Evolution of RNA-PUF protein networks

By

Daniel James Wilinski Jr.

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Cellular and Molecular Biology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2015

Date of final oral examination: May 6, 2015

The dissertation was reviewed by the following members of the Final Oral Committee:
Marvin Wickens, Max Perutz Professor of Biochemistry
Judith Kimble, Vilas Professor of Molecular Biology and Biochemistry
Audrey Gasch, Associate Professor of Genetics
Catherine Fox, Professor of Biomolecular Chemistry
Michael Sheets, Professor of Biomolecular Chemistry

ABSTRACT

Proteins bind and control mRNAs, directing their localization, translation, and stability. Members of the PUF family of RNA-binding proteins control multiple mRNAs in a single cell, and play key roles in development, stem cell maintenance and memory formation. Here we demonstrate the conservation of PUF regulatory networks by bioinformatic analysis. We identified the mRNA targets of two *Saccharomyces cerevisiae* PUF proteins, Puf3p and Puf5p, and two *Neurospora crassa* PUF proteins, PUF3 and PUF5, by UV-crosslinking-affinity purification and high-throughput sequencing (HITS-CLIP). Taken together these data allowed us to elucidate a rewiring of a nulcear-endcoded circuit of mitochondrial related mRNAs.

S. cerevisiae Puf5p associates with ~1,000 RNAs, suggesting that it is a broad spectrum regulator. The binding sites recognized by both Puf5 proteins are diverse, with variable spacer lengths between two specific sequences. *S. cerevisiae* Puf3p binds a group of RNAs significantly enriched for nuclear encoded RNAs that code for mitochondrial proteins. Many of these orthologous RNAs are bound by *N. crassa* PUF5. This constitutes a switch, from Puf3p to Puf5, regulation, when comparing *S. cerevisiae* and *N. crassa* PUF regulatory networks. Crystal structures of *S. cerevisiae* Puf5p-RNA complexes reveal that the protein scaffold presents an exceptionally flat and extended interaction surface relative to other PUF proteins. A single PUF protein repeat is sufficient to induce broadening of specificity. Changes in protein architecture, such as alterations in curvature, may have led to evolution of mRNA regulatory networks which we observe.

ACKNOWLEDGEMENTS

It would be impossible to complete a doctoral degree without support, emotional and intellectual, from many people. There is no such thing as a self-made person. I would like to acknowledge some of the more prominent members of my life in Madison.

I will start with my advisor, Marvin P. Wickens. Marv is a renaissance man who always contributed to my cultural as well as my scientific development and reminds me that my favorite people are not myopic in their pursuits. His gentle redirection and careful explanation always put me at ease even if we disagreed. Always in pursuit of the best in people and in the data—he allowed me to pursue my project long after we should have dropped it. I admire Marv a great deal and it was a great pleasure being in the Wickens lab.

I would like to thank my thesis committee. Audrey Gasch offered the only genomic perspective on my committee and provided, along with Catherine Fox, a second home when Marv was on sabbatical. Audrey pushed me to explain, understand, and defend my work. Catherine, a scientific child of Marv, was always excited about science and a joy to talk with; in constant pursuit of the mechanism. Mike Sheets for joining my committee in the last year. He had many thoughtful comments and questions. Judith Kimble was a true advocate of mine. She always had thoughtful focused contributions to make during meetings.

My parents Marian and Dan. Mom was always supportive though her own ups and downs. She believes in me always. Dad instilled a strong work ethic in me and always focused on the practical. John is one of the most Christ-like people I have ever had the privilege of knowing. He is the only one in my family who really tried to understand what I do (excluding Bethany). John and Judy were there to celebrate that day at Madison's—on the napkin. They are the parents that I got to choose, which I have never regretted.

My friends Danielle Smith and Cameron Smith the salt of the earth. Dear friends. We shared some of the most memorable times in Madison together at the terrace, traveling to Tanzania and Egypt and South Dakota, children, Risk, and in the endless pursuit of truth.

Elena Sorokin my first baymate. Bike rides, dreams, and the sign "The bay where dreams do not die". The Wickens lab: Amy (inspired me to join the lab), JJ (never afraid of the truth), Yvonne (the best "hands" ever), Craig (The Terrace), Alana (late night conservations), Cary (soccer and atomic details), Zak (the good idea), Shruti (patient troubleshooting), Melanie (controls and the Bishop), Doug (kindness), Markus (bad project-hard worker), and Caitlin (difficult lessons). Natascha Buter for all the kindness and keeping the lab running despite all the students' attempts to the contrary.

Manuel Llinas launched my scientific career by teaching me how to do science and more importantly love science. We met in Uganda and spent three days on safari together. Then he took a chance on me by inviting me to work in his new lab. For that, I will be forever grateful.

Chris Lapointe helped with my graduate work immeasurably. Chris is always willing to talk about science or football. A thoughtful person who forced me to drink coffee over which we had many fun conversations about science and life. We shared the fun and worked through the difficult together. A true friend.

Finally, Bethany Wilinski the most important person in my life. Our journey to graduate school started many years before in Tanzania; that day at Kibeta in her mzungu shirt. Madison is where we made a home together, we were married, and had our first child, Daniel "Danny" III. Bethany took the time to understand "science" and the good and the bad of graduate school. She, herself a graduate student, knew the struggles associated with publishing, advisors, class work, and research. She edited all of my papers and gently pushed me to be better both through

encouragement and by example. My wife, my confidant, my support, my love. Thank you and thank you all.

TABLE OF CONTENTS

Chapter 1

An Introduction to RNA-protein interactions in Ascomycota fungi

RNA is pervasive

Every aspect of an RNA's life is regulated. The role of RNA was historically thought of as simply transmitting information from DNA into protein. Our modern understanding is that this molecule is far more complex both in terms of its functional diversity and the numerous regulatory steps to which all RNAs are subject (Figure 1). RNA plays key roles in an array of functions including: gene expression, catalysis, structure, and regulation.

RNA types

Messenger RNAs (mRNAs) encode information required for protein biosynthesis. mRNAs generally consist of three regions: a protein coding open reading frame (ORF), 5' untranslated region (UTR), and a 3' UTR. Three-letter condons define the identity and order of amino acids which form the polypeptide chain and the UTRs contain regulatory signals. Synthesis of mRNAs from the DNA template occurs in the nucleus in Eukaryotes where the mRNAs are spliced, 5' 7-methylguanosine capped, and polyadenylated. Some of these steps occur co-transcriptionally while other processing events occur upon transcriptional termination, but all steps are regulated (Figure 1). Upon maturation, mRNAs are exported out of the nucleus, often localized to subcellular compartments, and then translated. Again, each of these steps is regulated (Figure 1).

Non-codng RNAs are a more recently identified group of RNAs that do not encode proteins. The synthesis and processing of these RNAs is highly regulated, an in-depth discussion of which is beyond the scope of this work. Non-coding RNAs include transfer, ribosomal, small nuclear, small nucleolar, ribozyme, and long noncoding RNAs. In addition, a subclass of "small RNAs" include: micro, pi, piwi, and small interfering RNAs. This diverse group of RNAs play a

critical role in cellular functions. For example, ribosomal RNAs provide a structural scaffold for the ribosome, ribozymes function as catalysts, and micro RNAs contribute to gene silencing.

RNA regulators

RNA-binding proteins are defined by the structural motifs they use to recognize and bind RNA. I will discuss some of the most prominent classes below and schematics of example proteins are shown in Figure 2.

RNA recognition motif (RRM). RRMs are the most abundant RNA binding protein motifs found in all domains of life (Cléry et al., 2008; Daubner et al., 2013; Lunde et al., 2007). About 2% of human proteins contain RRM domains (Maris et al., 2005). RRMs domains are characterized by a conserved fold consisting of four antiparallel beta-strands with two helices packed against the sheet (Lunde et al., 2007; Maris et al., 2005). This structure forms a platform where aromatic residues typically recognize RNA by stacking interactions or insertion intercalation between two sugar rings (Maris et al., 2005). Variations to this basic structure also exist, contributing to the vast number of RNA sequences and shapes recognized by these proteins (Daubner et al., 2013).

hnRNP K-homology (KH) domain. KH domains are characterized by a core, βααβ strands, forming a β-sheet flanked by either C-terminal, type I (Eukaryotes), or N-terminal, type II (Prokaryotes), α and β strands (Lunde et al., 2007; Trabucchi et al., 2009; Valverde et al., 2008). A GxxG loop links the two core α-helices and interacts with the nucleic acid backbone of either RNA or ssDNA (Duncan et al., 1994). This loop directs the nucleic acids towards a hydrophobic groove where base-specific contacts are made. Most KH domains bind short (4 bases) A/C rich sequences, but they have also been shown to bind G-rich sequences (Trabucchi

et al., 2009). KH domains are often repeated, up to 15, and are connected by flexible linkers that allow proteins to recognize longer sequences (Nicastro et al., 2015).

Zinc fingers. Classically thought of as DNA binding transcription factors, two classes--C_2H_2 and CCCH--can also bind RNA (Carballo et al., 1998; Hall, 2005; Lunde et al., 2007; Picard and Wegnez, 1979). The modular zinc fingers are typically about 30 amino acids and are named for the arrangement of residues that coordinate zinc ions. The polypeptide chains fold into a ββα structure that recognizes DNA, RNA, or proteins (Hall, 2005). Multiple zinc fingers are often found within a single protein.

PUF proteins. Pumilio and FBF (PUF) proteins will be discussed in depth below as they are the focus of this work.

Structure of PUF proteins

Pumilio homology domain. PUF stands for Pumilo and FBF the two founding members this family of RNA binding proteins (Zamore et al., 1997; Zhang et al., 1997). The distinguishing feature of PUF proteins is the Pumilio homology domain. The Pumilio homology domain consists of eight ~36 amino acid repeats arranged into 3 alpha helixes flanked by half repeasts. The alpha helixes are stacked next to each other into crescent shape (Miller and Olivas, 2011) (Figure 3A). The remaining residues typically do not have recognizable domains, though they often contain glutamine-rich regions which may promote protein aggregation (Salazar et al., 2010). The concave surface of the Pumilio homology domain makes base-specific contacts with RNA and the opposite convex side can make protein-protein contacts (Campbell et al., 2012; Miller and Olivas, 2011) (Figure 3A).

Recognition of RNA. The inner concave surface of PUF proteins has a net positive charge, but does not make contact with negatively charged phosphates of the RNA (Wang et al., 2001). Instead, three amino acid side chains at conserved positions (tripartite recognition motifs, or TRMs) make contact with a single RNA base through edge-on and stacking interactions (Wang et al., 2001; Wickens et al., 2002). These residues are conserved in PUF proteins and can be altered to recognize predicted RNA bases (Wang et al., 2013). This one-base one-repeat seems to be a general rule for all PUF proteins examined structurally (Miller and Olivas, 2011) (Figure 3B). RNA is bound in the low nanomolar range and is preferred to DNA. For example, Homo sapiens Pumilio binds RNA > 2500-fold more tightly than DNA (Wang et al., 2002).

PUF protein RNA recognition is based in the composition of tripartite recognition motifs, but often this does not completely explain the diversity of binding elements (Miller and Olivas, 2011). In particular, tripartite recognition motifs do not explain how *C. elegans* PUF-11 is able to bind two length binding elements. In addition to TRMs, *S. cerevisiae* Puf4p and *C. elegans* FBF are flattened and twisted (Miller et al., 2008; Wang et al., 2009).

Structural elements allow PUF proteins to “base flip” and flex. PUF proteins typically contain 8 repeats and therefore bind sequences 8nt in length. These binding elements can sometimes be longer for some PUF proteins. Longer binding elements are accommodated by “base flipping” with only one-repeat contacting one-base (Figure 3C). Two examples of this are *Saccharomyces cerevisiae* Puf4p and *Caenorhabditis elegans* FBF (Miller et al., 2008; Wang et al., 2009) (Figure 3C). Base flipping has been explained by the “two handed” model of binding. In this model one “hand” (repeats 8, 7, 6) binds the 5’ UGU and the second “hand” (repeats 1, 2, and 3) binds the 3’ AUA sequence (Valley et al., 2012). This allows one of the intervening bases to “flip” away from the protein, but each repeat only makes one contact .

C. elegans PUF-11 is an example of a PUF protein with 8 repeats that can bind to 9 and 10nt binding elements with similar affinity (Koh et al., 2009). It was postulated that *C. elegans* PUF-11 flexed in middle of the Pumilio homology domain combined with base flipping allowed the protein to accommodate either 9 or 10nt RNAs and maintain a one-base one-repeat recognition (Koh et al., 2009).

Function of PUF proteins

PUF proteins typically repress mRNA translation by stimulation of decay or inhibition of translation factors. PUF proteins accomplish this not directly, but through the recruitment of other protein partners. This process is context dependent and differs based on the composition of the protein complexes. Through mRNA regulation PUF proteins play key roles in gametogenesis/gamete maturation, embryogenesis, cell differentiation, and neural development and function (Miller and Olivas, 2011). Because of these biological roles a likely ancestral role for PUF proteins as stem cell maintenance and self-renewal has been hypothesized (Wickens et al., 2002).

Transcriptional regulators. S. cerevisiae Puf5p binds Pop2p to recruit the Ccr4p-Pop2p-Not deadenylase complex which deadenylates mRNAs leading to reduced translation and decay (Goldstrohm et al., 2006; Goldstrohm and Wickens, 2008). This appears to be a conserved mechanism in worms and humans (Goldstrohm et al., 2006; Quenault et al., 2011). Additional deadenylation independent mechanisms of repression have also been observed where PUF proteins likely interact with translation initiation proteins directly (Chagnovich and Lehmann, 2001; Goldstrohm et al., 2007; Hook et al., 2007). PUF proteins have also been show to stimulate decapping which stimulates rapid decay (Olivas and Parker, 2000). PUF proteins can

also promote repression through interactions with Argonaute possibly through the inhibition of the translation factor, eEF1A (Friend et al., 2012). Thus, PUF proteins repress translation through a variety of mechanisms depending on what protein partners are recruited by PUF.

Translational activator. PUF proteins, have in some cases, been shown to activate translation (Archer et al., 2009; Kaye et al., 2009; Piqué et al., 2008; Suh et al., 2009). Possible mechanisms of this activation include enhanced polyadenylation, and stabilization of CPEB independent of polyadenylation (Piqué et al., 2008; Quenault et al., 2011).

PUF proteins are versatile regulators of mRNAs. PUF proteins can play different roles in mRNA regulation depending on PUF-protein partners. Which protein partners are recruited is directed by mRNA sequence elements and the subcellular context. These varied partners allow PUF proteins to enhance translation in some cases while repressing RNAs in other situations.

Yeast PUF proteins

There are six PUF proteins in *S. cerevisiae* (Puf1-Puf6) and an additional protein with Pumilio repeats, nop9 (Dong et al., 2011). *S. cerevisiae* PUF proteins have been extensively studied including: structural determination, identification of RNA targets, and functional analyses have been performed (Miller and Olivas, 2011).

Structure. The Pumilio homology domain of *S. cerevisiae* Puf3p and Puf4p have been determined by X-Ray Crystallography (Miller et al., 2008; Zhu et al., 2009). *S. cerevisiae* Puf3p represents the likely ancestral PUF with orthologs found in *C. elegans*, *Drosophila melanogaster*, and *H. sapiens* (Wickens et al., 2002). The most distinguishing structural feature of *S. cerevisiae* Puf3p, compared to its orthologs, is an additional cytosine-binding pocket flanking the repeat 8 (Zhu et al., 2009). The cytosine-binding pocket recognizes a cytosine at the

5’ end of the “core” binding element. This feature of *S. cerevisiae* Puf3p is important for *in vitro* binding and *in vivo* regulation (Zhu et al., 2009).

S. cerevisiae Puf4p is capable of binding a 9nt binding element by “flipping” a base away from the concave surface of the protein, as discussed above (Miller et al., 2008) (Figure 3C).

RNA targets. Many of the targets for the conical (Puf1p-Puf5p) proteins were defined by RIP-chip (Gerber et al., 2004). Each protein binds a unique set of RNAs which are functionally related. For example, *S. cerevisiae* Puf3p binds nuclear-encoded RNAs that code for mitochondrial proteins, and Puf5p associates with RNAs that code for chromatin related proteins. The number of targets ranged from 40 RNAs for Puf1p to 224 RNAs for Puf5p. Many of the mRNA targets contained PUF binding elements in their 3’UTRs. For targets of *S. cerevisiae* Puf3p, the binding element was 8nt with an additional cytosine at the -2 position of the binding element. Puf4p and Puf5p targets were enriched for 9 and 10nt binding elements, respectively. The binding elements contained the conical 5’ UGUA bases followed by an A/U rich stretch then ended with UA.

Functional analysis. *S. cerevisiae* Puf3p localizes many of its target mRNAs to the mitochondria and interacts with the Arp2/3 complex which contributes to the motility and biogenesis of the mitochondria (Eliyahu et al., 2010; García-Rodríguez et al., 2007; Saint-Georges et al., 2008). Mutants, null or overexpression, have growth phenotypes on glycerol and increased sensitivity to oxidative stress which further implicate *S. cerevisiae* Puf3p in mitochondrial function (Eliyahu et al., 2010; Saint-Georges et al., 2008). In a carbon source dependent way, *S. cerevisiae* Puf3p has been shown to repress targets through decay (Miller et al., 2014; Rowe et al., 2014). In summary, *S. cerevisiae* Puf3p is a likely translational repressor

of nuclear-encoded mitochondria mRNAs, which contributes to the cell's response to oxidative stress and various carbon sources (Miller et al., 2014; Rowe et al., 2014).

The prototypical translational repressor *S. cerevisiae* Puf5p, Mpt5, plays a role in a diverse set of cellular functions including: cell wall integrity, chronological lifespan, chromatin modification, mating type switching, and spindle pole body architecture (Chen and Kurjan, 1997; Kaeberlein and Guarente, 2002; Kennedy et al., 1997; Stewart et al., 2007; Tadauchi et al., 2001; Traven et al., 2010). These diverse functions mirror the diversity of mRNA targets (Gerber et al., 2004).

Ascomycota fungi

The sac fungi are characterized by an ascus which contains ascosopores. The ascomycetes are a monophyletic group, derived from one common ancestor (Figure 4). The Phylum is relevant to humans because it includes the yeast that make beer, wine, cheese, and the most common human fungal pathogen. Additionally, the group includes species that are used in some industrial processes. The group contains the scientifically important containing a number of model organisms including "yeast", *S. cerevisiae*, *Saccharomyces pombe*, and *Neurospora crassa* (Figure 4).

Pezizomycotina is the largest subphylum, and includes the filamentous fungi (Spatafora et al., 2006). The group contains most fungi with ascocarps (fruiting bodies). Two well known species include *Neurospora crassa* and *Aspergillus nidulans* (Figure 4).

Saccharomycotina reproduce asexually by budding (Eriksson and Winka, 1997). The budding, or true yeasts, include: *S. cerevisiae*, baker's yeast, *Candia albicans*, the most common

human fungal pathogen, and Pezizomycotina and Saccharomycotina, which shared a common ancestor after Taphrinomycotina split from the group (James et al., 2006) (Figure 4).

Taphrinomycotina include the fission yeast, Schizosaccharomycetes pombe (Eriksson and Winka, 1997). This group is thought to be basal to the other groups (James et al., 2006) (Figure 4).

There is a wealth of functional information and sequenced genomes making the group one of the best phyla for comparative genomic studies.

Evolution of regulation

The focus of this dissertation was on the evolution of RNA control networks, but there are very few examples on this topic with supporting biochemical data (Beadell and Haag, 2015). In light of this, I will discuss the evolution of transcriptional regulation which will likely have many parallels.

The regulation of subcellular localization and concentration of proteins is vital to cellular life and is accomplished through many processes, including transcriptional and RNA control. Ever since the discovery of gene regulation, it was hypothesized to be a source of phenotypic variation. There is clear evidence of changes in non-coding sequences resulting in dramatic rewiring of transcriptional networks (Li and Johnson, 2010; Tautz, 2000).

Networks can be rewired either by the creation or destruction of DNA binding elements or by changing the binding specificity of the transcriptional regulator. Three general modes of changes observed in transcriptional network evolution result in new circuits (Li and Johnson, 2010). First, binding elements arise in new targets, but the transcription factor binding specificity is maintained (Borneman et al., 2007; Gasch et al., 2004; Tanay et al., 2005). Second,

the targets of the circuit are the same, but the transcription factor which is regulating them is different (Tsong et al., 2006). Recruitment of a new transcription factor to an existing circuit. Third, additional protein partners may recruit transcription factors to existing circuits allowing the old circuit to come under the control of a new (Tuch et al., 2008).

Genetic drift likely contributes to rewiring some transcriptional regulation. There are many ways to regulate a network to achieve a desired outcome, but because of the frequency of rewiring it seems likely some of this evolution is adaptive (Li and Johnson, 2010).

Summary

In this dissertation I have examined the evolution of PUF circuits in Phylum Ascomycota by defining PUF-RNA targets and predicting PUF binding elements. In addition, we have shown that *S. cerevisiae* Puf5p can bind a large and diverse collection of RNAs.

FIGURE LEGENDS

Figure 1. RNA regulation by RNA-binding proteins. RNA-binding proteins regulate each step of mRNA processing. The boxes represent processing steps in the life of an mRNA.

Figure 2. Domain structure of example RNA-binding proteins. Schematics of RNA-binding proteins domain structure. Representative examples of RNA-binding proteins. Each domain is represented by color and shape (see key). PTB, polypyrimidine-tract binding; PABP, poly(A)-binding protein; U2AF, U2 auxiliary factor; SF1, splicing factor-1; TFIIA Transcription factor IIIA; TTP, tristetraprolin. The figure was adapted from (Lunde et al., 2007).

Figure 3. Pumilio homology domain structure. (A) Cartoon of *S. cerevisiae* Puf3p structure in complex with binding element B from *cox17* (Zhu et al., 2009). Repeats 8-1 from top to bottom (R8-R1) are shown in blue. The RNA is shown in green and red along the concave surface. (B) Schematic of *S. cerevisiae* Puf3p TRM interactions (Valley et al., 2012). RNA is on the right and the TRMs are on the right. (C) Schematic of *S. cerevisiae* Puf4p TRM interactions (Valley et al., 2012).

Figure 4. Phylogenetic tree of Phylum Ascomycota. The branches for each Subphylum are colored; Saccharomycotina-red, Pezizomycotina-orange, Taphrinomycotina-blue.

