Supplementary Material

Figures

Figure S1. Promoter structure and transcriptional control in archaea. a) The promoter element is schematically aligned to the scale shown on top. The numbering refers to the co-ordinates of the base pairs with respect to transcription start site indicated as +1. BRE is a transcription factor binding (TFB) recognition element; TATA stands for the TATA box; INR refers to the initiator element and ORF for the open reading frame. b) Repression at archaeal promoter. The archaeal regulator TrmB is shown to bind to its operator overlapping the BRE-TATA binding site, inhibiting the RNApol recognition. c) Activation at archaeal promoter. The archaeal regulator Ptr2 binds to its activator binding sites upstream of the BRE-TATA box, facilitating the recruitment of TBP for transcription.

Figure S2. Distribution of TFs grouped by different lengths in *Escherichia coli* K12. Intervals of 50 amino acids were considered and the fraction of TFs per interval was computed. In addition, 1000 sets of 314 proteins retrieved randomly were considered to evaluate the significance of the distribution. TFs in *E .coli* were found to follow a similar distribution to the rest of the proteins.

Figure S3. Boxplot showing the distribution of the lengths of TFs from archaea and bacteria measured as the number of amino acids. Archaeal TFs were found to show significantly lower lengths compared to bacterial ones (median size of 179 *versus* 236 amino acids, $p < 2.2 \times 10^{-16}$, Wilcoxon test). Left most boxes correspond to the TFs from the three most abundant families as well as the remaining TF collection from the archaeal repertoire. Both the groups were found to show significantly lower lengths compared to bacterial ones ($p < 2.2 \times 10^{-16}$, Wilcoxon test).

Figure S4. Distribution of archaeal TFs shared by the three cellular domains, archaea, bacteria and eukarya. Homologues of archaeal TFs, across different major kingdoms, grouped by their genome. Blast searches were performed between all TFs previously identified against total sequences of bacterial and eukaryotic genomes. A protein was considered as homologue if the alignment covered at least ≥ 60% of the query sequence, with an E-value = 10^{-6}. The scale bar represents the proportion of TFs with homologues in bacteria and eukarya.

Tables

Table 1. TFs studied experimentally in archaea.

Name	ID	Role *	Family	Mode of action	Organism	Comments	References
Bat	Q9HPU8	+	HTH_10		*Halobacterium.salinarum*		Leong et al 1988
GvpE	P13044	+/-	PadR		*H.salinarum*		Krüger and Pfeifer, 1996
LrpA	P42180	-	FFRP	Blocks RNAP recruitment	*Pyrococcus furiosus*	Homotetramer	Dahlke and Thomm, 2002
Lrs14	O93774	+/-	FFRP	Blocks TBP and TFB recruitment	*Sulfolobus. solfataricus*	Homodimer	Bell and Jackson, 2000
LysM	Q980W9	+	FFRP		*S. solfataricus*	Homotetramer	Brinkman, et al 2002
NrpR	Q8J2Z0	-	PF01995	Probably blocks RNAP	*Methanococcus maripaludis*	Tetrameric	Lie, 2008
Ptr1, Ptr2	Q58133	+/-	FFRP	Facilitates TBP binding	*M. jannaschii*	Homotetramer	Ouhammouch, et al 2001
S7c3	P81551	-	SpoVT_AbrB				Knapp et al 1998
BldR		+	MarR	Transcriptional	Sulfolobus solfataricus		Di Fiore et al 2009
DmsR		+			*Halobacterium sp. NRC1*		Müller and DasSarma, 2005
GvpD	P13043	-			*Halobacterium salinarum*		Scheuch et al 2008
LrpB		-	FFRP		*Sulfolobus solfataricus*		Peeters et al 2007
Mbf1			MarR		Sulfolobus solfataricus		Fiorentino, et al 2007
Mdr1		-	DtxR	Blocks RNAP recruitment	*Archaeoglobus. fulgidus*		Bell et al., 1999
PhrA		-	ArsR	Blocks RNAP recruitment	*Pyrococcus furiosus*		Vierke et al 2003
SurR		-			*Pyrococcus furiosus*		Lipscomb, et al 2009
TrmB		-	TrmB	Probably TBP and TFB recruitment	*Thermococcus litoralis*	Tetramer/Octamer	Lee et al 2003
TrpY		U	?		*M. thermautotrophicus*		Karr et al 2008

ID: Swiss Prot identifier; * Regulatory roles are defined as: +, activator; -, repressor; +/-, dual; U, unknown. PF01995 is an unknown domain described in PFAM database. PhrA has been assigned to the ArsR family according to DBD database.

References

Bell SD, Cairns SS, Robson RL, Jackson SP (1999) Transcriptional regulation of an archaeal operon in vivo and in vitro. Mol Cell 4:971-82

Bell SD, Jackson SP (2000) Mechanism of autoregulation by an archaeal transcriptional repressor. J Biol Chem 275:31624-9

Brinkman AB, Bell SD, Lebbink RJ, de Vos WM, van der Oost J (2002) The Sulfolobus solfataricus Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability. J Biol Chem 277:29537-49

Dahlke I, Thomm M (2002) A Pyrococcus homolog of the leucine-responsive regulatory protein, LrpA, inhibits transcription by abrogating RNA polymerase recruitment. Nucleic Acids Res 30:701-10

Di Fiore A, Fiorentino G, Vitale RM, Ronca R, Amodeo P, Pedone C, Bartolucci S, De Simone G (2009) Structural analysis of BldR from Sulfolobus solfataricus provides insights into the molecular basis of transcriptional activation in archaea by MarR family proteins. J Mol Biol 388:559-69

Fiorentino G, Ronca R, Cannio R, Rossi M, Bartolucci S (2007) MarR-like transcriptional regulator involved in detoxification of aromatic compounds in Sulfolobus solfataricus. J Bacteriol 189:7351-60

Karr EA, Sandman K, Lurz R, Reeve JN (2008) TrpY regulation of trpB2 transcription in Methanothermobacter thermautotrophicus. J Bacteriol 190:2637-41

Kruger K, Pfeifer F (1996) Transcript analysis of the c-vac region and differential synthesis of the two regulatory gas vesicle proteins GvpD and GvpE in Halobacterium salinarium PHH4. J Bacteriol 178:4012-9

Lee SJ, Engelmann A, Horlacher R, Qu Q, Vierke G, Hebbeln C, Thomm M, Boos W (2003) TrmB, a sugar-specific transcriptional regulator of the trehalose/maltose ABC transporter from the hyperthermophilic archaeon Thermococcus litoralis. J Biol Chem 278:983-90

Leong D, Pfeifer F, Boyer H, Betlach M (1988) Characterization of a second gene involved in bacterio-opsin gene expression in a halophilic archaebacterium. J Bacteriol 170:4903-9

Lie TJ, Leigh JA (2007) Genetic screen for regulatory mutations in Methanococcus maripaludis and its use in identification of induction-deficient mutants of the euryarchaeal repressor NrpR. Appl Environ Microbiol 73:6595-600

