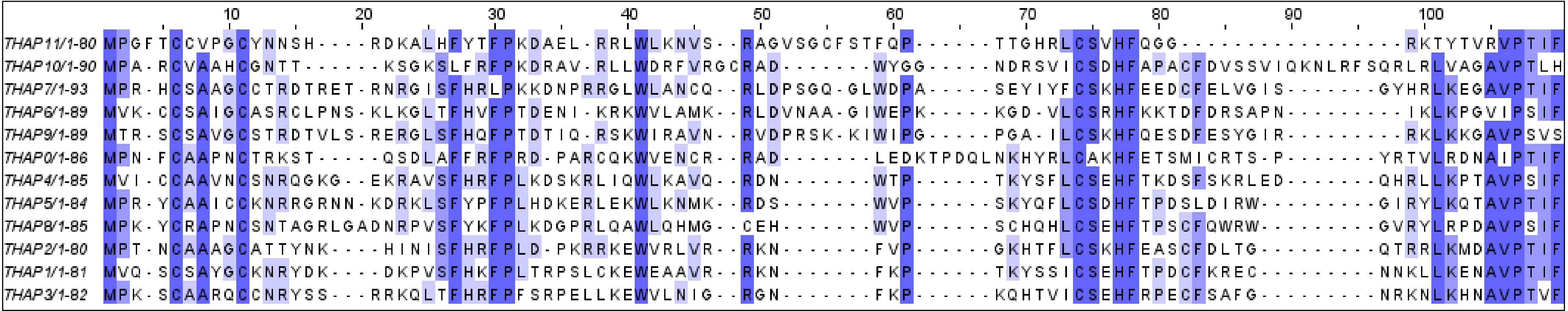
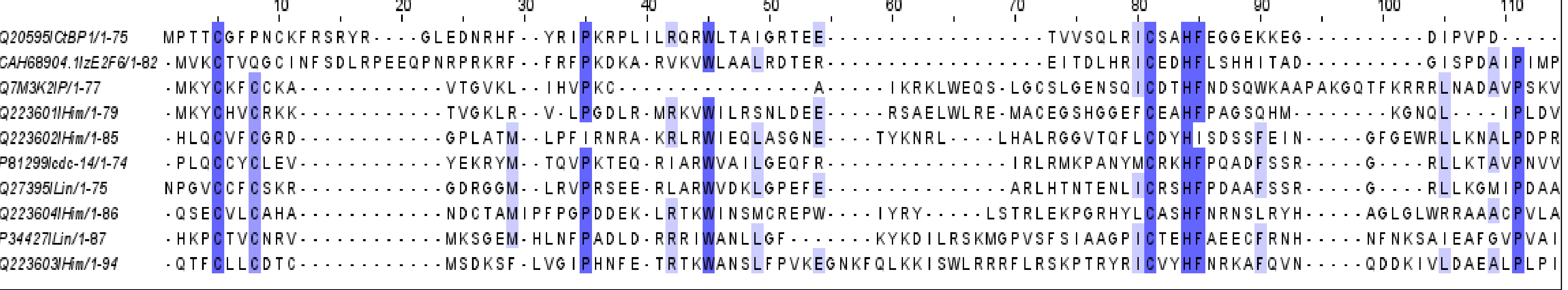


Figure 1

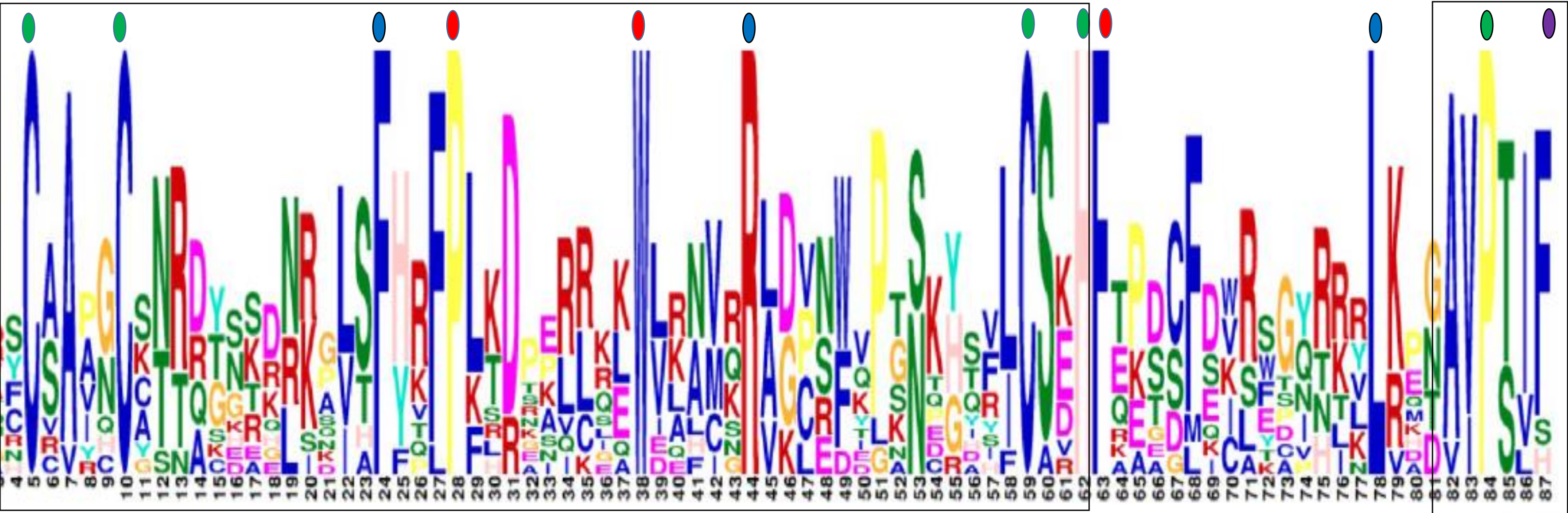
(A)



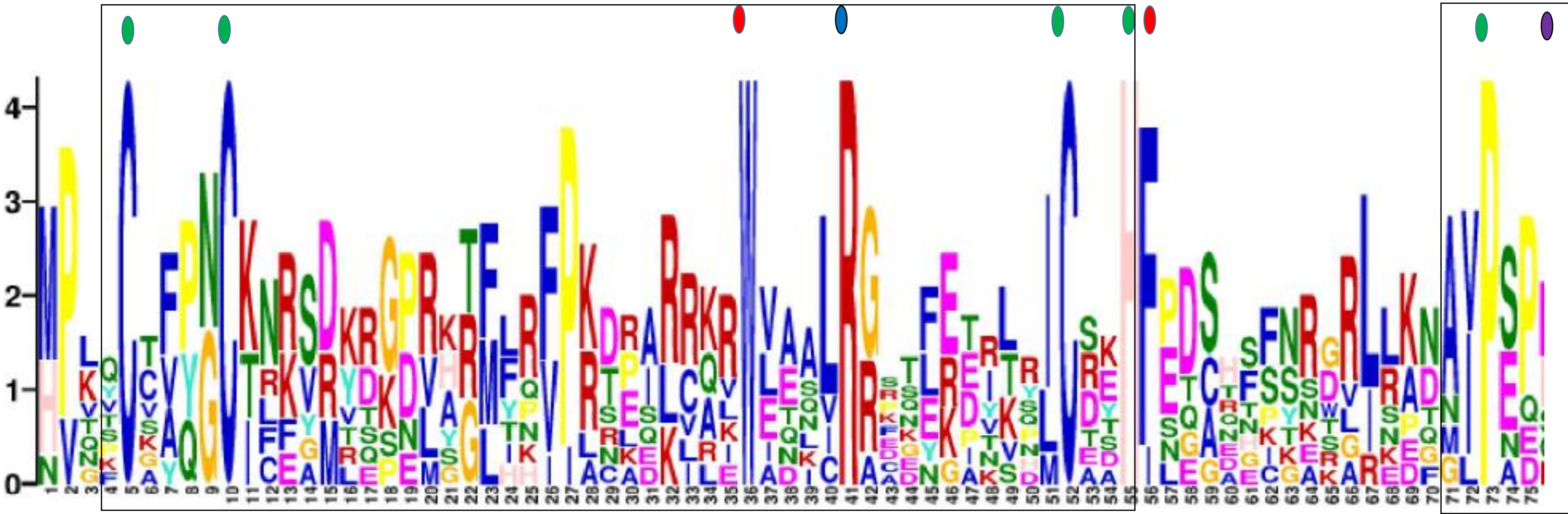
(B)