FIGURES

Figure 1

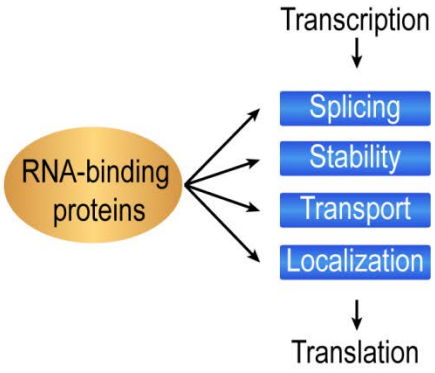


Figure 2

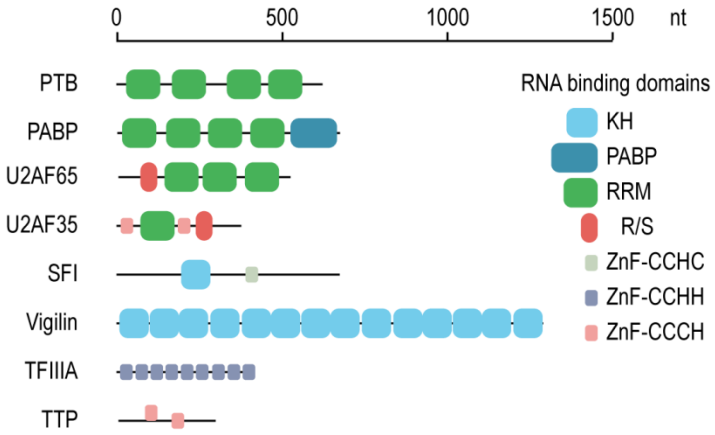

0
500
1000
1500
nt
PTB
PABP
U2AF65
U2AF35
SFI
Vigilin
TFIIIA
TTP
RNA binding domains
KH
PABP
RRM
R/S
ZnF-CCHC
ZnF-CCHH
ZnF-CCCH

Figure 3

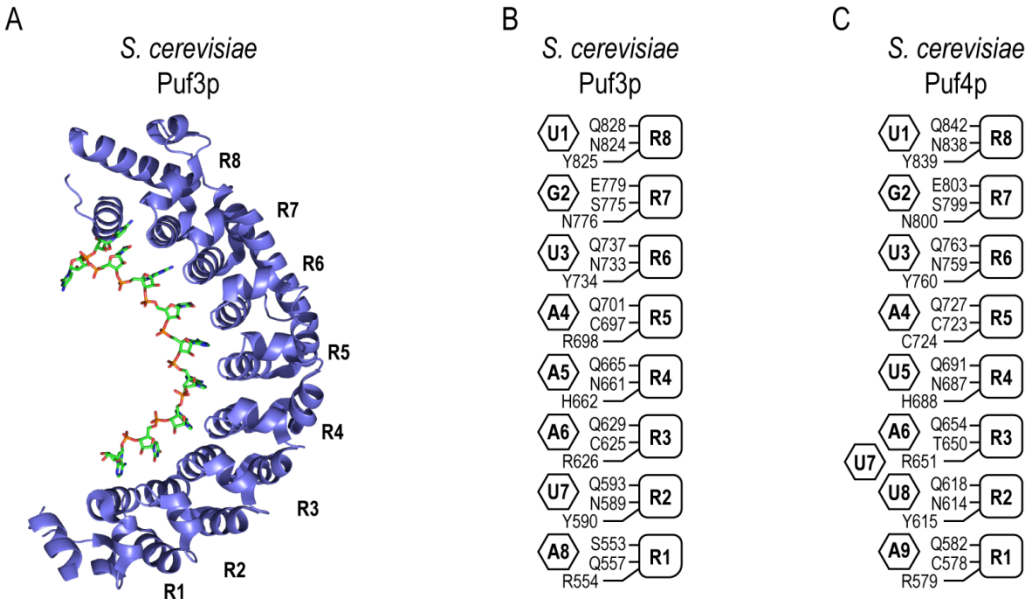
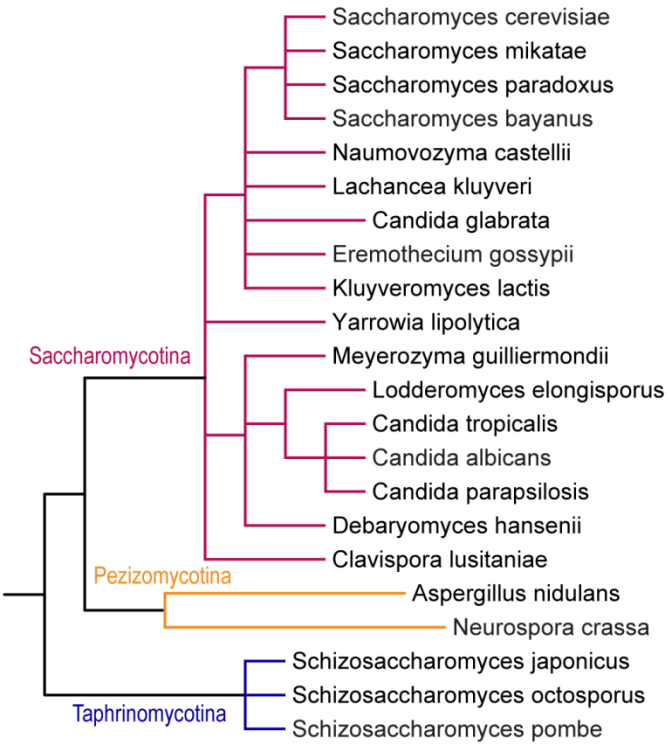

A
S. cerevisiae
Puf3p
R8
R7
R6
R5
R4
R3
R2
R1
B
S. cerevisiae
Puf3p
U1 Q828 N824 Y825 R8
G2 E779 S775 N776 R7
U3 Q737 N733 Y734 R6
A4 Q701 C697 R698 R5
A5 Q665 N661 H662 R4
A6 Q629 C625 R626 R3
U7 Q593 N589 Y590 R2
A8 S553 Q557 R554 R1
C
S. cerevisiae
Puf4p
U1 Q842 N838 Y839 R8
G2 E803 S799 N800 R7
U3 Q763 N759 Y760 R6
A4 Q727 C723 C724 R5
U5 Q691 N687 H688 R4
A6 Q654 T650 R651 R3
U7
U8 Q618 N614 Y615 R2
A9 Q582 C578 R579 R1

Figure 4


Saccharomyces cerevisiae
Saccharomyces mikatae
Saccharomyces paradoxus
Saccharomyces bayanus
Naumovozyma castellii
Lachancea kluyveri
Candida glabrata
Eremothecium gossypii
Kluyveromyces lactis
Yarrowia lipolytica
Meyerozyma guilliermondii
Lodderomyces elongisporus
Candida tropicalis
Candida albicans
Candida parapsilosis
Debaryomyces hansenii
Clavispora lusitaniae
Aspergillus nidulans
Neurospora crassa
Schizosaccharomyces japonicus
Schizosaccharomyces octosporus
Schizosaccharomyces pombe
Saccharomycotina
Pezizomycotina
Taphrinomycotina

REFERENCES

Archer, S.K., Luu, V.D., de Queiroz, R.A., Brems, S., and Clayton, C. (2009). Trypanosoma brucei PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle. PLoS Pathog *5*, e1000565.

Beadell, A.V., and Haag, E.S. (2015). Evolutionary Dynamics of GLD-1-mRNA complexes in Caenorhabditis nematodes. Genome Biol Evol *7*, 314-335.

Borneman, A.R., Gianoulis, T.A., Zhang, Z.D., Yu, H., Rozowsky, J., Seringhaus, M.R., Wang, L.Y., Gerstein, M., and Snyder, M. (2007). Divergence of transcription factor binding sites across related yeast species. Science *317*, 815-819.

Campbell, Z.T., Bhimsaria, D., Valley, C.T., Rodriguez-Martinez, J.A., Menichelli, E., Williamson, J.R., Ansari, A.Z., and Wickens, M. (2012). Cooperativity in RNA-protein interactions: global analysis of RNA binding specificity. Cell Rep *1*, 570-581.

Carballo, E., Lai, W.S., and Blackshear, P.J. (1998). Feedback inhibition of macrophage tumor necrosis factor-alpha production by tristetraprolin. Science *281*, 1001-1005.

Chagnovich, D., and Lehmann, R. (2001). Poly(A)-independent regulation of maternal hunchback translation in the Drosophila embryo. Proc Natl Acad Sci U S A *98*, 11359-11364.

Chen, T., and Kurjan, J. (1997). Saccharomyces cerevisiae Mpt5p interacts with Sst2p and plays roles in pheromone sensitivity and recovery from pheromone arrest. Mol Cell Biol *17*, 3429-3439.

Cléry, A., Blatter, M., and Allain, F.H.-T. (2008). RNA recognition motifs: boring? Not quite. Curr Opin Struct Biol *18*, 290-298.

Daubner, G.M., Cléry, A., and Allain, F.H.-T. (2013). RRM-RNA recognition: NMR or crystallography…and new findings. Curr Opin Struct Biol *23*, 100-108.

Dong, S., Wang, Y., Cassidy-Amstutz, C., Lu, G., Bigler, R., Jezyk, M.R., Li, C., Hall, T.M.T., and Wang, Z. (2011). Specific and modular binding code for cytosine recognition in Pumilio/FBF (PUF) RNA-binding domains. J Biol Chem *286*, 26732-26742.

Duncan, R., Bazar, L., Michelotti, G., Tomonaga, T., Krutzsch, H., Avigan, M., and Levens, D. (1994). A sequence-specific, single-strand binding protein activates the far upstream element of c-myc and defines a new DNA-binding motif. Genes Dev *8*, 465-480.

Eliyahu, E., Pnueli, L., Melamed, D., Scherrer, T., Gerber, A.P., Pines, O., Rapaport, D., and Arava, Y. (2010). Tom20 mediates localization of mRNAs to mitochondria in a translation-dependent manner. Mol Cell Biol *30*, 284-294.

Eriksson, O.E., and Winka, K. (1997). Supraordinal taxa of Ascomycota. Myconet *1*, 1-16.

Friend, K., Campbell, Z.T., Cooke, A., Kroll-Conner, P., Wickens, M.P., and Kimble, J. (2012). A conserved PUF-Ago-eEF1A complex attenuates translation elongation. Nat Struct Mol Biol *19*, 176-183.

García-Rodríguez, L.J., Gay, A.C., and Pon, L.A. (2007). Puf3p, a Pumilio family RNA binding protein, localizes to mitochondria and regulates mitochondrial biogenesis and motility in budding yeast. J Cell Biol *176*, 197-207.

Gasch, A.P., Moses, A.M., Chiang, D.Y., Fraser, H.B., Berardini, M., and Eisen, M.B. (2004). Conservation and evolution of cis-regulatory systems in ascomycete fungi. PLoS Biol *2*, e398.

Gerber, A.P., Herschlag, D., and Brown, P.O. (2004). Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. PLoS Biol *2*, E79.

Goldstrohm, A.C., Hook, B.A., Seay, D.J., and Wickens, M. (2006). PUF proteins bind Pop2p to regulate messenger RNAs. Nat Struct Mol Biol *13*, 533-539.

Goldstrohm, A.C., Seay, D.J., Hook, B.A., and Wickens, M. (2007). PUF protein-mediated deadenylation is catalyzed by Ccr4p. J Biol Chem *282*, 109-114.

Goldstrohm, A.C., and Wickens, M. (2008). Multifunctional deadenylase complexes diversify mRNA control. Nat Rev Mol Cell Biol *9*, 337-344.

Hall, T.M.T. (2005). Multiple modes of RNA recognition by zinc finger proteins. Curr Opin Struct Biol *15*, 367-373.

Hook, B.A., Goldstrohm, A.C., Seay, D.J., and Wickens, M. (2007). Two yeast PUF proteins negatively regulate a single mRNA. J Biol Chem *282*, 15430-15438.

James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J., *et al.* (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature *443*, 818-822.

Kaeberlein, M., and Guarente, L. (2002). Saccharomyces cerevisiae MPT5 and SSD1 function in parallel pathways to promote cell wall integrity. Genetics *160*, 83-95.

Kaye, J.A., Rose, N.C., Goldsworthy, B., Goga, A., and L'Etoile, N.D. (2009). A 3'UTR pumilio-binding element directs translational activation in olfactory sensory neurons. Neuron *61*, 57-70.

Kennedy, B.K., Gotta, M., Sinclair, D.A., Mills, K., McNabb, D.S., Murthy, M., Pak, S.M., Laroche, T., Gasser, S.M., and Guarente, L. (1997). Redistribution of silencing proteins from telomeres to the nucleolus is associated with extension of life span in S. cerevisiae. Cell *89*, 381-391.

Koh, Y.Y., Opperman, L., Stumpf, C., Mandan, A., Keles, S., and Wickens, M. (2009). A single C. elegans PUF protein binds RNA in multiple modes. RNA *15*, 1090-1099.

Li, H., and Johnson, A.D. (2010). Evolution of transcription networks--lessons from yeasts. Curr Biol *20*, R746-753.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat Rev Mol Cell Biol *8*, 479-490.

Maris, C., Dominguez, C., and Allain, F.H.-T. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J *272*, 2118-2131.

Miller, M.A., and Olivas, W.M. (2011). Roles of Puf proteins in mRNA degradation and translation. Wiley Interdiscip Rev RNA *2*, 471-492.

Miller, M.A., Russo, J., Fischer, A.D., Lopez Leban, F.A., and Olivas, W.M. (2014). Carbon source-dependent alteration of Puf3p activity mediates rapid changes in the stabilities of mRNAs involved in mitochondrial function. Nucleic Acids Res *42*, 3954-3970.

Miller, M.T., Higgin, J.J., and Hall, T.M.T. (2008). Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p. Nat Struct Mol Biol *15*, 397-402.

Nicastro, G., Taylor, I.A., and Ramos, A. (2015). KH-RNA interactions: back in the groove. Curr Opin Struct Biol *30C*, 63-70.

Olivas, W., and Parker, R. (2000). The Puf3 protein is a transcript-specific regulator of mRNA degradation in yeast. EMBO J *19*, 6602-6611.

Picard, B., and Wegnez, M. (1979). Isolation of a 7S particle from Xenopus laevis oocytes: a 5S RNA-protein complex. Proc Natl Acad Sci U S A *76*, 241-245.

Piqué, M., López, J.M., Foissac, S., Guigó, R., and Méndez, R. (2008). A combinatorial code for CPE-mediated translational control. Cell *132*, 434-448.

Quenault, T., Lithgow, T., and Traven, A. (2011). PUF proteins: repression, activation and mRNA localization. Trends Cell Biol *21*, 104-112.

Rowe, W., Kershaw, C.J., Castelli, L.M., Costello, J.L., Ashe, M.P., Grant, C.M., Sims, P.F.G., Pavitt, G.D., and Hubbard, S.J. (2014). Puf3p induces translational repression of genes linked to oxidative stress. Nucleic Acids Res *42*, 1026-1041.

Saint-Georges, Y., Garcia, M., Delaveau, T., Jourdren, L., Le Crom, S., Lemoine, S., Tanty, V., Devaux, F., and Jacq, C. (2008). Yeast mitochondrial biogenesis: a role for the PUF RNA-binding protein Puf3p in mRNA localization. PLoS One *3*, e2293.

Salazar, A.M., Silverman, E.J., Menon, K.P., and Zinn, K. (2010). Regulation of synaptic Pumilio function by an aggregation-prone domain. J Neurosci *30*, 515-522.

Spatafora, J.W., Sung, G.-H., Johnson, D., Hesse, C., O'Rourke, B., Serdani, M., Spotts, R., Lutzoni, F., Hofstetter, V., Miadlikowska, J., *et al.* (2006). A five-gene phylogeny of Pezizomycotina. Mycologia *98*, 1018-1028.

Stewart, M.S., Krause, S.A., McGhie, J., and Gray, J.V. (2007). Mpt5p, a stress tolerance- and lifespan-promoting PUF protein in Saccharomyces cerevisiae, acts upstream of the cell wall integrity pathway. Eukaryot Cell *6*, 262-270.

Suh, N., Crittenden, S.L., Goldstrohm, A., Hook, B., Thompson, B., Wickens, M., and Kimble, J. (2009). FBF and its dual control of gld-1 expression in the Caenorhabditis elegans germline. Genetics *181*, 1249-1260.

Tadauchi, T., Matsumoto, K., Herskowitz, I., and Irie, K. (2001). Post-transcriptional regulation through the HO 3'-UTR by Mpt5, a yeast homolog of Pumilio and FBF. EMBO J *20*, 552-561.

Tanay, A., Regev, A., and Shamir, R. (2005). Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. Proc Natl Acad Sci U S A *102*, 7203-7208.

Tautz, D. (2000). Evolution of transcriptional regulation. Curr Opin Genet Dev *10*, 575-579.

Trabucchi, M., Briata, P., Garcia-Mayoral, M., Haase, A.D., Filipowicz, W., Ramos, A., Gherzi, R., and Rosenfeld, M.G. (2009). The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. Nature *459*, 1010-1014.

Traven, A., Lo, T.L., Lithgow, T., and Heierhorst, J. (2010). The yeast PUF protein Puf5 has Pop2-independent roles in response to DNA replication stress. PLoS One *5*, e10651.

Tsong, A.E., Tuch, B.B., Li, H., and Johnson, A.D. (2006). Evolution of alternative transcriptional circuits with identical logic. Nature *443*, 415-420.

Tuch, B.B., Li, H., and Johnson, A.D. (2008). Evolution of eukaryotic transcription circuits. Science *319*, 1797-1799.

Valley, C.T., Porter, D.F., Qiu, C., Campbell, Z.T., Hall, T.M., and Wickens, M. (2012). Patterns and plasticity in RNA-protein interactions enable recruitment of multiple proteins through a single site. Proc Natl Acad Sci U S A *109*, 6054-6059.

Valverde, R., Edwards, L., and Regan, L. (2008). Structure and function of KH domains. FEBS J *275*, 2712-2726.

Wang, X., McLachlan, J., Zamore, P.D., and Hall, T.M.T. (2002). Modular recognition of RNA by a human pumilio-homology domain. Cell *110*, 501-512.

Wang, X., Zamore, P.D., and Hall, T.M. (2001). Crystal structure of a Pumilio homology domain. Mol Cell *7*, 855-865.

Wang, Y., Opperman, L., Wickens, M., and Hall, T.M.T. (2009). Structural basis for specific recognition of multiple mRNA targets by a PUF regulatory protein. Proc Natl Acad Sci U S A *106*, 20186-20191.

Wang, Y., Wang, Z., and Tanaka Hall, T.M. (2013). Engineered proteins with Pumilio/fem-3 mRNA binding factor scaffold to manipulate RNA metabolism. FEBS J *280*, 3755-3767.

Wickens, M., Bernstein, D.S., Kimble, J., and Parker, R. (2002). A PUF family portrait: 3'UTR regulation as a way of life. Trends Genet *18*, 150-157.

Zamore, P.D., Williamson, J.R., and Lehmann, R. (1997). The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. RNA *3*, 1421-1433.

Zhang, B., Gallegos, M., Puoti, A., Durkin, E., Fields, S., Kimble, J., and Wickens, M.P. (1997). A conserved RNA-binding protein that regulates sexual fates in the C. elegans hermaphrodite germ line. Nature *390*, 477-484.

Zhu, D., Stumpf, C.R., Krahn, J.M., Wickens, M., and Hall, T.M. (2009). A 5' cytosine binding pocket in Puf3p specifies regulation of mitochondrial mRNAs. Proc Natl Acad Sci U S A *106*, 20192-20197.

Chapter 2

RNA regulatory networks established through curvature of a PUF protein scaffold

This work has been submitted for publication: **Wilinski D***, Qiu C*, Lapointe CP, Nevil M, Campbell ZT, Hall TM, Wickens M. RNA regulatory networks established through curvature of a PUF protein scaffold. * Co-first authors.

I contributed to all the figures and writing. Chen Qiu preformed all of the structural work, analysis, and contributed to the writing; Chris Lapointe made intellectual contributions to all Figures except Figure 4 and planned Figure 6, Markus Nevil conducted early experiments that led to Figure 5, Zachary Campbell contributed Figure 2c, and Traci Hall and Marvin Wickens oversaw the work and contributed to writing the manuscript.

ABSTRACT

Proteins bind and control mRNAs, directing their localization, translation, and stability. Members of the PUF family of RNA-binding proteins control multiple mRNAs in a single cell, and play key roles in development, stem cell maintenance and memory formation. Here we identified the mRNA targets of a *S. cerevisiae* PUF protein, Puf5p, by UV-crosslinking-affinity purification and high-throughput sequencing (HITS-CLIP). Puf5p associates with ~1,000 RNAs, suggesting that it is a broad spectrum regulator. The binding sites recognized by Puf5p are diverse, with variable spacer lengths between two specific sequences. Each length of site correlates with a distinct biological function. Crystal structures of Puf5p-RNA complexes reveal that the protein scaffold presents an exceptionally flat and extended interaction surface relative to other PUF proteins. In complexes with RNAs of different lengths, the protein is unchanged. Rather, the RNAs adopt conformations that maintain atomic contacts with the protein that are similar to those in other RNA-PUF protein complexes. A single PUF protein repeat is sufficient to induce broadening of specificity. Changes in protein architecture, such as alterations in curvature, may lead to evolution of mRNA regulatory networks.

INTRODUCTION

RNA-binding proteins control an mRNA's life, including its translation, movement, and destruction. These events underlie diverse biological processes, ranging from early development to memory formation. Regulatory proteins bind simultaneously to short RNA sequences, typically in 3′ untranslated regions (3′UTRs), and to protein effectors that determine the RNA's fate. The RNA binding specificities of the proteins determine which mRNAs are controlled, while effectors determine the outcomes.

RNA regulatory networks, in which a single RNA binding protein controls multiple mRNAs, are widespread[1]. For example, Cytoplasmic Polyadenylation Element Binding protein, an RNA recognition motif-containing protein, binds to and regulates many mRNAs that participate in the regulation of embryonic cell cycles[2], while Nova co-regulates multiple mRNAs with roles in alternative polyadenylation and splicing[3]. As a result, RNA-binding proteins integrate post-transcriptional controls, as DNA-binding proteins coordinate transcriptional regulation. To understand RNA regulatory circuits in molecular terms, we need to know which mRNAs are controlled, how they are recognized, and how the networks change during evolution.

PUF proteins are exemplary mRNA regulators[4,5]. They bind to the 3′UTRs of many mRNAs and do so through single-stranded RNA binding elements. For example, Puf3p of yeast binds nuclear-encoded mRNAs with roles in mitochondria[6]. Similarly, PUF proteins in *C. elegans*, *Drosophila* and humans control an overlapping battery of mRNAs with established roles in stem cells[7]. The RNA-binding specificities of most PUF proteins are defined in part by three amino acids (tripartite recognition motifs, or TRMs) in each of eight tandemly reiterated PUF repeats[8]. TRMs, annotated as XY-Z, recognize specific bases through edge-on (residues X

and Y) and stacking interactions (residue Z)[9]. Specificity also can be achieved through the requirement for a base that does not contact the protein, but is solvent-exposed[8,9]. RNA-immunoprecipitation and microarray ("RIP-Chip") studies suggested that *Saccharomyces cerevisiae* Puf5p binds ~200 mRNAs that contain 10 nt long binding elements[6]. Genetic analysis has implicated Puf5p in multiple cellular functions, including lifespan[10], cell wall integrity[11], chromatin structure[10], and mating type switching[12], consistent with the view that it participates in the control of diverse groups of mRNAs.

In this work, we used UV-crosslinking and high-throughput sequencing to define ~1000 high-confidence RNA targets of Puf5p *in vivo*. These targets possess unexpected diversity of binding element lengths, with the same RNA sequence features at the two ends but varying numbers of nucleotides in between. The lengths of sites correlate with the biological functions of the targets. The crystal structures of Puf5p – RNA complexes revealed that the RNAs assume altered conformations to accommodate a fixed protein architecture. The plasticity in binding element length is driven by the flattened curvature of the PUF protein scaffold. The findings suggest ways in which alterations in protein curvature result in new specificities and enable the evolution of new RNA networks.

RESULTS

Identification of Puf5p RNA targets

Using *in vivo* UV crosslinking *and* high-throughput sequencing (HITS-CLIP[13,14]), we identified more than 1,000 mRNAs to which Puf5p binds in *S. cerevisiae*, representing 16% of the yeast transcriptome (Fig. 1a). Cells containing Puf5p fused to a tandem affinity purification (TAP) tag were irradiated during mid-log phase (Fig. 1a i). After lysis and mild RNase treatment (Fig. 1a ii), Puf5p was stringently purified through tandem affinity steps (Fig. 1a iii) and SDS PAGE electrophoresis. The purification of cross-linked complexes was effective, as evidenced by Western blotting (Fig. 1b). Complexes whose RNA components had been ^{32}P-end-labeled exhibited heterogenous, slower mobilities than Puf5p alone (Fig. 1c). To identify RNAs bound to Puf5p, adaptors were ligated, the protein digested, and the RNAs converted to cDNAs that were analyzed by high throughput sequencing (Fig. 1a iv). The adaptors contained random bar-codes, so that PCR duplication events could be discarded.