Lipscomb GL, Keese AM, Cowart DM, Schut GJ, Thomm M, Adams MW, Scott RA (2009) SurR: a transcriptional activator and repressor controlling hydrogen and elemental sulphur metabolism in Pyrococcus furiosus. Mol Microbiol 71:332-49

Muller JA, DasSarma S (2005) Genomic analysis of anaerobic respiration in the archaeon Halobacterium sp. strain NRC-1: dimethyl sulfoxide and trimethylamine N-oxide as terminal electron acceptors. J Bacteriol 187:1659-67

Oppermann UC, Knapp S, Bonetto V, Ladenstein R, Jornvall H (1998) Isolation and structure of repressor-like proteins from the archaeon Sulfolobus solfataricus. Co-purification of RNase A with Sso7c. FEBS Lett 432:141-4

Ouhammouch M, Dewhurst RE, Hausner W, Thomm M, Geiduschek EP (2003) Activation of archaeal transcription by recruitment of the TATA-binding protein. Proc Natl Acad Sci U S A 100:5097-102

Peeters E, Wartel C, Maes D, Charlier D (2007) Analysis of the DNA-binding sequence specificity of the archaeal transcriptional regulator Ss-LrpB from Sulfolobus solfataricus by systematic mutagenesis and high resolution contact probing. Nucleic Acids Res 35:623-33

Scheuch S, Marschaus L, Sartorius-Neef S, Pfeifer F (2008) Regulation of gvp genes encoding gas vesicle proteins in halophilic Archaea. Arch Microbiol 190:333-9

Vierke G, Engelmann A, Hebbeln C, Thomm M (2003) A novel archaeal transcriptional regulator of heat shock response. J Biol Chem 278:18-26

Table S2. Archaeal genomes analyzed in this work. Columns are as follows: Cellular division, Organism, GenBank identifiers, Life-Style, Number of predicted ORFs and Total number of transcription factors.

DIVISION	Organism	ID_GENBANK	LIFE STYLE	ORFS	TFS
Crenarchaeota	*Aeropyrum pernix K1*	NC_000854	Aerobic hyperthermophilic archaeon	1700	69
Crenarchaeota	*Caldivirga maquilingensis IC-167*	NC_009954	Hyperthermophilic crenarchaeota	1963	61
Crenarchaeota	*Hyperthermus butylicus DSM 5456*	NC_008818	Extreme hyperthermophilic archeon	1602	56
Crenarchaeota	*Ignicoccus hospitalis KIN4/I*	NC_009776	Isolated from gravel from the Kolbeinsey Ridge	1434	32
Crenarchaeota	*Metallosphaera sedula DSM 5348*	NC_009440	Acid-loving thermophile which can oxidize coal and pyrite	2256	82
Crenarchaeota	*Nitrosopumilus maritimus SCM1*	NC_010085	Autotrophic ammonia-oxidizing Archaea	1795	58
Crenarchaeota	*Pyrobaculum aerophilum str. IM2*	NC_003364	Nitrate-reducing hyperthermophilic archeon	2605	79
Crenarchaeota	*Pyrobaculum arsenaticum DSM 13514*	NC_009376	Arsenate-reducing hyperthermophile	2299	62
Crenarchaeota	*Pyrobaculum calidifontis JCM 11548*	NC_009073	Facultatively anaerobic hyperthermophile	2149	55
Crenarchaeota	*Pyrobaculum islandicum DSM 4184*	NC_008701	Hyperthermophilic anaerobe	1978	52
Crenarchaeota	*Staphylothermus marinus F1*	NC_009033	Hyperthermophilic archaeon	1570	48
Crenarchaeota	*Sulfolobus acidocaldarius DSM 639*	NC_007181	Extreme thermoacidophilic sulfur-oxidizing archeon	2223	95
Crenarchaeota	*Sulfolobus solfataricus P2*	NC_002754	Hyperthermophilic acidophilic sulfur-metabolizing archeon	2977	113
Crenarchaeota	*Sulfolobus tokodaii str. 7*	NC_003106	Hyperthermophilic acidophilic sulfur metabolizing archeon	2825	125
Crenarchaeota	*Thermofilum pendens Hrk 5*	NC_008698	Hyperthermophilic crenarchaeota	1876	65
Crenarchaeota	*Thermoproteus neutrophilus V24Sta*	NC_010525	Hyperthermophilic crenarchaeota	1966	59
Euryarchaeota	*Archaeoglobus fulgidus DSM 4304*	NC_000917	Hyperthermophilic sulfate-reducing archaeon	2420	106
Euryarchaeota	*Candidatus Methanoregula boonei 6A8*	NC_009712	Acidophilic methanogen	2450	67
Euryarchaeota	*Haloarcula marismortui ATCC 43049*	NC_006396, NC_006397	Chemoheterotrophic obligate extreme halophilic archeon	4240	204
Euryarchaeota	*Halobacterium salinarum R1*	NC_010364	Chemoheterotrophic obligate extreme halophilic archeon	2749	113

Euryarchaeota	*Halobacterium sp. NRC-1*	NC_002607	Chemoheterotrophic obligate extreme halophilic archeon	2622	122
Euryarchaeota	*Haloquadratum walsbyi DSM 16790*	NC_008212	A uniquely square archaeon	2646	94
Euryarchaeota	*Methanobrevibacter smithii ATCC 35061*	NC_009515	Sewage digester isolate\Human fecal bacterium	1793	41
Euryarchaeota	*Methanocaldococcus jannaschii DSM 2661*	NC_000909	Methane-producing thermophilic archeon	1786	64
Euryarchaeota	*Methanococcoides burtonii DSM 6242*	NC_007955	Psychrotolerant methanogenic archeon	2273	75
Euryarchaeota	*Methanococcus aeolicus Nankai-3*	NC_009635	Anaerobic methane-producing archeon	1490	34
Euryarchaeota	*Methanococcus maripaludis C5*	NC_009135	Anaerobic methane-producing archaeon	1822	52
Euryarchaeota	*Methanococcus maripaludis C6*	NC_009975	Isolated from intertidal sediments	1826	47
Euryarchaeota	*Methanococcus maripaludis C7*	NC_009637	Anaerobic methane-producing archeon	1788	48
Euryarchaeota	*Methanococcus maripaludis S2*	NC_005791	Anaerobic methane-producing archeon	1722	64
Euryarchaeota	*Methanococcus vannielii SB*	NC_009634.	A methane-producing motile coccus	1678	52
Euryarchaeota	*Methanocorpusculum labreanum Z*	NC_008942	Methanogen isolated from the LaBrea Tar Pits	1739	52
Euryarchaeota	*Methanoculleus marisnigri JR1*	NC_009051	Methanogen present in marine and freshwater sediments	2489	67
Euryarchaeota	*Methanopyrus kandleri AV19*	NC_003551	Thermophilic methanogenic archaeon	1687	36
Euryarchaeota	*Methanosaeta thermophila PT*	NC_008553	Produces methane from acetate	1696	34
Euryarchaeota	*Methanosarcina acetivorans C2A*	NC_003552	Anaerobic methane-producing archeon	4540	158
Euryarchaeota	*Methanosarcina barkeri str. Fusaro*	NC_007355	Anaerobic methane-producing archeon	3624	124
Euryarchaeota	*Methanosarcina mazei Go1*	NC_003901	Anaerobic methane-producing archeon	3370	127
Euryarchaeota	*Methanosphaera stadtmanae DSM 3091*	NC_007681	Methanogen that uses methanol and hydrogen to produce methane	1534	41
Euryarchaeota	*Methanospirillum hungatei JF-1*	NC_007796	Methanogenic archaeon with unusual filamentous structure	3139	73
Euryarchaeota	*Methanothermobacter thermautotrophicus str. Delta H*	NC_000916	Thermophilic methane-producing archaeon	1873	58
Euryarchaeota	*Natronomonas pharaonis DSM 2160*	NC_007426	Extreme haloalkaliphilic archeon	2822	122
Euryarchaeota	*Picrophilus torridus DSM 9790*	NC_005877	Thermoacidophilic organism that grows at extremes of temperature and pH	1535	64
Euryarchaeota	*Pyrococcus abyssi GE5*	NC_000868	Anaerobic hyperthermophilic archeon	1898	98
Euryarchaeota	*Pyrococcus furiosus DSM 3638*	NC_003413	Anaerobic hyperthermophilic	2125	85