(C)



(D)



- Experimentally validated functionally important conserved residues
- Experimentally validated functionally important but not conserved residues
- Predicted conserved residues in addition to important residues suggested earlier
- Earlier predicted important conserved residues, not yet functionally studied

Figure 2

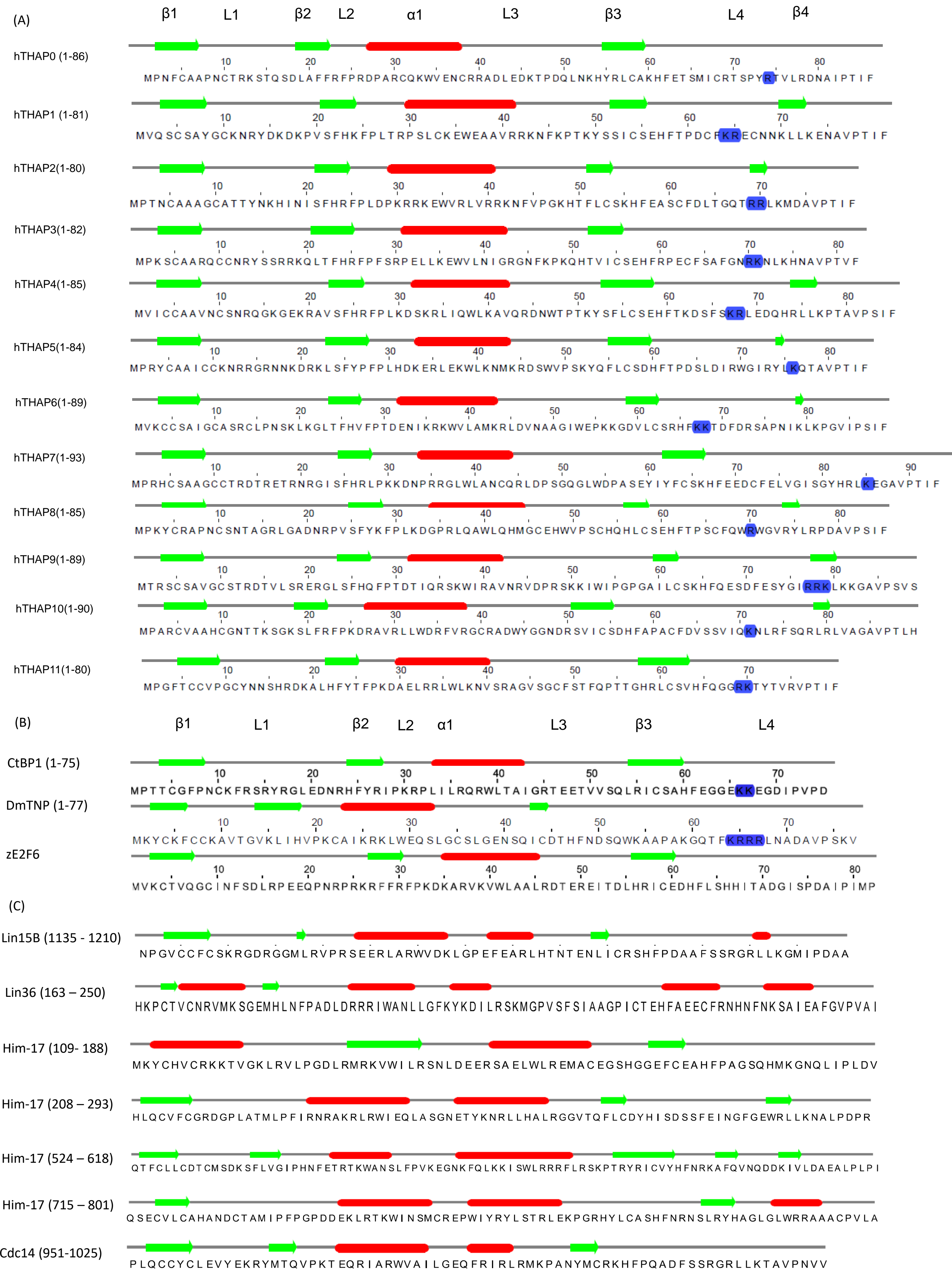


Figure 3

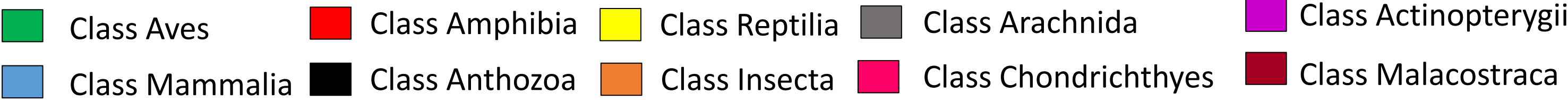
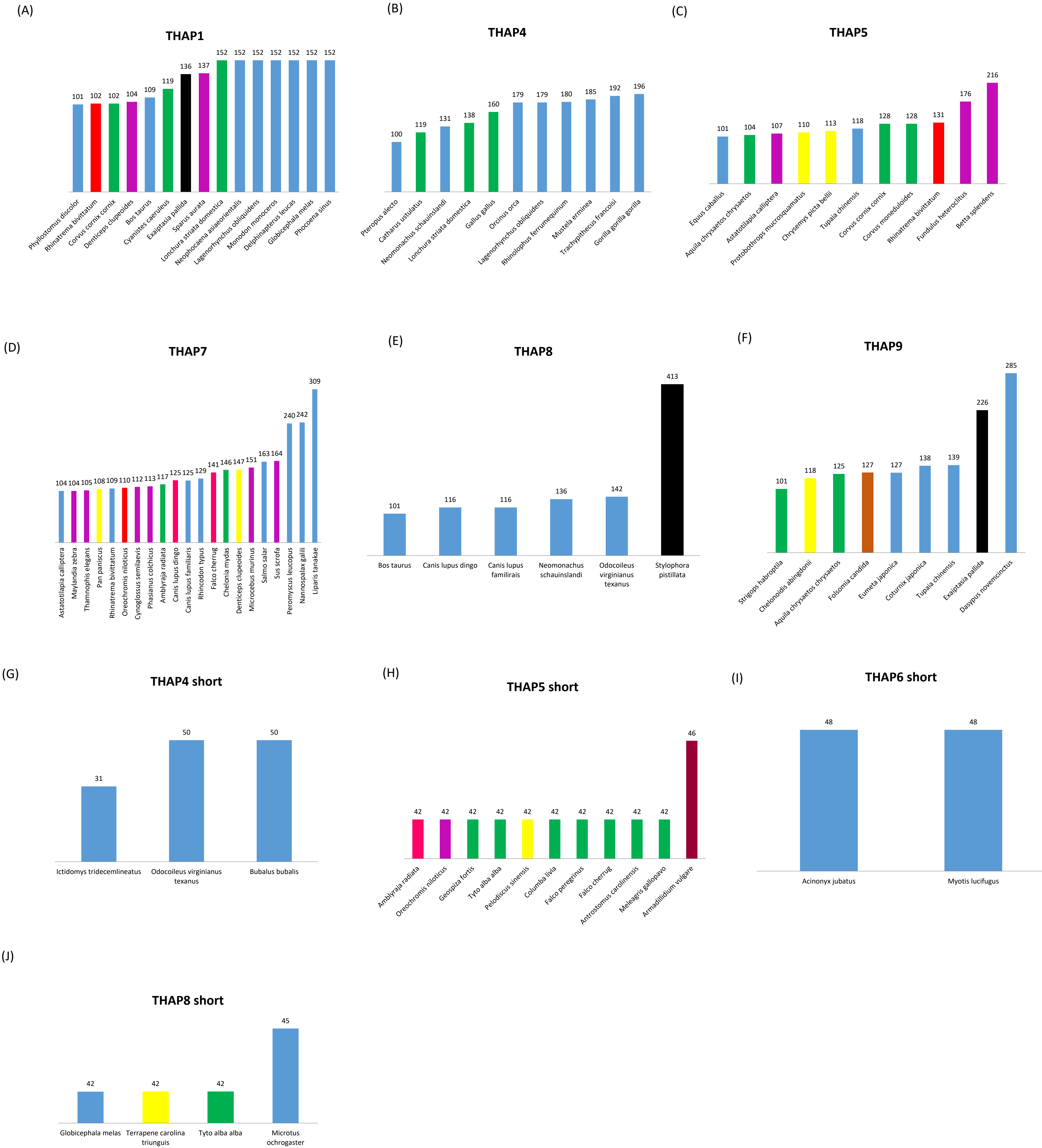


