New Transcripts Identified for S. cerevisiae, S. pombe and Drosophila Using Novel cDNA Cloning

Les M. Hoffman*, Janina Görnemann, Erica Rodriguez, and David A. Brow *EPICENTRE Biotechnologies, University of Wisconsin-Madison

Abstract

Unannotated transcripts, including antisense, noncoding, intergenic, and potential regulatory RNAs, were identified in three different eukaryotic model organisms by cDNA cloning and sequencing. A rapid double-stranded cDNA generation system will be described that uses modifications of standard techniques and a new cloning vector. Short reverse transcription times facilitated cloning short (250-750 nucleotide, nt) cDNAs that were sequenced by two schemes. No size selection of cDNAs other than shorter reverse transcription was used, and a high percentage of cDNAs were apparently full-length. The average cDNA length was approximately 400 nt, a size which is usually excluded by following standard cDNA cloning protocols. Putative capping (transcription initiation) sites and poly(A) addition sites were located for most clones by single-pass sequencing. Sequencing purified cDNA clone plasmids or cDNA library colony PCR products gave similar results. Several cDNAs correlated with ChIP-Chip data showing RNA polymerase II occupancy sites in the S. cerevisiae genome (Steinmetz, E.J., et al., Mol. Cell 24, 735-746 (2006)). For several transcription units, cDNA clones contained multiple 5' and/or 3' ends. Many transcripts overlapped with or corresponded to those found using a different cDNA cloning approach (Miura, F., et al., PNAS 103, 17846-17851 (2006)). There was a preponderance of *S. cerevisiae* cDNAs for nuclear-encoded mitochondrial (mt) proteins. We postulate that the mRNAs of nuclear-encoded mt proteins may be more accessible to extraction due to their proximity to the mt network structure near the cell membrane of S. cerevisiae. Approximately 50% of the cloned cDNAs from Drosophila melanogaster S2 cells encoded ribosomal protein genes, with a strong (11:1) bias toward small ribosomal subunit cDNAs over large subunit clones. The reason for such a discrepancy between small and large subunit cDNAs is under investigation.

