Deposited research article

Conserved protein domains are maintained in an average ratio to proteome size

Joel A Malek and Daniel H Haft

Address: The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA.

Correspondence: Joel A Malek. E-mail: jamalek@tigr.org

Posted: 9 April 2001

Genome ***Biology*** 2001, **2(5)**:preprint0004.1–0004.16

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2001/2/5/preprint/0004

Received: 5 April 2001

This is the first version of this article to be made available publicly.
This article has been submitted to *Genome* ***Biology*** for peer review.

Conserved Protein Domains are Maintained in an Average Ratio to Proteome Size

Running Title: **Protein Domain Ratios Across Eukaryotes**

Authors:

Joel A. Malek and Daniel H. Haft

Address of Authors:

The Institute for Genomic Research

9712 Medical Center. Dr.

Rockville, MD 20850

USA

Corresponding Author: Joel A. Malek

Phone: (310) 838-0200

Fax: (301) 838-0208

e-mail: jamalek@tigr.org

Abbreviations: conserved domain (CD); Proteome Analysis Database (PAD)

Note: the terms gene and protein, are frequently used interchangeably in this paper as from a conserved domain perspective they are similar.

Abstract

Background:

Conserved domains (CD) in proteins play a crucial role in protein interactions, DNA binding, enzyme activity, and other important cellular processes. We proposed to study ratios of genes containing these domains to ratios of proteome size of different eukaryotes.

Results:

We have calculated average occurrences of conserved domains in each of 5 eukaryote genomes. Ratios between two genomes of genes containing a conserved domain, on average, reflected the ratio of the predicted total genes between the two genomes. Using two different databases of conserved domains, these ratios have been verified.

Conclusions:

Conserved domains are maintained in an averaged ratio to proteome size across the 5 sequenced eukaryotic genomes. This finding raises the question whether this ratio is maintained out of functional constraints, or other unknown reasons. The universality of the ratio in the 5 eukaryotic genomes attests to its potential importance.

Background

Conserved domains (CD) in proteins play a crucial role in protein interactions, DNA binding, enzyme activity, and other important cellular processes. With recently released gene number predictions in the human genome [1,2] being less than many previous predictions, interactions among these domains may prove to be central to proteome complexity. Protein domains are often conserved across many species, and as such, they offer an interesting dataset in how genomes maintain them with relationship to other conserved domains, as well as to proteome size. Many groups have attempted to find, document and annotate these conserved domains. While most groups use a form of Hidden Markov Models [3,4] for profiling, each group approaches the problem in a unique way yielding a wide range of databases that can be used to verify each other.

For this study we used the SMART conserved domain database [5,6,8] to collect data on the number of genes containing each CD in each genome. We restricted our study to the 5 eukaryote genomes sequenced so far, those being *H. sapiens*, *D. melanogaster*, *A. thaliana*, *C. elegans*, and *S. cerevisiae*. We confirmed our results using an independent source of similar data called the Proteome Analysis Database [7,9] (abbreviated here as PAD) and checked the sequenced eukaryotic genomes available at the time of writing, those being *D. melanogaster*, *C. elegans*, and *S. cerevisiae*.

We have used this unique opportunity to compare conserved domains across different genomes, and validated the approach by using two separate databases. The findings reveal a close link between numbers of genes with a given CD and the total number of genes in each genome.

Results and Discussion

Our initial observation was: for many conserved domains, the ratio of the sum genes in genome 1 containing the conserved domain to the total number of predicted genes in genome 1 was proportional to the ratio of sum genes in genome 2 containing the conserved domain to the total number of predicted genes in genome 2. Or:

Given that:

A = sum proteins with given CD in genome 1

B = sum proteins with given CD in genome 2

C = sum predicted genes in genome 1

D = sum predicted genes in genome 2

Then on average:

$$A/C \approx B/D \qquad \text{(Relationship 1)}$$

Upon rearranging Relationship 1, it was noted that for many CD's the ratio of the number of genes containing the given CD in each genome accurately reflected the ratio of the total predicted number of genes of each genome. Or:

Given variables in Relationship 1,

Then on average:

$$A/B \approx C/D \qquad \text{(Relationship 2)}$$

To normalize the data we used a ratio of the sum genes with a given CD in a genome, to the sum genes with the given CD in all 5 (3 for PAD) genomes. This was used to minimize the effect that the predicted number of genes may be significantly wrong for one of the genomes while the others

may be more accurate. Relationship 1 was rewritten to reflect this normalization. This resulted in the relationship:

