Kobe J. Med. Sci., Vol. 53, No. 4, pp. 163-169, 2007

Studies on Fungal Pumilio Gene Family through Mining Multiple Genome-Scale Data Sets

YUTAKA MATSUSHIMA[1], REIKO SUGIURA[2] and TAKAYOSHI KUNO[1]

[1]Department of Genome Sciences, Kobe University Graduate School of Medicine, Kobe 650-0017, Japan, [2]Laboratory of Molecular Pharmacogenomics, School of Pharmaceutical Sciences, Kinki University, Higashi-Osaka, 577-8502, Japan

Received 1 March 2007 /Accepted 5 March 2007

Key words: Pumilio, fungi, RNA-binding protein, multi-genome analysis, public-access database, bioinformatics

The genes belonging to Pumilio gene family of the fission yeast *Schizosaccharomyces pombe* (*S. pombe*) were compared with genome information of the several fungi and their functions have been inferred using public-access databases disclosed on the internet websites.

The Pumilio-family genes are conserved from yeast to man and have been considered to suppress the expression of other genes by binding their specific target messenger RNAs. In *S. pombe* genome, nine genes belonging to Pumilio gene family were found and were clustered into four groups by homology of their amino acid sequences. Each gene has been analyzed to search 'close orthologue' in the genomes of *Saccharomyces cerevisiae, Candida parapsilosis, Aspergillus fumigatus, Phytophthora infestans, Pneumocystis carinii* and *Neurospora crassa,* using protein-protein blast P search.

Consequently, we found several genes belonging to Pumilio gene family in the genomes of *Saccharomyces cerevisiae, Candida parapsilosis,* and *Neurospora crassa.* Evolution and function of these genes can be expected from the analysis based on the molecular phylogenies.

Pumilio is an RNA-binding protein which associates with mRNA of the gene known as NANOS in *Drosophila* (1). It associates with the specific sequence of the nucleotide in 3'-UTR region of the mRNA. Pumilio-family proteins are conserved from yeast to man and have been postulated to bind specific sequences in mRNAs to regulate their translation or stability, or both. And such regulatory sequences have been identified in a wide variety of transcripts in numerous organisms, including mammals, *Drosophila melanogaster*, *Caenorhabditis elegans* and *Saccharomyces cerevesae* (*S. cerevisiae*) (for reviews see (2), (3)). Several types of RNA-binding proteins have been shown to regulate their target genes by biding to the 3'-UTR control elements in a sequence-specific manner (for review see (4)), although the mechanisms by which these RNA-binding proteins lead to functional changes in the behavior of the mRNAs remain largely unknown.

The Pumilio gene family of the fission yeast *Schizosaccharomyces pombe* (*S. pombe*) may also take part in control of translation of a specific gene by binding the transcript relating to the gene. The nine Pumilio-family genes exist in the genome of the fission yeast.

Phone: 81-78-382-5441 Fax: 81-78-382-5459 E-mail: tkuno@med.kobe-u.ac.jp

These were divided into several groups by homology clustering based on their amino acid sequences.

In the present study, we compared all the genes belonging to Pumilio gene family with the whole genome information of six fungi, and were successful to get useful information on their evolution and function.

MATERIALS AND METHODS

Definition of the Pumilio-family genes in *S. pombe* genome. We extracted the Pumilio-family genes of *S. pombe* from the gene database of the Sanger Institute (http://www.sanger.ac.uk) with keyword search for annotations. We defined a gene that was annotated as 'Pumilio' in this database was the gene belonging to Pumilio gene family.

Clustering of the Pumilio gene family of *S. pombe*. We clustered the Pumilio-family genes in *S. pombe* genome into several groups by their homology and drew the phylogram of the Pumilio-family genes with ClustalW (5) in the websites of GeneBee (http://www.genebee.msu.su/genebee.html).