Figure 4

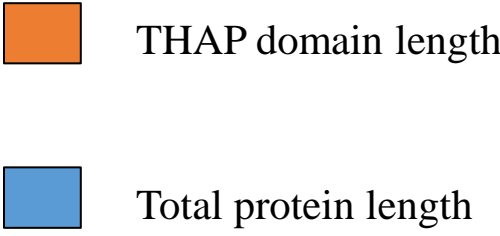
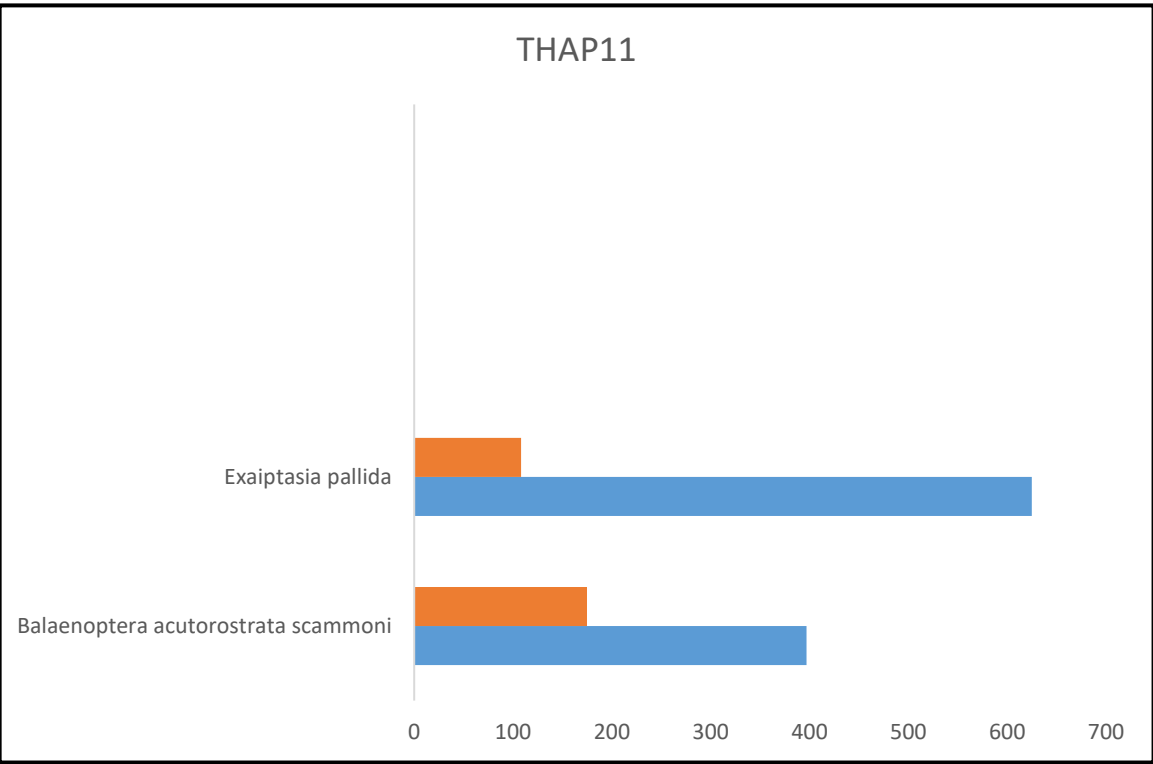
