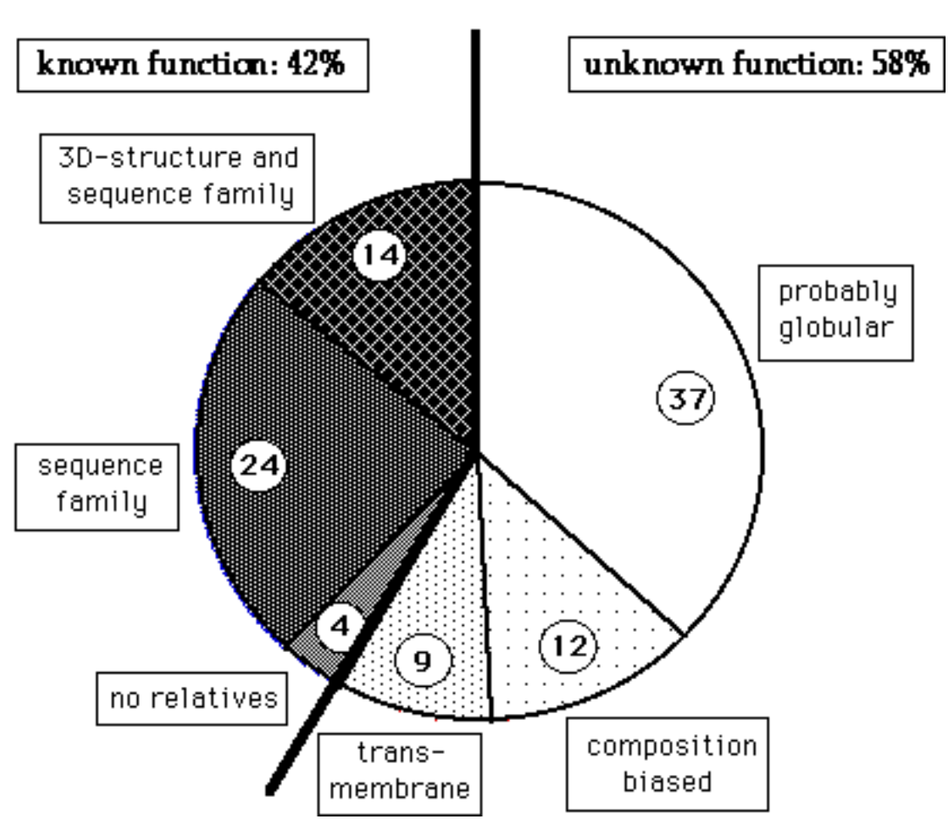
Nature, Vol 358, 23 July, 1992.

What's in a genome?

SIR -- We have taken up the challenge to help elucidate the function of the 182 predicted protein products derived from the complete DNA sequence of yeast chromosome III, a result of the European yeast genome project (Oliver et al. *Nature* **357**, 38-46; 1992). Some functional information is available for about 57 of these proteins, either determined by experiment or deduced by similarity searches in sequence databases.

We have identified the probable function of 17 additional protein products, using a combination of low-stringency sequence database searches, various ways of assessing significance, multiple sequence alignment, pattern searches and incorporation of prior knowledge about protein and domain families. The most interesting of these include a DNA polymerase of type X previously only found in mammalia, a new regulatory domain common to eukaryotes and prokaryotes (PILB), a methyltransferase, an acetolactate synthase and a GAL4-type transcriptional activator (see Table). In addition, we have determined that 25 of the chromosome III proteins have homologues of known three-dimensional structure. Adding it all up, as many as 42% of all proteins of yeast chromosome III have a known or probable function and 13% have an indirectly known three-dimensional structure. Of the remaining 58% of unknown function, about one third have one or more probable transmembrane segments (see Figure).



Information clock of yeast chromosome III proteins. Information accumulated to date by all methods, experimental and theoretical. Information content increases counterclockwise. The principal division is between known and unknown biological function. Numbers in percent. Composition bias indicates unusual amino acid composition, untypical of globular proteins, e.g., in coiled coils. The categories are approximate, but give an impression of the current state of the art.

Extrapolating from chromosome III to the entire yeast genome, we can expect that the white, unchartered, areas cover about 1/2 of the protein function map and 6/7 of the protein structure map. As genome projects pick up speed, more and more gene sequences need to be 'explained'. The challenge is clear: we must develop more efficient experimental and informatics methods for the determination of protein function and structure -- in parallel with genome sequencing projects.

Similarity of selected Chromosome III products to other proteins

---

ORF	length	family	closest	%id/len
YCL9c	309	prokaryotic acetolactate synthases, small subunit	ILVH_ECOLI	36%/170
YCL19w	1347	transposon B gene family,	COPI_DROME	18%/505
		related to pol genes	POLX_TOBAC	20%/393
YCL20w	438	transposon A gene family	YTY1_YEAST	49%/439
YCL33c	168	repressor of pilin promoter	PILB_NEIGO	33%/110
YCL75w	146	pol-like protein	S00954(P)	40%/74
YCR14c	582	type X DNA polymerases	DPOB_RAT	26%/393
YCR23c	611	tetracycline resistance proteins	TCR1_ECOLI	28%/150
YCR26c	743	mammalian PC1 plasma cell membr. prot.	PC1_HUMAN	38%/129
		phosphodiesterase family	PPD1_BOVIN	38%/ 66
YCR32w	2167	hypothetical protein rel. to C-term. "CDC4"-like human fragm.	HSCDC4A(E)	49%/316
YCR36w	333	ribokinase (other prok. sugar kinases)	RBSK_ECOLI	38%/96
YCR47c	275	ApoMet-methyltransferases	GLMT_RAT	26%/301
YCR64c	136	carboxypeptidases N	CBP8_HUMAN	27%/88
		dipeptidyl-peptidase IV	DPP_LACLA	25%/105
YCR69w	170	peptidyl-prolyl-cis-trans isomerases	CYPH_CANAL	37%/122
(YCR70w)1				
YCR72c	514	G-protein beta subunits	PR04_YEAST	23%/278
			TUP1_YEAST	32%/110
YCR98c	518	sugar transporter/symporter	A40260(P)	25%/179
YCR104w	124	glucose repressor/cold shock inducib.	SRP1_YEAST	27%/115
			SCTIPI(E)	27%/99
YCR106w	832	Gal4-like DNA/Zn binding domain	CYP1_YEAST	45%/47
			GAL4_YEAST	19%/168

Yeast chromosome III gene products with newly identified similarity to other proteins. ORF: name of the predicted open reading frame (Oliver et a

---

Peer Bork, Christos Ouzounis, Chris Sander, Michael Scharf, Reinhard Schneider, Erik Sonnhammer

European Molecular Biology Laboratory

Heidelberg, Germany