Upon aligning the sequence reads to the yeast genome, we found the majority of peaks were within 3´UTRs of mRNAs and that the set of target mRNAs were distinct from, but overlapped, those of other yeast PUF proteins. We obtained 16,300,145 and 11,100,468 reads from two biological replicates. Of these, 616,401 and 491,532 (6% and 7%) mapped to unique locations in the *S. cerevisiae* genome, after filtering by quality score and removing PCR duplicates (Fig. 1a v; Supplementary Fig. 1a). The functional enrichment of targets detected was only minimally affected by changing the filtering methods we used (Supplementary Fig. 1b). A total of 1,439 peaks was identified, representing a total of 1,190 RNAs (Supplementary Data Set 1). Of these, 1,043 (86%) mapped to mRNAs. The remaining 14% mapped to non-coding RNAs,

including snoRNA, snRNA, ncRNA, and tRNAs (Supplementary Fig. 2). Of the peaks in mRNAs, most resided in 3´UTRs (68%), but a fraction mapped to ORFs (28%) or 5´UTRs (4%) (Fig. 1d, see Supplementary Data Set 1 for complete gene list). Approximately 6% of Puf5p targets were shared with Puf4p (as deduced from RIP-Chip experiments[6]) and 27% with Puf3p (from CLIP studies[15]) (Supplementary Fig. 3). The 1,043 Puf5p targets represent 16% of yeast mRNAs, a five-fold increase relative to the 206 targets detected in earlier RIP-Chip studies[6] (Supplementary Fig. 3a) (see Discussion). We conclude that Puf5p is a broad regulator of a distinct set of mRNAs in *S. cerevisiae.*

Puf5p binding elements range in length from 8 to 12 nt

We developed stringent criteria to select a set of 1,043 high-confidence targets that we used to identify RNA sequence elements bound by Puf5p. We first defined significant peaks as an enrichment of independent reads in a specific genic region (modFDR <0.01)[16] (Supplementary Fig. 1a). To identify high-confidence targets, we required that a peak contain 1-nt or 2-nt deletions in multiple reads (a strong indicator that these RNAs had been cross-linked to Puf5p[17]) and a minimum of 10 reads per peak. In addition, these criteria had to be satisfied in both biological replicates (Supplementary Fig. 1a). Normalized peak heights at specific loci were reproducible between the biological replicates (Pearson correlation coefficient 0.90) (Fig. 1e). Previous studies validated several putative targets of Puf5p by showing they were regulated by that protein *in vivo*[18,19]. Among the best characterized are *SMX2* and *HO* mRNAs[20,21], which are used here as examples (Fig. 2a). With both mRNAs, peaks lay over the previously characterized binding elements in their 3´UTRs (Fig. 2a). (For additional examples, see Supplementary Fig. 5a-c.)

We identified five classes of Puf5p binding elements ranging from 8-12 nucleotides, each comprising a 5´-UGUA tetranucleotide sequence and a 3´UA with a variable length spacer region in between (Fig. 2b). We performed an unbiased search of the complete set of high-confidence targets for over-represented sequences in peaks using MEME[22]. The position weight matrix we obtained consists of a 5´UGUA tetranucleotide sequence followed by a degenerate 3´ end (Fig. 2b). However, we could de-convolute the complete set of 5´UGUA-containing sequences into five classes of binding elements, ranging in length from 8 to 12 nts beginning at the 5´UGUA, each with a 3´terminal UA sequence (Fig. 2b). 71% of the 1,043 targets, and 66% of the total number of peaks (1,439), contained at least one Puf5p binding element. Peaks without enriched sequences may reflect contacts that were less sequence-specific or mediated by interactions between Puf5p and RNA-bound factors. The sequences between UGUA and UA display little difference compared to the background nucleotide frequencies surrounding the sites (Supplementary Fig. 4), though adenosine was modestly enriched 1 or 2 nts upstream of the 3´terminal UA in 9 and 10 nt elements (Supplementary Fig. 4a), and guanosines were uncommon (Supplementary Fig. 4c).

The breadth of binding element lengths associated with Puf5p is unusual among PUF proteins (Fig. 2c). For example, three other PUF proteins – human PUM2[14], *S. cerevisiae* Puf4p[6] and *C. elegans* FBF-2[7] – show a single dominant length of site *in vivo*, measured either by CLIP methods[23] or inferred from RIP-Chip[24] (Fig. 2c). Essentially the same behavior was observed for each protein *in vitro*[25]. To examine the sequence preferences of Puf5p *in vitro*, the purified protein was incubated with an RNA library in which 20 consecutive nucleotides had been randomized, generating a theoretical complexity of 4^{20} RNA sequences[25]. Bound RNAs were eluted, and the process repeated five times. The RNAs were analyzed by high-throughput

sequencing. In this method, termed SEQRS, the number of reads obtained is a proxy for affinities measured *in vitro*[25]. Re-analysis of data obtained with Puf5p[25] revealed that the number of reads for each site length yielded a pattern similar to that seen in HITS-CLIP, in that 9-nt and 10-nt sites were the most abundant (Fig. 2c). 11-nt and 12-nt sites were less prevalent in SEQRS than *in vivo*.

Site length correlates with biological function

The majority of the sites bound by Puf5p consist of individual elements, in which only a single site of unambiguous length is present in the CLIP peak (Fig 3a and Supp Table 1). Gene Ontology (GO) analysis of the Puf5p targets suggests that each binding element length found in mRNAs correlates with a distinct biological role (Fig. 3b). Surprisingly, when we analyzed mRNAs with different binding site lengths separately, RNAs with 8-nt binding elements were overrepresented for mitochondrion organization (p-value 9.5e-4); 9-nt sites for ribosome biogenesis (p-value 3.6e-14); and 10-nt sites for regulation of gene expression (p value 2.9 e-6), 11-nt sites for translation (p-value 3.6e-3) (Fig. 3b). 12 nt sites did not correlate with a specific GO term.

The 8-nt Puf5p elements lie in a subset of mRNAs that also were bound to Puf3p in PAR-CLIP experiments[15]. Puf3p associated with ~1,000 nuclear-encoded mRNAs with mitochondrial functions[15]. Even if the criteria selecting high-confidence Puf5p targets are relaxed – not filtering for gapped reads, for example – a very similar enrichment emerges (Supplementary Fig. 1b). Similarly, 22% of the 9 nt Puf5p elements lie in mRNAs that bind Puf4p and are enriched for genes with ribosome assembly and nucleolar functions. We suggest that the restricted specificities of Puf3p and Puf4p for 8 and 9 nt sites, respectively, underlie the correlations

between Puf5p length of binding sites and their biological functions. In particular, we suggest that the broadened specificity of Puf5p enabled its recruitment to pre-existing RNA regulatory circuits in the *S. cerevisiae* lineage (see Discussion).

Alternate RNA conformations adapt Puf5p binding elements to a fixed protein scaffold

To understand how Puf5p accommodates a wide range of target site lengths, we determined crystal structures of the Puf5p RNA-binding domain bound to RNAs of 9-12 nucleotides (Fig. 4, Supplementary Table 2), including the recognition sequences in *SMX2* (9-nt, 2.35 Å resolution), *MFA2* (10-nt, 2.15 Å resolution), *AAT2* (11-nt, 2.5 Å resolution), and *AMN1* (12-nt, 2.8 Å resolution) mRNAs (sites of 8 nts did not crystallize). Each RNA corresponded to a high-confidence 3´UTR binding element in HITS-CLIP (Fig. 2a and Supplementary Fig. 5). The RNAs bound with affinities that corresponded to our structural observations, in that higher affinity RNA binding sites correlated with more RNA bases specifically recognized and/or larger numbers of protein:RNA contacts (Fig. 4, Supplementary Fig. 6, Supplementary Table 3).

Despite the varying lengths and sequences of the RNA binding sites, the overall conformation of Puf5p was unchanged in the four crystal structures (rmsd < 0.7 Å over all Cα atoms or <1.1 Å over all protein atoms). The protein scaffold comprises eight α-helical repeats flanked by a short N-terminal sequence and a C-terminal helix (R8´) (Fig. 4a). The C-terminal repeats 5-8 bound the 5´UGUA RNA sequence, while repeats 1 and 2 bound the UA 3´ element (Fig. 4a, b). Repeats 3 and 4 lie opposite the variable central regions of the RNAs (Fig. 4c-f). While the overall architecture of Puf5p resembles that of other PUF proteins, Puf5p's repeats are more irregular in length and structure than seen in human PUM1 and *S. cerevisiae* Puf3p and Puf4p (Supplementary Fig. 7a). For example, repeats 7 and 8 in Puf5p are unusually long (64

and 72 residues versus 36 in a typical repeat) with extended α2 and α3 helices and inter-helix loops. The positions of the α3 helices relative to the α1 and α2 helices are also more varied in Puf5p than Puf3p or Puf4p (Supplementary Fig. 7a).

Since the curvature of the Puf5p scaffold is fixed, RNAs of different lengths adopt different conformations, as described below. Recognition of the 5´UGUA and 3´UA elements by repeats 5-8 and 1-2, respectively, are identical in all structures. Differences in RNA conformation and recognition are found opposite the central repeats 3 and 4.

"5-parallel:" 9 nt site. Puf5p binds to the 9-nt *SMX2* RNA site by recognizing all but the central fifth base (Fig. 4c). Bases 1-4 and 6-9 are each recognized by a PUF repeat (Supplementary Fig. 5a). However, the fifth base, C5, lies in an atypical conformation, in which the plane of the base is parallel to the axis of the protein, within van der Waals bonding distance of the side chain of Cys 381 in repeat 5 (Fig. 4c). The ribose rings of C5 and U6 adopt C2´-endo conformations to accommodate positioning base C5.

"8-flipped:" 10 nt site. Puf5p binds the 10-nt *MFA2* site similarly to the 9-nt *SMX2* site, but an additional base is accommodated by turning the 8th base away from the RNA-binding surface opposite repeat 3 (Fig. 4d). The positions for all but the 8th base overlap with the 9 nt *SMX2* RNA, and the protein:RNA recognition pattern is similar, though the 7th base is a uracil in *MFA2* and an adenine in *SMX2* (Fig. 4b, Supplementary Fig. 5a).

"5-stacked:" 11 nt site. Puf5p appears to recognize only the 5´-UGUA conserved element and two additional 3´ bases of the 11-nt *ATT2* site (Fig. 4e, Supplementary Fig. 5b), consistent with weaker binding of Puf5p to this site than 9- or 10-nt sites (12- or 2-fold weaker binding, respectively, Supplementary Table 3). A 2.5 Å crystal structure of Puf5p:*ATT2* reveals electron density for bases 1-5 and for two 3´ bases bound to Puf5p repeats 1 and 2. In contrast to

the parallel orientation of base 5 in 9- and 10-nt sites, base A5 of *ATT2* stacks directly with base A4 and forms a van der Waals contact with the side chain of Cys381 in Puf5p repeat 5 (Fig. 4e). We will refer to this conformation as 5-stacked. Using the consensus sequences as a guide, we modeled the 3´ bases as the conserved U10 and A11 bases and did not model bases 6-9. However, alternate conformations of the RNA are possible, including a conformation similar to that of the 12-nt *AMN1* site.

"Triple-stacked:" 12 nt site. Puf5p binds to the longer 12-nt *AMN1* site with a distinct RNA conformation. Unlike the conformations of the shorter length binding sites, bases A4, A5, and C6 stack directly with each other opposite repeat 5. Residues in Repeat 5 (Cys381 and Lys385) contact bases A5 and C6 (Fig. 4f). Puf5p repeat 4 does not interact with an RNA base using its edge-interacting residues, but base U7 is bound to repeat 3 (Fig. 4f). Electron density was observed for bases 1-7 and two 3´ bases bound to Puf5p repeats 1 and 2. We modeled the 3´ bases as the conserved U11 and A12 bases, as we did for the 11-nt *ATT2* site, and bases 8-10 were not included in the model.

Curvature as a determinant of specificity

The flatter RNA-binding surface of Puf5p contributes to its specificity by creating a more extended RNA-binding surface. Puf5p possesses the least curved RNA-binding surface observed among PUF proteins to date (Supplementary Fig. 7b) and binds to the longest RNA target sequences identified thus far. Puf3p preferentially binds 8-nt sites and exhibits the greatest curvature among the yeast PUFs (Supplementary Fig. 7b-d); this reflects the regular spacing of RNA-binding helices, which matches the spacing of bases in an extended RNA chain[26]. Puf4p,

which binds 9-nt binding sites, is intermediate in curvature, between Puf3p and Puf5p (Supplementary Fig. 7b-c).

Extension of the Puf5p RNA-binding surface is produced by the structural arrangements in repeats 4 and 5 and corresponds to the variability in Puf5p target sequence length relative to other PUF proteins. The largest repeat-to-repeat angle in Puf5p is centered about repeat 5 (Supplementary Fig. 7e). Repeat 5 also lacks a large side chain capable of stacking with RNA bases and lies opposite several of the atypical RNA conformations (5-parallel, 5-stacked, and triple-stacked). The flatness combined with a protein surface lacking specificity allows "extra" RNA nucleotides, needed to span the distance between repeats with base specificity, to assume different conformations. These extra nucleotides may not contact the protein, but instead stack with one another or lie parallel to the RNA-binding surface.

Evolution of binding specificity of PUF proteins across Ascomycota

To examine the evolution of the broad specificity of Puf5p, we probed the RNA binding preferences of Puf5p proteins from representative species across Phylum Ascomycota. This group includes the budding yeasts, filamentous fungi, and fission yeasts (Fig. 5a). We used the yeast three-hybrid assay to measure the affinities of Puf5p orthologues from six different species – *S. cerevisiae*, *Saccharomyces bayanus*, *Eremothecium gossypii*, *Candida albicans*, *Neurospora crassa*, and *Schizosaccharomyces pombe* (Fig. 5a). These proteins were identified as orthologues using SYNERGY, which relies on the species tree, sequence similarity and synteny[27]. Their binding preferences versus length of site were evaluated using a set of RNAs 8-12 nt in length, conforming to the sequence UGUA$(A)_{2\text{-}5}$UA using the yeast three-hybrid system[28]. All the RNAs thus maintained the 5´UGUA and 3´UA critical for *S. cerevisiae* Puf5p interaction and

contained a single 3´UA element to define the target length unambiguously. In the three-hybrid assay, the level of expression of a reporter gene (LacZ) is a proxy for the affinity of the interaction[29].

Puf5p proteins across Ascomycota exhibited broad binding specificities. Puf5 proteins bound similarly to sites of 9, 10, 11 and 12 nts (Fig. 5b). The more restricted specificities of Puf4p for 9 nt sites (Fig. 5c), and Puf3p for 8 nt sites (Fig. 5d), also were conserved across the entire Phylum, with the exception of *S. pombe* Puf3. This protein bound a broad range of site lengths, unlike its orthologues in other species that showed preference for 8-nt sites.

The broadened specificity of *S. pombe* Puf3 appears to have arisen exclusively in the fission yeast lineage, which enabled us to probe how that broadening arose during evolution. We reasoned that the broadening was not due to the identity of the RNA-interacting TRMs, as the residues are identical among all the Puf3p orthologues (with the exception of Repeat 3 in *N. crassa*, with a Gln to Arg substitution). To identify the key regions of the proteins that confer specificity, we prepared chimeras in which segments of the *S. cerevisiae* and *S. pombe* proteins were exchanged (Fig. 6a). The specificity profile – broad or narrow – was conferred by PUF Repeats 6-8. A chimeric protein possessing Repeats 6-8 from *S. pombe* exhibited broad specificity, while a chimera with Repeats 6-8 of the *S. cerevisiae* protein had narrow specificity (Fig. 6a,b). The protein sequences in Repeat 6 contain a divergent region among Puf3p orthologues (Supplementary Fig. 8). Indeed, substitution of *S. pombe* Repeat 6 alone into an *S. cerevisiae* scaffold was sufficient to confer the broad specificity profile (Fig. 6b).

DISCUSSION

Puf5p is a broad regulator of RNAs in *S. cerevisiae*, binding to more than 1,000 RNA targets, constituting ~16% of the transcriptome. 71% of these targets possess recognizable binding elements beginning with a 5´ UGU sequence, which range in length from 8 to 12 nucleotides. The variations in length are accommodated by conformational adaptations of the RNA onto a fixed protein scaffold. The wide range of mRNA target site lengths is consistent with prior studies that linked Puf5p to a spectrum of functions, including cell wall integrity[11] and chromatin structure[10].

The biological functions of target mRNAs are correlated with the length of binding elements they possess. How does this correlation arise? We propose that the correlation is imposed by other RNA-binding proteins that recognize the same binding elements, and whose specificity is much more restricted than Puf5p (Fig. 5b-d). For example, Puf3p binds 8-nt sites that are largely in mRNAs with mitochondria-related functions, while Puf4p binds 9-nt binding elements in mRNAs with roles in ribosomal biogenesis and assembly[6].

Two PUF proteins that bind the same site could do so sequentially, competitively or cooperatively. Genetic studies demonstrate that Puf4p and Puf5p redundantly control the decay rate of common targets[30]. In the absence of one of the proteins, the other is sufficient. However, for other common targets, the actions of two PUF proteins may be sequential. For example, *MRPL8* mRNA is a target of both Puf5p and Puf3p, possesses a single binding element, and is localized to the mitochondrial periphery in a Puf3p-dependent fashion[31]. Puf5p could exchange with Puf3p, facilitating repression (Puf5p) en route to localization to mitochondria (Puf3p).

While 71% of Puf5p targets possess discernible binding elements, 29% do not. RNAs without binding elements may associate with Puf5p indirectly, perhaps through a protein to

which it and Puf5p are bound. Cross-linking to RNAs without sites could also be driven by their high concentrations in specific subcellular compartments (such as P-bodies), in which proteins and RNAs are present at high concentrations, and low complexity, Q/N-rich regions present in Puf3p, Puf4p, and Puf5p proteins that could facilitate aggregation[32].

RNAs of different lengths adopt a broad range of conformations when bound to Puf5p. The flatter, extended scaffold of Puf5p, combined with its specificity for 5′ and 3′ sequences, imposes the requirement for these RNA conformational variations and permits recognition of 8-12-nt length RNAs. The elegance of this arrangement is that very similar sets of atomic contacts between amino acids and RNA bases are maintained in the different complexes, despite the range of RNA lengths they possess. For example, 18 of the 21 edge-on contacts made between Puf3p and its RNA target are also made in Puf5p bound to a 10-nt length site. In an analogous manner, β-catenin maintains a fixed scaffold to recognize peptides from different ligands (reviewed in [33]). Its central α-helical Armadillo (ARM) repeats interact with conserved sequence elements in an extended peptide while N- and C-terminal ARM repeats bind elements unique to that ligand. The changes in repeat-to-repeat arrangement at the junctions between the central ARM repeats and N- or C-terminal repeats seem to mark the regions with different protein-binding functions. In the same fashion, changes in curvature at specific repeat junctions in PUF proteins correlate with specialization in RNA-binding specificity.

The fact that a single repeat can broaden or narrow specificity (Fig. 6) suggests that this sort of change may be common in evolution (Fig. 6b). The sixth PUF repeat of Puf3p determines whether that protein binds 8 nt sites (*S. cerevisiae*) or accommodates 8, 9 or 10 nt sites (*S. pombe*). *S. cerevisiae* Repeat 6, which induces narrow specificity, contains additional residues relative to the same region of the *S. pombe* protein, which, although not near the RNA-binding

residues, may alter the structure with corresponding effects on specificity (Supplementary Fig. 7a and 8).

From an evolutionary perspective, the broadening of Puf5p's specificity enabled new regulatory inputs into existing RNA circuits. In this view, the ability of Puf5p to recognize a wide array of target lengths arose after ancestral proteins (e.g., Puf3p) already regulated batteries of RNAs with related functions and conserved lengths of sites. Recruitment of Puf5p to these same targets, enabled by its flatter curvature, then provided new regulatory inputs and/or redundancy into that same circuit. For example, Puf5p binds regulatory kinases[34], whose input could be brought to bear on a pre-existing circuit. We suggest that curvature of the scaffold is critical in defining the RNAs that are controlled. Acquisition of new RNA specificities by alterations of the protein's architecture suggests ways in which new RNA circuits are established, expanded and contracted during evolution.

ACKNOWLEDGEMENTS

We thank Natascha Buter for help with experiments, and colleagues in the Wickens lab and NIEHS for comments during the work and discussions of the manuscript. We are very grateful to Drs. Robert Darnell and Aldo Mele for helpful suggestions in troubleshooting HITS-CLIP and Dr. Fabio Parmeggiani for help analyzing the β-catenin repeats. We thank Dr. Lars Pedersen and the staff of the Southeast Regional Collaborative Access Team beamlines for assistance with X-ray data collection. Work in the Wickens lab is supported by NIH (GM50942). ZKW was supported by an NIH Postdoctoral Fellowship. The work also was supported by the Intramural Research Program of the National Institutes of Health, National Institute of Environmental Health Sciences (TMTH). The Advanced Photon Source used for this study was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract no. W-31-109-Eng-38.

DATA DEPOSITION

The Illumina FASTQ sequence files have been deposited in the Gene Expression Omnibus (Accession # xxx). The atomic coordinates and structure factors for the crystal structures of Puf5p in complex with *SMX2* (PDB IDs: ??? and ???), *MFA2* (PDB ID: ???), *AAT2* (PDB ID: ???), and *AMN1* (PDB ID: ???) RNAs will be deposited in the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank prior to publication of the manuscript.

METHODS

HITS-CLIP

CLIP methods were adapted from Wolf et al.[35]. Modifications from the published protocols included disrupting the cells in the presence of liquid nitrogen, grinding with a mortar and pestle, and using 400 µl calmodulin Sepharose beads (GE 17-0529-01) and IgG beads (Life Technologies 11202D). Following isolation of RNA, methods published previously were used to prepare libraries for high-throughput sequencing[36], with the exception of the Illumina TruSeq small RNA adaptor and PCR primer sequences. Detailed methods have been deposited in Protocol Exchange.

Western Blot

50 µl IgG beads were removed from CLIP samples then incubated in 30 µl LDS sample buffer (Life Technologies NP0007). The whole reaction was run on a Novex 6% TBE gel then transferred to PDVF membrane (Millipore IPVH00010). The membrane was probed with TAP Tag Polyclonal Antibody (Pierce:CAB1001) primary antibody followed by goat anti mouse secondary antibody (KPL:074-1506).

Informatic pipeline

Pre-processing. FASTQ files were uploaded to the Galaxy server[37] (cite) and groomed (FASTQ Groomer)[38]. Adaptor sequences were then trimmed using Clip discarding sequences that contained the 5' adaptor or were too short after 3' adaptor clipping. The data were then filtered based on quality score using Filter FASTQ with a minimum length of 15 bases and a

minimum quality score of 20. The 5´ adaptor included a 3´ random bar code that was used to remove PCR duplicates by discarding any read with a perfect duplicate.

Mapping and defining peaks. The filtered reads were mapped to the *S. cerevisiae* genome using Bowtie2[39] (bowtie2 -x /Scgenome -q filename.fastq -S filename.sam -5 5 -N 1 -p 8). The .sam files were used to create .bam and indexed .bam files using samtools for visualization of the data in Artemis Genome browser. Peaks were defined using Pyicoteo[16] (python pyicoclip filename.sam -f filename.pk --region Sc.bed --stranded). The .bed file required for Pyicoteo was downloaded from the Saccharomyces Genome Database (SGD)[40]. Next, the the duplicate peaks were removed from the .pk file. Using the Pyicoteo defined summit, each peak was assigned to a genomic feature using the features table from the SGD. Sequences 200 bases upstream of the ORF and 300 bases downstream of the ORF were used as 5´UTRs and 3´UTRs, respectively and then added to the SGD features table. The number of gapped reads for each peak was defined. Kurtosis was calculated for each peak using the peak profile defined by Pyicoteo. 25 bases of genomic sequence flanking each peak summit was retrieved to define binding elements in two ways. MEME was used as an unbiased search and direct searches were used for known binding elements.

Filtering peaks. The biological replicates were combined into one list based on: 1. Each peak had a summit within 10 bases in both replicates, 2. Each peak contained a gapped read in both replicates, and 3. Each peak had a height greater than 10 reads (third quartile) in both replicates.

Protein purification

The RNA-binding domain of yeast Puf5p (residues 201-600) was subcloned into the pSMX vector with an N-terminal His_6-SUMO tag[41]. *E. coli* cells BL21 Star (DE3) carrying the Puf5p plasmid were grown in Terrific Broth media to OD_{600} = ~0.8 and then protein expression was

induced with 0.4 mM IPTG for 20 hours at 18 °C. The cell pellet was resuspended in lysis buffer containing 20 mM Tris, pH 8.0; 0.5 M NaCl; 20 mM imidazole; 5% (v/v) glycerol and 0.1% (v/v) β-mercaptoethanol and sonicated on ice. The lysate was cleared by centrifugation and loaded onto a Ni-chelating gravity column (Thermo Scientific). His-SUMO-tagged Puf5p was eluted with a buffer containing 20 mM Tris, pH 8.0; 50 mM NaCl; 0.2 M imidazole and 1 mM DTT. Ulp1 protease was added to remove the His_6-SUMO tag, and the protein solution was loaded onto a Hi-Trap heparin column (GE Healthcare) and eluted with a gradient from 0-1 M NaCl in buffer containing 20 mM Tris, pH 8.0, 1 mM DTT. The fractions containing Puf5p were pooled and concentrated by Amicon filters and loaded onto a Superdex 200 16/60 column equilibrated in 20 mM HEPES, pH 7.4; 0.15 M NaCl and 2 mM DTT. The Puf5p peak fractions were pooled and concentrated in column buffer for crystallization and RNA-binding assays.