Search for orthologues of the Pumilio gene family of *S. pombe* in the fungal genomes. SGD™ *Saccharomyces* Genome Database was used for *S. cerevisiae* (budding yeast). The gene database of the Sanger Institute was used for *Candida parapsilosis* (*C. parapsilosis*, pathogenic yeast), *Aspergillus fumigatus* (*A. fumigatus*, filamentous fungus), *Phytophthora infestans* (*P. infestans*, plant pathogenic fungus), and *Pneumocystis carinii* (*P. carinii*, pathogenic filamentous fungus). *Neurospora crassa* Database for *Neurospora crassa* (*N. crassa*, filamentous fungus). We summarize the databases used in this study in Table 1.

Protein-protein blast P search (6) was done over the gene databases for the above-mentioned organisms. The threshold of P(N) value for 'orthologue' was set at less than 1.0^{-10}. A gene in the fungal genomes, of which P(N) value of protein-protein blast P search score with the Pumilio-famiily genes in *S. pombe* genome was less than the threshold, was defined as a 'orthologue'.

Table 1. Databases used in this study

Organism	Database Name	URL
S. pombe	The Sanger Institute	http://www.sanger.ac.uk
S. cerevisiae	SGD™ *Saccharomyces* Genome Database	http://www.yeastgenome.org
C. parapsilosis	The Sanger Institute	http://www.sanger.ac.uk
A. fumigatus	The Sanger Institute	http://www.sanger.ac.uk
P. infestans	The Sanger Institute	http://www.sanger.ac.uk
P. carinii	The Sanger Institute	http://www.sanger.ac.uk
N. crassa	*Neurospora crassa* Database	http://www.broad.mit.edu/annotation/genome/neurospora/Home.html

RESULTS AND DISCUSSION

Pumilio-family genes in *S. pombe* genome. Nine genes were extracted from the gene database of the Sanger Institute with keyword search of 'Pumilio' for annotations.

Clustering of the Pumilio gene family of *S. pombe*. We classified the nine genes of *S. pombe* into several groups by their homology using ClustalW (http://www.genebee.msu.su/genebee.html). The result is shown in the Figure 1. It was categorized and classified into

the four groups. The first group includes the five genes (SPAC4G8.03c, SPAC6G9.14, SPAC1687.22c, SPAC4G9.05 and SPCC1682.08c). The second group includes the two genes (SPBC56F2.08c and SPBP35G2.14). The only one gene (SPCP1E11.11) belongs to the third group and the last gene (SPAC6G9.02c) belongs to the forth, respectively.

In the first group (named as Group 1), SPAC4G8.03c and SPAC6G9.14 were considered to be functionally similar to each other from their remarkably low P(N) value (1.5^{-70}, Table 2). Likewise, SPAC4G9.05 and SPCC1682.08c were considered to be very similar to each other from their low P(N) value (2.8^{-79}). Therefore, two pairs of similar genes existed in Group I (Figure 1).

In the second group (Group II), P(N) value between SPBC56F2.08c and SPBP35G2.14 was remarkably low (2.4^{-89}), then they were considered to form a pair.

Table 2. Probability of blast P search (P(N)) of the orthologues with homology to the Pumilio-family genes in *S. pombe*

Group	**I**	**I**	**I**	**I**	**I**
Gene	SPAC4G8.03c	SPAC6G9.14	SPAC1687.22c	SPAC4G9.05	SPCC1682.08c
SPAC4G8.03c	0.0	1.5^{-70}	1.2^{-36}	2.4^{-17}	4.7^{-16}
SPAC6G9.14	1.5^{-70}	0.0	7.9^{-52}	8.0^{-20}	8.9^{-19}
SPAC1687.22c	1.2^{-36}	7.9^{-52}	0.0	2.0^{-29}	4.0^{-36}
SPAC4G9.05	2.4^{-17}	8.0^{-20}	2.0^{-29}	0.0	2.8^{-79}
SPCC1682.08c	4.7^{-16}	8.9^{-19}	4.0^{-36}	2.8^{-79}	0.0
SPBC56F2.08c	1.5^{-11}	2.6^{-16}	1.8^{-11}		
SPBP35G2.14	4.5^{-11}	4.3^{-21}	6.9^{-14}		
SPCP1E11.11					
SPAC6G9.02c					

