01S080	593	170	67.1	A. <i>thaliana</i> transcribed sequence; clone PAPI875, mRNA sequence	gb	ATTS4842	<i>Arabidopsis thaliana</i>	
M01B047	550	282	62.8	A_IG002N01.18 gene product	gb	IG002N01	<i>Arabidopsis thaliana</i>	
M01H043	524	114	73.7	A_IG002N01.21 gene product	gb	IG002N01	<i>Arabidopsis thaliana</i>	F01P013
M01I039	516	387	33.3	ABA-inducible protein, landform-specific	pir	S58219	<i>Riccia fluitans</i>	
M01G085	549	318	72.6	acetolactate synthase	pir	S22490	<i>Zea mays</i>	
M01-015	611	297	69.7	acidic ribosomal protein P1	sp	RLA1_CHLRE	<i>Chlamydomonas reinhardtii</i>	F01-052
M01H031	476	315	56.2	acidic ribosomal protein P2	sp	RLA2_PARAR	<i>Parthenium argentatum</i>	F01-057
M01U026	506	363	85.1	actin 93	sp	ACT5_TOBAC	<i>Nicotiana tabacum</i>	
M01T080	558	201	65.7	aquaporin-like protein	gb	PAAJ5078	<i>Picea abies</i>	F01E025
M01F068	621	159	75.5	ascorbate peroxidase	prf	2209436A	<i>Gossypium hirsutum</i>	
M01A068	486	270	83.3	ATPase gamma	prf	1407163A	<i>Spinacia oleracea</i>	F010004
M01-084	580	132	77.3	beta-ketoacyl-ACP synthase	gb	RCCBKACPSA	<i>Ricinus communis</i>	F01-080
M01E038	518	291	95.9	caltractin (centrin)	sp	CATR_SCHDU	<i>Scherffelia dubia</i>	F01B112
M01-098	611	267	58.4	casein kinase I, delta isoform like	sp	KCID_ARATH	<i>Arabidopsis thaliana</i>	
M01F094	344	306	91.2	catalase	sp	CATA_PEA	<i>Pisum sativum</i>	
M01A079	650	264	45.5	cell wall protein	gb	ATHAPTR	<i>Arabidopsis thaliana</i>	
M01N016	572	438	89.7	chlorophyll a/b binding protein	pir	S37099	<i>Amaranthus hybridus</i> var.	
M01-049	965	618	77.7	chlorophyll a/b-binding protein precursor	gb	AF094776	<i>Oryza sativa</i>	
M01-069	508	174	79.3	chloroplast FtsH protease	gb	AB017480	<i>Nicotiana tabacum</i>	
M01D063	604	594	71.7	chloroplast ribosomal protein L3	gb	AB010877	<i>Nicotiana tabacum</i>	
M01E093	312	258	77.9	chorismate synthase precursor	sp	AROC_CORSE	<i>Corydalis sempervirens</i>	
M01H092	586	120	68.3	Cicer arietinum epicotyl EST clone Can117-2, mRNA sequence	gb	CAR5971	<i>Cicer arietinum</i>	
M01-109	736	240	83.8	Cks1 protein	gb	ATCKS1SUC	<i>Arabidopsis thaliana</i>	
M01-079	598	198	50.0	class I chitinase	gb	MSU83591	<i>Medicago sativa</i>	
M01S096	528	285	49.5	cystatin	pir	JC4882	<i>Zea mays</i>	F01-095
M01I064	344	342	63.5	cysteine proteinase	gb	CAR9878	<i>Cicer arietinum</i>	
M01B076	649	153	56.9	cysteine-rich extensin-like protein 2 precursor	pir	B48232	<i>Nicotiana tabacum</i>	F01K174
M01E077	254	174	87.9	cytochrome b5	gb	OJE001369	<i>Olea europaea</i>	
M01-077	624	195	46.2	cytochrome c oxidase:subunit=8.2kD	prf	2107179A	<i>Solanum tuberosum</i>	F01Y065
M01B039	570	276	40.2	cytochrome P450	gb	GMC450CP1	<i>Glycine max</i>	
M01D042	650	540	60.0	cytosolic fructose-1,6-bisphosphate aldolase	gb	CAAJ5041	<i>Cicer arietinum</i>	
M01F081	650	342	80.7	defender against cell death 1 homolog	gb	AF051247	<i>Picea mariana</i>	
M01D045	650	318	58.5	early light-inducible protein	pir	JC5876	<i>Glycine max</i>	
M01E035	486	486	91.4	elongation factor 1-alpha	sp	EF1A_VICFA	<i>Vicia faba</i>	F01-109
M01T018	523	126	54.8	endomembrane protein EMP70 precursor isolog	gb	ATU95973	<i>Arabidopsis thaliana</i>	
M01H076	550	123	48.8	endoxylglucan transferase-like protein	gb	ATF6H11	<i>Arabidopsis thaliana</i>	
M01Q078	638	438	41.1	eukaryotic translation initiation factor 5	sp	IF5_PHAVU	<i>Phaseolus vulgaris</i>	
M01Y009	620	162	38.9	F12F1.21 gene product	gp	F12F1	<i>Arabidopsis thaliana</i>	
M01B046	650	258	66.3	F19F24.7 gene product	gb	ATAC003673	<i>Arabidopsis thaliana</i>	
M01G081	379	147	65.3	F19P19.2 gene product	gb	F19P19	<i>Arabidopsis thaliana</i>	
M01I083	376	159	47.2	F20P5.28 gene product	gb	F20P5	<i>Arabidopsis thaliana</i>	
M01A003	521	438	69.9	F27F23.14 gene product	gb	ATAC003058	<i>Arabidopsis thaliana</i>	
M01D019	504	210	47.1	F41I.17 gene product	gb	ATAC004521	<i>Arabidopsis thaliana</i>	
M01-071	650	195	72.3	F8A5.25 gene product	gb	AC002292	<i>Arabidopsis thaliana</i>	F01M158
M01H035	354	117	64.1	F8K4.3 gene product	gb	F8K4	<i>Arabidopsis thaliana</i>	
M01B070	650	285	90.5	ferredoxin	sp	FER_MARPO	<i>Marchantia polymorpha</i>	F01F019
M01D005	650	633	79.1	fructose-bisphosphate aldolase - like protein	gb	ATM3E9	<i>Arabidopsis thaliana</i>	
M01H072	468	174	67.2	glossyl homolog	gb	OSU57338	<i>Oryza sativa</i>	

Table 1. A. Continued.