Protein-RNA crystallization

RNAs were purchased from Thermo Scientific. Puf5p (4 mg/ml) was mixed with each of the four different RNAs at a protein:RNA molar ratio of 1:1.2 and incubated on ice for 1 hour. Crystals were obtained at 20 °C by hanging drop vapor diffusion, mixing 1 µl Puf5p-RNA complex with 1 µl reservoir solution of 15-20% (w/v) PEG 3350 and 0.1 M citrate Bis-Tris propane (CBTP), pH 7.6. Microseeding was performed to grow larger single crystals. Crystals were cryo-protected in crystallization solution supplemented with 15% (v/v) glycerol and flash frozen in liquid nitrogen. For phasing, a Puf5p:*SMX* RNA complex crystal was soaked in 17.5% (w/v) PEG 3350, 0.5 M KI, 0.1 M CBTP, and 15% (v/v) glycerol for 5 minutes and then flash frozen.

X-ray data collection

X-ray data for structures of the 9-nt, 10-nt and 12-nt RNA complexes were collected at the SER-CAT beamline at the Advanced Photon Source, Argonne National Laboratory. Data for the 11-nt RNA complex and the iodide-soaked crystal were collected at the NIEHS in-house facility equipped with a Rigaku 007HF rotating anode generator and a Saturn 92 charge-coupled device (CCD) area detector system. All data were processed using HKL2000[42].

Structure determination

The crystal structure of a Puf5p:*SMX2* RNA complex (space group $P2_12_12$) was determined by combining molecular replacement (MR) with iodide-single-wavelength anomalous diffraction (SAD)) phasing. The Phenix software suite was used throughout the process of structure determination[43]. The anomalous signal of the SAD data extended only to 5.0 Å, and MR or SAD alone failed to solve the structure. A truncated Puf4p structure (PDB: 3BX2) containing repeats 4-8 (residues 684-887) was used as the MR search model. Following MR, AutoSol identified eight iodide sites with FOM of 0.33. Running AutoBuild after MR-SAD phasing produced a model with R_{free} = 44%. The model was further improved to R_{free} = 38% by using the EMBL-Hamburg Auto-Rickshaw web server[44]. Electron density for the 9-nt RNA was clearly visible. Iterative cycles of manual model building in Coot[45] and refinement with Phenix led to the final model with R_{free} = 28% (Supplementary Table 2).

Crystals of the 10-nt, 11-nt and 12-nt RNA complexes and some crystals of the 9-nt *SMX2* RNA complex belonged to space group $P6_122$, although all crystals were grown in the same conditions as the $P2_12_12$ *SMX2* crystals. These structures were determined by MR using the Puf5p coordinates from the initial Puf5p:*SMX2* structure as the search model. Data and refinement statistics are shown in Supplementary Table 2. All models show good geometry according to

MolProbity[46]: 95-98% of the residues are in favored regions of the Ramachandran plot, and there are no outliers.

Electrophoretic mobility shift assays

RNA oligonucleotides were radiolabeled using ^{32}P-γ-ATP and T4 polynucleotide kinase (New England Biolabs) following manufacturer instructions. Serially-diluted Puf5p was mixed with 100 pM labeled RNA in buffer containing 10 mM HEPES, pH 7.4; 50 mM NaCl; 1 mM EDTA; 0.1 mg/ml BSA; 0.01% (v/v) Tween 20 and 0.1 mg/ml yeast tRNA. After overnight incubation at 4 °C, 4 µl loading dye (15% v/v Ficoll 400 and 0.01% bromophenol blue) was added to each 20-µl reaction prior to gel loading. 10% Novex TBE gels (Invitrogen) were run at 100 V at 4 °C for 30 minutes to resolve the samples. The gels were dried and exposed to storage phosphor screens. The screens were scanned using a Molecular Dynamics Typhoon phosphorimaging system (GE Healthcare). The band intensities were analyzed with ImageQuant. K_d values were calculated with GraphPad Prism by fitting the data assuming one-site specific binding and a Hill coefficient of 1. ~93% of Puf5p was active, as determined using the method described in reference[47]. The reported K_d values were not adjusted.

Yeast three-hybrid.assays

Each orthologous PUF RNA-binding domain was cloned into activation domain–protein fusion plasmid, pGADT7[28]. Oligonucleotides representing each RNA sequence were ordered from IDT and cloned into the Hybrid RNA plasmid, p3HR2[28]. All experiments were conducted in the *Saccharomyces cerevisiae* strain YBZ-1 (MATa, ura3-52, leu2-3, -112, his3-200, trp1-1, ade2, LYS2::(LexAop)-HIS3, URA3::(lexAop)-lacZ, and LexA-MS2 MS2 coat (N55K)). Strains were Lithium Acetate transformed with appropriate combinations of plasmids and plated

on synthetic dextrose media lacking uracil and leucine. Single colonies were selected and allowed to grow to stationary phase, then diluted and grown for about 4 hours. Optical density$_{660}$ (OD) for each culture was measured then 50 µl of culture was added to 50 µl of Beta-Glo (Promgea E4720) then incubated for 1 hr in the dark. Luminescence was measured by microplate reader (BioTech Synergy 4). Raw luminescence was normalized to OD_{660}, and each biological replicate (n=3) was then averaged and standard deviation was calculated.

FIGURE LEGENDS

Figure 1. Puf5p HITS-CLIP. a. Summary of HITS-CLIP protocol. (i). *S. cerevisiae* cells were isolated and UV irradiated (254 nm wavelength). (ii) Cell lysate was subjected to gentle RNase A digestion and then (iii) TAP-tagged Puf5p was affinity purified sequentially with calmodulin and IgG resins. (iv) RNA adaptors were ligated to RNA fragments. (v) Libraries were PCR amplified and high-throughput sequenced. b. Western blot of WT and epitope-tagged Puf5p. c. Autoradiogram of ^{32}P-labelled RNA crosslinked to Puf5p. Complexes migrate higher than protein alone. d. Pie chart of Puf5p CLIP peaks found in mRNA regions. e. Reproducibility of peaks from each biological replicate. Normalized $\log_2$ reads/peak for the two experiments are plotted.

Figure 2. Puf5p binding element exhibits flexibility. a. Puf5p interaction peaks map to previously characterized Puf5p binding sites. Orange and green lines represent reads mapped for each replicate. *SMX2* has a peak over the 9 nt binding element. *HO* has a broad peak over two binding elements: a 9 nt lower affinity site and a 8/10 nt higher affinity site. b. Binding elements identified in high-confidence Puf5p target mRNAs. The MEME-derived logo is shown on the left, which was deconvoluted into 5 binding elements of 8-12 nts in length. c. Distribution of binding element lengths for 4 PUF proteins representing 3 species. Results from CLIP (red), SEQRS[25] (light blue), RIP-Chip[6,7] (green), and PAR-CLIP[14] (dark blue) experiments are compared, where available, and shown as enrichment relative to the predominant length for each protein, which is set to 1. The consensus RNA sequence element for each protein is shown, where N is A, C, G or U.

Figure 3. Binding element lengths correlate with biological functions. a. Network representation of Puf5p mRNA targets visualized in Cytoscape 3.0.1[48]. Binding elements were defined as $UGUN_{(x)}UA$ within 25 bp of the peak summit. Large red hubs indicate the length of the binding element. Each node (small circle or square) represents one HITS-CLIP peak in an mRNA. Green square nodes represent mRNAs containing an individual binding element of either 8, 9, 10, 11 or 12 nts; nodes with only one binding element (edge) are placed at the outer periphery of the diagram. Green square nodes with two lines indicate that the HITS-CLIP peak contained two non-overlapping binding elements. Most binding elements were unambiguously of a single length. In a minority of elements, two different lengths of binding elements co-reside in a single sequence. Circles represent these mRNAs with "overlapping" binding elements: for example, "8-10" means a single site of the sequence UGUNNNUAUA, which possesses both 8 and 10 nt elements depending on the 3´UA used, and either sequence may be used *in vivo*, and "8-10, 9-11" means that two distinct overlapping sites are present under the peak. The key to the right is a color code for each combination of overlapping binding element lengths (nt). The numbers of mRNAs containing overlapping binding elements are provided in Supplementary Table 1. b. Gene Ontology term enrichment for mRNAs belonging to each length of binding element using SGD YeastMine[49]. GO terms that are significantly overrepresented in the gene list are bolded for each binding element length. Numbers of genes in the most enriched GO terms are in parentheses.

Figure 4. Crystal structures of Puf5p in complex with representative 9-12 nt target mRNAs. a. Crystal structure of Puf5p in complex with 10-nt *MFA2* RNA. Puf5p is shown as a ribbon diagram with PUM repeats colored alternately blue and red. Two disordered loops are indicated with dotted lines. RNA interacting residues and *MFA2* RNA are shown as stick models

with atoms colored by element (carbon, grey; nitrogen, blue; oxygen, red; sulfur, yellow; phosphorus, orange). b. Schematic diagram of interactions between Puf5p repeats (rectangles) and *MFA2* RNA bases (ovals). Hydrogen bonds are indicated by dotted lines and van der Waals contacts are indicated by >>>>. c-f. Interactions between repeats 3-5 of Puf5p and 9-nt *SMX2* RNA (c), 10-nt *MFA2* RNA (d), 11-nt *AAT2* RNA (e) or 12-nt *AMN1* RNA (f). Hydrogen bonds are indicated with dotted lines. Discontinuous F_o-F_c electron density (contoured at 3σ) following base A5 in the 11-nt *AAT2* RNA is shown in panel e.

Figure 5. Binding element preference across Phylum Ascomycota. a. Phylogeny of Ascomycota fungi. Each subphylum is indicated: budding yeast-yellow, filamentous fungi – green, and the fission yeast – blue. Species used for RNA-binding studies below are highlighted. b-d RNA binding element length preferences for Puf5p, Puf4p, and Puf3p orthologues. RNA binding was assayed using the yeast three–hybrid system[50]. RNAs tested were 8 (UGUAAAUA), 9 (UGUAAAAUA), 10 (UGUAAAAAUA), 11 (UGUAAAAAAUA), or 12 (UGUAAAAAAAUA) nts in length. Raw luminescence values per cell for each biological replicate (n=3) were averaged then normalized to controls where the 5´UGU sequence was mutated to ACA. Error bars represent standard deviation.

Figure 6. Evolution of Puf3p broadened RNA specificity. a. Schematic representations of chimeras tested in b. PUM repeats are represented by circles: *S. cerevisiae*, blue; *S. pombe*, green. b. Broadened *S. pombe* Puf3 specificity is linked to Repeat 6. RNA binding element length preferences for the chimeric proteins were assayed using the yeast 3-hybrid system with 8, 9 or 10 nt RNAs as in Figure 5. Raw luminescence values per cell for each biological replicate (n=3) were averaged then normalized to an acaAAAUA mutant negative control, which depresses binding more than 100-fold[51]. Error bars represent standard deviation.

Supplementary Figure 1. HITS-CLIP data analysis pipeline. a. Flow chart of the data analysis. b. GO analysis of target RNA list using various criteria for filtering peaks using SGD YeastMine[49].

Supplementary Figure 2. Puf5p binds predominantly mRNAs. Distribution of Puf5p HITS-CLIP peaks in RNA types. Numbers of peaks in each category are indicated.

Supplementary Figure 3. Overlap of Puf5p HITS-CLIP targets with Puf3p and Puf4p targets. a. mRNA overlap between Puf5p HITS-CLIP and RIP-Chip[6]. b. mRNA overlap between Puf5p HITS-CLIP and Puf4p RIP-Chip[6]. c. mRNA overlap between Puf5p HITS-CLIP and Puf3p PAR-CLIP[15]. Numbers of mRNAs in each subset are indicated.

Supplementary Figure 4. Nucleotide composition of Puf5p binding elements. a. Enrichment for A's (blue), or U's (red) at the 3´ terminus of each length binding element. b. The frequency of each nucleotide flanking CLIP-defined peaks: A's (green), U's (red), C's (blue), and G's (orange). c. Positional analysis of intervening nucleotides (excluding the fixed 5´-UGUA tetranucleotide sequence and the 3´UA sequences) for each length binding element.

Supplementary Figure 5. Crystal structures of Puf5p in complex with 9-nt *SMX2*, 11-nt *AAT2*, and 12-nt *AMN1* RNAs. HITS-CLIP peaks (a), ribbon drawings (b), and schematic diagrams of protein:RNA interactions (c) are shown.

Supplementary Figure 6. Electrophoretic mobility shift assay (EMSA) of Puf5p. Representative EMSA of Pufp5 with *SMX2* binding element (UGUACUAUA) RNA (top) and data analysis (bottom) are shown. All binding assays were performed in triplicate and the mean K_d and standard error of the mean are reported in Supplementary Table 2.

Supplementary Figure 7. Structural differences in yeast PUF proteins Puf3p, Puf4p, and Puf5p. a. Superposition of PUM repeats in Puf3p, Puf4p, and Puf5p. Cα traces of the eight repeats from each protein are superimposed with human PUM1 repeat 1 (blue). Divergent structures following helix α2 are colored teal (R6), yellow (R7), and orange (R8) for all proteins and green (R3) for Puf5p. b-d. Superposition of RNA-binding helices in Puf3p (yellow), Puf4p (green) and Puf5p (red). Repeats 5-8 of each structure were aligned in the superposition. Cα traces are shown in b, and the α2 helices only are shown as cylinders in c and d. Dotted lines connecting the Cα atoms of base-stacking residues are shown, and ∠R1-2-3 is indicated (d). e. Repeat-to-repeat angles in Puf3p, Puf4p and Puf5p. The angles formed by lines between the Cα atoms of stacking residues in sets of three successive repeats for Puf3p, Puf4p and Puf5p are plotted.

Supplementary Figure 8. Sequence analysis of Puf3p Repeat 6. a. Sequence alignment of Puf3p Repeat 6 from Ascomycota based on a MUSCLE multiple sequence aligment[52]. Positions of α helices in the crystal structure of *S. cerevisiae* Puf3p are indicated above the sequence alignment. Magenta highlighted residues represent the RNA-binding motif in *S. cerevisiae* Repeat 6 that is conserved in all species shown. The blue box indicates the additional residues at the interface of α helices 2 and 3 in Repeat 6. b. Ribbon drawing of the crystal structure of *S. cerevisiae* Puf3p in complex with a *COX17* binding element. RNA-binding motif residues of Repeat 6 are colored magenta and the additional loop residues are colored blue.

Supplementary Table 1 – Number of peaks containing binding elements. Left columns show the numbers of peaks posses only one binding element. The middle and right columns contain the number of peaks that posses non-overlapping or overlapping binding elements, respectively.

Supplementary Table 2. Data collection and refinement statistics.

Supplementary Table 3. Puf5p Electrophoretic Mobility Shift Assays (EMSAs). EMSAs were performed in triplicate and the mean K_d ± standard error of the mean are reported as well as K_d's relative to binding to the 9 nt *SMX2* RNA, which was set to K_{rel}=1.

Supplementary Data Set 1. List of Puf5p target RNAs. Columns indicate properties of each peak or gene information contained in the header. Each row represents a peak.

Supplementary Data Set 2. Cytoscape network file.

FIGURES

Figure 1
Wilinski et al.

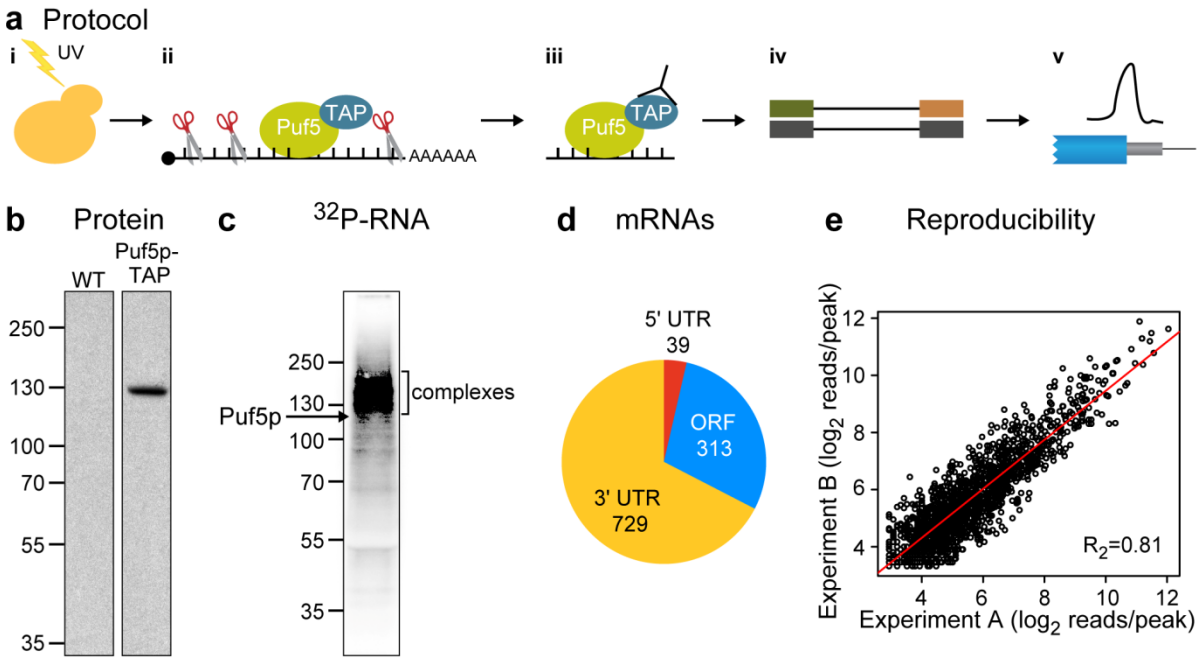


Figure 2
Wilinski et al.

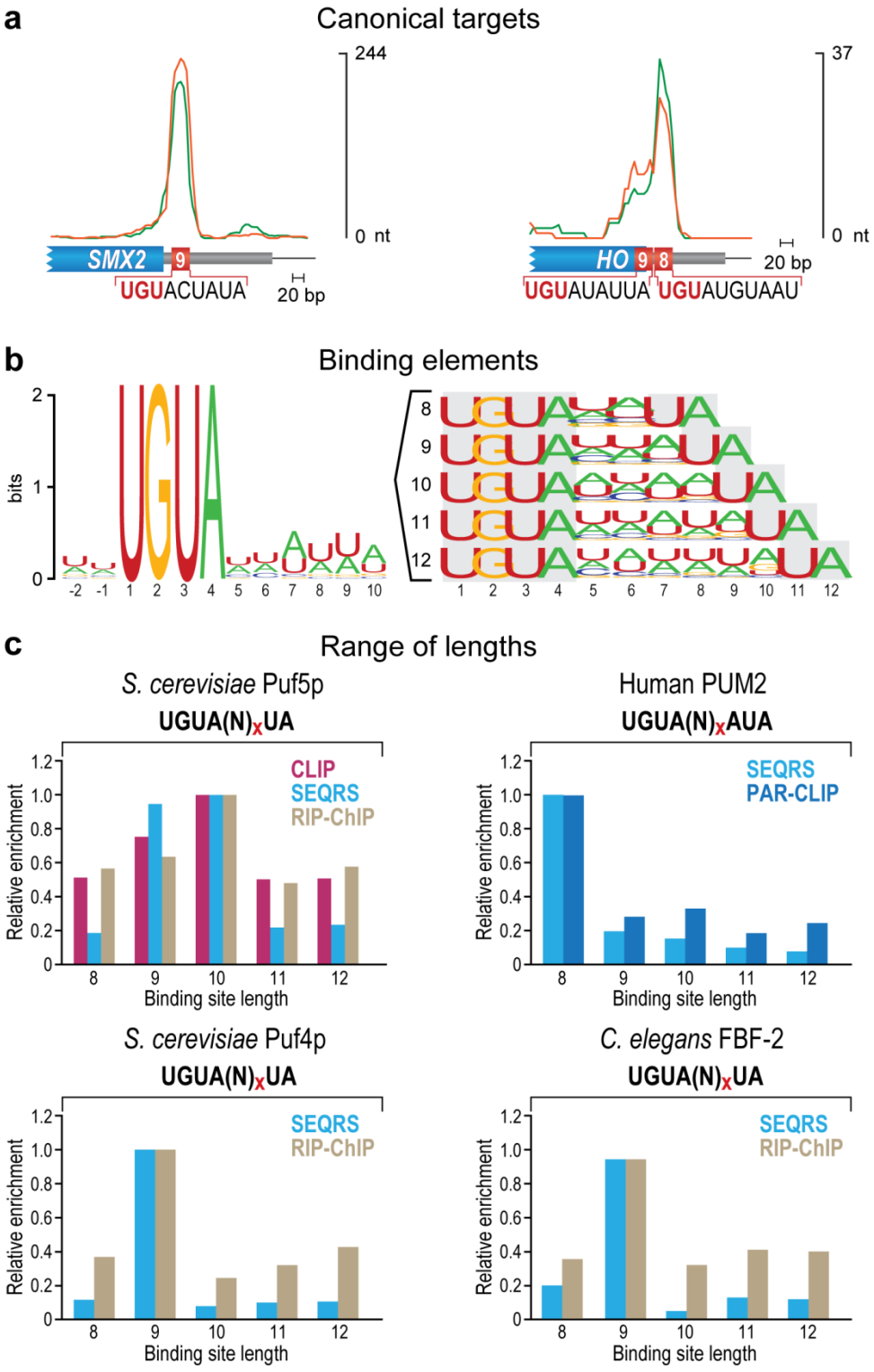
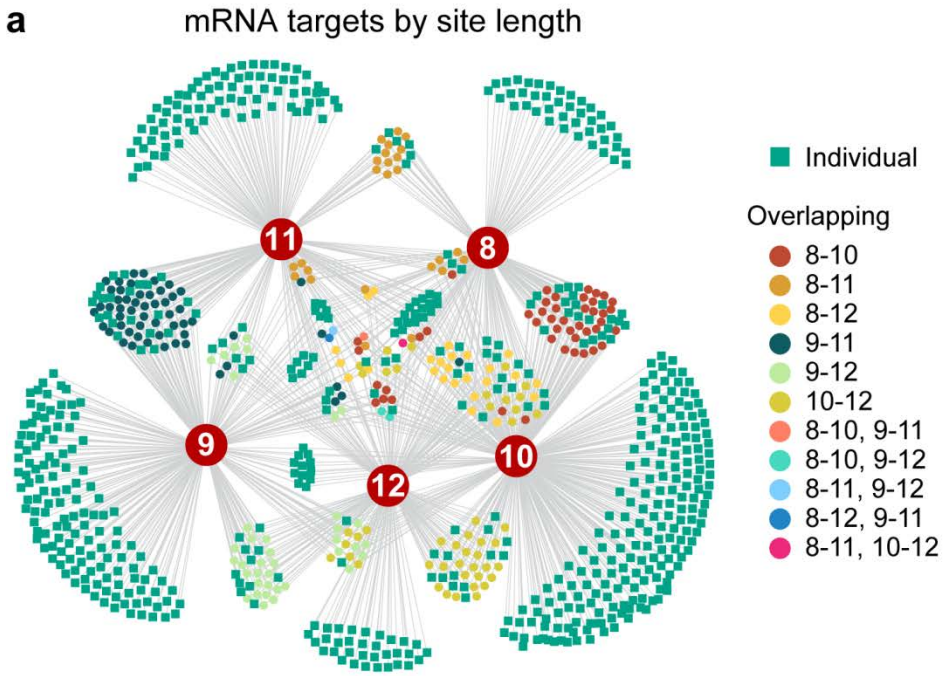


Figure 3
Wilinski et al.