Given that:

A = sum proteins with given CD in genome 1

E = sum proteins with given CD in 5(3 for PAD) genomes

C = sum predicted genes in genome 1

F = sum predicted genes for all 5(3 for PAD) genomes

Then on average:

$$A/E \approx C/F \qquad \text{(Relationship 3)}$$

The sums of CDs in each Relationship 3 ratio range were graphed for each genome, and are displayed in Figure 1 (SMART database) and Figure 2 (Proteome Analysis Database). The average ratio for each genome was calculated and multiplied against the sum predicted genes of all 5 genomes, yielding a number close to the predicted genes in each respective genome (Table 1).

Relationship 2 could be used to predict total genes in a genome given the other variables are reasonably well known, such as from Express Sequence Tag data. More importantly, this raises the question whether conserved domains are maintained in this ratio due to functional constraints or some other unknown reason. The fact that this ratio is maintained fairly well in all 5 eukaryotic genomes attests to its potential importance.

While there is much disagreement on total number of genes for the different genomes, similar gene finding methods were used for each of the 5 published eukaryotic genomes. It can therefore be assumed that ratios of predicted genes between the genomes will remain similar to present ratios as the gene numbers for each genome are further clarified. Likewise, neither SMART nor the Proteome Analysis Database claim to have found all occurrences of each CD in each genome. However, due to similar strategies used for CD finding in different genomes within each database, the ratio of total genes found with a given CD in each genome is likely to remain near constant as gene prediction improves.

An interesting finding from this research was that while ratios for *H. sapiens*, *A. thaliana*, and *S. cerevisiae* corresponded closely to total predicted genes for each organism, both databases gave a ratio that exchanged total predicted gene numbers between *D. melanogaster* and *C. elegans* (Figure 1, Figure 2, Table1). While this exchange cannot be explained presently, it may offer insight into distinctions between the genomes, and genes that remain unidentified.

It has been shown that conserved domains in proteins are maintained in proteome specific ratio for the 5 eukaryotic genomes sequenced so far. The reasons for this ratio are unclear, but it would not be unreasonable to suspect functional interaction of these domains requires they be kept in a specific ratio. Further research will be needed to understand the reasons for, and universality of this ratio in eukaryotic genomes.

Materials and Methods

For searches against the SMART database, we limited our data to conserved domains occurring at least once in each of the 5 genomes [8]. For the Proteome Analysis Database we restricted our search to those conserved domains listed in the top 200 occurring domains for which there was at least one occurrence in each of the 3 genomes [9]. This strategy of limiting the study to more global CD's was used to increase the chance that the conserved domains were constructed correctly and to increase statistical reliability of the results.

Data gathering was carried out as follows, a perl script was written to submit requests to the SMART database [8] for number of genes with each of 519 CDs in each genome. Information in the Proteome Analysis Database [9] is already in genome specific columns for the top 200 occurring CDs, and, as such, was downloaded directly. The information was parsed and stored for each genome. From the SMART database 211 conserved domains were selected based on the fact that they occurred at least once in each of the 5 genomes (see SMART_CDs.txt for information on these domains). From the Proteome Analysis Database 147 conserved domains were selected based on the fact that they occurred at least once in each of the 3 genomes (see PAD_CDs.txt for information on these domains).

The total number of predicted genes for each genome was as follows: *H. sapiens*, 35,000 [1,2]; *D.melanogaster*, 14,100 [10,11]; *A. thaliana*, 26,000 [11,12,13]; *C. elegans*, 19,100 [11,14]; *S. cerevisiae*, 6,300 [11]. This yielded a total of 100,500 genes for all 5 genomes, and a total of 39,500 for *D. melanogaster*, *C.elegans*, and *S.cerevisiae* alone. The number of genes in each of the eukaryotic genomes is an approximate number because the number of genes predicted is always a changing estimate constantly being clarified (13).

Acknowledgements

Thank you to those at TIGR who reviewed the ideas presented here. Thank you to S. Malek for critical review, SDG.

References:

1. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, *et al*.: **The sequence of the human genome.** *Science* 2001, **291**(5507):1304-1351.

