

Table 1A

abbrev.	Species Name	ORF Coverage			Amino Acid Coverage			Domain Matches			Domain Length	Duplication	
		Total	Matching	m/t	Total	Matching	m/t	Folds	Sfam	Dom		Fold	Sfam
Aao	<i>Aquifex aeolicus</i>	1522	527	34.6%	482512	116664	24.2%	162	205	690	169.1	4.26	3.37
Aful	<i>Archaeoglobus fulgidus</i>	2409	650	27.0%	663320	146655	22.1%	147	186	849	172.7	5.78	4.56
Bbur	<i>Borrelia burgdorferi</i>	1638	289	17.6%	432219	65816	15.2%	126	151	369	178.4	2.93	2.44
Bsub	<i>Bacillus subtilis</i>	4100	1121	27.3%	1217000	276596	22.7%	208	276	1460	189.4	7.02	5.29
Cele	<i>Caenorhabditis elegans</i>	19099	4586	24.0%	8096713	1136801	14.0%	247	304	7803	145.7	31.59	25.67
Cpne	<i>Chlamydia pneumoniae</i>	1052	274	26.0%	361694	66160	18.3%	136	165	367	180.3	2.70	2.22
Ctra	<i>Chlamydia trachomatis</i>	894	259	29.0%	312553	60295	19.3%	134	163	348	173.3	2.60	2.13
Ecol	<i>Echerischia coli</i>	4290	1191	27.8%	1363501	296762	21.8%	229	303	1611	184.2	7.03	5.32
Hinf	<i>Haemophilus influenzae Rd</i>	1707	528	30.9%	520930	125776	24.1%	190	243	710	177.1	3.74	2.92
Hpyl	<i>Helicobacter pylori</i>	1577	381	24.2%	500616	89025	17.8%	152	193	495	179.8	3.26	2.56
Mthe	<i>Methanobacterium thermoautotrophicum</i>	479	164	34.2%	174566	39680	22.7%	95	111	228	174.0	2.40	2.05
Mjan	<i>Methanococcus jannaschii</i>	1771	470	26.5%	501793	93299	18.6%	128	164	613	152.2	4.79	3.74
Mtub	<i>Mycobacterium tuberculosis</i>	677	178	26.3%	237651	43222	18.2%	101	118	251	172.2	2.49	2.13
Mgen	<i>Mycoplasma genitalium</i>	1871	522	27.9%	526205	105553	20.1%	135	179	675	156.4	5.00	3.77
Mpne	<i>Mycoplasma pneumoniae</i>	3924	1198	30.5%	1335687	291496	21.8%	199	253	1587	183.7	7.97	6.27
Phor	<i>Pyrococcus horikoshii</i>	2064	461	22.3%	568544	97276	17.1%	121	155	555	175.3	4.59	3.58
Rpro	<i>Rickettsia prowazekii</i>	837	264	31.5%	280233	60285	21.5%	135	160	350	172.2	2.59	2.19
Scer	<i>Saccharomyces cerevisiae</i>	6218	1699	27.3%	2906890	434481	14.9%	215	273	2346	185.2	10.91	8.59
Syne	<i>Synechocystis sp.</i>	3168	882	27.8%	1119717	196041	17.5%	199	255	1131	173.3	5.68	4.44
Tpal	<i>Treponema pallidum</i>	1031	252	24.4%	350676	58542	16.7%	123	150	346	169.2	2.81	2.31

Table 1B: Represented Superfamilies and Their Average Duplication

Levels in the Soluble Fold Classes in *A.fulgidus*, *E.coli*, Yeast, Worm and the Total of the 20 Genomes