			archeon		
Euryarchaeota	*Pyrococcus horikoshii OT3*	NC_000961	Anaerobic hyperthermophilic archeon	1955	94
Euryarchaeota	*Thermococcus kodakarensis KOD1*	NC_006624	Hyperthermophilic archaeon	2306	105
Euryarchaeota	*Thermoplasma acidophilum DSM 1728*	NC_002578	Thermophilic acidophilic mycoplasma	1482	64
Euryarchaeota	*Thermoplasma volcanium GSS1*	NC_002689	Thermoacidophilic mycoplasma	1499	56
Euryarchaeota	*uncultured methanogenic archaeon RC-I*	NC_009464	A key methano producer in the rhizome of rice	3085	103
Korarchaeota	*Candidatus Korarchaeum cryptofilum OPF8*	NC_010482	An isolated korarchaeote	1602	55
Nanoarchaeum	*Nanoarchaeum equitans Kin4-M*	NC_005213	Hyperthermophilic anaerobic obligate symbiont	536	8

Figure S1

Promoter Structure in Archaea

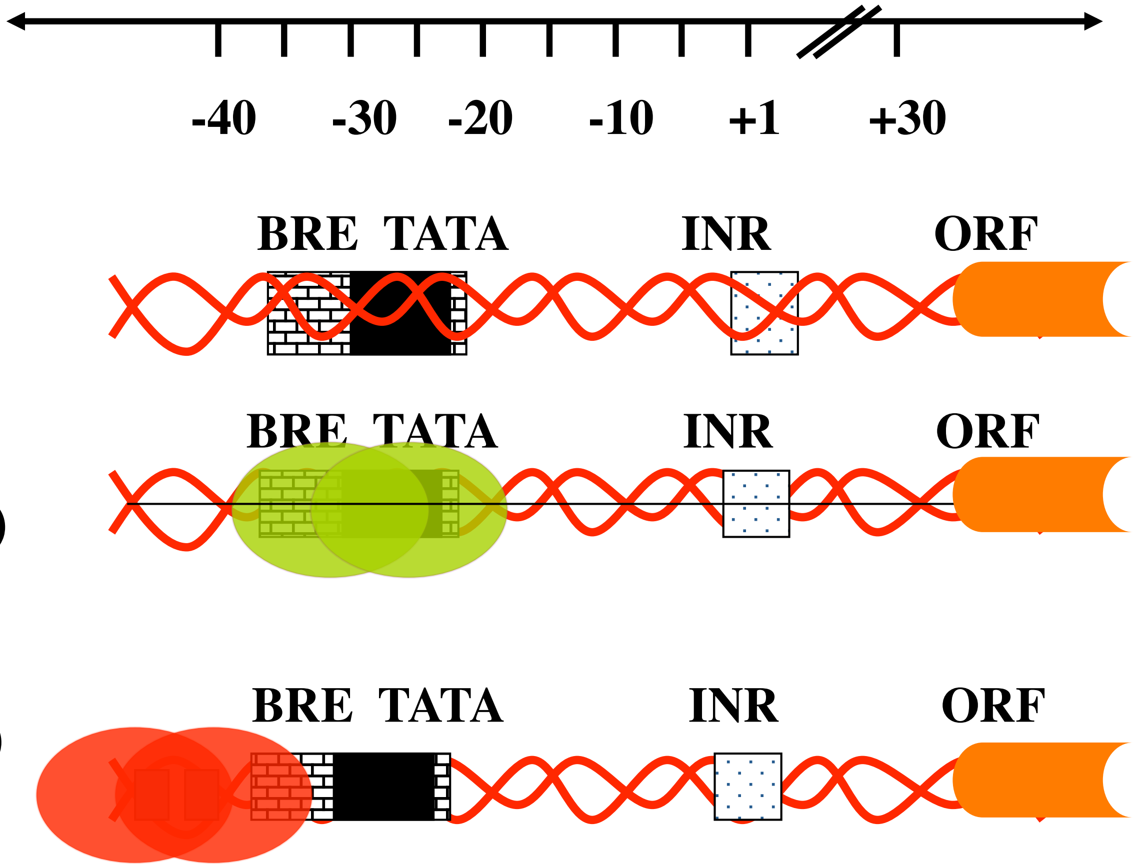
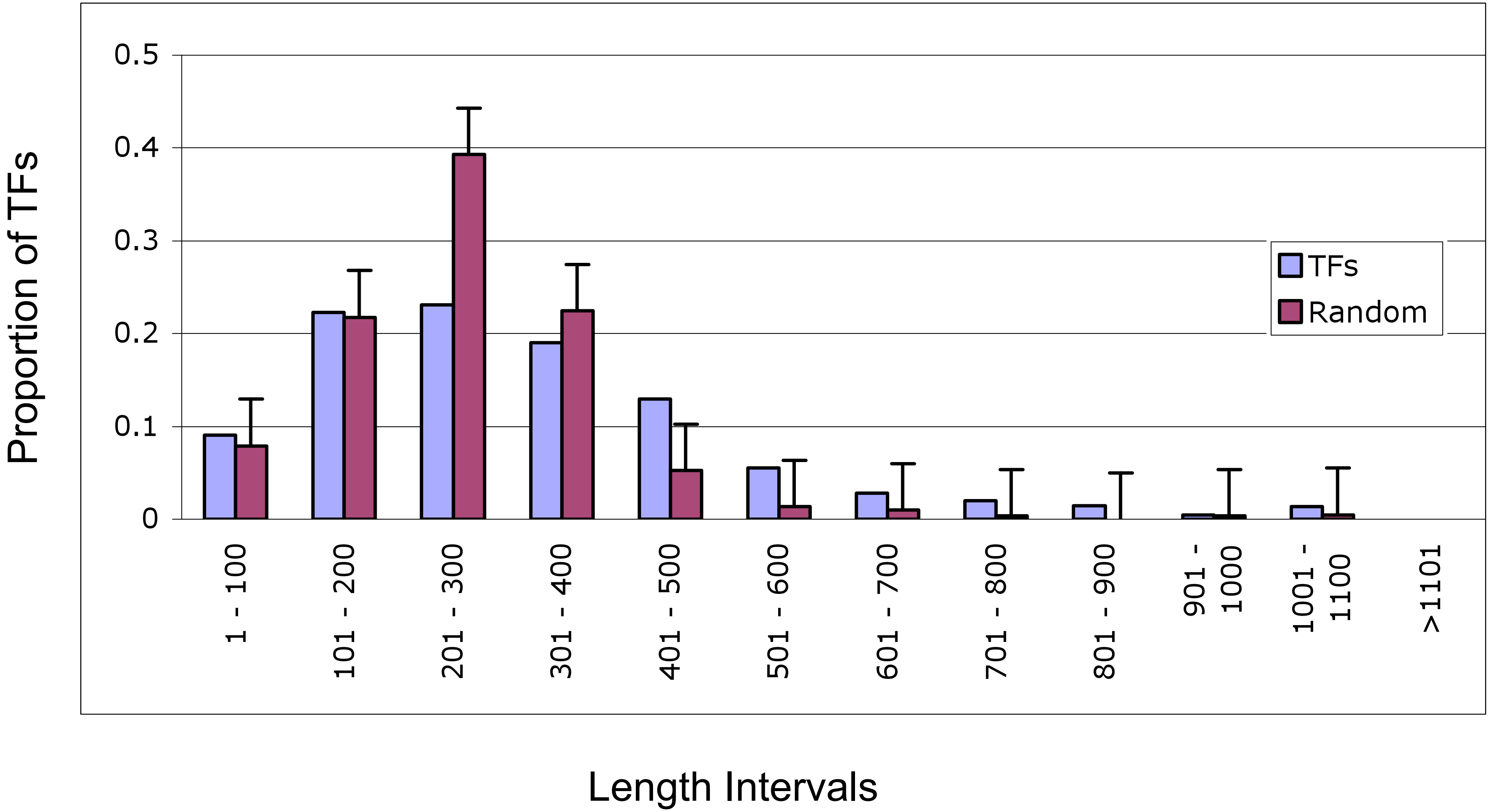