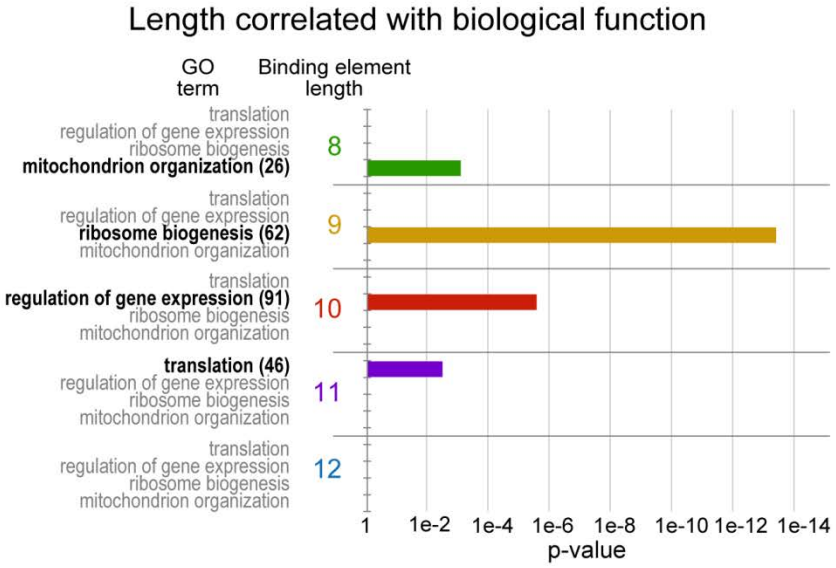


Figure 4
Wilinski et al.

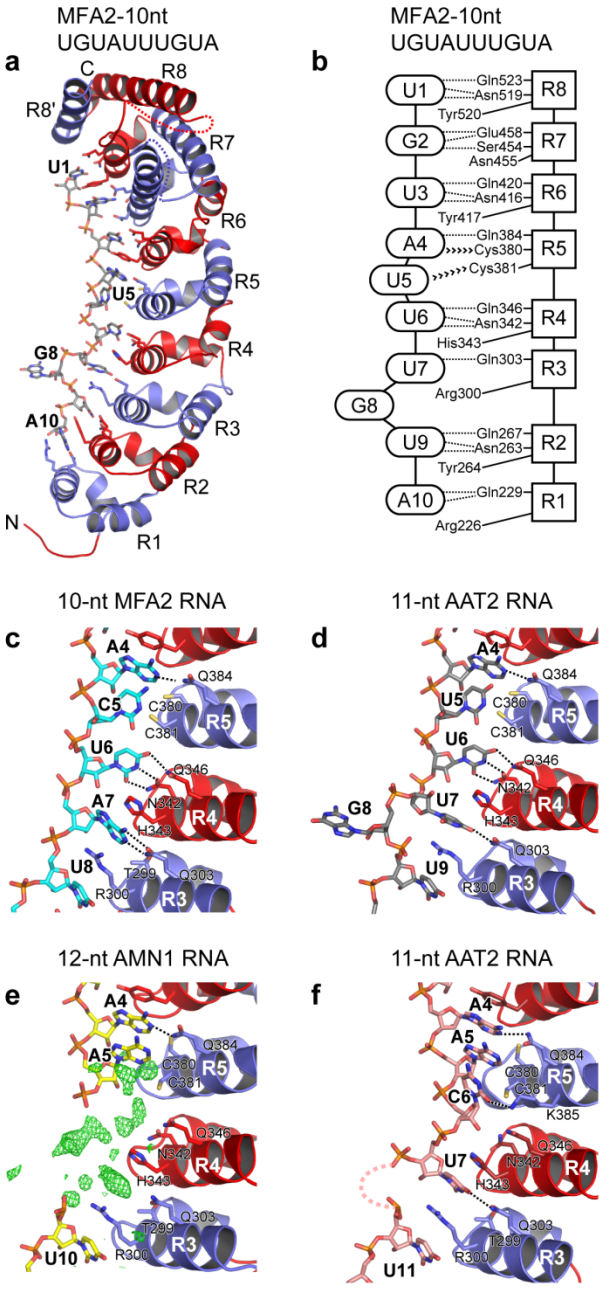


Figure 5
Wilinski et al.

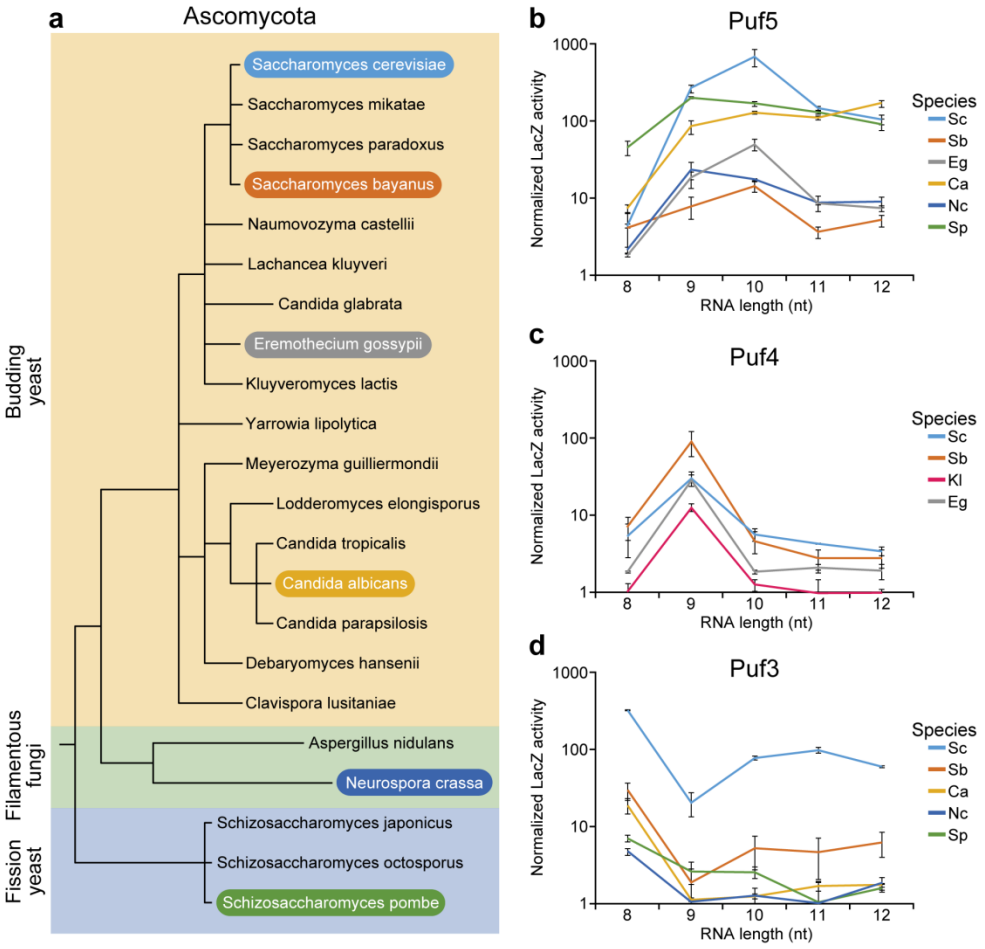
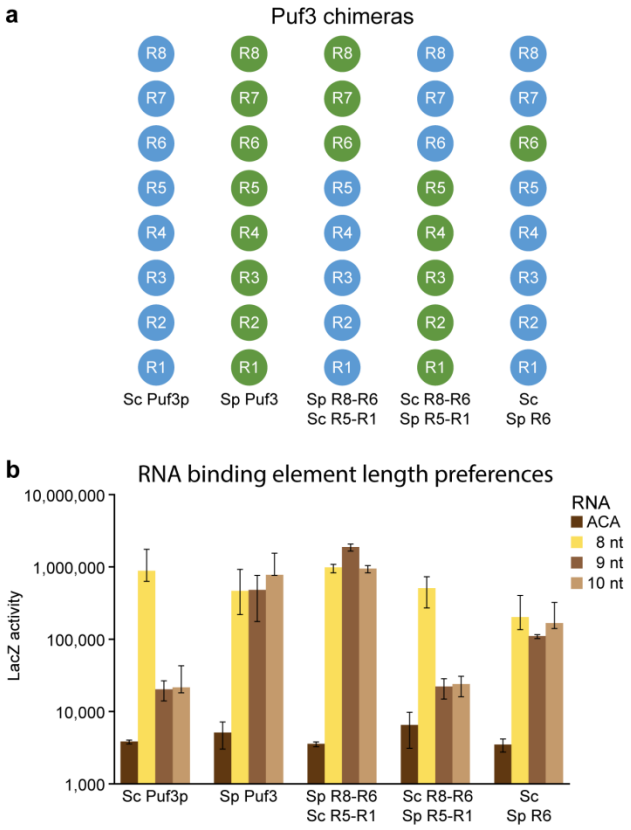
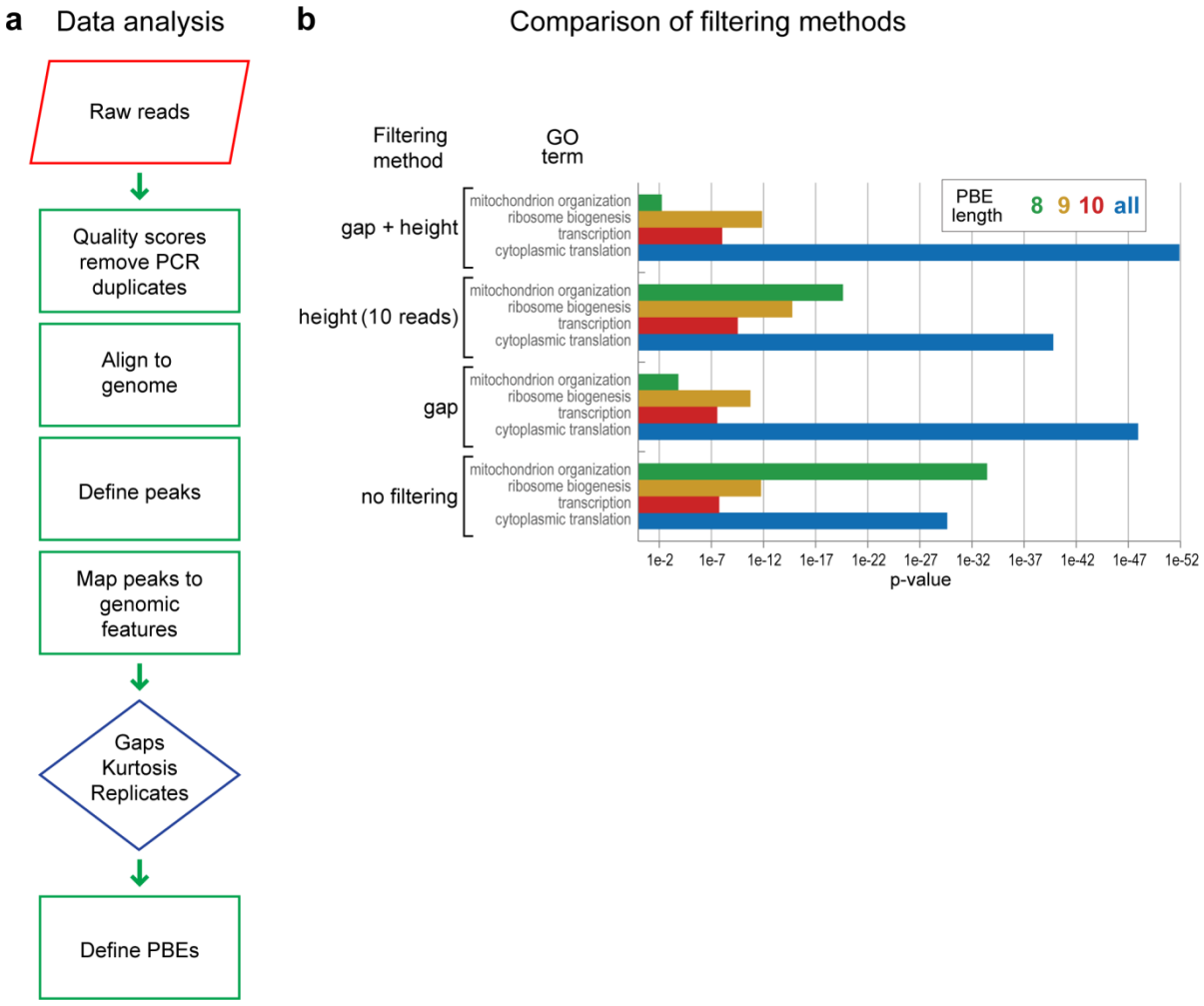


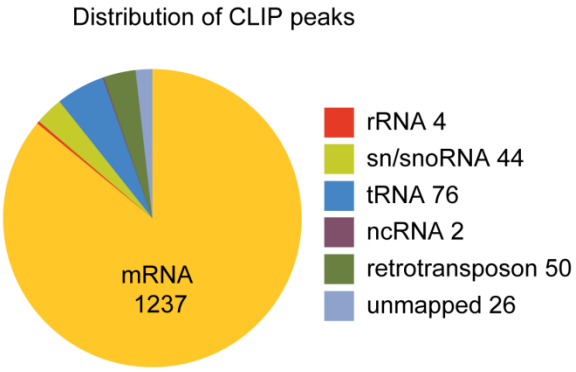
Figure 6
Wilinski et al.



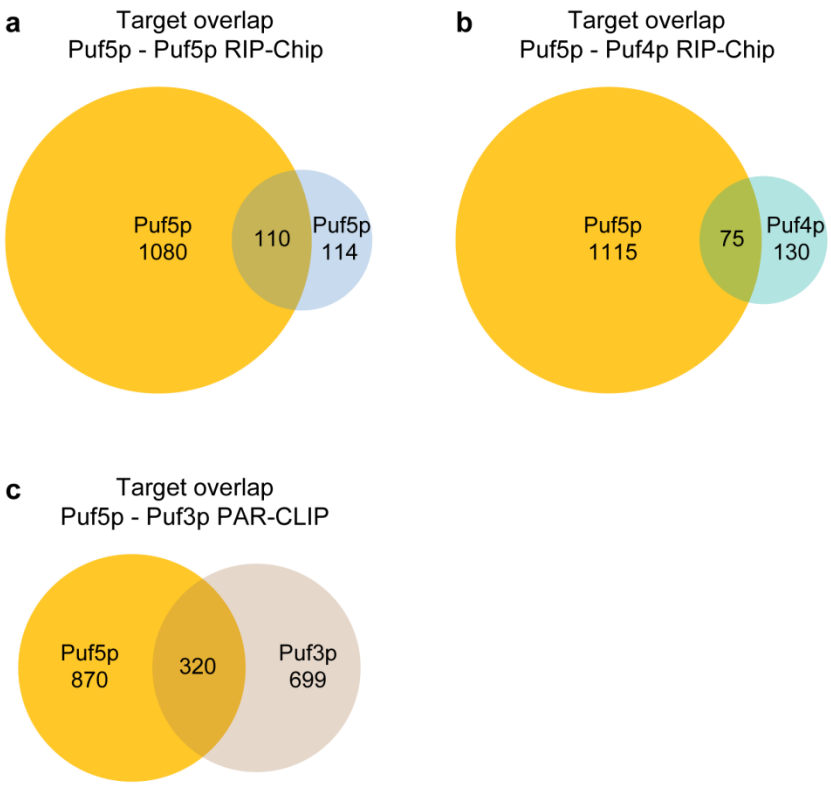
Supp Figure 1
Wilinski et al.



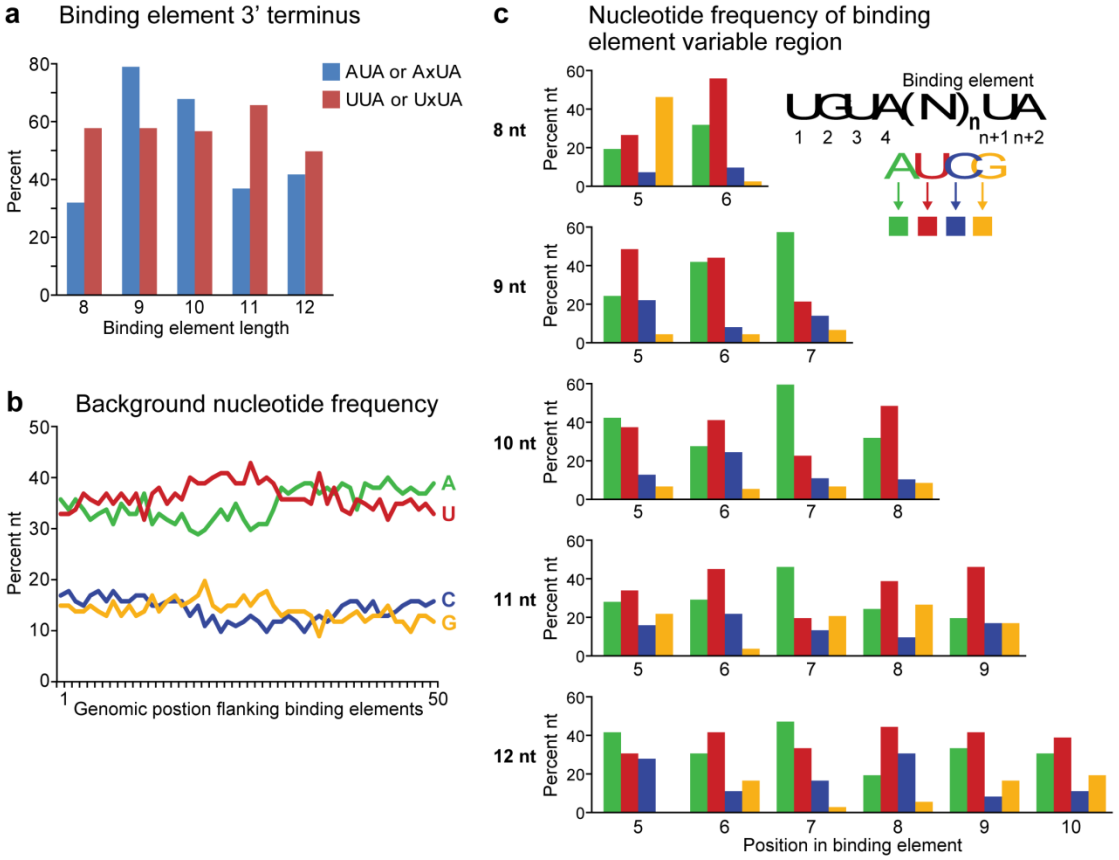
Supp Figure 2
Wilinski et al.



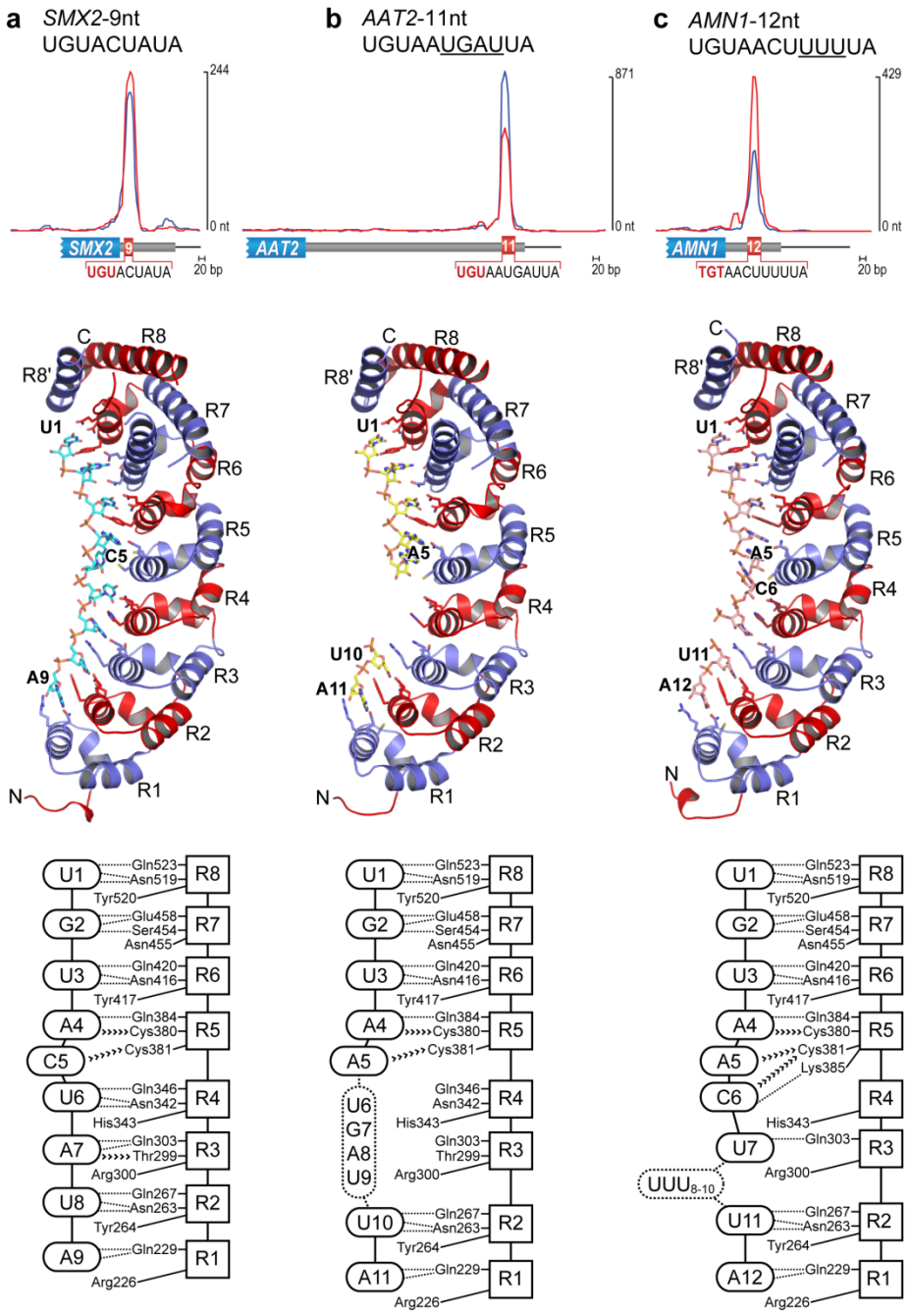
Supp Figure 3
Wilinski et al.



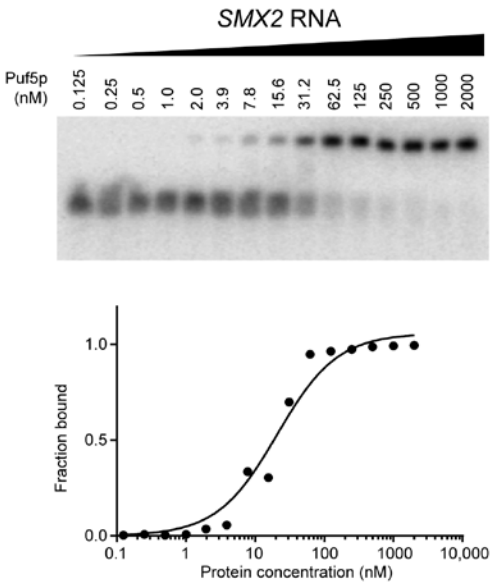
Supp Figure 4
Wilinski et al.



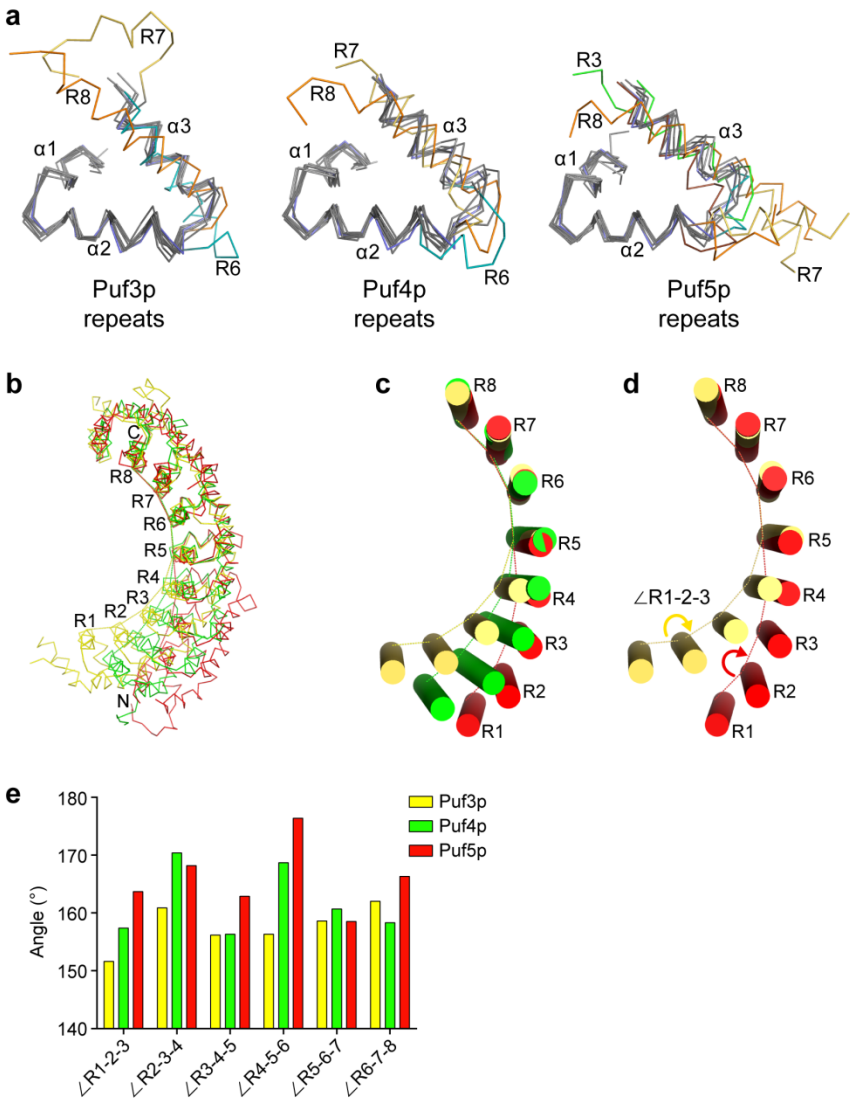
Supp Figure 5
Wilinski et al.