	All-alpha			All-beta			Alpha/Beta			Alpha+Beta			Multidomain			Small			Total
	Sfams	Dup	Copy#	Sfams	Dup	Copy#	Sfams	Dup	Copy#	Sfams	Dup	Copy#	Sfams	Dup	Copy#	Sfams	Dup	Copy#	
iful	29	2.5	73	18	2.1	37	74	6.1	453	49	4.2	207	12	5.7	68	4	2.8	11	186
ecol	55	2.9	159	44	4.1	181	105	8.3	872	78	3.9	313	16	5.0	81	5	1.0	5	303
scer	56	7.9	448	35	9.5	333	88	9.3	823	72	5.2	351	14	14.6	204	13	14.3	187	273
cele	62	20.3	1319	52	27.8	1633	81	18.1	1482	72	15.7	1140	14	42.6	598	23	63.7	1631	304
20	97		3197	83		3069	117		8976	120		4046	19		1598	35		1898	471

Table 2: Patterns of Complementary Superfamilies and Horizontal Transfer
A/ Complementary clades between bacterial/archaeal and eukaryotic genes
B/ Complementary clades between bacterial and eukaryotic/archaeal genomes
C/ Other complementary patterns
D/ Horizontal Transfer between Archaea and Bacteria
E/ Horizontal Transfer between Eukaryotes and Bacteria

	afu	mjan	mthe	phor	scer	cele	aaeo	syne	ecol	bsub	mtub	hirf	hpyl	ingen	mpne	bbur	tpal	cpne	rpro	Sfam	domain	SCOP Function	Swissprot	Swissprot Function	
A	3.25.1	d1fsz_1	Tubulin, GTPase domain	FTSZ_ECOLI	CELL DIVISION PROTEIN FTSZ
	4.57.1	d1puc_	Cell cycle regulatory proteins	CKS1_YEAST	CELL DIVISION CONTROL PROTEIN CKS1
B	1.22.1	d1tafb_	Histone-fold	T2D5_YEAST	TRANSCRIPTION INITIATION FACTOR TFIID
	1.63.1	d1kxu_2	Cyclin-like	TF2B_RAT	TRANSCRIPTION INITIATION FACTOR IIB
	7.35.3	d1qyp_	Rubredoxin-like transcriptional factor domain	TFS2_YEAST	TRANSCRIPTION ELONGATION FACTOR S-II
	1.34.1	d1coo_	C' domain of RNA polymerase alpha subunit	RPOA_TREPA	DNA-DIRECTED RNA POLYMERASE
	1.75.1	d1gln_1	Glu-tRNA synthetase AC-binding domain	SYE_BACSU	GLUTAMYL-TRNA SYNTHETASE (EC 6.1.1.17)
	1.88.1	d1sig_	RNA polymerase, sigma70 subunit	RPOS_ECOLI	RNA POLYMERASE SIGMA FACTOR
