Large-scale cDNA analysis of the maternal genetic information in the egg of *Halocynthia roretzi* for a gene expression catalog of ascidian development

Kazuhiro W. Makabe[1,*], Takeshi Kawashima[1], Shuichi Kawashima[2], Takuya Minokawa[3], Asako Adachi[1], Hiroshi Kawamura[4], Hisayoshi Ishikawa[4], Riho Yasuda[1], Hiroki Yamamoto[3], Keisuke Kondoh[3], Sachiko Arioka[3], Yasunori Sasakura[1], Ako Kobayashi[1], Kasumi Yagi[1], Keisuke Shojima[1], Yuki Kondoh[4], Sumina Kido[4], Mayu Tsujinami[4], Nahoko Nishimura[1], Miyuki Takahashi[1], Tetsuji Nakamura[1], Minoru Kanehisa[2], Michio Ogasawara[1], Takahito Nishikata[4,5] and Hiroki Nishida[3]

[1]Department of Zoology, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan
[2]Institute of Chemical Research, Kyoto University, Uji 611-0011, Japan
[3]Department of Biological Sciences, Tokyo Institute of Technology, Yokohama 226-8501, Japan
[4]Department of Biology, Faculty of Science, Konan University, Kobe 658-8501, Japan
[5]High Technology Research Center, Konan University, Kobe 658-8501, Japan
*Author for correspondence (e-mail: kwmakabe@ascidian.zool.kyoto-u.ac.jp)

SUMMARY

The ascidian egg is a well-known mosaic egg. In order to investigate the molecular nature of the maternal genetic information stored in the egg, we have prepared cDNAs from the mRNAs in the fertilized eggs of the ascidian, *Halocynthia roretzi*. The cDNAs of the ascidian embryo were sequenced, and the localization of individual mRNA was examined in staged embryos by whole-mount in situ hybridization. The data obtained were stored in the database MAGEST (http://www.genome.ad.jp/magest) and further analyzed. A total of 4240 cDNA clones were found to represent 2221 gene transcripts, including at least 934 different protein-coding sequences. The mRNA population of the egg consisted of a low prevalence, high complexity sequence set. The majority of the clones were of the rare sequence class, and of these, 42% of the clones showed significant matches with known peptides, mainly consisting of proteins with housekeeping functions such as metabolism and cell division. In addition, we found cDNAs encoding components involved in different signal transduction pathways and cDNAs encoding nucleotide-binding proteins. Large-scale analyses of the distribution of the RNA corresponding to each cDNA in the eight-cell, 110-cell and early tailbud embryos were simultaneously carried out. These analyses revealed that a small fraction of the maternal RNAs were localized in the eight-cell embryo, and that 7.9% of the clones were exclusively maternal, while 40.6% of the maternal clones showed expression in the later stages. This study provides global insights about the genes expressed during early development.

Key words: Ascidian, Maternal RNA, cDNA analysis, EST, Expression profile, Database

INTRODUCTION

The fertilized egg is a single totipotent cell that cleaves many times to give rise to a multicellular organism. Within an embryo, embryonic cells develop into various tissue types, and ultimately into all cell types and structures in a series of specification and differentiation steps. Ascidian embryos develop into tadpole-shaped swimming larvae, which have a primitive chordate body. Ascidians have been the subject of embryological studies for more than a hundred years (Chabry, 1887) and have been used as classic organisms to study what is known as 'mosaic development' (Conklin, 1905a; Conklin, 1905b). Of the major organs and tissues, the epidermis, muscle and endoderm are formed by cell-autonomous processes. The mosaicism of the ascidian embryo is evidence that the localized cytoplasmic factors in the egg specify the tissue precursor cells during embryogenesis (Reverberi and Minganti, 1946; see

Satoh, 1994 for a review). In addition, a series of experiments has shown that localized maternal factors are responsible for controlling the axis specification, cleavage pattern, morphogenetic movements such as gastrulation (for a review, see Nishida, 1997) and the responsiveness of the notochord and mesenchyme precursor blastomeres (Kim et al., 2000). These factors are segregated into particular blastomeres by an invariant cleavage pattern of the ascidian embryo. Thus, various processes of ascidian embryogenesis are mediated by localized maternal factors in the egg. We are interested in the maternal factors in the egg cytoplasm that provide information for the diversification of cell types and for the morphogenesis during the successive stages of embryogenesis.

In the early 1980s, it was reported that actin mRNA accumulated preferentially in a region of the ascidian egg called the myoplasm, providing the first evidence for polarized mRNA distribution within a cell (Jeffery et al., 1983). This has

been followed by numerous reports of mRNAs localized in the oocytes, eggs and embryos of various organisms, including both protostomes, such as the fruit fly and nematode, and deuterostomes, such as *Xenopus* (for reviews, see Wilhelm and Vale, 1993; Ding and Lipshitz, 1993; Glotzer and Ephrussi, 1996; Gavis, 1997; Schnapp et al., 1997). The RNAs present in a cell determine the range of functions the cell can perform. In most of the species studied, including the fruit fly, nematode, sea urchin, frog and possibly in fish, the establishment of axes, diversification of cell types and morphological changes during early embryogenesis rely partly on the maternally stored mRNAs (e.g. Davidson, 1986; Wieschaus, 1996; Wylie et al., 1996; Bowerman et al., 1997; Yamaha et al., 1998). A large-scale analysis of the functional requirements in the *Drosophila* germline has suggested that 75% of the 3600 lethal loci in the genome are functionally required during oogenesis (Perrimon et al., 1989). Therefore, large-scale identification of RNAs stored in the eggs and analysis of the abundance of the RNAs can give valuable insights into the basic molecular processes. Reflecting the enormous developmental potential of the egg, the egg contains the greatest transcript complexity of any known cell type (Davidson, 1986; Poustka et al., 1999; Ko et al., 2000). This fact also makes the egg a particularly useful source of cDNA clones for the identification of genes.

Because of its phylogenetic position, the ascidian would be a key organism for large-scale genome-wide research. In the phylum Chordata, genomic duplication events took place twice during the process of vertebrate evolution. The ascidian, which is a lower chordate, however, has a non-duplicated genome that can be regarded as a basic set of the chordate-type genome (Holland et al., 1994; Sidow, 1996). The haploid genome of the ascidian consists of about 160 Mb, and contains about 15,000 genes, which are approximately one-twentieth and one-fifth of the respective amounts in mammals (Laird, 1971; Simmen et al., 1998). As a result of the absence of genome duplication in the ascidian lineage, the number of genes and the genome size of the ascidian are in the same range as those of the fruit fly and nematode (Miklos and Rubin, 1996), for which the genomes have already been sequenced. The RNA complexity in the egg of the ascidian is therefore expected to be lower than those of other chordates. Therefore, the ascidian is a good model system with which to investigate the functions of chordate genomes, especially as the function of genes can easily be analyzed by gene introduction and gene disruption in ascidians, which are suitable for microsurgical manipulation and modern molecular techniques.

One approach to the comprehensive study of RNA functions is to sequence randomly isolated cDNAs and analyze the expression pattern of the corresponding mRNAs by whole-mount in situ hybridization on a large scale, because it is generally accepted that differential gene expression is a major mechanism in cellular differentiation and development. Description of the gene expression pattern is now considered to be an essential part of the characterization of genes because it clarifies their functions. There are examples of similar research and related databases for several organisms (Tabara et al., 1996; Gawantka et al., 1998; Ko et al., 2000).

We have therefore constructed a database named MAGEST, an acronym for Maboya (the ascidian *Halocynthia roretzi*) gene expression patterns and sequence tags (Kawashima et al., 2000), to analyze the data produced in this project, which is implemented in the Sybase relational database system and is accessible through the Internet. Our aim is to achieve an all-inclusive and systematic description of maternal transcripts stored in the fertilized eggs of the ascidian, *Halocynthia roretzi*, in terms of the cDNA sequences and their localization and zygotic expression patterns during embryogenesis. Data collection is ongoing and we describe an overview of the large-scale, global analysis so far carried out on the identification and localization/expression of the maternal mRNAs stored in the ascidian.

MATERIALS AND METHODS

Eggs and embryos

Adult ascidians, *Halocynthia roretzi*, were obtained from fishermen near the Otsuchi Marine Research Institute, the University of Tokyo, Iwate, Japan and the Asamushi Marine Biological Station, Tohoku University, Aomori, Japan. Naturally spawned eggs were artificially fertilized with a suspension of non-self sperm. The fertilized eggs were immediately collected or cultured at 12°C in Millipore-filtered sea water containing 50 µg/ml streptomycin up to the eight-cell, 110-cell and early tailbud stages, at about 4 hours, 9 hours and 24 hours after fertilization, respectively. The eggs and embryos were collected by low-speed centrifugation.