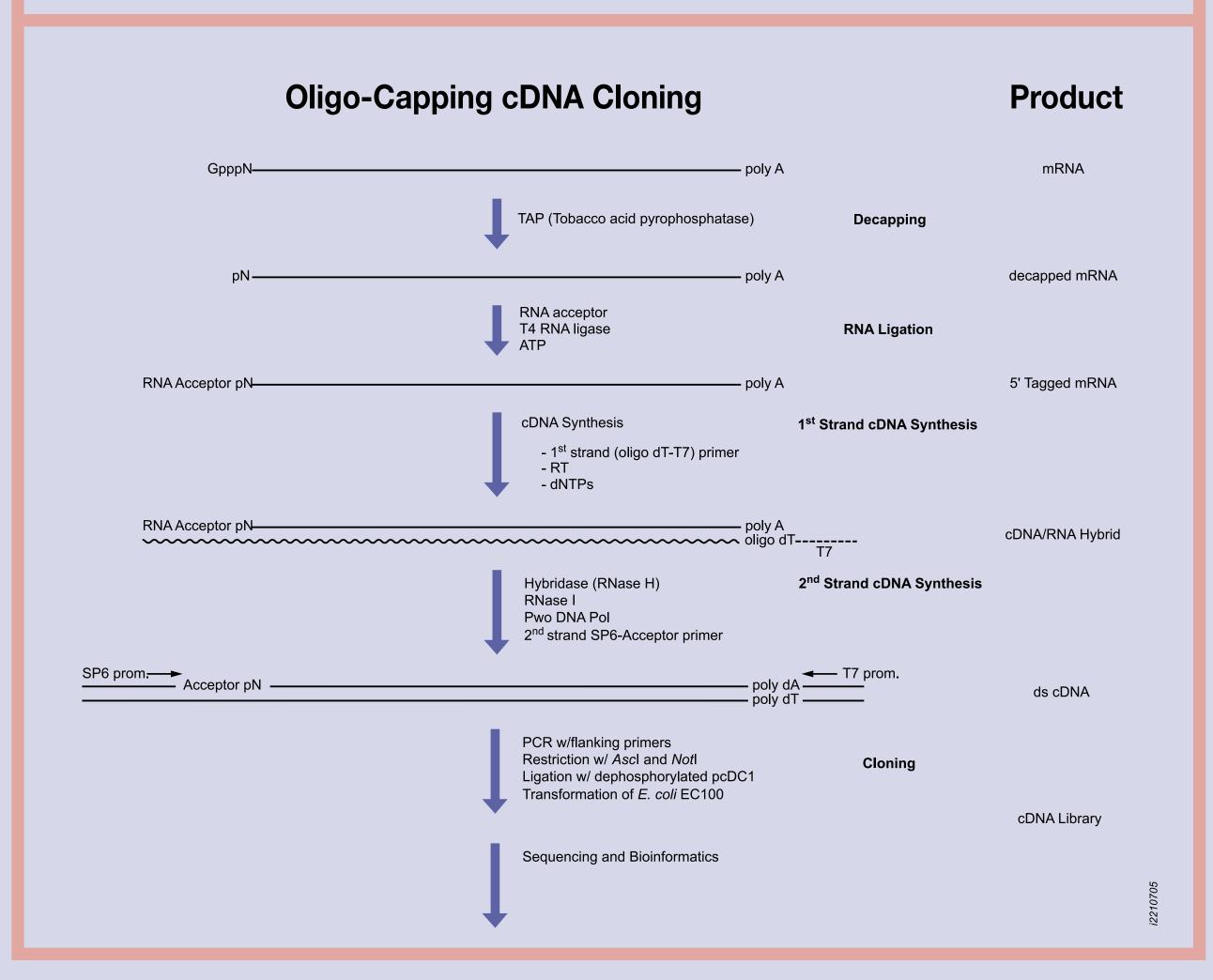
Introduction

cDNA cloning methods are being developed to meet the needs of bioinformatics and the evolving landscape of the transcriptome. The traditional view of neatly annotated transcripts representing the open reading frames (ORFs) of a genome have given way to a more fluid concept of the transcriptome. Recent articles (1-8) have noted noncoding, antisense, cryptic, unannotated or regulatory transcripts in virtually every system investigated. Some S. cerevisiae transcripts appear to be promiscuous intergenic, capped transcripts, called CUTs (1,2). CUTs are rapidly degraded, however, in wild-type yeast. Other yeast studies, such as the large-scale cDNA cloning and sequencing by Miura et al. (3), have identified a plethora of new yeast transcripts that are not "CUTs". Analyses of steady-state RNA (4) and RNA polymerase Il occupancy (5) using high-density tiled DNA microarrays have also indicated the presence of many intergenic transcripts in S. cerevisiae.

We present data for three different eukaryotic model organisms that show a range of new RNAs can be found by cloning cDNAs of "small, but not very small" RNAs, rather than selecting for larger, "gene-sized" cDNAs, as is traditionally done. The repertory of short-ish RNAs (named shishRNAs), both annotated and unannotated, varies between organisms. If longer (1-3 kb) cDNA is selected prior to cloning, few unannotated or CUT-like transcripts are found. The "oligo capping" method presented is a new technique in the transcriptome analysis arsenal that is fast and straightforward. To show the robustness of the technique, we made no attempt to exclude uncapped or degraded mRNAs from the template pools. Very few degraded, and necessarily 5' phosphorylated, RNAs were cloned by the new method. Several of the newly cloned Saccharomyces cDNAs correspond to transcripts found by analysis of RNA polymerase II occupancy sites by ChIP-Chip assays (5). The transcripts are located in regions of the genome that appear to have significant RNA Pol II occupancy, and several are in conserved regions of the yeast genome.

1. Wyers, F., et al., Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase.

- 2. Davis, C. A and Ares, M. Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6 in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. 1003: 3262-3267.
- 3. Miura, F. et al., A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. PNAS 103:17846-17851 (2006). 4. David, L., et al., A high-resolution map of transcription in the yeast genome. PNAS 103: 5320-5325 (2006).
- 5. Steinmetz, E.J., et al., Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 Helicase. Mol. Cell 24:735-
- 6. Furuno, M., et al., Clusters of internally primed transcripts reveal novel long noncoding RNAs. PLoS Genetics 2:537-553 (2006). 7. Yuan, G. et al., RNomics in Drosophila melanogaster, Nucleic. Acids Res. 31:2495-2501 (2003).
- 8. Watanabe, T., et al., Abundant poly(A)-bearing RNAs that lack open reading frames in Schizosaccharomyces pombe. DNA Res. **9**:209-215 (2002).