Figure S2



Proportion of TFs
0.5
0.4
0.3
0.2
0.1
0
TFs
Random
1 - 100
101 - 200
201 - 300
301 - 400
401 - 500
501 - 600
601 - 700
701 - 800
801 - 900
901 - 1000
1001 - 1100
>1101
Length Intervals

Figure S3

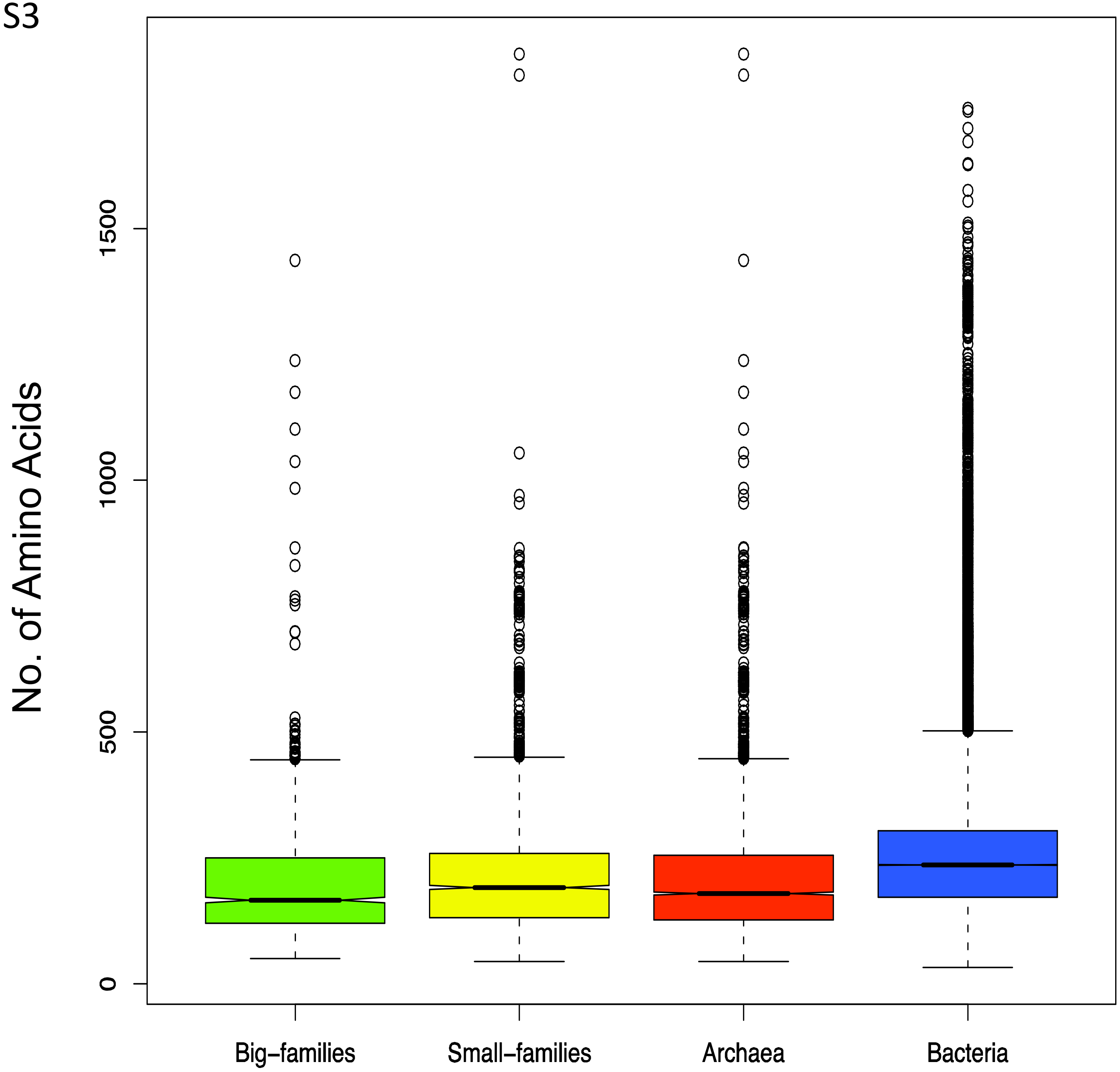

No. of Amino Acids
0
500
1000
1500
Big–families
Small–families
Archaea
Bacteria

Figure S4

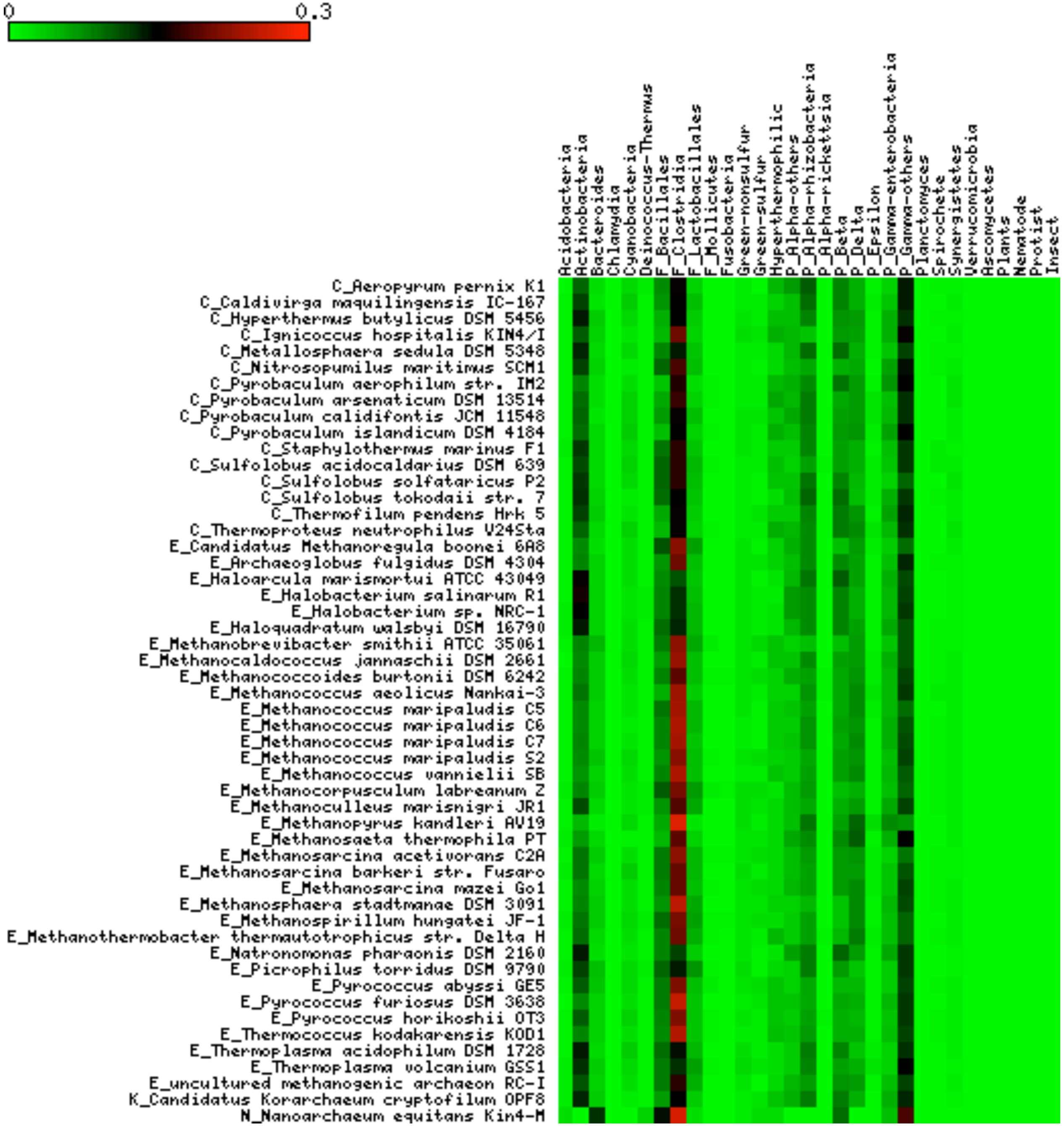

0
0.3
Acidobacteria
Actinobacteria
Bacteroides
Chlamydia
Cyanobacteria
Deinococcus-Thermus
F_Bacillales
F_Clostridia
F_Lactobacillales
F_Mollicutes
Fusobacteria
Green-nonsulfur
Green-sulfur
Hyperthermophilic
P_Alpha-others
P_Alpha-rhizobacteria
P_Alpha-rickettsia
P_Beta
P_Delta
P_Epsilon
P_Gamma-enterobacteria
P_Gamma-others
Planctomyces
Spirochete
Synergistetes
Verrucomicrobia
Ascomycetes
Plants
Nematode
Protist
Insect
C_Aeropyrum pernix K1
C_Caldivirga maquilingensis IC-167
C_Hyperthermus butylicus DSM 5456
C_Ignicoccus hospitalis KIN4/I
C_Metallosphaera sedula DSM 5348
C_Nitrosopumilus maritimus SCM1
C_Pyrobaculum aerophilum str. IM2
C_Pyrobaculum arsenaticum DSM 13514
C_Pyrobaculum calidifontis JCM 11548
C_Pyrobaculum islandicum DSM 4184
C_Staphylothermus marinus F1
C_Sulfolobus acidocaldarius DSM 639
C_Sulfolobus solfataricus P2
C_Sulfolobus tokodaii str. 7