Construction of the arrayed cDNA library and EST sequencing

Ascidian eggs are surrounded with the chorion, attached to which are many follicle cells. Within the perivitelline space, numerous test cells are found. The presence of the follicle and the test cells makes it difficult to analyze the property of the eggs when it is homogenized with the chorion (Jeffery, 1980). Therefore, we used eggs that were chemically dechorionated, as described by Mita-Miyazawa et al. (Mita-Miyazawa et al., 1985) prior to RNA extraction. Poly(A) RNA was extracted from the fertilized eggs using AGPC (Chomczynski and Sacchi, 1987) and purified using Oligotex-dT30 beads (Roche Japan, Tokyo). An oligo(dT)-primed cDNA library was constructed from 5 µg of poly(A) RNA. For the cDNA synthesis, a lambda uni-ZAP cDNA synthesis kit (Stratagene) was used until purification of the adaptor-ligated cDNA. The cDNAs of the largest and second-largest fraction obtained in the fractionation step were ligated into pBluescript(SK-) and electroporated into *Escherichia coli* DH10B cells (Gibco) using a Gene Pulser II electroporation system (Pharmacia). A total of 86,000 independent cDNA clones of the library were picked randomly and arrayed in over two hundred 384-well plates in a Q-Bot robot (Genetix, UK). All the clones had a clone ID as their identification number, each of which was a combination of a plate number and a well number of the clone. The arrayed library was stored at -80°C. Single-pass cDNA sequencing was conducted. The 384-well plates were thawed and the clones were withdrawn, inoculated directly into a PI-50 six-hole tube that contained 3 ml of LB broth and cultured overnight, after which the cDNAs were isolated in a PI-50 mini-prep machine (Kurabo). Their 5′ and 3′ termini were sequenced in a half-scale reaction by conventional procedures in an automated ABI PRISM 377 sequencer (Perkin Elmer Japan), using Big-Dye dideoxy chain terminators. The sequencing primers were T3 (5′-ATTAACCCTCACTAAAGGGA-3′) or SK20 (5′-CGCTCTAG-AACTAGTGGATC-3′) for sequencing the 5′ termini and T7 (5′-GCGTAATACGACTCACTATA-3′) for the 3′ termini.

Sequence data analysis

After removing the vector sequences and ambiguous regions that contained stretches of N (ambiguous nucleotide) from the raw sequence data, the processed sequences were registered and used as query sequences for BLAST homology searches against GenBank

(Benson et al., 2000) including dbEST (Boguski et al., 1993) at the nucleotide sequence level and also against nr-aa, a non-redundant protein sequence database constructed from SWISS-PROT (Bairoch and Apweiler, 2000), PIR (Barker et al., 1999), PRF (Protein Research Foundation, Osaka, Japan) and GenPept (translated GenBank) (Benson et al., 2000) at the amino acid sequence level. Up to 10 entries above a given threshold (blast E value<e^{-8}) were retrieved from the original databases by the DBGET/LinkDB system (Fujibuchi et al., 1998) and stored in our database, named MAGEST (Kawashima et al., 2000).

All the 3′ sequences were used as fingerprints of the genes. Clustering of the 3′-sequences was carried out by searching for sequence similarities in MAGEST using the entire length of the sequences, in order to examine the number of redundant clones. The clustered clones derived from the same gene were given a cluster ID number, which began at C. As to clones whose 3′ sequence was accidentally not available, some of them were clustered by the 5′ sequences, but the other clones that stood alone were temporarily given cluster ID numbers which began at H.

Whole-mount in situ hybridization

Whole-mount specimens were mechanically dechorionated by a chorion peeler (H. I., H. K. and T. N., unpublished) after fixation. The specimens were hybridized in situ at 45°C using digoxigenin-labeled antisense probes, according to a newly developed method described by Ogasawara et al. (Ogasawara et al., 2001). We analyzed three developmental stages: the eight-cell, 110-cell and early tailbud embryos. To render the specimens transparent when necessary, they were dehydrated in ethanol and cleared in a 1:2 mixture (v/v) of benzyl alcohol and benzyl benzoate. Images of stained embryos were acquired with an HC-300Z/OL-2 digital CCD camera (Olympus) mounted on an SZX-12-3121 dissecting microscope (Olympus) and processed using Adobe-Photoshop software. These expression data were compiled in MAGEST. The photo images and the classification data can be found at http://www.genome.ad.jp/magest/

RESULTS AND DISCUSSION

Statistical overview of the maternal RNA repertoire stored in the egg

The ascidian embryo has several advantages as an experimental system (see Satoh et al., 1996 for a review). In particular, one ascidian species, *Halocynthia roretzi*, possesses a few additional advantages. First, they are the largest species among all the ascidians widely used in biological studies and they produce a large number of eggs. Second, in most cases of in situ hybridization with whole-mount *Halocynthia* specimens, signals are first detected in the nuclei of certain blastomeres. This enables us to ascertain clearly when and which cells of the early

embryos express the gene, together with the complete description of the cleavage pattern up to the gastrula stage in *Halocynthia* (Nishida, 1987). Third, more detailed experimental and embryological studies have been possible in this than in any other species (see Nishida, 1997 for a review). Fourth, the egg of this species is as large as 280µm in diameter and easy to manipulate in future functional analyses. Therefore, we constructed a directionally cloned, arrayed plasmid cDNA library from the RNA of the uncleaved fertilized egg of *H. roretzi*.

We determined the sequences of the 4240 cDNA clones from the library. The length of inserts in the library was approximately 2 kb on average and lay mainly in the range from 0.5 kb to 5 kb. The average readable sequence length, on which the following analysis is based, was about 500 bp. All the sequences obtained have been deposited in the public sequence database and made available since the summer of 1999. The sequences obtained were divided into 2221 C clusters consisting of 3832 clones and 408 H clusters. Unless otherwise specified, the following sequence analyses were done based on the C clusters.

The GC contents of all the sequences obtained were calculated for each terminus. The 5′ terminus sequences, which mainly consisted of open reading frames, had 700859 GC bases and 1048999 AT bases(40.1% GC). In contrast, the 3′ terminus sequences without poly(A) tails, most of which are considered to be 3′ UTRs, had 539096 GC bases and 1048707 AT bases (34.0% GC). According to the data collected by Nakamura et al. (Nakamura et al., 2000), the overall GC content of 48447 codons from the coding sequences of *Halocynthia roretzi* is 43.45%, while that of the third codons, which are relatively free from GC/AT-usage restriction, is 41.38%. This suggests that the genome of *Halocynthia roretzi* has a tendency to prefer AT to GC. Taking the lack of general restriction of the use of nucleotides in the 3′ UTRs into account, this may explain the lower GC content in the 3′ termini we obtained.

As we used a non-normalized primary library without amplification, the cluster size, namely the clone abundance, will reflect the relative population of the corresponding

Fig. 1. Sequence prevalence distributions for the fertilized egg ESTs of known and putative mRNA classes. The black bars represent mRNAs with high similarity to the GenBank data. The hatched bars represent sequences with no similarity to any known sequences. The white bars represent sequences with homology to the ESTs or cosmids from other organisms.
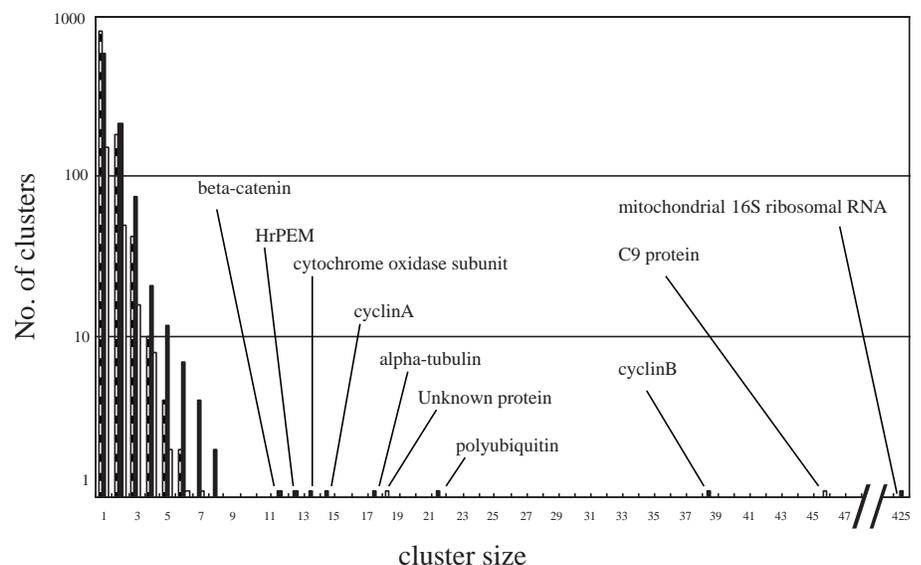
maternal RNAs. Fig. 1 illustrates the prevalence distribution of the cluster sizes. The total numbers of clusters in each cluster size are shown. Four general abundance classes were identified.

(1) The most frequently represented among the redundant clones was mitochondrial 16S ribosomal RNA.

(2) There were only nine clusters that contained more than 12 clones, except for the mitochondrial 16S ribosomal RNA. These nine clusters, representing the most abundant transcripts, constituted 4.7% of the total clones (198 of 4240 clones), but represented just 0.41% of the total clusters (nine of 2221 clusters). The genes of higher prevalence classes encoded proteins such as the C9 protein (unknown mammalian protein located on human chromosome 12 and mouse chromosome 6)(Ansari-Lari et al., 1998), cyclins, ubiquitin and tubulin, most of which are involved in cell division (Fig. 1).

(3) There were 658 medium-sized clusters with two to eight members. These 658 clusters together contained 1699 clones (40.0% of the starting clones) and constituted 29.6% of the total clusters. Interestingly, a cDNA encoding a homolog of CsEPI-1 in *Ciona savignyi*, which is reported to be expressed only zygotically after the 110-cell stage in the epidermal blastomeres (Chiba et al., 1998), was maternally recovered as a member of this class in this study.

(4) The vast majority of the egg poly (A) RNAs were of the rare sequence class, consisting of RNA species appearing only once in the database, which were 1556 clusters containing only 36.7% of the total clones but 70.0% of the total clusters. In Fig. 1, the clusters were separated into sequences of putative mRNAs encoding known peptides, sequences with homology to ESTs or cosmids of other organisms, and sequences with no known similarity. A similar tendency was observed in the medium and low prevalence classes.

Overall identification and categorization of the sequences

The nucleotide sequences obtained and the predicted amino acid sequences were compared with those in the public databases, such as GenBank, to identify significant matches.

