

GNATHOSTOMES

DBD

COE1															
Mm	COE1	-----	-M-FGIQES-	-----	I-QRS	G-SSMKEE	P L-GS-	GMN	--AVR	TWMQ	GAGVLDANTA	AQSGVGLARA	HFEKQPPSNL	RKSNFFHFVL	
Hs	COE1	-----	-M-FGIQES-	-----	I-QRS	G-SSMKEE	P L-GS-	GMN	--AVR	TWMQ	GAGVLDANTA	AQSGVGLARA	HFEKQPPSNL	RKSNFFHFVL	
Md	COE1	-----	-M-FGIQES-	-----	I-QRS	G-SSMKEE	P L-GS-	GMN	--AVR	TWMQ	GAGVLDANTA	AQSGVGLARA	HFEKQPPSNL	RKSNFFHFVL	
Xt	COE1	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	
Dr	COE1	-----	-M-FGIQDS-	-----	I-QRS	G-SSMKEE	P I-GA-	GMN	--AVR	TWMQ	GAGVLDANTA	AQSGVGLARA	HFEKQPPSNL	RKSNFFHFVL	
COE2															
Mm	COE2	-----	-M-FGIQDT-	-----	L-GR-	G-PALKDK-	S L-GA-	EMD	--SVR	SWVR	NVGVVDANVA	AQSGVALSRA	HFEKQPPSNL	RKSNFFHFVL	
Hs	COE2	-----	-M-FGIQDT-	-----	L-GR-	G-PTLKEK-	S L-GA-	EMD	--SVR	SWVR	NVGVVDANVA	AQSGVALSRA	HFEKQPPSNL	RKSNFFHFVL	
Xt	COE2	-----	-M-FGVQDT-	-----	L-ARS	G-TALRDK-	A L-GV-	GMD	--PVR	SWVR	NVGVVDAKVA	AQSGVAVSRA	HFEKQPPSNL	RKSNFFHFVL	
Dr	COE2	-----	-M-FESQDH-	-----	S-IRT	I-TELKVR-	T F-EE-	EMD	--PVR	SWVR	NVGVIDANIA	AQSGVALSRA	HFEKQPPSNL	RKSNFFHFVL	
COE3															
Mm	COE3	-----	-M-FGIQEN-	-----	I-PRG	G-TTMKEE-	P L-GS-	GMN	--PVR	SWMH	TAGVVDANTA	AQSGVGLARA	HFEKQPPSNL	RKSNFFHFVL	
Hs	COE3	-----	-M-FGIQEN-	-----	I-PRG	G-TTMKEE-	P L-GS-	GMN	--PVR	SWMH	TAGVVDANTA	AQSGVGLARA	HFEKQPPSNL	RKSNFFHFVL	
Xt	COE3	-----	-M-FGIQEN-	-----	I-PRG	G-TTMKEE-	P L-GG-	GMN	--PVR	SWMH	TAGVVDANTA	AQSGVGLARA	HFEKQPPSNL	RKSNFFHFVL	
Dr	COE3	-----	-M-FGIQEN-	-----	I-PRG	G-TTMKEE-	P L-GS-	GMN	--PVR	SWMH	TAGVVDANTA	AQSGVGLARA	HFEKQPPSNL	RKSNFFHFVL	
EBF4															
Mm	EBF4	-----	-M-FPAQDA-	-----	L-PRG	G-LHLKEE-	P LLPS-	SLG	--SVR	SWMQ	SAGILDSNTA	AQSGVGLARA	HFEKQPPSNL	RKSNFFHFVL	
Hs	EBF4	-----	-----	-----	L-PRS	G-LNLKEE-	P LLPA-	GLG	--SVR	SWMQ	GAGILDASTA	AQSGVGLARA	HFEKQPPSNL	RKSNFFHFVL	
Md	EBF4	-----	-M-FSMQDS-	-----	L-PRG	G-LSLKEE-	P L-PA-	GLG	--SVR	SWMQ	GTGILDATTA	AQSGVGLARA	HFEKQPPSNL	RKSNFFHFVL	
Unassigned															
Dr	COE	-----	-----	-----	---	--MKQE-	V R-GS-	SAG	--SAV	SWLH	THSSL--ST-	-QRD	VGLARA	HFEKQPPANL	RKSNFFHFVL
LAMPREYS															
Pm	COE-A	-----	-M-FGIQES-	-----	L-PRG	G-SGLKEE-	P L-GG-	GMP	--TVR	SWMH	S-GVVDASTA	AQSGIALARA	HFEKQPPSNL	RKSNFFHFVL	
NON-VERTEBRATES															
Ci	COE	-----	-----	-----	---	---	---	---	---	---	---	---	VGLHRA	HFEKQPPNLL	RKSNFFHFVL
Bf	COE	-----	-----	-----	L-PRG	T-TTLKEE-	P L-GS-	QIG	--SVR	NWMQ	P-SIVDQT-A	TSGSVGLARA	HFEKQPPSNL	RKSNFFHFVI	
Lg	COE	-----	-M-FVT-	-----	D-QET	T-SPMKEE-	S L-GS-	QLS	--AVR	NWMQ	PT-VLDQN--	--GSVGLSRA	HFEKQPPSNL	RKSNFFHFVL	
Cc	COE	-----	-M-FSLQDS-	-----	L-QPRG	P-LSMKEE-	P L-GAPQAPLG	---	---	---	---	---	ATASVGLGRA	HFEKQPPSNL	RKSNFFHFVL
Dm	COE	MEWGRKLYPS	AVSGPSRSAGG	LM-FGLPPTA	AVDMNQ	PRG	PMTSLKEE-	P L-GS-	---	---	RWAM	QP-VVDQ---	--SNLIGIRA	HFEKQPPSNL	RKSNFFHFVI
Nv	COE	-----	-MSFDMQE-	-----	D---	E-NSMKNDRP	V-V-	---	---	---	GCLQ	HR-VLEPQVT	STASGLLARC	HFEKQPPANL	RKSNFFHFVL