M-EST*	L*	OL*	ID% ^a	Putative Identification	Source*	Organism ¹	F-EST*
M01E032	392	306	70.6	glutaredoxin	sp	GLRX_RICCO <i>Ricinus communis</i>	
M01M043	643	273	54.9	glutaredoxin	gb	AF047694 <i>Vernicia fordii</i>	
M01-062	649	372	83.9	glyceraldehyde 3-phosphate dehydrogenase, cytosolic	sp	G3PC_ANTMA <i>Antirrhinum majus</i>	
M01C069	650	237	92.4	glyceraldehyde-3-phosphate dehydrogenase	pir	S51836 <i>Pinus sylvestris</i>	
M01Q074	650	117	92.3	glyceraldehyde-3-phosphate-dehydrogenase	gb	SLU96623 <i>Selaginella lepidophylla</i>	
M01H066	650	432	95.8	GTP-binding nuclear protein RAN-1	sp	RANI_ARATH <i>Arabidopsis thaliana</i>	
M01-002	810	351	87.2	histone H2A	sp	H2A_PICAB <i>Picea abies</i>	
M01C030	596	294	85.7	histone H2A	sp	H2A_PICAB <i>Picea abies</i>	
M01-086	468	63	100.0	histone H3	sp	H3_HORVU <i>Hordeum vulgare</i>	
M01A016	575	408	97.8	histone H3.3-like protein	pir	S24346 <i>Arabidopsis thaliana</i>	
M01-010	668	351	100.0	histone H3.3-like protein	pir	S24346 <i>Arabidopsis thaliana</i>	F01-058
M01-014	647	408	100.0	histone H3.3-like protein	pir	S24346 <i>Arabidopsis thaliana</i>	F01-031
M01-083	509	309	99.0	histone H4	pir	HSZM4 <i>Zea mays</i>	
M01-104	582	222	67.6	hydroxypyruvate reductase	gb	D85339 <i>Arabidopsis thaliana</i>	
M01F006	295	153	41.2	hypothetical protein	gb	ATF4D11 <i>Arabidopsis thaliana</i>	
M01Q049	544	246	54.9	hypothetical protein	prf	2405340LJ <i>Arabidopsis thaliana</i>	
M01-090	575	222	90.5	initiation factor 5A-2	sp	IF52_NICPL <i>Nicotiana plumbaginifolia</i>	F01F125 F01K169
M01-085	549	219	75.3	lactoylglutathione lyase	sp	LGUL_LYCES <i>Lycopersicon esculentum</i>	
M01-009	681	180	45.0	LEA protein	gb	CAR224518 <i>Cicer arietinum</i>	F01-010
M01L069	650	528	69.9	LHCI-680, photosystem I antenna protein	pir	S52341 <i>Hordeum vulgare</i>	
M01A064	650	288	40.6	mannose-specific lectin (agglutinin)	sp	LEC_ALOAR <i>Aloe arborescens</i>	
M01F002	310	309	72.8	methionine synthase	gb	MCU84889 <i>Mesembryanthemum crystallinum</i>	
M01-031	651	219	30.1	MtN5 gene product	gb	MTY15371 <i>Medicago truncatula</i>	
M01E060	231	120	67.5	NADH dehydrogenase	gb	ATTS1882 <i>Arabidopsis thaliana</i>	
M01P095	633	309	95.1	NADH-ubiquinone oxidoreductase 20 kD subunit precursor	sp	NUKM_ARATH <i>Arabidopsis thaliana</i>	
M01O068	442	147	49.0	nuclear encoded chloroplast protein CP12	prf	2306420C <i>Pisum sativum</i>	
M01C051	557	168	62.5	oxygen-evolving enhancer protein 2 precursor	sp	PSBP_LYCES <i>Lycopersicon esculentum</i>	
M01D021	650	204	58.8	pathogenesis-related protein 4A	pir	S23799 <i>Nicotiana tabacum</i>	
M01Q036	608	339	53.1	peroxidase	prf	2304179B <i>Medicago sativa</i>	
M01A070	326	113	83.2	Phaseolus vulgaris clone XZT-205 mRNA, partial cds	gb	PVU28922 <i>Phaseolus vulgaris</i>	F01A003
M01Y017	519	225	68.0	phosphoglucomutase	gp	MCU84888 <i>Mesembryanthemum crystallinum</i>	
M01D048	584	255	62.4	photosystem I complex PsaN subunit precursor	gb	AI001278 <i>Zea mays</i>	
M01I057	359	285	77.9	photosystem I reaction centre subunit II precursor	sp	PSAD_CUCSA <i>Cucumis sativus</i>	