Chromosome 4 Chromosome 15 LAS17 V Chromosome 12 CCW14 192 nt. YLR392C Chromosome 9 Chromosome 7 LTR CRM1 Chromosome 10 VTC4 = Pol II M = cDNAclear RT-PCR product.

Examples of *S. cerevisiae* shish RNAs Sense transcripts that overlap the 3' (and 5') ends of ORFs: While these could be degradation fragments of mRNAs or failed primer extension products the cloning strategy and Pol II ChIP-Chip data suggest otherwise. However, preliminary Northern blot analysis has not yet detected these RNAs. A) dSNA2 (3D01): This 371 nt. cDNA starts in RBA50 ORF, which encodes an RNA polymeras interacting protein of unknown function. Pol II occupancy correlates better with the cDNA than with the SNA2 ORF. Conservation is poor except for the last 100 nt. RT-PCR with oligo(dT) gives major 3' ends near 3' end of cDNA and minor 3' ends about 200 nt. upstream (SNA2 mRNA?). **B) dLAS17** (3C04): This 365 nt. cDNA starts near the end of the LAS17 ORF, which encodes a WASP protein that nucleates branched actin filaments. It underlies a strong Pol II peak that overlaps the convergent RPS30B Pol II peak. C) CCW14 (2E04): This 192 nt. cDNA just over laps the CCW14 ORF, which encodes a cell wall mannoprotein. The DNA sequence is moderately conserved for about 200 bp downstream of the CCW14 ORF. **D) uPIG2** (3G02) and **dPIG2** (2G06): Strikingly, the PIG2 gene, which encodes a targeting subunit for the Glc7 protein phosphatase, exhibits short cDNAs at both the upstream and downstream ends of the ORF. Each cDNA correlates with a peak of Pol II, and both cDNAs are also found in the Miura et al. library. **Intergenic transcripts:** These transcripts are potential regulatory RNAs for the genes immediately downstream, as has been established previously for SRG1/SER3 (Martens et al., Nature 429:571-574, 2004). uCRM1 is the strongest candidate so far for such **E) uCRM1** (2F10): This 116 nt. cDNA lies upstream of the CRM1 ORF, which encodes a protein required for export of RNAs and proteins from the nucleus to the cytoplasm. RT-PCR with an oligo(dT) primer gave a major cDNA product about 200 bp longer than expected, which would reach the CRM1 ORF. Sequences within and upstream of the 2F10 cDNA are fairly well F) uVTC4 (1C10): This 272 nt. cDNA lies upstream of the VTC4 ORF, which encodes a vacuolar membrane protein. It did not give a

l	Well	Length	Ch.#	Location (bp)	ORF Name (Standard)	ORF Name (Systematic)	ORF location	Sense or Anti-sense to ORF	Intergenic	Watson or Crick Strand	Poly (A) Tail approx	Comments
ı	E04	191 BP	XII	904413-904604	CCW14	YLR390W-A	903724-904440	SENSE	Intergenic	Watson	19	Barely skims CCW14, no other ORF on 3' side
	F10	115 BP	VII	932276-932391	CRM1	YGR218W	932544-935798	SENSE	Intergenic	Watson	22	Near the 5' end of CRM1
	G06	222 BP	IX	272692-272914	PIG2	YIL045	271160-272776	SENSE		Watson	20	Covers a small part of the 3' end and then beyond
ı	G02	140	IX	271066-271206	PIG2	YIL045W	271160-272776	SENSE		Watson	YES	PARTIAL poly(A) 44 nt after ATG; overlaps F Miura transcrs
ı	D01	380	IV	1490684-1491054	RBA50	YDR527W	1491086-1492405	SENSE		Watson	YES	PARTIAL, but protein is not well-annotated
	C04A	375	XV	677607-677971	RPS30B	YOR182C	678794-678192	ANTI		Watson	YES	unannotated RNA near tRNA- lys gene, SAGE tags present. Covers the 3´ end of LAS17 and then is anti sense to the 5´ end of RPS30B.
	C10	290 BP	Х	413610-413882	VTC4	YJL012C	413393-411228	SENSE	Intergenic	Crick	40	Upstream of VTC4

Methods

RNA Isolation

For all RNAs except the D.melanogaster, the EPICENTRE MasterPure™ Yeast DNA Purification Kit was used as recommended in the product literature. D. melanogaster S2 tissue culture total RNA was obtained from Ambion. T4 DNA ligase and cloning reagents were from EPICENTRE. cDNA Synthesis

See figure at left for graphics

cDNA regions of colony swipes were amplified using condition "B" of the FailSafe™ PCR Kit (EPICENTRE) with the M13 forward and reverse primers. PCR products were purified by precipitation with the DNA Fragment Precipitation Solution (EPICENTRE) followed by one wash with cold 70% ethanol.