C_Thermofilum pendens Hrk 5
C_Thermoproteus neutrophilus V24Sta
E_Candidatus Methanoregula boonei 6A8
E_Archaeoglobus fulgidus DSM 4304
E_Haloarcula marismortui ATCC 43049
E_Halobacterium salinarum R1
E_Halobacterium sp. NRC-1
E_Haloquadratum walsbyi DSM 16790
E_Methanobrevibacter smithii ATCC 35061
E_Methanocaldococcus jannaschii DSM 2661
E_Methanococcoides burtonii DSM 6242
E_Methanococcus aeolicus Nankai-3
E_Methanococcus maripaludis C5
E_Methanococcus maripaludis C6
E_Methanococcus maripaludis C7
E_Methanococcus maripaludis S2
E_Methanococcus vannielii SB
E_Methanocorpusculum labreanum Z
E_Methanoculleus marisnigri JR1
E_Methanopyrus kandleri AV19
E_Methanosaeta thermophila PT
E_Methanosarcina acetivorans C2A
E_Methanosarcina barkeri str. Fusaro
E_Methanosarcina mazei Go1
E_Methanosphaera stadtmanae DSM 3091
E_Methanospirillum hungatei JF-1
E_Methanothermobacter thermautotrophicus str. Delta H
E_Natronomonas pharaonis DSM 2160
E_Picrophilus torridus DSM 9790
E_Pyrococcus abyssi GE5
E_Pyrococcus furiosus DSM 3638
E_Pyrococcus horikoshii OT3
E_Thermococcus kodakarensis KOD1
E_Thermoplasma acidophilum DSM 1728
E_Thermoplasma volcanium GSS1
E_uncultured methanogenic archaeon RC-I
K_Candidatus Korarchaeum cryptofilum OPF8
N_Nanoarchaeum equitans Kin4-M