Group	**II**	**II**	**III**	**VI**	**Annotation***
Gene	SPBC56F2.08c	SPBP35G2.14	SPCP1E11.11	SPAC6G9.02c	
SPAC4G8.03c	1.5^{-11}	4.5^{-11}			RNA-binding protein
SPAC6G9.14	2.6^{-16}	1.1^{-23}			RNA-binding protein
SPAC1687.22c	1.8^{-11}	6.9^{-14}			RNA-binding protein Puf3
SPAC4G9.05					meiotic PUF family protein 1
SPCC1682.08c					RNA-binding protein *Mcp2*
SPBC56F2.08c	0.0	2.4^{-89}			RNA-binding protein
SPBP35G2.14	2.4^{-89}	0.0			RNA-binding protein
SPCP1E11.11			0.0		RNA-binding protein
SPAC6G9.02c				0.0	RNA-binding protein *Nop9*

*: Annotation was from quotations of the database of the Sanger Institute.

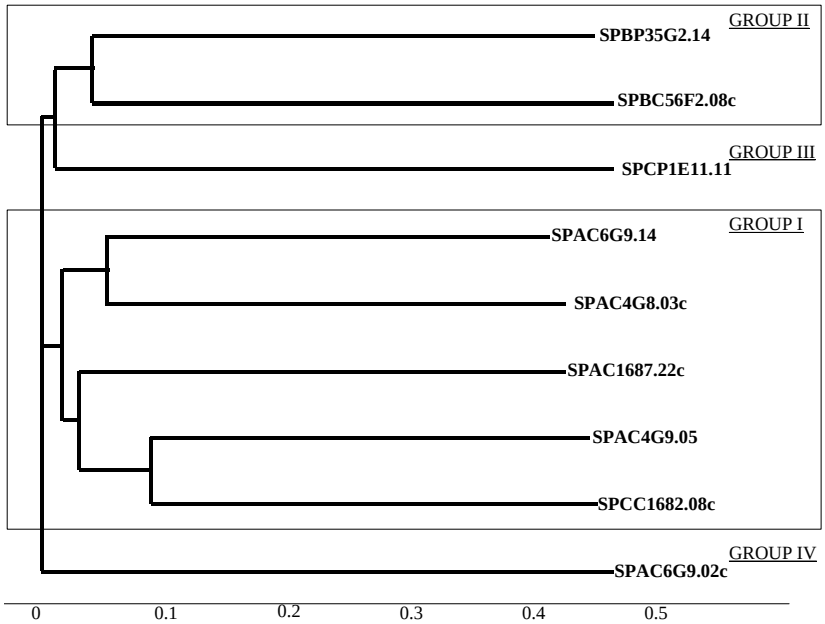

Figure 1. Phylogram of Pumilio gene family of the fission yeast *Schizosaccharomyces pombe*. The genes were classified into four groups by their homology. The graduation line below the chart shows the calculated distance between all pairs of the sequences from the multiple alignments.