Sequencing

1/16 X BigDye sequencing reactions were in a 96 well plate with conical wells (ABI). One primer, the M13 Reverse primer, was used for all sequencing reactions in a single pass. Sequencing from the poly(A) end of the cDNAs was problematic due to long oligo(T) regions. The reactions were purified by ethanol precipitation and a 70% ethanol wash, and were sequenced on an ABI 3730 at the University of Wisconsin-Madison Biotechnology Center.

Data Analysis

Sequences were analyzed with Chromas Pro software (www.technelysium.com.au/chromas. html). The browsers for each species are listed under "WEB RESOURCES".

Summaries of cDNA Cloning

Unsized <i>S. cerevisiae</i> cDNAs							
lone	TAG	poly(A)	RNA length	BLASTN best match	Chr. Location	<u>COMMENTS</u>	
A1				BAD sequence			
A2	yes	yes	425	TOM6, mito protein			
A3	yes	yes	280	YGR174W-A protein of unknown function		full length	
A4	yes	yes	140	upstream of TOM6, mito protein	XV 413544 - 413681	SAGE tags present	
4 5	yes	yes	700	QCR6/YFR033C, Subunit 6 of the ubiquinol cytochrome-c reductase		full length	
46	yes	yes	230	COR1/YBL045C ubiquinol-cytochrome c reductase complex (mito)	II 134229- 134001	partial 3' end; overlaps F Miura transcripts	
47	yes	yes	115	dubious ORF, 28 AA hypothetical	VII 149552- 149437	F Miura transc. Terminates in same region	
4 8	yes	yes	780	ARF2/YDL137W, ADP-ribosylation factor		full-length	
49	yes	yes	410	TOM6, mito protein		full length	
10	yes	yes	195	UGX2/YDL169C unknown prot,. mRNA accumulates in stress		Partial, 3´ end, no SAGE tags	
11	yes	yes	525	SCT1/YBL011W		Partial, 3' end, no SAGE tags	
12	yes	yes	200	CPB6/YBR120C		Partial, 3' end, no SAGE tags; no F Miura transc.	
31	yes	yes	440	SEM1 /YDR363W-A		full-length ORF	
B2	yes	yes	750	VPS51/YKR020W		full-length ORF	
B3	yes	yes	530	Ty1 LTR/YLRWdelta6			
B4	yes	yes	340	NEW transcript upstram of HXT4 high-affinity glucose transporter	VIII 289264 -288926	overlaps F Miura transcr; no SAGE tags , no ORFs	
B5	700	700		BAD sequence	200201 200020	Transport initial durinos; no or the tago; no office	
B6	yes	yes	90	MSC2/YDR205W	IV 861523 -861605	in 3' untranslated;no SAGE tags; NO F Miura transc.	
B7	yes	yes	88	ARK1/YNL020C Serine/threonine protein kinase	XIV 597583-597497	poly(A) ~42 nt after ATG start; NO F Miura transc.	
38	yes	ves	280	GLR1/YPL091W; glutathione oxidoreductase,	XVI 375231-375512	ends 12 nt after ATG start, no F Miura annotated transc.	
39	,	,		weird segence, no Sc hits		,	
10	yes	yes	380	INH1/YDL181W		starts 39 nt after ATG start, no other ATGs	
11	yes	yes	300	3' of ORF for YDR090C hypothetical protein	IV 624994- 624726	SAGE tag present, no F Miura transcr	
12	yes	yes	275	3' end of HEF3/YNL014W	XIV 609247- 609519	partial, no F miura transc	
71	yes	yes	290	YGR174W-A protein of unknown function		full length	
2	yes	yes		DAD4/YDR320C-A, starts after ATG	IV 1108482-1108198	many cap sites in F Miura transcripts	
; <u> </u>	yes	yes	375	LAS17/YOR181W, unannotated		PARTIAL, 3' END, NO SAGE tag	
4A	yes	yes	375	numbered ???? unannotated RNA near tRNA-lys gene	XVI 582555 -582228	SAGE tags present	
5	yes	yes	334	ATP19/YOL077W-A; subunit k mito F1F0 ATP Synthase	AVI GOZOGO GOZZZO	full-length	
6	you	you	001	strange clone, no cap oligo or poly(A)		Tuli longui	
.7 .7	V00	1/00	455	DDR2/YOL052C-A		full-length	
	yes	yes	433			iuii-ieiigui	
08 09	yes	yes	150	only 43 A's as poly(A) 3' of STF1 gene, unannotated	IV 229459- 229593	CACE tog procent	
10	yes	yes	130	strange clone, BAD sequence	IV 223403- 223030	SAGE tag present	
						DADTIAL OF END PROCESSOR TO	
211	V00	1/00	170	GDH3/YAL062W	VIII 676906 677060	PARTIAL, 3' END, near SAGE tag	
12	yes	yes		3' of YMR206W	XIII 676896- 677062	no SAGE tags	
01	yes	yes	380	SNA2 YDR525W-A		PARTIAL, but protein is not well-annotated	
)2	yes	yes	595	YJR104C/S0D1 cytosol superoxide dismutase		full length	
)3	yes	yes	510	DDR2/YOL052C-A	VIII 505151	full-length	
)4	yes	yes	370	downstream antisense from tK(CUU)P tRNA	XVI 582131-582059 -	21.65	
15	yes	yes	410	3' end of YJL016W putative GFP-fusion protein	X 405583 - 407268	no SAGE tags	
)6				A-rich, weird, can't BLAST			
D7	yes	yes	740	TSA1/ YML028W ; thioredoxin peroxidase		full length	
D8	yes	yes	~200	YDL169C; transcript accum. In response to stress; partial 3' end	IV 158195-158001	no SAGE tags	
09	yes	yes	110	3' end of FES1/YBR101C; HSP70 nucleotide exchange factor	II 443846- 443740	no SAGE tags	
10	yes	yes	535	GSF2/YML048W		partial., 3' end, near SAGE tag	
11	yes	yes	284	No annotated transc, near left telomere end	XV 5833- 5546	overlaps F Miura transc 3' of BDS1, bact-derived sulfatase	
12	yes	yes	90	3' OF IST1/ YNL265C, putative translation initiator	XIV 144211-144140	no F Miura transc	

