DNA Research **9**, 209–215 (2002) Short Communication

Abundant Poly(A)-Bearing RNAs That Lack Open Reading Frames in *Schizosaccharomyces pombe*

Takanori Watanabe, Kazuyuki Miyashita, Takamune T. Saito, Kentaro Nabeshima, and Hiroshi Nojima*

Department of Molecular Genetics, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamadaoka, Suita City, Osaka 565-0871, Japan

(Received 18 October 2002; revised 25 November 2002)

Abstract

We report here that 6.9% (68/987) of randomly selected cDNA clones from an *S. pombe* cDNA library lack apparently long open reading frames which we denote *prl*. One of them, *prl1*, was examined further because multiple bands were observed when it was used as a probe in northern blot analysis. These multiple bands appear to be derived from overlapping transcripts from both DNA strands, including non-coding RNAs and antisense RNAs in addition to mRNA. Such mechanisms may increase the transcriptional variation in *S. pombe* cells.

Key words: non-coding RNA; anti-sense RNA; overlapping transcripts; multiplex transcription; *S. pombe*

Eukaryotic genomes adopt the following strategies to increase the variation of transcripts; overlapping transcription derived from both strands,[1,2] overlapping reading frame in one strand,[3] transcripts derived from an intron of another transcript,[4] alternative splicing, *trans*-splicing[5–8] and translational frame shifting.[9] The production of non-coding/antisense RNAs that do not code for proteins also increases the transcriptional repertoire. Non-coding RNAs have been found in many organisms and are known to play critical roles in many biological phenomena.[10–12] In humans and *Drosophila*, non-coding RNAs are important in the regulation of dosage compensation in X chromosome.[13,14] In mice, H19 RNA expressed maternally are essential, acting negatively for growth, whereas Igf2 expressed paternally acts positively for growth.[15]

In *S. pombe*, binding of meiRNA to Mei2 protein is required for progression in meiosis I, which promotes translocation of Mei2 protein from the cytoplasm to the nucleus.[16,17] The coding region of *spo6*$^+$ is transcribed bi-directionally,[18] and three kinds of transcript of the complementary strand of *rec7*$^+$ have been detected.[19] The complete genome sequence of *S. pombe*[20] revealed 4824 annotated protein-coding genes, the smallest among eukaryotes examined to date, amounting to only 85% of the number found in the budding yeast *S. cerevisiae*. The following possibilities were considered to explain this finding: 1) there are fewer duplicated genes (361) than that in *S. cerevisiae* (716), and 2) the distance between the protein-coding regions is generally longer than that in *S. cerevisiae*, and this may contribute to the complex regulation of gene expression.

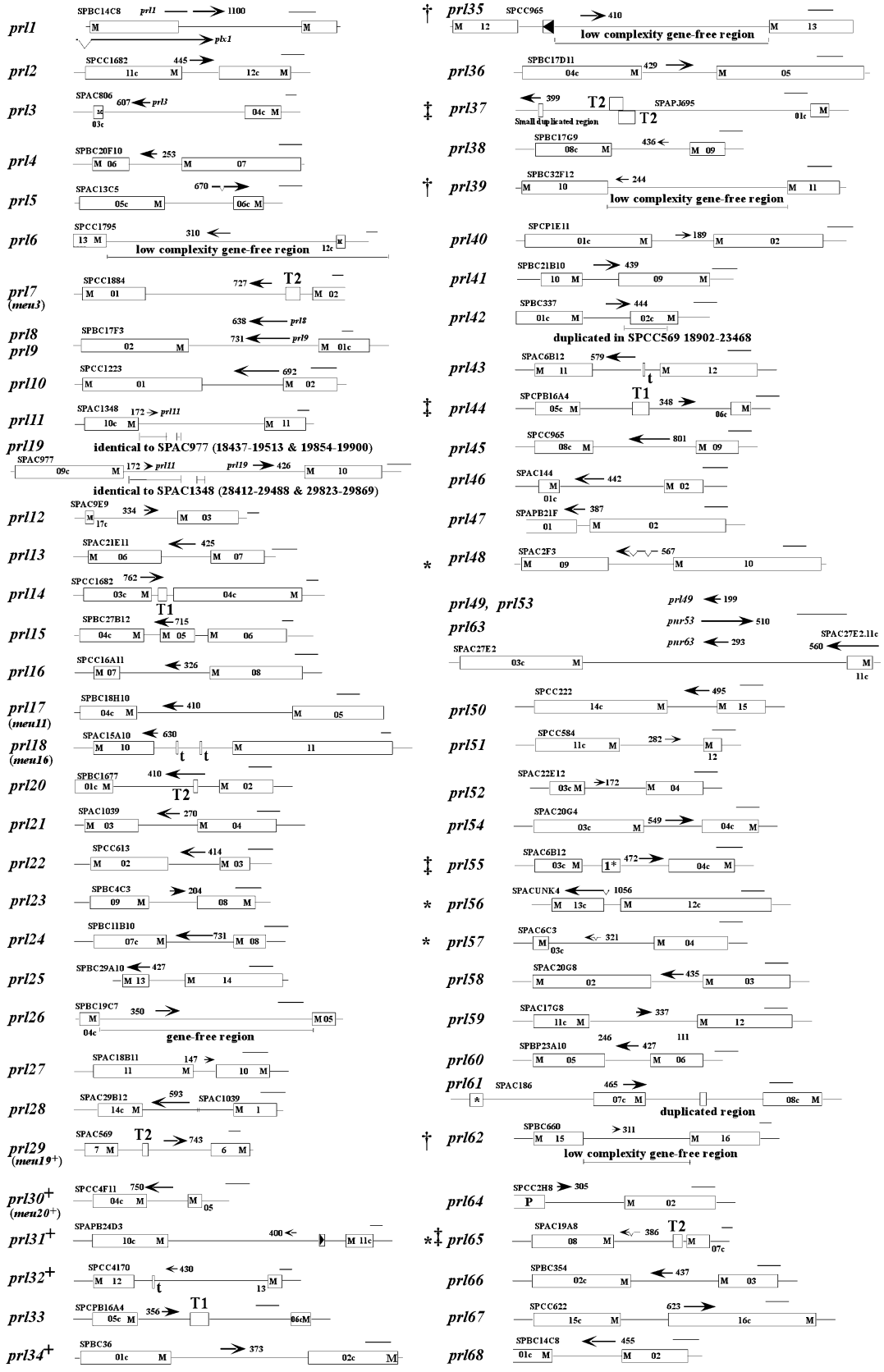
Previously, we isolated 31 kinds of meiosis-specific transcripts named *meu* in *S. pombe* by a cDNA subtraction method.[21] Unexpectedly, 5 out of the 31 *meu* transcripts were estimated to be non-coding/antisense RNAs. Since this finding seems important, we searched for more examples of such non-coding/antisense RNAs in *S. pombe*. Here, we report isolation of cDNA for 68 non-coding/antisense RNAs as well as an example of multiply overlapping transcripts in *S. pombe*.