M01A040	591	258	79.1	photosystem II 22 kD protein precursor	sp	PSBS_SPIOL <i>Spinacia oleracea</i>	F01H116
M01A051	623	285	82.1	plasma membrane intrinsic protein 1C	sp	WCIC_ARATH <i>Arabidopsis thaliana</i>	
M01T077	600	507	50.3	probable glutathione transferase	pir	S39542 <i>Arabidopsis thaliana</i>	
M01-063	645	450	50.0	probable glutathione transferase	pir	S39542 <i>Arabidopsis thaliana</i>	F01K103
M01-092	535	447	76.5	probable ribosomal protein L13a	gb	AF051212 <i>Picea mariana</i>	F01H041
M01-019	623	486	92.0	probable ribosomal protein L15	gb	AF051207 <i>Picea mariana</i>	
M01-054	646	360	81.7	probable ribosomal protein L31	gb	AF051232 <i>Picea mariana</i>	F01-066
M01-048	562	453	90.1	probable ribosomal protein S15	gb	AF051217 <i>Picea mariana</i>	F01X007
M01-102	650	339	80.5	protein kinase C inhibitor	pir	S45368 <i>Zea mays</i>	
M01E088	287	240	77.5	protein translation factor SUI1 homolog	sp	SUI1_ARATH <i>Arabidopsis thaliana</i>	F01C151
M01W030	270	201	44.8	PSI-H subunit	gb	BRU92504 <i>Brassica rapa</i>	
M01F039	341	150	40.0	putative chloroplast nucleoid DNA binding protein	gb	ATAC005313 <i>Arabidopsis thaliana</i>	
M01-064	635	120	65.0	putative permease	gb	ATAC004481 <i>Arabidopsis thaliana</i>	
M01-051	598	180	91.7	putative ribosomal protein	gb	ATAC005169 <i>Arabidopsis thaliana</i>	F01Q076
M01-011	649	420	94.3	putative ribosomal protein L17	gb	ATAC002332 <i>Arabidopsis thaliana</i>	F01E113
M01-045	593	366	82.8	putative ribosomal protein L35	gb	ATAC004218 <i>Arabidopsis thaliana</i>	F01-006
M01F010	325	177	79.7	putative ribosomal protein L7A	gb	ATAC002535 <i>Arabidopsis thaliana</i>	
M01A083	561	138	58.7	putative ribosomal protein S16	gb	ATT4L20 <i>Arabidopsis thaliana</i>	
M01-022	657	252	86.9	putative ribosomal protein S27	gb	ATAC003680 <i>Arabidopsis thaliana</i>	F01-009
M01F059	567	459	72.5	putative UDP-galactose-4-epimerase	gb	ATAC004238 <i>Arabidopsis thaliana</i>	
M01-091	650	270	72.2	putative vacuolar ATPase	gb	AC002330 <i>Arabidopsis thaliana</i>	
M01S037	650	648	68.1	pyruvate kinase, cytosolic isozyme	sp	KPYC_SOYBN <i>Glycine max</i>	
M01-087	565	429	42.7	pyruvate kinase, cytosolic isozyme	sp	KPYC_SOLTU <i>Solanum tuberosum</i>	
M01X012	123	117	87.2	RAS-related protein RAB1BV	sp	RAB1_BETVU <i>Beta vulgaris</i>	
M01-018	623	558	56.5	rehydrin	sp	REHY_TORRU <i>Torula ruralis</i>	
M01C095	464	129	55.8	replication protein A1	gb	AF009179 <i>Oryza sativa</i>	
M01B031	156			ribosomal protein S30	(sp)	(RS30_ARATH) (<i>Arabidopsis thaliana</i>)	F01-112
M01-068	644	627	92.3	ribosomal protein L10	sp	RL10_SOLME <i>Solanum melongena</i>	F01-094
M01N030	401	306	93.1	ribosomal protein L12	gb	PAU93168 <i>Prunus armeniaca</i>	F01-036
M01F077	444	240	67.5	ribosomal protein L13	sp	RL13_ARATH <i>Arabidopsis thaliana</i>	
M01I024	419	369	78.9	ribosomal protein L13-1	sp	R131_BRANA <i>Brassica napus</i>	F01L049
M01E011	502	441	87.1	ribosomal protein L17	gb	AF034948 <i>Zea mays</i>	F01F140

