Evolutionary expansion of the Ras switch regulatory module in eukaryotes

(Supporting Material)

Diego Díez[1, 2 *], Francisca Sánchez-Jiménez[3] and Juan A. G. Ranea[3]

[1]Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011 Japan

[2]Instituto de Investigaciones Biomédicas, CSIC-UAM, 28029 Madrid, Spain

[3]Department of Molecular Biology and Biochemistry-CIBER de Enfermedades Raras, University of Malaga, Malaga, Spain

***Corresponding author:**

Diego Diez
Bioinformatics Center, Institute for Chemical Research, Kyoto University
Gokasho, Uji, Kyoto 611-0011, JAPAN
Phone: +81-0774-38-3296
Fax: +81-0774-38-3269
E-mail: diez@kuicr.kyoto-u.ac.jp

Table S1 List of species included in this analysis. Most genomes were downloaded from EnsEMBL. Scientific name, short name, source database, assembly version and URL are indicated.

Species	***Alias***	***Assembly***	***Source/Url***
S. cerevisiae	yeast	SGD1.01	
C. elegans	worm	WS190	
A. gambiae	mosquito	AgamP3	
D. melanogaster	fly	BDGP 5.4	
D. rerio	zebrafish	Zv8	
T. rubripes	fugu	FUGU 4.0	
C. intestinalis	ciona	JGI 2	
X. tropicalis	xenopus	JGI 4.1	
G. gallus	chicken	WASHUC2	EnsEMBL
T. guttata	finch	TG-3.2.4	http://www.ensembl.org
A. carolinensis	anolis	AnoCar1.0	
O. anatinus	platypus	OA-5.0	
M. domestica	monodephils	monDom5	
B. taurus	cow	Btau_4.0	
R. norvegicus	rat	RGSC 3.4	
M. musculus	mouse	NCBIM37	
P. troglodites	chimp	CHIMP2.1	
H. sapiens	human	NCBI36	
C. albicans	candida	21	CandidaGenomeDB http://www.candidagenome.org
S. pombe	pombe		Sanger http://www.sanger.ac.uk/Projects/S_pombe
D. discoideum	slime		DictyBase http://dictybase.org
S. purpuratus	seaurchin	2.1	SpBase http://www.spbase.org/SpBase/index.php
M. brevicollis	monosiga	1	http://genome.jgi-psf.org/Monbr1
A. thaliana	arabidopsis	TAIR8	Gramene http://www.gramene.org

Table S2 Rates of duplication relative to Ras for individual species.

Organism	*GEF*	*RasGAP*	*RapGAP*
human	0.78	0.41	0.27
chimp	0.75	0.42	0.28
mouse	0.68	0.37	0.26
rat	0.62	0.33	0.23
cow	0.79	0.39	0.30
opossum	0.58	0.30	0.20
platypus	0.67	0.44	0.37
chicken	0.65	0.35	0.26
finch	0.53	0.41	0.26
anolis	0.69	0.26	0.23
xenopus	0.70	0.30	0.18
zebrafish	0.59	0.35	0.30
fugu	0.71	0.37	0.34
ciona	0.58	0.42	0.33
seaurchin	0.19	0.15	0.09
mosquito	0.45	0.20	0.30
fly	0.50	0.25	0.20
worm	0.73	0.45	0.27
monosiga	1.86	1.43	0.57
pombe	1.00	1.00	0.50
yeast	0.80	0.80	0.00
candida	1.33	1.00	0.33
slime	1.45	0.70	0.55

Figure S1 Correlation between gene numbers and genome size/divergence time before removing the effect of outliers. The linear fit includes all species except *Arabidopsis*.