DBD

GNATHOSTOMES

COE1													
Mm	COE1	ALYDRQGQPV	EIERTAFVGF	VEKEK	---E	ANSEKTNNGI	HYRLQLLYSN	GIRTEQDFYV	RLIDSMTKQA	IYVEGQDKNP	EMCRVLLT--	-HEIMCSRCC	DKKSCGNRNE
Hs	COE1	ALYDRQGQPV	EIERTAFVGF	VEKEK	---E	ANSEKTNNGI	HYRLQLLYSN	GIRTEQDFYV	RLIDSMTKQA	IYVEGQDKNP	EMCRVLLT--	-HEIMCSRCC	DKKSCGNRNE
Md	COE1	ALYDRQGQPV	EIERTAFVGF	VEKEK	---E	ANSEKTNNGI	HYRLQLLYSN	GIRTEQDFYV	RLIDSMTKQA	IYVEGQDKNP	EMCRVLLT--	-HEIMCSRCC	DKKSCGNRNE
Xt	COE1	-----	-----	-----	---	---	---	---	---	---	---	---	---
Dr	COE1	ALYDRQGQPV	EIERTSFVGF	VEKEK	---A	RIGEKTNNGI	HYRLQLLYSN	GIRTEQDFYV	RLIDSMTKQA	IYVEGQDKNP	EMLTALFLSS	SPSVSISRCC	DKKSCGNRNE
COE2													
Mm	COE2	ALYDRQGQPV	EIERTAFVDF	VENDK	---E	QGNEKTNNGT	HYKLQLLYSN	GVRTEQDLYV	RLIDSVTQKQ	IAYEGQNKNP	EMCRVLLT--	-HEVMCSRCC	EKKSCGNRNE
Hs	COE2	ALYDRQGQPV	EIERTAFVDF	VENDK	---E	QGNEKTNNGT	HYKLQLLYSN	GVRTEQDLYV	RLIDSVTQKQ	IAYEGQNKNP	EMCRVLLT--	-HEVMCSRCC	EKKSCGNRNE
Xt	COE2	ALYDRQGQPI	EIERTAFVDF	VENEK	---E	LSTEKTNNGT	HYKLQLLYSN	GVRTEQDLYV	RLIDSVTQKQ	IAYEGQNKNP	EMCRVLLT--	-HEVMCSRCC	EKKSCGNRNE
Dr	COE2	ALYDRNGQPV	EVERTSFVDF	VEQDK	---	-TGEKTNNGT	HYKLQLLYSN	GVRTEQDLYA	RLIDSVTQKQ	ISYEGQNKNP	EMCRVLLT--	-HEVMCSRCC	EKKSCGNRNE
COE3													
Mm	COE3	ALYDRQGQPV	EIERTAFVDF	VEKEK	---E	PNNEKTNNGI	HYKLQLLYSN	GVRTEQDLYV	RLIDSMTKQA	IYVEGQDKNP	EMCRVLLT--	-HEIMCSRCC	DKKSCGNRNE
Hs	COE3	ALYDRQGQPV	EIERTAFVDF	VEKEK	---E	PNNEKTNNGI	HYKLQLLYSN	GVRTEQDLYV	RLIDSMTKQA	IYVEGQDKNP	EMCRVLLT--	-HEIMCSRCC	DKKSCGNRNE
Xt	COE3	ALYDRQGQPV	EIERTAFVDF	VEKDK	---E	PNSEKTNNGI	HYKLQLLYSN	GVRTEQDLYV	RLIDSMTKQA	IYVEGQDKNP	EMCRVLLT--	-HEIMCSRCC	DKKSCGNRNE