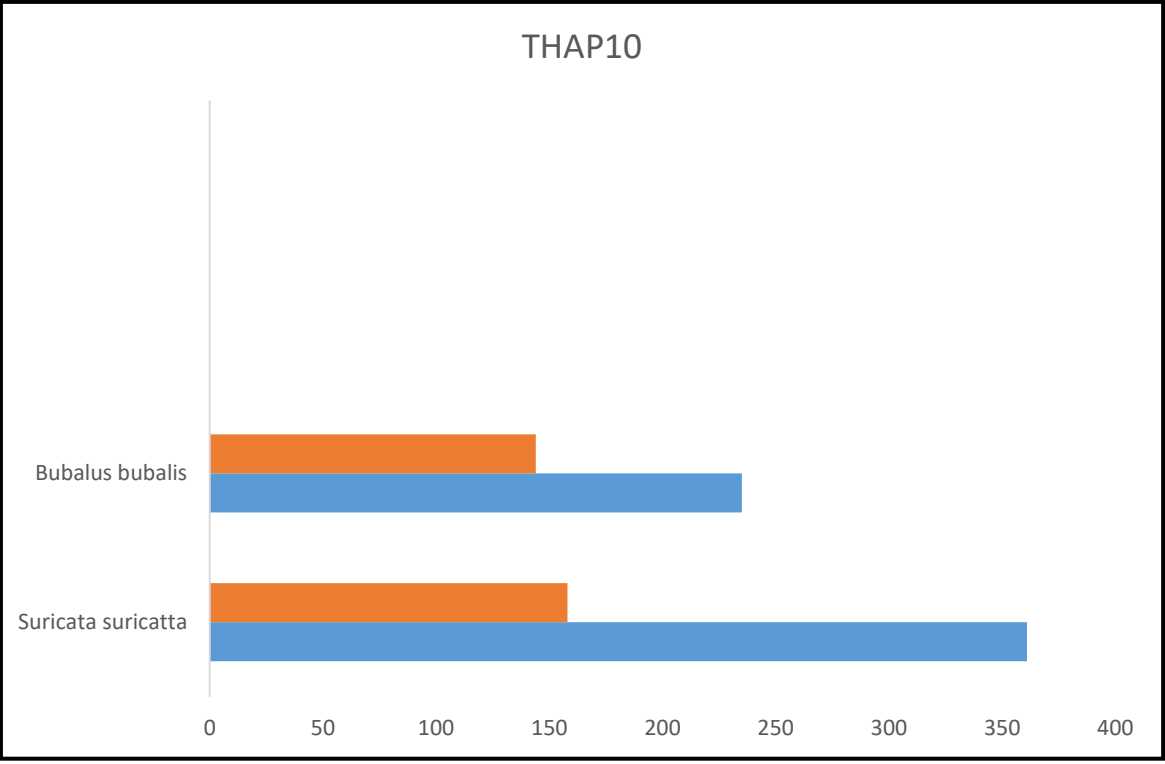
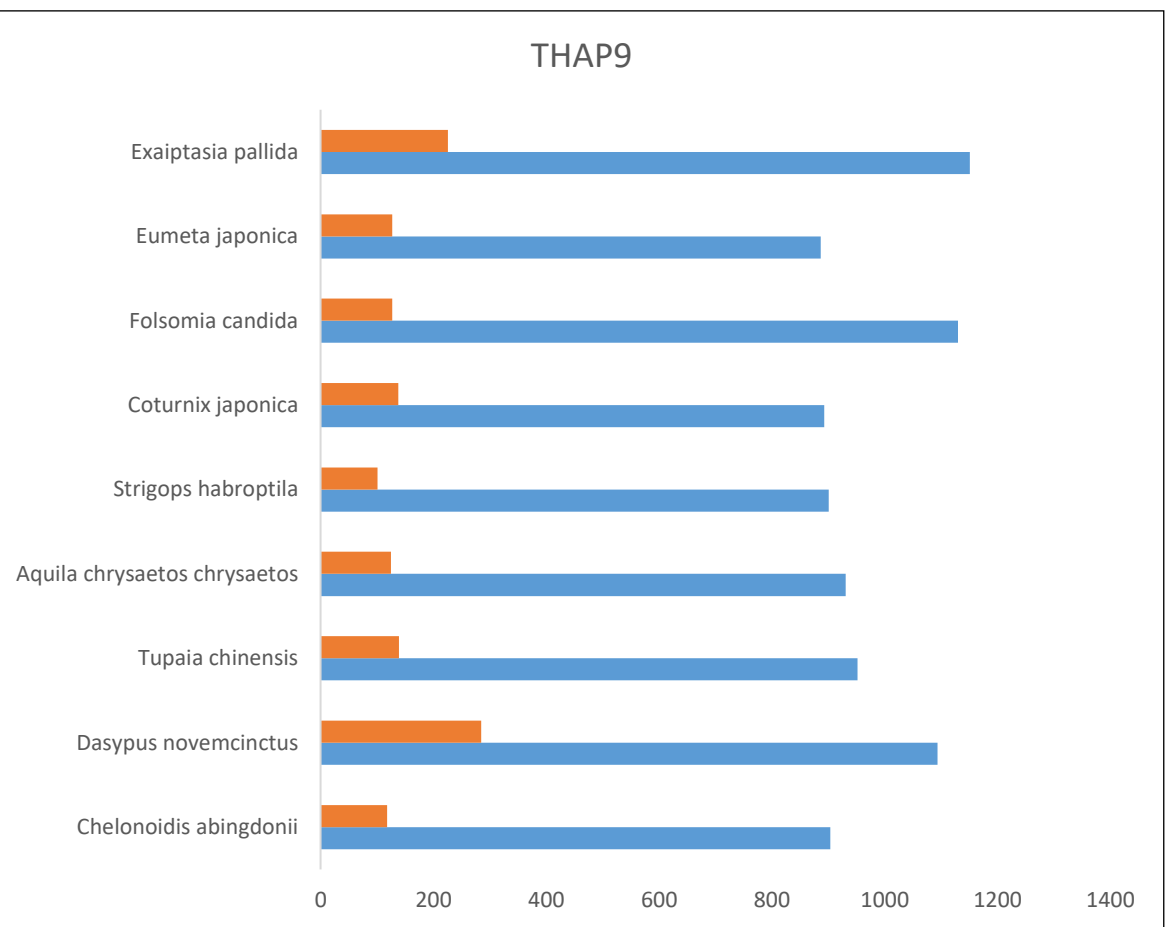
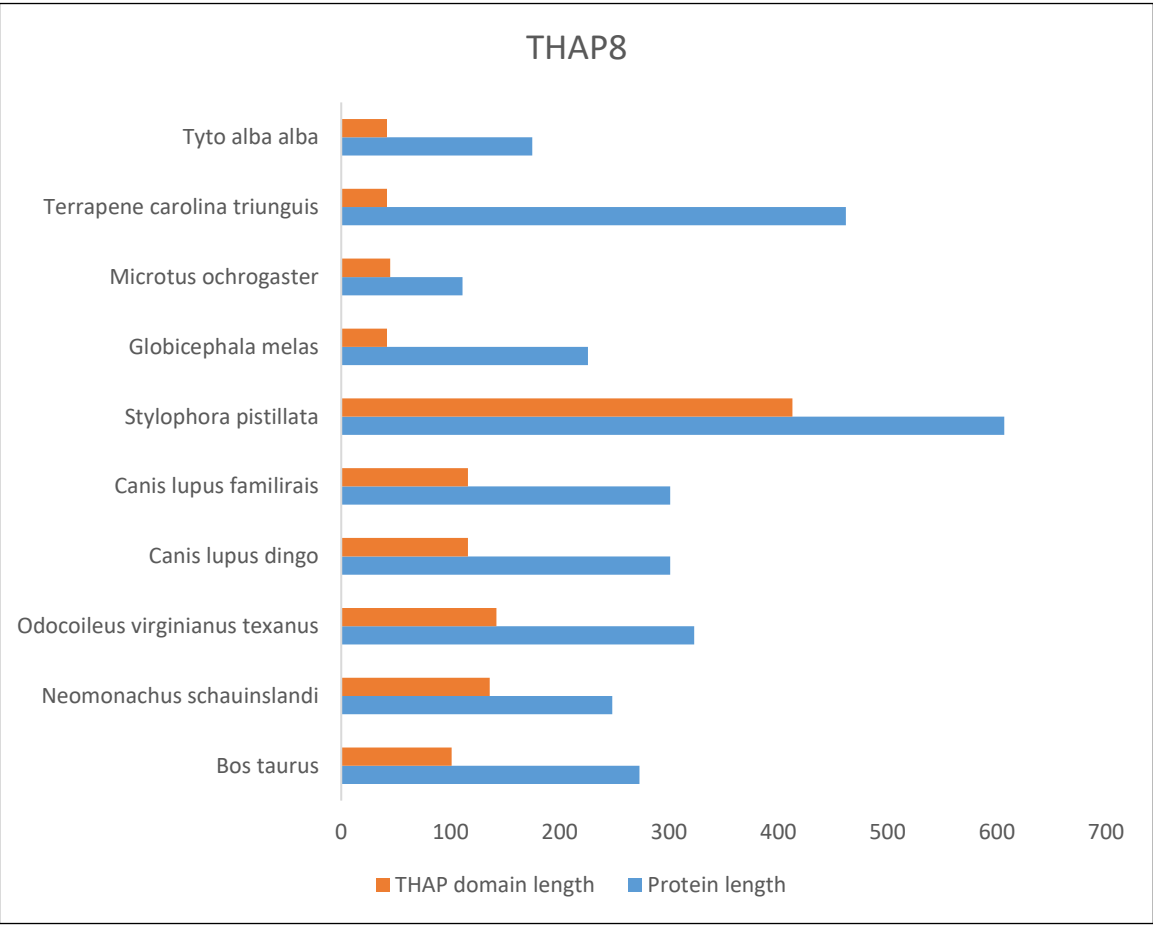