2. International Human Genome Sequencing Consortium : **Initial sequencing and analysis of the human genome.** *Nature* 2001 **409:** 860-921.

3. Sonnhammer EL, Eddy SR, Birney E, Bateman A, Durbin R: **Pfam: multiple sequence alignments and HMM-profiles of protein domains.** *Nucleic Acids Res* 1998, **26**(1):320-322.

4. 2. Eddy SR, Mitchison G, Durbin R: **Maximum discrimination hidden Markov models of sequence consensus.** *J Comput Biol* 1995 **2**(1):9-23.

5. 3. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P: **SMART: a web-based tool for the study of genetically mobile domains.** *Nucleic Acids Res* 2000 **28**(1):231-234.

6. Schultz J, Milpetz F, Bork P, Ponting CP: **SMART, a simple modular architecture research tool: identification of signaling domains.** *Proc Natl Acad Sci U S A* 1998, **95**(11):5857-5864.

7. Apweiler R, Biswas M, Fleischmann W, Kanapin A, Karavidopoulou Y, Kersey P, Kriventseva EV, Mittard V, Mulder N, Phan I, Zdobnov E.: **Proteome Analysis Database: online application of InterPro and CluSTr for the functional classification of proteins in whole genomes.** *Nucleic Acids Res* 2001, **29**(1):44-48.

8. **SMART - Simple Modular Architecture Research Tool** [http://smart.embl-heidelberg.de]

9. **Proteome Analysis Database top 200 Domains** [www.ebi.ac.uk/proteome/DROME/interpro/comparison/top200.html]

10. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, *et al.*: **The genome sequence of Drosophila melanogaster.** *Science* 2000, **287**(5461):2185-2195.

11. Drosophila gene numbers from [www.fruitfly.org/sequence/download.html], C. elegans gene numbers from [www.sanger.ac.uk/Projects/C_elegans/wormpep], and S. cerevisiae gene numbers from[ftp://genome-ftp.stanford.edu/pub/yeast/yeast_ORFs](12).

12. The Arabidopsis Genome Initiative: **Analysis of the genome sequence of the flowering plant Arabidopsis thaliana.** *Nature* 2000 **408:** 796-815

13. Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, *et al.*: **Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes.** *Science* 2000, **290**(5499):2105-2110.

14. The C. elegans Sequencing Consortium: **Genome Sequence of the Nematode C. elegans: A Platform for Investigating Biology.** *Science* 1998, **282**:2012-2018.

Additional Files

1. SMART_CDs.txt is a text, tab delimited file containing all 211 conserved domain names from the SMART database used in this study. For each conserved domain name, the corresponding number of genes containing the CD in each genome is listed.

2. PAD_CDs.txt is a text, tab delimited file containing all 147 InterPro entry numbers for the domains in the Proteome Analysis Database used in this study. For each InterPro entry number, the corresponding number of genes containing the CD in each genome is listed.

	SMART Database			Proteome Analysis Database			
Organism	Average ratio of genes with CD in organism to total genes with CD in 5 organisms	Sum of predicted genes for all 5 organisms	Product	Average ratio of genes with CD in organism to total genes with CD in 3 organisms	Sum of predicted genes for all 3 organisms	Product	Predicted total Genes in Genome
H.sapiens	0.386	100500	38793				35000
D.melanogaster	0.172	100500	17286	0.454	39500	17933	14100
A.thaliana	0.283	100500	28442				26000
C.elegans	0.158	100500	15879	0.405	39500	15998	19100
S.cerevisiae	0.076	100500	7638	0.141	39500	5569.5	6300

Table 1. Comparison of averaged conserved domain ratios between SMART and Proteome Analysis Database and their relationship to Proteome size. Relationship 3 was used for all conserved domains and the results were averaged for each genome. A comparison between both databases yields very similar results for *D. melanogaster* and *C. elegans*. *S. cerevisiae* data differed between the databases but the differences are centered around the predicted total number of genes. Interestingly, both databases gave a ratio of *D. melanogaster* and *C. elegans* that exchanged their predicted total genes.