Dr	COE3	ALYDRQGQPV	EIERTAFVDF	VEKEK	---E	PNSEKTNNGI	HYKLQLLYSN	GVRTEQDLYV	RLIDSMTKQA	IYVEGQDKNP	EMCRVLLT--	-HEIMCSRCC	DKKSCGNRNE
EBF4													
Mm	EBF4	AMYDRQGQPV	EVERTAFIDF	VEKDR	---E	PGTEKTNNGI	HYRLRLVYNN	GLRTEQDLYV	RLIDSMSKQA	IYVEGQDKNP	EMCRVLLT--	-HEIMCSRCC	DRKSCGNRNE
Hs	EBF4	AMYDRQGQPV	EVERTAFIDF	VEKDR	---E	PGAETNNGI	HYRLRLVYNN	GLRTEQDLYV	RLIDSMSKQA	IYVEGQDKNP	EMCRVLLT--	-HEIMCSRCC	DRKSCGNRNE
Md	EBF4	AMYDRQGQPV	EIERTAFIDF	VEKER	---E	PSGERTNNGI	HYRLQLLYSN	GIRTEQDLYV	RLIDSMSKQA	IYVEGQDKNP	EMCRVLLT--	-HEIMCSRCC	DKKSCGNRNE
Unassigned													
Dr	COE	ALYDSQGQSV	EIEQTTFVDF	VEKDK	---E	PVSEKTNNGI	HYKLQLLYNN	GVRTEQDLYI	RLIDSVTKQA	IVFEGQDKNP	EMCRVLLT--	-HETMCSRCC	DKKSCGNRNE
LAMPREYS													
Pm	COE-A	ALYDRQGQPV	EIERTAFVDF	VEKDR	---E	PNSEKTNNGI	HYRLQLLYSN	GVRTEQDLYV	RLIDSMTKQX	XXXXXXXXXX	XXXXXXXXXX	-XXXXXSRCC	DKKSCGNRNE
NON-VERTEBRATES													
Ci	COE	ALYDRQGQPV	EIERSAFVGF	VENET	---E	IAGEKTNNGI	QYRLQLLYHS	GVRTEQDVVF	RLIDSATKQS	ITYEGQDKNP	EMRRVLLT--	-HEIMCSRCC	DKKSCGNRNE
Bf	COE	ALYDRQGQPV	EIERTAFIDF	VEKDR	ISPOE	TAGENTRNNGI	HYRLQLLYSN	GVRTEQDLYV	RLIDSMTKQA	IYVEGQDKNP	EMQVLLT--	-HEIMCSRCC	DKKSCGNRNE
Lg	COE	ALYDRQGQPV	EIERTAFIDF	VEKDO	---E	HESQKTNNGI	HYRLQLLYSN	GVRQEIEYV	RLIDSVSKQA	IYVEGQDKNP	EMCRVLLT--	-HEVMCSRCC	DKKSCGNRNE
Cc	COE	ALYDRQGQPV	EVERTAFIDF	VEKER	---E	PEG-KTNNGI	HYRLQLLYAN	GVRQEQDLYV	RLIDSVTQKQA	IYVEGQDKNP	EMCRCLLT--	-HEIMCSRCC	DKKSCGNRNE
Dm	COE	ALYDRAGQPI	EIERTAFIGF	IEKDS	---E	SDATKTNNGI	QYRLQLLYAN	GARQEQDIFV	RLIDSVTQKQA	IYVEGQDKNP	EMCRVLLT--	-HEVMCSRCC	DKKSCGNRNE
Nv	COE	AFFDHNGQAV	EIERASFIDF	IEHLP	---E	--DKNVRNGV	LYRVLLLYSN	GIRTEQDIFV	RLIDAATKI	VPYEGQDKNP	EMRRVLLT--	-HEVMCSRCC	EKKSCGNRNE