	4.104.1	d2def_	Peptide deformylase catalytic core	DEF_HAEIN	POLYPEPTIDE DEFORMYLASE (EC 3.5.1.31)
	4.11.7	d1tif_	Translation initiation factor, N' domain	IF3_BORBU	TRANSLATION INITIATION FACTOR IF-3
	4.28.1	d2reb_2	RecA protein, C-terminal domain	RECA_HAEIN	RECA PROTEIN (RECOMBINASE)
	4.36.1	d1ife_	Translation initiation factor IF3	IF3_BORBU	TRANSLATION INITIATION FACTOR IF-3
.	4.88.1	d1div_	Ribosomal protein L9	RL9_BACSU	50S RIBOSOMAL PROTEIN L9	
C	4.40.1	d2chsa	Chorismate mutase	CHMU_BACSU	CHORISMATE MUTASE (EC 5.4.99.5)	
	1.101.1	d5csma	Chorismate mutase II	CHMU_YEAST	CHORISMATE MUTASE (EC 5.4.99.5)	
	1.81.1	d1cem_	Glycosyltransferases of the superhelical fold	GUN6_DICDI	ENDOGLUCANASE (EC 3.2.1.4)	
	2.21.1	d1yna_	ConA-like lectins/glucanases	GUN1_TRIRE	ENDOGLUCANASE EG-1 (EC 3.2.1.4)	
	3.1.1	d1edt_	Glycosyltransferases	GUNB_NEOPA	ENDOGLUCANASE B (EC 3.2.1.4)	
	4.2.1	d153l_	Lysozyme-like	CHIT_SOLTU	ENDOCHITINASE PRECURSOR (EC 3.2.1.14)	
.	2.65.2	d1hcz_2	Rudiment single hybrid motif	PYC_PICPA	PYRUVATE CARBOXYLASE (EC 6.4.1.1)		
.	2.65.3	d1f3z_	Duplicated hybrid motif	PTGA_BACSU	PTS SYSTEM, GLUCOSE-SPECIFIC IIABC COMP		
D	1.86.1	d1aora1	Aldehyde FerOR C' domain	AOR_PYRFU	ALDEHYDE:FERREDOXIN OXIDOREDUCTASE	
	4.94.1	d1aora2	Aldehyde FerOR N' domain	AOR_PYRFU	ALDEHYDE:FERREDOXIN OXIDOREDUCTASE	
	3.1.10	d5rubal1	RuBisCo, C' domain	RBL_NITVU	RUBISCO LARGE SUBUNIT	
E	1.101.1	d5csma	Chorismate mutase II	CHMU_ARATH	CHORISMATE MUTASE (EC 5.4.99.5)	
	1.37.1	d1rec_	EF-hand	TPC2_DROME	TROPONIN C	
	2.1.5	d1suh_	Cadherin	CAD5_HUMAN	VASCULAR ENDOTHELIAL-CADHERIN	
	2.45.1	d1eal_	Lipocalins	PGHD_HUMAN	PROSTAGLANDIN-H2	
	3.7.1	d2bnh_	Leucine-rich repeats	RINI_PIG	RIBONUCLEASE INHIBITOR	
	4.70.1	d1axx_	Cytochrome b5	NI1_MAIZE	NITRATE REDUCTASE (EC 1.6.6.1)	
.	4.112.1	d1toh_	Tyrosine hydroxylase	TY3H_HUMAN	TYROSINE 3-HYDROXYLASE (EC 1.14.16.2)		