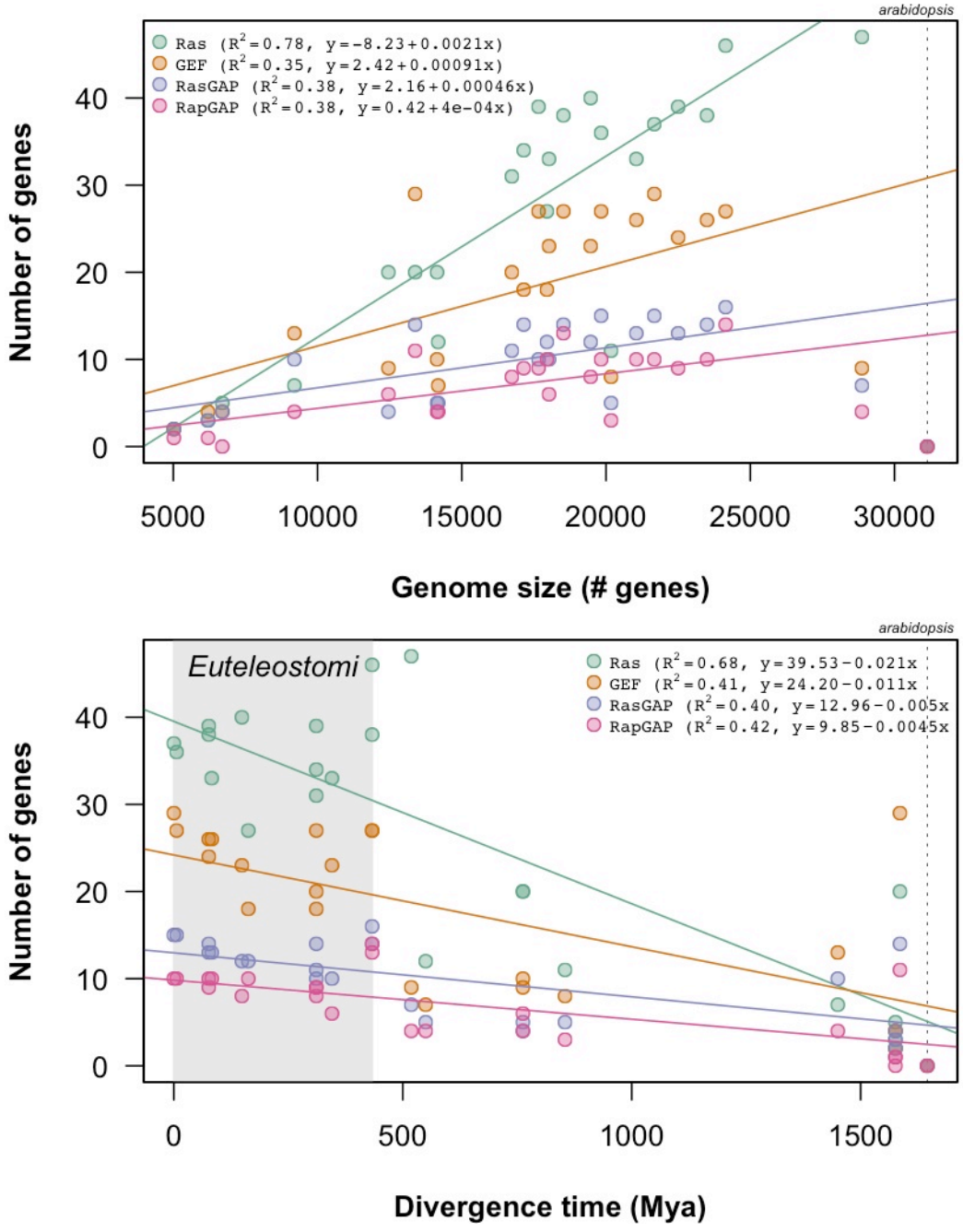


Figure S6 Frequency of domains (unique) per architecture (number of genes with each number of domains in their sequence).