DBD

GNATHOSTOMES

COE1														
Mm	COE1	TPSDPVIIDR	FFLKFFLKC	QNCLKNAGNP	RDMRRFQVVV	STTVNVGDHV	LAVSDNMFVH	NNSKHGRRAR	RL-DPSE	GT-PSYLEH	-----	-----	-----	ATPC

Hs	COE1	TPSDPVIIDR	FFLKFFLKC	QNCLKNAGNP	RDMMRFQVVV	STTVNVDGHV	LAVSDNMFVH	NNSKHGRRAR	RL	DPSEGT-	PSYLEH----	-----	-----	ATPC	
Md	COE1	TPSDPVIIDR	FFLKFFLKC	QNCLKNAGNP	RDMMRFQVVV	STTVNVDGHV	LAVSDNMFVH	NNSKHGRRAR	RL	DPSEGT-	PSYLEH----	-----	-----	ATPC	
Xt	COE1	-----	FFLKFFLKC	QNCLKNAGNP	RDMMRFQVVV	STTVNVDGHV	LAVSDNMFVH	NNSKHGRRAR	RL	DPSEGT-	PSYLEHEA--	-----	-----	ATPC	
Dr	COE1	TPSDPVIIDR	FFLKFFLKC	QNCLKNAGNP	RDMMRFQVVV	STTVSVDGHV	LAVSDNMFVH	NNSKHGRRAR	RL	DPSEGT-	PSYLEH----	-----	-----	ATPC	
	COE2														
Mm	COE2	TPSDPVIIDR	FFLKFFLKC	QNCLKTAGNP	RDMMRFQVVL	STTVNVDGHV	LAVSDNMFVH	NNSKHGRRAR	RL	DPSE---	-----	-----	-----	ATPC	
Hs	COE2	TPSDPVIIDR	FFLKFFLKC	QNCLKTAGNP	RDMMRFQVVL	STTVNVDGHV	LAVSDNMFVH	NNSKHGRRAR	RL	DPSE---	-----	-----	-----	ATPC	
Xt	COE2	TPSDPVIIDR	FFLKFFLKC	QNCLKTAGNP	RDMMRFQVVL	STTVNVDGHV	LAVSDNMFVH	NNSKHGRRAR	RL	DPSE---	-----	-----	-----	ATPC	
Dr	COE2	TPSDPVIIDR	FFLKFFLKC	QNCLKTAGNP	RDMMRFQVVL	STTVSVDGPV	LAISDNMFVH	NNSKHGRRSR	RM	DPNETV-	ENNMEY----	-----	-----	ATPC	
	COE3														
Mm	COE3	TPSDPVIIDR	FFLKFFLKC	QNCLKNAGNP	RDMMRFQVVV	STTVNVDGHV	LAVSDNMFVH	NNSKHGRRAR	RL	DPSE---	-----	-----	-----	ATPC	
Hs	COE3	TPSDPVIIDR	FFLKFFLKC	QNCLKNAGNP	RDMMRFQVVV	STTVNVDGHV	LAVSDNMFVH	NNSKHGRRAR	RL	DPSEGT-	PSYLEN----	-----	-----	ATPC	
Xt	COE3	TPSDPVIIDR	FFLKFFLKC	QNCLKNAGNP	RDMMRFQVVV	STTVNVDGHV	LAVSDNMFVH	NNSKHGRRAR	RL	DPSEGT-	PSYLENV---	-----	-----	ATPC	
Dr	COE3	TPSDPVIIDR	FFLKFFLKC	QNCLKNAGNP	RDMMRFQVVV	STTVNVDGHV	LAVSDNMFVH	NNSKHGRRAR	RL	DPSEGT-	TPYLENA---	-----	-----	ATPC	
	EBF4														
Mm	EBF4	TPSDPVIIDR	FFLKFFLKC	QNCLKNAGNP	RDMMRFQVVV	STTVSVDGHV	LAVSDNMFVH	NNSKHGRRAR	RL	DPSEA--	-----	-----	-----	ATPC	
Hs	EBF4	TPSDPVIIDR	FFLKFFLKC	QNCLKNAGNP	RDMMRFQVVV	STTVSVDGHV	LAVSDNMFVH	NNSKHGRRAR	RL	DPSEA--	-----	-----	-----	ATPC	
Md	EBF4	TPSDPVIIDR	FFLKFFLKC	QNCLKNAGNP	RDMMRFQVVV	STTVSVDGHV	LAVSDNMFVH	NNSKHGRRAR	RL	DPSEA--	-----	-----	-----	ATPC	
	Unassigned														
	DrCOE	TPSDPVIIDR	FFLKFFLKC	QNCLKNAGNP	RDTRRFQVLV	STTASVKAHI	LAISDNMFVH	NNSKHGRRAR	RF	ELIEA--	-----	-----	-----	EIPC	
	LAMPREYS														
	PmCOE-A	TPSDPVIIDR	FFLKFFLKC	QNCLKNAGNP	RDMMRFQVVV	STTVHVDGHV	LAVSDNMFVH	NNSKHGRRAR	RI	DPSE---	-----	-----	-----	ATPC	
	NON-VERTEBRATES														
	Ci	COE	TPSDPVVIDR	YFLKFFLKC	QNCLKNAGNP	RDMMRFQVVV	STTVHVDGHV	LAVSDNMFVH	NNSKHGRRAR	RV	DPSE---	-----	-----	ASPT	
	Bf	COE	TPSDPVVIDR	FFLKAFLKC	QNCLKNAGNP	RDMMRFQVVI	ATTNVDGHV	LAVSDNMFVH	NNSKHGRRAR	RL	DPSEGE	MSFCGDSPI	NRGQK----	-----	ATPC
	Lg	COE	TPSDPVVIDR	FFLKFFMCK	QNCLKNAGNP	RDMMRFQVAV	STSVADGPL	LTISDNMFVH	NNSKHGRRAR	RL	DPTEGGD	LRFSHLRPKL	SPFAA----	-----	TTPC
	Cc	COE	TPSDPVVIDR	FFLKFFMCK	QNCLKNAGNP	RDMMRFQVSI	GTSSSVEGPL	LAVSDNMFVH	NNSKHGRRAR	RL	DPGDAGF	LCA-----	-----	-----	ATPT
	Dm	COE	TPSDPVVIDR	FFLKFFLKC	QNCLKNAGNP	RDMMRFQVVI	STQVAVDGPL	LAISDNMFVH	NNSKHGRRAR	RL	DTTEGTG	NTSLSISGHP	LAPDSTYDGL	YPPLPV	ATPC
	Nv	COE	TPSDPVVIDR	YCLKFFMCK	QNCLKNAGNP	KDMMRFQVHV	STSDPVMYGM	IACSDNMFVH	NNSKHGRRAR	SR	ADGNEQDT	SA-----	-----	-----	DDPC