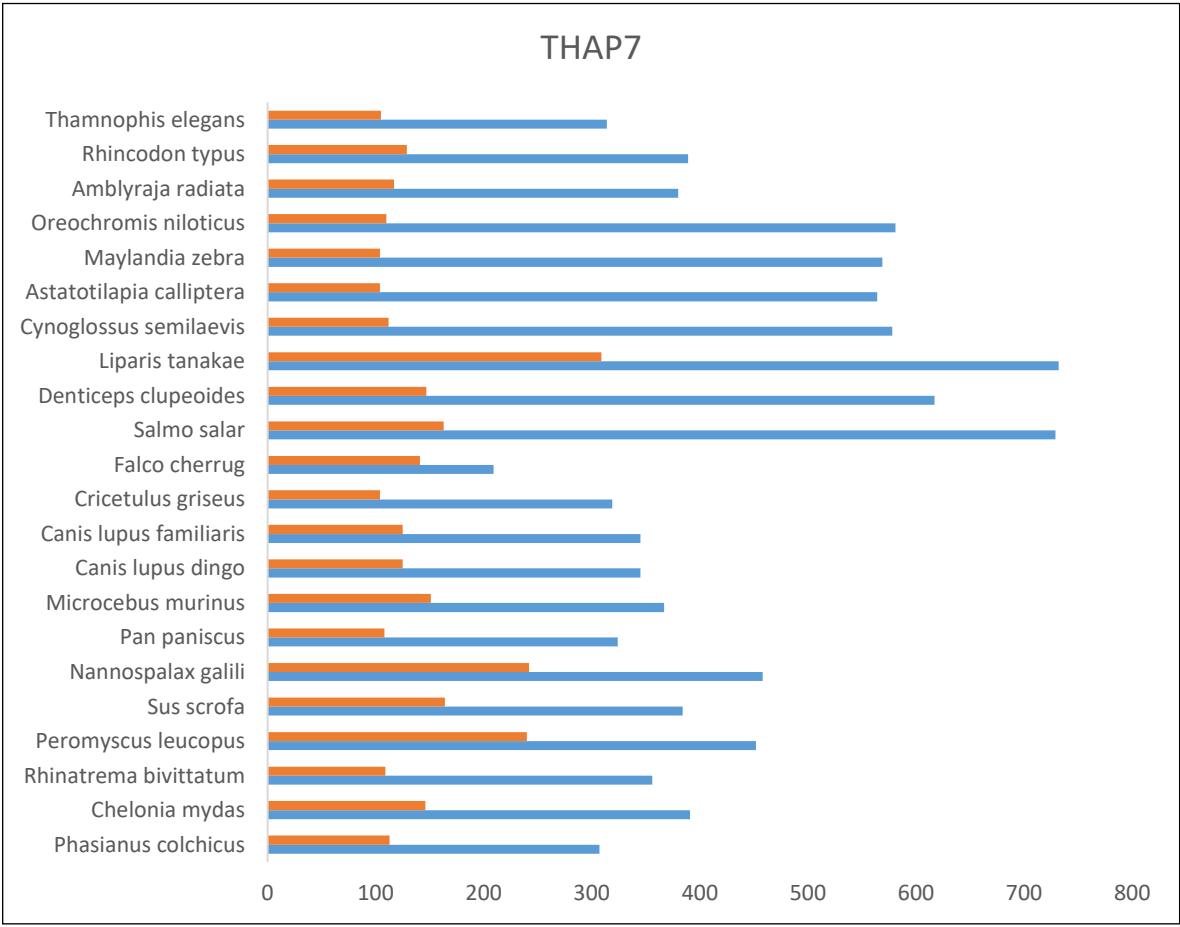
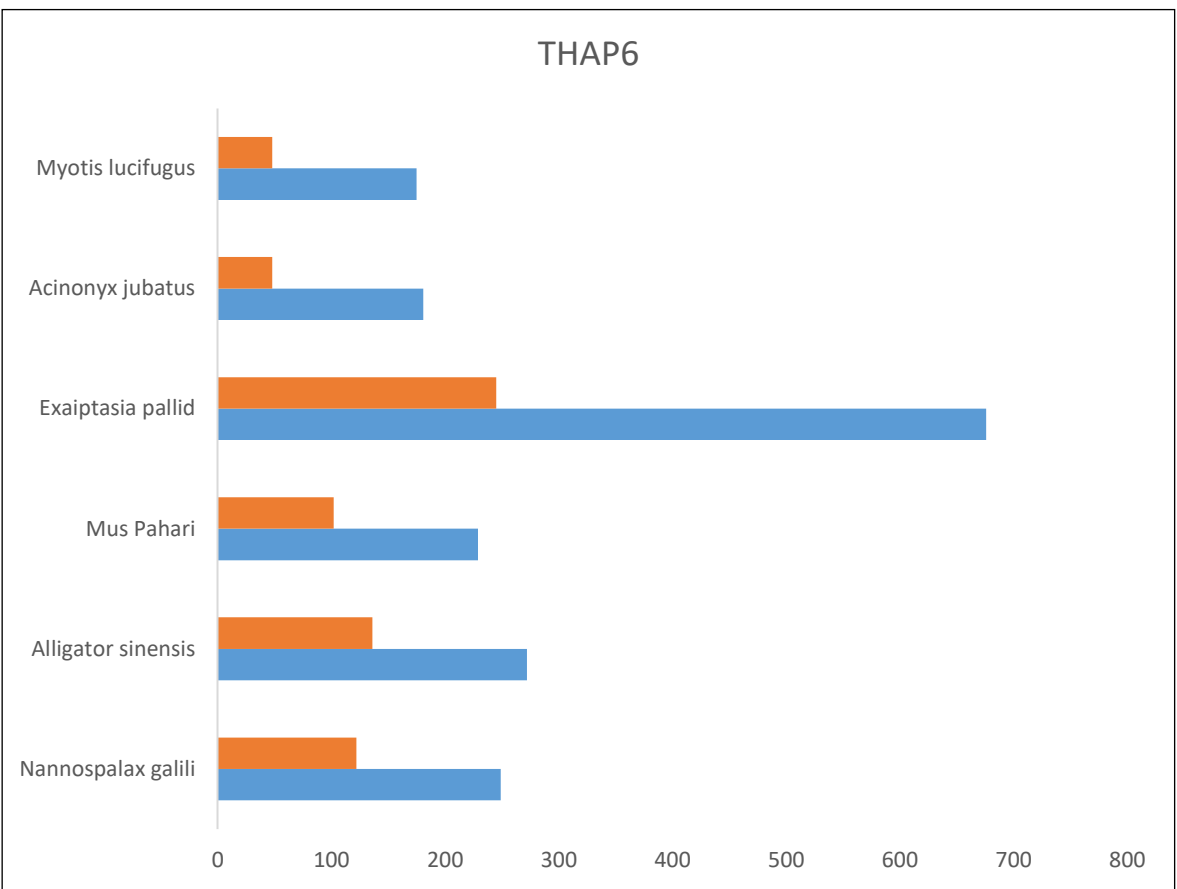