Of the 987 different cDNA clones, we randomly selected and sequenced cells that are either in mitotic growth phase or in meiosis from a cDNA library prepared using mRNA transcribed in *S. pombe* (see legend for Fig. 1). We found 68 unique clones lacking significant open reading frames (ORFs) that appear to generate non-coding or antisense RNAs species. We denoted these clones *prl* (*poly(A)-bearing RNA without long open reading frames*). Here, "long" means 100 or more amino acids except for *prl25*, which is an antisense RNA of SPBC29A10. Figure 1 demonstrates the locations in the *S. pombe* genome from which these *prl* clones are derived as well as the direction of their transcription (Table 1). The potential CDSs (protein-coding sequence) around these *prl* transcripts as determined by the

Communicated by Hideo Shinagawa

* To whom correspondence should be addressed. Tel. +81-6-6875-3980, Fax. +81-6-6875-5192, E-mail: hnojima@biken.osaka-u.ac.jp

† The sequences reported here have been deposited in the DDBJ database under accession numbers AB084813 to AB084880.



DNA sequence database of *S. pombe* (The Sanger Centre, UK) are shown by boxes with the initiation methionine sites indicated (M) (Fig. 1). We noticed that the Sanger Centre database also contains 43 non-coding RNA species that are annotated as 'mRNA-like miscellaneous RNA' based on comparison of the genomic database with the cDNA clone database. We denoted these species as *prl* here for convenience.

Curiously, several of the *prl*s (*prl5*, *prl48*, *prl56*, *prl57*, and *prl65*) have introns. The maximum ORFs of these *prl*s would code for small proteins that are 29, 85, 70, 25, and 18 amino acids long, respectively. These sequences could actually be translated into these small proteins since the putative peptide from *prl48* has weak similarity to the 49C12.12p protein of *Caenorhabditis elegans*. However, as it is known that the U6 snRNA gene of *S. pombe* also has an intron,[22] these *prl*s could also be such intron-charged non-coding RNA genes.

Using Zuker's computer program with the parameters in the algorithm presented in Jaeger et al.,[23] we found that all of the 68 *prl* transcripts form stable hairpin structures (data not shown), suggesting the idea that the gene products are stable RNA molecules. It is noteworthy that about 20% (14 of the 68 clones) of the *prl* transcripts are derived from gene-free regions of the *S. pombe* genome or near the repetitive sequence, long terminal repeat (LTR). They are indicated in Fig. 1 by daggers (†) and double daggers (‡), respectively. The *prl11* sequence is found in both SPAC1348 and SPAC977, and this is probably a case of transcriptional duplication as is the case of $meu3^+$ (*prl7*) and $meu19^+$ (*prl29*).[21]