Table 1. A. Continued.

M-EST*	L*	OL*	ID% ⁴	Putative Identification	Source*	Organism ¹	F-EST*
M01G039	560	354	80.5	ribosomal protein L17	gb	TIF15	<i>Arabidopsis thaliana</i>
M01C033	615	492	86.6	ribosomal protein L18A	gb	ATAC004077	<i>Arabidopsis thaliana</i>
M01E075	331	324	80.6	ribosomal protein L19	prf	2111458B	<i>Nicotiana tabacum</i>
M01B051	611	507	85.2	ribosomal protein L2	sp	RL2_LYCES	<i>Lycopersicon esculentum</i>
M01F084	563	309	71.8	ribosomal protein L21	sp	RL21_ARATH	<i>Arabidopsis thaliana</i>
M01-020	583	393	71.8	ribosomal protein L24	sp	RL24_HORVU	<i>Hordeum vulgare</i>
M01-046	599	312	85.6	ribosomal protein L26	gb	AF093540	<i>Zea mays</i>
M01-053	590	405	72.6	ribosomal protein L27	prf	2111451A	<i>Pisum sativum</i>
M01-007	584	282	87.2	ribosomal protein L27A	sp	RL2A_ARATH	<i>Arabidopsis thaliana</i>
M01-025	753	327	86.2	ribosomal protein L30	gb	LLAJ3316	<i>Lupinus luteus</i>
M01-038	519	276	83.7	ribosomal protein L37	sp	RL37_ARATH	<i>Arabidopsis thaliana</i>
M01-013	611	243	91.4	ribosomal protein L37a, cytosolic	pir	S34661	<i>Brassica rapa</i>
M01-097	540	207	85.5	ribosomal protein L38	sp	RL38_ARATH	<i>Arabidopsis thaliana</i>
M01B059	466	150	92.0	ribosomal protein L39	sp	RL39_MAIZE	<i>Zea mays</i>
M01-024	508	75	100.0	ribosomal protein L41	sp	RL41_TOBAC	<i>Nicotiana tabacum</i>
M01-108	219	135	66.7	ribosomal protein L6	sp	RL6_MESCR	<i>Mesembryanthemum crystallinum</i>
M01D072	650	498	84.3	ribosomal protein L7A	sp	RL7A_ORYSA	<i>Oryza sativa</i>
M01-041	949	417	81.3	ribosomal protein L9	sp	RL9_PEA	<i>Pisum sativum</i>
M01E013	463	411	79.6	ribosomal protein S11	sp	RS11_SOYBN	<i>Glycine max</i>
M01D040	650	450	86.7	ribosomal protein S13	sp	RS13_ARATH	<i>Arabidopsis thaliana</i>
M01-057	664	405	90.4	ribosomal protein S14	sp	RS14_CHLRE	<i>Chlamydomonas reinhardtii</i>
M01U023	532	390	87.7	ribosomal protein S15a	pir	S20945	<i>Brassica napus</i>
M01N001	624	375	87.2	ribosomal protein S16	sp	RS16_GOSHI	<i>Gossypium hirsutum</i>
M01A046	584	420	90.7	ribosomal protein S16	sp	RS16_GOSHI	<i>Gossypium hirsutum</i>
M01D068	650	384	89.1	ribosomal protein S18	sp	RS18_ARATH	<i>Arabidopsis thaliana</i>
M01N065	618	456	88.8	ribosomal protein S18	sp	RS18_ARATH	<i>Arabidopsis thaliana</i>
M01E019	532	429	70.6	ribosomal protein S19	sp	RS19_ORYSA	<i>Oryza sativa</i>
M01X045	627	390	78.5	ribosomal protein S19	prf	I909359A	<i>Solanum tuberosum</i>
M01-037	612	348	81.9	ribosomal protein S20	sp	RS20_ARATH	<i>Arabidopsis thaliana</i>
M01I033	474	402	96.3	ribosomal protein S23	sp	RS23_FRAAN	<i>Fragaria ananassa</i>
M01H030	392	315	74.3	ribosomal protein S25	sp	RS25_LYCES	<i>Lycopersicon esculentum</i>
M01W046	556	192	85.9	ribosomal protein S28	gb	HVJ001161	<i>Hordeum vulgare</i>
M01-039	556	444	78.4	ribosomal protein S8	gb	AF071889	<i>Prunus armeniaca</i>
M01L032	408	345	87.8	ribosomal protein SA	sp	RSP4_SOYBN	<i>Glycine max</i>
M01E064	95			ribulose-bisphosphate carboxylase activase	(pir)	(S25482)	<i>(Nicotiana tabacum)</i>
M01D073	650	339	77.0	ribulose-bisphosphate carboxylase small chain	gb	AB004883	<i>Marchantia paleacea</i>
M01-001	660	339	80.5	ribulose-bisphosphate carboxylase small chain	gb	AB004883	<i>Marchantia paleacea</i>
M01-106	524	465	87.1	ribulose-bisphosphate carboxylase small chain	gb	AB004883	<i>Marchantia paleacea</i>
M01-107	511	339	84.1	ribulose-bisphosphate carboxylase small chain	gb	AB004883	<i>Marchantia paleacea</i>
M01-026	642	528	86.4	ribulose-bisphosphate carboxylase small chain	gb	AB004883	<i>Marchantia paleacea</i>
M01G041	422			ribulose-bisphosphate carboxylase small chain	(gb)	(AB004883)	<i>(Marchantia paleacea)</i>
M01A035	650	214	60.3	Rice cDNA, partial sequence (C2657_1A), mRNA sequence	gb	RICC2657A	<i>Oryza sativa</i>
M01G087	443	110	66.4	Rice cDNA, partial sequence (C52831_1A), mRNA sequence	gb	C27773	<i>Oryza sativa</i>
M01D014	569	315	80.0	RNA helicase isolog	gb	ATAC002337	<i>Arabidopsis thaliana</i>
M01N014	650	438	67.1	RNA polymerase I, II and III 24.3 kDa subunit	gb	AF019248	<i>Arabidopsis thaliana</i>
M01A010	503	270	72.2	soluble inorganic pyrophosphatase	sp	IPYR_SOLTU	<i>Solanum tuberosum</i>
M01F055	609	102	64.7	T08I13.19 gene product	gb	ATAC002337	<i>Arabidopsis thaliana</i>
M01-094	554	174	41.4	T2L5.6 gene product	gb	T2L5	<i>Arabidopsis thaliana</i>
M01-089	469	123	56.1	T32F6.5 gene product	gb	ATAC005700	<i>Arabidopsis thaliana</i>
M01B063	650	339	58.4	thioredoxin H	gb	AF051206	<i>Picea mariana</i>
M01-017	651	525	56.0	transformer-2-like SR-related protein	prf	2402344B	<i>Nicotiana tabacum</i>
M01B033	598	270	51.1	translation initiation factor EIF-2B-epsilon subunit	gb	ATAC004238	<i>Arabidopsis thaliana</i>
M01A045	650	504	72.6	translationally controlled tumor protein homolog	sp	TCTP_ORYSA	<i>Oryza sativa</i>
M01B067	447			tubulin alpha-5 chain	(sp)	(TBA5_MAIZE)	<i>(Zea mays)</i>
M01E031	409	198	71.2	ubiquinol--cytochrome-c reductase (EC 1.10.2.2) 11K protein	pir	S48690	<i>Solanum tuberosum</i>
M01-004	526	384	97.7	ubiquitin / ribosomal protein CEP52	pir	S28420	<i>Nicotiana sylvestris</i>
M01-036	767	348	94.0	ubiquitin extension protein	gb	STUBIEP	<i>Solanum tuberosum</i>
M01B092	543	324	65.7	UDP-glucose/sterol glucosyltransferase	prf	2402364A	<i>Avena sativa</i>
M01-016	654			VVTL1	(prf)	(2322376A)	<i>(Vitis vinifera)</i>

Table 1.

