

- Eliminate Structurally Characterized Regions (to what threshold?)

- Decide if going for typical or atypical (feasibility), lots of low-hanging v. a little high-up fruit. Then refine, rank...

- Rank based on phylogenetic distribution

- Rank based on membership in interesting expression cluster (yeast)

- Rank based membership in phenotypes cluster (essential, assoc. with cold sensitivity, &c)

Target Selection Strategies

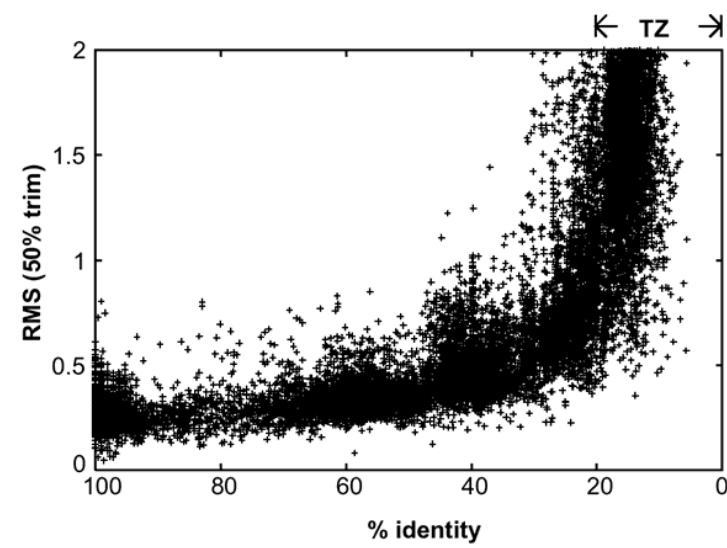
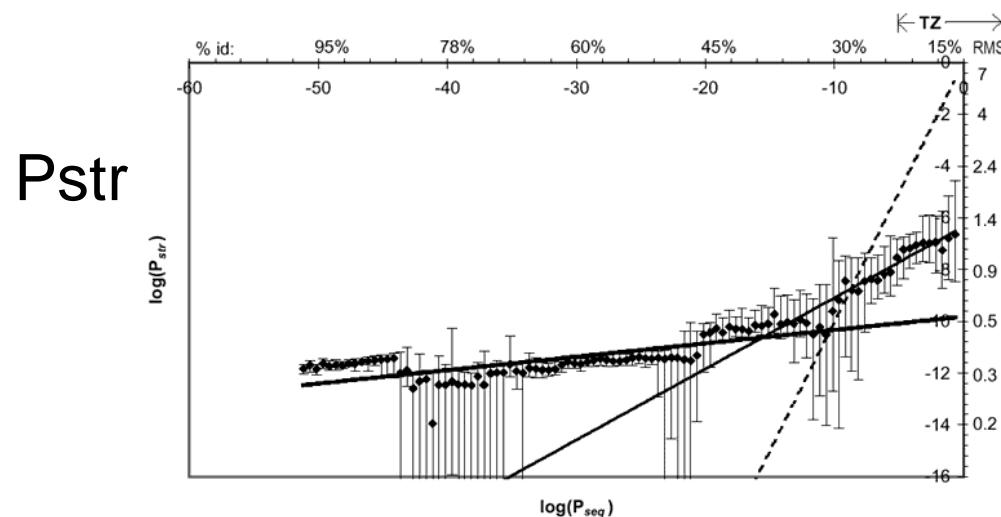
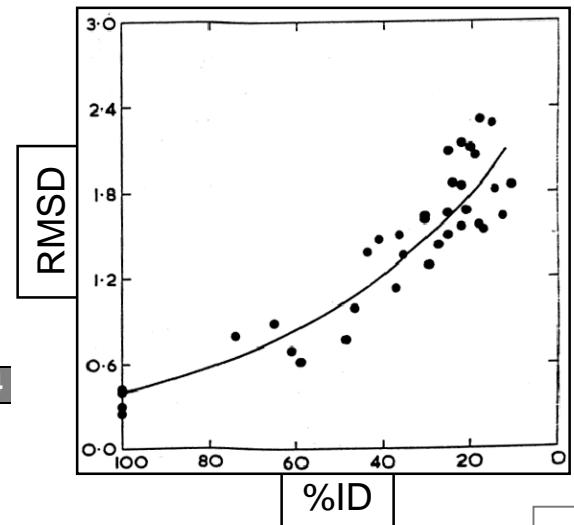
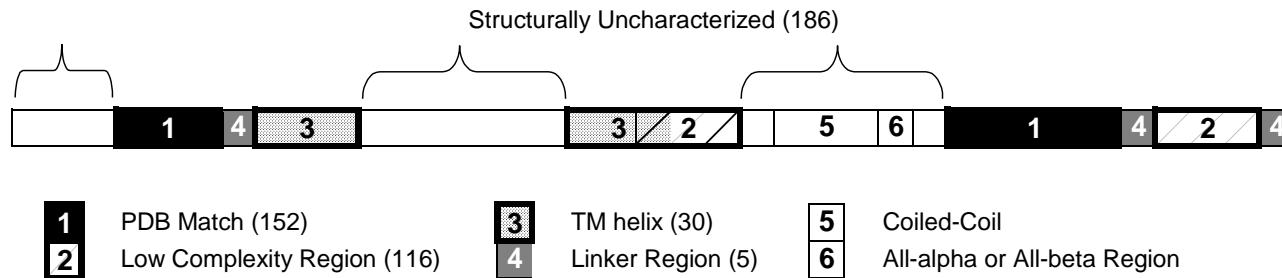
Not Discussed

Structure Prediction -- solve structures predicted to be interesting

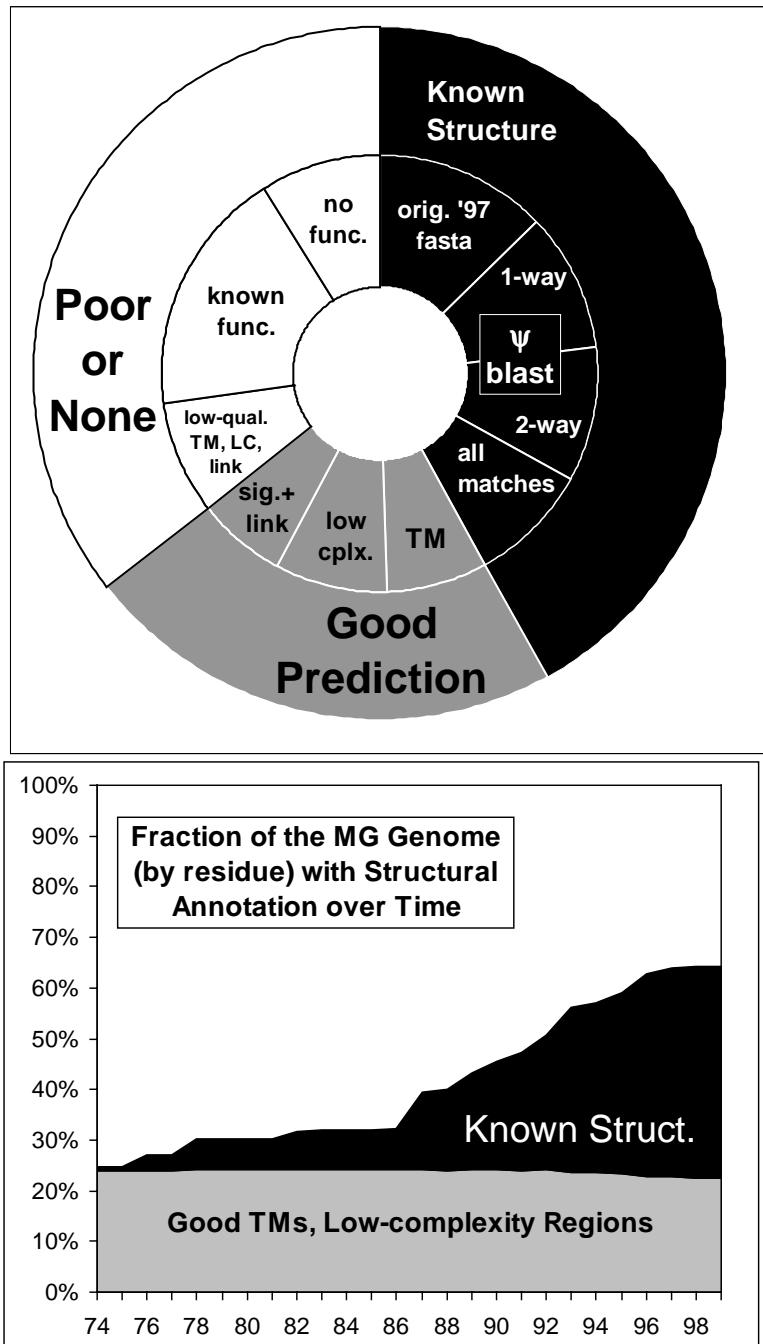
Relationship to disease (human homologs, pathogens)



What Parts of the Genome are Structurally Uncharacterized?

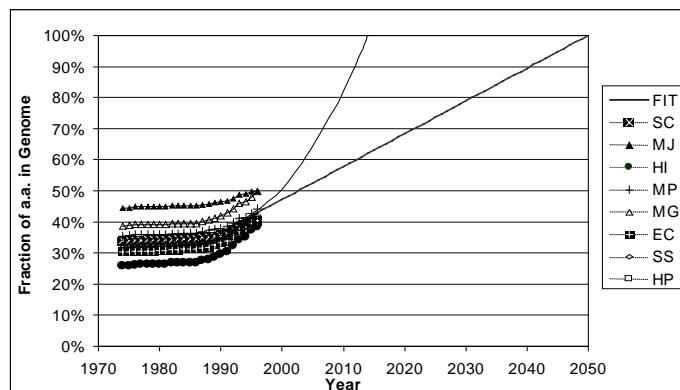
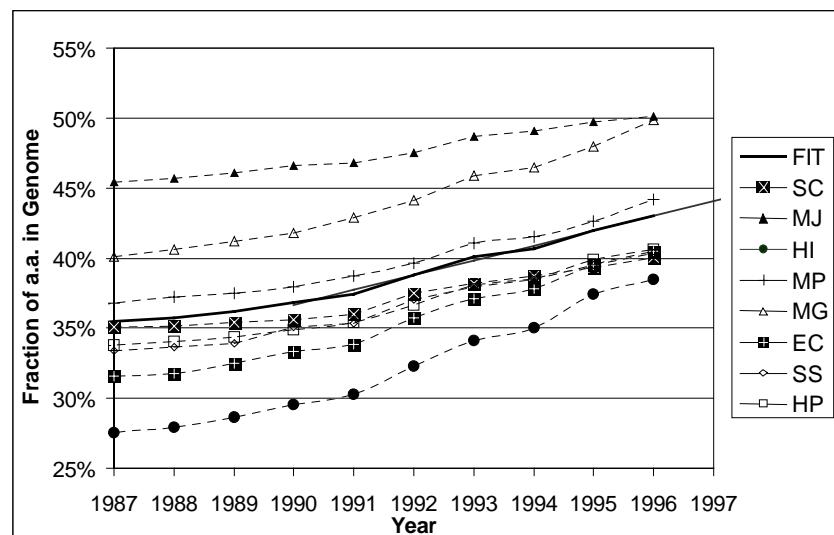


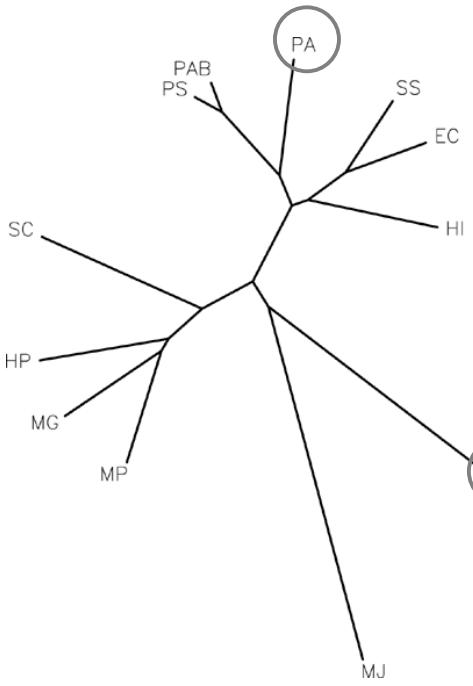
Pseq



How well are we doing?