We chose *prl1* for further analysis because multiple bands were observed when it was used as a probe in northern blot analysis for *S. pombe* RNA. We dissected the genomic DNA fragment around the *prl1* region with appropriate restriction enzymes to generate the DNA fragments denoted a–i (Fig. 2A). Northern blot analysis using these fragments as probes showed that the probes detected two or more bands with distinct sizes and various patterns for appearance and disappearance in mitotic or meiotic cells (Fig. 2B). The longest transcript, which is 5.8 kb, appears to contain two ORFs (Plx1 and Pac2). The downstream ORF coincides with Pac2 that controls the onset of sexual development.[24] Thus, the 5.8-kb transcript may be a bicistronic mRNA. Transcripts of smaller sizes (less than 0.8 kb) are expected to harbor no apparent ORFs longer than 30 amino acids. Genomic Southern blot analysis using the DNA fragments from this region as probes reveals that the probes all recognize a single band. Thus, it is not likely that the multiple bands in the northern blots arise from cross-reaction with transcripts from other genomic regions (Fig. 2C).

To accurately identify the size of these transcripts and the direction they are transcribed, we screened the cDNA library[21] (see legend for Fig. 1) by colony hybridization and isolated five different cDNA clones that may correspond to each transcript (Table 2). These cDNA clones are called *plx1* to *plx5* after *multiplex transcripts* (Fig. 2A). The *plx1* transcript encodes a protein (Plx1) that is homologous to the Myb-like transcriptional factor. A cDNA clone for *prl1* is not the partial cDNA clone of *plx1* because the band at 1.1 kb by northern blot (Fig. 2B) for *prl1* is detected only when probe e or f is used. It is also unlikely that the 1.1-kb transcript is a breakdown product because the time course of transcription is distinct from that of *plx1* transcript (2.6 kb) which is meiosis specific (Fig. 2B-b, c, and d). *plx2* is a small transcript with its end in the intron of *plx1*. It remains to be examined whether *plx2* is the breakdown product derived from the mRNA precursor of *plx1*. We failed to obtain the cDNA for the longest sense transcript (*plx6*) that corresponds to the band at 5.8 kb (Fig. 2B-c–h).

It is notable that *plx3*, *plx4*, and *plx5* cDNA clones are derived from the antisense RNAs of *plx1*. Namely, *plx3*, *plx4*, and *plx5* transcripts cannot be the breakdown products of *plx1*. The facts that northern blot confirms that these cDNAs are transcribed (Fig. 2B-c, d, and e) and that each cDNA clone has a poly(A) tail about 30 nucleotides downstream of the putative poly(A) signal (AAUAAN) indicate that these cDNAs are not the artifacts generated during the cDNA preparation.[21] We also detected similar kinds of antisense RNAs in the $rec7^+$ gene region, previously.[19] It remains to be examined whether these antisense transcripts are functional gene products or merely the junk mRNA-like transcripts.

The upstream region of *plx1* contains a TR-box sequence that is also contained in the target sequence of Ste11,[25] a transcriptional factor that regulates the entry of the cells into meiosis. As shown in Fig. 2B, the expression profiles of each transcript in northern blot differ. Expression of the 5.8-kb (*plx6*) and 2.6-kb (*plx1*) bands appear 4 h after nitrogen starvation in a meiosis-specific manner. Their levels peak at 6 h and then decrease. In contrast, *prl1* (1.1 kb), *plx3* (0.8 kb), *plx4* (0.4 kb),

Figure 1. The locations and the directions of the *prl* transcripts in the *S. pombe* genome. The sharp end of each horizontal arrow denotes the location of the poly(A) tail for each *prl* transcript. The numbers beside the arrows are the estimated sizes of the isolated cDNA inserts (nucleotides). Each bar indicates 500 base pairs. The T1, T2, $1^{asterisk}$, asterisk, P and filled triangle in the box indicated TF1-LTR (LTR retrotransposon of the Tf1/sushi group), TF2-LTR (LTR retrotransposon of the Tf2 group), TF1-107-like LTR, LTR-like, pseudogene and 5s rRNA, respectively. The filled rectangle near the t indicates tRNA. Asterisks, daggers and double daggers signify the *prl*s that harbor introns, that locate in the gene-free region and that situate near the LTR of the *S. pombe* genome, respectively. To make *S. pombe* cDNA libraries that are derived from mRNA transcribed in both mitotic and meiotic cells, CD16-1 (h^+/h^- *ade6M-210/ade6-M216 cyh1/++/lys5-391*) cells were directed to meiosis by nitrogen starvation and collected at one hour intervals (1, 2, 3, 4, 5, 6 hrs).[21] Note that not all cells proceed to meiosis under this condition, and a part of the cell population remains at the mitotic phase. Using poly(A) plus RNA purified from these cells, the cDNA library was constructed by a linker-primer method with the pAP3neo vector, as described previously.[30]

Table 1. Characterization of *prl* genes.