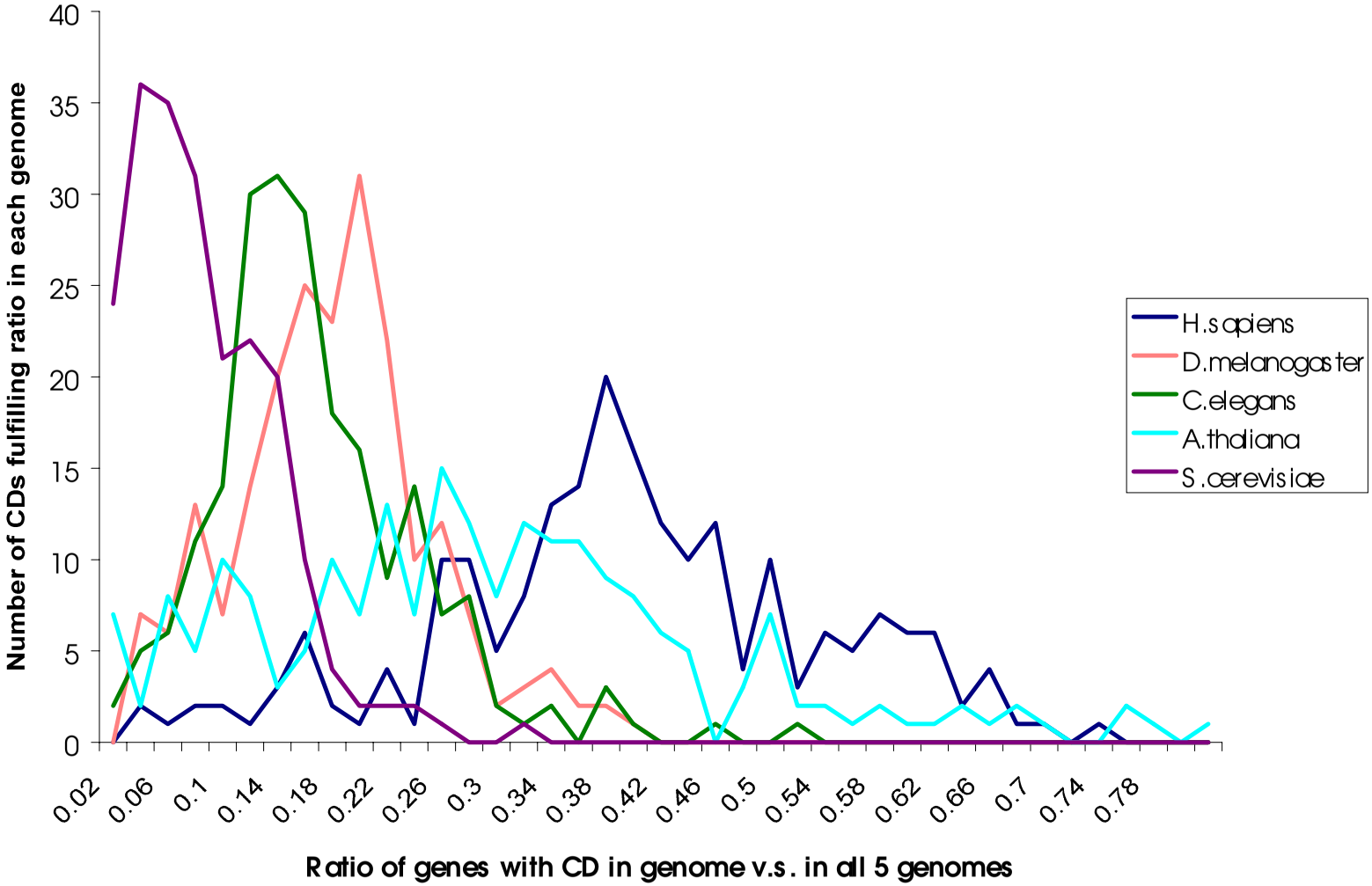

Figure 1. Sum CDs in each ratio of conserved domains in genome to occurrences in all 5 genomes (211 CDs considered) - SMART database. Relationship 3 was used for all conserved domains for each genome. The number of conserved domains in each ratio for each genome were summed and graphed. The sum of all predicted genes for the 5 genomes was 100,500. It is apparent that each genome peaks, and is averaged near their respective proteome size (multiply average ratio for each genome by 100,500 as in Table 1).

