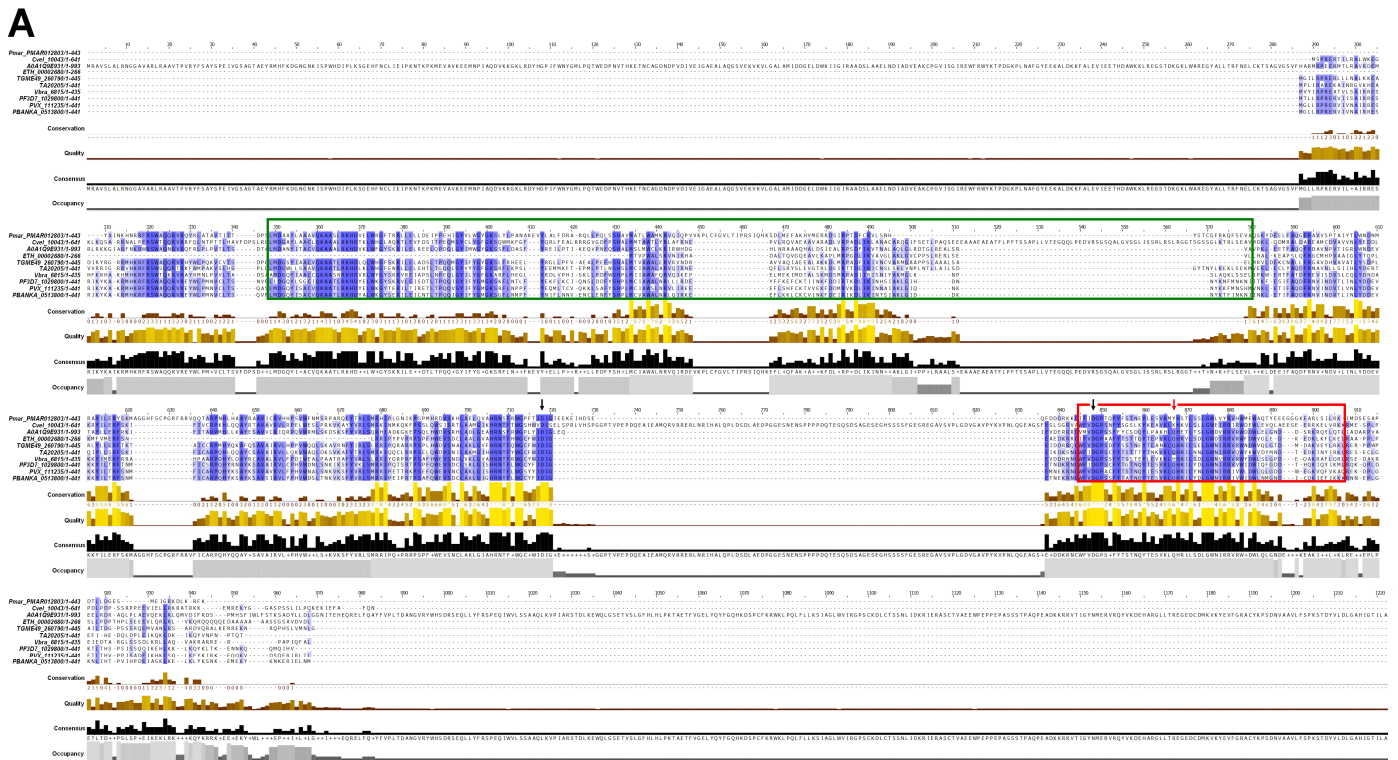


Figure S1. Relative position of the RAP domain. Tukey's box and whiskers show the positioning of the N-terminal and C-terminal ends of the RAP domains on their respective proteins. The relative values indicate N-terminal (0) and the C-terminal (1) of the full-length proteins. Proteins below 200 amino acids have been excluded (7/267).