Name	Cosmid number	G enomic sequence	Class	Amino acids	Base	Inter CDS	Accession #
prl01	SPAC31G5.10/11	14046(14518)→15146	Non-coding	48	1100	2371	AB084813
prl02	SPCC1682.11c/12c	24926→25370	Non-coding	51	445	699	AB084814
prl03	SPAC806.03c/04c	7406→6800	Non-coding	72	607	5410	AB084815
prl04	SPBC20F10.06/07	12596→12344	Non-coding	43	253	1021	AB084816
prl05	SPAC13C5.06c A/S	9962→10125, 10196→10701	Antisense	29	668	1320*	AB084817
prl06	SPCC1795.13/12C	5143→4834	Non-coding	36	310	8851 †	AB084818
prl07/meu3	SPCC1884.01/02	9595→8852	Non-coding	36	727	5065 ‡	AB084819
prl08	SPBC17F3.02/01c	5638→5000	Non-coding	45	638	2486	AB084820
prl09	SPBC17F3.02/01c	5742→5012	Non-coding	45	731	2486	AB084821
prl10	SPCC1223.01/02	5037→4346	Non-coding	49	692	1570	AB084822
prl11	SPAC1348.10c/11, SPAC977.09c/10	28624→28796	Non-coding	23	172	4858	AB084823
prl12	SPAC9E9.17c/03	4375→4708	Non-coding	10	334	2941	AB084824
prl13	SPAC21E11.06/07	9944→9520	Non-coding	15	425	961 ‡	AB084825
prl14	SPAC1486.03c/04c	7001→7762	Antisense	69	762	961	AB084826
prl15	SPBC27B12.05 A/S	7546→6832	Antisense	22	715	611	AB084827
prl16	SPCC16A11.07/08	15361→15045	Non-coding	30	326	1731	AB084828
prl17/meu11	SPBC18H10.04c/05	10576→10163	Non-coding	38	410	3116	AB084829
prl18/meu16	SPAC15A10.10 A/S	23807→23178	Antisense	14	630	2957/827	AB084830
prl19	SPAC977 09c/10	21395→21820	Non-coding	34	426	3956	AB084831
prl20	SPBC1677.01c/02	2394→1942	Non-coding	38	410	2030 ‡	AB084832
prl21	SPAC1039.03/04	10523→10254	Non-coding	3	270	1146	AB084833
prl22	SPCC613.02/03	5296→4883	Non-coding	23	414	1009	AB084834
prl23	SPBC4C3.09/08	5926→6129	Non-coding	14	204	928	AB084835
prl24	SPBC11B10.07c/08	12246→11516	Non-coding	99	731	1309	AB084836
prl25	SPBC29A10.13 A/S	34312→34738	Antisense	102	427	709	AB084837
prl26	SPBC19C7.04c/05	14485→14834	Non-coding	19+19	350	4122 †	AB084838
prl27	SPAC18B11.11/10	4985→5131	Non-coding	17	147	841	AB084839
prl28	SPAC29B12.14c/SPAC1039.01	36000→35407	Non-coding	19	593	3756	AB084840
prl29/meu19	SPCC569.07/06	6657→7399	Non-coding	36	743	3602 ‡	AB084841
prl30/meu20	SPCC4F11.04c/-	8882→8278	Non-coding	36	750	1305	AB084842
prl31	SPAPB24D3.10c/11c	26459→26060	Non-coding	49	400	6822/5801	AB084843
prl32	SPCC417.12/13	34568→34139	Non-coding	43	430	5159/4489	AB084844
prl33	SPCPB16A4.05c/06c	11431→11786	Non-coding	6	356	2887 ‡	AB084845