Optimist vs. Pessimist

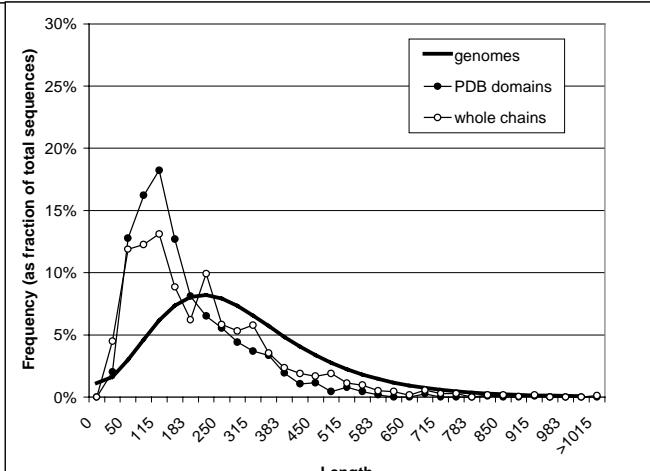
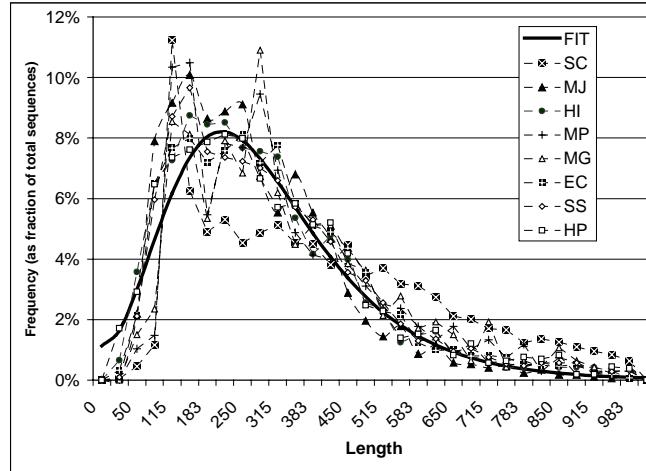




Name	Soluble PDB	= all- β	+ all- α
A	8.40%	6.8%	9.2%
C	1.72%	1.6%	1.4%
D	5.91%	5.9%	5.8%
E	6.29%	5.2%	7.3%
F	3.94%	4.2%	4.2%
G	7.79%	8.4%	6.4%
H	2.19%	2.1%	2.2%
I	5.54%	5.4%	5.1%
K	6.02%	5.6%	6.5%
L	8.37%	7.3%	9.6%
M	2.15%	1.7%	2.4%
N	4.57%	5.3%	4.4%
P	4.70%	5.1%	4.4%
Q	3.73%	3.5%	4.2%
R	4.78%	4.2%	5.4%
S	5.97%	7.2%	5.7%
T	5.87%	7.2%	5.2%
V	6.96%	7.6%	5.7%
W	1.46%	1.7%	1.5%
Y	3.64%	3.8%	3.5%

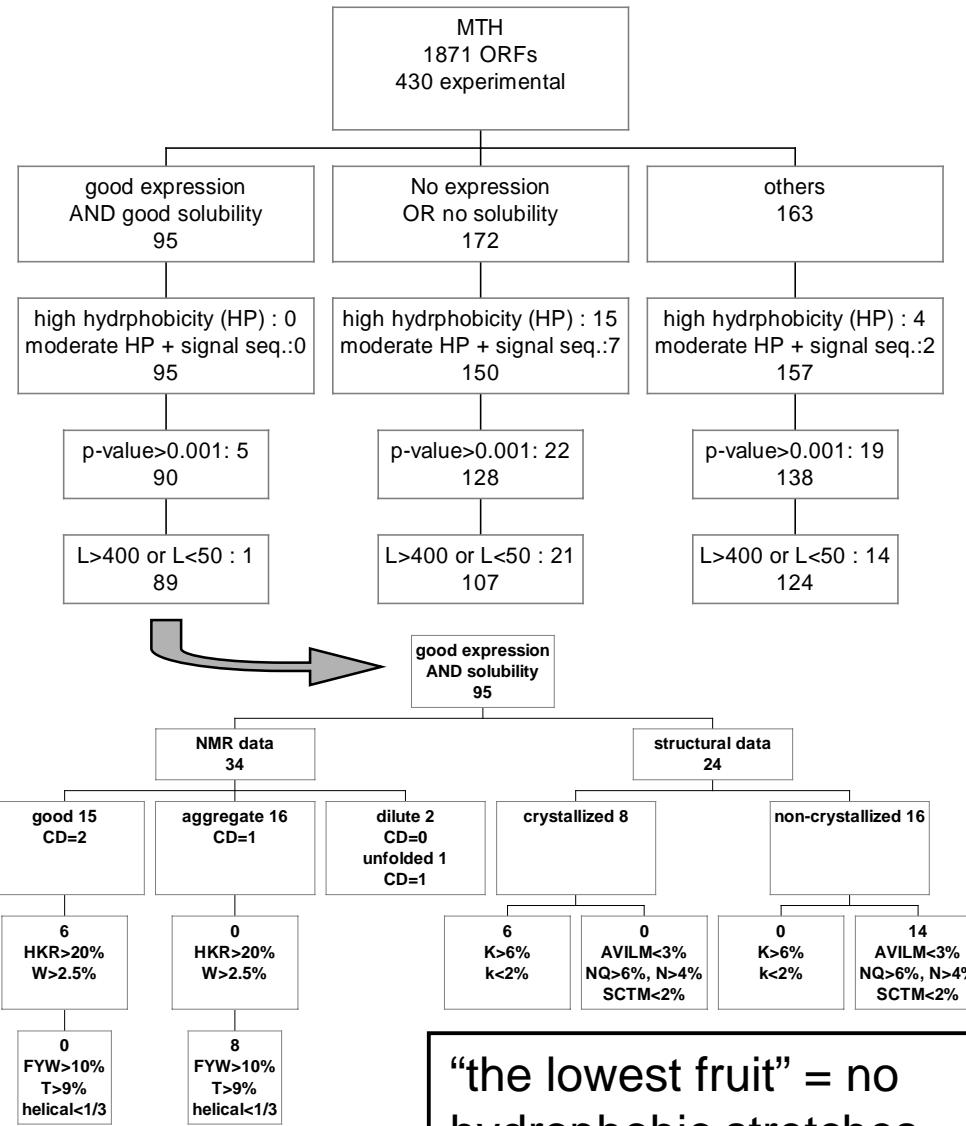
Focus on Typical or Atypical Proteins?

ABS.	rms	K	I	C	Q	W	N	F	L	G	A	P	S	R	H	M	E	D	T	Y	V
EC		4.4	6.0	1.2	4.4	1.5	4.0	3.9	10.6	7.4	9.5	4.4	5.8	5.5	2.3	2.8	5.7	5.1	5.4	2.9	7.1
HI		6.3	7.1	1.0	4.6	1.1	4.9	4.5	10.5	6.6	8.2	3.7	5.8	4.5	2.1	2.4	6.5	5.0	5.2	3.1	6.7
SS		4.2	6.3	1.0	5.6	1.6	4.0	4.0	11.4	7.4	8.5	5.1	5.8	5.1	1.9	2.0	6.0	5.0	5.5	2.9	6.7
SC		7.3	6.6	1.3	3.9	1.0	6.1	4.5	9.6	5.0	5.5	4.3	9.0	4.5	2.2	2.1	6.5	5.8	5.9	3.4	5.6
HP		8.9	7.2	1.1	3.7	.7	5.9	5.4	11.2	5.8	6.8	3.3	6.8	3.5	2.1	2.2	6.9	4.8	4.4	3.7	5.6
MP		8.6	6.6	.8	5.4	1.2	6.2	5.6	10.3	5.5	6.7	3.5	6.5	3.5	1.8	1.6	5.7	5.0	6.0	3.2	6.5
MG		9.5	8.2	.8	4.7	1.0	7.5	6.1	10.7	4.6	5.6	3.0	6.6	3.1	1.6	1.5	5.7	4.9	5.4	3.2	6.1
MJ		10.4	10.5	1.3	1.5	.7	5.3	4.2	9.5	6.3	5.5	3.4	4.5	3.8	1.4	2.2	8.7	5.5	4.0	4.4	6.9
Avg		7.5	7.3	1.1	4.2	1.1	5.5	4.8	10.5	6.1	7.0	3.8	6.4	4.2	1.9	2.1	6.5	5.1	5.2	3.3	6.4
SD		2.3	1.4	.2	1.3	.3	1.2	.8	.7	1.0	1.5	.7	1.3	.9	.3	.4	1.0	.3	.7	.5	.6
Diff.																					
EC	16	-25	8	-29	19	7	-15	-2	28	-6	13	-5	-3	16	3	28	-7	-14	-7	-22	1
HI	17	8	27	-38	24	-21	6	12	26	-15	-2	-20	-2	-6	-7	10	5	-17	-11	-14	-4
SS	20	-29	13	-39	49	9	-13	1	37	-6	1	11	-3	6	-15	-8	-2	-16	-6	-20	-4
SC	21	24	18	-21	5	-27	31	14	15	-36	-34	-7	51	-7	-2	-4	5	-4	0	-8	-20
HP	27	52	29	-34	0	-51	27	36	34	-26	-18	-29	14	-28	-4	2	11	-20	-25	1	-20
MP	28	45	18	-55	44	-17	35	41	24	-29	-20	-25	8	-27	-18	-28	-8	-17	2	-11	-7
MG	36	61	48	-50	27	-32	62	53	28	-41	-33	-36	11	-35	-28	-30	-8	-18	-8	-11	-12
MJ	38	77	88	-23	-61	-49	14	6	14	-19	-35	-28	-25	-20	-35	1	40	-8	31	20	-2
Avg		26	31	-36	13	-23	19	20	26	-22	-16	-17	6	-13	-13	-4	4	-14	-11	-8	-9
RMS		45	39	38	35	31	30	28	27	25	24	23	21	21	18	18	16	15	15	15	11

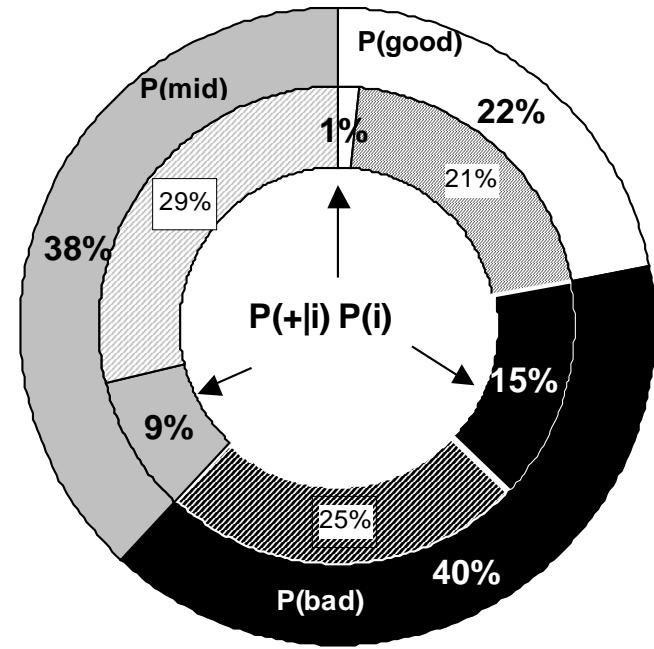


Typical: MT
(Edwards,
Arrowsmith)
Atypical: MG
(Regan)

How Representative are the Known Structures of the Proteins in Complete Genome?



Helping to Find the “Lowest-hanging Fruit,” Speeding High- Throughput Analysis



A Edwards
C Arrowsmith
500 proteins!

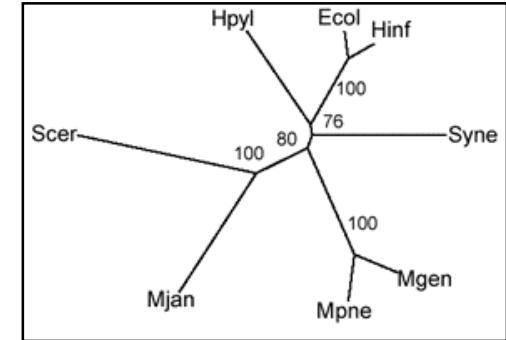
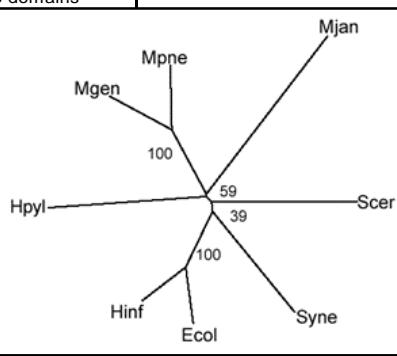
“the lowest fruit” = no hydrophobic stretches, homologs in DB, $50 < \text{len} < 400$ (low-complexity not useful, sec. struc.); find ~100 bads with 6 FPs

Also, compositions (HKR, α , NQ, FYW, SCTM) useful for finding those with good NMR or xtal results.