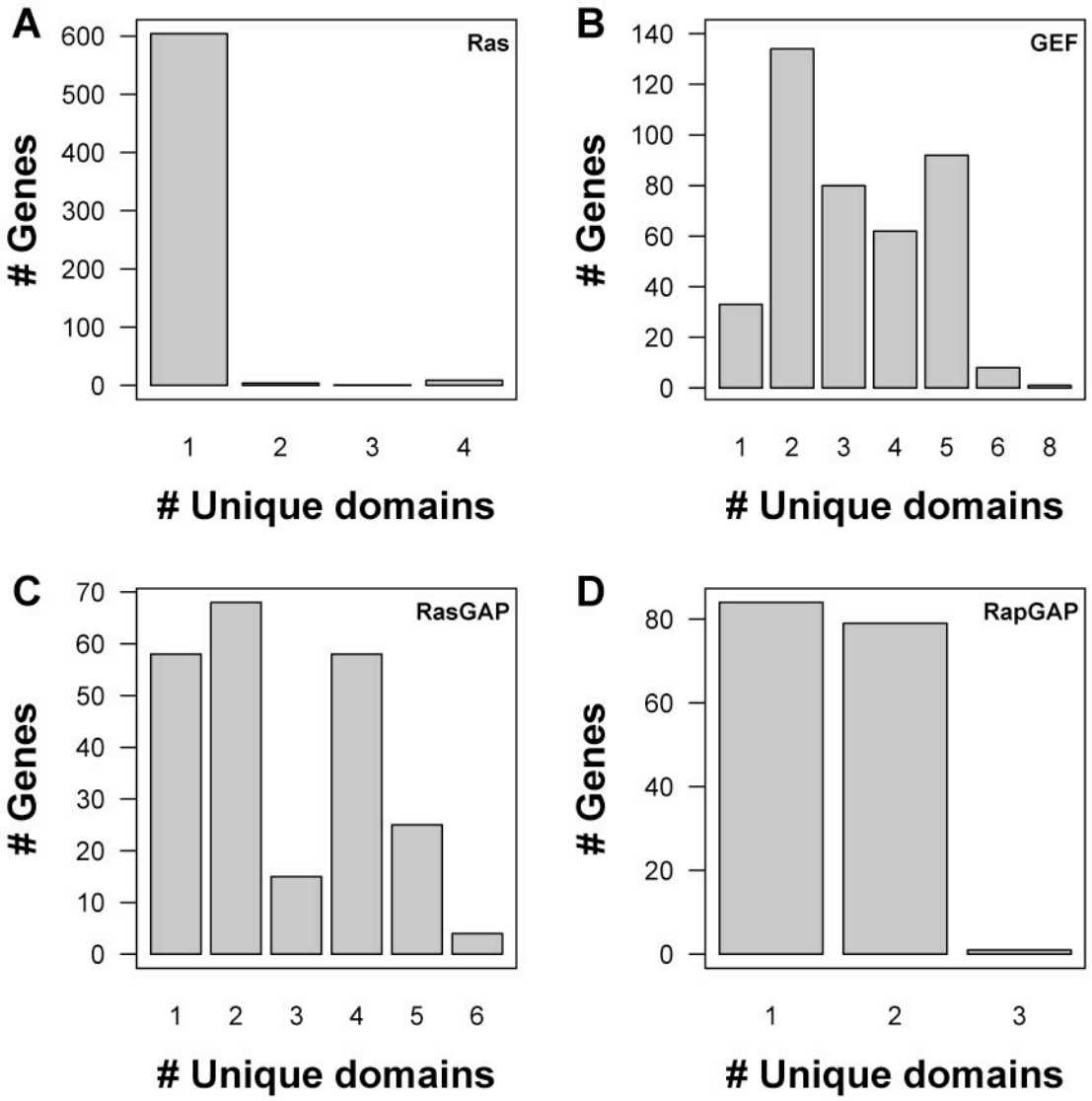


Figure S7 Frequency of domains per architecture (number of genes with each number of domains in their sequence).