prl34	SPBC36.01c/2c	4806→5178	Non-coding	15	373	2647	AB084846
prl35	SPCC965.12/13	31707→32116	Non-coding	58	410	4937/4160 †	AB084847
prl36	SPBC17D11.04c/5	9240→9668	Non-coding	9	429	1457	AB084848
prl37	SPAPJ695.-/01c	1905→1507	Non-coding	21	399	8639 ‡	AB084849
prl38	SPBC17G9.08c/09	20116→19682	Non-coding	64	436	3035	AB084850
prl39	SPBC32F12.10/11	19419→19176	Non-coding	50	242	3455 †	AB084851
prl40	SPCP1E11.01c/02	4408→4598	Non-coding	19	189	1200	AB084852
prl41	SPBC21B10.10/09	17800→18238	Non-coding	12+12	439	708	AB084853
prl42	SPBC337.01c/02	1662→2109	Non-coding	28	444	918	AB084854
prl43	SPAC6B12.11/12	26150→25569	Non-coding	66	579	1284/940	AB084855
prl44	SPCPB16A4.05c/06c	12868→13215	Non-coding	37	348	2887 ‡	AB084856
prl45	SPCC965.08/09	21539→20737	Non-coding	24	801	2005	AB084857
prl46	SPAC144.01c/02	1527→1086	Non-coding	18	442	2030	AB084858
prl47	SPAPB21F2.01A/S	1483→1098	Antisense	40	387	480	AB084859
prl48	SPAC2F3.09/10	20491→20360, 20258→20089, 19949→19713	Non-coding	85	552	1254*	AB084860
prl49	SPAC27E2.03c/11c	7812→7614	Non-coding	33	199	2544	AB084861
prl50	SPACC222.14c/15	34830→34337	Non-coding	19	495	975	AB084862
prl51	SPCC584.11c/12	14047→14328	Non-coding	18	282	1598	AB084863
prl52	SPAC22E12.03C/04	3719→3890	Non-coding	46	172	1183	AB084864
prl53	SPAC27E2.03C/11c	7602→8111	Non-coding	21	510	2544	AB084865
prl54	SPAC20G4.02C/03C	8631→9179	Non-coding	64	549	1117	AB084866
prl55	SPAC6B12.03c/04c	9463→9934	Non-coding	73	472	1644 ‡	AB084867
prl56	SPACUNK4.13c A/S	15893→15892, 15804→14771	Antisense	70	1055	317*	AB084868
prl57	SPAC6C3.03c/04	8572→8493, 8441→8201	Non-coding	25	321	2040*	AB084869
prl58	SPAC20G8.02/03	4265→3819	Non-coding	30	435	989	AB084870
prl59	SPAC17G8.11c/12	18820→19156	Non-coding	27	337	2054	AB084871
prl60	SPBP23A10.05/06	8007→7581	Non-coding	36	427	868	AB084872
prl61	SPAC186.07c A/S	21421→21885	Antisense	38	465	2270	AB084873
prl62	SPBC660.15/16	35312→35700	Non-coding	30	311	4128 †	AB084874
prl63	SPAC27E2.03C/11c	7853→7614	Non-coding	33	239	2544	AB084875
prl64	SPCC2H8.psudo/02	499→791	Non-coding	6+6	293	1520	AB084876
prl65	SPAC19A8.08/07c	16608→16572, 16460→16110	Non-coding	18	388	2773* ‡	AB084877
prl66	SPBC354.02c/03	3008→3444	Non-coding	48	437	1703	AB084878
prl67	SPCC622.16c A/S	30852→31474	Antisense	39	623	919	AB084879
prl68	SPBC14C8.01c/02	1735→1281	Non-coding	33	455	1332	AB084880