B

	<i>P. vivax</i> (PVX_111235)	<i>P. berghei</i> (PBANKA_0513800)	<i>T. annulata</i> (TA20205)	<i>T. gondii</i> (TGME49_260790)	<i>E. tenella</i> (ETH_00002680)	<i>C. velia</i> (Cvel_10043)	<i>V. brassicaformis</i> (Vbra_6815)	<i>P. marinus</i> (Pmar_PMAR012803)	<i>S. microadriaticum</i> (ADA1Q9E931)
<i>P. falciparum</i> (PF3D7_1029800)	73%	76%	49%	48%	26%	27%	47%	32%	19%

Figure S2. Global alignment of PF3D7_1029800 and its homologs. A. Global alignment between PF3D7_1029800 and its homologs from *P. vivax*, *P. berghei*, *T. annulata*, *T. gondii*, *E. tenella*, *C. velia*, *V. brassicaformis*, *P. marinus*, and *S. microadriaticum*. The positions of the peptide repeats (green box) and the RAP domain (red box) are indicated according to the respective predictions of [21] and hmmscan (Table S1). The black arrows point to the two residues of the PD-(D/E)XK endonuclease superfamily while the red arrow indicates the predicted position of the residue K, not conserved for these RAP proteins. B. Percentage of identity between PF3D7_1029800 and its homologs in the different species.