B. ESTs similar to sequences from non-plant species

M-EST*	L*	OL*	ID%*	Putative Identification	Source*	Organism*	F-EST*
M01D070	562	447	20.1	4.6 kD protein in TOR2-PAS1 intergenic region	sp	YKU1_YEAST	<i>Saccharomyces cerevisiae</i>
M01F051	589	549	55.7	5-methyltetrahydrofolate-homocysteine methyltransferase	sp	METH_ECOLI	<i>Escherichia coli</i>
M01Q012	476	303	40.6	Ats1	gb	SPU82218	<i>Schizosaccharomyces pombe</i>
M01A034	650	318	47.2	bacterioferritin comigratory protein	gb	MTV009	<i>Mycobacterium tuberculosis</i>
M01A089	650	261	32.2	C01B10.8 gene product	gb	CELC01B10	<i>Caenorhabditis elegans</i>
M01I070	498	192	53.1	C11D2.4 gene product	gb	CELC11D2	<i>Caenorhabditis elegans</i>
M01A077	638	267	30.3	C25A1.13	gb	CEC25A1	<i>Caenorhabditis elegans</i>
M01D093	650	375	46.4	C27F2.5 gene product	gb	CELC27F2	<i>Caenorhabditis elegans</i>
M01R051	143	102	67.6	calmodulin	prf	I102228A	<i>Gallus gallus</i>
M01R052	461	81	100.0	calmodulin	gb	AF007889	<i>Symbiodinium microadriaticum</i>
M01-040	793	105	97.1	calmodulin	gb	AF007889	<i>Symbiodinium microadriaticum</i>
							F01-019 F01S007 F01W064
M01N026	383	198	39.4	CBY protein, chromosomal	sp	CBYC_ALCEU	<i>Alcaligenes eutrophus</i>
M01W070	530	162	53.7	conserved hypothetical protein ybgG	pir	B69751	<i>Bacillus subtilis</i>
M01A047	557	98	81.6	Crassostrea gigas clone AR86x microsatellite sequence.	gb	AF051181	<i>Crassostrea gigas</i>
M01-101	636	234	42.3	CRTK protein	sp	CRTK_RHOCA	<i>Rhodobacter capsulatus</i>
M01A038	584	42	85.7	CTG26 alternate open reading frame	gb	HSU80761	<i>Homo sapiens</i>
M01D060	537	108	83.3	DNA-directed RNA polymerase II 14.4 kD polypeptide	sp	RPB6_HUMAN	<i>Homo sapiens</i>
M01I077	512	213	66.2	dolichyl-diphosphooligosaccharide-protein glycosyltransferase 48 kD subunit	sp	OST4_CHICK	<i>Gallus gallus</i>
M01F036	111	90	73.3	EBNA-2 nuclear protein	sp	EBN2_EBV	<i>Human herpesvirus 4</i>
M01B072	600	432	68.1	electron transfer flavoprotein beta-subunit	sp	ETFB_BRAJA	<i>Bradyrhizobium japonicum</i>
M01E049	494	354	86.4	F55G1.11 gene product	gb	CEL55G1	<i>Caenorhabditis elegans</i>
M01L074	335	105	65.7	FBRL_HUMAN gene product	gb	AC005393	<i>Homo sapiens</i>
M01J066	625	408	98.5	histone H3	pir	S07350	<i>Neurospora crassa</i>
M01-030	650	186	37.1	Homo sapiens chromosome 17, clone hCIT54K19, complete sequence	gb	AC003664	<i>Homo sapiens</i>
M01B084	556	95	69.5	Homo sapiens clone DJ0852O24; HTGS phase 1, 2 unordered pieces	gb	AC004906	<i>Homo sapiens</i>
M01N083	630	240	60.0	hypothetical 12.9 kD protein SLR1417	sp	YE17_SYNY3	<i>Synechocystis sp.</i>
M01G033	650	306	45.1	hypothetical 15.9 kD protein in OLE1-DUP1 intergenic region	sp	YGF4_YEAST	<i>Saccharomyces cerevisiae</i>
M01X039	594	369	51.2	hypothetical 37.4 kD protein in EXUR-TDCC intergenic region (O328)	sp	YQJG_ECOLI	<i>Escherichia coli</i>
M01G063	491	195	46.2	hypothetical protein	gb	AF011338	<i>Dictyostelium discoideum</i>
M01P050	325	210	37.1	hypothetical protein	gb	AF025662	<i>Vibrio cholerae</i>
M01F022	382	126	54.8	hypothetical protein MTCY441.43	gb	MTCY441	<i>Mycobacterium tuberculosis</i>
M01A088	566	426	48.6	hypothetical protein MTV005.09	gb	MTV005	<i>Mycobacterium tuberculosis</i>
M01I001	587	237	43.0	hypothetical protein slr1046	pir	S74990	<i>Synechocystis sp.</i>
M01Y015	573	54	90.7	Mouse mRNA for P100 serine protease of Ra-reactive factor (RaRF), complete cds	gb	MUSCRARF	<i>Mus musculus</i>
M01H053	443	147	50.0	N4-(beta-N-acetylglucosaminyl)-L-asparaginase precursor nascent polypeptide associated complex alpha chain (KIAA0363 gene product)	sp	ASPG_HUMAN	<i>Homo sapiens</i>
M01A084	326				(gp)	(AB002361)	<i>(Homo sapiens)</i>
							F01G122
M01H054	487	294	46.9	nuclear transport factor 2	sp	NTF2_YEAST	<i>Saccharomyces cerevisiae</i>
M01N040	445			PDR5-like ABC transporter	(prf)	(2221393A)	<i>(Spirodela polyrrhiza)</i>
							F01O027
M01G086	589	69	69.6	pk34c08.s1 Caenorhabditis briggsae cDNA similar to SP	gb	R04891	<i>Caenorhabditis briggsae</i>
M01-023	648	291	87.6	PRCDNA10	gb	A63899	<i>Phaffia rhodozyma</i>
M01S021	571	363	44.6	probable ribosomal protein S7	sp	RS7_SCHPO	<i>Schizosaccharomyces pombe</i>
M01-099	521	309	48.5	riboflavin synthase alpha chain	sp	RISA_ACTPL	<i>Actinobacillus pleuropneumoniae</i>
M01B045	578	513	65.5	ribosomal protein L10A	sp	R10A_RAT	<i>Rattus norvegicus</i>
M01I027	650	591	65.0	ribosomal protein L10A	sp	R10A_RAT	<i>Rattus norvegicus</i>
							F01I014
M01-088	587	351	52.1	ribosomal protein L13	pir	H70461	<i>Aquifex aeolicus</i>
M01Y014	651	354	55.9	ribosomal protein L22	sp	RL22_TRIGR	<i>Tripneustes gratilla</i>
							F01F027
M01A026	575	288	35.4	ribosomal protein L28	pir	I48738	<i>Mus musculus</i>
M01G043	475	153	76.5	ribosomal protein L29, cytosolic	pir	JC2012	<i>Mus musculus</i>
M01-012	561	390	71.5	ribosomal protein L32	pir	R5HU32	<i>Homo sapiens</i>
							F01-021
M01-027	548	315	58.1	ribosomal protein L35a.e.c16, cytosolic	pir	S18431	<i>Saccharomyces cerevisiae</i>
							F01I135
M01F003	329	279	92.5	ribosomal protein L41	gb	YSAQLAG	<i>Candida maltosa</i>
							F01-013
M01-008	635	333	65.8	ribosomal protein S12	pir	JC4159	<i>Sus scrofa domestica</i>
							F01-015
M01-043	653	369	65.9	ribosomal protein S17	gb	SPBC24E9	<i>Schizosaccharomyces pombe</i>
							F01-123
M01A081	577	426	82.4	ribosomal protein S2	sp	RS2_DROME	<i>Drosophila melanogaster</i>
M01E003	462	312	83.7	ribosomal protein S3	sp	RS3_AMBME	<i>Ambystoma mexicanum</i>
M01K056	501	390	71.5	ribosomal protein S9	sp	RS9_PODAN	<i>Podospira anserina</i>
							F01X081

Table 1. B. Continued.