Inter CDS (protein coding sequence) signifies the length between the CDS that are registered in the database of Sanger Center.

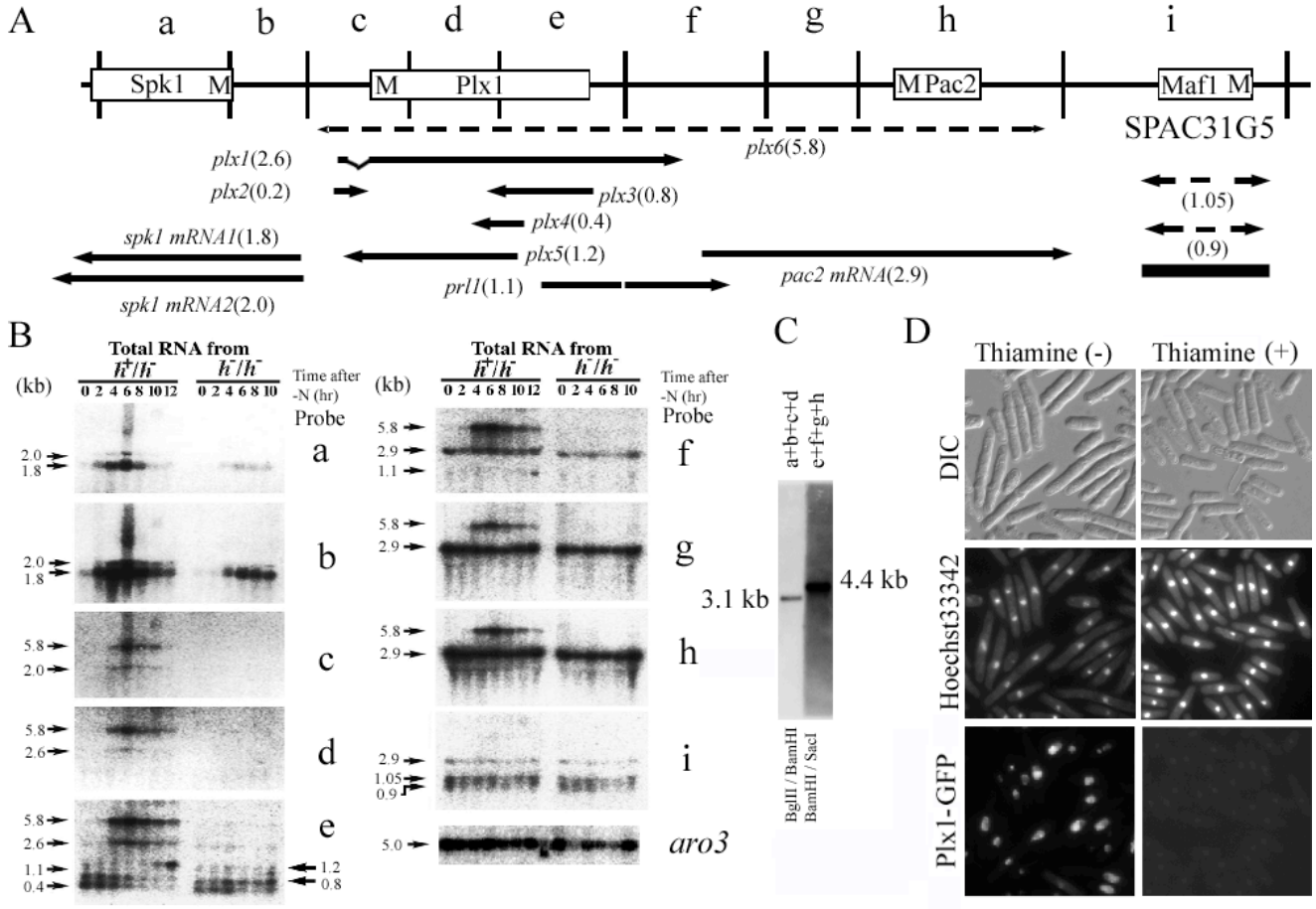


Figure 2. Isolation and characterization of multiplex RNA species in the genome of *S. pombe*. (A) A schematic presentation of the regions around the genomic fragment of the *plx1*$^+$ gene that transcribes *plx1* mRNA and encodes Plx1 protein, showing its multiplex transcription. The locations and directions of the isolated cDNA inserts are indicated by thick horizontal arrows. Boxes represent potential CDSs as assessed in the database (The Sanger Centre, UK). The initiation methionine site (M) in these CDSs is indicated. The bar indicates 1000 base pairs. Accession numbers for the *plxes* are registered as AB084881–AB084888 (See also Table 2). (B) Northern blot analyses using the DNA fragment around the *prl1* transcript. [α-^{32}P]dCTP probes were used as described before.[21] The locations of the nine different DNA fragments (a–i) used as probes are shown above. Thick horizontal arrows denote the orientation and size of the cDNA clones. The amount of loaded RNA was monitored by using the ^{32}P-labeled *aro3*$^+$ gene probe. (C) Southern blot analysis of whole genomic *S. pombe* DNA digested with relevant restriction enzymes. The probes consisted of two pools of probes (a + b + c + d and e + f + g + h). (D) Microscopic views of the enlarged cells generated by ectopically expressing *plx1* cDNA fused with *gfp* gene in mitotic cells. Shown are DIC (differential interference contrast) and fluorescence photographs of Hoechst33342-stained cells or *plx1*$^+$-*gfp*-bearing cells. Fluorescence from GFP-Plx1 fusion protein was observed predominantly in the nucleus only when the *nmt* promotor was induced by depletion of thiamine. The profiles of meiotic progression after nitrogen starvation using CD16-1 and CD16-5 (*h*$^-$/*h*$^-$ *ade6-M210/ade6-M216 cyh1/++/lys5-391*) strains were reported previously.[21]

and *plx5* (1.2 kb) are expressed not only during meiosis but also at 0 h when cells are in the mitotic phase. Thus, these latter transcripts are not specifically produced during meiosis alone.

It is interesting to note that, in the heterozygous strain CD16-1 that enters meiosis, the expression level of the antisense RNA species (*plx3* and *plx4*) decreases as the amount of the sense transcript (*plx1*) increases. However, the levels of *plx3* and *plx4* transcripts do not change in the homozygous strain CD16-5 that does not enter meiosis after nitrogen starvation. This suggests that the Plx1 protein may be harmful for mitotic cell growth. To test this possibility, we examined mitotic cells carrying *plx1* cDNA fused with green fluorescent protein (gfp) gene whose expression is driven by the *nmt* promoter. After 22 h of thiamine depletion to induce the activity of the *nmt* promoter, enlarged cells were observed under a microscope (Fig. 2D). This suggests that the expression of *plx1* mRNA in mitosis may indeed be harmful to the cell.

In the vicinity of *plx1*, other genes such as *spk1*$^+$ (SPAC5G10.9c), *pac2*$^+$ (SPAC5G10.11) and *maf1*$^+$ (SPAC5G10.12c) also display two, two and three bands in northern blots (Fig. 2B, probe a, h, and i), respectively. We have isolated two kinds of cDNA clones that are derived from the region of the Spk1 protein-coding

Table 2. List of transcripts in the vicinity of *prl1*$^+$ genes.