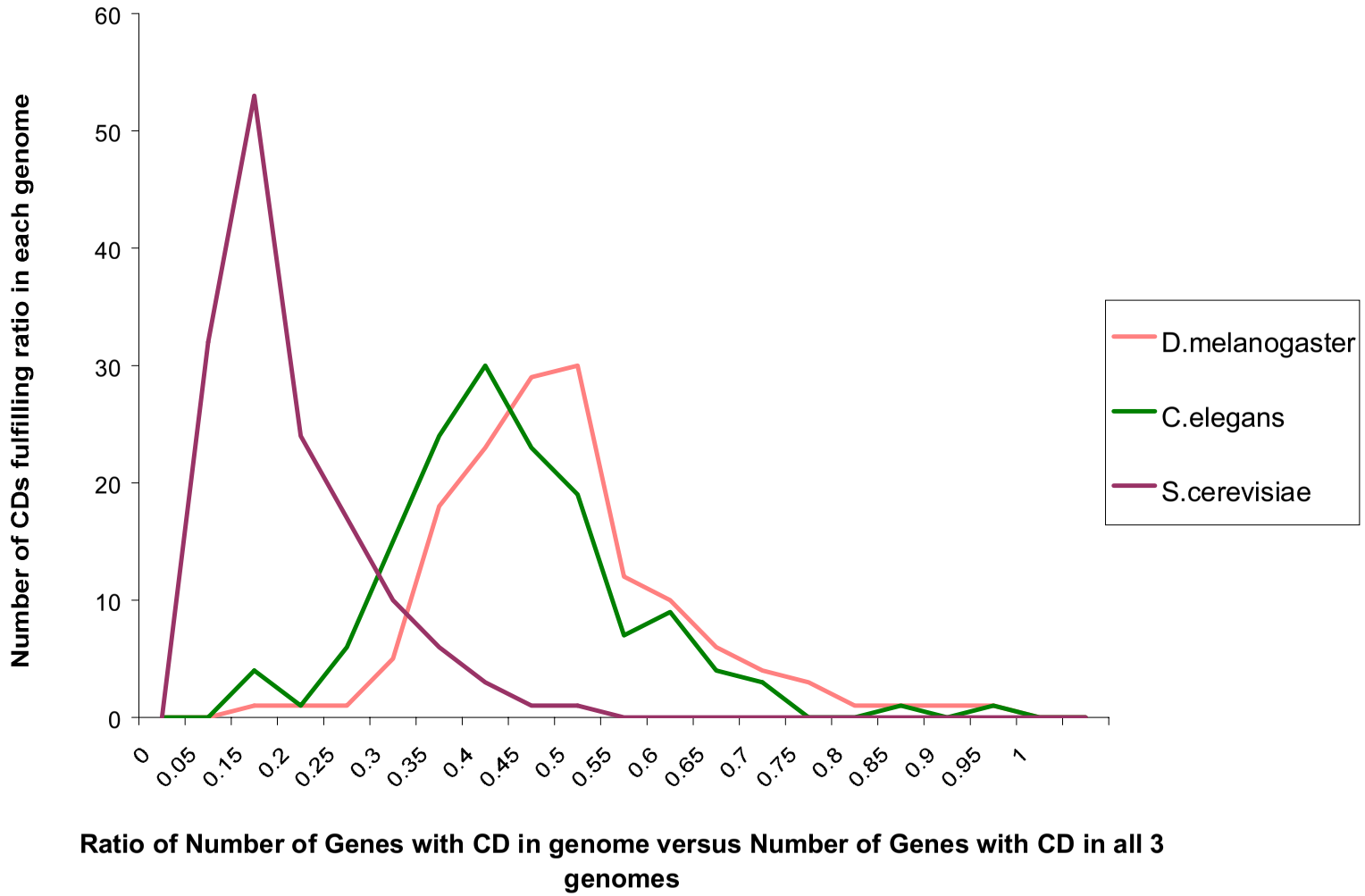

Figure 2. Sum CDs in each ratio of conserved domains in genome to 3 organisms (147 CDs considered) - Proteome Analysis Database. Relationship 3 was used for all conserved domains for each genome. The number of conserved domains in each ratio for each genome were summed and graphed. The sum of all predicted genes for the 3 genomes was 39,500. It is apparent that each genome peaks, and is averaged near their respective proteome size (multiply average ratio for each genome by 39,500 as in Table1). Compare the results of the 3 genomes here with those in Figure 1.