Figure 1

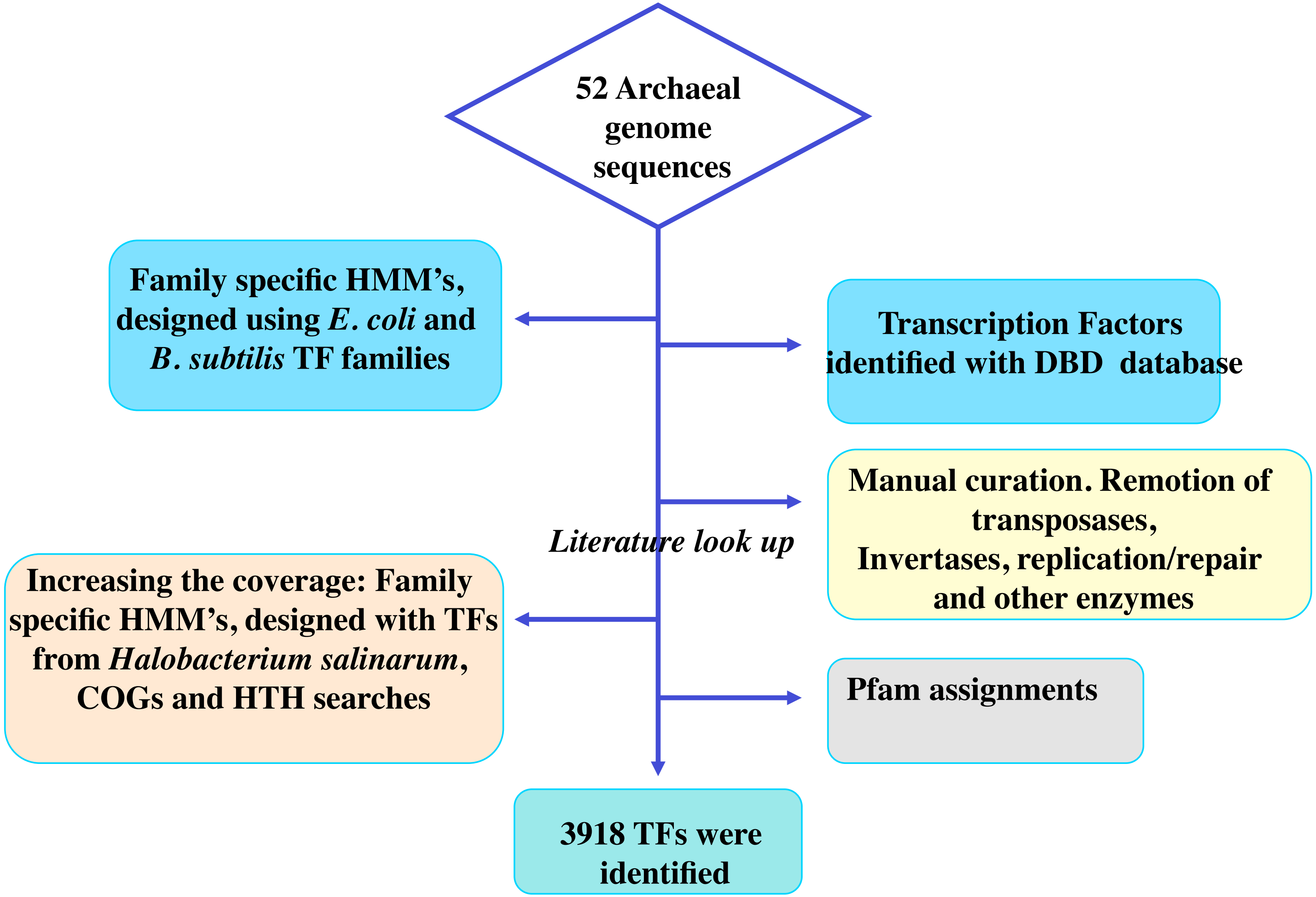

52 Archaeal genome sequences
Family specific HMM's, designed using *E. coli* and *B. subtilis* TF families
Transcription Factors identified with DBD database
Manual curation. Remotion of transposases, Invertases, replication/repair and other enzymes
Literature look up
Increasing the coverage: Family specific HMM's, designed with TFs from *Halobacterium salinarum*, COGs and HTH searches
Pfam assignments
3918 TFs were identified