Name	class	site (SPAC31G5)	TR-box	TATA-box	polyA additional signal	b.p.	A. N.
prl1 $^+$	non-coding	14518-15146		13992TATAAT/14131TATAAT	15122AATAAt	629	AB084813
plx1+	mRNA	11986-12034 12269-14773	11927TTCTTTGTTT	11777 TATAAc	14748AATAAA	2554	AB084881
plx2 $^+$	non-coding	11991-12160		11777 TATAAc	12132AATAAt	170	AB084882
plx3 $^+$	antisense	13988-13206		14257TATAAa	13222AATAAc/13229AATAAt	783	AB084883
plx4 $^+$	antisense	13485-13064		14257TATAAa	13085AATAAg	422	AB084884
plx5 $^+$	antisense	13333-12055		14257TATAAa/15007TATAtt	12119AATtAA	1279	AB084885
spk1+	mRNA1	11715-9893		11780TATAAa	9919AATAAA	1823	AB084886
spk1+	mRNA2	11723-9776		11780TATAAa	9799AATgAA	1948	AB084887
pac2+	mRNA	14854-17761		14623gATAAT	17727AATAcA	2908	AB084888

region and that have different poly(A) sites. The variation in the size of the 3′ UTR may also be caused by post-transcriptional regulation.[26] Furthermore, we found that *prl49*, *prl53*, and *prl63* are derived from the same genomic region in SPAC27E2 (Fig. 1). Northern blots using this region as a probe also showed three bands that are not meiosis-specific (data not shown), representing another case of multiplex transcription.

Whole genome DNA sequencing of *S. pombe* revealed that the spaces between protein coding genes are longer than that of *S. cerevisiae*. About ten gene-free regions per chromosome are found, which are usually flanked by tandemly oriented genes. It has been pointed out that one of them corresponds to a prominent meiotic DNA break site or cluster of such sites.[20] The average length of the spaces between the genes (including ORFs, tRNA, or rRNA) where *prl* transcripts (*prl1–prl68*) are detected is 2483 bp. This value is larger than that of the average length of the spaces between the genes in the whole genome (about 900 bp for *S. pombe* and 800 bp for *S. cerevisiae*). The result indicates that *prl* transcripts are preferentially situated in longer inter-CDS regions to increase the variation of gene expression in *S. pombe*.

It should be borne in mind, of course, that such poly(A)-bearing RNAs without long ORFs may encode small peptides because the smallest protein-coding gene so far identified encodes an amino acid only 7 peptides long.[27] The annotated *S. pombe* genes includes the 147 genes that are confirmed or predicted to encode proteins of 25–99 amino acids, and the 116 genes that are treated as low coding potential because the gene products are too small to display any significant homology. Notably, our *prl* transcripts are not included in any of these categories.

Considering that we identified *prl* transcripts by cDNA clonings and they represent only a part of such clones in the cDNA library, it is evident that a large number of such non-coding poly(A)-bearing RNAs are transcribed in *S. pombe* cells and the corresponding cDNA clones remain to be discovered. Since we have isolated 68 kinds of *prl* transcripts from 987 randomly selected cDNA clones in the library (6.9%), we surmise that nearly 300 *prl* transcripts remain undiscovered. It should be pointed out that functional analysis of such non-coding RNAs has escaped classical genetics analysis so far, because genes without protein-coding regions tended to be ignored. In *S. pombe*, meiRNA is an exceptional case, which is shown to be essential for commitment of meiosis.[16] We therefore anticipate that many non-coding RNAs will be rediscovered among the ignored functional genes.

Considering that positional cloning in human genetics has identified non-coding RNA species as the cause of cartilage-hair hypoplasia[28] and autosomal-dominant congenital dyskeratosis,[29] it is probable that *prl* transcripts play a variety of roles in many aspects of cellular function. DNA chip array analysis targeting the *prl* transcripts would be helpful to analyze their functions, thus complementing the functional genomic analysis of *S. pombe*.

Acknowledgements: We thank Profs. C. Shimoda, M. Yamamoto, M. Yanagida, and Dr. Y. Watanabe for the *S. pombe* strains and plasmids, and Dr. Valerie Wood for the gift of cosmid clones. This work was supported by Grant-in-Aid for Scientific Research on Priority Areas (C) "Genome Biology" from the Ministry of Education, Culture, Sports, Science and Technology of Japan to HN. TW is a Research Fellow of the Japan Society for the promotion of Science.

References

1. Adelman, J. P., Bond, C. T., Douglass, J., and Herbert, E. 1987, Two mammalian genes transcribed from opposite strands of the same DNA locus, *Science*, **235**, 1514–1517.
2. Williams, T. and Fried, M. 1986, A mouse locus at which transcription from both DNA strands produces mRNAs complementary at their 3′ ends, *Nature*, **322**, 275–279.
3. Kozak, M. 2001, Extensively overlapping reading frames in a second mammalian gene, *EMBO Rep.*, **2**, 768–769.
4. Henikoff, S., Keene, M. A., Fechtel, K., and Fristrom, J. W. 1986, Gene within a gene: nested *Drosophila* genes encode unrelated proteins on opposite DNA strands, *Cell*, **44**, 33–42.
5. Casci, T. 2001, RNA splicing. Chance findings, *Nat. Rev. Genet.*, **2**, 162.
6. Labrador, M., Mongelard, F., Plata-Rengifo, P., Baxter, E. M., Corces, V. G., and Gerasimova, T. I. 2001, Protein encoding by both strands, *Nature*, **409**, 1000.
7. Nilsen, T. W. 2001, Evolutionary origin of SL-addition *trans*-splicing: still an enigma, *Trends Genet.*, **17**, 678–