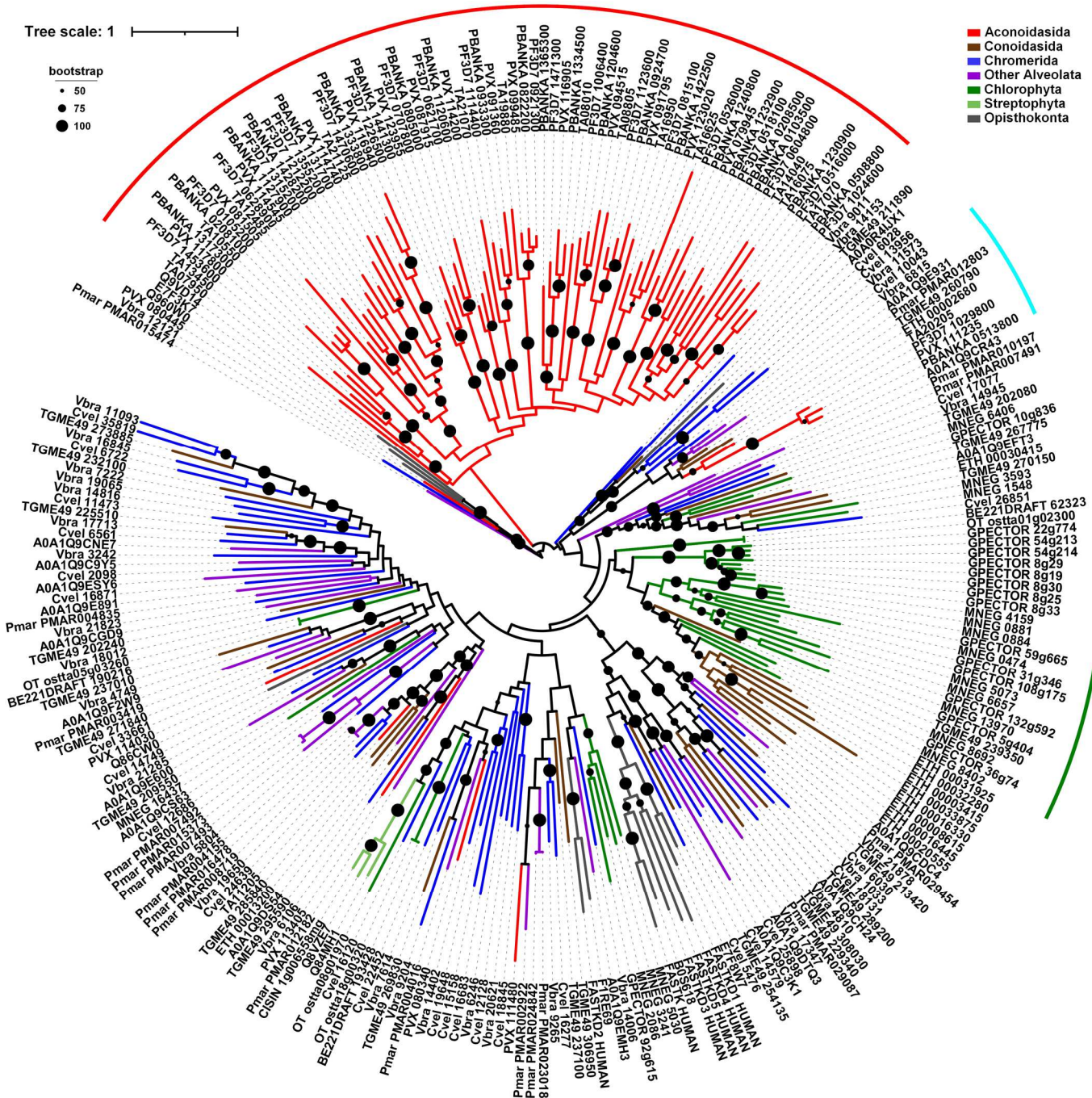


Figure S3. Phylogenetic analysis of full-length RAP proteins in Eukaryota. The maximum likelihood tree is built from 267 protein sequences corresponding to 19 different species. Alveolates are represented by Aconoidasida (*P. falciparum*, *P. vivax*, *P. berghei*, and *T. annulata*) in red, Conoidasida (*T. gondii* and *E. tenella*) in brown, Chromerida (*V. brassicaformis* and *C. velia*) in blue and *P. marinus* and *S. microadriaticum* in purple. Viridiplantae are represented by Chlorophyta (*G. pectorale*, *M. neglectum*, and *O. tauri*) in green and Streptophyta (*A. thaliana*, *O. sativa*, and *C. sinensis*) in pale green. Opisthokonta (*H. sapiens*, *D. rerio*, and *D. melanogaster*) are indicated in grey. Bootstrap values (> 50%) are shown on respective branches.

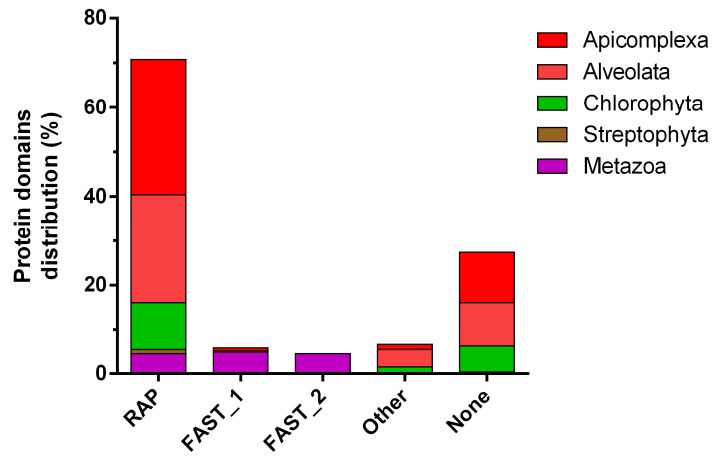
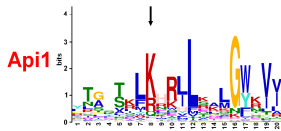
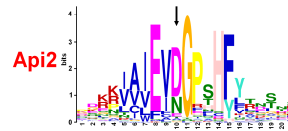


Figure S4. Domains identified in RAP proteins by SMART. The complete set of full-length proteins selected in this study (267 sequences) were analyzed by SMART. The results obtained for each protein are available in Table S1.