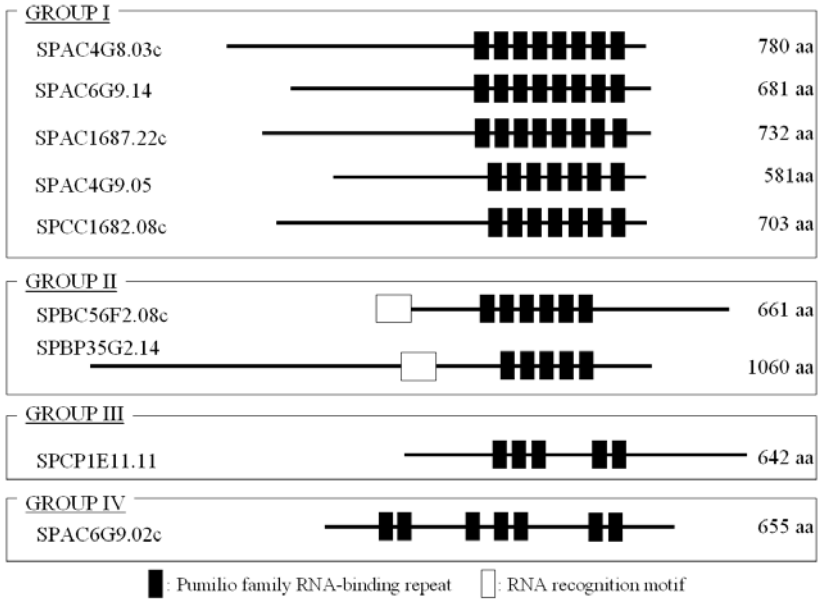

FIG. 2. Pumilio family RNA-binding repeats in Pumilio gene family of *Schizosaccharomyces pombe*. Location data of Pumilio family RNA-binding repeats and RNA recognition motifs were obtained from the gene database of the Sanger Institute. Location of both of 'Pumilio family RNA-binding repeat (PUF)' assigned by Pfam-A (9) and 'Pumilio-like repeats' assigned by SMART (10) was aligned as 'Pumilio family RNA-binding repeat'. Location of 'RNA recognition motif ('RRM_1) assigned by Pfam-A was aligned as 'RNA recognition motif'.

Location of two types of the putative RNA-binding motifs ('Pumilio-family RNA-binding repeat' (PUF) and 'RNA recognition motif' (RRM)) is illustrated in Figure 2. In Group I, the Pumilio-family genes include tandem eight or seven PUF repeats in their sequence. In Group II, the two genes include tandem six or five PUF repeats and one RRM. The genes in Group III and in Group IV had five and seven PUF repeats with intervals, respectively. Based on the number of PUF repeats, Group I could be divided into two subgroups; (SPAC4G8.03c, SPAC6G9.14 and SPAC1687.22c) and (SPAC4G9.05 and SPCC1682.08c).

Search for close orthologues of the *S. pombe* Pumilio-family genes in a wide variety of the fugal genomes. We extracted orthologue of the Pumilio-family genes of *S. pombe* from the six fungal genomes (*S. cerevisiae*, *C. parapsilosis*, *A. fumigatus*, *P. infestans*, *P. carinii* and *N. crassa*) from the databases by protein-protein blast P search, and their number below the threshold of P(N) value was counted and listed. The results are shown in Table 3.

Table 3. The number of the orthologue with homology to the Pumilio-family genes in *S. pombe*

Group	**I**	**I**	**I**	**I**	**I**	**II**	**II**	**III**	**VI**
Gene	SPAC 4G8.03c	SPAC 6G9.14	SPAC 1687.22c	SPAC 4G9.05	SPCC 1682.08c	SPBC 56F2.08c	SPBP 35G2.14	SPCP 1E11.11	SPAC 6G9.02c
S. cerevisiae	3	4	3	2	3	3	3	1	1
C. parapsilosis	3	3	3	1	1	1	2	1	1
A. fumigatus	0	0	0	0	0	0	0	0	0
P. infestans	0	0	0	0	0	0	0	0	0
P. carinii	0	0	0	0	0	0	0	0	0
N. crassa	3	4	4	3	3	2	2	1	1

To our surprise, homology search in only the three fungal organisms (*S. cerevisiae*, *C. parapsilosis* and *N. crassa*) was able to detect *S. pombe* Pumilio orthologues that show low P(N) values (less than 1.0^{-10}) of blast P search as shown in Table 3 and Table 4. The absence of Pumilio orthologues in *A. fumigatus*, *P. infestans*, and *P. carinii* may be due to their incomplete genome sequencing projects.