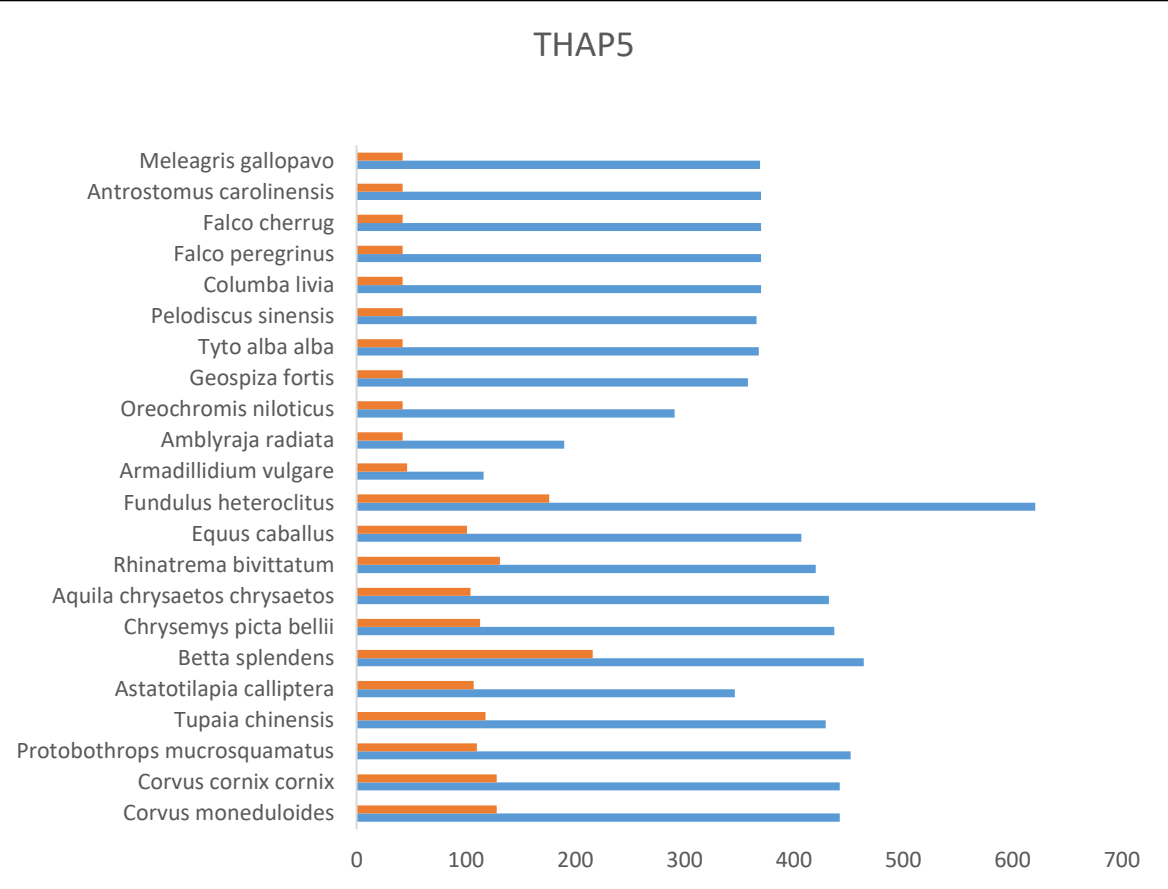
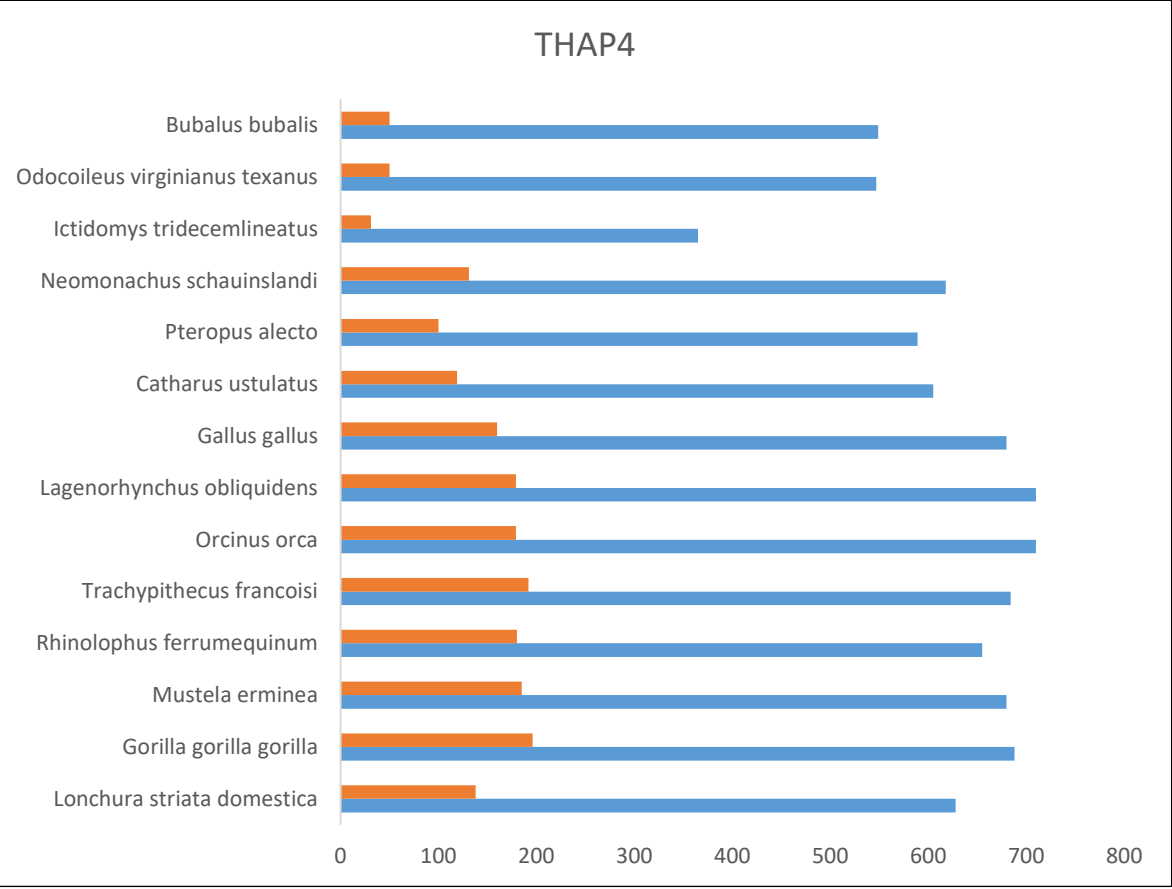
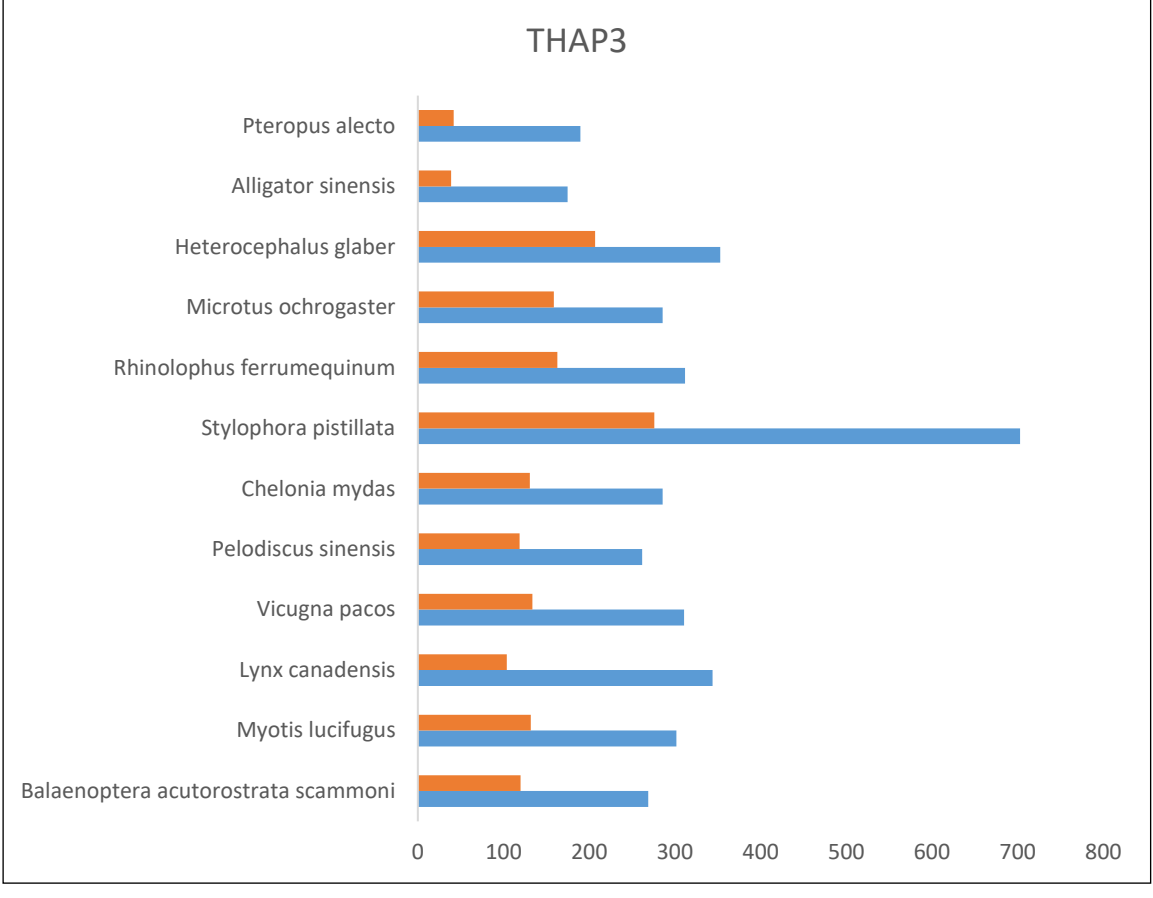