Id#	EC	SC	HI	SS	HP	MJ	MPMG	total	Fam.	PDB	Rep.	Struc.	Name
60	46	23	40	19	7	4	3	202	16	183	1xel	-	NAD(P)-binding Rossmann Fold
20	69	17	19	17	16	10	11		13	132	1gky	-	
37	28	18	16	12	40	3	3		23	160	1fxd	-	
45	36	13	22	11	10	5	4		37	399	1byb	-	
18	17	7	9	4	8	2	2		5	36	1pyd	a:2-181	
15	11	7	10	1	9	5	5		13	132	2tmd	a:490-645	
8	9	7	8	9	3	6	6		4	23	1sry	a:111-421	
7	10	8	8	4	4	3	3		5	19	1fnb	19-154	
13	7	4	3	3	3	3	3		18	177	1snc	-	
10	8	4	8	2	2	2	1		11	48	1igd	-	
9	10	5	5	2	2	2	2		7	19	1bdo	-	
5	5	4	4	5	6	3	3		3	22	2ts1	1-217	
10	4	2	4	2	2	2	3		4	35	1zym	a:	
5	7	4	6	3	2	1	1		3	18	3pmg	a:1-190	The "swivelling" beta/beta/alpha domain
4	2	3	6	4	2	4	3		2	3	1mat	-	Phosphoglucomutase, first 3 domains
6	4	3	4	4	1	1	1		42	1gad	o:149-312	like G3P dehydrogenase, C	Creatinases/methionine amin
5	4	4	1	2	2	1	2		23	1fkd	-	FKBP-like	
3	3	3	3	1	3	1	1		16	1opr	-	Phosphoribosyltransferases	
1	9	1	2	1	1	1	1		23	1oel	a:(*)	GroEL, the ATPase domain	
2	2	2	4	2	1	2	2		5	1dar	477-599	Ribosomal protein S5 doma	
4	3	2	2	1	1	2	2		50	3grs	364-478	FAD/NAD-linked reductases	
3	4	3	1	2	1	1	1		12	1kpa	a:	HIT-like	
4	2	3	1	2	1	1	1		10	1lub	-	Purine and uridine phosphor	
3	1	3	3	2	1	1	1		3	1tig	-	IF3-like	
2	3	1	2	2	1	1	1		4	1stu	-	dsRBD & PDA domains	
2	5	1	1	1	1	1	1		26	1one	a:1-141	like Enolase, Nt-dom.	
2	1	2	1	2	2	1	1		1	1	1ecl	-	type I DNA topoisomerase
1	3	1	1	1	1	1	1		1	1	1whi	-	Ribosomal protein L14
2	2	1	1	1	1	1	1		10	10	1ltrk	a:535-680	Transketolase, Ct-dom.
1	1	1	1	1	1	1	1		4	3pgk	-	Phosphoglycerate kinase	
49	8	14	57	12	5		1		15	100	3chy	-	Flavodoxin-like
24	54	15	11	4		4	5		19	112	2rn2	-	Ribonuclease H-like motif
7	18	6	9	4			5	5					
14	13	3	3	2			2						
7	13	4	10	2			1						
8	4	2	4	2			2						
3	6	3	3	3			3	3					
2	1	3	5	6			4	2					
4	1	4	3	2			4	4					
3	10	2	3	1			1	1					
4	2	3	3	3			2	1					
4	2	4	3	2			1	1					
3	6	1	2	1			2	1					
3	2	3	3	2			2	1					
2	2	2	2	2			2	2					
4	4	1	2	1			1	1					
4	1	2	3	1			1	1					
2	1	2	2	2			1	1					
2	2	1	2				1	1					
2	1	2	1	1			1	1					
2	2	1	1	1			1	1					
1	3	1	1				1	1					
1	1	1	1	1			1	1					
1	1	1	1	1			1	1					

Rank Based on Phylogenetic Distribution: Normal or Unique?

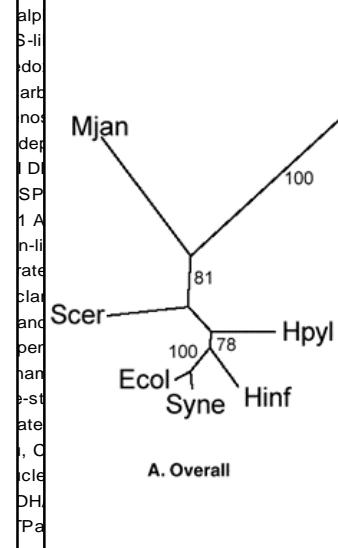
The "swivelling" beta/beta/alpha domain
Creatinases/methionine amin
like G3P dehydrogenase, Ct
FKBP-like
Phosphoribosyltransferases
GroEL, the ATPase domain
Ribosomal protein S5 doma
FAD/NAD-linked reductases
HIT-like
Purine and uridine phosphor
IF3-like
dsRBD & PDA domains
like Enolase, Nt-dom.
type I DNA topoisomerase
Ribosomal protein L14
Transketolase, Ct-dom.
Phosphoglycerate kinase



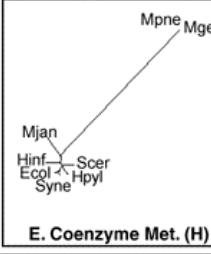
TIM

ribosomal

Homolog,
Ortholog
COGs
Sequence
similarity
of indiv.
protein



A. Overall



E. Coenzyme Met. (H)

B. meta.
info.
cell.

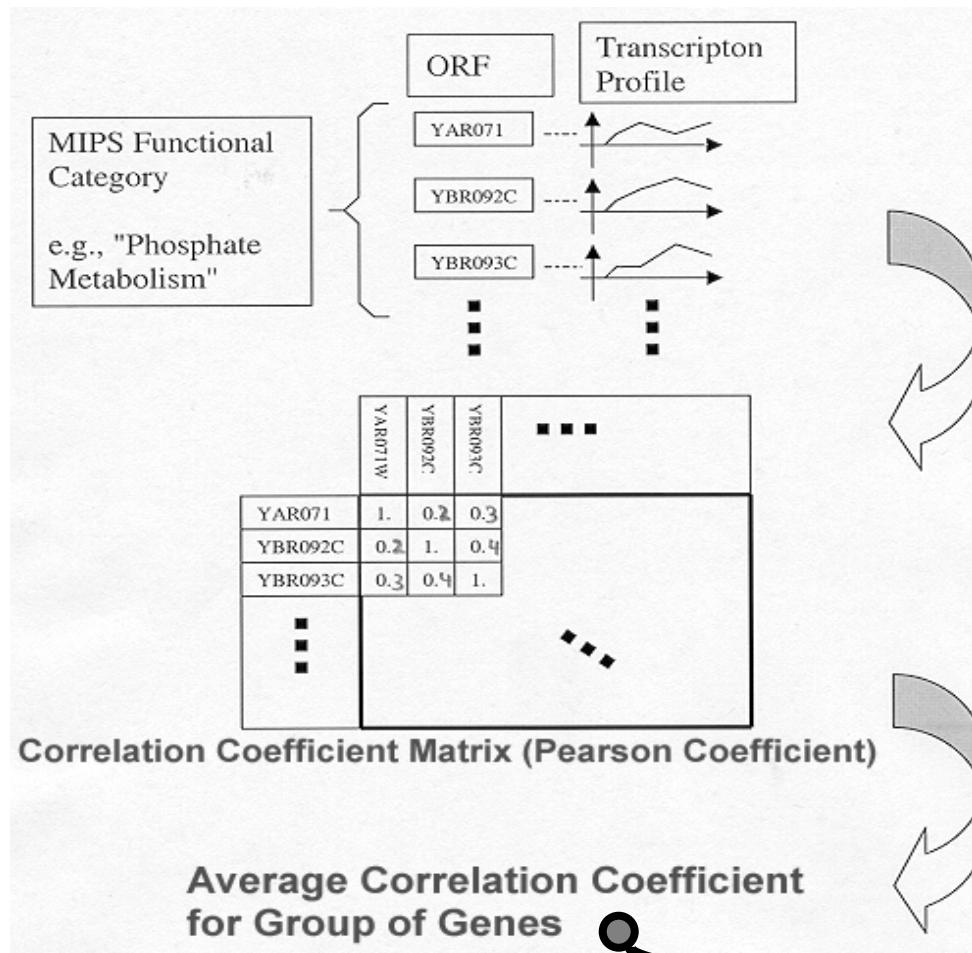
D.

F. Transcription (J)

C.

Functional category number	Function	Average correlation	# ORFs
01	METABOLISM	0.1001	1005
01.01	amino-acid metabolism	0.1488	199
01.01.01	amino-acid biosynthesis	0.239	114
01.01.04	regulation of amino-acid metabolism	0.23	32

MIPS YFC: 66 bottom classes, 10 top classes
 Average correlation of uncharacterized genes is 0.16
 Similar to Botstein analysis.

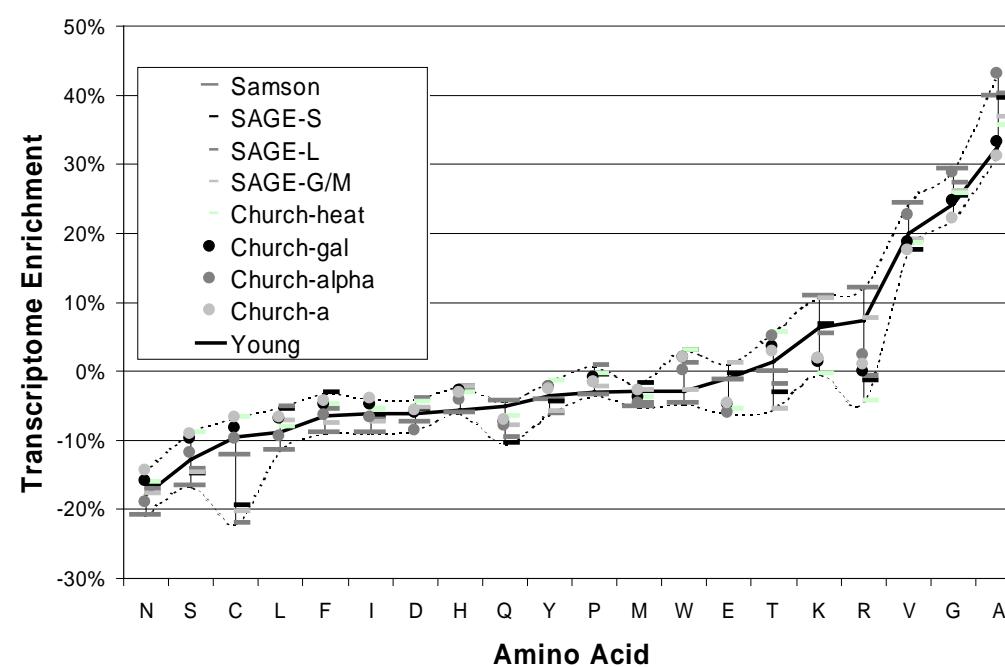
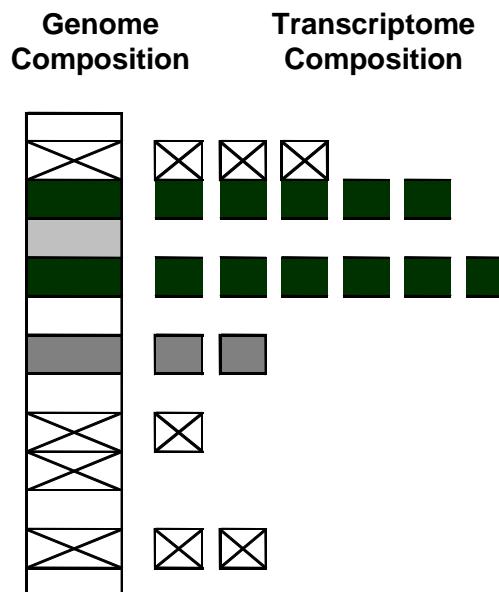