M-EST ^a	L ^b	OL ^c	ID% ^d	Putative Identification	Source ^e	Organism ^f	F-EST ^g
M01E029	466	243	30.9	Schizosaccharomyces pombe mRNA, partial cds, clone: SY1105	gb	D89201	<i>Schizosaccharomyces pombe</i>
M01-072	535	228	51.3	sensor protein LEMA	sp	LEMA_PSESY	<i>Pseudomonas syringae</i>
M01-082	650	318	27.4	syntaxin-like protein 3135	gb	AF093064	<i>Mus musculus</i>
M01A085	650	576	44.8	T20H4.3 gene product	gb	CEL20H4	<i>Caenorhabditis elegans</i>
M01A061	650	261	37.9	vacuolar proton pump:subunit=G2	prf	2312368C	<i>Bos primigenius taurus</i>
M01E030	451	381	31.5	vitellogenin	gb	APU46857	<i>Anolis pulchellus</i>
M01H063	633	120	57.5	Z.ramigera DNA for 23S ribosomal RNA	gb	ZR23SRRNA	<i>Zoogloea ramigera</i>
M01X034	649	120	42.5	zj05a04.s1 Soares fetal liver spleen 1NFLS S1 Homo sapiens cDNA clone 449358 3', mRNA sequence	gb	CELF08B4	<i>Caenorhabditis elegans</i>
M01W031	650	216	34.7	ZK675.1 gene product	gb	CEZK675	<i>Caenorhabditis elegans</i>

^aNon-redundant ESTs, ^blength of reads in bp without poly(A)⁺, ^coverlap in amino acid or in nucleotide residue, ^d% sequence identity, ^esource database (abbreviations: gb, GenBank; pir, PIR; sp, SwissProt; prf, PRF) and locus number of the database entry, ^forganism for which similarity was found, ^gcorresponding female ESTs. When a male EST has no similarity to known genes but its corresponding female EST shows significant similarity, putative identification, source and organism are given in parentheses.

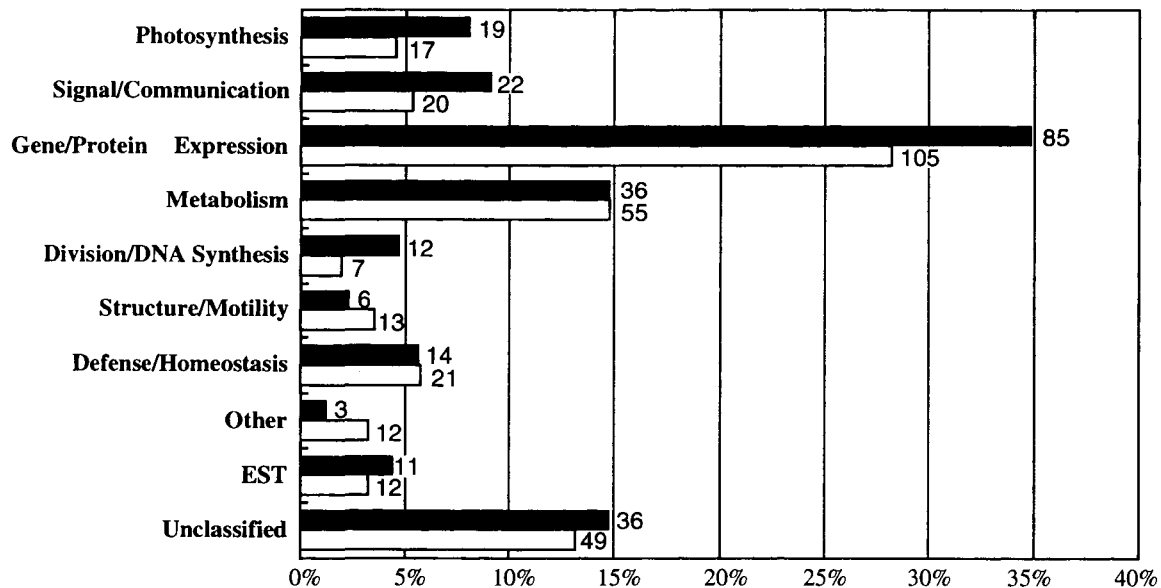


Figure 1. Classification of the male identified ESTs according to putative gene functions and comparison with the female ESTs. Black and white bars refer to percentages of classified ESTs among the male and female unique EST sequences, respectively, for which similarities to known sequences were identified. The category “unclassified” includes ESTs homologous to mRNA sequences of non-plant species or functionally unknown proteins such as hypothetical proteins. ESTs similar to those from other plants of unknown function are classified in the “EST” category.

served in their amino acid sequences and also at the nucleotide sequence level (sequence data not shown). Two ESTs (M01-055 and M01D090) show no significant similarity to known genes, both nucleotide sequences being nearly identical in the putative coding region, and divergent in the 5'-untranslated region (UTR) and thus are considered as members of a gene family (sequence data not shown). Transcripts from various members of the *rbcS* gene family were found most frequently among the male ESTs, bringing the total number of members to eight, including two found among the female ESTs.

The EST analysis from *A. thaliana* has revealed that many of its genes consist of multigene families.¹⁴ Similarly, when including the gene families found in the female ESTs,⁸ the presence of 19 gene families were shown in *M. polymorpha*. Divergence in UTRs, especially 3'-UTRs, among a gene family will make it possible to distinguish its individual members. However, not all ESTs examined in this study contain UTRs, thus the number of family members may be underestimated.