As described above, in the second group (Group II), P(N) value between SPBC56F2.08c and SPBP35G2.14 was remarkably low (2.4^{-89}). Consistently, only one 'close' orthologue is found in the genome of *C. parapsilosis*, cpara996f03.qlk, as well as in the genome of *N. crassa*, NCU06199.3 as shown in Table 4. These results suggest that SPBC56F2.08c and SPBP35G2.14 have redundant function.

As described below, SPCP1E11.11 in Group III has one orthologue in the genome of *S. cerevisiae*, *C. parapsilosis* and *N. crassa*, and SPAC6G9.02c in Group IV also has one orthologue in these three fungal genomes as shown in Table 4. Therefore, variation in the number of Pumilio gene family seems to be caused by redundancy in Group I and Group II.

Each of the three *S. pombe* genes (SPAC4G8.03c, SPAC6G9.14, SPAC1687.22c) of Group I has more than three orthologues in the genomes of *S. cerevisiae*, *C. parapsilosis* and *N. crassa*. Notably, SPAC4G8.03c has one close orthologue in *S. cerevisiae* with very low P(N) value (2.2^{-52}), while SPAC6G9.14 has two close orthologues in *S. cerevisiae* (1.7^{-69} and 1.1^{-48}).

The five genes in Group I would be divided into the same two subgroups as these depended on the number of PUF repeats. It might be shown that the ancestors of Pumilio-family genes in the two subgroups would diverse and become redundant before *S. pombe* had been divided from the other species.

Table 4. Probability of blast P search (P(N)) of the orthologues in the other fungi with homology to the Pumilio-family genes in *S. pombe*

Group		I	I	I	I	I
Organism	**Gene**	SPAC4G8.03c	SPAC6G9.14	SPAC1687.22c	SPAC4G9.05	SPCC1682.08c
S. cerevisiae	YGL014W	2.2^{-52}	1.7^{-69}	4.6^{-38}	9.4^{-15}	1.1^{-16}
	YGL178W	1.6^{-30}	1.1^{-48}	1.9^{-32}		8.6^{-14}
	YLL013C	1.0^{-35}	1.3^{-37}	4.1^{-72}	2.8^{-20}	1.2^{-28}
	YJR091C		8.2^{-13}			
	YPR042C					
	YDR496C					
	YJL010C					
C. parapsilosis	cpara817g10.q1k	1.9^{-30}	1.7^{-50}	3.5^{-22}		
	cpara755c10.q1k	3.0^{-30}	5.9^{-55}	4.9^{-24}		
	cpara600c09.q1k	2.2^{-18}	1.2^{-30}	1.9^{-61}	5.1^{-15}	1.3^{-18}
	cpara996f03.q1k					
	cpara1143e11.q1k					
	cpara842b07.p1k					
N. crassa	NCU01775.3	$<1.0^{-50}$	$<1.0^{-50}$	$<1.0^{-40}$	3.7^{-24}	2.4^{-20}
	NCU06511.3	5.0^{-27}	1.3^{-30}	$<1.0^{-50}$	2.7^{-22}	2.2^{-26}
	NCU01760.3	2.3^{-11}	2.8^{-18}	2.5^{-36}	$<1.0^{-50}$	$<1.0^{-50}$
	NCU06199.3		2.9^{-23}	1.5^{-12}		
	NCU09380.3					
	NCU06118.3					