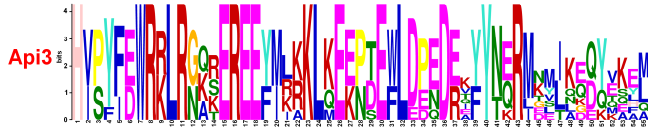
Apicomplexa-Chromerida



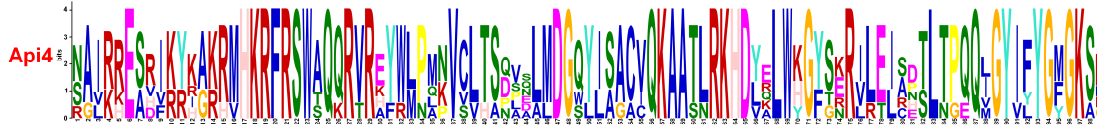
E-value: 1.5e-465
Number of sites: 169
Width: 20



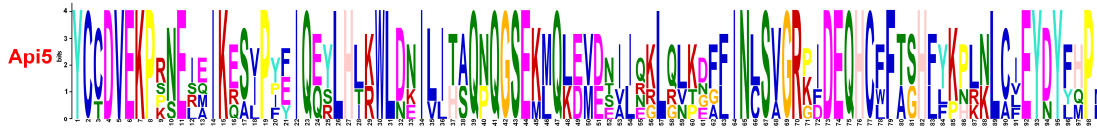
E-value: 1.2e-332
Number of sites: 92
Width: 21



E-value: 3.1e-131
Number of sites: 8
Width: 55



E-value: 2.1e-124
Number of sites: 6
Width: 99

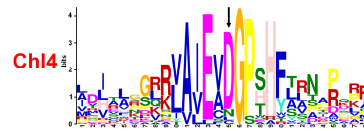


E-value: 1.2e-110
Number of sites: 5
Width: 100

Chlorophyta



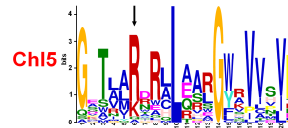
E-value: 6.0e-261
Number of sites: 9
Width: 57



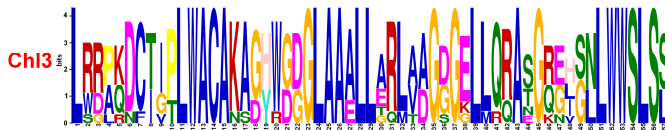
E-value: 1.0e-160
Number of sites: 25
Width: 28



E-value: 5.1e-189
Number of sites: 8
Width: 55

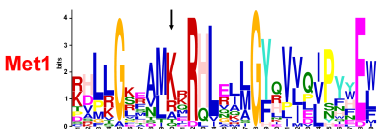


E-value: 6.6e-150
Number of sites: 29
Width: 22

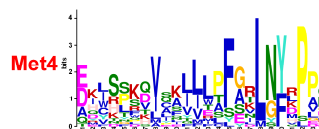


E-value: 5.1e-166
Number of sites: 8
Width: 57

Metazoa



E-value: 1.5e-174
Number of sites: 20
Width: 29



E-value: 5.4e-060
Number of sites: 19
Width: 24



E-value: 1.2e-099
Number of sites: 15
Width: 32



E-value: 3.1e-053
Number of sites: 20
Width: 29



E-value: 4.7e-071
Number of sites: 16
Width: 24

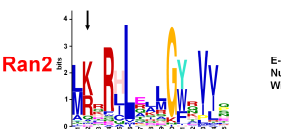
Randomised group



E-value: 6.8e-137
Number of sites: 31
Width: 15



E-value: 2.0e-075
Number of sites: 3
Width: 33



E-value: 1.7e-145
Number of sites: 50
Width: 15



E-value: 4.9e-078
Number of sites: 14
Width: 29



E-value: 5.3e-127
Number of sites: 51
Width: 21

Figure S5. Motifs identified in RAP proteins by MEME Suite. The sequences of RAP proteins from three groups, Apicomplexa-Chromerida, Chlorophyta and Metazoa, were analyzed by MEME Suite. An additional group was made and regrouped 20 random sequences from each previous group. The e-values, number of sites detected and the sequence width are indicated. The arrows point to the two residues of the PD-(D/E)XK endonuclease superfamily.