680.
8. Mongelard, F., Labrador, M., Baxter, E. M., Gerasimova, T. I., and Corces, V. G. 2002, Trans-splicing as a novel mechanism to explain interallelic complementation in *Drosophila*, *Genetics*, **160**, 1481–1487.
9. Matsufuji, S., Matsufuji, T., Miyazaki, Y. et al. 1995, Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme, *Cell*, **80**, 51–60.
10. Eddy, S. R. 2001, Non-coding RNA genes and the modern RNA world, *Nat. Rev. Genet.*, **12**, 919–929.
11. Mattick, J. S. 2001, Non-coding RNAs: the architects of eukaryotic complexity, *EMBO Rep.*, **11**, 986–991.
12. Moss, E. G. 2000, Non-coding RNA's: lightning strikes twice, *Curr. Biol.*, **10**, 436–439.
13. Brown, C. J., Ballabio, A., Rupert, J. L. et al. 1991, A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome, *Nature*, **349**, 38–44.
14. Franke, A. and Baker, B. S. 1999, The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in *Drosophila*, *Mol. Cell*, **4**, 117–122.
15. Wolffe, A. P. 2000, Transcriptional control: imprinting insulation, *Curr. Biol.*, **10**, 463–465.
16. Watanabe, Y. and Yamamoto, M. 1994, *S. pombe mei2*$^+$ encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species meiRNA, *Cell*, **78**, 487–498.
17. Yamashita, A., Watanabe, Y., Nukina, N., and Yamamoto, M. 1998, RNA-assisted nuclear transport of the meiotic regulator Mei2p in fission yeast, *Cell*, **95**, 115–123.
18. Nakamura, T., Kishida, M., and Shimoda, C. 2000, The *Schizosaccharomyces pombe spo6*$^+$ gene encoding a nuclear protein with sequence similarity to budding yeast Dbf4 is required for meiotic second division and sporulation, *Genes Cells*, **6**, 463–479.
19. Molnar, M., Parisi, S., Kakihara, Y. et al. 2001, Characterization of REC7, an early meiotic recombination gene in *Schizosaccharomyces pombe*, *Genetics*, **157**, 519–532.
20. Wood, V., Gwilliam, R., Rajandream, M. A. et al. 2002, The genome sequence of *Schizosaccharomyces pombe*, *Nature*, **415**, 871–880.
21. Watanabe, T., Miyashita, K., Saito, T. T. et al. 2001, Comprehensive isolation of meiosis-specific genes identifies novel proteins and unusual non-coding transcripts in *Schizosaccharomyces pombe*, *Nucleic Acids Res.*, **29**, 2327–2337.
22. Tani, T. and Ohshima, Y. 1989, The gene for the U6 small nuclear RNA in fission yeast has an intron, *Nature*, **337**, 87–90.
23. Jaeger, J. A., Turner, D. H., and Zuker, M. 1989, Improved predictions of secondary structures for RNA, *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 7706–7710.
24. Kunitomo, H., Sugimoto, A., Wilkinson, C. R., and Yamamoto, M. 1995, *Schizosaccharomyces pombe pac2*$^+$ controls the onset of sexual development via a pathway independent of the cAMP cascade, *Curr. Genet.*, **28**, 32–38.
25. Sugimoto, A., Iino, Y., Maeda, T., Watanabe, Y., and Yamamoto, M. 1991, *Schizosaccharomyces pombe ste11*$^+$ encodes a transcription factor with an HMG motif that is a critical regulator of sexual development, *Genes Dev.*, **11**, 1990–1999.
26. Edwalds-Gilbert, G., Veraldi, K. L., and Milcarek, C. 1997, Alternative poly(A) site selection in complex transcription units: means to an end?, *Nucleic Acids Res.*, **25**, 2547–2561.
27. Gonzalez-Pastor, J. E., San Millan, J. L., and Moreno, F. 1994, The smallest known gene, *Nature*, **369**, 281.
28. Ridanpaa, M., van Eenennaam, H., Pelin, K. et al. 2001, Mutations in the RNA component of Rnase MRP cause a pleiotropic human disease, cartilage-hair hypoplasia, *Cell*, **104**, 195–203.
29. Vulliamy, T., Marrone, A., Goldman, F. et al. 2001, The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenital, *Nature*, **413**, 432–435.
30. Kobori, M., Ikeda, Y., Nara, H. et al. 1998, Large scale isolation of osteoclast-specific genes by an improved method involving the preparation of a subtracted cDNA library, *Genes Cells*, **3**, 459–475.