Fig. 2 shows the general categories. The total complete sequence data already analyzed were divided into four general categories, defined by the sequence similarity searches: (1) 1610 recognized protein-coding sequences for which there is a high probability that the identifications are meaningful, totaling 42.0% of the 3832 clones and representing 934 clusters (42.1% of 2221 clusters); (2) 403 sequences showing significant similarities to unidentified ESTs or cosmids of other organisms, totaling 10.5% of the total clones, and representing 233 clusters (10.5% of the total clusters); (3) 425 sequences representing the mitochondrial 16S ribosomal RNA, totaling about 11.1% of the total clones; and (4) 1394 sequences belonging to neither of the above classes, totaling about 36.4% of the total clones, and representing 1053 clusters (47.4% of the total clusters). These represent unidentified mRNAs. Comparison of the number of cDNAs analyzed here versus the number of genes estimated to be expressed in the egg shows that the above-obtained percentages must be close to those of the RNA repertoire in the egg.

Table 1 lists examples of the most significant matches discovered by comparing the ascidian sequences to the public databases. Except for the previously isolated ascidian proteins and proteins highly conserved beyond species or phyla, they are not necessarily the *Halocynthia* homologs, but in general they suggest the probable protein families to which the ascidian proteins belong or the proteins sharing some motifs. For further analysis, we attempted a functional classification of the 934 identified clusters containing 1610 clones with significant matches, according to attributable function. We basically made the classification according to a system described by Lee et al. (Lee et al., 1999). Category A includes the proteins required for housekeeping functions including the cell cycle, metabolism and energy transfer. Category B consists of signaling molecules and intercellular communication proteins. Category C contains transcription factors and other nuclear proteins that affect gene regulation. Category D consists of the other proteins that are not classified into the above categories. Table 2 summarizes the proportion and the number of diverse mRNA species represented in each of these categories. Though
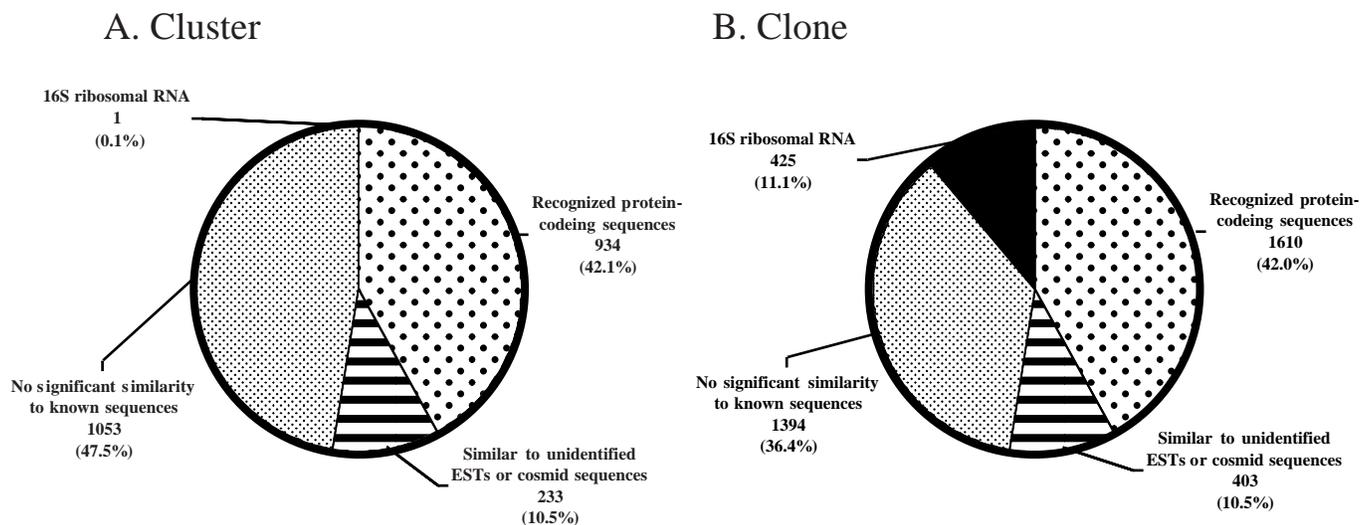
A. Cluster

B. Clone

Fig. 2. Summary of EST sequences. A total of 4240 cDNA clones (B) and 2221 clusters (A) were divided into four general categories, defined by sequence similarity searches.

long suspected, this provided the direct evidence for the high fraction of the maternal RNAs encoding housekeeping machinery. Irrespective of the large numbers of clones and clusters in category A, it consisted mainly of a relatively high prevalence, low complexity sequence set (i.e. low cluster/clone ratio), in contrast to category C, which consisted of a low prevalence, high complexity sequence set (high cluster/clone ratio). Strikingly, 12.8% of the clusters with attributable function represent genes with a potential regulatory role in

development (namely, genes in categories B and C). Comparison of the proportion of the four categories obtained from the ascidian one-cell embryo with that of the sea urchin 7 hour cleavage-stage embryo (Lee et al., 1999) showed a significant difference in the number of genes in category B. Unlike categories A and C, whose numbers were comparable in the two analyses, the ascidian embryo expressed fewer RNAs for cell-cell communication proteins. This may reflect both the stage difference between the one-cell and cleavage-

Table 1. Clones with significant database matches

Cluster ID	Sequence ID	Description	Similarity
C00001	gb:AB024528	*Halocynthia roretzi* mitochondrial DNA	0.0
C00036	gb:AB003083	*Halocynthia roretzi* mRNA for calmodulin, complete cds	0.0
C00043	gb:D89794	*Halocynthia roretzi* mRNA for beta-tubulin, complete cds	0.0
C00170	gb:AB018106	*Halocynthia roretzi* mRNA for HrSmad1/5, complete cds	0.0
C00238	gb:AB029334	*Halocynthia roretzi* mRNA for HrPET-1, complete cds	0.0
C00673	gb:AB029335	*Halocynthia roretzi* mRNA for HrPET-3, complete cds	0.0
C00309	gb:AB029333	*Halocynthia roretzi* mRNA for HrPET-2, complete cds	0.0
C02016	gb:AB001327	*Halocynthia roretzi* mRNA for HrNotch protein, complete cds	0.0
C02137	gb:AB018796	*Halocynthia roretzi* gene for calmodulin A, complete cds	0.0
C01859	gb:AB003695	*Halocynthia roretzi* mRNA for RBP-Jkappa protein, complete cds	0.0
C00221	gb:AF076517	*Molgula occulta* clone MoccCA1 cytoskeletal actin 1 mRNA, complete cds	1e-127
C00045	gp:HROHRATL1_1	ADT/ATP translocase (*Halocynthia roretzi*)	1e-121
C00065	gp:AB030275_1	Musashi (*Halocynthia roretzi*)	1e-113
C00184	gp:HUMCSNK1A_1	Casein kinase I-alpha (*Homo sapiens*)	1e-115
C00027	gp:AF001363_1	Protein phosphatase-1 (*Herdmania curvata*)	1e-112
C02480	gp:AF044169_1	Erythrocyte histone deacetylase (*Gallus gallus*)	1e-106
C00075	gpu:AF180350_1	Protein phosphatase 2A B regulatory subunit delta isoform (*Rattus norvegicus*)	1e-105
C00075	gpu:AF180350_1	Protein phosphatase 2A B regulatory subunit delta isoform (*Rattus norvegicus*)	1e-105
C02325	gp:AB012667_1	CsCDC42 (*Ciona savignyi*)	1e-101
C00090	gpu:AF190909_1	α actinin 4 (*Rattus norvegicus*)	9e-98
C00140	gp:HSALDC_1	Aldolase C (*Homo sapiens*)	3e-96
C00085	gp:HSTIF_1	Translation initiation factor (*Homo sapiens*)	4e-94
C00239	gp:HSPOLYAB_1	Human mRNA for polyA binding protein (*Homo sapiens*)	7e-87
C00122	gpu:AE003766_2	Noa36 gene product (*Drosophila melanogaster*)	2e-84
C00200	prf:0710290A	Synthase, citrate - *Sus scrofa* domestica (domestic pig)	1e-83
C00216	gp:HSSADSYNA_1	S-adenosylmethionine synthetase (*Homo sapiens*)	6e-82
C00226	gp:PIGMOESB_1	Moesin B (*Sus scrofa*)	2e-73
C00113	gp:RAT90KDGK_1	90 kDa-diacylglycerol kinase (*Rattus norvegicus*)	3e-72
C00171	gpu:AB031544_1	Glycogen synthase kinase (*Ciona intestinalis*)	1e-70
C00225	gp:BTIDNADP_1	Isocitrate dehydrogenase (NADP+) (*Bos taurus*)	2e-69
C00128	gp:MABHKP23_1	Integral membrane protein BHKp23 (*Mesocricetus auratus*)	3e-68
C00142	gp:AF042838_1	MEK kinase 1 (*Homo sapiens*)	3e-68
C00079	gp:HUMNPC15_1	Niemann-Pick C1 protein (*Homo sapiens*)	6e-66
C00247	gp:AF038009_1	Tyrosylprotein sulfotransferase-1 (*Homo sapiens*)	2e-65
C00154	gp:AF134570_1	Exonuclease ExoI (*Xenopus laevis*)	7e-63
C00145	gp:MMU7012_1	Fish protein (*Mus musculus*)	4e-62
C00223	gpu:AE003500_38	Gbeta13F gene product (*Drosophila melanogaster*)	6e-61
C00215	gpu:AF178987_1	Protein-tyrosine phosphatase D (*Homo sapiens*)	1e-57
C00220	gp:AB003102_1	26S proteasome subunit p44.5 (*Homo sapiens*)	2e-54
C00180	gpu:AE003532_8	Pdi gene product (*Drosophila melanogaster*)	2e-53
C00209	gp:MMU26460_1	ICAp69 (*Mus musculus*)	2e-53
C00176	gpu:HSM800543_1	Hypothetical protein (*Homo sapiens*)	8e-53
C00150	gpu:AF095639_1	Uroporphyrinogen decarboxylase (*Danio rerio*)	7e-53
C00127	gpu:CFU55935_1	ZO-1 MDCK (*Canis familiaris*)	6e-52
C00240	gp:HSU94831_1	Multispanning membrane protein (*Homo sapiens*)	2e-51
C00146	gpu:AE003819_27	CG4062 gene product (*Drosophila melanogaster*)	7e-50
C00108	gp:AB032081_1	Phosphoacetylglucosamine mutase (*Homo sapiens*)	5e-50
C00125	gpu:AF233334_1	Soluble epoxide hydrolase (*Homo sapiens*)	7e-49
C00126	gp:AB020319_1	Ximpact (*Xenopus laevis*)	4e-49
C00178	gp:AB000407_1	DNA topoisomerase I (*Gallus gallus*)	3e-48
C00103	gp:AF118128_1	Nuclear RNA helicase Bat1 (*Mus musculus*)	4e-46
C00138	gp:HUMALAD_1	ALAD gene product (*Homo sapiens*)	1e-45
C00089	pir:A54601	RNA-binding protein AUF1 - human	4e-44
C00101	gpu:CEW02B12_8	W02B12.9 gene product (*Caenorhabditis elegans*)	4e-42
C00217	gpu:AE003661_5	CG10621 gene product (*Drosophila melanogaster*)	3e-41
C00242	gpu:AE003745_6	Pli gene product (*Drosophila melanogaster*)	4e-38
C00218	gp:HUMECHD_1	Enyol-CoA: hydratase/3-hydroxyacyl-CoA dehydrogenase (*Homo sapiens*)	1e-31
C00235	gp:HUMTLEI_1	Transducin-like enhancer protein (*Homo sapiens*)	1e-31