Table III: List of the Unique Superfamilies in 19 Genomes (worm excluded)

Species	ORF	SCOP dom	SCOP #	copy	Stat. Signif.	SCOP domain function
bbur	BBA15	d1ospo	2.58.1.1.1	2	5.00E-88	NONENZ Outer surface protein A
bsub	NprE	d1ezm_1	1.57.1.1.1	2	2.00E-47	ENZYME Extracellular elastase; Bacillolysin
bsub	PelB	d1idk_	2.62.1.2.1	2	1.00E-70	ENZYME Pectin lyase A
bsub	MtrB	d1wapa_	2.64.6.1.1	1	1.00E-32	NONENZ Trp RNA-binding attenuation protein
bsub	YrdF	d1brsd_	3.6.1.1.1	1	1.70E-14	NONENZ Barstar (barnase inhibitor)
cpne	AAD18679	d1kpta	4.37.1.1.1	1	0.00096	NONENZ Virally encoded KP4 toxin
ecol	tar	d2asr_	1.24.2.1.1	3	2.00E-57	NONENZ Aspartate receptor, ligand-binding domain
ecol	cybC	d256ba_	1.24.3.1.1	1	1.00E-52	NONENZ Cytochrome b562
ecol	arcB	d2a0b_	1.24.9.1.1	1	1.40E-48	NONENZ Aerobic respiration control sensor protein
ecol	holB	d1a5t_1	1.93.1.1.1	1	1.00E-80	ENZYME DNA polymerase III, delta subunit, C' dom
ecol	eco	d1slua_	2.12.1.1.1	1	3.70E-61	NONENZ Ecotin, trypsin inhibitor
ecol	lacZ	d1bglA4	2.22.1.1.1	2	2.50E-136	ENZYME Beta-Galactosidase, domain 5
ecol	tynA	d1oaca1	2.22.2.1.1	1	0	ENZYME Copper amine oxidase, domain 3 (catalytic)
ecol	b0717	d3dpa_2	2.6.2.1.1	10	2.00E-25	NONENZ Chaperone protein, PapD, C-domain
ecol	tynA	d1oaca3	4.13.2.1.1	2	1.00E-62	ENZYME Copper amine oxidase, domains 1 and 2
ecol	cheA	d1eayc_	4.34.20.1.1	1	1.00E-23	NONENZ CheY-binding domain of CheA
ecol	tynA	d1oaca4	4.43.1.1.1	1	1.50E-36	ENZYME Copper amine oxidase, domain N
ecol	tus	d1ecra	5.3.1.1.1	1	2.00E-130	NONENZ Replication terminator protein (Tus)
hinf	HI1478	d1bco_1	2.36.1.1.1	1	2.00E-29	ENZYME Mu transposase, C-terminal domain
mtub	Rv1353c	d2tct_2	1.94.1.1.1	1	8.00E-31	NONENZ Tetracyclin repressor, C-terminal domain
mtub	Rv1758	d1cex_	3.14.7.1.1	7	3.00E-40	ENZYME Cutinase, closest homolog in Penicillium
mtub	Rv0062	d1tml_	3.2.1.1.1	1	5.00E-66	ENZYME Cellulase E2
mtub	Rv0316	d1mli_	4.34.4.1.1	1	9.00E-07	ENZYME Muconalactone isomerase
mtub	Rv1919c	d1bv1_	4.79.3.1.1	1	0.0018	NONENZ Major birch pollen allergen Bet v 1
rpro	RP396	d3pcca_	2.3.3.1.1	1	1.00E-06	ENZYME Protocatechuate-3,4-dioxygenase
scer	YDL185W	d1vdea3	4.55.2.2.1	4	2.90E-43	NONENZ Homothallic switching nuclease
scer	YLR014C	d1pyia2	7.32.1.1.2	47	1.50E-20	NONENZ Zn/Cys6 DNA binding domain
scer	YHR053C	d1aoo	7.38.1.1.5	2	2.80E-22	NONENZ Metallothionein
syne	slr1317	d1hcz_1	2.2.5.1.1	1	3.00E-90	NONENZ Cytochrome f, large domain
syne	ssr2831	d1pse_	2.24.5.1.1	1	5.00E-33	NONENZ Photosystem I accessory protein E (PSAE)
syne	slr1028	d1jpc_	2.60.1.1.1	1	0.0035	NONENZ Lectin (agglutinin)
syne	slr0012	d3rubs_	4.38.1.1.1	2	1.00E-39	ENZYME RuBisCO, small subunit