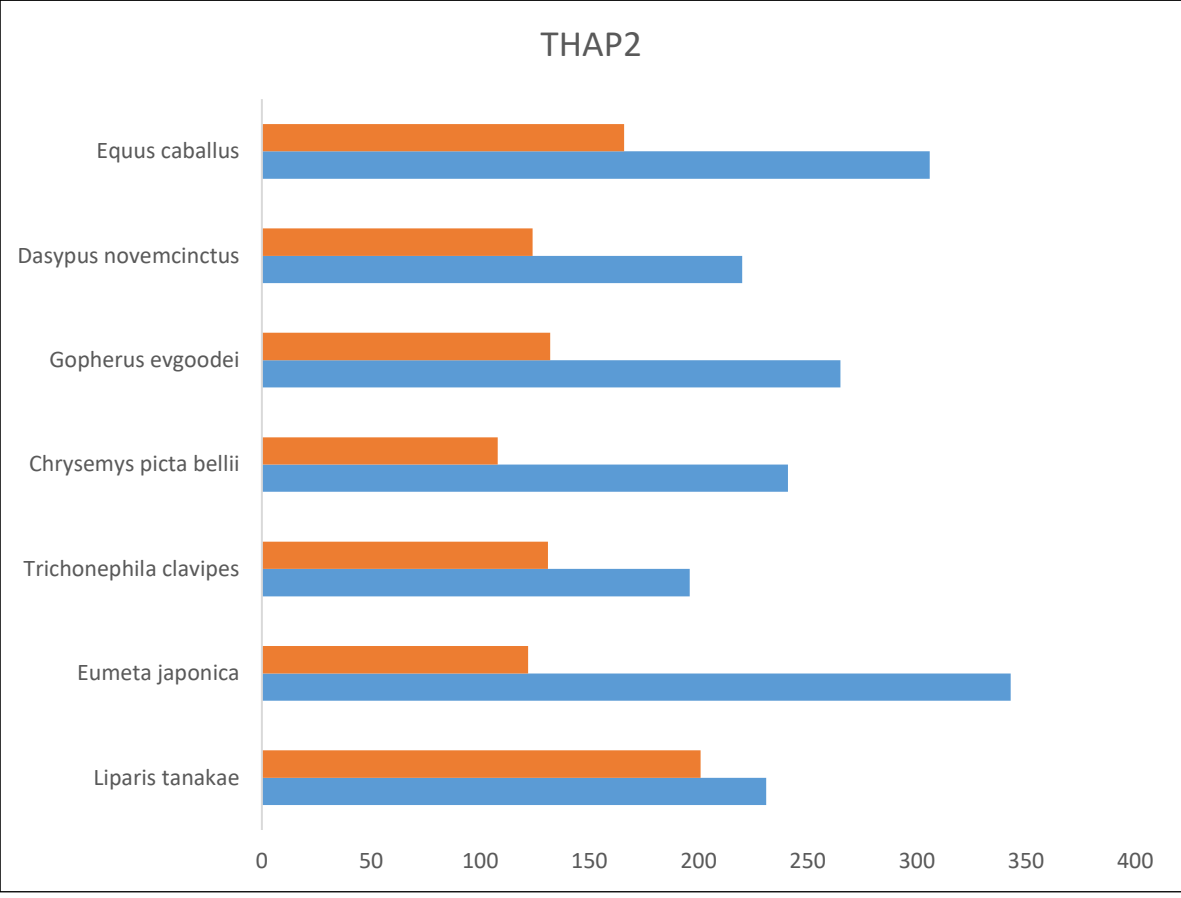
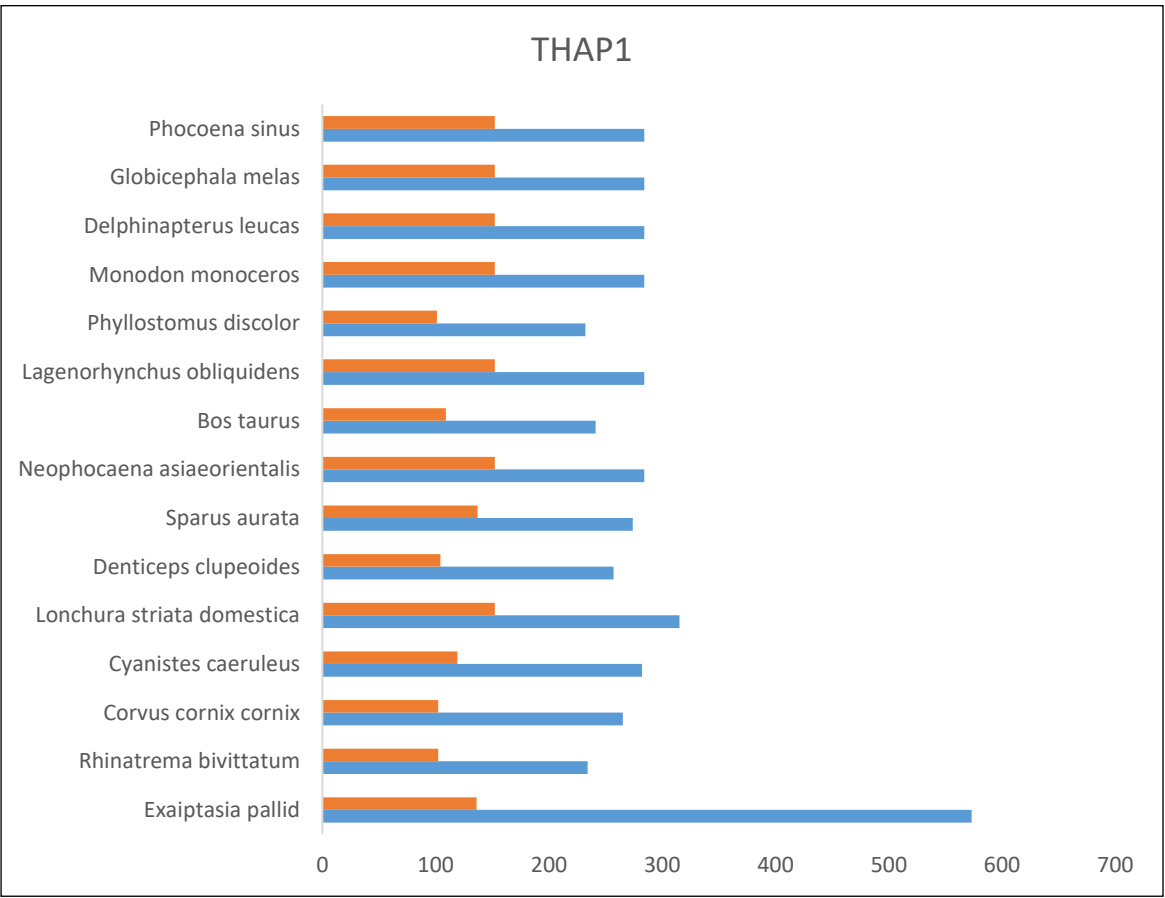