GNATHOSTOMES

IPT

H1

GNATHOSTOMES		IPT										H1		
COE1														
Mm	COE1	IKAISPSEGW	TTGGATVII	GDNFFDGLQV	IFGTMLVWSE	LITPHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGTPGRFI	YT	ALNEPTI	DYGFQRLQKV	IPRHGDDPER	
Hs	COE1	IKAISPSEGW	TTGGATVII	GDNFFDGLQV	IFGTMLVWSE	LITPHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGTPGRFI	YT	ALNEPTI	DYGFQRLQKV	IPRHGDDPER	
Md	COE1	IKAISPSEGW	TTGGATVII	GDNFFDGLQV	IFGTMLVWSE	LITPHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGTPGRFI	YT	ALNEPTI	DYGFQRLQKV	IPRHGDDPER	
Xt	COE1	IKAISPSEGW	TTGGATVII	GDNFFDGLQV	IFGTMLVWSE	LITPHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGTPGRFI	YT	ALNEPTI	DYGFQRLQKV	IPRHGDDPER	
Dr	COE1	IKAISPSEGW	TTGGATVII	GDNFFDGLQV	IFGTMLVWSE	LITPHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGTPGRFI	YT	ALNEPTI	DYGFQRLQKV	IPRHGDDPER	
COE2														
Mm	COE2	IKAISPSEGW	TTGGAMVII	GDNFFDGLQV	VFGTMLVWSE	LITPHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGAPGRFI	YT	ALNEPTI	DYGFQRLQKV	IPRHGDDPER	
Hs	COE2	IKAISPSEGW	TTGGAMVII	GDNFFDGLQV	VFGTMLVWSE	LITPHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGAPGRFI	YT	ALNEPTI	DYGFQRLQKV	IPRHGDDPER	
Xt	COE2	IKAISPSEGW	TTGGAMVII	GDNFFDGLQV	VFGTMLVWSE	LITPHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGAPGRFI	YT	ALNEPTI	DYGFQRLQKV	IPRHGDDPER	
Dr	COE2	IKAISPSEGW	TTGGAMVII	GENFFDGLQV	VFGSMLVWSE	LITPHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGAPGRFI	YT	ALNEPTI	DYGFQRLQKL	IPRHGDDPK	
COE3														
Mm	COE3	IKAISPSEGW	TTGGATVII	GDNFFDGLQV	VFGTMLVWSE	LITPHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGAPGRFV	YT	ALNEPTI	DYGFQRLQKV	IPRHGDDPER	
Hs	COE3	IKAISPSEGW	TTGGATVII	GDNFFDGLQV	VFGTMLVWSE	LITPHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGAPGRFV	YT	ALNEPTI	DYGFQRLQKV	IPRHGDDPER	
Xt	COE3	IKAISPSEGW	TTGGATVII	GDNFFDGLQV	VFGTMLVWSE	LITPHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGAPGRFV	YT	ALNEPTI	DYGFQRLQKV	IPRHGDDPER	
Dr	COE3	IKAISPSEGW	TTGGATVII	GDNFFDGLQV	VFGTMLVWSE	LITPHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGAPGRFV	YT	ALNEPTI	DYGFQRLQKV	IPRHGDDPER	
EBF4														
Mm	EBF4	IKAISPSEGW	TTGGATVII	GDNFFDGLQV	VFGNVLWSE	LITPHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGAPGRFV	YT	ALNEPTI	DYGFQRLQKV	IPRHGDDPER	
Hs	EBF4	IKAISPSEGW	TTGGATVII	GDNFFDGLQV	VFGNVLWSE	LITPHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGAPGRFV	YT	ALNEPTI	DYGFQRLQKV	IPRHGDDPER	
Md	EBF4	IKAISPSEGW	TTGGATVII	GDNFFDGLQV	VFGTMLVWSE	LITPHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGAPGRFV	YT	ALNEPTI	DYGFQRLQKV	IPRHGDDPER	
Unassigned														
Dr	COE	IKAISPSEGW	TSGGASVII	GDQFFSGLQV	VFGTMLVWSE	VITPHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCRGAPGRFV	YT	ALNEPTI	DYGFQRLQKV	IPHHGDDVER	
LAMPREYS														
Pm	COE-A	IKAISPSEGW	TTGGATVII	GDNFFDGLQV	VFGTMLVWSE	LITPHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGAPGRFV	YT	ALNEPTI	DYGFQRLQKL	IPRHGDDPER	
NON-VERTEBRATES														
Ci	COE	IKAINPAEGW	TTGGATVII	GENFFDGLQV	VFGSMVWSE	LITQHAIRVQ	TPPRHLPGVV	EVTLSYKSKQ	FCSGAPGRFV	YT	ALNEPTL	DYGFQRLQKL	VPRHPGDDPER	
Bf	COE	IKAISPSEGW	TTGGATVII	GDNFFDGLQV	VFGTMLVWSE	LITPHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGAPGRFV	YT	SLNEPTI	DYGFQRLQKL	VPRHPGDDPER	
Lg	COE	IKAICPSEGW	TSGGATVII	GDNFFDGLQV	VFGTMLVWSE	LITSHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGAPGRFV	YT	ALTEPTI	DYGFQRLQKL	VPRHPGDDPER	
Cc	COE	IKAICPSEGW	TTGGATVII	GDNFFDGLQV	VFGTMLVWSE	LITSHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGAPGRFV	YT	ALTEPTI	DYGFQRLQKL	VPRHPGDDPER	
Dm	COE	IKAISPSEGW	TTGGATVII	GDNFFDGLQV	VFGTMLVWSE	LITSHAIRVQ	TPPRHIPGVV	EVTLSYKSKQ	FCKGSPGRFV	YV	SALNEPTI	DYGFQRLQKL	IPRHGDDPER	
Nv	COE	IKAICPNEGW	TIGGSNVILI	GDNFFDGLQV	VFGSFIWSE	FITPHALRVQ	APPSMPGVV	KVYLMHKEKQ	YCKYAPAKFG	YT	ALVEPTI	DYGFQRLSKL	IPRHGDDPER	