stage embryos and the embryological difference between the mosaic and regulative natures of the two species.

In Table 3, several examples of the *Halocynthia* clones classified into the above categories are listed. Several cDNA clones that have been reported to be expressed maternally in the egg were reasonably recovered. One of the most practical products of any EST project is the new probes obtained for interesting gene homologs of the animal. Among these, the most interesting genes are likely to be those concerning cell signaling and transcriptional control. There are several such discoveries listed in Table 3, among which we have reported 001A03, *HrWnt-5* (Sasakura et al., 1998a), and 002C13, *Hrsmad1/5*, which is assumed to transduce BMP signaling (Kobayashi et al., 1999), for example. The presence of these maternal transcripts encoding signaling molecules suggests that signaling mediated by these proteins will soon occur in cleavage-stage embryos, and that these proteins may be involved in early specification events. We also found sequences similar to those of various transcription factors that have not been previously isolated from the ascidian embryos, such as groucho and Sin3. In addition, 25 clusters whose products were predicted to contain zinc-finger motifs were found (Table 4).

Ten of them had zinc-finger motifs of the C2H2 type, which are expected to bind to DNA as transcription factors. As one of the most important roles performed by maternal factors stored in eggs is to control subsequent zygotic gene expression during early embryogenesis as transcriptional regulators, they may function in the cell-type diversification processes. All of these findings have new implications for the functional activities of the ascidian egg. The mechanism(s) by which these maternal transcription factors are modified and/or sequestered so that they become functional at the proper developmental stages in the proper blastomeres remains elusive.

In category D, we found a cluster consisting of two clones with a high similarity to the transposase MER37 of the human transposable element, *Tigger1*. *Tigger1* belongs to the pogo-class DNA transposon superfamily, which is distantly related to the *Tc1/mariner* superfamily (Smit and Riggs, 1996). This is the first evidence that a sequence encoding a transposase, which may originate from an endogenous transposon, is transcribed in ascidian cells.

Sequences constituting about 47.5% of the total clusters displayed no significant similarity to any known sequences or ESTs in the public databases. This is probably due to two reasons: first, the ascidian sequences are too divergent to have any similarity to the sequences registered in the public
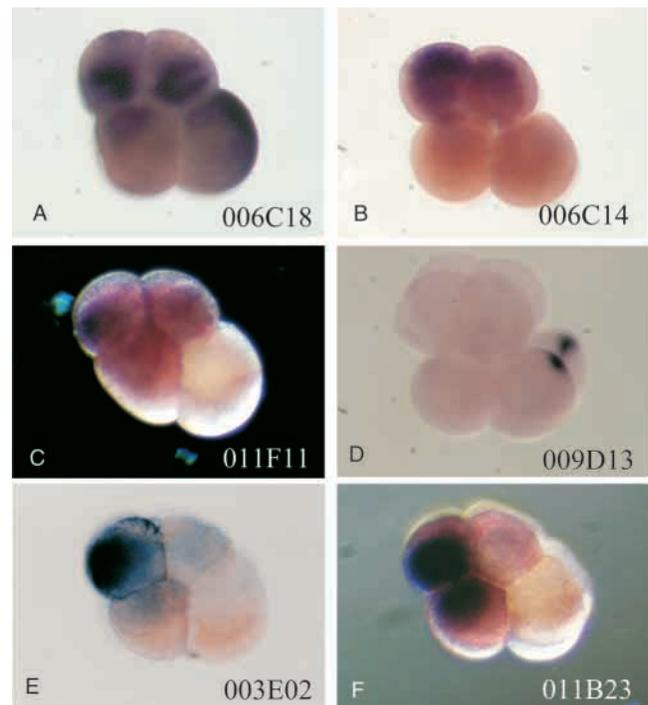
Table 2. The number of different genes expressed by class

Class		Number of clones	Number of clusters
	A. Functions that many kinds of cells use		
I	Metabolic enzymes: ion or electron transfer proteins (mostly mitochondria genome genes)	60	23
II	Metabolic enzymes: not categorized in AI	368	242
III	Cytoskeleton and motor protein	83	37
IV	Chromosome modification	6	5
V	DNA modification	18	12
VI	Cell cycle	111	35
VII	RNA processing	33	17
VIII	Translation	21	14
IX	Protein degradation	41	11
X	Kinase	23	16
XI	Phosphatase	38	21
	Total	**802 (49.8%)**	**433 (46.3%)**
	B. Signaling molecules and inter/intra-cellular communication proteins		
I-a	Signaling molecules, (e.g. Wnt or Smad)	58	27
I-b	Signaling molecules: MAP cascade	18	11
I-c	Signaling molecules: G-protein related	5	4
II	Cell-cell adhesion	23	8
	Total	**104 (6.5%)**	**50 (5.4%)**
	C. Transcription factors and other gene regulatory proteins		
I	Basic transcription factors	15	12
II	Transcription factors	45	33
III	DNA-binding motif containing proteins	9	8
IV	RNA-binding proteins	23	16
	Total	**92 (5.7%)**	**69 (7.4%)**
	D. Others: not enough information to classify	**612 (38.0%)**	**382 (40.9%)**
	Total number of different proteins recognized	**1610 (100.0%)**	**934 (100.0%)**

Fig. 3. Examples of clones showing maternal localization at the eight-cell stage. (A) 006C18. A clone categorized as mitochondria-like. (B) 006C14. The signals are seen only in the animal blastomeres. (C) 011F11. The signals are detected in the a4.2, b4.2 and A4.1 blastomere pairs, but not in the B4.1 blastomere pair. (D) 009D13, which encodes HrPEM. The signals are localized in the postplasm. (E) 003E02. The signals are detected only in the a4.2 blastomere pair. (F) 011B23. The signals are seen in the a4.2 and A4.1 anterior blastomere pairs. Lateral views. Anterior is towards the left. The animal pole is upwards.

databases represented by GenBank, which consist mainly of mammalian genes. Second, this reflects the high complexity of mRNA species in the egg, as is the case for the mammalian preimplantation embryo (Ko et al., 2000).

General classification of the localization of maternal RNAs in early embryos

Whole-mount in situ hybridization was carried out to collect information about the localization and/or expression sites of maternal mRNAs, which is important for understanding the outline of the gene functions in developmental mechanisms. To determine the distribution of the maternal RNAs, the eight-cell embryo was used, because it is easier to identify the orientation along the animal-vegetal and anterior-posterior axes in the eight-cell embryo than in the uncleaved egg. We must emphasize that large-scale expression screening by whole-mount in situ hybridization is extremely effective for gaining a global overview and identifying candidate clones that show particular localization, but is not very suitable for determining detailed expression profiles of the genes. These features are analogous to single-pass sequencing in EST projects. We classified the expression data according to each bilateral

Table 3. Examples of ESTs based on sequence similarity

A. Functions that many kinds of cell use

AI: Metabolic enzymes: ion or electron transfer proteins (mainly mitochondria genome genes)

Cluster	Description	Cluster size
C00009	Cytochrome oxidase subunit II	14
C00021	NADH dehydrogenase subunit 5	7
C00710	Apocytochrome b	2
C01023	ATP synthase subunit 6	1

AII: Metabolic enzymes: not categorized in AI

Cluster	Description	Cluster size
C00045	ADP/ATP translocase HrATL1	8
C00110	Ribonucleoside reductase	3
C00049	Acyl-coA dehydrogenase	4
C00125	Epoxide hydrolase	3
C00130	β-Hydroxyisobutyryl CoA hydrolase	2
C00138	δ-Aminolevulinate dehydrogenase	2

AIII: Cytoskeleton and motor protein

Cluster	Description	Cluster size
C00019	α-tubulin	18
C00221	β-actin	2
C00078	Dynein light chain	3
C01874	Kinesin light chain	1
C00453	Kinesin-related protein XKIF2	1
C01201	Myosin heavy chain non muscle form A	1

AIV: Chromosome modification

Cluster	Description	Cluster size
C00756	HistoneH3	2
C01192	HistoneH2A	1
C02164	CRM1/XPO1	1