Supp Figure 6
Wilinski et al.



Supp Figure 7
Wilinski et al.



Supp Figure 8
Wilinski et al.

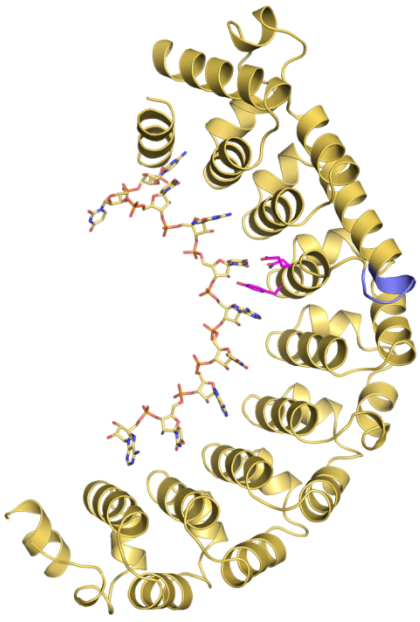
a

Puf3 Repeat 6

```
                          Helix          Helix                    Helix
Scer|YLL013C          IPYLIQDQYGNYVIQYVLQQDQ---FTNKEMVDIKQE--IIETVANN
Smik|smik1210-g1.1    IPYLIQDQYGNYVIQYILQQNQ---FTNKEMVDVKQE--IIETVANN
Spar|spar52-g2.1      IPYLIQDQYGNYVIQYILQQDQ---FTNKEMVDIKQE--IIETVANN
Sbay|sbayc566-g3.1    IPYLIQDQYGNYVIQYILQQDQ---FTNKEMVDVKQE--IVETVADN
Ncas|Scas688.28       IPYLIQDQYGNYVIQHILEQQDNNPNVSQEMMNTKQE--IVNIVSQN
Sklu|SAKL0G13002g     IPYLIQDQYGNYVIQHILQHGG--EHTNIHIGSTKQN--IVDIVSKS
Cgla|CAGL0D05544g     IPYLIQDQYGNYVIQHILQHGSDVNLASEHMRVIKQE--IINNVADN
Egos|AAL152W          IPYLVQDQYGNYVIQHILQHGG-DNPAENHID--KSKQDIVDTISKT
Klac|KLLA0F15477g     IPFLIQDQYGNYVIQHILQHGT-EDTSSHIGMS-KQN--IIDIIRKN
Ylip|YALI0E12001g     AYHLIQDQYGNYVIQHVLEQGA---------PDDKEA--MMLVIKQH
Mgui|PGUG05371.1      IFYLIQDQYGNYVMQHILERGS---------SKDREA--ILEVVLGS
Ctro|CTRG03880.3      LYYLILDQYGNYVIQHILENGT---------PEEKEP--ILEIVLGS
Calb|orf19.1795       LYYLILDQYGNYVIQHILENGT---------QEEKEP--ILEIVLGS
Cpar|CPAG05281        LGFLITHKFGNYVIQACLENQ----------LREQD--IFTTVVCK
Dhan|DEHA2C07128g     IFYLIQDQYGNYVMQHTLERGN---------PEDREE--ILKIVLGS
Clus|CLUG04502        LFYLIQDQYGNYVIQHILERGT---------PSEKEE--IFEVAFSS
Anid|AN6587           APRLIEDQYGNYVIQHIIQSGE---------EEDRSF--MIEMVKQK
Ncra|NCU06511         AHTLITDAYGNYVAQHIIEAGK---------PEDRAR--MIAAVMSQ
Sjap|SJAG00686        ILHLAQDQYGNYVIQHLMKKGS---------PSEQRE--IVEVVLGN
Soct|SOCG02114        SLQLTQGQYGNYVVQHILKEGS---------EKDKKF--VFNLIAKN
Spom|SPAC1687.22c     ILKLTQDQYGNYVVQHILRTGS---------ESDKKY--IFDLMIDH
```

b

Sc Puf3p



REFERENCES

1. Keene, J.D. RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* **8**, 533-43 (2007).
2. Richter, J.D. CPEB: a life in translation. *Trends Biochem Sci* **32**, 279-85 (2007).
3. Darnell, R.B. RNA protein interaction in neurons. *Annu Rev Neurosci* **36**, 243-70 (2013).
4. Quenault, T., Lithgow, T. & Traven, A. PUF proteins: repression, activation and mRNA localization. *Trends Cell Biol* **21**, 104-12 (2011).
5. Wickens, M., Bernstein, D.S., Kimble, J. & Parker, R. A PUF family portrait: 3'UTR regulation as a way of life. *Trends Genet* **18**, 150-7 (2002).
6. Gerber, A.P., Herschlag, D. & Brown, P.O. Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS Biol* **2**, E79 (2004).
7. Kershner, A.M. & Kimble, J. Genome-wide analysis of mRNA targets for Caenorhabditis elegans FBF, a conserved stem cell regulator. *Proc Natl Acad Sci U S A* **107**, 3936-41 (2010).
8. Campbell, Z.T., Valley, C.T. & Wickens, M. A protein-RNA specificity code enables targeted activation of an endogenous human transcript. *Nat Struct Mol Biol* **21**, 732-8 (2014).
9. Hall, T.M. Expanding the RNA-recognition code of PUF proteins. *Nat Struct Mol Biol* **21**, 653-5 (2014).
10. Kennedy, B.K. et al. Redistribution of silencing proteins from telomeres to the nucleolus is associated with extension of life span in S. cerevisiae. *Cell* **89**, 381-91 (1997).
11. Kaeberlein, M. & Guarente, L. Saccharomyces cerevisiae MPT5 and SSD1 function in parallel pathways to promote cell wall integrity. *Genetics* **160**, 83-95 (2002).
12. Tadauchi, T., Matsumoto, K., Herskowitz, I. & Irie, K. Post-transcriptional regulation through the HO 3'-UTR by Mpt5, a yeast homolog of Pumilio and FBF. *EMBO J* **20**, 552-61 (2001).
13. Ule, J. et al. CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212-5 (2003).

14. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-41 (2010).
15. Freeberg, M.A. et al. Pervasive and dynamic protein binding sites of the mRNA transcriptome in Saccharomyces cerevisiae. *Genome Biol* **14**, R13 (2013).
16. Althammer, S., Gonzalez-Vallinas, J., Ballare, C., Beato, M. & Eyras, E. Pyicos: a versatile toolkit for the analysis of high-throughput sequencing data. *Bioinformatics* **27**, 3333-40 (2011).
17. Zhang, C. & Darnell, R.B. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat Biotechnol* **29**, 607-14 (2011).
18. Seay, D., Hook, B., Evans, K. & Wickens, M. A three-hybrid screen identifies mRNAs controlled by a regulatory protein. *RNA* **12**, 1594-600 (2006).
19. Prinz, S., Aldridge, C., Ramsey, S.A., Taylor, R.J. & Galitski, T. Control of signaling in a MAP-kinase pathway by an RNA-binding protein. *PLoS One* **2**, e249 (2007).
20. Valley, C.T. et al. Patterns and plasticity in RNA-protein interactions enable recruitment of multiple proteins through a single site. *Proc Natl Acad Sci U S A* **109**, 6054-9 (2012).
21. Chritton, J.J. & Wickens, M. Translational repression by PUF proteins in vitro. *RNA* **16**, 1217-25 (2010).
22. Bailey, T.L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**, 28-36 (1994).
23. Licatalosi, D.D. et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464-9 (2008).
24. Tenenbaum, S.A., Carson, C.C., Lager, P.J. & Keene, J.D. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc Natl Acad Sci U S A* **97**, 14085-90 (2000).
25. Campbell, Z.T. et al. Cooperativity in RNA-protein interactions: global analysis of RNA binding specificity. *Cell Rep* **1**, 570-81 (2012).
26. Zhu, D., Stumpf, C.R., Krahn, J.M., Wickens, M. & Hall, T.M. A 5' cytosine binding pocket in Puf3p specifies regulation of mitochondrial mRNAs. *Proc Natl Acad Sci U S A* **106**, 20192-7 (2009).
27. Wapinski, I., Pfeffer, A., Friedman, N. & Regev, A. Natural history and evolutionary principles of gene duplication in fungi. *Nature* **449**, 54-61 (2007).

28. Koh, Y.Y. & Wickens, M. Determining the RNA specificity and targets of RNA-binding proteins using a three-hybrid system. *Methods enzymol* **539**, 163-81 (2014).
29. Hook, B., Bernstein, D., Zhang, B. & Wickens, M. RNA-protein interactions in the yeast three-hybrid system: affinity, sensitivity, and enhanced library screening. *RNA* **11**, 227-33 (2005).
30. Ulbricht, R.J. & Olivas, W.M. Puf1p acts in combination with other yeast Puf proteins to control mRNA stability. *RNA* **14**, 246-62 (2008).
31. Saint-Georges, Y. et al. Yeast mitochondrial biogenesis: a role for the PUF RNA-binding protein Puf3p in mRNA localization. *PLoS One* **3**, e2293 (2008).
32. Kato, M. et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753-67 (2012).
33. Reichen, C., Hansen, S. & Pluckthun, A. Modular peptide binding: from a comparison of natural binders to designed armadillo repeat proteins. *J Struct Biol* **185**, 147-62 (2014).
34. Chen, T. & Kurjan, J. Saccharomyces cerevisiae Mpt5p interacts with Sst2p and plays roles in pheromone sensitivity and recovery from pheromone arrest. *Mol Cell Biol* **17**, 3429-39 (1997).
35. Wolf, J.J. et al. Feed-forward regulation of a cell fate determinant by an RNA-binding protein generates asymmetry in yeast. *Genetics* **185**, 513-22 (2010).
36. Green, M.R. & Sambrook, J. *Molecular cloning: a laboratory manual*, (Cold Spring Harbor Laboratory Press New York, 2012).
37. Goecks, J., Nekrutenko, A. & Taylor, J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **11**, R86 (2010).
38. Blankenberg, D. et al. Manipulation of FASTQ data with Galaxy. *Bioinformatics* **26**, 1783-5 (2010).
39. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-9 (2012).
40. Cherry, J.M. et al. Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* **40**, D700-5 (2012).

41. Mossessova, E. & Lima, C.D. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol Cell* **5**, 865-76 (2000).
42. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data. *Methods enzymol* **276**, 307-326 (1997).
43. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).
44. Panjikar, S., Parthasarathy, V., Lamzin, V.S., Weiss, M.S. & Tucker, P.A. Auto-rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr D Biol Crystallogr* **61**, 449-57 (2005).
45. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).
46. Davis, I.W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375-83 (2007).
47. Cheong, C.G. & Hall, T.M. Engineering RNA sequence specificity of Pumilio repeats. *Proc Natl Acad Sci U S A* **103**, 13635-9 (2006).
48. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-504 (2003).
49. Balakrishnan, R. et al. YeastMine--an integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose tool-kit. *Database (Oxford)* **2012**, bar062 (2012).
50. SenGupta, D.J. et al. A three-hybrid system to detect RNA-protein interactions in vivo. *Proc Natl Acad Sci U S A* **93**, 8496-501 (1996).
51. Goldstrohm, A.C., Seay, D.J., Hook, B.A. & Wickens, M. PUF protein-mediated deadenylation is catalyzed by Ccr4p. *J Biol Chem* **282**, 109-14 (2007).
52. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-7 (2004).

Chapter 3

Rewiring a fungal RNA regulatory network through the evolution of binding elements

This work will be submitted for publication: Wilinski D, Buter N, Wickens M. Rewiring a fungal RNA regulatory network through the evolution of binding elements

This work will be submitted after the completion of analysis and perhaps a final experiment.
I contributed to all of this work and wrote the manuscript. Natascha Buter preformed the immunoprecipitation of Neurospora crassa Puf3 and Puf5. Marvin Wickens oversaw the work.

ABSTRACT

Proteins bind and control mRNAs to, direct their localization, translation, and stability. PUF RNA-bining proteins control circuits of mRNAs, and play key roles in development, stem cell maintenance, and memory formation. Here we look at the conservation of a collection of PUF-RNA circuits. Through a combiniation of bioinformatic analysises and *in vivo* assays,
we demonstrate conservation of three circuits in budding yeast and elucidate a rewiring of a nulcear-endcoded circuit of mitochondrial related mRNAs.

INTRODUCTION

Organisms devote a substantial fraction of their genomes to the regulation of protein expression. The regulatory sequences often found in the untranslated regions of the genome allow for the coordinated timing, subcelluar localization, and abundance of proteins. These processes are tightly regulated through many mechanisms including transcriptional and RNA control. The evolution of protein control is an import source of phenotypic variations between species. Homeobox transcription factors, hox genes, are a dramatic example of how changes in protein abundance or localization of proteins leads to phenotypic variation (Pick and Heffer, 2012). Hox proteins specify anterior-posterior segment identity in the embryo and are largely functionally conserved in metazoan, but changes over time in their expression and targets contribute to the vast array of body plans observed in nature (Mallo and Alonso, 2013; Pick and Heffer, 2012).

To understand the molecular events that underlie the evolution of protein regulation, we investigated a RNA-PUF protein network in a large collection of fungal species. PUF proteins are sequence specific RNA binding proteins which typically repress translation of target RNAs (Wickens et al., 2002). They have been found in all Eukaryotes and play roles the maintenance of stem cells and memory formation (Dubnau et al., 2003; Kershner and Kimble, 2010). Many of the RNA targets for the canonical Saccharomyces cerevisiae PUF proteins have been identified by RIP-chip (Gerber et al., 2004). These data show that each PUF protein binds a group of RNAs, or circuit, that are enriched for a molecular function. For example, of the 220 RNAs in the Puf3p circuit 131 of them encode proteins that contribute to mitochondrial biogenesis. In addition, previous work demonstrated the conservation of putative PUF binding

elements in some functional groups such as mitochondrion for Puf3p binding elements (Gasch et al., 2004; Jiang et al., 2010; Jiang et al., 2012). To the best of our knowledge PUF proteins from other fungal species have not been examined biochemically.

We considered species from three subphylum of Phylum Ascomycoata: Saccharomycotina (budding yeast), Pezizomycotina (filamentous fungi), and Taphrinomycotina (fission yeast) (Sugiyama et al., 2006) (Figure 1A). The phylum represents at least 400 million years of evolution, and provides a wealth of genomic sequences and functional information (Taylor and Berbee, 2006).

Two general mechanisms may be responsible for the evolution of RNA-protein networks. In the first, RNA binding elements may change over time allowing for the regulation of new proteins. Alternatively, the RNA binding protein may evolve new sequence specificity to select new targets. Both of these mechanisms have been observed in transcription factor evolution (Li and Johnson, 2010). Intermediate states also likely exist to allow the transition from the ancestral state to the derived state.

We define putative RNA-PUF protein circuits across Ascomycota fungi, define targets and binding preferences for 3 PUF proteins, and elucidate a rewired RNA-PUF circuit.

RESULTS

Binding element conservation

To probe the conservation of an RNA-protein network, we identified putative PUF protein binding elements in 3'UTRs of orthologous RNAs in Ascomycota fungi (Figure 1A). We generated position weight matrixes (PWM) of 8, 9, and10nt in length to model PUF binding elements, which were based upon previous work (Figure 1B) (Gerber et al., 2004). Each PWM model was used as parameters for the log likelihood function to determine the probability of each k-mer being a binding element. For each binding element length, the highest log-likelihood score for each gene was retained then all orthologous genes were clustered and visualized by heat map (Figure 1C-1E).

We observed two modes of binding element conservation. First, we observed conservation of binding elements in clusters of orthologous genes across all budding yeast, which likely represents functional conservation (same function—same genes). For example, there is strong conservation of the putative 8nt binding elements in a cluster of genes in all budding yeast, which is depicted by a bright yellow cluster (Figure 1C). This cluster was significantly enriched for nuclear-encoded genes annotated as mitochondrion organization (e-value 1.2 E-178). We will refer to this cluster of genes below as the "mitochondrial cluster".

Second, we observed dynamic clusters of functional conservation where gene identity sometimes changed, yet the functional conservation was maintained (same function—different genes). This mode of conservation is exemplified by the 9nt and 10nt putative binding elements (Figure 1C-1D). For the 9nt putative binding elements, two clusters are significantly enriched for ribosomal biogenesis (p-value 1.0 E-35 and 5.4E-8), and exhibit different patterns of putative

binding element conservation (Figure 1D). For the top highlighted cluster, the highest likelihood values (bright yellow) are from the Candida clade, and for the lower highlighted cluster, the highest likelihood values are found in the Saccharomyces clade (Figure 1D). Likewise, putative10nt binding element conservation resolves into two clusters, and both are enriched for chromatin related genes (Figure 1E). Most other clusters for each of the binding element lengths did not have obvious patterns of conversation or enriched gene ontology terms (See Supplementary Table 1 for a complete list of GO enrichments for each cluster.) One exception to this, is the top highlighted cluster of putative 9nt binding elements, which was enriched for cytoplasmic translation (p-value 1.3E-33) in addition to ribosome biogenesis (Figure 1D). This suggests PUF-RNA regulation is tied to the broad function of a network rather than specific genes.

Previous work demonstrated *S. cerevisiae* Puf3p targets are significantly enriched for nuclear-encoded mitochondrial RNAs containing 8nt binding elements (Gerber et al., 2004). In addition, *S. cerevisiae* Puf4p and Puf5p targets are significantly enriched for RNAs related to ribosome biogenesis and chromatin, respectively (Gerber et al., 2004). These observations suggest that our bioinformatic analyses are elucidating biologically relevant principles.

Filamentous mitochondrial RNAs contain conserved putative binding elements

Conspicuously absent from our bioinformatic results are clusters of putative binding elements in either the filamentous fungi or fission yeast (Figure 1C-1E). One possible explanation for this is that we over-sampled the budding yeast in our clustering analysis. To address this directly, we re-clustered the data after removing the budding yeast. Interestingly, filamentous fungi and fission yeast still lacked significantly conserved clusters (Figure S1). An

alternative and more attractive explanation for the lack of enriched clusters could be that the binding elements in functionally related genes have completely diverged from budding yeast.

To probe the divergence of binding elements in filamentous fungi, we focused on the mitochondrial cluster because it was the most prominent cluster from our putative binding element analysis. Most orthologous genes from budding yeast in the mitochondrial cluster contain an 8nt binding element while very few putative 8nt binding elements are present in the same genes from filamentous fungi (Figure 2A). However, when we probed the 3'UTRs of genes in the mitochondrial cluster directly for 9nt and 10nt putative PUF binding elements, we found 35% (80/228) of *Aspergillus nidulans* and 52% (119/228) of *Neurospora crassa* orthologous genes contain either a 9nt or 10nt PUF binding element (Figure 2B). No other cluster exhibited such large percentage of putative binding elements (data not shown). When we refined our list to consider only genes that had orthologs in all three species (*S. cerevisiae, A. nidulans*, and *N. crassa*), 81% (p-value 1.6 E-26, hypergeometric distribution test) of *A. nidulans* and 80% (p-value 2.2E-46, hypergeometric distribution test) of *N. crassa* genes contained putative binding elements (Figure 2C). Importantly, these results were supported by an unbiased sequence pattern search (Bailey and Elkan, 1994). This search identified 9 and 10 nt binding elements in *A. nidulans* and *N. crassa* genes from the mitochondrial cluster (Figure 2D). Thus, the putative PUF binding element conservation in the mitochondrial cluster appears to extend to filamentous fungi, through the use of different length PUF binding elements.

Putative binding elements are evolving (Transition from Puf3 to Puf5 control-PUF handoff)

We wondered why the two filamentous fungi motifs were different lengths (Figure 2D). To address this, we performed a phylogenetic analysis of Ascomycota PUF proteins and found that *A. nidulans* has a single *puf4* gene no *puf5*, while *N. crassa* has one *puf5* gene and no *puf4* (each also have one *puf3*) (Figure 2E). This suggests that the difference we observe in putative binding element length is due to the identity of the PUF protein, since *S. cerevisiae* Puf4p binds a 9nt binding element and *S. cerevisiae* Puf5p binds a 9 or 10nt binding elements (Figure 2D). Alternatively, the PUF specificity may have evolved to accept different length binding elements.

To test the binding element length preference of the filamentous PUF proteins, we assayed the proteins in the yeast three-hybrid assay against a battery of RNAs ranging in length from 8-10nt. *A. nidulans* PUF4 bound the best to the shorter 9nt binding element and *N. crassa* PUF5 bound 9nt and 10nt long binding elements similarly (Figure 2F).

We conclude that *A. nidulans* and *N. crassa* putative binding elements in orthologous mitochondrial RNAs have evolved from 8nt to 9nt or 10nt constituting a switch and are likely bound by *A. nidulans* PUF4 and *N. crassa* PUF5.

Identification of PUF RNA targets

Based on our bioinformatic analyses and yeast three-hybrid results, we hypothesized that mitochondrial genes are bound by *N. crassa* PUF5, not *N. crassa* PUF3, which would represent rewiring of the RNA-PUF regulatory circuit. To test this, we determined the *in vivo* targets and binding preference of *S. cerevisiae* Puf3p, *N. crassa* PUF3, and *N. crassa* PUF5 by UV crosslinking and high-throughput sequencing (HITS-CLIP) (Licatalosi et al., 2008). We obtained data from three biological replicates for each endogenously tagged protein (Figure 3A). Table S2 is a summary of the sequencing depth and mapping statistics. The HITS-CLIP data

were very reproducible between biological replicates (Pearson's r values: *S. cerevisiae* Puf3p > 0.77, *N. crassa* PUF3 > 0.99, and N crassa PUF5 > 0.67) (Figure 3A and Table S2). We identified 503, 670, and 1044 CLIP peaks for *S. cerevisiae* Puf3p, *N. crassa* PUF3, and *N. crassa* PUF5, respectively.

The peaks, predominately found in the 3'UTRs of mRNAs, represent 465, 803, and 686 genes for ScPUF3, NcPUF3, and NcPUF5, respectively (Figure S1). Figure 3B shows representative CLIP peaks for each protein. *cox17*, a characterized *S. cerevisiae* Puf3p target and predicted *N. crassa* PUF5 target, is bound by both *S. cerevisiae* Puf3p and *N. crassa* PUF5, but not *N. crassa* PUF3 (Figure 3B). In both cases, there are two peaks centered over binding elements. The two binding elements for *S. cerevisiae* Puf3 are 8nt and 9/10nt for *N. crassa* PUF5. An example of a *N. crassa* PUF3 target is the light induced gene *albino-2 (al-2*) which has an 8nt binding element under the peak (Figure 3B).