SMART	Conserved Domain name	Number of genes with conserved domain in : H.sapiens	D.melanogaster	A.thaliana	C.elegans	S.cerevisiae
14_3_3		20	4	25	2	3
ADF		18	7	16	7	5
ANK		253	104	123	100	19
ARF		33	11	19	10	5
ArfGap		29	10	22	12	6
ARM		51	19	45	4	2
BAG		7	2	6	4	2
BIR		10	6	2	2	1
C1		80	48	72	42	1
C2		196	57	112	66	11
CH		84	43	16	28	4
cNMP		47	33	41	22	2
CULLIN		14	8	6	10	3
DAGKc		18	11	12	8	2
DEP		33	8	1	14	6
DSPc		37	9	5	10	7
DYNc		14	4	27	6	3
EFh		178	73	137	46	7
EH		23	7	3	8	5
ENTH		16	8	25	10	8
FBOX		54	20	492	174	9
FF		13	3	4	3	2
FH2		13	6	19	8	3
FHA		45	21	19	15	14
FYVE		38	13	14	20	6
G-alpha		31	16	7	24	2
GAF		16	2	14	2	1
GED		15	4	25	5	2
GGL		28	6	1	5	1
GRAM		16	4	14	4	6
GuKc		47	13	6	9	1
HATPase_c		15	5	39	6	5
HECTc		50	15	9	8	5
IPPc		23	6	22	7	4
IQ		101	46	71	22	4
KISc		48	39	71	32	6
LIM		124	48	12	41	4
LMWPc		3	3	1	1	1
MIR		21	7	2	10	7
MYSc		32	17	22	22	5
OPR		15	17	17	2	1
PAS		61	25	18	12	4
PBD		32	14	14	12	5
PDZ		236	85	18	76	2
PH		353	102	37	92	29
PI3K_C2		11	6	2	5	1
PI3Ka		13	6	4	7	2
PI3Kc		31	13	18	15	8
PIPKc		13	5	21	4	2
PLCXc		19	8	15	10	1

PLCYc	21	6	12	10	1	
PLDc	5	6	16	6	3	
PP2Ac	27	23	33	52	12	
PP2Cc	20	16	85	12	8	
PTPc	78	29	3	90	3	
PTPc_DSPc		73	14	21	21	3
PX	50	17	9	13	15	
PXA	3	1	2	3	1	
RAB	82	31	72	30	9	
RAN	5	3	7	2	2	
RanBD	13	4	10	3	4	
RGS	47	13	1	23	4	
RHO	36	11	20	12	6	
RhoGAP		86	25	9	22	12
RIIa	13	4	1	2	1	
RING	367	140	491	160	42	
RIO	4	3	3	3	2	
S_TK_X		83	50	16	35	12
S_TKc	389	167	312	157	67	
SAM	104	41	10	17	4	
SAR	5	2	6	1	1	
SEC14	25	37	38	14	6	
Sec7	23	7	8	7	5	
SH2	175	54	3	71	1	
SH3	354	106	4	65	25	
small_GTPase			66	28	23	46	12
SPRY	83	13	7	11	5	
t_SNARE		27	16	41	15	14
TBC	47	23	24	19	13	
UBA	53	35	27	13	9	
UBCc	72	39	76	25	17	
UBQ	59	24	70	25	9	
UBX	18	10	15	3	7	
VHS	19	3	10	4	4	
VPS9	9	4	2	3	2	
WD40	405	196	270	140	102	
WW	79	27	11	17	6	
ZnF_AN1		7	3	12	3	2
ZnF_RBZ		35	19	30	8	2
ZnF_UBP		14	6	12	5	4
ZnF_ZZ		39	36	18	14	2
AAA	141	101	209	77	62	
acidPPc		17	12	9	5	6
ACTIN	43	19	26	13	12	
BTB	175	124	57	133	3	
CBS	27	6	28	18	10	
CLH	11	6	9	3	3	
CUE	13	4	1	4	7	
CYCLIN		46	34	73	23	18
DnaJ	50	37	110	35	22	
EZ_HEAT		2	1	2	1	1
GYF	7	3	8	3	3	
HAT	21	11	14	8	7	
HDc	57	11	13	7	4	
JAB_MPN		21	13	19	7	4
LON	5	2	10	3	1	
LRRcap		27	9	4	6	1