Cluster Yeast ORFs
based on Expression
Level, Changes in
Expression, Other whole
genome data... Rank
based on cluster

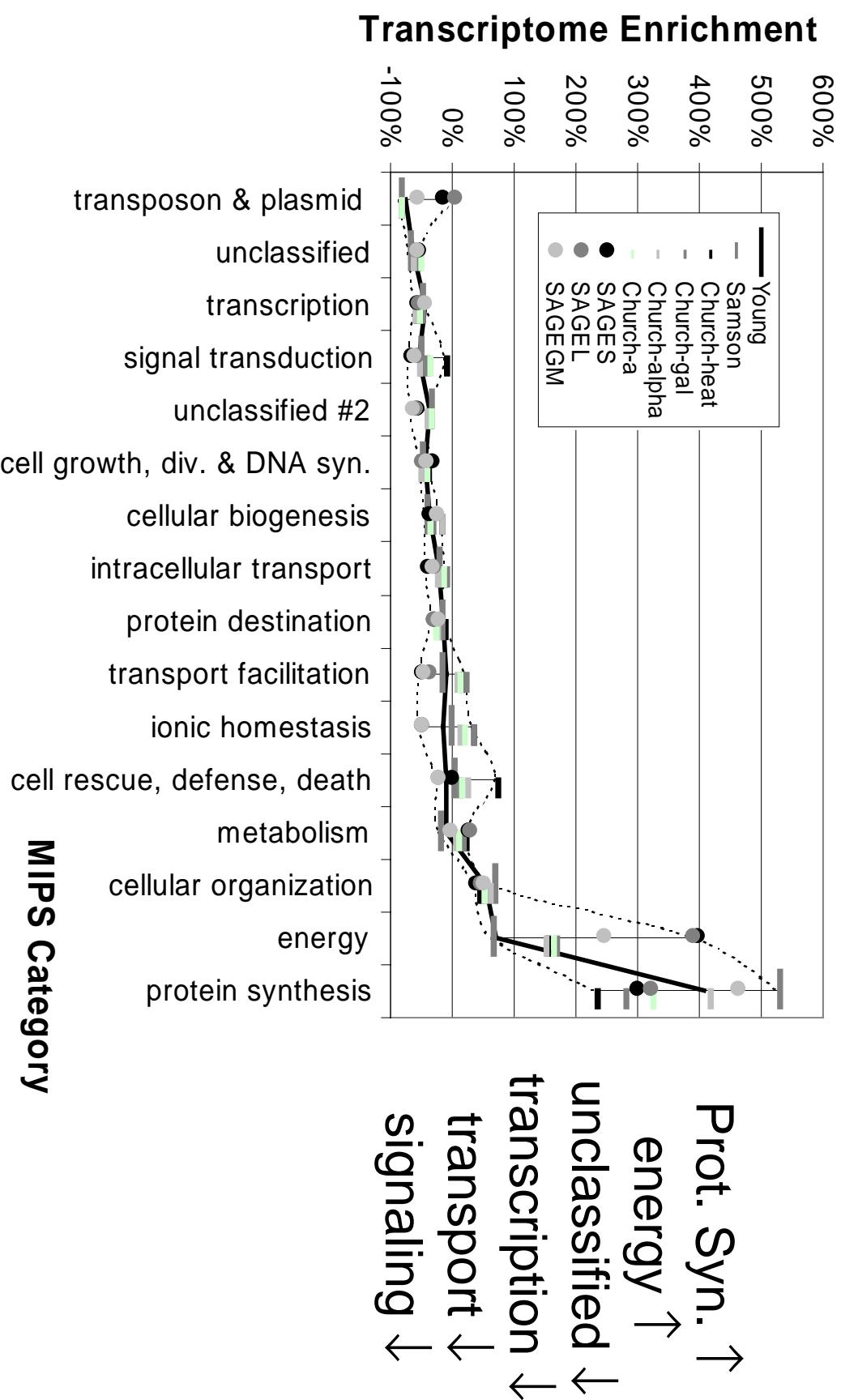
Functional category number	Function	Average correlation	# ORFs
01	METABOLISM	0.1001	1005
01.01	amino-acid metabolism	0.1488	199
01.01.01	amino-acid biosynthesis	0.239	114
01.01.04	regulation of amino-acid metabolism	0.23	32
01.01.07	amino-acid transport	0.1198	23
01.01.10	amino-acid degradation	0.0524	36
01.01.99	other amino-acid metabolism activities	0.2205	4
01.02	nitrogen and sulphur metabolism	0.1869	73
01.02.01	nitrogen and sulphur utilization	0.0726	37
01.02.04	regulation of nitrogen and sulphur utilization	0.3715	28
01.02.07	nitrogen and sulphur transport	0.2829	8
01.03	nucleotide metabolism	0.1708	134
01.03.01	purine-ribonucleotide metabolism	0.3639	42
01.03.04	pyrimidine-ribonucleotide metabolism	0.176	28
01.03.07	deoxyribonucleotide metabolism	0.1095	12
01.03.10	metabolism of cyclic and unusual nucleotides	0.2848	8
01.03.13	regulation of nucleotide metabolism	0.2696	13
01.03.16	polynucleotide degradation	0.2461	7
01.03.19	nucleotide transport	0.1187	12
01.03.99	other nucleotide-metabolism activities	-0.0328	7
01.04	phosphate metabolism	0.1348	31
01.04.01	phosphate utilization	0.16	13
01.04.04	regulation of phosphate utilization	0.0399	8
01.04.07	phosphate transport	0.0724	10
01.05	carbohydrate metabolism	0.0779	409
01.05.01	carbohydrate utilization	0.075	256
01.05.04	regulation of carbohydrate utilization	0.1174	120

Composition of Genome vs. Transcriptome

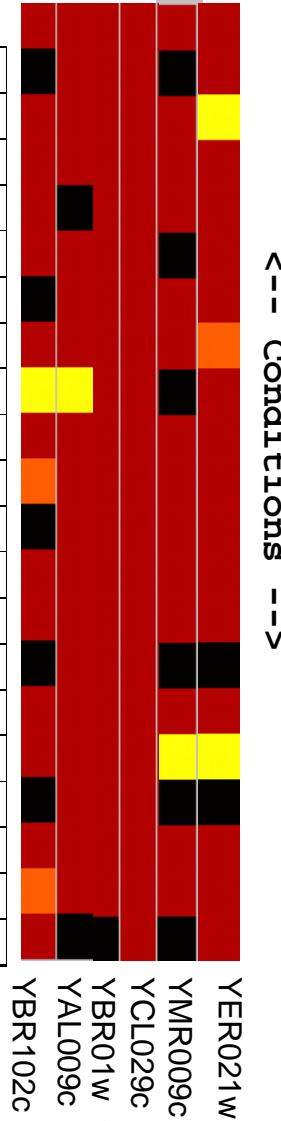
	$\sum_{\text{orf } i} n_i(F)$	$\sum_{F \text{ orf } i} n_i(F)$	$G(F)$	$\sum_{\text{orf } i} e_i n_i(F)$	$\sum_{F \text{ orf } i} e_i n_i(F)$	$T(F)$	$D(F)$
Feature F is Amino acids, in particular Ala	Number of Ala in yeast	Number of amino acids in yeast	Genome composition of Ala in yeast	Number of Ala weighted by expression	Number of amino acids weighted by expression	Transcriptome composition of Ala in yeast	Relative enrichment of Ala in transcriptome
Spec. Num.	141890	2574876	5.5%	347807	4758441	7.3%	32.7%
Feature F is Folds, in particular the TIM-barrel (3.1)	Number of TIM-barrel fold matches in yeast genome	Number of matches with all folds in yeast genome	Genome composition of TIM-barrel fold matches	Number of TIM-barrel fold matches weighted by expression	Number of matches with all folds weighted by expression	Transcriptome composition of TIM-barrel fold matches	Relative enrichment of TIM-barrel matches in transcriptome
Spec. Num.	65	1560	4.2%	389	4709	8.3%	97.8%



Composition of Transcriptome in terms of Functional Classes



YPD + 8mM caffeine	Caff
Cycloheximide hypersensitivity: YPD + 0.08 g/ml cycloheximide at 30°C	Cyc ^S
White/ red color on YPD	W/R
YPGlycerol	YPG
Calcofluor hypersensitivity: YPD + 12 g/ml calcofluor at 30°C	Calc ^S
YPD + 46 g/ml hygromycin at 30°C	Hyg
YPD + 0.003% SDS	SDS
Benomyl hypersensitivity: YPD + 10 g/ml benomyl	Ben ^S
YPD + 5-bromo-4-chloro-3-indolyl phosphate 37°C	BCIP
YPD + 0.001% methylene blue at 30°C	MB
Benomyl resistance: YPD + 20 g/ml benomyl	Ben ^R
YPD at 37°C	YPD ³⁷
YPD + 2 mM EGTA	EGTA
YPD + 0.008% MMS	MMS
YPD + 75 mM hydroxyurea	HU
YPD at 11°C (COLD)	YPD¹¹
Calcofluor resistance: YPD + 66.7 g/ml calcofluor at 30°C	Calc ^R
Cycloheximide resistance: YPD + 0.3 g/ml cycloheximide	Cyc ^R
Hyperhaploid invasive growth mutants	HHIG
YPD + 0.9 M NaCl	NaCl