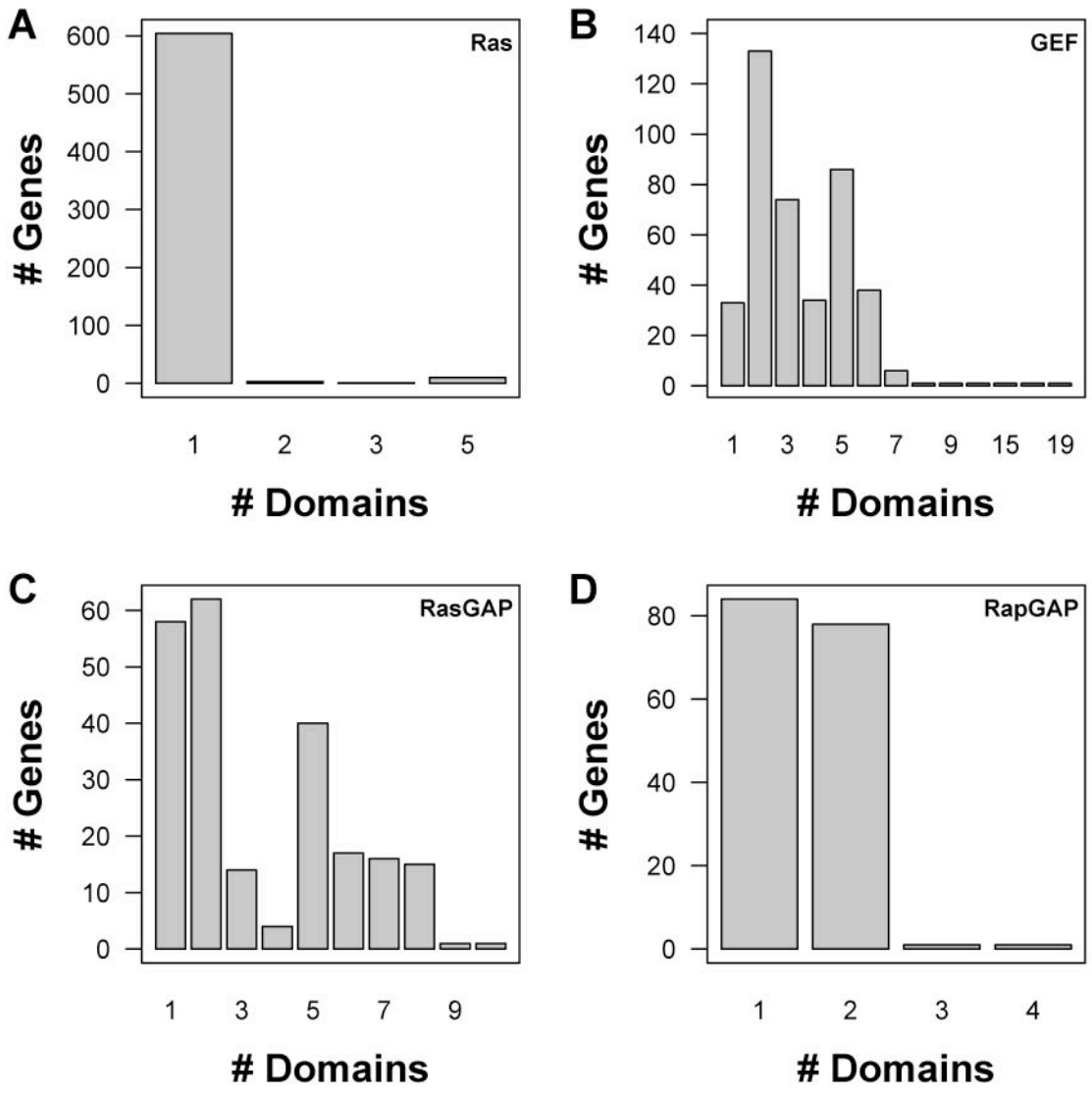


Figure S8 Distribution of domain counts across species for A) Ras, B) RapGAP, C) RasGAP and D) RasGEF. The color scale indicates the percentage of sequences with that domain in the corresponding species. Domain frequency distribution is shown for each family.