GNATHOSTOMES

H2d

H2a

GNATHOSTOMES		H2d										H2a									
COE1																					
Mm	COE1	LPKEVILKRA	ADLVEALYGM	PHNNQ--EII	LKRAADIAEA	LYSVPRNHQ	LPALA-NTSV	HA-----	-----	-----	GMMG	VNS	-----	F	SSQLAVNVSE						
Hs	COE1	LPKEVILKRA	ADLVEALYGM	PHNNQ--EII	LKRAADIAEA	LYSVPRNHQ	LPALA-NTSV	HA-----	-----	-----	GMMG	VNS	-----	F	SSQLAVNVSE						
Md	COE1	LPKEVILKRA	ADLVEALYGM	PHNNQ--EII	LKRAADIAEA	LYSVPRNHQ	LPALA-NTSV	HA-----	-----	-----	GMMG	VNS	-----	F	SSQLAVNVSE						
Xt	COE1	LPKEVILKRA	ADLVEALYGM	PHNNQ--EII	LKRAADIAEA	LYSVPRNHSQ	LPALT-NSSV	HP-----	-----	-----	GMMG	VNS	-----	F	SSQLAVNVSE						
Dr	COE1	LPKEVILKRA	ADLVEALYGM	PHNNQ--EII	LKRAADIAEA	LYNVPRGHQ	LPGLT-NSSV	HS-----	-----	-----	GMMG	VNS	-----	F	HSQLAVNVSD						
COE2																					
Mm	COE2	LAKEMLLKRA	ADLVEALYGT	PHNNQ--DII	LKRAADIAEA	LYSVPRNPSQ	IPALS-SSPA	HS-----	-----	-----	GMMG	INS	-----	Y	GSQGVISISE						
Hs	COE2	LAKEMLLKRA	ADLVEALYGT	PHNNQ--DII	LKRAADIAEA	LYSVPRNPSQ	IPALS-SSPA	HS-----	-----	-----	GMMG	INS	-----	Y	GSQGVISISE						
Xt	COE2	LAKEMLLKRA	ADLVEALYGT	PHNNQ--DII	LKRAADIAEA	LYSVPRNHQ	IPALS-SSPV	HS-----	-----	-----	GMMG	INS	-----	Y	GGQLGVISISE						
Dr	COE2	LAKEMLLKRA	ADVVEESLYGN	TTSNQ--DML	LKRAADIAEA	LYSVPRPHSQ	LQAMP-SSPV	HG-----	-----	-----	SVMG	LSS	-----	Y	PTQLGVISIGE						
COE3																					
Mm	COE3	LPKEVLLKRA	ADLVEALYGM	PHNNQ--EII	LKRAADIAEA	LYSVPRNHQ	IPTLG-NTPA	HT-----	-----	-----	GMMG	VNS	-----	F	SSQLAVNVSE						
Hs	COE3	LPKEVLLKRA	ADLVEALYGM	PHNNQ--EII	LKRAADIAEA	LYSVPRNHQ	IPTLG-NNPA	HT-----	-----	-----	GMMG	VNS	-----	F	SSQLAVNVSE						
Xt	COE3	LPKEVLLKRA	ADLVEALYGM	PHNNQAKEII	LKRAADIAEA	LYSVPRNHQ	IPSLA-NTPS	HS-----	-----	-----	GMMG	VNS	-----	F	SSQLAVNVSE						
Dr	COE3	LPKEVLLKRA	ADLVEALYGM	PHNNQ--EII	LKRAADIAEA	LYSVPRNHQ	IPSLA-NTAS	HG-----	-----	-----	GMMG	VNS	-----	F	SSQLAVNVSE						
EBF4																					
Mm	EBF4	LPKEVLLKRA	ADLAEALYGV	PSSNQ--ELL	LKRAADVAEA	LYSAPRAPAP	LGPLA-PSHP	HP-----	-----	-----	AVVG	INA	-----	F	SSPLAIAVGD						
Hs	EBF4	LPKEVLLKRA	ADLAEALYGV	PGSNQ--ELL	LKRAADVAEA	LYSTPRAPGP	LAPLA-PSHP	HP-----	-----	-----	AVVG	INA	-----	F	SSPLAIAVGD						
Md	EBF4	LPKEVLLKRA	ADLVEALYGM	PHSNQ--DLI	LKRAADIAEA	LYSVPRTPSQ	LGSLA-PGHP	HT-----	-----	-----	AMMG	INS	-----	F	GSQAVNIGD						
Unassigned																					
Dr	COE	LPKEVLLKRA	ADLIETFYGA	PQNNQ--EII	LKRASDIAEA	LNSIPRNPS-	-----HTNP	H-----	-----	-----	GMMG	VNS	-----	Y	DGQLSLNTEV						
LAMPREYS																					
Pm	COE-A	LPK-----	-----	-----Q--EII	LKRAADIAEA	LYSVPRNHQ	LPAIN-NSPP	GLP-----	-----	-----	GMMG	MNS	-----	F	GGQLAISVSE						
NON-VERTEBRATES																					
Ci	COE	LPK-----	-----	-----EII	LKRAADVMEA	VISRQYAPPS	QMPPSAGITP	PAPHLAAAPC	APPGSFVPS	ASAA--	MAVA	MNGYAAAA	-V	SSQFGGTPDR							
Bf	COE	LPK-----	-----	-----EII	LKRAADLAEA	LYSMRPNPL	PALTPRPSA	VNNSVGAG--	-----	-----	GMMG	MNS	-----	F	GNQLAVSVDP						
Lg	COE	LPK-----	-----	-----EII	LKRAADLAEA	LYSMRPNHNP	LSLPAPRSPA	MNNTS----	-----	-----	PMNG	FNA	-----	Y	PSQLAVSVGS						
Cc	COE	LPKGICEGHV	GPLSSQIFLS	CPQ---EII	LKRAADLAEA	LYSMRPNPNQ	LSLPAPRSPA	MN-----	-----	-----	NTTG	FNS	-----	Y	TGQLAVSVQD						
Dm	COE	LQK-----	-----	-----EII	LKRAADLVEA	LYSMRSPG-	-----	-----	-----	-----	GSTG	FNS	-----	Y	AGQLAVSVQD						
Nv	COE	IPK-----	-----	-----EIV	LKRAADLAET	LYQMPRTPTQ	VHPSQYASTG	LATPKSPALV	GGQHYVTLPL	DSMTNVFTQS	ITQDNQPIAY	SSHNDNASDQ									