<-- Conditions -->

M Snyder

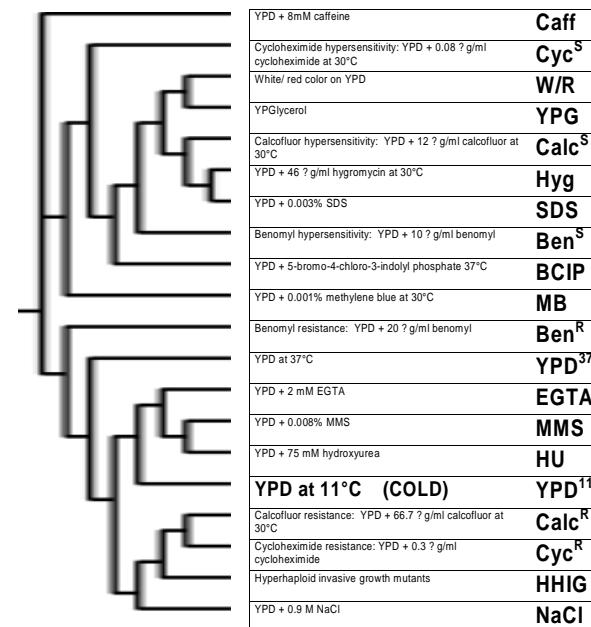
Affected by Another Condition

WT

Affected by Cold

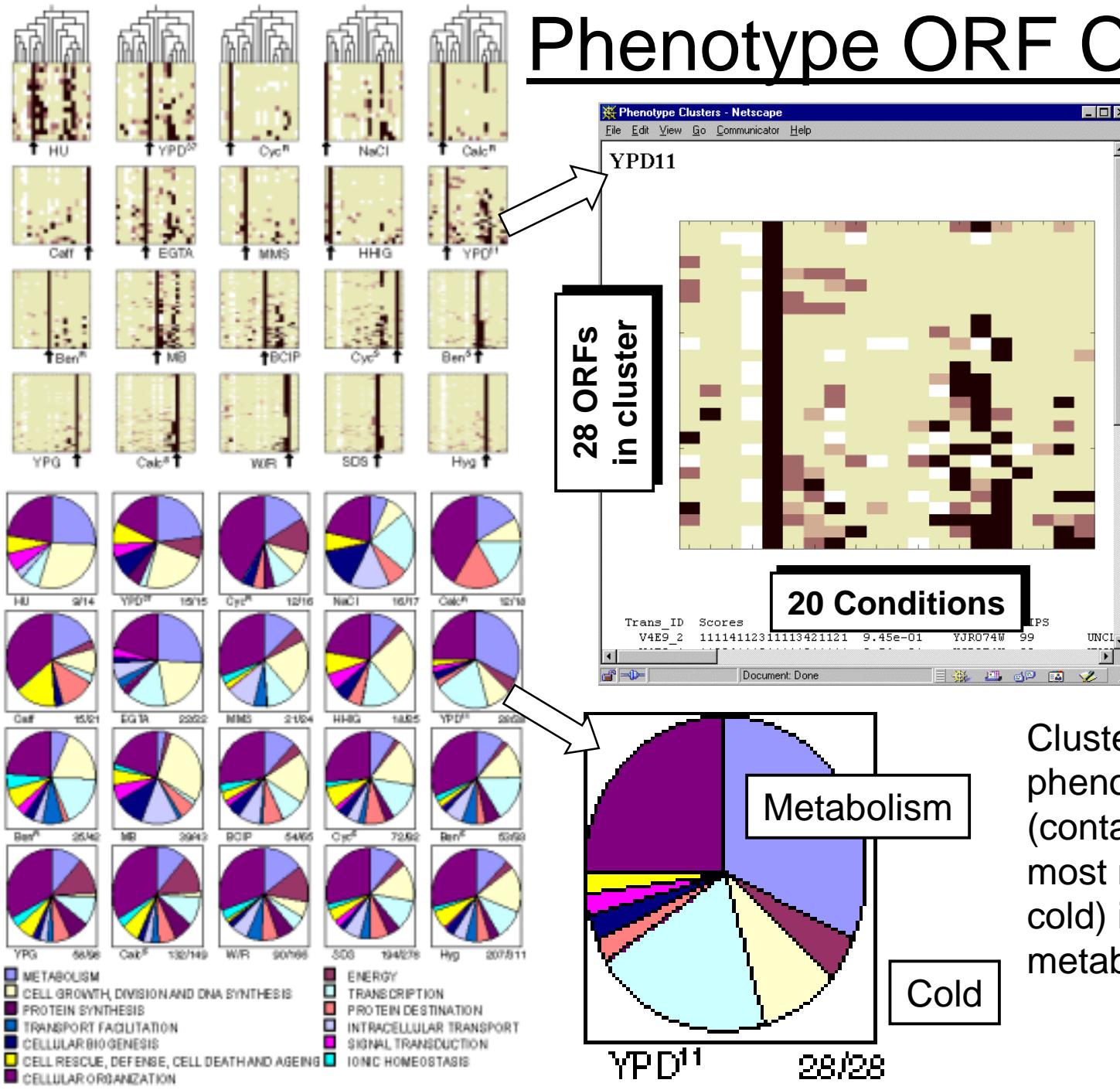
Whole Genome Phenotype Profiles

Transposon insertions into (almost) each yeast gene to see how yeast is affected in 20 conditions. Generates a phenotype pattern vector, which can be treated **similarly to expression data**



Clustering Conditions

Phenotype ORF Clustering



k-means clustering of ORFs based on “phenotype patterns,” cross-ref. to MIPs Functional Classes

- Eliminate Structurally Characterized Regions (to what threshold?)

- Decide if going for typical or atypical (feasibility), lots of low-hanging v. a little high-up fruit. Then refine, rank...

- Rank based on phylogenetic distribution

- Rank based on membership in interesting expression cluster (yeast)

- Rank based membership in phenotypes cluster (essential, assoc. with cold sensitivity, &c)

Target Selection Strategies

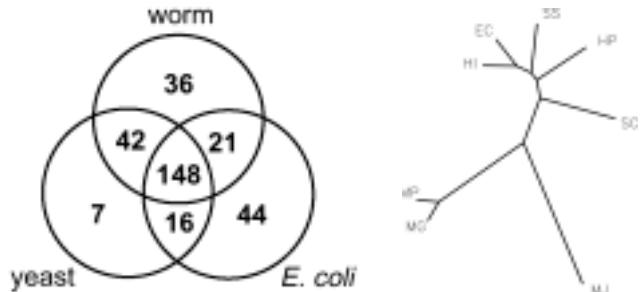
Not Discussed

Structure Prediction -- solve structures predicted to be interesting

Relationship to disease (human homologs, pathogens)



Integrated Database Analysis



Fold	Fold Class	Rep. PDB	Composition		
			Genome [%]	Transcriptome [%]	Rel. Diff. [%]
TIM barrel	α/β	1byb	4.2	8.3	+98
P-loop NTP hydrolases	α/β	1gky	5.8	5.2	-11
Ferrredoxin like	α/β	1fxd	3.9	3.4	-14
Rossmann fold	α/β	1xel	3.3	3.3	0
7-bladed beta-propeller	β	1mda*	6.4	2.9	-55
alpha-alpha superhelix	α	2bct	4.4	2.7	-37
Thioredoxin fold	α/β	2trx	1.7	2.7	+63
G3P dehydrogenase-like	α/β	1drw1	0.2	2.7	+1316
beta grasp	α/β	1lqd	0.6	2.6	+348
HSP70 C-term. fragment	multi	1dky	0.8	2.6	+231
long helices oligomers	α	1zta	3.8	2.1	-46
Protein kinases (cat. core)	multi	1hcl	6.8	1.6	-77
alpha/beta hydrolases	α/β	2ace	2.2	0.9	-62
Zn2/C6 DNA-bind. dom.	sml	1aw6	2.6	0.3	-89

Rank	Genome											
	Young	Sanson	Church-a	Church-gal	Church-neat	SAGE-GW	SAGE-L	SAGE-S	EC	HI	SS	SC
5	1	1	1	1	1	1	1	1	1	1	1	1
3	2	2	4	4	4	5	5	5	6	7	7	7
6	3	3	3	3	2	2	2	2	2	2	2	2
8	4	4	5	6	6	7	9	9	9	9	9	9
2	5	5	5	6	6	7	9	9	9	9	9	9
4	6	6	11	15	16	12	12	8	5	8	8	8
7	7	7	6	8	2	5	4	11	10	6	6	6
14	8	8	7	7	2	5	3	3	3	3	3	3
1	9	9	10	21	9	18	21	82	122	120	120	120
36	10	10	16	17	11	16	12	48	25	56	56	56
31	11	11	12	12	12	12	12	12	12	12	12	12
7	12	12	8	14	21	15	19	21	20	33	33	33
1	13	13	13	13	13	13	13	13	13	13	13	13
10	12	12	12	12	12	12	12	12	12	12	12	12
9	15	15	15	25	25	21	23	26	26	26	26	26
75	16	16	16	27	50	32	40	48	39	50	50	50

prototype: bioinfo.mbb.yale.edu/genome

- Typical Structural Genomics Analyses

- ◊ Shared and unique folds

- Relating Results Across Genomes
- Coverage of a number of organisms, biases

- ◊ Common folds

- Relationship to target selection (typical vs atypical)

- Integration with other whole genome data

- ◊ X-ref expression results, functional classifications

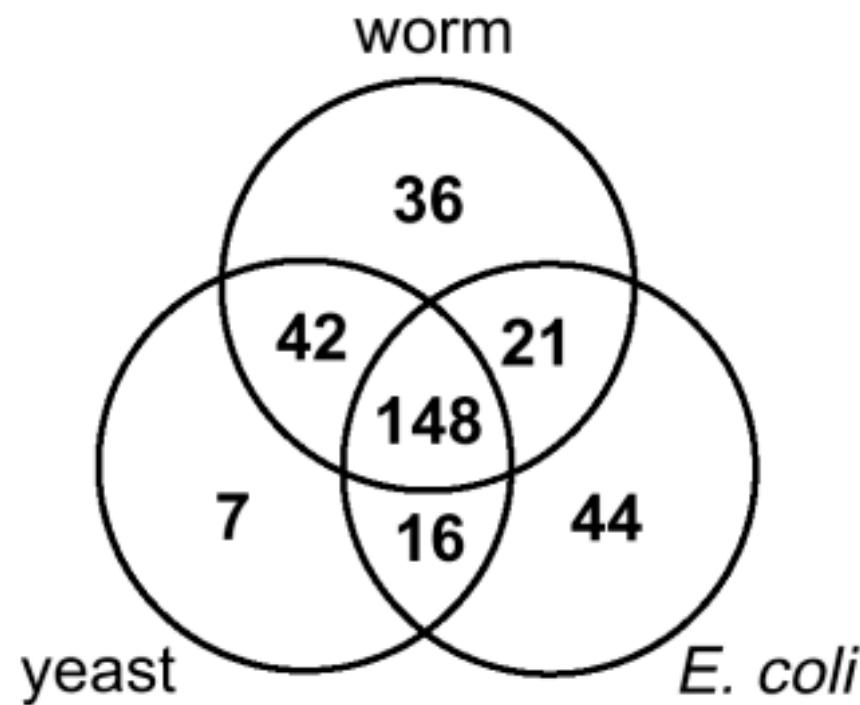
- ◊ Importance of selecting an organism for which there is much of this data

- Implementation issues

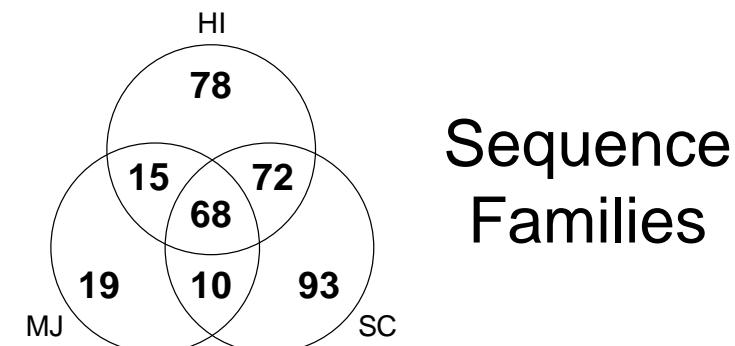
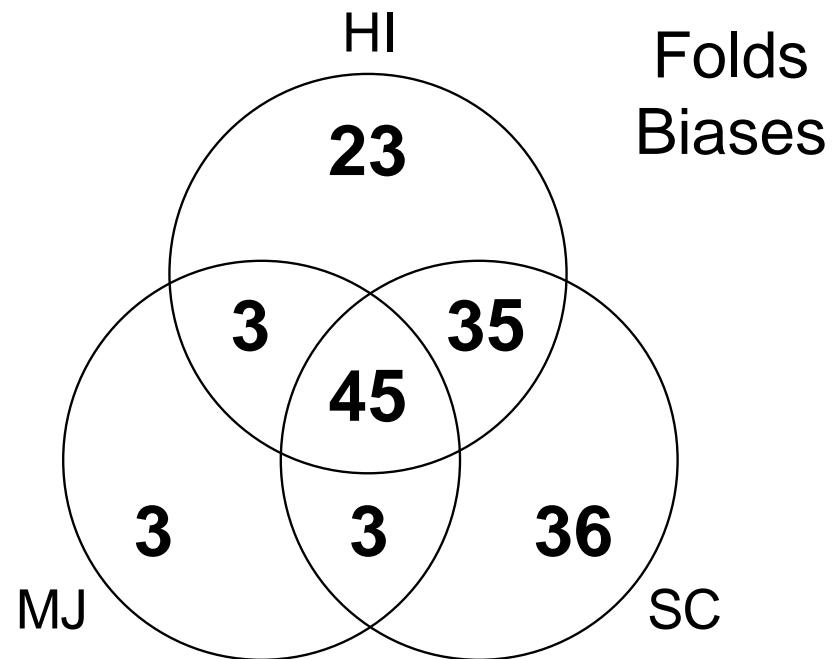
- ◊ Consistent identifier schemes