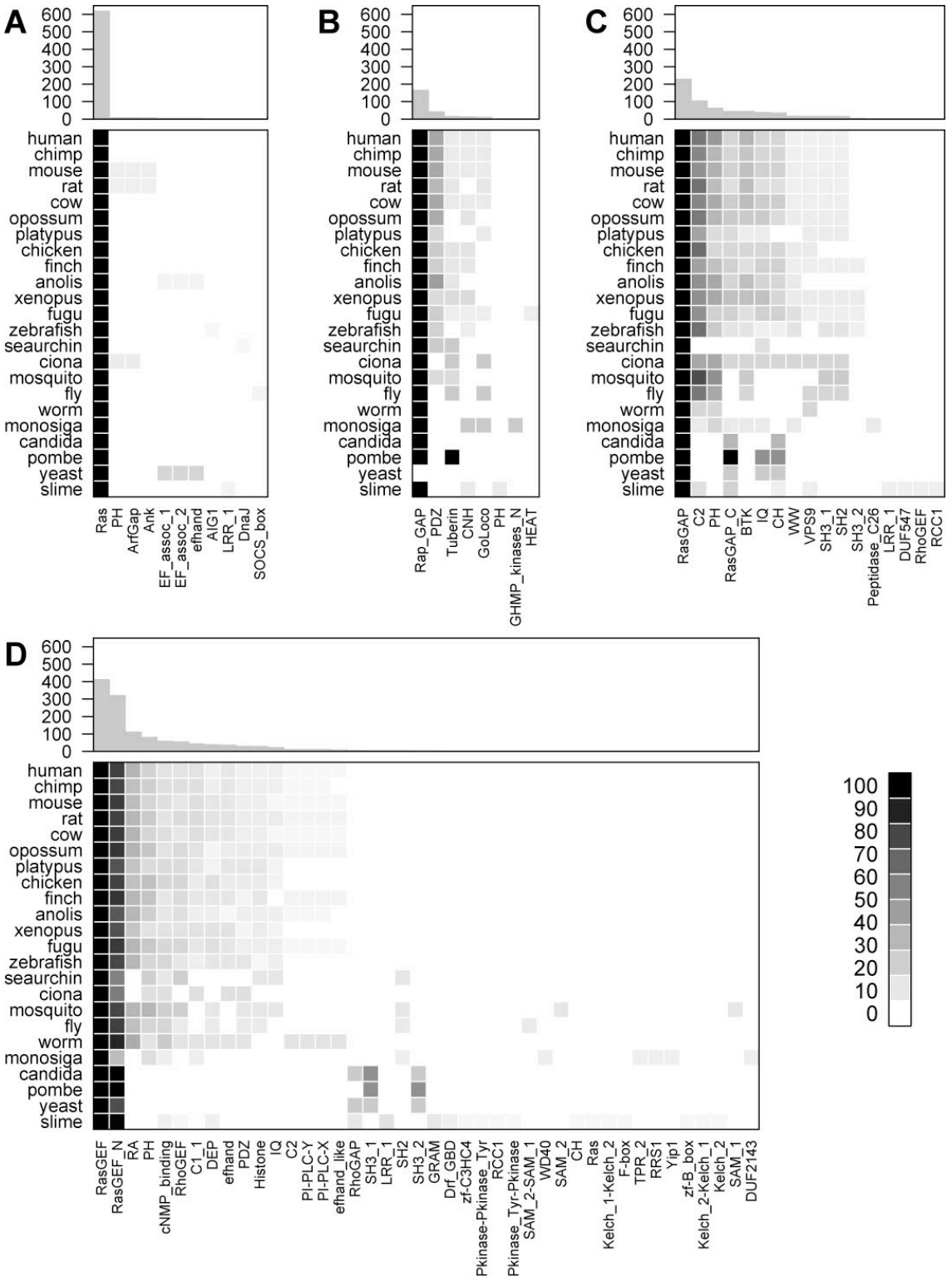
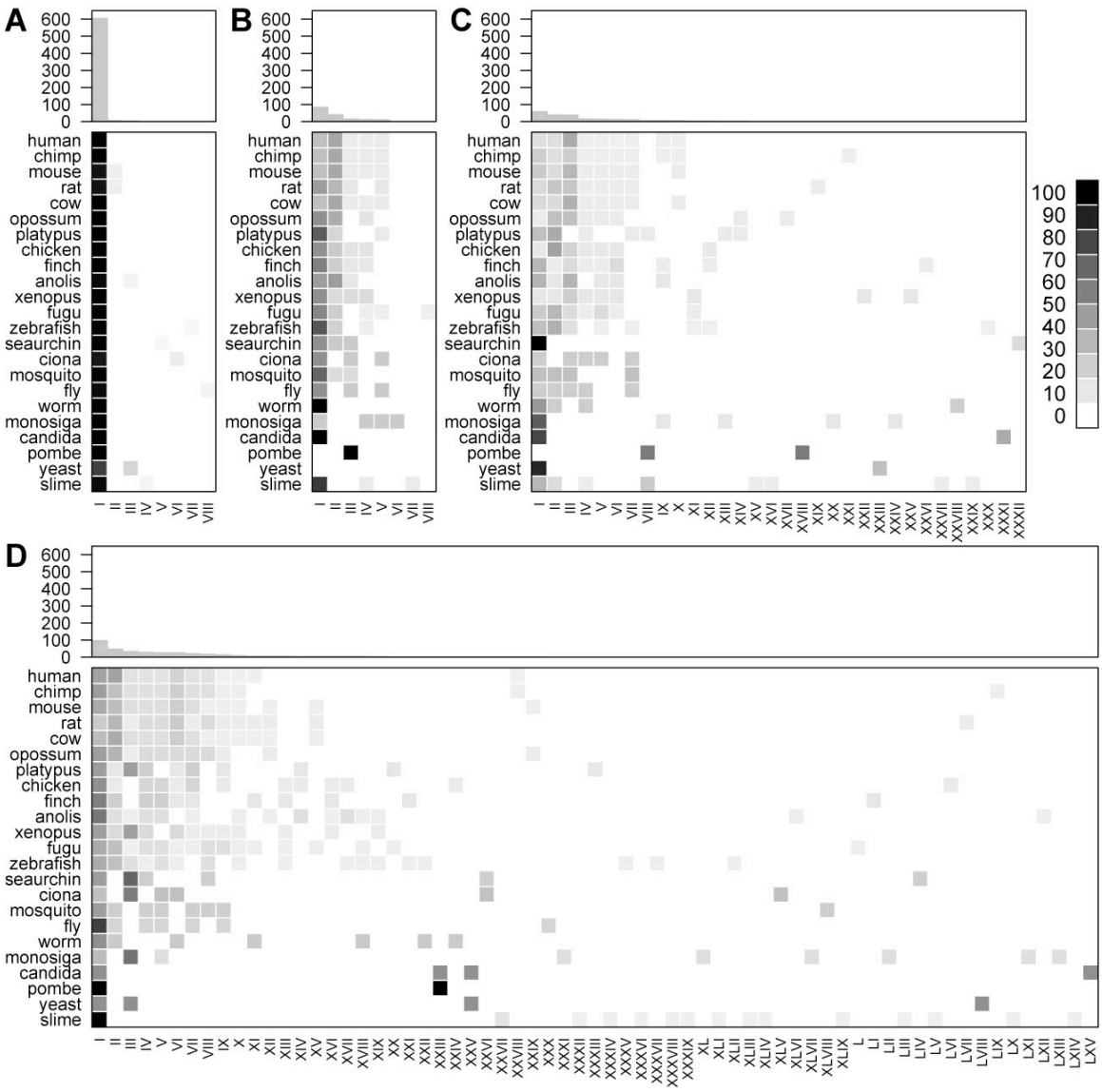