<u>clone</u>	<u>TAG</u>	poly(A)	RNA length	BLASTN best match	Chr. Location	COMMENTS	
A2	yes		430	DDR2/Y0L052C-A	XV, 231374 bp to 231781 bp	full length	
A4	yes	yes	650	SPI1	V, 468330 bp to 468976	full length	
A 5	yes	?	~600	TSA1		full length	
A6	yes	?	600	TSA1	XIII, 220214 bp to 220694	full length	
A7	yes		1200	PBH2, member of Prohibitin complex of inner MITO membrane, see E8	VII, 953531-953142?	full length	
A8	yes	?	~960	APD1, YBR151W, see C1	II, 544876 bp to 545273 ?	?	
A10	yes	yes	61	down from GLN1 gene, same dir.	XVI, 643349 bp to 643417	why so smal	
A12	yes	19, yes	290	3' end of MSP1			
B4	yes	?	~470	YFR033c, QCR6 Subunit 6 of the ubiquinol cytochrome-c reductase	VI, 224757 to 224314	full length	
B5	yes	?	~1925	GYP1/YOR070C	XV, 457822 to 455909	fulllength	
B6	yes	?	1480	YBR283c, SSH1, cotranslational protein translocation complex	II	full length	
B8	? Prob.	?		down from GLN1 gene, same dir., see A10			
В9	yes			Copia transposon LTR			
B10	yes	?	~1925	GYP1/YOR070C, see B5	XV, 457822 to 455909	full length	
B11	yes	?	~600	TSA1			
B12	yes	?	~360	FPR1, YNL135c Peptidyl-prolyl cis-trans isomerase	XIV, 372228 to 371884	full length	
C1	?	?	~960	APD1, YBR151W, see A8			
C2	?	?		lysostaphin clone, a visitor from another plate			
C3	yes	?	600	TSA1			
C4			600	TSA1			
C5	yes		~1390	TEF1, YPR080w Translational elongation factor EF-1 alpha	XVI, 700592 to 701968		
C6			600	TSA1			
C7	yes	?	~1040	DFM1/YDR411C, derlin family ER protein	IV, 1294386 to 1293361.	full length	
C8	yes	yes	470	YHR055C, CUP1-2, metallothionine	VIII, 214720 to 214535	full length	
C9			600	TSA1			
C10	yes		600	TSA1			
C11	yes	yes	~300	RPL43a, RIBO protein	XVI, 654163 to 654844	full length	
C12	yes	yes	470	YHR005C-A, MRS11, MITO TIM10	VIII, 115896 to 115615	full length	
D1	yes	yes	325	TRX2, cytoplasmic thioredoxin	VII, 913232 to 912918.	full length	
D2	yes		600	TSA1			
D3	yes	?	~1300	LYS20, homocitrate synthase	IV, 133438 to 134724	full length	
D4				TSA1	·		
D5	yes	yes	460	YDR115W, putative MITO ribosomal protein Large subunit,	IV, 682172 to 682489	full length	
D6	yes	?	~1000	SBP1, Nucleolar single-strand NA binding protein, binds snRNAs	VIII, 34075 to 33191	full length	
D7	yes	?	~1500	ATG4, Cysteine protease req. for autophagy	XIV, 227371 to 228855	full length	
D8	yes		~1000 ?	ARG80 Transcription factor (arginine)	XIII, 352602 to 353135	full length	
D9	?	?	~2000	SAC7, GTPase activating protein	IV, 1252529 to 1254493	full length ?	
D10	yes	?	~1500	SSH1, cotransl. Protein translocation element	II, 770411 to 768939	full length	
D11	?	?	~2000	GYP1, Cis-golgi GTPase-activating protein	XV, 457822 to 455909	full length ?	
D12	yes	?	~1500	DLD3, D- lactate dehydrogenease	V, 16355 to 17845	full length	