GNATHOSTOMES

GNATHOSTOMES	
COE1	
Mm	COE1
	ASQATNQ--

	-----GFTN--NSS SV-----
	SPHG-YV-P- STTPQQTN--

	YNS-VTTSMN GYGSA--MSN LG--GSPTFL

Hs	COE1	ASQATNQ---	----	----	----	GFTR-NSS SV	----	SPHG-YV-P	STTPQQTN	----	----	YNS-VTSMN	GYGSA	-MSN	LG--GSPTFL
Md	COE1	ASQATNQ---	----	----	----	GFTR-NSS SV	----	SPHG-YV-P	STTPQQTN	----	----	YNS-VTSMN	GYGNAG	-MSN	LG--GSPTFL
Xt	COE1	PSQVNTQ---	----	----	----	GFSR-NTS SV	----	SPHG-YA-P	STTPQQTN	----	----	YSS-VTSMN	GYGNTA	-MSN	LA--GSPSFL
Dr	COE1	STQAANQ---	----	----	----	GFSR-NTS SV	----	SPHG-YV-P	STTPQQSS	----	----	YST-VTSMN	GYGNAG	-MTT	LG--GSPNFL
COE2															
Mm	COE2	STQGNQ---	----	----	----	GYIR-NTS SI	----	SPRG-YV-S	SSTPQQSN	----	----	YST-SSNSMN	GYSNVP	-MAN	LGVPGSPGFL
Hs	COE2	STQGNQ---	----	----	----	GYIR-NTS SI	----	SPRG-YV-S	SSTPQQSN	----	----	YST-SSNSMN	GYSNVP	-MAN	LGVPGSPGFL
Xt	COE2	SQAN-NQ---	----	----	----	GYIR-NTS SI	----	SPRG-YV-S	SSTPQQSN	----	----	YST-PSNSMN	GYSNVP	-MSN	LGVPGSPGFI
Dr	COE2	PGQTSQV---	----	----	----	GYTR-NSS SL	----	SPRG-YE-S	SSTPQQSA	----	----	YGS-NGGM--	SYGAVP	-MSS	LGVSGSPGFN
COE3															
Mm	COE3	TSQANDQV--	----	----	----	GYSR-NTS SV	----	SPRG-YV-P	SSTPQQSN	----	----	YNT-VTSMN	GYGSA	-MAN	LGVPGSPGFL
Hs	COE3	TSQANDQV--	----	----	----	GYSR-NTS SV	----	SPRG-YV-P	SSTPQQSN	----	----	YNT-VTSMN	GYGSA	-MAS	LGVPGSPGFL
Xt	COE3	TSQANDQV--	----	----	----	GYSR-NTS SV	----	SPRG-YV-P	SSTPQQSN	----	----	YNT-VNSMN	GYGNAG	-MPN	LGVPGSPGFL
Dr	COE3	TSQVFLSTV-	----	----	----	GYSR-NTS SV	----	SPRG-YV-P	SSTPQQSN	----	----	YNT-VNSMN	GYGNTG	-MPN	LGVPSPPGFL
EBF4															
Mm	EBF4	TTP--EP---	----	----	----	GYAR-SCG SA	----	SPR--FA-P	SPGSQQSS	----	----	YGSGLGAGLG	SYGAPG	-VTG	LGVPGSPSFL
Hs	EBF4	ATPGPEP---	----	----	----	GYAR-SCS SA	----	SPRG-FA-P	SPGSQQSG	----	----	YGGGLGAGLG	GYGAPG	-VAG	LGVPGSPSFL