Table 2. Possible gene families in the ESTs from male *M. polymorpha*.

Gene	No. of ESTs	EST
glyceraldehyde-3-phosphate dehydrogenase	2	M01-062, M01C069
histone H2A	2	M01-002, M01C030
histone H3	2	M01-086, M01J066
histone H3.3-like protein	3	M01-010, M01-014, M01A016
ribosomal protein L10a	2	M01B045, M01I027
ribosomal protein L17	2	M01E011, M01G039
ribosomal protein S16	2	M01A046, M01N001
ribosomal protein S18	2	M01D068, M01N065
RuBisCO small chain	6	M01-001, M01-026, M01-106, M01-107, M01D073, M01G041 (female ESTs; F01C167, F01G168)
unknown	2	M01M-55, M01D090

3.5. Comparison of male and female ESTs

To identify ESTs specific to a respective sex, ESTs were compared between the male and female libraries. ESTs from the opposite sexes were considered to tag the same gene if the sequences showed 97% or greater identity within the overlap. This relatively low stringency of the identity level was adopted since the female ESTs generated from field populations of *M. polymorpha* might contain polymorphism. Altogether 99 independent ESTs of the male library appear to originate from those genes found also in the female library, 81 of them show significant similarities to known genes (Table 1). On the other hand, 103 female-independent ESTs are shared with the male library. This discrepancy is due to the evidence that in four cases a single male EST, such as M01-090, bridges two non-overlapping female ESTs, F01F125 and F01K169, to form a single group. This makes the total number of non-redundant female ESTs 713 instead of 717 reported previously.⁸ Therefore, the total number of the male ESTs unique to the male library was 346 (445–99). Consequently, the overall number of non-redundant EST sequences from both sexual organs of *M. polymorpha* now totals 1059 (713+346).

3.6. Male/female unique ESTs

Comparison of the relative representation of functional groups of genes from the male and female EST libraries shows that the overall patterns are similar (Fig. 1). The frequencies of ESTs for genes involved in photosynthesis, signal transduction and cell division/DNA synthesis are apparently higher in the male library mainly due to the smaller number of ESTs obtained.

Many different species of histone genes related with cell division and DNA synthesis were found more often in the male ESTs than in the female ESTs. Furthermore, one of the male ESTs (M01B076) equivalent to the female EST F01K174 had a significant similarity to a cystein-

rich extensin-like protein specifically expressed in pistil.¹⁵ Of the ESTs unique to either library, several ESTs, whose homologs are either conspicuously involved in sexual differentiation processes or absent from it, are particularly noteworthy though their sex specificity should be examined at the DNA/RNA levels. In the female EST collection, five ESTs tag members of a lectin gene family, two of which were abundant in the female library while none of these being a few in the male ESTs. Further functional analysis is necessary to elucidate the role of lectin genes in the development of the female sexual organs. Among the ESTs only found in the male collection, two species shared sequence similarities to genes associated with sexual reproduction in other organisms; the EST M01-017 identifies *transformer-2* (*tra2*), which is involved in sex determination of *Drosophila melanogaster*¹⁶ (Fig. 2), and the EST M01E030 is similar to vitellogenin, which is a female-specific lipoprotein of a species of iguana, *Anolis pulchellus*,¹⁷ where it plays a role as an estrogen and is connected with vitellus development.

3.7. Characterization of a contig with similarity to a *transformer-2* like protein

A reading frame of 244 amino acids in one of the male ESTs, M01-017, showed a high level of similarity to the *transformer-2* (*tra2*)-like SR-related protein of *Nicotiana tabacum*¹⁸ with 53.6% amino acid identity and to the *tra2* protein of *D. melanogaster*¹⁶ with 30.0% amino acid identity (Fig. 2). Sequence analysis of the EST M01-017 revealed a highly conserved serine-arginine (SR)-rich repeat region at the N-terminal region and a divergent SR-rich region at the C-terminal region, both of which are characteristic of the group of SR proteins.¹⁹ The sequence also shows significant similarity to the RNA-binding domain of the *tra2*-like protein, with well-conserved RNP1 and RNP2 consensus motifs. Nineteen residues of the 23 consensus amino acids in the RNA-binding domain are identical (highlighted in Fig. 2).²⁰ The linker region,

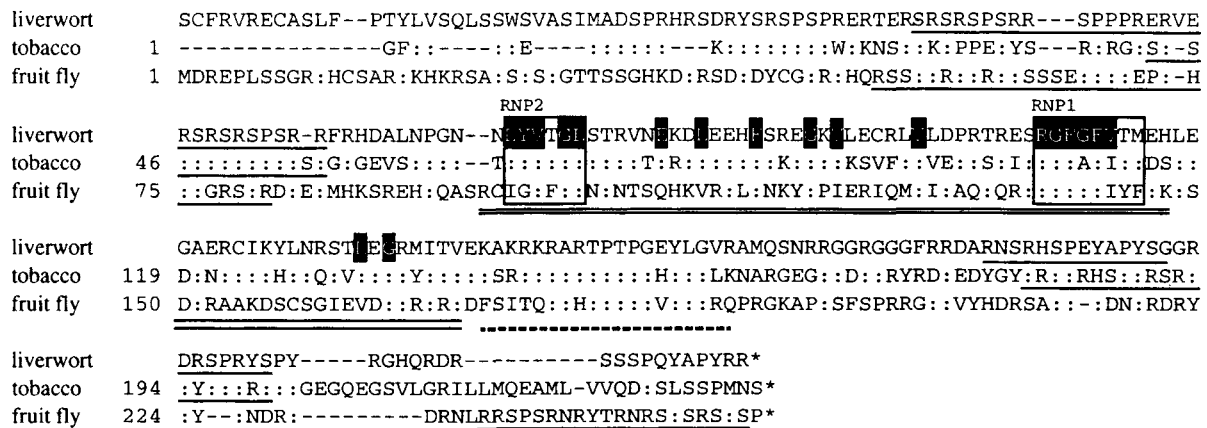


Figure 2. Alignment of deduced amino acid sequences of the transformer-2 (tra2)-like protein of *M. polymorpha* with those of the tra2-like SR-related protein of *N. tabacum*¹⁸ and the tra2 protein of *D. melanogaster*.¹⁶ Asterisks indicate stop codons. Hyphens indicate gaps. Amino acid residues identical to the EST are represented as colons. Predicted SR domains are underlined. RNA-binding domains are doubly underlined, and RNP1 and RNP2 consensus motifs are boxed. Consensus amino acid residues in RNA-binding domain of the tra2-like protein of *M. polymorpha* are highlighted. A broken line indicates the linker region.