"2-4 Kb" Sized *S. cerevisiae* cDNAs

	Color Le	egend Unannotated untranscripts Antisense transcripts poly(a) only

	llnei:	zed <i>S. pombe</i> cDNAs (Largely Ribo	neomal Protein cDNAc)
	011312		
		Gene Description	<u>Features</u>
A6		rpp101 Ribo, acidic ribo prot rpp1-1, spliced gene	full length
A7		rpp101 Ribo, acidic ribo prot rpp1-1, spliced gene	full length
A8	660	rps802 40 S ribo S8	full length
A9	~900	rpa12 DNA-directed RNA polymerase complex I subunit	full length
A10	77	gene PARTIAL 7SL signal recognition particle component	poly (A) within ORF
A11	680	sbpc2d10.03c conserved hypothetical	full length
B1	510	rps1901, ribo S19	full length
B2	500	972h- 40S ribo S10	full length
В3	510	rpl2402, ribo L24	full length
B4	220	rp14102, ribo L41	full length, spliced
B5	580	rp11202, ribo L12	full length, spliced
B6	750	UNANNOTATED RNA, no ORFs of any size	
B7	480	rps2201, ribo S22	full length
B8	365	ANTISENSE to sequence orphan SPACUNK4.11C	no ORFs > 40 AA
B9	450	rp122	full length
B10	187	UNANNOTATED RNA, no ORFs of any size >40 AA	
B11	125	ANTISENSE to sequence orphan SPBC1921.04c	
B12	215	UNANNOTATED RNA, no ORFs of any size	downstr of enhancer of RNA-mediated gene silencing
C2	520	rps23, ribo S23	full length
C3	~450	rpp101 Ribo, acidic ribo prot rpp1-1, spliced gene	full length
C4	500	rps1901, ribo S19	full length
C5	~550	IES6 chromatin remodeling complex subunit	full length
C6	60	60 A's caught in act of degrad?	
C7	440	rpl3201, ribo L32	full length
C8	765	rpl100, ribo L10	full length
C9	400	rpl3001, ribo L30	full length
C10	660	rps902, ribo S9	full length
C11	610	UNANNOTATED RNA, no ORFs >40 AA	
C12	530	sui1, translation initiation factor eIF1	full length
D2	290	ANTISENSE to SPCC825.08C , blocks initiator ATG	N-acetyltransferase (predicted)
D3	~380	ANTISENSE to 3' end of SPBC13G1.10c	DEAD/DEAH box helicase
D4	460	rps1602, ribo S16, spliced transc	full length
D5	480	rps2201, ribo S22	full length
D6	385	rpl4301, ribo L43, spliced transc	full length
D7	335	partial transc TPI1, triose phosphate isomerase	no strong promoters
D8	400	rpl31, ribo L31, spliced transc	full length
D9	515	partial, 3' end of SPCC569.05c	spermidine family transporter (predicted)
D10	495	qcr9, ubiquinol-cytochrome-c reductase complex sub. 10	PARTIAL transc., extends 3' of gene, 3 introns
D11	580	alr2 alanine racemase ANTISENSE, blocks initiator ATG	adios, oxionos o di gono, o intiono
D12	230	rpl4102, ribo L41	full length, spliced
U 1 Z	230	1µ14102,1100 L41	iuli icrigui, spilceu