Figure S9 Distribution of architecture counts across species for A) Ras, B) RapGAP, C) RasGAP and D) RasGEF. Architectures types are named with roman figures. Architectures types are named with roman figures (translation to architecture is provided below). The color scale indicates the percentage of sequences with that architecture in the corresponding species. Architectures frequency distribution is shown for each family.



Translation from roman labels to architectures. Individual domains are separated by " | "

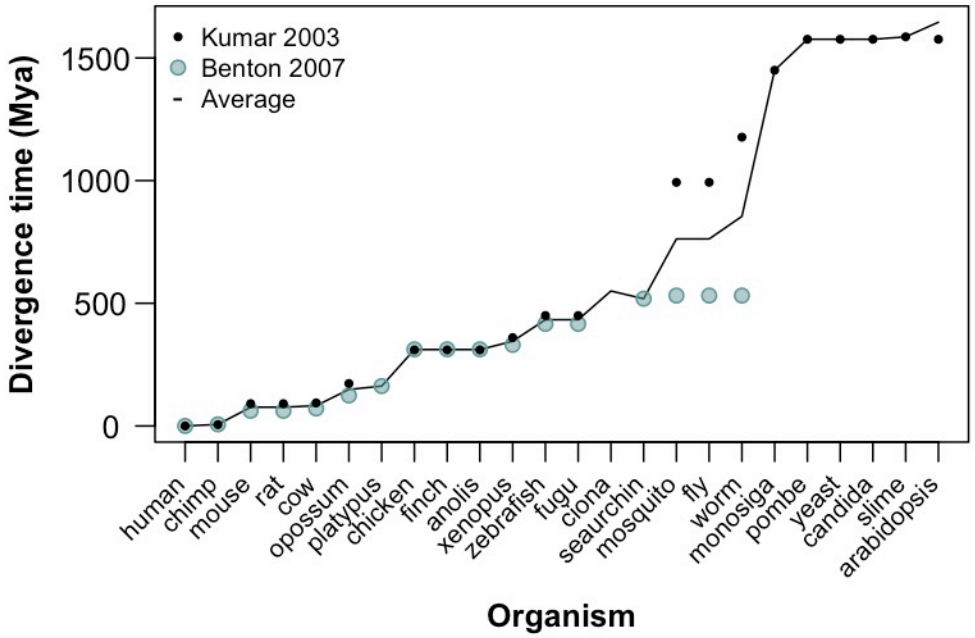
Key	*Description*
RAS	
I	Ras
II	Ras | PH | ArfGap | Ank | Ank