To assess whether targets were co-shared between *S. cerevisiae* Puf3p and *N. crassa* PUFs, we determined the overlapping genes for all three proteins (Figure 3C). The overlap between *S. cerevisiae* Puf3p and *N. crassa* PUF5 was more significant (6.1E-41, hypergeometric distribution test) than the overlap between *S. cerevisiae* Puf3p and *N. crassa* PUF3 (5.0E-3, hypergeometric distribution test) (Figure 3C). We conclude that *S. cerevisiae* Puf3p and *N. crassa* PUF5 proteins bind a more similar group of RNAs than the two orthologous Puf3 proteins.

in vivo PUF binding elements

To determine the sequence each PUF protein binds, we searched for enriched sequences within each HITS-CLIP peak (Bailey and Elkan, 1994). *S. cerevisiae* Puf3p peaks were enriched

for an 8nt binding element comprised of a 5’ UGU and a 3’ AUA (Figure 4A). In addition to the core PUF binding element, ~50% of the sequences for *S. cerevisiae* Puf3 contain a cytosine at the minus 2 position (Figure 4A). The enriched sequence for *S. cerevisiae* Puf5p is longer and more degenerate (Figure 4B) (Wilinski et al., 2015). The 5’ end contains the canonical UGUA, but the 3’ end is a mix of uridines and adenines (Figure 4B). We have previously shown the degeneracy in the 3’ end of the *S. cerevisiae* Puf5p element is due to the offset UA’s of various length binding element lengths (Wilinski et al., 2015).

The binding preferences of the two *N. crassa* PUFs were strikingly similarity to their orthologous *S. cerevisiae* PUF proteins (Figure 4C-4D). *N. crassa* PUF3 peaks were enriched for an 8nt binding element, but lacked the -2C enrichment (Figure 4C). The *N. crassa* PUF5 enriched sequence element had a prominent UGUA at the 5’ end and a degenerate 3’ end (Figure 4D), which agrees well with our bioinformatic predictions (Figure 4D and 2D). To see if the conservation of binding element preference extends outside of fungi, we examined par-CLIP data for the orthologous human Puf3 protein, PUM2 (Hafner et al., 2010). Indeed, human PUM2 peaks were also enriched for 8nt binding elements (Figure 4E). We conclude that the in vivo binding preferences of orthologous PUF proteins are highly conserved.

Mitochondrial target RNAs

To determine the functional enrichment for each protein’s targets, we performed Gene Ontology analyses on the targets of the each PUF protein. As expected, there was a strong enrichment for nuclear-encoded mitochondrial genes for *S. cerevisiae* Puf3p targets (p-value 2E-110) (Figure 5A). To our surprise, *N. crassa* PUF3 and PUF5 targets were enriched for nuclear-encoded mitochondrial genes (p-value Puf3 1E-5, Puf5 1E-25), albeit to different degrees (Figure

5A). Our Gene Ontology analysis did not suggest another strong candidate function for the regulation of *N. crassa* PUF3 targets.

This mitochondrial enrichment for both *N. crassa* PUF proteins led us to more closely examine the overlap between each PUF protein for genes in the mitochondrial cluster. Of the genes in the mitochondrial cluster, 77% (188/244), 14% (34/244), and 59% (143/244) are *S. cerevisiae* Puf3p, *N. crassa* PUF3 and *N. crassa* PUF5 targets, respectively. The overlap between *N. crassa* PUF3 and PUF5 is 88% (30/34). We suspect this overlap represents a hand-off from one *N. crassa* PUF protein to another PUF protein of the conserved PUF regulated mitochondrial genes.

DISCUSSION

We identified a conserved PUF regulatory network structure consisting of: circuits that retain chromatin and ribosomal functional conservation, but have changed the identity of genes; and a circuit of mitochondrial biogenesis genes that have retained coordinated RNA control across at least 400 million years of evolution (Taylor and Berbee, 2006).

PUF proteins retain binding preference

A striking feature of all PUF regulated circuits is the conservation of broad functional enrichment. Unlike some transcription factor networks we do not observe great differences in binding element preference between orthologous PUF proteins across species (Li and Johnson, 2010; Villar et al., 2014; Wohlbach et al., 2009). Puf3 is an extreme example of binding element preference conservation. Every Puf3 ortholog studied to date, from yeast to humans, bind a very similar 8nt binding element (Figure 4A, C, E) (Gerber et al., 2006; Hafner et al., 2010). An exception of to this rule is the additional base at the -2 position in *S. cerevisiae* Puf3 preferred binding elements (Gerber et al., 2004; Zhu et al., 2009). It will be interesting to see in the future if RNA binding protein binding element preference is more fixed than rapidly evolving mammalian transcription factor preference (Villar et al., 2014).

A molecular rewiring

A significant number of binding elements in the 3'UTRs of genes in the mitochondrial cluster are conserved in all the budding yeast and filamentous fungi species we examined. We favor the model where Puf3 was the ancestral regulator of mitochondrial genes (Figure 6A).

Based on parsimony neither model, Puf3 or Puf4/5 ancestral regulation, can be ruled out as both require two changes (see numbered grey circles Figure 6A-B).

The alternative Puf4/5 ancestral model presents more caveats in terms of the timing of divergence (Figure 6B). First, the switch from Puf4/5 to Puf3 likely occurred at the root of budding yeast because we observe complete penetrance of 8nt putative binding elements (Figure 2A). Second, the switch from Puf4 to Puf5 in *N. crassa* happened (or vice versa for Puf5 to Puf4 in *A. nidulans*) before the loss of Puf4 and after the split between the *N. crassa* and A. nidulians lineages (Figure 6B). On the other hand, the divergence from Puf3 to Puf4/5 regulation in the filamentous linage could have occurred before or after the split between *A. nidulans* and *N. crassa* and would not require the switch to occur before the loss of Puf4/5 paralogs (Figure 6A). Therefore, we reason that the simpler Puf3 ancestral regulator model is more likely.

The switch from Puf3 to Puf4/5 regulation occurred through the lengthening of binding elements. The nearly complete overlap between mitochondrial cluster target genes we observe for *N. crassa* PUF3 and PUF5 likely represents a transitional state of the circuit.

There are two ways the circuit could evolve to accommodate the modern regulation. First, the 8nt binding element is followed by an additional UA dinucleotide allowing both Puf3 and Puf5 to bind the same site. Second, an additional binding element could emerge to accommodate both PUFs on one RNA. The overlap between N. crassa Puf3 and Puf5 mitochondrail cluster targets may present a place to trace the change from 8nt to 9/10nt binding elements and resolve whether binding elements lengthened or new ones evolved.

RNA regulons

We observe two modes of conservation of putative regulatory circuits. The modes of conservation are unified by the fact that both include the maintenance of RNA regulons. Keene (2007) proposed that RNA binding proteins bind circuits of RNAs that are functionally related, termed regulons. Theses regulons may be co-regulated by RNA binding proteins to maintain stoichiometric concentrations of multi-protein complexes or multi-component processes. Our survey of the fungal PUF network suggests that certain regulons have retained PUF regulation. Despite long evolutionary distance and examples of displacement of specific genes, particular regulons appear to be under conserved PUF control. The molecular specifics have diverged in the case of the mitochondrial cluster genes with a derived Puf4/5 regulatory state, but the PUF regulon is maintained.

MATERIALS AND METHODS

Bioinformatics

Ascomycota fungi were chosen based on previously determined orthology (Wapinski et al., 2007). 300 bases downstream of the translation termination codon were obtained from organism specific databases listed in Table S3. Each 3'UTR sequence was probed for putative binding elements using a custom perl script. The PWMs used to determine the log likelihood scores were generated based on RIP Chip (Gerber et al., 2004). Genes were filtered (> 50% species representation) then k-means (k=13) clustered. Clustering was visualized using Java Treeveiw (Saldanha, 2004). Gene ontology enrichments were generated using DAVID and Benjamini corrected p-values are reported (Balakrishnan et al., 2012; Dennis et al., 2003). Unbiased motif analysis was done with MEME (Bailey and Elkan, 1994).

Yeast three-hybrid

The RNA-binding domains of A. nidulans PUF4 and N. crassa PUF5 were cloned into activation domain-protein fusion plasmid, pGADT7 (Koh and Wickens, 2014). Sequences for the RNA were cloned into the Hybrid RNA plasmid, p3HR2 (Koh and Wickens, 2014). S. cerevisiae strain YBZ-1 1 (MATa, ura3-52, leu2-3, -112, his3-200, trp1-1, ade2, LYS2::(LexAop)-HIS3, URA3::(lexAop)-lacZ, and LexA-MS2 MS2 coat (N55K)) was transformed using standard lithium Acetate procedures. Luminescence was assayed following the manufactures instructions by micorpalte reader (BioTech Synergy 4). Raw luminescence was normalized to OD_{660}, and each biological replicate (n=3) was averaged and standard deviation was calculated.

HITS-CLIP

S. cerevisiae HITS-CLIP was preformed essentially as in (Wilinski et al., 2015). One modification was a supplemental treatment of the immunoprecipitated material with RNase R following overnight 3' DNA adaptor ligation was preformed.

N. crassa strains harboring a chromosomally integrated C-terminal FLAG tag at each PUF loci were generated by the Selker Lab as described (Honda and Selker, 2009). Strains were inoculated from freezer stocks and grown on slants (VM agar supplemented with L-Histidine) for 5 days at 25 °C. Conidia from slants were used to inoculate liquid cultures (VM liquid supplemented with L-Histidine). Mycelia were harvested after 2 days of growth shaking at 30 °C in a thin layer onto filter paper, UV-crosslinked for 7 min on each side, then frozen in liquid nitrogen. After cell disruption with mortar and pastel, lysate was cleared, and RNase A treated as above and immunoprecipitated with anti-FLAG M2 Magnetic Beads (Sigma) equilibrated according to the manufacturer's protocol. The *S. cerevisiae* protocol was then followed beginning with the final IgG bead washing step.

FIGURE LEGENDS

Figure 1. Conservation of putative PUF binding elements in Asycomocota fungi. (A) Phylogenetic tree of selected Ascomycota Fungi. The tree was constructed using NCBI Taxonomy Browser. Colored branches represent each subphylum, Hemi (Budding yeast) Pink, asdf (Filamentous fungi) orange, and Schozoascomycoata (Fission yeast) blue. (B) Representation of 8, 9, 10nt PWM models used to paramertize the log-likelihood function. Height of base represents probability of a base at each position in the binding element. (C-E) Log likelihood scores for putative PUF binding elements in 3'UTRs for species in A. Yellow is a high and blue is a low likelihood score. Log-likelihood scores for orthologous 3'UTRs are plotted in rows and each column represents a species. Clustering was done independently for each heat map. Species phylogenic tree is on top of each heat map. (C) Putative 8 nt binding elements from 4425 genes. (D) Putative 9nt binding elements from 4898 genes. (E) Putative 10nt binding elements from 4423 genes.

Figure 2. Evolution of PUF-mitochondrial RNA network. (A) Zoomed-in of mitochondrial cluster from Figure 1C. 8nt log-likelihood values from budding yeast and filamentous fungi 3'UTRs. (B) 9nt and 10nt log-likelihood values for 3'UTRs of mitochondrial cluster genes for filamentous fungi. (C) Overlap of genes containing putative binding element for three species. (D) MEME derived PWMs of enriched sequences found in 3'UTRs of mitochondrial cluster genes. (E) Phylogenetic tree of parlogous Puf4-Puf5 proteins. Puf5s have a blue background and Puf4s have a green background. The two filamentous PUFs are highlighted with orange

boxes. (F) Relative luminescence values as a proxy for binding affinity for each PUF protein assayed by yeast-three hybrid.

Figure 3. HITS-CLIP data. (A) Reproducibility of biological replicates. Height of CLIP peaks were log transformed and then plotted on each axis. Data colored based on the precipitated protein (*S. cerevisiae* Puf3p: blue, *N. crassa* PUF3: red, and *N. crassa* PUF5: brown) (B) Example peaks for each protein. Targets shown: YLL009C (*cox17*), NCU0058 (*al-2*), NCU02530 (*cox17*). Y-axis is the number of stacked sequencing reads, peak height. On x-axis blue bars are ORFs and the grey bars are the UTRs. The sequence is the likely binding element. (C) Overlapping CLIP targets for each protein. Only genes with orthologs in *S. cerevisiae* and *N. crassa* are included.

Figure 4. *In vivo* binding elements of PUF proteins. MEME derived PWMs from CLIP data. (A) *S. cerevisiae* Puf3p. (B) Previously defined *S. cerevisiae* Puf5p (Wilinski et al., 2015). (C) *N. crassa* PUF3. (D) *N. crassa* PUF5. (D) The previously defined Par-CLIP of H. sapiens Puf3 ortholog, PUM2, defined by PhyloGibbs (Hafner et al., 2010; Siddharthan et al., 2005).

Figure 5. Connect bioinfromatics and *in vivo*. (A) Gene Ontology (GO) enrichment of *S. cerevisiae* Puf3p, *N. crassa* PUF3 and *N. crassa* PUF5 CLIP targets. (B) Overlapping CLIP targets in mitochondrial cluster. Only genes with orthologs in *S. cerevisiae* and *N. crassa* were considered.

Figure 6. Alternative models of mitochondrial cluster circuit evolution. Grey circles represent relevant protein loss or circuit re-wiring. Blue "mito" bars represent genes in mito cluster. (A) Puf3 ancestral model. (B) Puf4 ancestral model. A Puf5 ancestral model would have the same features as the Puf5, thus it is not shown for simplicity.

Supplementary Figure 1. Putative binding elements in filamentous fungi and fission yeast. Log-likelihood scores for putative PUF binding elements in 3'UTRs for species filamentous fungi and fission yeast. Yellow is a high and blue is a low likelihood score. Log-likelihood scores for orthologous 3'UTRs are plotted in rows and each column represents a species. Clustering was done independently for each heat map. Species phylogenic tree is on top of each heat map. (A) Putative 8 nt binding elements from 4409 genes. (B) Putative 9nt binding elements from 4392 genes. (C) Putative 10nt binding elements from 4423 genes.

Supplementary Figure 2. Distribution of peaks in RNAs. mRNA regions where peaks are found for *S. cerevisiae* Puf3p, *N. crassa* PUF4 and *N. crassa* PUF5.

Supplementary Table 1. Gene Ontology enrichments for each binding element cluster.

Supplementary Table 2. Summary of high throughput sequencing reads.

Supplementary Table 3. 3'UTR sequence sources.

FIGURES

Figure 1
Figure 20
Wilinski et al.

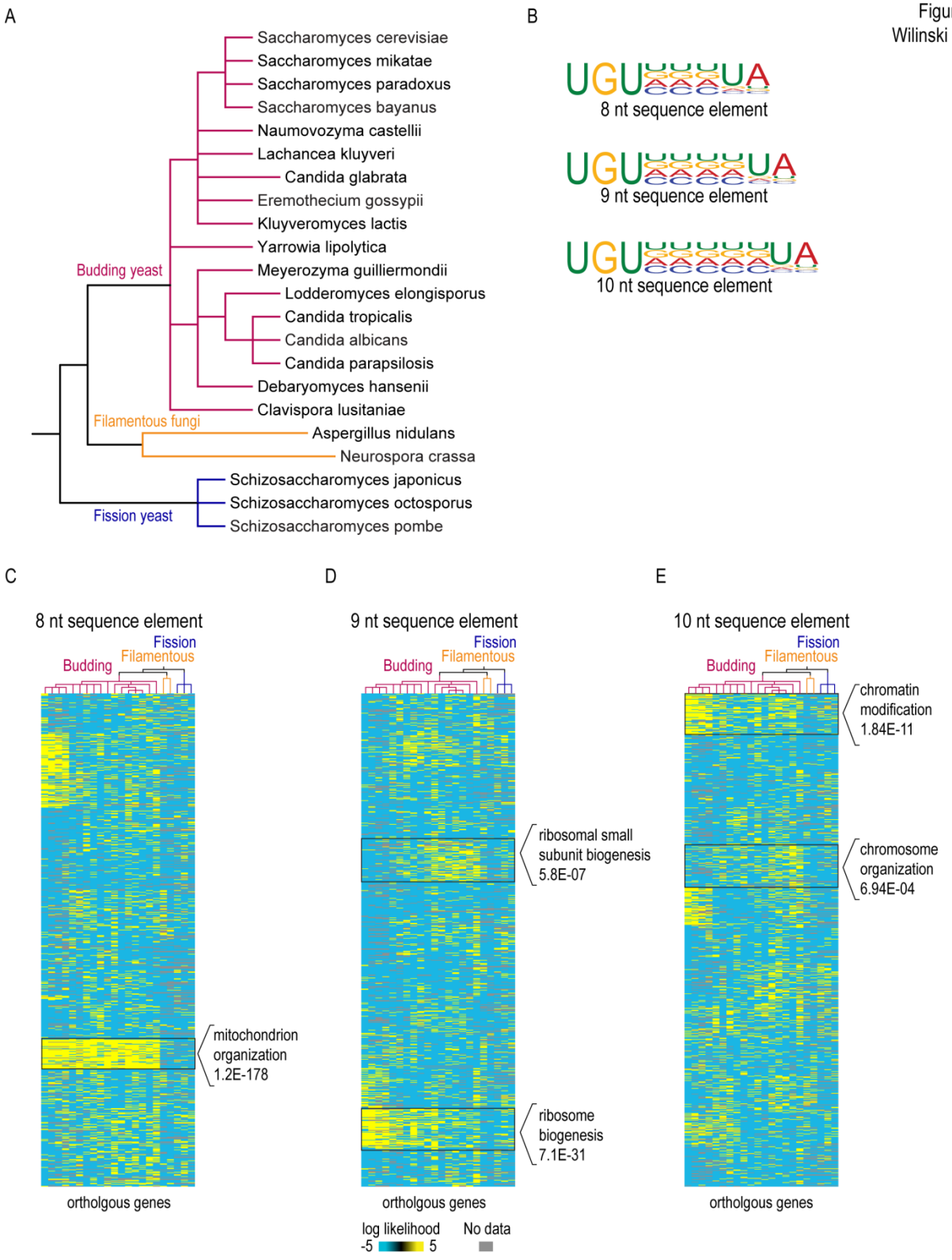


Figure 2
Figure 21
Wilinski et al.

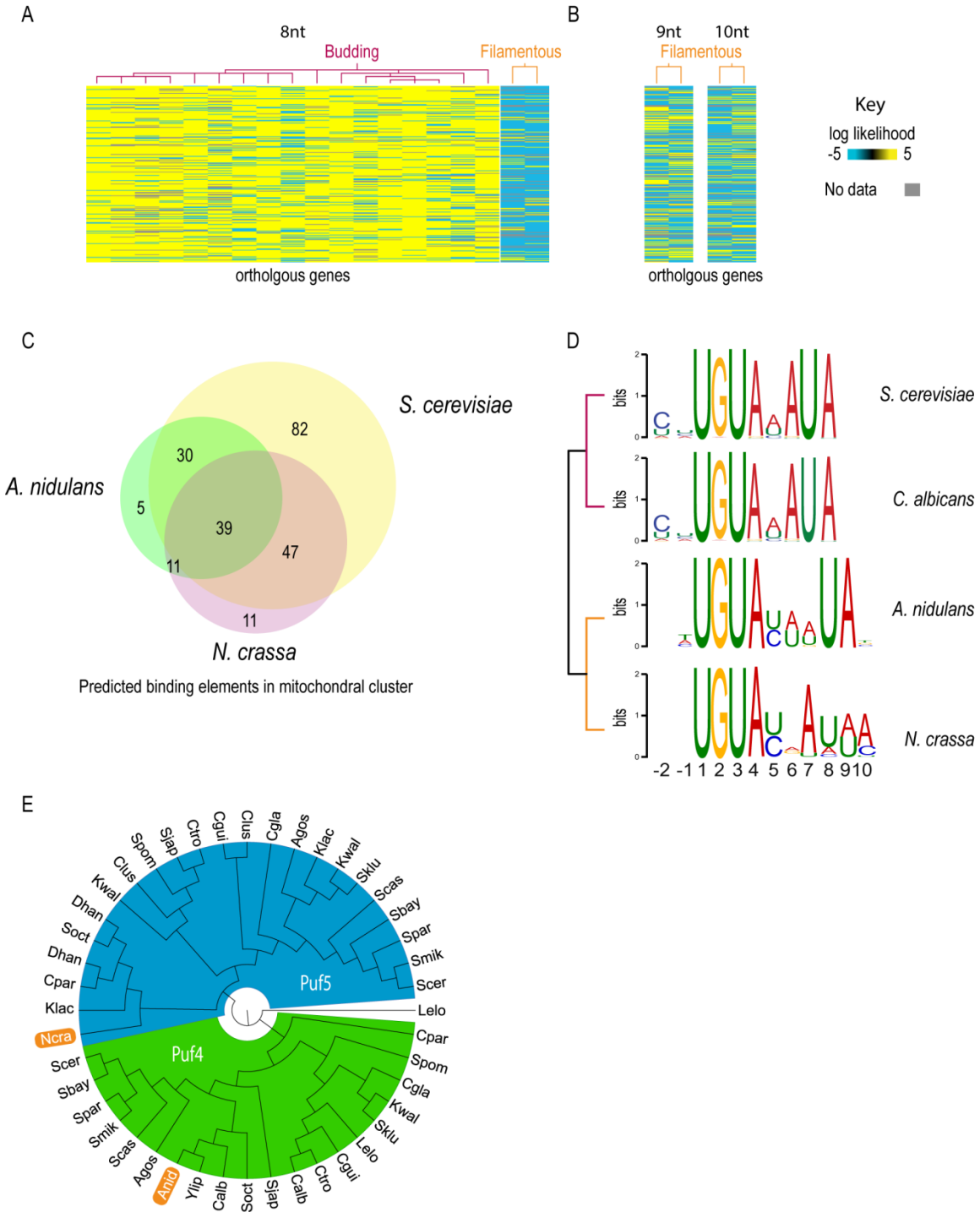


Figure 3
Figure 22
Wilinski et al.

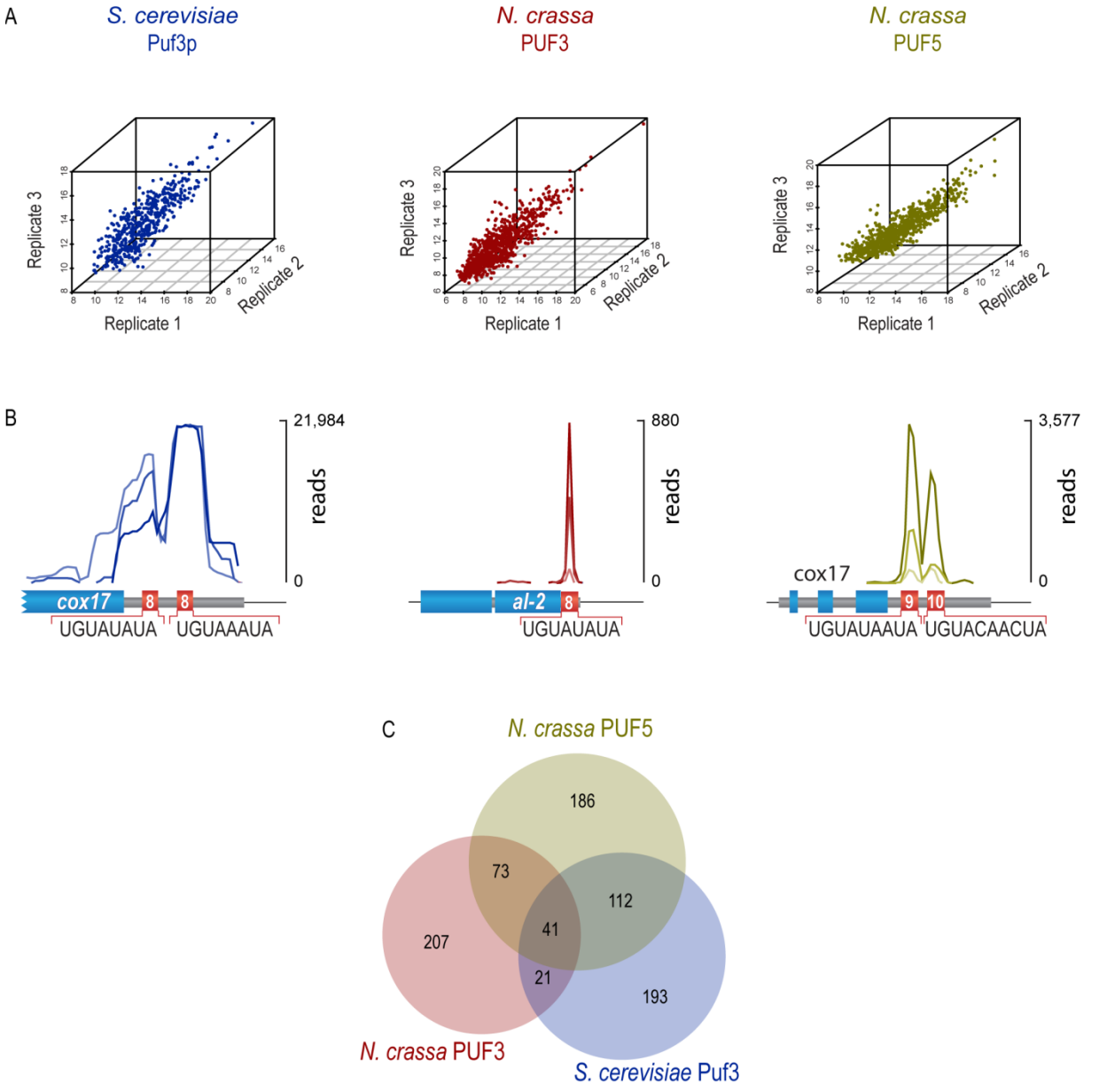


Figure 4
Figure 23
Wilinski et al.