Figure 2a

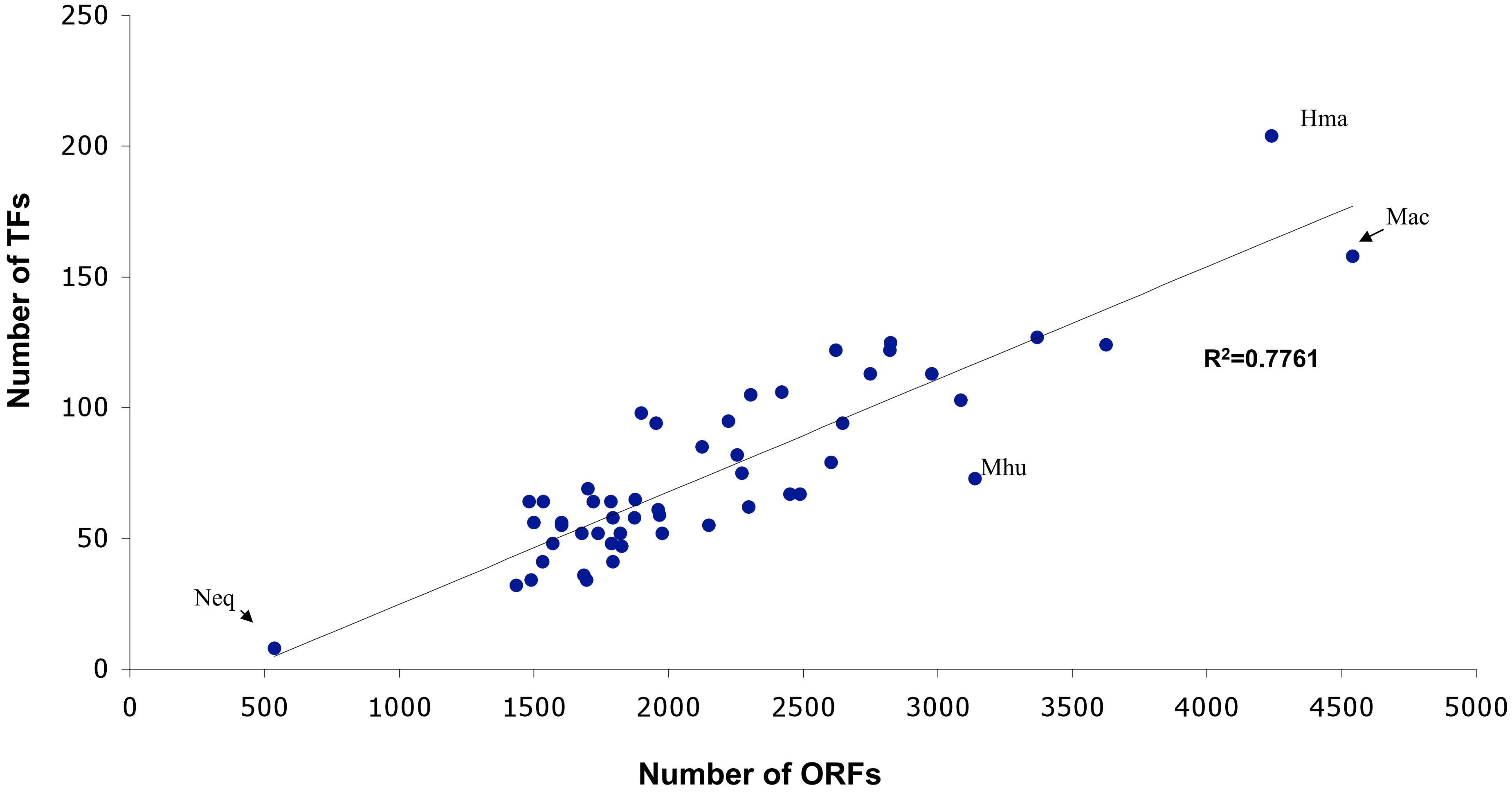

250
200
150
100
50
0
Number of TFs
Hma
Mac
R²=0.7761
Mhu
Neq
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
Number of ORFs

Figure 2b

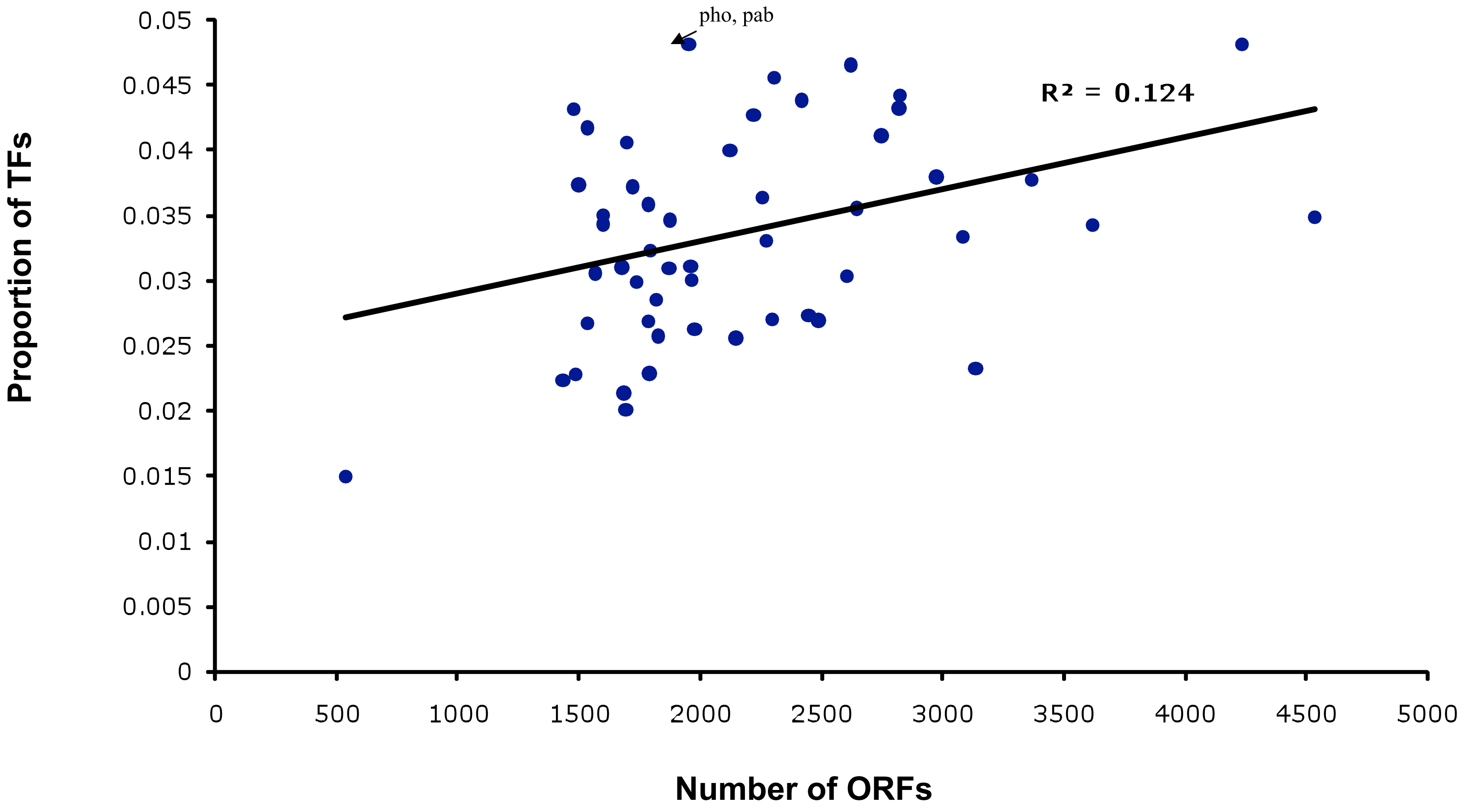

pho, pab
R² = 0.124
Proportion of TFs
0.05
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
Number of ORFs