AV: DNA modification

Cluster	Description	Cluster size
C00178	DNA topoisomerase I	2
C00199	Replication factor C	2
C00257	Calumenin	6
C01762	Exonuclease	1

AVI: Cell cycle

Cluster	Description	Cluster size
C00005	Cyclin B	39
C00006	Cyclin A	15
C00430	Cyclin C	1
C01766	Brahma protein	1
C00415	CDC2	1

B. Signaling molecules and inter/intra-cellular communication proteins

BI-a: Signaling molecules

Cluster	Description	Cluster size
C00170	HrSmad1/5	5
C02016	HrNotch	1
C00264	Hrwnt-5	3
C00299	STAM	3
C01859	RBP-Jkappa protein	2
C02392	SH3 domain binding protein	1

BI-b: Signaling molecules: MAP cascade

Cluster	Description	Cluster size
C00104	MAP kinase	3
C00803	Rab6	2
C01522	Rab15	2
C01057	Rho1	1

BI-c: Signaling molecules: G-protein related

Cluster	Description	Cluster size
C00787	G-protein-coupled receptor	1
C00958	G-protein signaling regulator	2

BII: Cell-cell adhesion

Cluster	Description	Cluster size
C00010	β-catenin	12
C01269	α N-catenin	3
C00405	M cadherin	6
C01532	Integrin α6	1

C. Transcription factors and other gene regulatory proteins

CI: Basic transcription factors

Cluster	Description	Cluster size
C00864	TFIID	2
C01429	RNA polymerase II subunit 4	1
C00901	RNA polymerase II SIII p18 subunit	1

CII: Transcription factors

Cluster	Description	Cluster size
C00109	Sp3	2
C00235	Groucho	2
C00754	Transcription repressor CTCF	1
C01409	Sin3	1
C00504	As-mT	1
C02276	Y box-binding protein 1	2
C01487	MTF-1	1

CIII: DNA-binding motif-containing proteins

Cluster	Description	Cluster size
C00524	PurA	1
C00869	NEFA	1
C01155	CDR62	1
C01674	MEL-18	1

CIV: RNA-binding proteins

Cluster	Description	Cluster size
C00065	HrMsi1	3
C00401	RNA-binding protein TIAR	2
C00089	RNA-binding protein AUF1	4
C00239	Poly(A)-binding protein	2
C01816	Apoptosis-related RNA binding protein	1

D. Others: not enough information to classify

Cluster	Description	Cluster size
C00312	Transposase MER37	2

Table 4. List of clusters carrying zinc-finger motifs

Cluster ID	Cluster size	Sequence ID	Description	Type*
C00033	6	sp:Z207_HUMAN	Zinc-finger protein 207	C2H2
C00109	2	sp:SP4_HUMAN	Transcription factor SP4 (SPR-1)	C2H2
C00241	2	prf:2421453B	Zf9 protein	Kruppel-like C2H2
C00595	2	sp:MTF1_MOUSE	Transcription factor MTF-1	C2H2
C01487	1	pir:S35335	Transcription factor MTF-1	C2H2
C01613	1	pir:S35335	Transcription factor MTF-1	C2H2
C00382	1	gp:AF192913_1	Zinc-finger protein ZNF180	C2H2
C00754	1	sp:CTCF_HUMAN	Transcription repressor CTCF	C2H2
C01109	1	prf:2007304A	ZFM1 gene	(CCHC)
C02056	1	prf:2310402C	Bop gene	mynd C6
C02062	1	gp:AB015427_1	Zinc-finger protein 219	Kruppel-like C2H2
C02175	1	pir:A33633	Transcription repressor protein GCF	(leucine zipper C4)
C02178	1	pir:G02919	Transcription factor ZFM1	(CCHC)
C02357	1	prf:2404365A	Zinc-finger protein Noc2	PHD C6
C02421	1	pir:S69214	Deformed epidermal autoregulatory factor-1	(mynd C6)
C02436	1	prf:2312346A	Core promoter-binding protein	(C2H2)
C01047	1	gp:MMY12880_1	Polycomb-M33 interacting protein Ring1B	Ring C3HC4
C01674	1	pir:A37991	DNA-binding protein Mel-18	C3HC4
C00687	2	pir:A47156	Hexamer-binding protein HEXBP	CCHC
C00122	1	gp:CGZNFING_1	Zinc-finger factor	CCCH
C02069	1	pir:T41264	zinc-finger protein	CCCH
C02144	1	pir:A40448	DNA-binding protein CDR-62	(leucine zipper C4)
C02306	1	prf:1905390A	Zinc-finger protein	(C6)
C02458	1	gp:AF056617_1	BWSCR2 associated zinc-finger protein BAZ1	C2H2
C01649	1	gb:AB029332	*Halocynthia roretzi* mRNA for HrZF-1	CCCH

*Types of zinc-finger motifs are indicated. If the zinc-finger motif is not included in the EST, the types seen in the sequence showing the highest similarity are indicated in parentheses.

blastomere pair at this stage, a4.2 (anterior-animal), b4.1 (posterior-animal), A4.2 (anterior-vegetal) and B4.1 (posterior-vegetal). Furthermore, categories of 'mitochondria-like' and 'postplasm' were added, based on the initial preliminary results, in which we found many clones that otherwise would be classified into the category B4.1. Mitochondrial large ribosomal RNA (mtlrRNA) encoded by the mitochondrial genome contains a poly (A) tail and is abundant in egg cytoplasm. It is abundant in the myoplasm of the posterior-vegetal region in eggs and embryos of *H. roretzi* (Oka et al., 1998; Fig. 3A). The postplasm is a small region located at the posterior-most part of the embryo (Sasakura et al., 1998a; Fig. 3D). Photo images of the maternal localization of each clone are available and can also be searched for, according to localization patterns, using MAGEST via the Internet.

As shown in Table 5, among the 2626 clones representing 1206 clusters for which we examined the localization sites at the eight-cell stage, the vast majority (40.0%: 1047/2626) were distributed ubiquitously in the

embryo. The largest category showing localization was mitochondria like, and contained 494 mitochondrial cDNA clones (Fig. 3A). There were, however, three large localization clusters containing heterogenous populations of clones; the first showed localization in the animal half of the embryo, totaling 43 clusters of the analyzed clones, although this number was vague and can be perturbed by the threshold employed for localization (Fig. 3B). The second group was 58 clusters that were widely distributed throughout the eight-cell embryo except in the B4.1 blastomere pair (Fig. 3C). The third group showed clear localization in the postplasm of the posterior-vegetal blastomere, and totaled 28 clusters, including type I and II postplasmic RNAs such as *HrWnt-5*, *HrPOPK-1*, *HrZF-1*, and *HrPet-1, HrPet-2 and HrPet-3* – reported previously (Sasakura et al., 1998a; Sasakura et al., 1998b; Sasakura et al., 2000). Table 6 lists all the clones localized in the postplasm and their molecular identities. The most prevalent in this group was *Hrpem*, which is a *Halocynthia* homolog of the gene reported in *Ciona savignyi* by Yoshida et al. (Yoshida et al., 1996; Fig. 3D). The maternal transcripts of all the genes in this category showed the same distribution as *pem* throughout embryogenesis: the transcripts were segregated into the posterior-most blastomere pair during the cleavage stage and were partitioned into a pair of endodermal strand cells in the posterior tip of the tail. As the posterior cytoplasm of the eight-cell embryo possesses multiple functions, such as muscle formation, anterior-posterior axis specification (Nishida, 1996), and generation of the difference in the responsiveness of the notochord and mesenchyme precursor blastomeres (Kim et al., 2000), these postplasmic clones are expected to play some roles in these processes. In addition to these major localization patterns, a few minor localization groups such as the a4.2 blastomere pair

Table 5. Expression patterns in the eight-cell embryos

Expression patterns	Number of clones		Number of clusters	
Overall staining	75	(2.9%)	52	(4.3%)
Almost all regions	972	(37.0%)	458	(38.0%)
Low level expression (not stained)	734	(28.0%)	502	(41.8%)
Mitochondria like	494	(18.8%)	31	(2.6%)
Animal half (a4.2, b4.2)	33	(1.3%)	43	(3.6%)
Except B4.1 (a4.2, b4.2, A4.1)	119	(4.5%)	58	(4.8%)
Postplasm	92	(3.5%)	28	(2.3%)
a4.2	28	(1.1%)	16	(1.3%)
Anterior half (a4.2, A4.1)	79	(3.0%)	16	(1.3%)
Total	**2626**	**(100.0%)**	**1204**	**(100.0%)**

(Fig. 3E) and the two anterior blastomere pairs (Fig. 3F) were found.

Contrary to our expectations at the beginning that many RNA species might be localized in each of the cytoplasmic domains such as myoplasm, ectoplasm, endoplasm, chymoplasm, caudal chymoplasm and chorda-neuroplasm (Conklin , 1905b), there were only three major and a few minor localization patterns in the eight-cell embryo, as mentioned above. Therefore, localization of maternal RNAs seems to follow a few simple patterns, and these simple patterns may be enough to generate complicated diversification of embryonic cells during development. However, this apparent simplicity may have resulted because we missed other patterns, owing to the limits of sensitivity and resolution in our in situ experiments, or because the localized activities of the ascidian embryo are not due to RNA localization at this stage, but are rather controlled mainly at the protein level.