Figure 1A

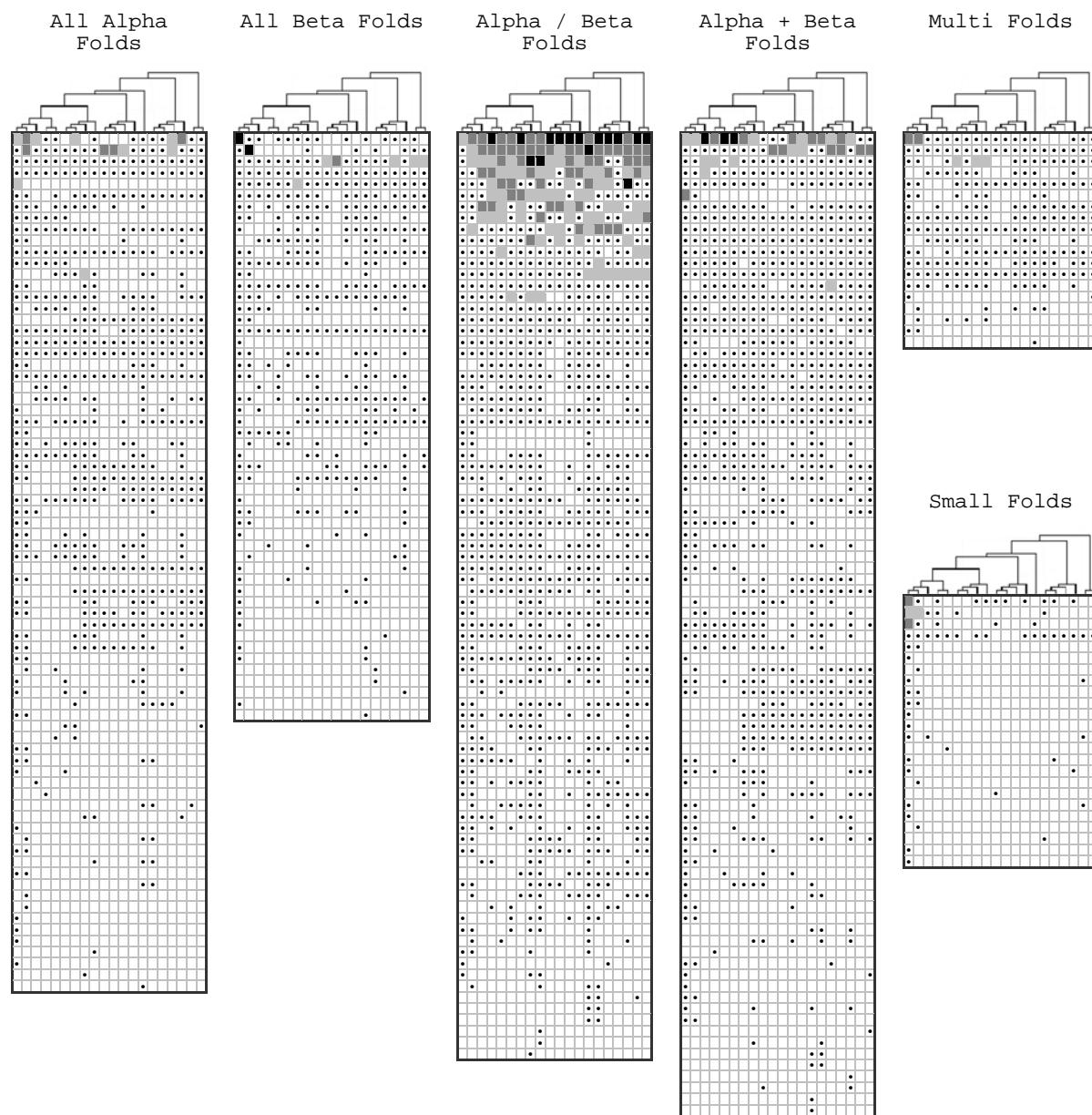


Figure 2

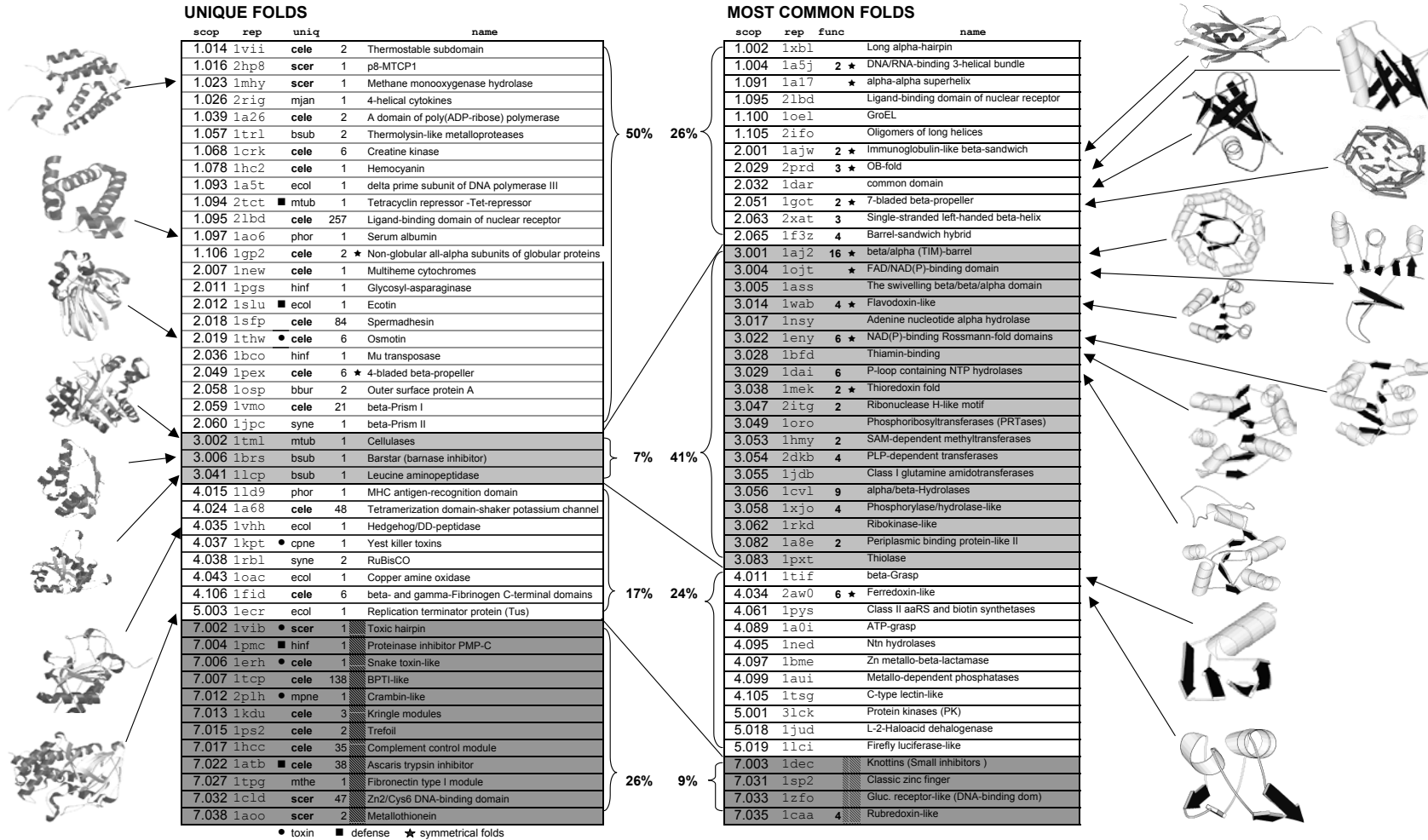


Figure 3 -- Conservation of the superfamilies in the 6 structural classes

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	cele	scer	mjan	phor	mthe	aful	aaeo	mtub	bsub	mpne	mgen	hpyl	rpro	ecol	hinf	bbur	tpal	syne	ctra	cpne
A	19	16	8	6	3	6	3	3	1	3	3	3	3	4	0	5	2	3	1	5
B	22	13	8	2	4	3	1	7	1	5	2	1	2	1	3	1	3	0	0	4
A/B	4	6	7	4	2	4	3	2	1	5	5	15	5	8	7	6	4	5	8	16
A+B	14	14	10	6	11	3	0	1	7	6	2	7	2	7	4	6	5	2	3	10
M	1	2	0	1	1	0	1	0	0	2	0	0	1	1	0	1	4	1	0	3
S	15	10	1	3	2	1	0	1	0	0	2	0	0	0	0	0	0	0	0	0