Figure 3

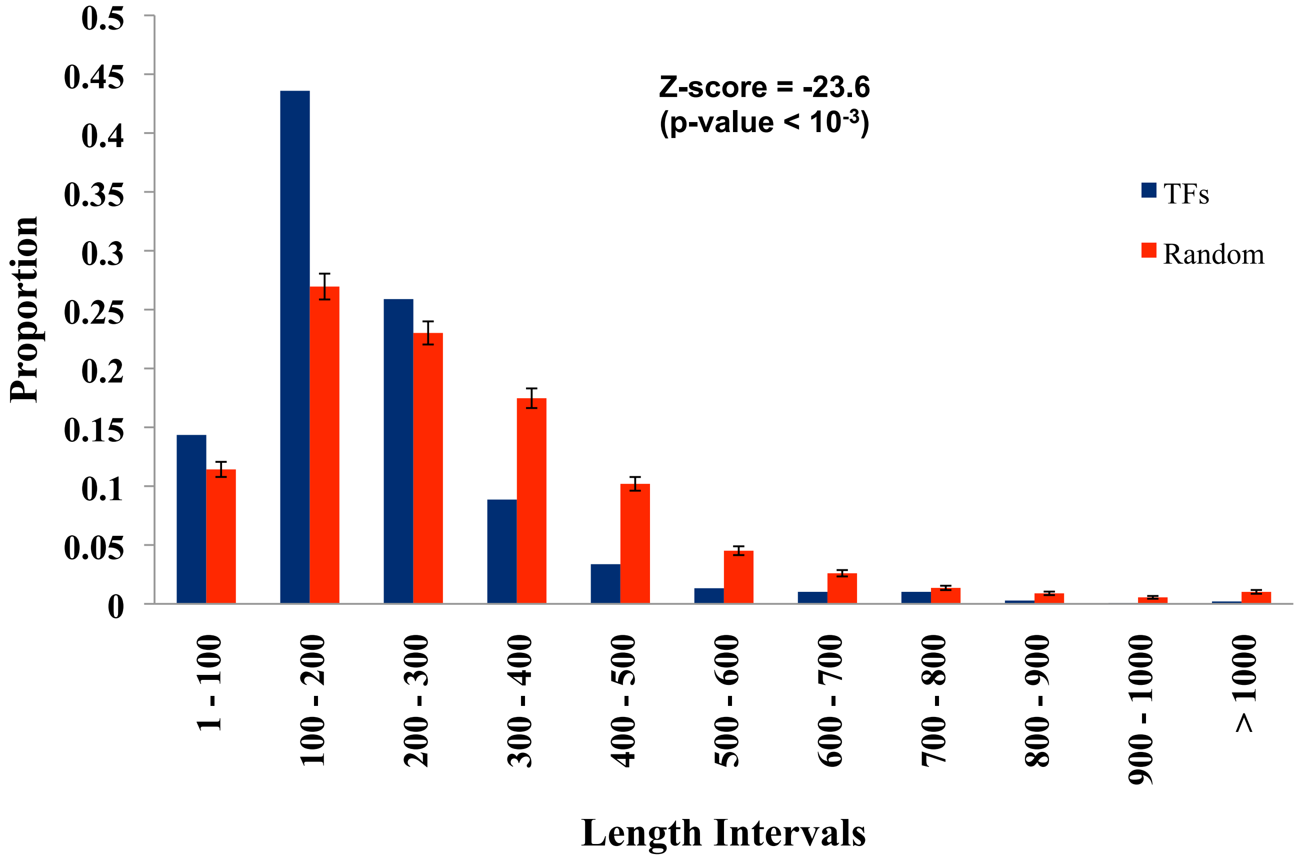
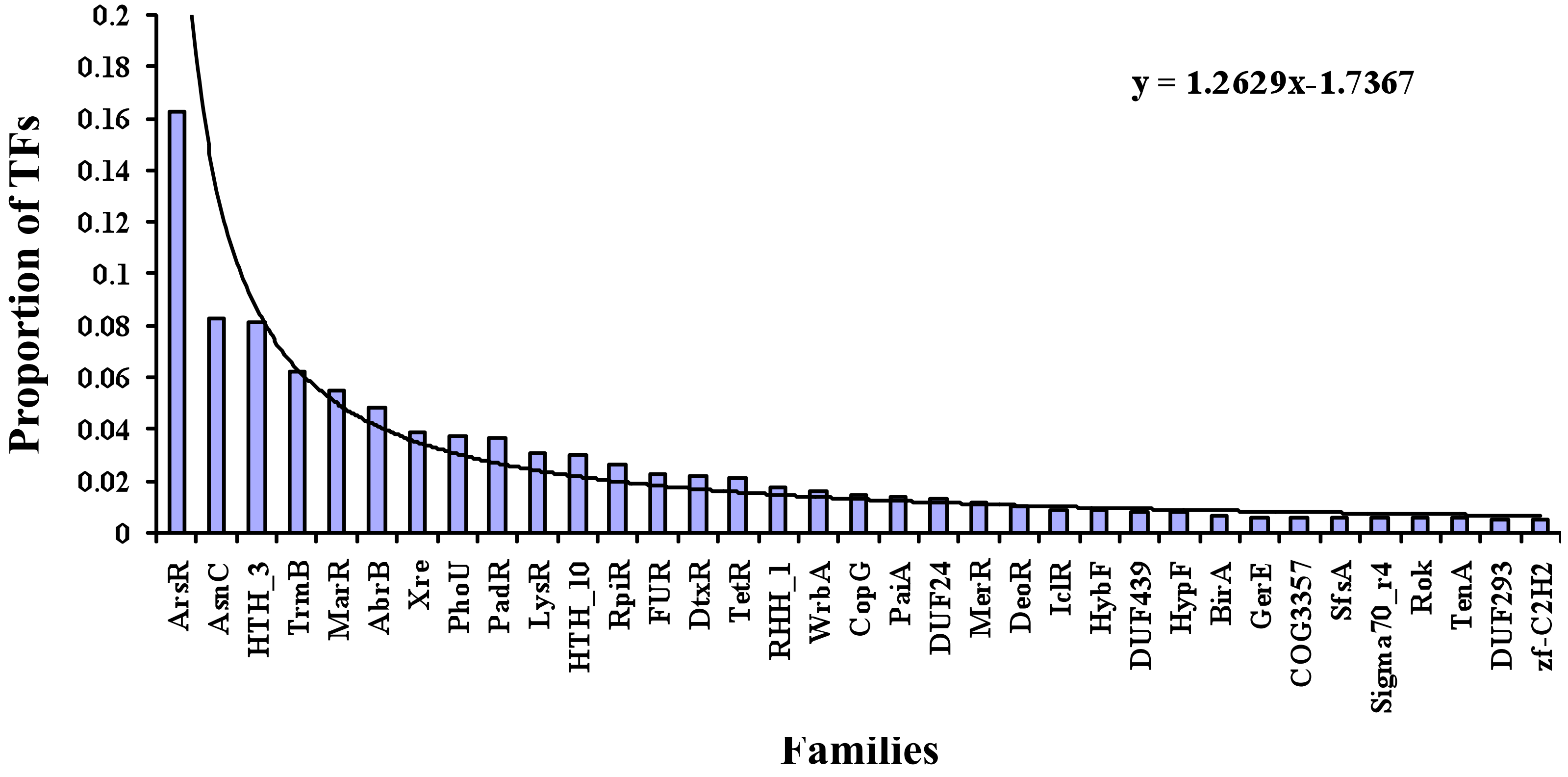

Z-score = -23.6
(p-value < 10^{-3})
TFs
Random
Proportion
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1 - 100
100 - 200
200 - 300
300 - 400
400 - 500
500 - 600
600 - 700
700 - 800
800 - 900
900 - 1000
> 1000
Length Intervals

Figure 4


y = 1.2629x-1.7367
Proportion of TFs
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
ArsR
AsnC
HTH_3
TrmB
MarR
AbrB
Xre
PhoU
PadR
LysR
HTH_10
RpiR
FUR
DtxR
TetR
RHH_1
WrbA
CopG
PaiA
DUF24
MerR
DeoR
IclR
HybF
DUF439
HypF
BirA
GerE
COG3357
SfsA
Sigma70_r4
Rok
TenA
DUF293
zf-C2H2
Families