- ◊ History, fold and ORF assignments change

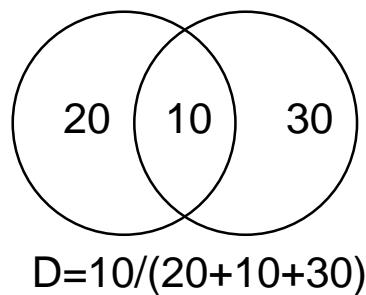
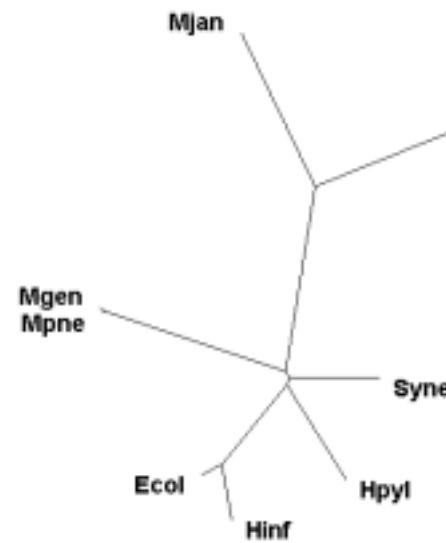
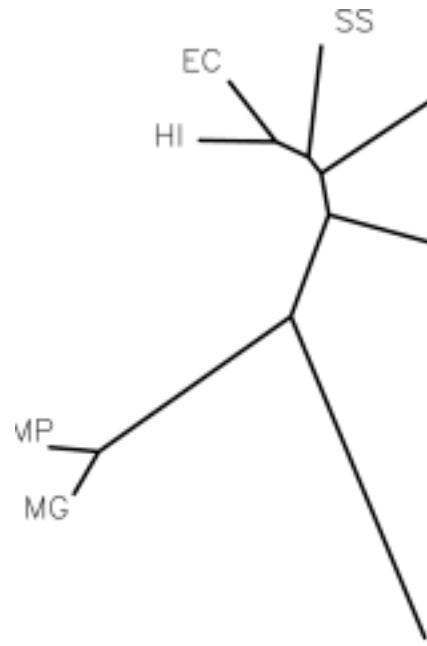
Comparing Genomes in terms of Protein Folds



of 339



Cluster Trees Grouping Initial Genomes on Basis of Shared Folds

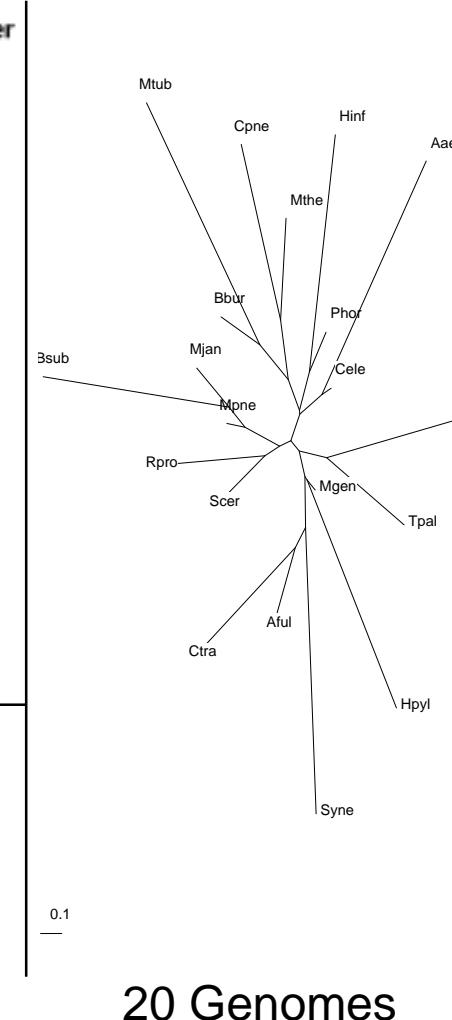


$$D = S/T$$

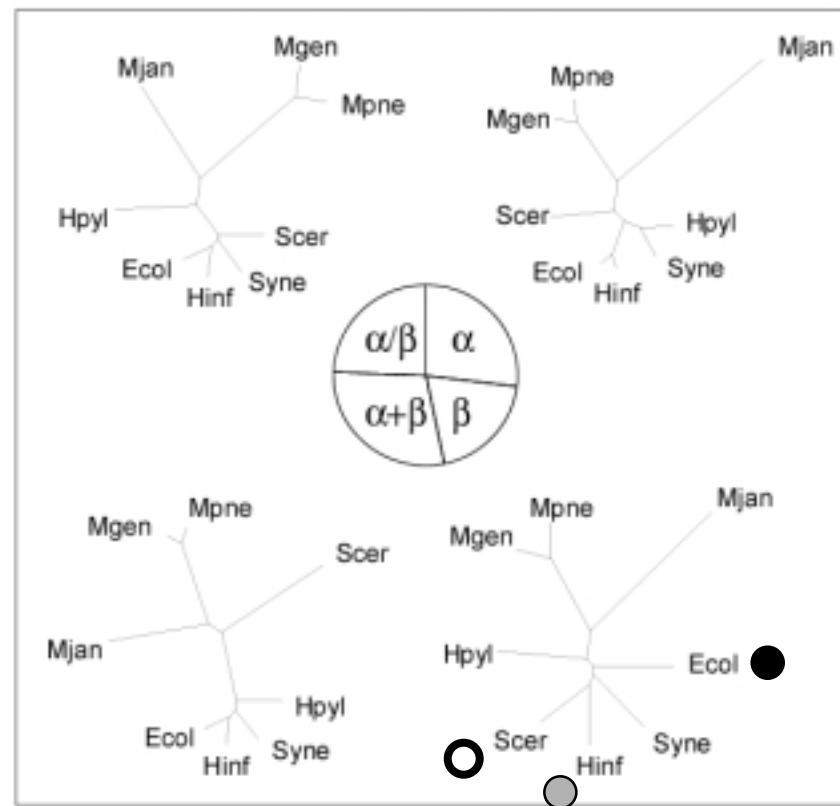
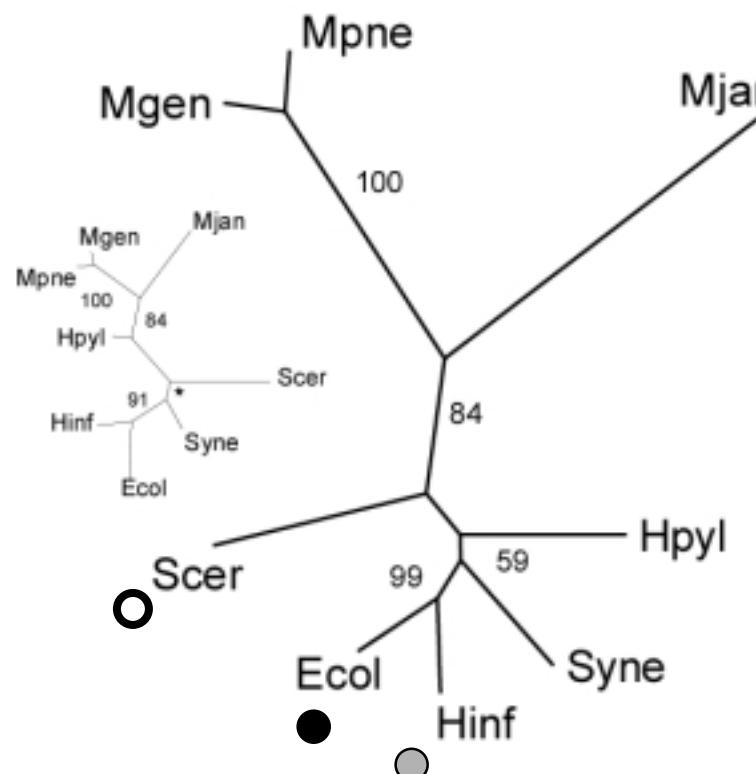
D = shared fold dist.
betw. 2 genomes

$$S = \# \text{ shared folds}$$

T = total #
folds in both

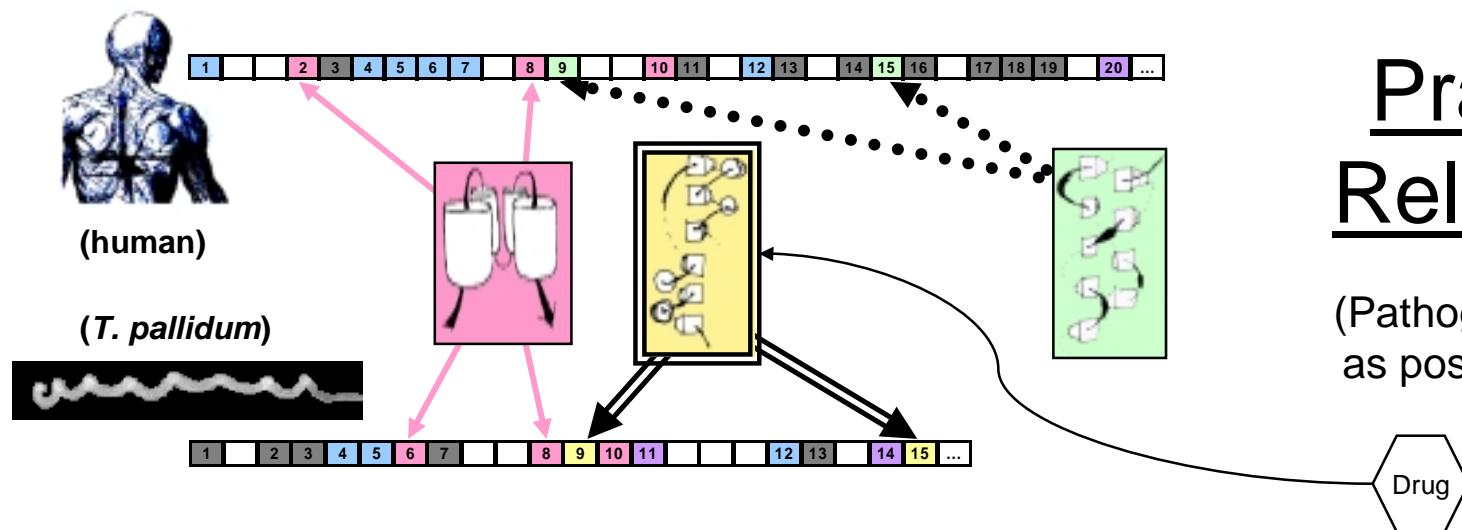
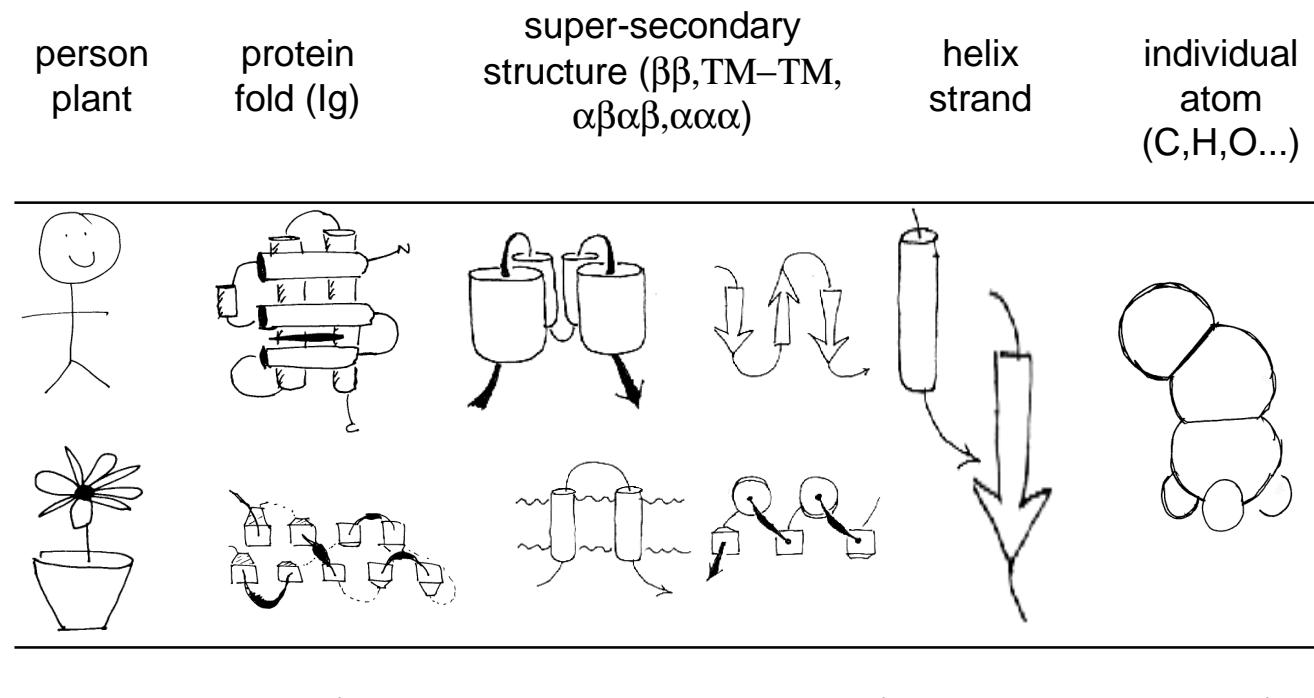


Distribution of Folds in Various Classes



Unusual distribution of all-beta folds

At What Structural Resolution Are Organisms Different?