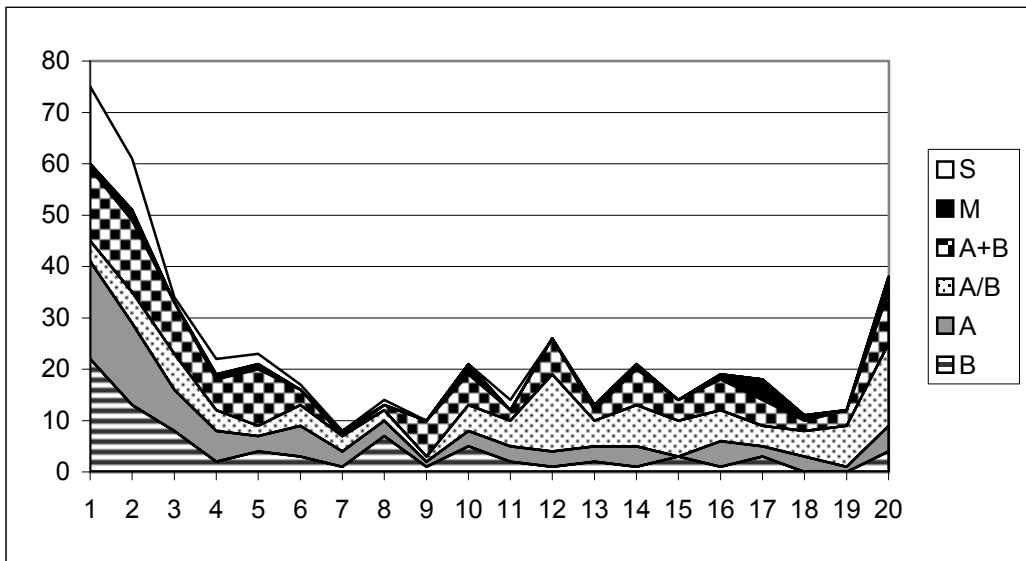
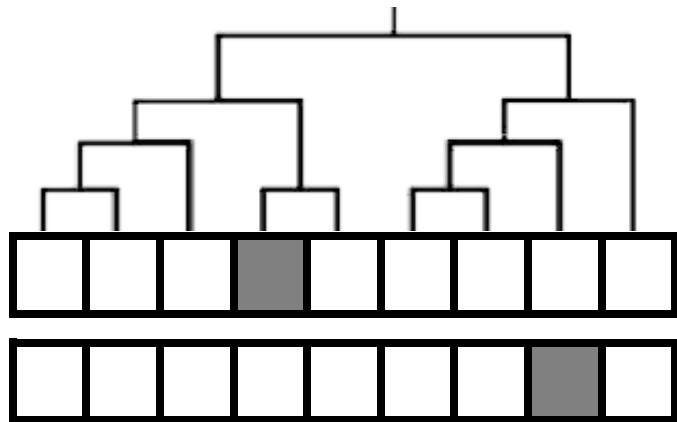
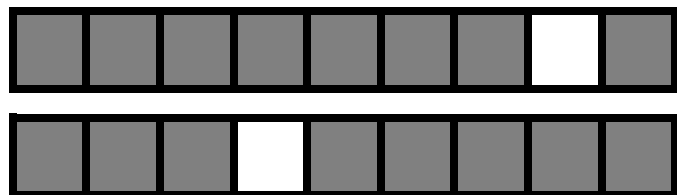


Figure 4 - Schematic of the different fold patterns



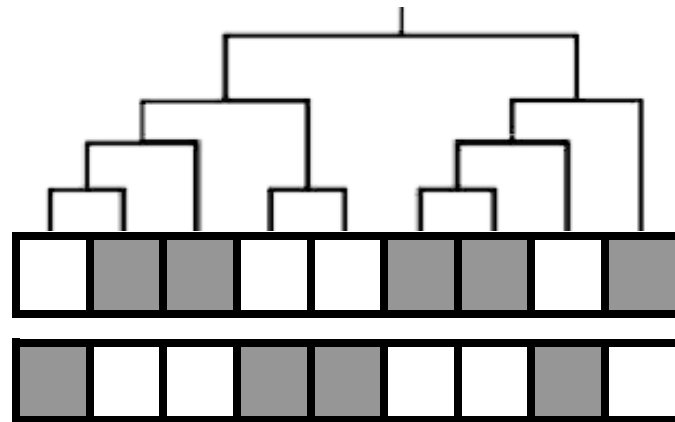
PRESENT IN A SINGLE ORGANISM



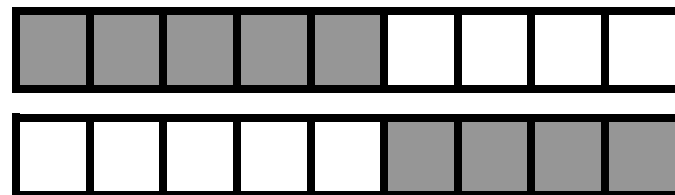
ABSENT FROM A SINGLE ORGANISM



ABUNDANCE OF FOLDS OR ORGANISMS



COMPLEMENTARY PATTERN



SINGLE CLADE (COMPLEMENTS)



LATERAL GENE TRANSFER



GENE LOSS