MIF4G	16	11	11	6	6
NDK	18	7	7	2	1
PHB	19	13	15	13	2
PINT	39	17	23	14	8
PlsC	21	12	26	16	7
PolyA	10	3	9	3	1
PP2C_SIG		14	14	71	8	6
PROF	5	1	6	3	1
Pumilio		5	3	23	12	7
PWI	8	2	8	3	1
PWWP	44	7	15	1	2
R3H	8	7	8	3	3
RHOD	37	10	22	12	11
S1	14	15	25	10	7
SNc	1	1	4	1	1
SWIB	5	3	18	3	2
SynN	15	4	24	7	3
TGc	14	4	1	3	3
TPR	73	39	48	22	12
ZnF_UBR1		7	10	3	6	2
CysPc	21	8	1	15	1
EGF_like		200	63	30	63	1
FN3	254	96	2	60	2
LDLa	57	55	2	41	1
LRR	188	110	423	48	6
LRR_TYP		124	88	25	29	1
MATH	16	6	70	93	1
PQQ	2	3	1	2	1
SCP	12	23	25	36	3
VWA	109	13	22	54	4
35EXOc		3	6	25	5	1
A1pp	15	2	4	2	2
ADEAMc		12	4	1	5	1
AT_hook		46	35	42	16	6
BAH	16	9	29	4	5
BRCT	34	17	19	29	11
BRIGHT		26	8	9	5	2
BRLZ	93	43	93	27	16
BROMO	70	27	29	17	9
CBF	1	1	13	2	1
CHROMO		49	24	26	26	4
DEXDc	144	86	148	84	58
DSRM	56	27	19	13	2
eIF5C	13	7	8	2	2
ENDO3c		15	2	14	1	4
EXOIII		22	9	14	12	5
FES	13	1	7	1	1
G-patch		35	18	15	14	5
H15	20	3	19	12	1
H2A	29	2	13	16	3
H2B	27	2	15	7	2
H3	23	4	16	14	2
H4	13	1	10	14	1
HELICc		176	90	180	82	74
HhH1	16	4	11	2	2
HhH2	11	9	9	4	4
HLH	171	82	154	43	10

HMG	103	31	20	18	10	
HOX	293	174	120	118	10	
HRDC	4	3	6	3	2	
HSF	7	1	31	2	5	
HTH_XRE		2	1	4	1	1
JmjC	46	12	20	11	5	
JmjN	15	4	9	2	3	
KH	64	45	39	33	9	
Ku78	2	5	5	2	2	
MA3	10	6	12	2	2	
MADS	6	5	138	3	4	
MCM	18	12	10	7	6	
MUTSac		11	2	9	7	6
MUTSd	10	2	8	6	6	
PHD	164	57	190	45	16	
POLAc	3	3	3	2	1	
POLBc	10	9	6	4	4	
PostSET		29	13	19	10	2
PUA	8	7	4	3	5	
RIBOc	9	3	10	3	3	
RRM	346	167	283	121	54	
S4	9	5	17	4	7	
SANT	67	41	402	22	19	
SAP	32	11	10	7	5	
SET	62	37	46	33	8	
SFM	5	2	2	2	1	
Skp1	5	11	23	22	3	
SMR	1	4	15	6	2	
TFIIE	2	1	3	1	1	
TFS2M	9	3	7	1	2	
TFS2N	9	3	14	1	1	
TOP1Ac		3	3	3	3	1
TOP1Bc		3	3	3	3	1
TOP2c	9	1	4	4	1	
TOP4c	7	1	5	4	2	
TOPEUc		4	1	3	3	1
TOPRIM		3	3	5	3	1
Ubox	12	7	56	5	2	
XPGI	10	9	6	4	7	
XPGN	11	9	8	4	6	
ZnF_C2C2		9	6	5	4	5
ZnF_C2H2		868	429	185	212	51
ZnF_C2HC		64	51	261	50	15
ZnF_C3H1		55	23	62	35	7
ZnF_GATA		18	12	26	15	11
ZnF_NFX		7	3	2	2	1
ZnF_U1		26	16	10	7	5