Global complexity of gene expression during embryogenesis

It has been reported that about 90% of zygotically active genes are also functional during oogenesis in the fruit fly, nematode and sea urchin (Davidson, 1986). Simultaneously with the analysis of the eight-cell embryos, we investigated the spatial expression patterns of zygotic transcription in the 110-cell and early tailbud embryos using the MAGEST cDNAs as probes. The 110-cell embryo represents the stage just prior to gastrulation, and all of its blastomeres face the outer surface of the embryo. The developmental fates of most of the blastomeres are completely restricted to give rise to a single tissue type at this stage (Nishida, 1987). The 110-cell embryo was therefore used to determine the lineage specificity of gene expression. In the early tailbud embryo, the basic chordate body plan is established. The embryo at this stage was used to determine the tissue specificity of gene expression. To make a compromise between large-scale throughput and the depth of analysis, we had to restrict scoring only to major embryonic tissues and two additional locations: epidermis, adhesive organ, brain, nerve cord, A-line notochord, B-line notochord, muscle, mesenchyme, trunk lateral cell, trunk ventral cell, endoderm, endodermal strand, mitochondria-like and postplasm. All the photo images of the in situ results obtained are accessible and can be searched according to expression patterns using MAGEST through the Internet.

As summarized in Table 7, a substantial proportion (29.3%) of the total clusters, including 420 maternal clones, did not show any staining at any stage, including the eight-cell stage (000 in Table 7). This is mainly due to the fact that many of maternal messages belong to the rare abundance class, as mentioned above, for which the expression level is below the limit of detection of the whole-mount in situ hybridization technique. Of the genes analyzed, 7.9% of the clusters seemed to be exclusively maternal (100 and 200 in Table 7), while the remaining 40.6% of the clones showed signals in later stages. There were almost all the patterns in the temporal expression profiles, ranging from 000 to 222, shown in Table 7. Interestingly, for 14 genes, expression was detected only at the 110-cell stage (010 and 020 in Table 7). For example, one of these clones, 001D02,

Table 6. Postplasmic RNAs

Cluster ID	Description	Cluster size
C00011	HrPEM	13
C01333	No similarity	6
C00238	HrPET-1	4
C00202	No similarity	3
C00203	Claudin	3
C00232	HrPOPK-1	3
C00264	Hrwnt-5	3
C00309	HrPET-2	3
C00673	HrPET-3	3
C00954	C-termini PDZ ligand	3
C01181	Galactokinase	3
C00124	No similarity	2
C00133	Similar to some hypothetical protein	2
C00188	Epidermis-specific protein 1 (*Ciona savignyi*)	2
C00191	No similarity	2
C00248	Glucose transporter	2
C00261	Adenylate cyclase	2
C00769	Polypeptide N-acetylgalactosaminyl-transferase	2
C01587	No similarity	2
C02478	No similarity	2
C00613	Casein kinase I alpha S	1
C00730	No similarity	1
C01141	Ximpact protein	1
C01611	MAK11	1
C01649	HrZF-1	1
C01703	No similarity	1
C02347	Syndecan (*Ciona savignyi*)	1
C02356	Similar to some hypothetical protein	1
		Total 73
		(28 clusters)

showed transient expression exclusively in the presumptive notochord cells in the 110-cell embryo (see Fig. 6A).

Fig. 4 shows an overview of the frequency of gene expression in the different tissues scored in the early tailbud embryos. The largest proportion (31%) of the genes were expressed in the central nervous system (brain and nerve cord), followed by mesenchyme and the epidermis. It should be noted
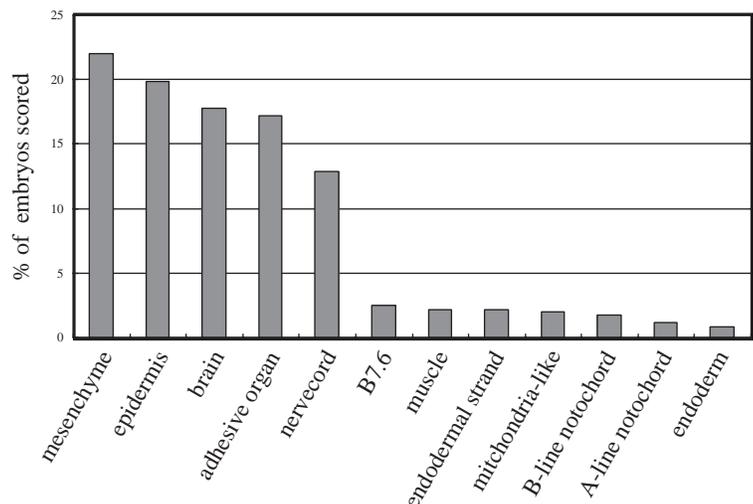

Fig. 4. Frequency of zygotic gene expression in different tissues. The 1206 unique genes that are expressed in the indicated tissues of the tailbud embryos are represented. Note that a single gene can be expressed in multiple tissues.

Table 7. Temporal expression patterns*

Eight cell	110 cell	Early tailbud	Number of clusters
0	0	0	293
0	0	1	8
0	0	2	58
0	1	0	10
0	1	1	9
0	1	2	14
0	2	0	4
0	2	1	0
0	2	2	23
1	0	0	17
1	0	1	13
1	0	2	8
1	1	0	5
1	1	1	23
1	1	2	8
1	2	0	0
1	2	1	2
1	2	2	6
2	0	0	62
2	0	1	14
2	1	2	43
2	1	0	21
2	1	1	31
2	1	2	89
2	2	0	29
2	2	1	25
2	2	2	184
			Total 999

*0, Not stained; 1, weakly stained; 2, Intensely stained.

that a single gene is sometimes expressed in multiple tissues. Table 8 classifies the zygotic expression according to how many tissues and which tissues the staining was detected in. One of the most useful aspects of this kind of project is finding new probes for interesting genes that are expressed in a cell type-specific manner. Of the genes mentioned above, 13.2% of the clones showed tissue-specific expression. Fig. 5 shows examples of such clones. Many genes were identified whose expression pattern is so restricted that they may serve as useful markers of tissue differentiation. Among them were tissues for which no molecular probes were previously available, such as the adhesive organ, a subset of neural cells and trunk lateral cells, for example (Fig. 5B,C,F, respectively).

The complexity of the zygotic gene expression at the 110-cell and early tailbud stage is broad, from the cell-type specific, as mentioned above, to the complex expression in multiple, ontogenetically unrelated regions (e.g. 003O15, Fig. 6B). Some genes showed widespread expression in many tissues but lack of expression in one tissue (e.g. 004D21, Fig. 6C). This observation must be attributed to the pleiotropic and/or borderless functions of the gene products in the context of tissue types. For example, the zygotic expression of genes whose maternal RNA showed postplasmic localization was detected in a variety of cell types in the early tailbud embryo (Table 6, Fig. 6D-F). These maternal gene products may serve various purposes in the postplasm in the early stages, and the zygotic gene products may function differently in various cell types.

A discussion of the genes showing a variety of interesting expression patterns is largely beyond the scope of this report. However, as one of fruitful results of the clustering analysis of the expression profiles, we refer to a group of genes that are zygotically expressed in a pair of B7.6 cells in the early tailbud embryo. The endodermal strand is a line of cells located in the ventral midline just beneath the notochord in the tail. These cells are derived from three pairs of blastomeres of the 110-cell embryo, B7.2, b8.17 and B7.6 (Nishida, 1987). The main source of the tissue is B7.2, which contributes most of the endodermal strand cells. There are another two supplementary sources of this tissue. The first is a cell located at the tip of the tissue, which is derived from one of the b8.17 pair, and the second is B7.6, which is the posterior-most blastomere at the 64-cell stage. The bilateral pair of B7.6 does not further divide after the 64-cell stage, and gives rise to the two endodermal strand cells in the posterior region of the tail. All these cells are identified as endodermal strand cells by their position and expression of endoderm/endodermal strand markers, such as clone 003D22, for example (Fig. 5H). However, little is known about the cellular differences among the endodermal strand

Table 8. Classification of zygotic expression

Number of stained tissues	Number of clusters	(%)		Number of clusters	(%)
	Total 999	**(100.0)**			
Not expressed	565	(56.6)			
Expressed in					
>4 tissues	50	(5.0)			
4 tissues	68	(6.8)			
3 tissues	93	(9.3)			
2 tissues	91	(9.1)			
1 tissue	132	(13.2)	**Subtotal 132**		**(100.0)**
	Ectoderm				
			Epidermis	60	(45.5)
			Adhesive organ	6	(4.6)
			Brain	8	(6.1)
			Nervecord	2	(1.5)
	Mesoderm				
			Muscle	7	(5.3)
			Notochord	2	(1.5)
			Mesenchyme	36	(27.3)
	Endoderm				
			Endoderm	0	(0.0)
			Endodermal strand	1	(0.8)
			Postplasm	10	(7.6)

cells, except for their lineal origins. As a localization study has shown that B7.6 carries many maternal products, such as postplasmic RNAs, including *vasa* RNA (Fujimura and Takamura, 2000), which is a germ-cell marker, there should be specific characteristics of their cellular functions. It was therefore interesting to identify subset-specific zygotic genes by which endodermal strand cells can be distinguished from each other. We found 20 genes that were zygotically expressed in the B7.6 pair, but not in the other cells in the endodermal strand of the tailbud embryo (Table 9). Maternal mRNAs of most of these genes were not localized at the posterior end, but rather were present ubiquitously during the early stages (Fig. 7A). All these genes had various expression sites other than B7.6: some genes were expressed in neural tissues, some in mesenchyme and others in the epidermis (e.g. 003D19, 003C20, 003H03 in Fig. 7B-D, respectively). Except for the expression in the B7.6 blastomere pair, there was no common feature among these genes in terms of other expression sites or the timing of the onset of zygotic expression. These observations suggest that these genes do not define cell types, but rather that they play pleiotropic roles and/or roles

independent of the tissue type. To investigate the characteristics of B7.6 endodermal strand cells, further studies of these genes will be necessary.