a 19-amino-acid conserved region, displays a high level of local similarity with the analogous domain of the tra2-like protein in *N. tabacum* (13 of 19 amino acid residues are identical) and the tra2 protein of *D. melanogaster* (10 of 19 amino acid residues identical).

Tra2-like proteins are SR proteins that belong to a ribonucleoprotein family of pre-mRNA splicing factors. It has been reported that female-specific splicing of *doublesex* (*dsx*) pre-mRNA is regulated by transformer (*tra*) protein and tra2 protein in *Drosophila*.²¹ The gene coding for a tra2-like SR-related protein of *N. tabacum* was reported to be transcriptionally activated by treating with cryptogin but its involvement in sex determination is as yet unclear.¹⁸ Further examination of the male EST is required to reveal whether the tra2-like gene of *M. polymorpha* plays a role in sexual differentiation in this liverwort.

Acknowledgments: We thank Axel Brennicke, Ulm University, Germany, for his critical reading of the manuscript and invaluable discussions. This work was supported by Grants-in-Aid from the Ministry of Education, Science, Sports and Culture of Japan (No. 06454076 and No. 07281101), and by Program for Promotion of Basic Research Activities for Innovative Biosciences (PROBRAIN).

References

- Marra, M. A., Hillier, L., and Waterston, R. H. 1998, Expressed sequence tags-ESTablishing bridges between genomes, *Trends in Genet.*, **14**, 4-7.
- Affara, N. A., Bentley, E., Davey, P., Pelmear, A., and Jones, M. H. 1994, The identification of novel gene sequences of the human adult testis, *Genomics*, **22**, 205-210.
- Starborg, M., Brundell, E., and Hoog, C. 1992, Analysis of the expression of a large number of novel genes isolated from mouse prepubertal testis, *Mol. Reprod. Dev.*, **33**, 243-251.
- Höfte, H., Desprez, T., Amselem, J. et al. 1993, An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*, *Plant J.*, **4**, 1051-1061.
- Cooke, R., Raynal, M., Laudie, M. et al. 1996, Further progress towards a catalogue of all *Arabidopsis* genes: analysis of a set of 5000 non-redundant ESTs, *Plant J.*, **9**, 101-124.
- Yamamoto, K. and Sasaki, T. 1997, Large-scale EST sequencing in rice, *Plant Mol. Biol.*, **35**, 135-144.
- Lim, C. O., Kim, H. Y., Kim, M. G. et al. 1996, Expressed sequence tags of Chinese cabbage flower bud cDNA, *Plant Physiol.*, **111**, 577-588.
- Nagai, J., Yamato, K. T., Sakaida, M., Yoda, H., Fukuzawa, H., and Ohyama, K. 1999, Expressed sequence tags from immature sexual organ of a liverwort, *Marchantia polymorpha*, *DNA Res.*, **6**, 1-11.
- Smith, G. M. 1955, Cryptogamic botany: Bryophytes and Pteridophytes, McGraw-Hill Publications in the Botanical Sciences, New York.
- Lorbeer, G. 1934, Die Zytologie der Lebermoose mit besonderer Berücksichtigung allgemeiner Chromosomenfragen, *Jahrb. Wiss. Bot.*, **80**, 565-818.
- Ono, K., Ohyama, K., and Gamborg, O. 1979, Regeneration of the liverwort *Marchantia polymorpha* L. from protoplasts isolated from cell suspension culture, *Plant Sci. Lett.*, **14**, 225-229.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990, Basic Local Alignment Search Tool, *J. Mol. Biol.*, **215**, 403-410.
- Adams, M., Kerlavage, A., Fleischmann, R. et al. 1995, Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence, *Nature*, **377**, 3-174.
- Cooke, R., Raynal, M., Laudie, M., and Delseny, M. 1997, Identification of members of gene families in *Arabidopsis*

- thaliana* by contig construction from partial cDNA sequences: 106 genes encoding 50 cytoplasmic ribosomal proteins, *Plant J.*, **11**, 1127–1140.
15. Goldman, M. H., Pezzotti, M., Seurinck, J., and Mariani, C. 1992, Developmental expression of tobacco pistil-specific genes encoding novel extensin-like proteins, *Plant Cell*, **4**, 1041–1051.
 16. Amrein, H., Gorman, M., and Nöthiger, R. 1988, The sex-determining gene *tra-2* of *Drosophila* encodes a putative RNA binding protein, *Cell*, **55**, 1025–1035.
 17. Morales, M. H., Osuna, R., and Sanchez, E. 1991, Vitellogenesis in *Anolis pulchellus*: induction of VTG-like protein in liver explants from male and immature lizards, *Exp. Zool.*, **260**, 50–58.
 18. Petitot, A. S., Blein, J. P., Pugin, A., and Suty, L. 1997, Cloning of two plant cDNAs encoding a β -type proteasome subunit and a transformer-2-like SR-related protein: early induction of the corresponding genes in tobacco cells treated with cryptogein, *Plant Mol. Biol.*, **35**, 261–269.
 19. Fu, X. D. 1995, The superfamily of arginine/serine-rich splicing factors, *RNA*, **1**, 663–680.
 20. Burd, C. G. and Dreyfuss, G. 1994, Conserved structures and diversity of functions of RNA-binding proteins, *Science*, **265**, 615–621.
 21. Tian, M. and Maniatis, T. 1993, A splicing enhancer complex controls alternative splicing of *doublesex* pre-mRNA, *Cell*, **74**, 105–114.