III	Ras | efhand | EF_assoc_2 | efhand | EF_assoc_1
IV	Ras | LRR_1 | LRR_1 | LRR_1 | LRR_1
V	Ras | DnaJ
VI	Ras | PH | ArfGap
VII	AIG1 | Ras
VIII	Ras | SOCS_box
RAPGAP	
I	Rap_GAP
II	Rap_GAP | PDZ
III	Tuberin | Rap_GAP
IV	Rap_GAP | CNH
V	GoLoco | Rap_GAP
VI	GHMP_kinases_N | Rap_GAP
VII	PH | PH | Rap_GAP
VIII	HEAT | HEAT | Tuberin | Rap_GAP
RASGAP	
I	RasGAP
II	C2 | RasGAP
III	C2 | C2 | RasGAP | PH | BTK
IV	RasGAP | VPS9
V	CH | WW | IQ | IQ | IQ | IQ | RasGAP | RasGAP_C
VI	CH | IQ | IQ | IQ | IQ | RasGAP | RasGAP_C
VII	SH2 | SH3_1 | SH2 | PH | C2 | RasGAP
VIII	RasGAP | RasGAP_C
IX	PH | C2 | RasGAP
X	CH | IQ | IQ | IQ | RasGAP | RasGAP_C
XI	SH2 | SH3_1 | SH3_2 | SH2 | PH | C2 | RasGAP
XII	C2 | C2 | RasGAP
XIII	RasGAP | PH | BTK
XIV	IQ | IQ | RasGAP | RasGAP_C
XV	CH | CH | RasGAP
XVI	RCC1 | RCC1 | RCC1 | RCC1 | RCC1 | RhoGEF | RasGAP
XVII	SH3_1 | SH2 | PH | C2 | RasGAP
XVIII	CH | IQ | IQ | IQ | IQ | IQ | IQ | IQ | IQ | IQ | IQ | RasGAP | RasGAP_C
XIX	C2 | C2 | RasGAP | BTK
XX	WW | IQ | IQ | IQ | RasGAP | RasGAP_C
XXI	C2 | RasGAP | BTK
XXII	C2 | RasGAP | PH | BTK
XXIII	CH | IQ | IQ | IQ | IQ | IQ | RasGAP | RasGAP_C
XXIV	RasGAP | Peptidase_C26
XXV	IQ | IQ | IQ | IQ | RasGAP | RasGAP_C
XXVI	SH3_1 | SH3_2 | SH2 | PH | C2 | RasGAP
XXVII	RasGAP | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1
XXVIII	RasGAP | PH
XXIX	RasGAP | DUF547
XXX	WW | RasGAP | RasGAP_C
XXXI	CH | RasGAP | RasGAP_C
XXXII	IQ | IQ | IQ | IQ | RasGAP