		_					
D. melanogaster Unsized cDNAs							
<u>clone</u>	TAG	poly(A)	RNA length	BLASTN best match	Chr. Location	COMMENTS	
A2	у	у	665	cytc-p , cytochrome c, somatic cell	2L, 16,720,015 to 16,724,648	full length	
A3	у	у	565	Copia element LTR, see A8 & A9			
A4	У	у	550	CG3922-PB, RpS17 RIBO Prot	3L, 9,419,895 to 9,424,208	full length	
A5	у	у	530	CG15693-RA, RpS20 RIBO Prot	3R	full length	
A6	у		~750	CG4294-RA, RpS16 RIBO Prot	2R, 18,492,080 to 18,496,135	full length	
A7	у	у	80	5' long terminal repeat LTR and untransl leader, Copia element		~50 A's	
A8	У	у	570	Copia element LTR, almost ID to A3			
A9	у	у	565	Copia element LTR, almost ID to A3	3R, 27,158,768 to 27,163,287		
A11	у		>1 KB	partial CG11055-RB, esterase, contained in ORF			
A12	у	у	525	CG15693-RA, RpS20 RIBO Prot	3R	full length	
B2	у	у	380	5.8S rRNA , 17 A's		full length	
В3	у	у	530	CG15693-RA, RpS20 RIBO Prot	3R	full length	
B4	у	у	524	CG6684-RA, RpS25-RA RIBO Prot	3R	full length	
B5	у	у	530	CG15693-RA, RpS20 RIBO Prot	3R	full length	
B6	У	у	890	Rps3A-RA RIBO Prot	4 (left end)	full length	
B7	У	у		PARTIAL Rm62 helicase involved in dsRNA-mediated silencing	3R		
B8	У	у	~330	CG11710-RB, 3' end downstr of ORF	X		
B10	у	у	565	CG12740-RD, RpL28 RIBO Prot	3L	full length	
B11	?	у	435	Copia element 1391 transcript, many A's	3R		
B12	у	y ?	~600	CG2998-RA, RpS28A RIBO Prot	Х	full length	
C1	У	у	590	RpS16 RIBO Prot	2R	full length	
C2	У	у	370	CG9032-PA RpS28b RIBO Prot	X	full length	
C3	у	у	420	CG7630-RA uncharacterized protein	3L	full length	
C4	у	у	600	RpS16 RIBO Prot	2R	full length	
C5	У	у	345	rel, <i>Relish</i> , 3´end of ORF; in intron of nfdc, NF-kappa B-like	3R		
C6	у	у	735	CG8331-PA, small GTPase regulatory activity	2R	full length	
C7	У	у	95	Copia element 5' long terminal repeat LTR and untranslated leader	140 A's		
C8	?	у	450	unannotated transc, probably full-length, similar to one that is	3LHet:604499608751	full length	
C9	у	у	525	RpS20 RIBO prot	3R	full length	
C10	у	у	120	3' end of transc, CG6105-RA, hydrogen-transporting ATP synth activ	2L		
C11	у	у	515	RpS28b RIBO prot	X	full length	
C12	У	у	910	RpS3A-RA RIBO Prot	3R	full length	
D1	У	у	535	RpS20 RIBO prot	3R	full length	
D2		у	225	cloned at Not I site, 3' end of HmgZ ORF, internal oligo (A) priming	2R	9	
D3	у	у	375	partial ANTISENSE transc, Tenascin major (Ten-m) secreted prot.glycan	antisense to INTRON, 3L		
D4	У	у	535	RpS20 RIBO prot	3R	full length	
D5		у	635	MAYBE deleted cDNA clone- Partial CG4840-PA	2R	91	
D6		у	360	cloned at Not I site of cDNA, partial RpS24 RIBO prot	2R, long poly(A)		
D8	у	у	980	UTR3 gene Saccharomyces cerevisiae, VISITOR from another organism	,g p// y	?	
D9	У	у	695	cytochrome C proximal	2L	full length	
D10	У	у	380	oho23b-RC, overgrown hematopoietic organs at 23B	2L	full length	
D10	у	у	405	Copia element LTR		Tun tongui	
D12	У	у	635	RpL24 RIBO prot	2L	full length	