Perspectives

We have described the large-scale analysis of the maternal

Fig. 6. Examples of whole-mount in situ hybridization. (A) 001D02 in the 110-cell embryo. Vegetal view. Anterior is upwards. (B) 003O15 in the early tailbud embryo. The signals are seen in multiple regions in a complex manner in terms of cell lineage or germ layers. Anterior is towards the left. (C) 004D21 in the 110-cell embryo. The expression is detected in every blastomere except the endodermal cells. (D) 011F04. (E) 008B07. (F) 006L20. The maternal RNA of each clone is localized in the postplasm while the zygotic expression is seen in the epidermis and adhesive organ (D), the nervous system, adhesive organ and endodermal strand in (E) and in the trunk lateral cells (F), respectively.
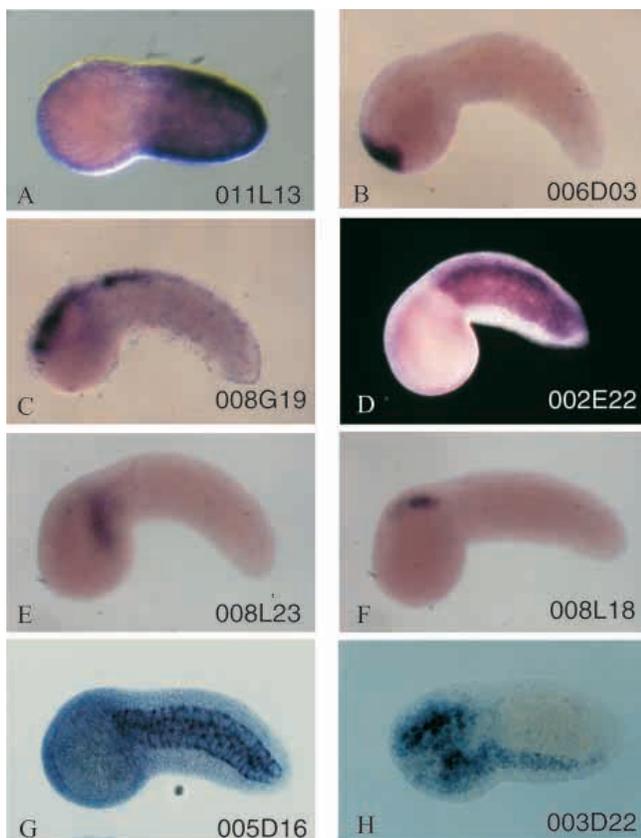

Fig. 5. Examples of the genes zygotically expressed in a single tissue in the tailbud embryos. (A) 011L13, which is expressed in the tail epidermis. (B) 006D03, which is expressed in the presumptive adhesive organ. (C) 008G19, which is expressed in a subset of the neural cells in the central nervous system. (D) 002E22, which is expressed in the larval tail muscle. (E) 008L23, which is expressed in the mesenchyme cells. (F) 008L18, which is expressed in the trunk lateral cells. (G) 005D16, which is expressed in the notochord. (H) 003D22, which is expressed in the endoderm/endodermal strand. Anterior is towards the left.
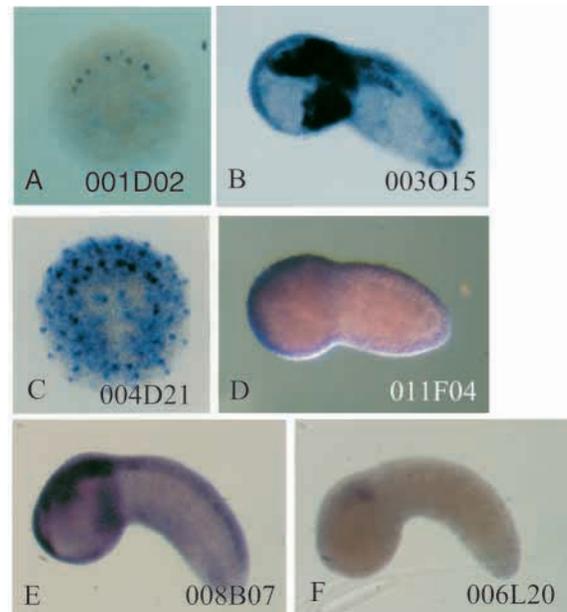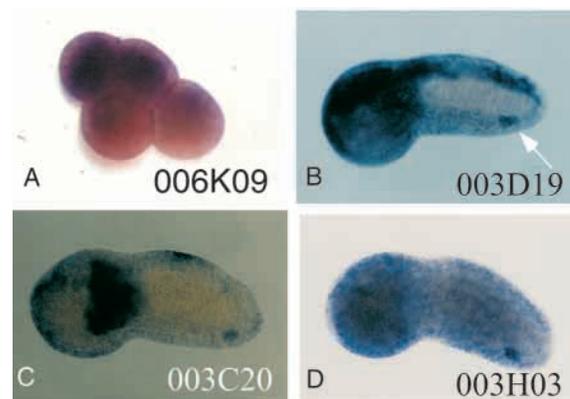

Fig. 7. Examples of the genes zygotically expressed in B7.6 cells. (A) 006K09 (belonging to the same cluster as 003D22 shown in Fig. 5H) expression in the eight-cell embryo. (B) 003D19, which is expressed in B7.6 and the nervous system. Arrow indicates B7.6. (C) 003C20, which is expressed in B7.6 and the mesenchyme. (D) 003H03, which is expressed in B7.6, the epidermis and the endoderm.

Table 9. The clusters expressed in B7.6 cells

Cluster ID	Description	Cluster size
C00045	HrATL1	8
C00027	Protein phosphatase	6
C00172	No similarity	4
C00277	No similarity	3
C01137	Ubiquitin-like protein SMT3A	3
C00837	No similarity	2
C01079	Golgi stacking protein	2
C01236	No similarity	2
C02075	Exostose-related protein 1	2
C00375	Germ cell nuclear factor	1
C00542	No similarity	1
C00941	No similarity	1
C00979	No similarity	1
C01042	No similarity	1
C01765	Golgi Sec1p-like protein	1
C02076	26S proteasome regulatory chain 4	1
C02098	Exostosin 1	1
C02121	Uracil-DNA glycosylase	1
C02148	No similarity	1
C02175	Transcription repressor protein GCF	1
		Total 43
		(20 clusters)

genetic repertoire stored in the ascidian egg in terms of the sequences and localization and/or expression patterns in early development. We analyzed 4240 cDNA clones representing the 2629 clusters according to our definition. Many genes with interesting similarities to known proteins and with clearly discernable expression patterns were identified. These data will also serve as resources for conventional studies to elucidate the molecular mechanisms of development. Therefore, we next need to study the functions of these genes by tests of physiological relevance during development, which will be approached by gain- or loss-of-function experiments. In the past decade, several significant advances in technology have allowed us to modify gene expression: gene introduction by microinjection (Hikosaka et al., 1992) and electroporation (Corbo et al., 1997), the suppression of zygotic gene expression by an antisense oligonucleotide (Swalla and Jeffery, 1996) and the overexpression of an exogenous gene by synthetic mRNA injection (Yoshida et al., 1996) have all been applied to the ascidian embryology. However, further technical innovation and the development of new methods for suppressing the activities of specific maternal RNAs and proteins are needed in order to assess their functions. Recently, we have been able to deplete specific maternal RNAs using antisense oligonucleotides (Nishida and Sawada, 2001). Extensive studies of the maternal genetic information using loss-of-function techniques of this kind will shed light on the molecular mechanisms of ascidian embryogenesis.

This project will provide a resource for future research that will facilitate the investigation of the maternal factors involved in early developmental events, and consequently, comprehensive studies of the maternal genetic information. This cannot otherwise be achieved by specific investigations on particular phenomena during embryogenesis. Finally, this study should also facilitate the elucidation of the molecular mechanisms of the establishment of embryonic body plans of chordates and lead to further understanding of the evolution from invertebrates to vertebrates.

REFERENCES

Ansari-Lari, M. A., Oeltjen, J. C., Schwartz, S., Zhang, Z., Muzny, D. M., Lu, J., Gorrell, J. H., Chinault, A. C., Belmont, J. W., Miller, W. and Gibbs, R. A. (1998). Comparative sequence analysis of a gene-rich cluster at human chromosome 12p13 and its syntenic region in mouse chromosome 6. *Genome Res.* **8**, 29-40.

Bairoch, A. and Apweiler, R. (2000). The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res.* **28**, 45-48.

Barker, W. C., Garavelli, J. S., McGarvey, P. B., Marzec, C. R., Orcutt, B. C., Srinivasarao, G. Y., Yeh, L. L., Ledley, R. S., Mewes, H., Pferffer, F. et al. (1999). The PIR-International Protein Sequence Database. *Nucleic Acids Res.* **27**, 39-43.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A. and Wheeler, D. L. (2000). GenBank. *Nucleic Acids Res.* **28**, 15-18.

Boguski, M. S., Lowe, T. M. J. and Tolstoshev, C. M. (1993). DbEST-database for 'expressed sequence tags'. *Nature* **4**, 332-333.

Bowerman, B., Ingram, M. K. and Hunter, C. P. (1997). The maternal par genes and the segregation of cell fate specification activities in early Caenorhabditis elegans embryos. *Development* **124**, 3815-3826.

Chomczynski, P. and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156-159.

Chabry, L. (1887). Contribution à l'embryologie normale et teratologique des Ascidies simples. *J. Anat. Physiol.* **23**, 167-319.

Chiba, S., Satou, Y., Nishikata, T. and Satoh, N. (1998). Isolation and characterization of cDNA clones for epidermis-specific and muscle-specific genes in *Ciona savignyi* embryos. *Zool. Sci.* **15**, 239-246.

Conklin, E. G. (1905a). Mosaic development in ascidian eggs. *J. Exp. Zool.* **2**, 145-223.

Conklin, E. G. (1905b). Organ-forming substances in the eggs of ascidians. *Biol. Bull.* **8**, 205-230.