Group		II	II	III	VI	Annotation*
Organism	**Gene**	SPBC56F2.08c	SPBP35G2.14	SPCP1E11.11	SPAC6G9.02c	
S. cerevisiae	YGL014W	3.8^{-15}	2.6^{-14}			*PUF4*
	YGL178W					*PUF5, MPT5*
	YLL013C					*PUF3*
	YJR091C	5.7^{-70}	3.2^{-75}			*PUF1, JSN1*
	YPR042C	1.2^{-62}	6.8^{-69}			*PUF2*
	YDR496C			3.2^{-91}		*PUF6*
	YJL010C				1.1^{-32}	*NOP9*
C. parapsilosis	cpara817g10.q1k					
	cpara755c10.q1k		5.6^{-17}			
	cpara600c09.q1k					
	cpara996f03.q1k	4.1^{-33}	2.4^{-26}			
	cpara1143e11.q1k			6.4^{-78}		
	cpara842b07.p1k				1.1^{-14}	
N. crassa	NCU01775.3	1.8^{-17}	6.8^{-22}			
	NCU06511.3					
	NCU01760.3					
	NCU06199.3	$<1.0^{-50}$	$<1.0^{-50}$			
	NCU09380.3			$<1.0^{-50}$		
	NCU06118.3				4.5^{-42}	

*: Annotation was from quotations of the database of the SGD™.

In contrast to other genes in Group I, SPAC4G9.05 and SPCC1682.08c have no close orthologue in the genomes of *S. cerevisiae* and *C. parapsilosis*, while NCU01760.3 in *N. crassa* genome seems to be the close orthologue. These results suggest that these two genes in Group I have specific function in these fungal species (*S. pombe* and *N. crassa*).

The genes of Group III and Group VI are well conserved all over the fungal genomes and their P(N) values were very low. These suggest that they may have important roles. The *S. cerevisiae*'s close orthologue of the gene (SPCP1E11.11) of Group III was *PUF6* (YDR496C), of which protein represses *ASH1* (a protein dominant for mating-type switching) mRNA translation (8). Deletion mutant of *NOP9*, the *S. cerevisiae*'s close orthologue of the gene (SPAC6G9.02c) of Group VI, is inviable (E. Thomson and D. Tollervey, 2005 Personal communication to SGD™), suggesting that SPAC6G9.02c is also essential for *S. pombe*.

REFERENCES

1. **D. D. Barker, C. Wang, J. Moore, L. K. Dickinson, and R. Lehmann.** 1992. Pumilio is essential for function but not for distribution of the Drosophila abdominal determinant Nanos. Genes Dev. **6**: 611-625.
2. **C. J. Deker, and R. Parker.** 1995. Diversity of cytoplasmic functions for the untranslated region of eukaryotic transcript. Curr. Opin. Cell Biol. **7**: 386-392.
3. **N. K. Gray, and M. Wickens.** 1998. Control of translation initiation in animals. Annu. Rev. Cell Dev. Biol. **14**: 399-458.
4. **M. Derrego, A. Cesareni, G. Savettieri, and I. Di Loegro.** 2000. RNA-protein interactions in the control of stability and localization of messenger RNA. Int. J. of Mol. Med. **5**: 111-123.
5. **J. D Thompson, DG Higgins, T. J Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. Nov 11: **22(22)** 4673-4680.
6. **S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. Oct 5; **215(3)**: 403-410.
7. **W. Olivas, and R. Parker.** 2000. The Puf3 protein is a transcript-specific regulator of mRNA degradation in yeast. The ENBO J. Vol. 19 No. **23**: 6602-6611.
8. **W. Gu, Y. Deng, D. Zenklusen and R. H. Singer.** 2004. A new yeast PUF family protein, Puf6p, represses *ASH1* mRNA translation and is required for its localization. GENES & DEVELOPMENT **18**: 1452-1465.
9. **R. D. Finn, J. Mistry, B. Schuster-Böckler, S. Griffiths-Jones, V. Hollich, T. Lassmann , S. Moxon, M. Marshall, A. Khanna, R. Durbin, S. R. Eddy, E. L. L. Sonnhammer, and A. Bateman.** 2006. Pfam: clans, web tools and services. Nucleic Acids Res. (Database Issue) **34**: D247-D251.
10. **I. Letunic, R. R. Copley, B. Pils, S. Pinkert, J. Schultz and P. Bork.** 2006. SMART 5: domains in the context of genomes and networks. Nucleic Acids Res. (Database Issue) **34**: D257-D260.