RASGEF	
I	RasGEF_N | RasGEF
II	RasGEF_N | RasGEF | RA
III	RasGEF
IV	Histone | RhoGEF | PH | RasGEF_N | RasGEF
V	RasGEF | PH
VI	RasGEF_N | RasGEF | efhand | efhand | C1_1
VII	cNMP_binding | RasGEF_N | PDZ | RA | RasGEF
VIII	PH | IQ | RhoGEF | PH | RasGEF_N | RasGEF
IX	cNMP_binding | DEP | cNMP_binding | RasGEF_N | RA | RasGEF
X	DEP | cNMP_binding | RasGEF_N | RasGEF
XI	RasGEF | efhand_like | PI-PLC-X | PI-PLC-Y | C2 | RA | RA
XII	RasGEF_N | RasGEF | C1_1
XIII	RasGEF_N | RasGEF | efhand | C1_1
XIV	IQ | RhoGEF | PH | RasGEF_N | RasGEF
XV	DEP | RasGEF_N | RasGEF
XVI	DEP | RasGEF_N | RA | RasGEF
XVII	RasGEF | RA
XVIII	cNMP_binding | cNMP_binding | RasGEF_N | PDZ | RA | RasGEF
XIX	RA | RasGEF
XX	RasGEF | efhand | C1_1
XXI	DEP | cNMP_binding | RasGEF_N | RA | RasGEF
XXII	cNMP_binding | DEP | cNMP_binding | RasGEF_N | RasGEF
XXIII	SH3_1 | SH3_2 | RasGEF_N | RasGEF
XXIV	RhoGEF | PH | RasGEF_N | RasGEF
XXV	RasGEF | RasGEF_N | RhoGAP
XXVI	cNMP_binding | RasGEF_N | RasGEF
XXVII	RhoGAP | RasGEF_N | RasGEF
XXVIII	RasGEF | C1_1
XXIX	RasGEF | efhand_like | PI-PLC-X | PI-PLC-Y | C2 | RA
XXX	SAM_2-SAM_1 | SH2 | RasGEF
XXXI	PH | RasGEF
XXXII	F-box | F-box | RasGEF_N | RasGEF
XXXIII	RasGEF_N | RA | RasGEF
XXXIV	LRR_1 | LRR_1 | LRR_1 | LRR_1 | Ras | Pkinase_Tyr-Pkinase | RasGEF_N | DEP | RasGEF | cNMP_binding | GRAM | cNMP_binding
XXXV	RasGEF_N | RasGEF | efhand | efhand
XXXVI	LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | RasGEF_N | RasGEF
XXXVII	cNMP_binding | RasGEF_N | PDZ | RasGEF
XXXVIII	GRAM | RhoGAP | RasGEF
XXXIX	RasGEF_N | DEP | RasGEF
XL	cNMP_binding | cNMP_binding | RasGEF
XLI	RCC1 | RCC1 | RCC1 | RhoGEF | RasGEF_N | RasGEF
XLII	RasGEF_N | PDZ | RA | RasGEF
XLIII	LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | Drf_GBD | LRR_1 | RasGEF_N | RasGEF | LRR_1
XLIV	RasGEF_N | RasGEF | cNMP_binding | GRAM | cNMP_binding

XLV	RasGEF_N | PDZ | RasGEF
XLVI	RasGEF | efhand | efhand | C1_1
XLVII	RasGEF_N | RasGEF | WD40 | WD40 | WD40 | WD40 | WD40 | WD40 | WD40
XLVIII	SAM_2 | SAM_1 | SH2 | RasGEF
XLIX	CH | RasGEF_N | RasGEF
L	PH | RhoGEF | PH | RasGEF_N | RasGEF
LI	PH | RhoGEF | RasGEF_N | RasGEF
LII	TPR_2 | TPR_2 | RasGEF | RRS1
LIII	Pkinase-Pkinase_Tyr | RasGEF_N | RasGEF
LIV	SH2 | RasGEF
LV	Kelch_1-Kelch_2 | Kelch_2-Kelch_1 | Kelch_1-Kelch_2 | Kelch_1-Kelch_2 | Kelch_2-Kelch_1 | Kelch_2 | RasGEF_N | RasGEF
LVI	DEP | cNMP_binding | RasGEF
LVII	PDZ | RA | RasGEF
LVIII	SH3_2 | SH3_1 | RasGEF_N | RasGEF
LIX	RasGEF | PI-PLC-X | PI-PLC-Y | C2 | RA
LX	LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | LRR_1 | Drf_GBD | RasGEF_N | RasGEF
LXI	RasGEF_N | RasGEF | SH2 | SH2
LXII	RasGEF | PI-PLC-X | PI-PLC-Y | C2 | RA | RA
LXIII	Yip1 | DUF2143 | RasGEF | C1_1
LXIV	zf-C3HC4 | zf-B_box | RasGEF_N | RasGEF
LXV	SH3_1 | RasGEF_N | RasGEF

Figure S10 Comparison of different divergence times (MYA) for the species included in this study. Blue points are fossil record estimates extracted from Benton 2007 (1). Black points are molecular distance estimates from Kumar 2003 (2). The average value used in this study is indicated with a black solid line. In general there is concordance between both measures, with the main difference in the insect's clade divergence times.



1. Donoghue, P.C. and Benton, M.J. (2007) Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends Ecol Evol*, **22**, 424-431.

2. Blair Hedges, S. and Kumar, S. (2003) Genomic clocks and evolutionary timescales. *Trends Genet*, **19**, 200-206.