Corbo, J. C., Levine, M. and Zeller, R. W. (1997). Characterization of a notochord-specific enhancer from the *Brachyury* promoter region of the ascidian, *Ciona intestinalis*. *Development* **124**, 589-602.

Davidson, E. H. (1986). *Gene Activity in Early Development*. 3rd edn. Orlando, FL: Academic Press.

Ding, D. and Lipshitz, H. D. (1993). Localized RNAs and their functions. *BioEssays* **15**, 651-658.

Fujibuchi, W., Goto, S., Migimatsu, H., Uchiyama, I., Ogiwara, A., Akiyama, Y.,and Kanehisa, M. (1998). DBGET/LinkDB: an integrated database retrieval system. In *Pacific Symposium on Biocomputing '98* (ed. R. B. Altman, A. K. Dunker, L. Hunter and T. E. Klein), pp. 683-694. Singapore: World Scientific.

Fujimura, M. and Takamura, K. (2000). Characterization of an ascidian DEAD-box gene, *Ci-DEAD1*: specific expression in the germ cells and its RNA localization in the posterior-most blastomeres in early embryos. *Dev. Genes Evol.* **210**, 64-72.

Gavis, E. R. (1997). Expeditions to the pole: RNA localization in *Xenopus* and *Dorosophila*. *Trends Cell Biol.* **7**, 485-492.

Gawantka, V., Pollet, N., Delius, H., Virgron, M., Pfister, R., Nitsch, R., Blumenstock, C. and Niehrs, C. (1998). Gene expression screening in *Xenopus* identifies molecular pathways, predicts gene function and provides a global view of embryonic patterning. *Mech. Dev.* **77**, 95-141.

Glotzer, J. B. and Ephrussi, A. (1996). mRNA localization and the cytoskeleton. *Semin. Cell Dev. Biol.* **7**, 357-365.

Hikosaka, A., Kusakabe, T., Satoh, N. and Makabe, K. W. (1992). Introduction and expression of recombinant genes in ascidian embryos. *Dev. Growth Differ.* **34**, 627-634.

Holland, P. W. H., Garcia-Fernandes. J., William, N. A.,and Sidow, A. (1994). Gene duplications and the origins of vertebrate development. *Development* **120 Suppl.**, 125-133.

Jeffery, W. R. (1980). The follicular envelope of ascidan eggs: a site of messenger RNA and protein synthesis during early embryogenesis. *J. Exp. Zool.* **216**, 545-547.

Jeffery, W. R., Tomlinson, C. R. and Brodeur. R. D. (1983). Localization of actin messenger RNA during early ascidian development. *Dev. Biol.* **99**, 408-417.

Kawashima, T., Kawashima, S., Kanehisa, M., Nishida, H. and Makabe, K. W. (2000). MAGEST: maboya gene expression patterns and sequence tags. *Nucleic Acids Res.* **28**, 133-135.

Kim, G. J., Yamada, A. and Nishida H. (2000). An FGF signal from endoderm and localized factors in the posterior-vegetal egg cytoplasm pattern the mesodermal tissues in the ascidian embryo. *Development* **127**, 2853-2862.

Ko, M. S. H., Kitchen, J. R., Wang, X., Threat, T. A., Wang, X., Hasegawa, A., Sun, T., Grahovac, M. J., Kargul, G. J., Lim, M. K. et al. (2000). Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development. *Development* **127**, 1737-1749.

Kobayashi, A., Sasakura, Y., Ogasawara, M. and Makabe, K. W. (1999). A maternal RNA encoding *smad1/5* is segregated to animal blastomeres during ascidian development. *Dev. Growth Differ.* **41**, 419-427.

Laird, C. D. (1971). Chromatid structure: relationship between DNA content and nucleotide sequence diversity. *Chromosoma* **32**, 378-406.

Lee, Y., Huang, G. M., Cameron, R. A., Graham, G., Davidson, E. H., Hood, L. and Britten, R. J. (1999). EST analysis of gene expression in early cleavage-stage sea urchin embryos. *Development* **126**, 3857-3867.

Miklos, G. L. G. and Rubin, G. M. (1996). The role of the genome project in determining gene function: insights from model organisms. *Cell* **86**, 521-529.

Mita-Miyazawa, I., Ikegami, S. and Satoh, N. (1985). Histo-specific acetylcholinesterase development in the presumptive muscle cells isolated from 16-cell stage ascidian embryos with respect to the number of DNA replication. *J. Embryol. Exp. Morphol.* **87**, 1-12.

Nakamura, Y., Gojobori, T. and Ikemura, T. (2000). Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* **28**, 292.

Nishida, H. (1987). Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. III. Up to the tissue-restricted stage. *Dev. Biol.* **121**, 526-541.

Nishida, H. (1996). Vegetal egg cytoplasm promotes gastrulation and is responsible for specification of vegetal blastomeres in embryos of the ascidian, *Halocynthia roretzi*. *Development* **121**, 1271-1279.

Nishida, H. (1997). Cell fate specification by localized cytoplasmic determinants and cell interactions in ascidian embryos. *Int. Rev. Cytol.* **176**, 245-306.

Nishida, H. and Sawada, K. (2001). macho-1 encodes a localized mRNA in ascidian eggs that specifies muscle fate during embryogenesis. *Nature* **409**, 724-729.

Ogasawara, M., Minokawa, T., Sasakura, Y., Nishida, H. and Makabe, K. W. (2001). A large-scale whole-mount in situ hybridization system: rapid one-tube preparation of DIG-labeled RNA probes and high throuhput hybridization using 96-well silent screen plates. *Zool. Sci.* **18**, 187-193.

Oka, T., Amikura, R., Kobayashi, S., Yamamoto, H. and Nishida, H. (1998). Localization of mitochondrial large ribosomal RNA in myoplasm of the early ascidian embryo. *Dev. Growth Differ.* **41**, 1-8.

Perrimon, N., Engstrom, L. and Mahowald, A. P. (1989). Zygotic lethals with specific maternal effect phenotypes in *Drosophila melanogaster*. I. Loci on the X chromosome. *Genetics* **121**, 333-352.

Poustka, A. J., Herwig, R., Krause, A., Hennig, S., Meier-Ewert, S and Lehrach, H. (1999). Toward the gene catalogue of sea urchin development: the construction and analysis of an unfertilized egg cDNA library highly normalized by oligonucleotide fingerprinting. *Genomics* **59**, 122-133.

Reverberi, G. and Minganti, A. (1946). Fenomeni di evocatione nelle sviluppo dell'uovo di ascidie. *Pubbl. Stn. Zool. Napoli* **20**, 199-252

Sasakura, Y., Ogasawara, M. and Makabe, K. W. (1998a). *HrWnt-5*: a maternally expressed ascidian *Wnt* gene with posterior localization in early embryos. *Int. J. Dev. Biol.* **42**, 573-579.

Sasakura, Y., Ogasawara, M. and Makabe, K. W. (1998b). Maternally localizedRNA encoding a serine/threonine protein kinase in the ascidian, *Halocynthia rorezti*. *Mech. Dev.* **76**, 161-163.

Sasakura, Y., Ogasawara, M. and Makabe, K. W. (2000). Two pathways of maternal RNA lozalization at the posterior-vegetal cytoplasm in early ascidian embryos. *Dev. Biol.* **220**, 365-378.

Satoh, N. (1994). *Developmental Biology of Ascidians.* Cambridge, UK: Cambridge University Press.

Satoh, N., Makabe, K. W., Katsuyama, Y., Wada, S. and Saiga, H. (1996). The ascidian embryo: an experimental system for studying genetic circuitry for embryonic cell specifivation and morphogenesis. *Dev. Growth Differ.* **38**, 325-340.

Schnapp, B. J., Arn, E. A., Deshler, J. O. and Highett, M. I. (1997). RNA localization in Xenopus oocyte. *Semin. Cell Dev. Biol.* **8**, 529-540.

Sidow, A. (1996). Gen(om)e duplications in the evolution of early vertebrates. *Curr. Opin. Genet. Dev.* **6**, 715-722.

Simmen, M. W., Leitgeb, S., Clark, V. H. and Jones, S. J. M. and Bird, A. (1998). Gene number in an invertebrate chordate, *Ciona intestinalis*. *Proc. Natl. Acad. Sci. USA* **95**, 4437-4440.

Smit, A. F. and Riggs, A. D. (1996). Tiggers and DNA transposon fossils in the human genome. *Proc. Natl. Acad. Sci. USA* **93**, 1443-1448

Swalla, B. J. and Jeffery, W. R. (1996). Requirement of the *Manx* gene for expression of chordate features in a tailless ascidian larva. *Science* **274**, 1205-1208.

Tabara, H., Motohashi, T. and Kohara, Y. (1996). A multi-well version of in situ hybridization on whole mount embryos of *Caenorhabditis elegans*. *Nucleic Acids Res.* **24**, 2119-2124.

Wieschaus, E. (1996). Embryonic transcription and the control of developmental pathways. *Genetics* **142**, 5-10.

Wilhelm, J. E. and Vale, R. D. (1993). RNA on the move: the mRNA localization pathway. *J. Cell Biol.* **123**, 269-274.

Wylie, C., Kofron, M., Patne, C., Anderson, R., Hosobuchi, M., Joseph, E. and Heasman, J. (1996). Maternal beta-catenin establishes a 'dorsal signal' in early *Xenopus* embryos. *Development* **122**, 2987-2996.

Yamaha, E., Mizuno, T., Hasebe, Y., Takeda, H. and Yamazaki, F. (1998). Dorsal specification in blatoderm in the brastula stage in the goldfish, *Carassius auratus*. *Dev. Growth Differ.* **40**, 267-275.

Yoshida, S., Marikawa, Y. and Satoh, N. (1996). *Posterior end mark*, a novel maternal gene encoding a localized factor in the ascidian embryo. *Development* **122**, 2005-2012.