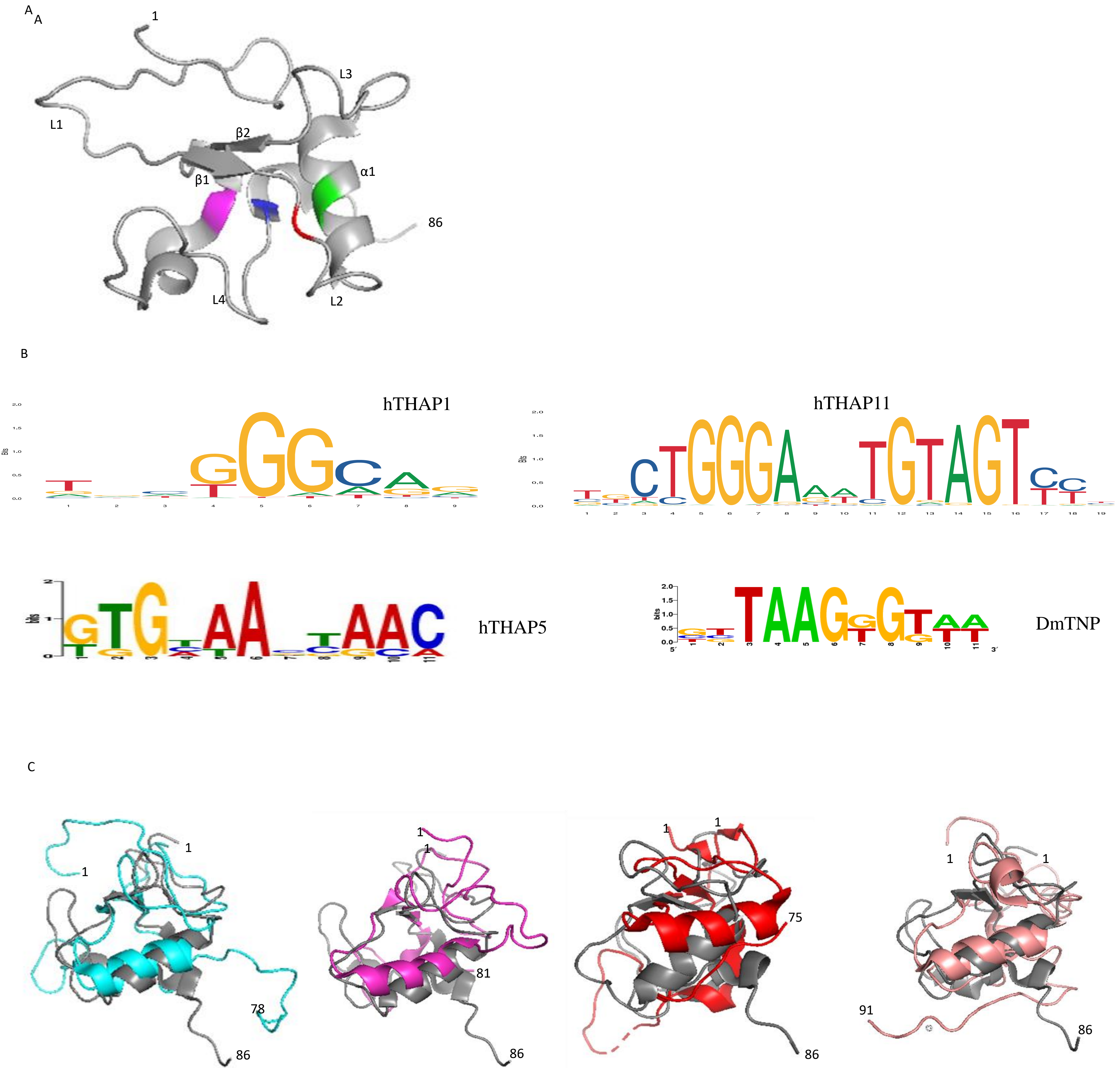
Figure 5

(A) Predicted gapped THAP motif for longer THAP domains

(B) Predicted gapped THAP motif for ~90 residues long THAP domains



Supplementary figure 1



(B)

(C)

THAP5

025755623.1	Oreochromis	MKREQWTPSRHQYLCSEHFTADCFDIRWGIRYLKNTAIP	4
006122567.1	Pelodiscus	MKRDTWTPSKHQLLCSDHFTPDSDLVVRWGIRYLKHTAIP	4
005426394.1	Geospiza	MKRDSWTPSKHQLLCSDHFTPDSDLVVRWGIRYLKNTAVPT	IF
032855683.1	Tyto	MKRDAWMPKSHQLLCSDHFTPDSDLVVRWGIRYLKHTAVPT	IF
005508832.1	Columba	MKRDAWTPSKHQLLCSDHFTPDSDLVVRWGIRYLKHTAVPT	IF
005242159.1	Falco	MKRDAWTPSKHQLLCSDHFTPDSDLVVRWGIRYLKHTAVPT	IF
005444088.1	Falco	MKRDAWTPSKHQLLCSDHFTPDSDLVVRWGIRYLKHTAVPT	IF
028943284.1	Antrostomus	MKRDAWTPSKHQLLCSDHFTPDSDLVVRWGIRYLKHTAVPT	IF
010705265.1	Meleagris	MKRDAWTPSKHQLLCSDHFTPDSDLVVRWGIRYLKHTAVPT	IF
		***. * ** ** ** **	

Supplementary figure 3

021403661.1 Lonchura	MQSSKVSPQNGNMLKLLKINLFVTTNGFVETSIYCLLMHLHTTFSRRDGTLGA---ACCR	57
024620782.1 Neophocaena	MKS-----CVSRTAAAWAVAG-YRAWGR-ALSEGTRPTALGAASSGRCR	42
026978759.1 Lagenorhynchus	MKS-----CVSRTAAAWAVAG-YRAWGR-ALSEGTRPTALGAASSGRCR	42
022427054.1 Delphinapterus	MKS-----CVSRTAAAWAVAG-YRAWGR-ALSEGTRPTALGAASSGRCR	42
030690375.1 Globicephala	MKS-----CVSRTAAAWAVAG-YRAWGR-ALSEGTRPTALGAASSGRCR	42
032474305.1 Phocoena	MKS-----CVSRTAAAWAVAG-YRAWGR-ALSEGTRPTALGAASSGRCR	42
029082374.1 Monodon	MKS-----CVSRTAAAWAVAG-CRAWGR-ALSEGTRPTALGAASSGRCR	42
	: : * : . : . : : : . * :*** . **	
021403661.1 Lonchura	ATASNE--TPFPPHPDEGQE-----DFWLRL----SSFPEKLNK-VPLACRFPLTRPD	102
024620782.1 Neophocaena	GRSPRRDGLLLRRKPQNGMIVIVRCLHGAIIFIQPNVFTSFTTLRKCGRNLRLFFPLTRPS	102
026978759.1 Lagenorhynchus	GRSPRRDGLLLRRKPQNGMIVIVRCLHGAIIFIQPNVFTSFTTLRKCGRNLRLFFPLTRPS	102
022427054.1 Delphinapterus	GRSPRRDGLLLRRKPQNGMIVIVRCLHGAIIFIQPNVFTSFTTLRKCGRNLRLFFPLTRPS	102
030690375.1 Globicephala	GRSPRRDGLLLRRKPQNGMIVIVRCLHGAIIFIQPNVFTSFTTLRKCGRNLRLFFPLTRPS	102
032474305.1 Phocoena	GRSPRRDGLLLRRKPQNGMIVIVRCLHGAIIFIQPNVFTSFTTLRKCGRNLRLFFPLTRPS	102
029082374.1 Monodon	GRSPRRDGLLLRRKPQNGMIVIVRCLHGAIIFINAVFTSFTTLRKCGRNLRLFFPLTRPS	102
	. : .. : :*::* :*: :** *	*****.
021403661.1 Lonchura	LCKKWEAAVKKRKNFKPTKYSSICSEHFTPDCKFKRECNNKLLKENAVPTIF	152
024620782.1 Neophocaena	LCKKWEAAVRRKNFKPTKYSSICSEHFTPDCKFKRECNNKLLKEDAVPTIF	152
026978759.1 Lagenorhynchus	LCKKWEAAVRRKNFKPTKYSSICSEHFTPDCKFKRECNNKLLKEDAVPTIF	152
022427054.1 Delphinapterus	LCKKWEAAVRRKNFKPTKYSSICSEHFTPDCKFKRECNNKLLKEDAVPTIF	152
030690375.1 Globicephala	LCKKWEAAVRRKNFKPTKYSSICSEHFTPDCKFKRECNNKLLKEDAVPTIF	152
032474305.1 Phocoena	LCKKWEAAVRRKNFKPTKYSSICSEHFTPDCKFKRECNNKLLKEDAVPTIF	152
029082374.1 Monodon	LCKKWEAAVRRKNFKPTKYSSICSEHFTPDCKFKRECNNKLLKEDAVPTIF	152
	*****.******.******.******.******	

Supplementary figure 4

THAP1