Practical Relevance

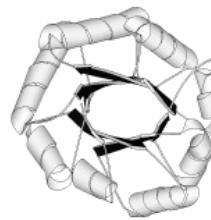
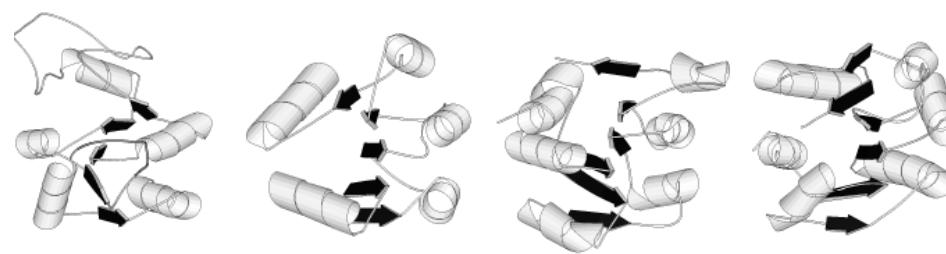
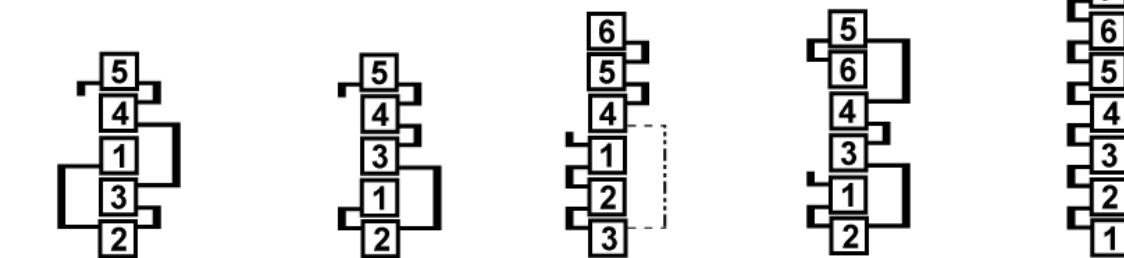
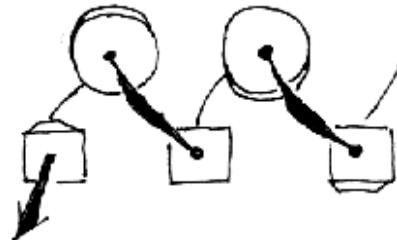
(Pathogen only folds as possible targets)

Top-10 Worm Folds

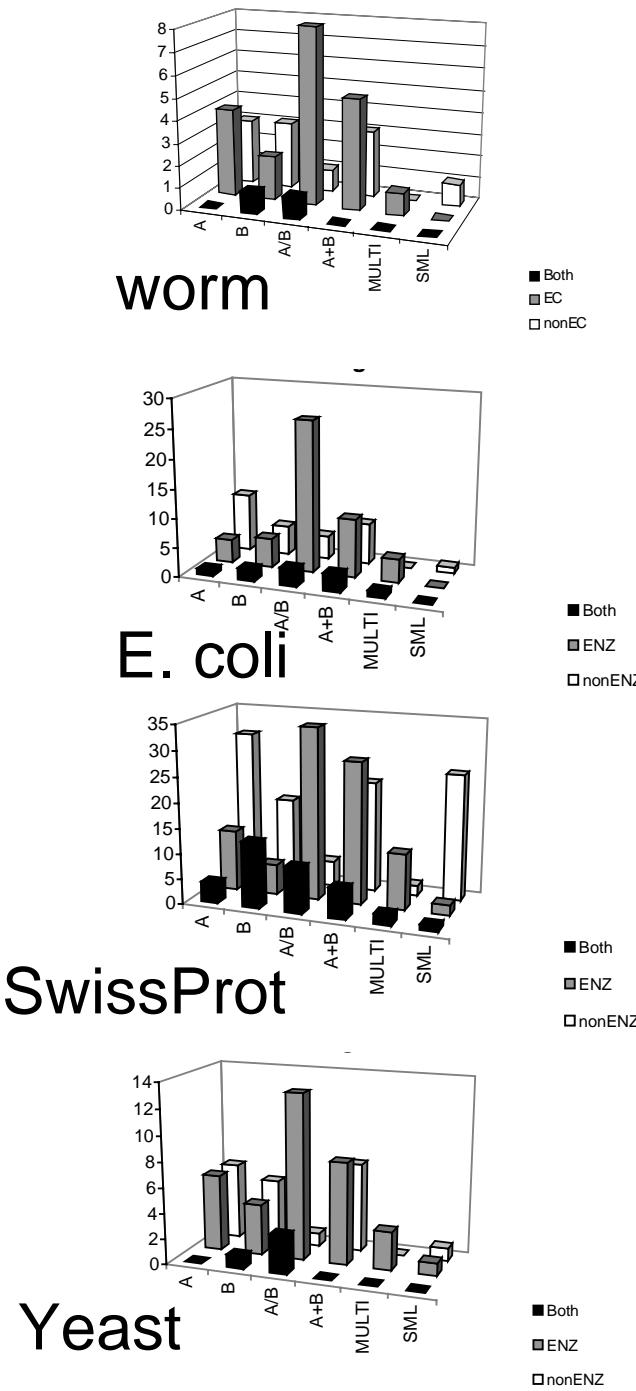
Top-10 Worm Folds		num. matches in worm genome (N)	frac. all worm dom. (F)	in EC?	in SC?
	class				
Ig	B	830	1.7%		
Knottins	SML	565	1.1%		
Protein kinases (cat. core)	MULT	472	0.9%		
C-type lectin-like	A+B	322	0.6%		
corticoid recep. (DNA-bind dom.)	SML	276	0.5%		
Ligand-bind dom. nuc. receptor	A	257	0.5%		
alpha-alpha superhelix	A	247	0.5%		
C2H2 Zn finger	SML	239	0.5%		
P-loop NTP Hydrolase	A/B	235	0.5%		
Ferrodoxin	A+B	207	0.4%		

Most Common Folds

$\beta\alpha\beta$ unit common in microbial genomes



TIM
Barrel



Integrated Databases:

Cross-reference

Structure Information

MIPS Functional Cat.

		A	B	A/B	A+B	MULTI	SML
metabolism	1	3.5	2.3	10	4.5	1.3	0.8
energy	2	1.1	1.2	5	1.5	0.3	0.2
growth, div., DNA syn.	3	4.9	3.6	4	4.5	1.8	1.2
transcription	4	1.5	1.3	2.2	1.5	0.5	0.8
protein synthesis	5	1	0.9	0.7	1.3	0.3	0.2
protein targetting	6	1.2	1.7	2	1.6	0.5	0.3
transport facilitation	7	0.9	0.5	0.7	0.6	0.4	
intracellular transport	8	1.8	2.1	1.6	0.6	1	
cellular biogenesis	9	0.9	0.7	1.2	0.3	0.3	0.1
signal transduction	10	1	1	1.1	0.3	0.7	0.3
cell rescue, defense...	11	1.5	1	2.6	1.9	0.7	0.5
ionic homeostatis	13	0.5	0.3	0.4	0.4	0.2	

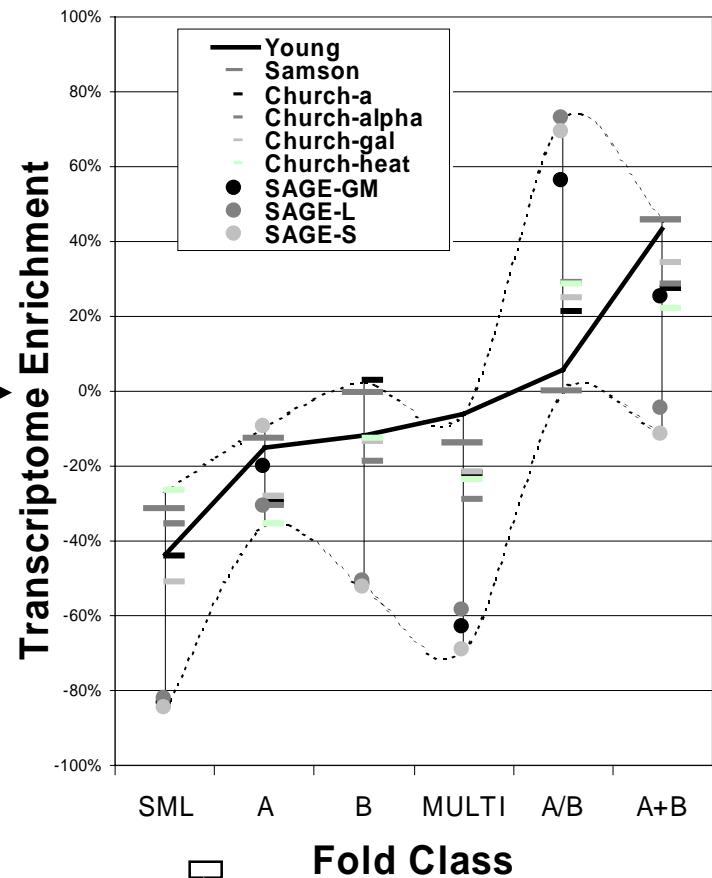
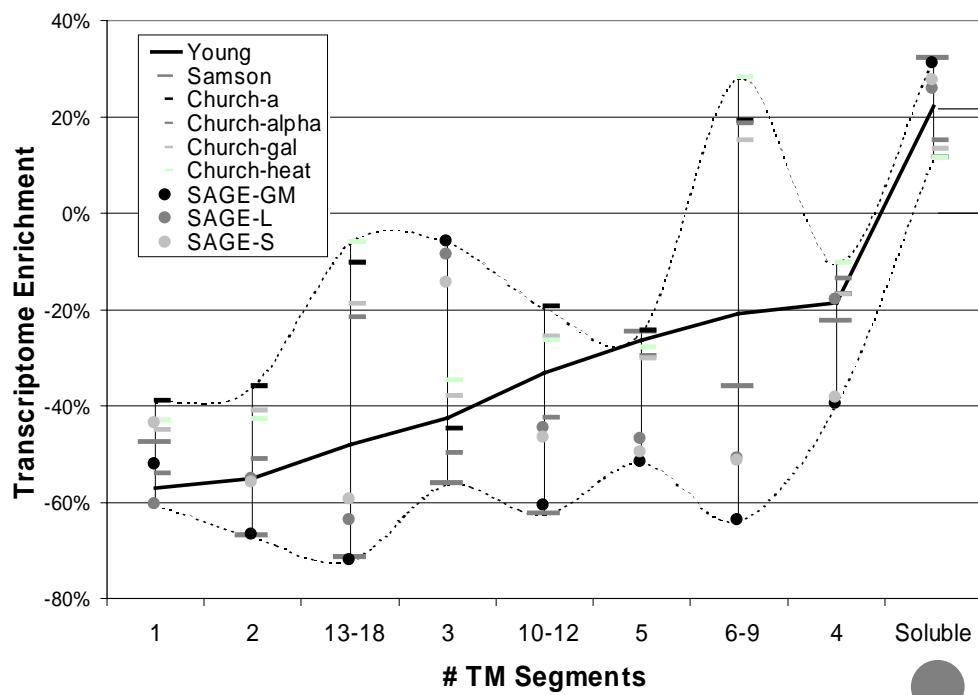
MIPS YFC (Mewes)

Which Protein Folds are Highly Expressed?