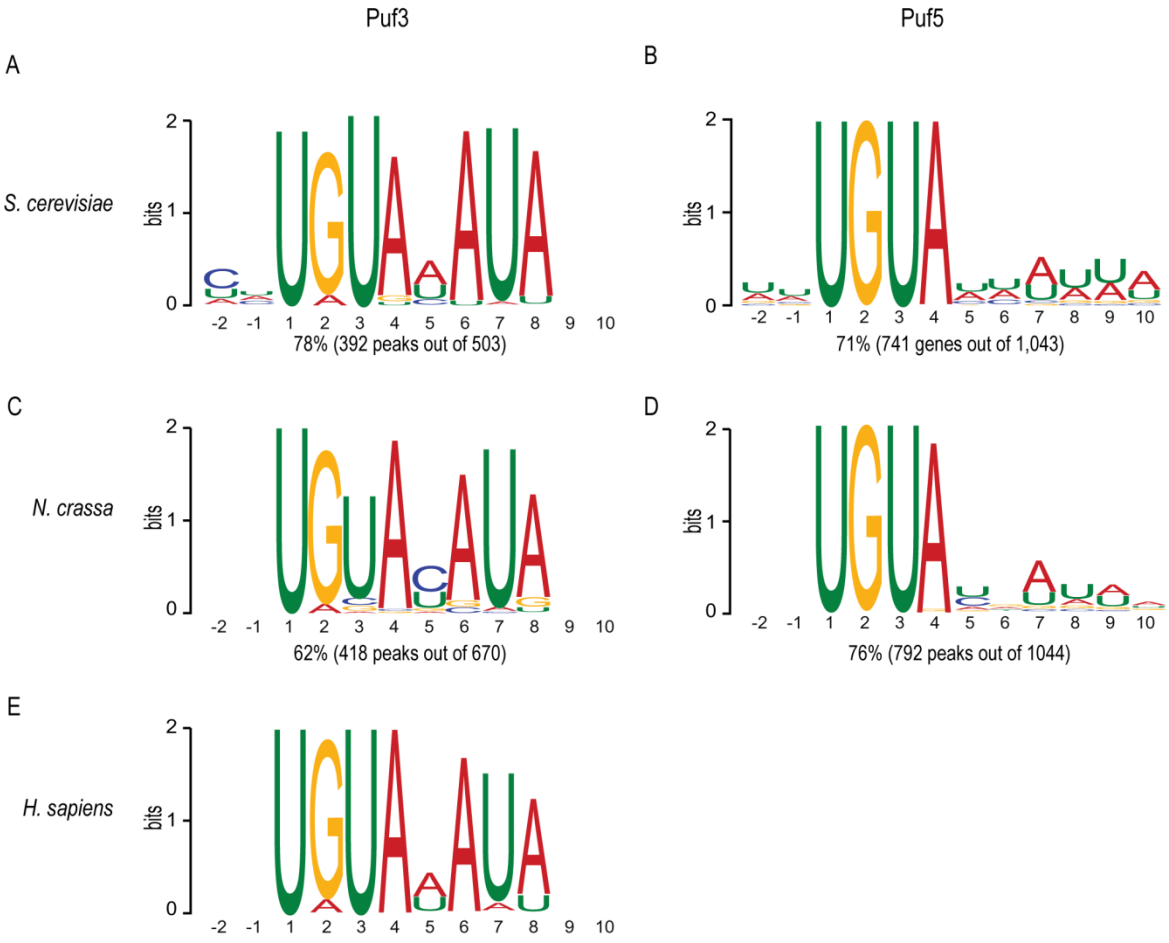
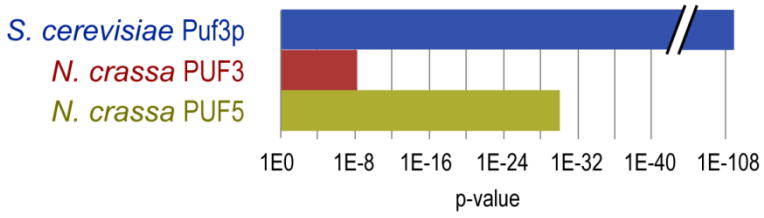


Figure 5
Figure 24
Wilinski et al.

A



B

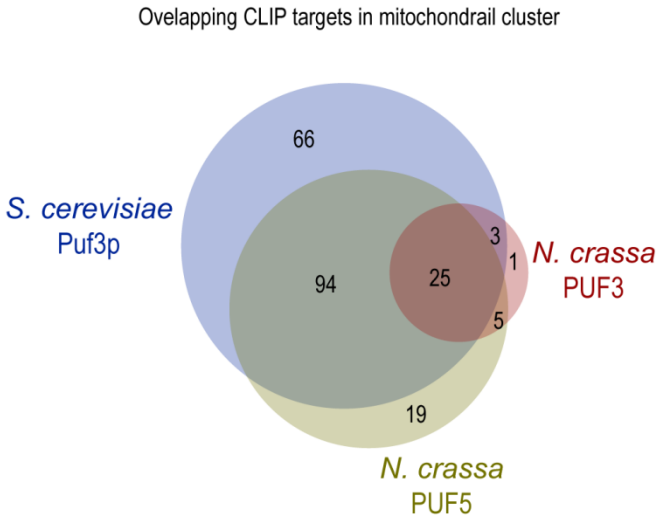
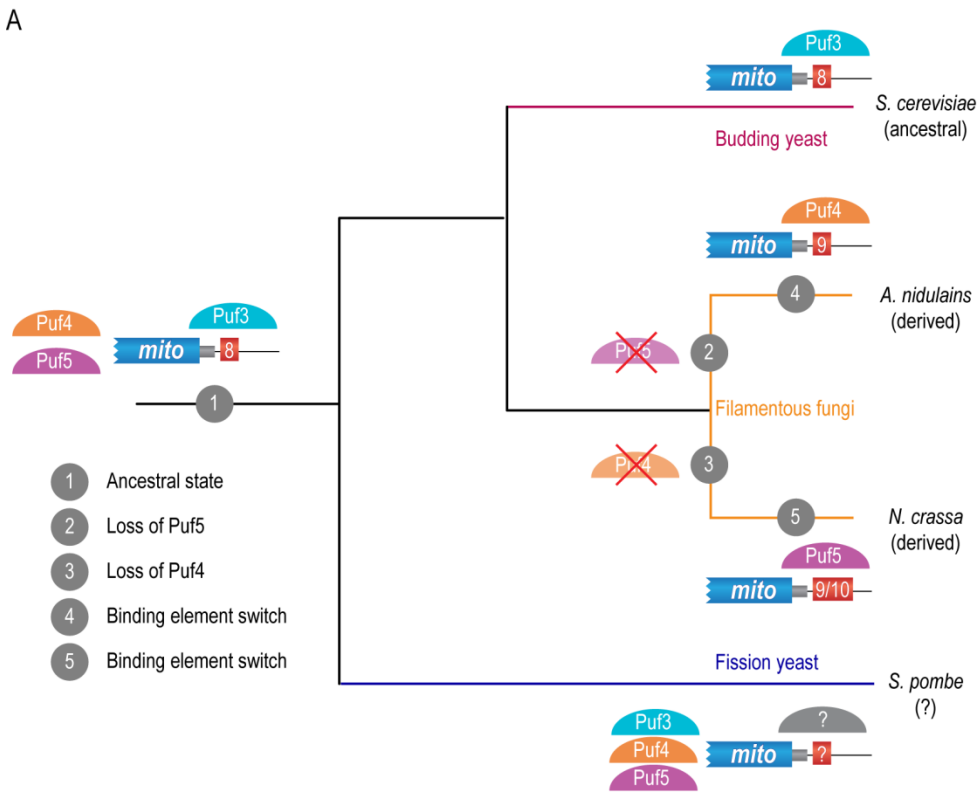
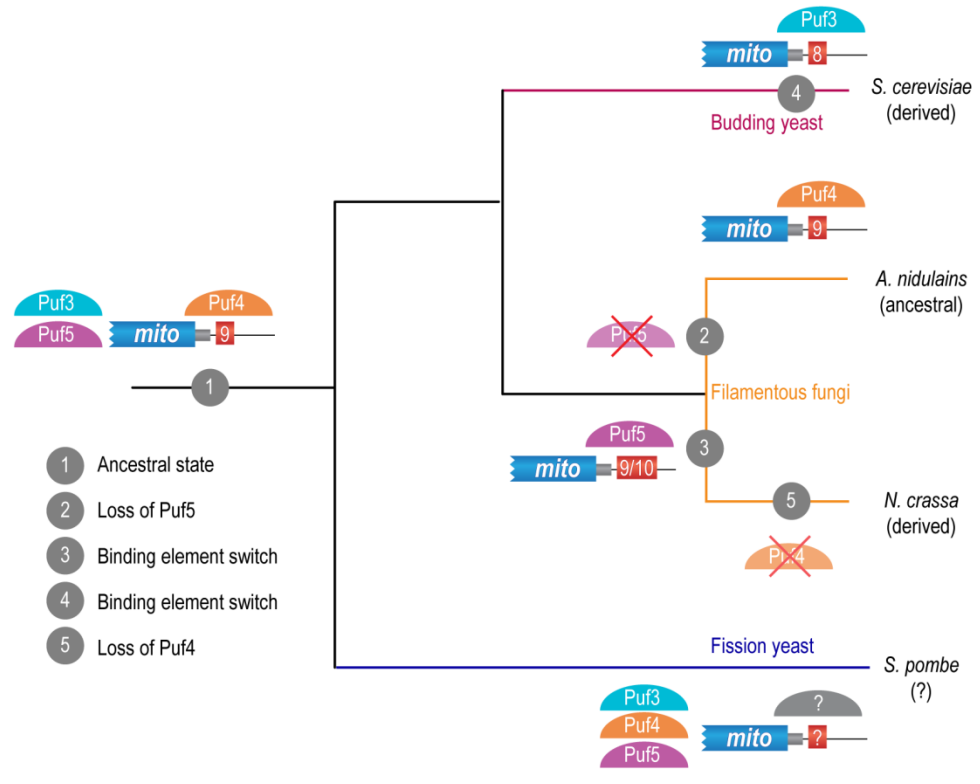
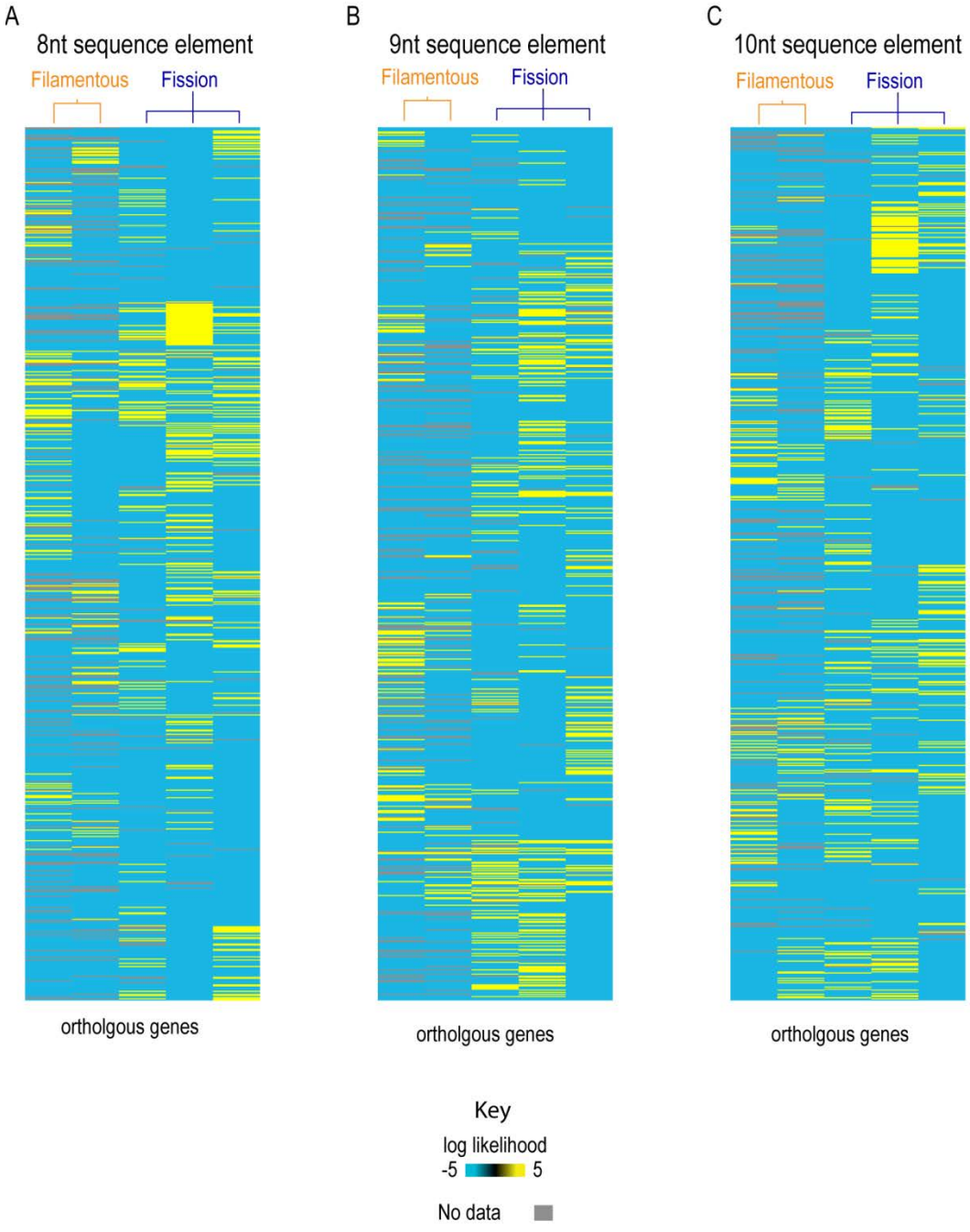


Figure 6
Figure 25
Wilinski et al.

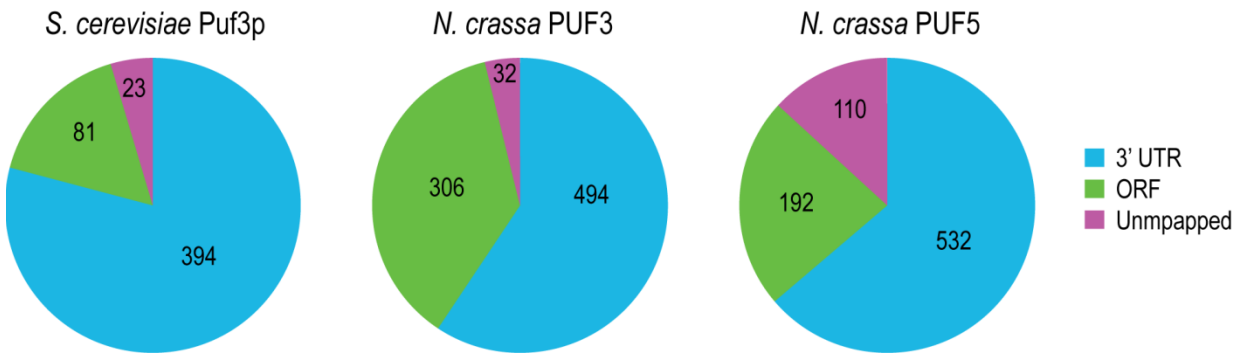




Supp Figure 1
Figure 27
Wilinski et al.



Supp Figure 2
Figure 28
Wilinski et al.



Supp Table 1
Wilinski et al.

8 nt putative binding element		
Cluster #	GO term	p-value
0	none	
1	none	
2	none	
3	none	
4	none	
5	none	
6	vesicle-mediated transport	5.63E-05
	Golgi vesicle transport	0.025405
7	none	
8	none	
9	mitochondrion organization	1.19E-178
	mitochondrial translation	1.22E-102
10	none	
11	cytoplasmic translation	1.22E-12
12	none	

9 nt putative binding element		
Cluster #	GO term	p-value
0	single-organism process	0.000581
1	transmembrane transport	1.46E-15
2	none	
3	none	
4	cytoplasmic translation	1.32E-33
	ribosomal small subunit biogenesis	5.48E-08
5	viral life cycle	9.73E-42
	symbiosis, encompassing mutualism	9.73E-42
	transposition, RNA-mediated	2.23E-34
	transposition	3.45E-33
6	none	
7	meiotic cell cycle	0.001215
	sporulation	0.008913
8	regulation of pheromone-dependent	0.008338
	regulation of signal transduction	0.008338
9	none	
10	none	
11	ribosome biogenesis	1.10E-35
	ribonucleoprotein complex biogenesis	1.08E-32
	rRNA processing	7.17E-24
12	none	

10 nt putative binding element		
Cluster #	GO term	p-value
0	chromatin modification	1.84E-11
	chromatin organization	3.85E-10
1	none	
2	none	
3	none	
4	chromosome organization	6.94E-04
	chromatin organization	1.12E-03
5	none	
6	none	
7	cytoplasmic translation	2.28E-06
	anion transport	3.61E-04
	nitrogen compound transport	3.98E-04
8	positive regulation of Rho GTPase activity	2.58E-03
9	none	
10	mitochondrial translation	2.95E-05
11	none	
12	none	

Supp Table 2
Wilinski et al.

Internal ID	Sample ID	Raw Reads	Reads after dup removal	% remaining	Reads after clipping	Mapped (unique)	% mapped	Notes
DW115	Sc Puf3 A	19,151,614	7,735,761	40	6,723,006	5,260,964	78	
DW116	Sc Puf3 B	14,524,322	5,782,164	40	3,161,397	1,709,752	54	
DW125	Sc Puf3 C	7,626,399	3,099,312	41	1,557,163	634,971	41	
DW100_102	Nc Puf3 A	1,179,860	456,082	39	402,543	354,391	88	*
DW111	Nc Puf3 B	14,899,440	11,974,763	80	7,856,464	5,162,488	66	
DW112	Nc Puf3 C	12,072,745	9,770,367	81	5,285,092	2,843,762	54	
DW101_103	Nc Puf5 A	1,355,794	1,096,425	81	983,281	877,507	89	*
DW113	Nc Puf5 B	18,748,631	14,990,087	80	12,394,324	10,541,429	85	
DW114	Nc Puf5 C	11,393,103	9,959,443	87	5,274,058	2,918,173	55	

* samples were sequenced on MiSeq

REFERENCES

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol *2*, 28-36.

Balakrishnan, R., Park, J., Karra, K., Hitz, B.C., Binkley, G., Hong, E.L., Sullivan, J., Micklem, G., and Cherry, J.M. (2012). YeastMine--an integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose tool-kit. Database (Oxford) *2012*, bar062.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol *4*, P3.

Dubnau, J., Chiang, A.S., Grady, L., Barditch, J., Gossweiler, S., McNeil, J., Smith, P., Buldoc, F., Scott, R., Certa, U., *et al.* (2003). The staufen/pumilio pathway is involved in Drosophila long-term memory. Curr Biol *13*, 286-296.

Gasch, A.P., Moses, A.M., Chiang, D.Y., Fraser, H.B., Berardini, M., and Eisen, M.B. (2004). Conservation and evolution of cis-regulatory systems in ascomycete fungi. PLoS Biol *2*, e398.

Gerber, A.P., Herschlag, D., and Brown, P.O. (2004). Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. PLoS Biol *2*, E79.

Gerber, A.P., Luschnig, S., Krasnow, M.A., Brown, P.O., and Herschlag, D. (2006). Genome-wide identification of mRNAs associated with the translational regulator PUMILIO in Drosophila melanogaster. Proc Natl Acad Sci U S A *103*, 4487-4492.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., *et al.* (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129-141.

Jiang, H., Guan, W., and Gu, Z. (2010). Tinkering evolution of post-transcriptional RNA regulons: puf3p in fungi as an example. PLoS Genet *6*, e1001030.

Jiang, H., Guo, X., Xu, L., and Gu, Z. (2012). Rewiring of posttranscriptional RNA regulons: Puf4p in fungi as an example. Mol Biol Evol *29*, 2169-2176.

Kershner, A.M., and Kimble, J. (2010). Genome-wide analysis of mRNA targets for Caenorhabditis elegans FBF, a conserved stem cell regulator. Proc Natl Acad Sci U S A *107*, 3936-3941.

Koh, Y.Y., and Wickens, M. (2014). Determining the RNA specificity and targets of RNA-binding proteins using a three-hybrid system. Methods enzymol *539*, 163-181.

Li, H., and Johnson, A.D. (2010). Evolution of transcription networks--lessons from yeasts. Curr Biol *20*, R746-753.

Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., *et al.* (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature *456*, 464-469.

Mallo, M., and Alonso, C.R. (2013). The regulation of Hox gene expression during animal development. Development *140*, 3951-3963.

Pick, L., and Heffer, A. (2012). Hox gene evolution: multiple mechanisms contributing to evolutionary novelties. Ann N Y Acad Sci *1256*, 15-32.

Saldanha, A.J. (2004). Java Treeview--extensible visualization of microarray data. Bioinformatics *20*, 3246-3248.

Siddharthan, R., Siggia, E.D., and van Nimwegen, E. (2005). PhyloGibbs: a Gibbs sampling motif finder that incorporates phylogeny. PLoS Comput Biol *1*, e67.

Sugiyama, J., Hosaka, K., and Suh, S.-O. (2006). Early diverging Ascomycota: phylogenetic divergence and related evolutionary enigmas. Mycologia *98*, 996-1005.

Taylor, J.W., and Berbee, M.L. (2006). Dating divergences in the Fungal Tree of Life: review and new analyses. Mycologia *98*, 838-849.

Villar, D., Flicek, P., and Odom, D.T. (2014). Evolution of transcription factor binding in metazoans - mechanisms and functional implications. Nat Rev Genet *15*, 221-233.

Wapinski, I., Pfeffer, A., Friedman, N., and Regev, A. (2007). Natural history and evolutionary principles of gene duplication in fungi. Nature *449*, 54-61.

Wickens, M., Bernstein, D.S., Kimble, J., and Parker, R. (2002). A PUF family portrait: 3'UTR regulation as a way of life. Trends Genet *18*, 150-157.

Wilinski, D., Chen, Q., Lapointe, C.P., Nevil, M., Campbell, Z.T., Hall, T.M., and Wickens, M. (2015). RNA regulatory networks established through curvature of a PUF protein scaffold.

Wohlbach, D.J., Thompson, D.A., Gasch, A.P., and Regev, A. (2009). From elements to modules: regulatory evolution in Ascomycota fungi. Curr Opin Genet Dev *19*, 571-578.

Zhu, D., Stumpf, C.R., Krahn, J.M., Wickens, M., and Hall, T.M. (2009). A 5' cytosine binding pocket in Puf3p specifies regulation of mitochondrial mRNAs. Proc Natl Acad Sci U S A *106*, 20192-20197.

List of Publications

Wilinsk D, Buter N, Wickens M. Rewiring a fungal RNA regulatory network through the evolution of binding sites. [In preparation]

Lapointe CP, **Wilinski D,** Campbell ZT, Wickens M. RNA Tagging identifies and reveals new features of RNA regulatory networks. [In preperation]

Wilinski D*, Qiu C*, Lapointe CP, Nevil M, Campbell ZT, Hall TM, Wickens M. RNA regulatory networks established through curvature of a PUF protein scaffold. [Under reivew—Nat Commun]

Qiu C, Kershner A, Wang Y, Holley CP, **Wilinski D**, Keles S, Kimble J, Wickens M, Hall TM. Divergence of Pumilio/fem-3 mRNA binding factor (PUF) protein specificity through variations in an RNA-binding pocket. J Biol Chem. 2011 Dec 28. [Epub ahead of print]