PAD_CDs

PAD Domains Number of genes with Conserved Domain in: InterPro Domain Entry	Number	D.melanogaster	C. elegans	S.cerevisiae
IPR000822	356	221	54	
IPR000719	240	417	118	
IPR001254	207	15	1	
IPR001680	184	148	102	
IPR000504	156	131	60	
IPR002290	156	214	113	
IPR000379	134	118	37	
IPR002048	134	119	19	
IPR001841	117	152	40	
IPR001611	108	55	7	
IPR000561	106	191	1	
IPR001356	103	99	7	
IPR003662	99	74	45	
IPR001128	91	80	3	
IPR003592	90	42	6	
IPR002110	86	98	18	
IPR001440	83	56	33	
IPR000130	81	110	8	
IPR001245	81	115	4	
IPR002403	81	10	1	
IPR001410	78	77	76	
IPR000210	77	135	3	
IPR003593	77	69	57	
IPR001452	75	61	24	
IPR001849	75	76	29	
IPR003439	75	68	37	
IPR001650	74	73	74	
IPR001806	70	59	23	
IPR001478	69	72	2	
IPR003015	69	67	8	
IPR003591	69	26	1	
IPR001993	67	59	40	
IPR000734	62	39	8	
IPR001092	61	39	8	
IPR001777	58	47	2	
IPR002106	58	108	22	
IPR002198	56	84	13	
IPR001005	48	34	21	
IPR001064	46	78	6	
IPR002347	46	33	2	
IPR002172	45	29	1	
IPR000008	43	47	11	
IPR000345	43	46	6	
IPR001622	43	79	1	
IPR001965	43	43	16	
IPR000063	40	47	12	
IPR000521	40	54	6	
IPR000051	39	47	20	
IPR001939	39	33	33	
IPR003880	38	23	6	

IPR001251	37	16	6
IPR000169	36	47	1
IPR001781	36	30	4
IPR000387	35	93	10
IPR000980	34	67	1
IPR001623	34	37	22
IPR000608	32	23	15
IPR001660	32	17	4
IPR002173	32	6	3
IPR000834	31	25	3
IPR000873	31	28	10
IPR000886	31	26	16
IPR000934	31	66	20
IPR001412	31	22	15
IPR001454	31	35	28
IPR000047	30	13	2
IPR000048	30	18	4
IPR000910	30	13	9
IPR000958	30	27	9
IPR001878	30	43	14
IPR002086	30	27	10
IPR002219	30	34	1
IPR000961	29	25	12
IPR001214	29	31	7
IPR001715	29	23	4
IPR001871	29	31	17
IPR003579	29	27	9
IPR000408	28	21	4
IPR000595	28	14	2
IPR000073	27	18	15
IPR000637	27	17	6
IPR003560	27	33	7
IPR000205	26	23	10
IPR000629	26	23	20
IPR001752	26	20	6
IPR002067	26	10	7
IPR000194	25	16	27
IPR000182	24	13	13
IPR000219	24	20	6
IPR000449	24	13	8
IPR000626	24	30	8
IPR001054	24	37	1
IPR001202	24	23	9
IPR001353	24	13	14
IPR003575	24	34	12
IPR000217	23	17	4
IPR001810	23	353	11
IPR002113	23	5	4
IPR001140	22	30	10
IPR001594	22	16	7
IPR002017	22	13	2
IPR000195	21	19	11
IPR000198	21	19	11
IPR002111	21	23	2
IPR000553	20	20	18
IPR000795	20	17	14
IPR000953	20	24	4

IPR001283	20	36	3
IPR001394	20	28	18
IPR001930	20	18	4
IPR002453	20	17	1
IPR000225	19	6	2
IPR000531	19	19	2
IPR000571	19	32	6
IPR000690	19	9	7
IPR001395	19	25	6
IPR001766	19	18	4
IPR001969	19	31	43
IPR002422	19	15	7
IPR003095	19	7	3
IPR000166	18	33	15
IPR000242	18	88	3
IPR000330	18	22	17
IPR000340	18	28	7
IPR000636	18	23	1
IPR002130	18	18	8
IPR000253	17	13	15
IPR000467	17	14	5
IPR001173	17	14	4
IPR001487	17	16	10
IPR001734	17	3	1
IPR001798	17	14	6
IPR002076	17	8	3
IPR000014	16	10	4
IPR000717	16	16	8
IPR001163	16	13	16
IPR001223	16	34	1
IPR001327	16	10	9
IPR001609	16	17	5
IPR002078	16	11	1
IPR002293	16	19	25
IPR000279	15	10	10
IPR000577	15	5	3
IPR000727	15	15	15
IPR000901	15	9	8
IPR001357	15	31	10
IPR001683	15	12	15