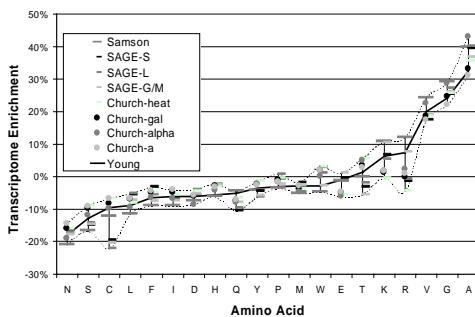
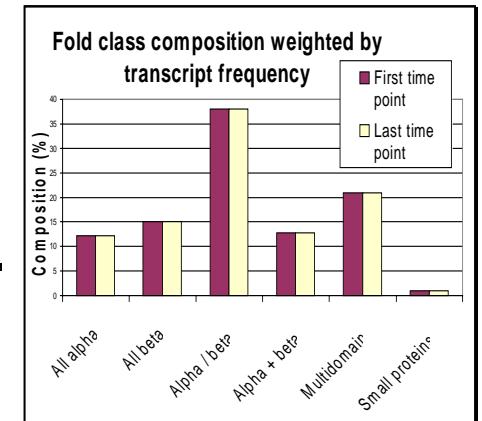
Fold	Fold Class	Rep. PDB	Composition			Rank										
			Genome [%]	Transcriptome [%]	Rel. Diff. [%]		Genome	Young	Samson	Church-a	Church-alpha	Church-gal	Church-heat	SAGE-GM	SAGE-L	SAGE-S
TIM barrel	α/β	1byb	4.2	8.3	+98	5	1	1	1	1	1	1	1	1	1	1
P-loop NTP hydrolases	α/β	1gky	5.8	5.2	-11	3	2	2	4	4	4	5	5	6	7	
Ferredoxin like	$\alpha\beta$	1fxd	3.9	3.4	-14	6	3	3	3	3	3	2	2	19	15	9
Rossmann fold	α/β	1xel	3.3	3.3	0	8	4	4	5	5	5	6	6	7	10	11
7-bladed beta-propeller	β	1mda*	6.4	2.9	-55	2	5	5	4	5	6	6	7	9	9	16
alpha-alpha superhelix	α	2bct	4.4	2.7	-37	4	6	6	11	15	16	12	12	8	5	8
Thioredoxin fold	α/β	2trx	1.7	2.7	+63	14	7	7	6	8	2	5	4	11	10	6
G3P dehydrogenase-like	$\alpha\beta$	1drwt	0.2	2.7	+1316	81	8	8	12	2	5	3	3	35	19	30
beta grasp	$\alpha\beta$	1igd	0.6	2.6	+348	36	9	9	10	21	9	18	21	82	122	120
HSP70 C-term. fragment	multi	1dky	0.8	2.6	+231	31	10	10	16	17	11	16	12	48	25	56
long helices oligomers	α	1zta	3.8	2.1	-46	7	15	15	8	14	21	15	19	21	20	33
Protein kinases (cat. core)	multi	1hcl	6.8	1.6	-77	1	18	18	19	9	16	11	15	13	16	17
alpha/beta hydrolases	α/β	2ace	2.2	0.9	-62	10	32	32	31	25	26	21	23	26	26	26
Zn2/C6 DNA-bind. dom.	sml	1aw6	2.6	0.3	-89	9	75	75	94	27	50	32	40	48	39	50

Top-10
folds in
genome
and tran-
scriptome

Composition of Transcriptome in terms of Broad Structural Classes

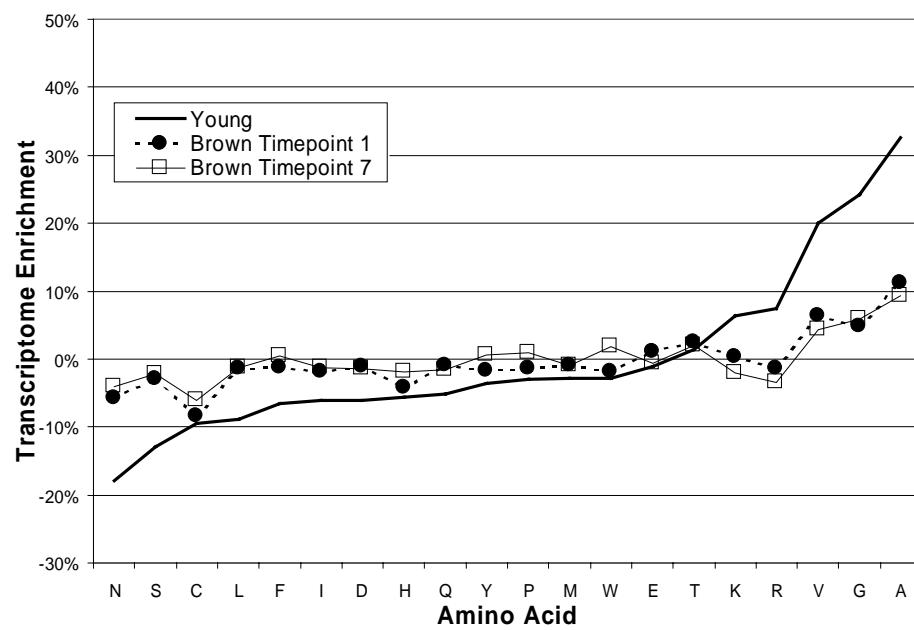


Broad Categories Const. in Transcriptome over Timecourse, Not Specific Genes (or Folds)



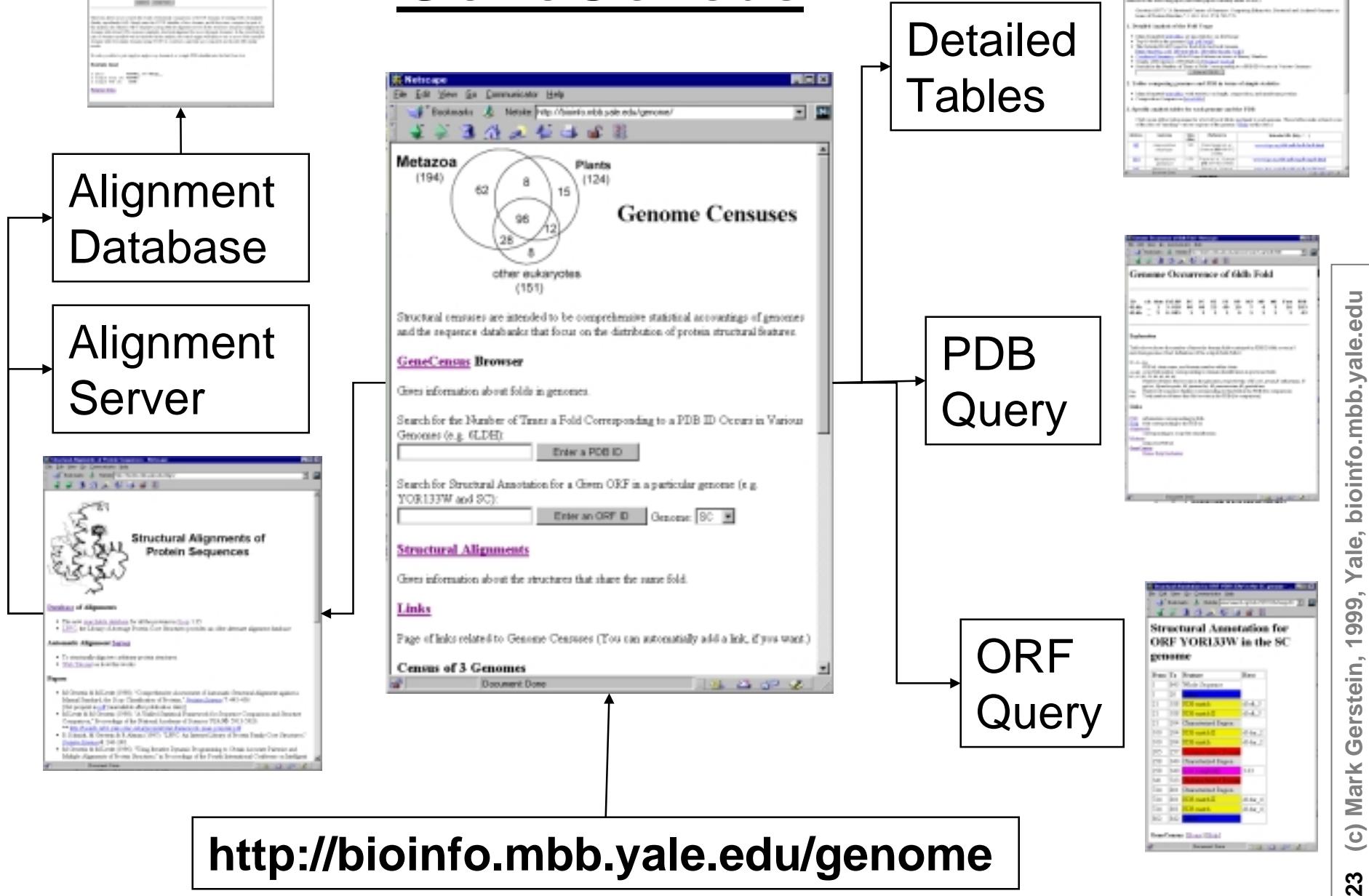
Brown cDNA microarray expts. not as useful for X-ref. at individual timepts

Nevertheless, they show same aa composition and fold class usage at different timepts. However, top fold changes and also specific TM proteins....

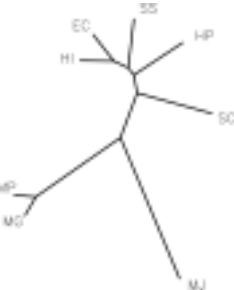
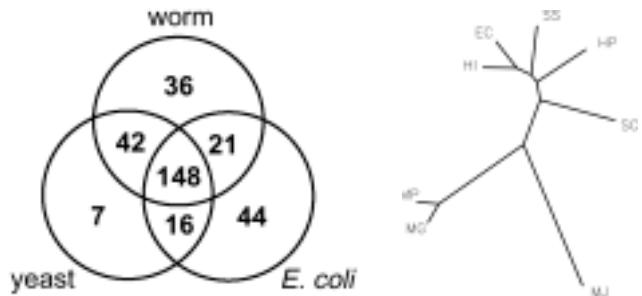


Common Yeast Folds (scop)	Rep. Structure	Genome Duplication	Expression (aerobic)	Expression (anaerobic)
Protein kinases (cat. core)	1hc1	1	3	4
NTP Hydrolases with P-loop	1gky	2	1	2
Classic Zn finger	1ard	3	9	5
Ribonuclease H-like motif	2rn2	4	2	1
Rossmann Fold	1xel	5	4	3
Zn2/Cys6 DNA-binding dom.	125d	6	6	7
7-bladed beta-propeller	2bbk-H	7	8	16
TIM-barrel	1byb	8	5	6
like Ferrodoxin	1fxd	9	7	10
DNA-binding 3-helix bundle	1enh	10	30	36
...		...		
GroES-like	1lep-A	17	10	9
...		...		
like HSP70, Ct-dom.	1dkz-A	22	11	8

GeneCensus



Integrated Database Analysis



Fold	Fold Class	Rep. PDB	Composition			Rank					
			Genome [%]	Transcriptome [%]	Rel. Diff. [%]		Young	Sanson-a	Church-alpha	Church-gal	Church-neat
TIM barrel	α/β	1byb	4.2	8.3	+98	5	1	1 1 1 1 1 1 1			
P-loop NTP hydrolases	α/β	1gky	5.8	5.2	-11	3	2	2 4 4 4 4 5 6	7		
Ferrredoxin like	α/β	1fxd	3.9	3.4	-14	6	3	7 11 9 8 10 4 11			
Rossmann fold	α/β	1xel	3.3	3.3	0	8	4	3 3 3 2 2 19 19	9		
7-bladed beta-propeller	β	1mda*	6.4	2.9	-55	2	5	4 5 6 6 7 9 9	16		
alpha-alpha superhelix	α	2bct	4.4	2.7	-37	4	6	11 15 16 12 12 8 5	8		
Thioredoxin fold	α/β	2trx	1.7	2.7	+63	7	7	6 8 2 5 4 11 10	6		
G3P dehydrogenase-like	α/β	1drw1	0.2	2.7	+1316	14	8	7 2 5 3 3 39 30			
beta grasp	α/β	1lqd	0.6	2.6	+348	31	10	10 21 9 18 21 82 122 120			
HSP70 C-term. fragment	multi	1dky	0.8	2.6	+231	31	10	16 17 11 16 12 48 25 56			
long helices oligomers	α	1zta	3.8	2.1	-46	7	11	8 14 21 15 19 21 30 33			
Protein kinases (cat. core)	multi	1hcl	6.8	1.6	-77	1	1	10 9 10 11 12 13 14	17		
alpha/beta hydrolases	α/β	2ace	2.2	0.9	-62	10	12	11 25 26 21 23 26 26			
Zn2/C6 DNA-bind. dom.	sml	1aw6	2.6	0.3	-89	9	15	14 27 30 32 21 48 39 50			

prototype: bioinfo.mbb.yale.edu/genome

- Typical Structural Genomics Analyses

- ◊ Shared and unique folds

- Relating Results Across Genomes
- Coverage of a number of organisms, biases

- ◊ Common folds

- Relationship to target selection (typical vs atypical)

- Integration with other whole genome data

- ◊ X-ref expression results, functional classifications

- ◊ Importance of selecting an organism for which there is much of this data

- Implementation issues

- ◊ Consistent identifier schemes

- ◊ History, fold and ORF assignments change