Plate Position	Ribosomal Prot.	Polypyrimidine end?	End Sequence
B10	RpL28	YES	стттсстт
D12	RpL24	YES	ATTTGG
F12	RpL13A	YES	сттстттсс
H10	RpL27A	YES	стттсттт
A4	RpS17	YES	сттттстттс
A 5	RpS20	YES	стсттсстттс
A6	RpS16	YES	стттт
A12	RpS20	YES	стсттсстттс
В3	RpS20	YES	стсттсстттс
B4	RpS25	YES	стттствстт
B5	RpS20	YES	тстсттссттт
B6	RpS3A	?	
B12	RpS28A	YES	сттттс
C1	RpS16	YES	стттттсст
C2	RpS28B	?	
C4	RpS16	YES	стстттсст
C9	RpS20	YES	стсттстттс
C11	RpS28B	YES	СТТТТТССТТ
C12	RpS3A	YES	CTTTTTCCGGTTT
D1	RpS20	YES	стсттстттс
D4	RpS20	YES	стсттстттс
D6	RpS24	YES	CTTTGCCTCTTT

Drosophila RP Clone Transcription

cDNA Cloning Results

1. Saccharomyces cerevisiae

With unsized cDNA synthesized from total yeast RNA, there are many unannotated transcripts, including antisense, noncoding and potentially regulatory RNAs. A few of these transcripts are described in more detail in the following panel. Multiple clones from the same transcription unit but with different cap sites and/or polyadenylation sites were found.

We found an unusually large representation of nuclear gene products encoding mitochondrial proteins. Although we hypothesized that the RNA extraction method might favor the isolation of mRNAs associated with the reticular network of yeast mitochondria, in fact most of the mRNAs isolated do not appear to co-localize with mitochondria (Garcia et al., Mol. Biol. Cell 18:362-368 2007). Thus, the reason for enrichment of mRNAs encoding mitochondrial proteins is unknown. Using purified 1-3 kb cDNA, there are few unannotated transcripts present in the library. Instead, mostly full-length annotated transcripts predominate. The average cDNA length was 1 kb.

2. Schizosaccharomyces pombe

A relatively large number of antisense transcripts were cloned, about 6% of the total. Two of the antisense RNAs could block initiation of transcripts from the canonical transcription start sites. Around 60% of the cDNAs encoded ribosomal proteins, one-third for the small ribosomal subunit and two-thirds for the large subunit. Four of the unannotated RNAs contained no ORFs coding for more than 40 amino acids. The average S. pombe cDNA length was 470 nt.

3. Drosophila melanogaster

About 55% of the cDNA clones encoded ribosomal proteins, similar to the fraction of S. pombe clones. A very high percentage of the cDNAs were apparently full-length. However, there was an 11:1 ratio of small subunit to large subunit clones. The average fruitfly cDNA length was 535 nt.

Summary

A modified cDNA cloning technique allows novel cDNAs to be cloned from three model organisms. Many of the newly described RNAs are worthy of further study. We did not establish that any of these new RNAs are functional, but many are similar to RNAs described in the current literature.

WEB Resources

The University of Tokyo yeast genome browser: http://yeast.utgenome.org/ The SGD Saccharomyces browser: http://www.yeastgenome.org/

The S. pombe browser: http://www.genedb.org/perl-gb/gbrowse/S.pombe/ The D. melanogaster browser: http://flybase.bio.indiana.edu/cgi-bin/gbrowse/dmel/