Md	EBF4	STQGL----	----	----	----	GYSR-NTS SV	----	SPRG-YV-P	SSTPQQSS	----	----	YSS-ITSSIN	GYGATG	-MAG	LGVPSPPSFL
Unassigned															
Dr	COE	SQODAERA--	----	----	----	SFNR-S--	----	SNTG-FI-E	GSTSQQSE	----	----	FNN-TNMCMN	G
LAMPREYS															
Pm	COE-A	ASQAADQ---	----	----	----	GYTR-NSS SV	----	SPHG-YV-P	SSTPQQSH	----	----	YSS-VSSSMN	GYGVS	GGMG	LG.....
NON-VERTEBRATES															
Ci	COE	FD-TGSDS--	----	----	----	GYSR--GN SV	----	SPRNGYS-P	QTPHSLNSG	SIGSMVGLTT	VGA-VPAPAP	YHCAP	S-FNS	YSSASVLKYN	
Bf	COE	NSSTNGNGQV	QGFVVFVFTYS	GLKEDCAEQT	NHGYSR-QSN SV	----	SPRG-YG-A	-STPHSTN-G	SISSYST---	----	SASGLN	GYGSSG	-NLS	NMPVPSPPGF	
Lg	COE	VA-DTASQW	-EEGWCKLRT	SR-----	SYNRSQSS SV	----	SPRG-YG-SN	GSTPHSNG-S	YT-----	----	STPAMN	GYHHGG	-GAL	GNMNVNAPHSP	
Cc	COE	TA-N--GQW	-EE-----	----	GWSH-LTS SLLGYSSRMT	PPHG-YP-M	CSGSLSTS-C	SAYSGSS---	----	STYSRS	GYIDGC	-TSS	IPIGTPHSP		
Dm	COE	GS-----GQW	TED-----	----	DYQRAQSS SV	SPRGGYCSS	ASTPHSSG-G	SY-----	----	-----	--GATA	-ASA	AVAATANGYA		
Nv	COE	TD-MNNNVVQ	SS-----	----	NSSR-LTS NA	TLSG-VI-N	MSNHDSHNG	VMASDEGLS	AIATVNSCY	GNGSPA	-LSC	MAVPASPGYF			

GNATHOSTOMES

COE1															
Mm	COE1	NGSAANSPLY	IVPSSPTMA-	-SSTSLPNC	SSSSGIFSPS	PANMVSAVKQ	KSA--FAPVV	RPQTS-PPPT	CTST-NGNSL	QAISGMIVPP	M				
Hs	COE1	NGSAANSPLY	IVPSSPLMA-	-SSTSLPNC	SSSSGIFSPS	PANMVSAVKQ	KSA--FAPVV	RPQTS-PPPT	CTST-NGNSL	QAISGMIVPP	M				
Md	COE1	NGSAANSPLY	IVPSSPTMA-	-SSTSLPNC	SSSSGIFSPS	PANMVSAVKQ	KSA--FAPVV	RPQTS-PPPT	CTST-NGNSL	QAISGMIVPP	M				
Xt	COE1	NGSAANSPLY	IVPSSPTMA-	-SSTSLPNC	SSSSGIFSPS	PANMVSAVKQ	KSA--FAPVV	RPQAS-PPPT	CSSS-SATS	QAISGMIA	PP M				
Dr	COE1	NGSAANSPLY	IVPSSPTMA-	-SSTSLPNC	SSSSGIFSPS	PANMVSAVKQ	KSA--FAPVV	RPQAS-PPPT	CTSA-NGNGL	QGESG----	-				
COE2															
Mm	COE2	NGSPTGSPYG	IMSSSPTV--	-----GS	SSTSILPFS	SSVFP-AVKQ	KSA--FAPVI	RPQGS-PSPA	CSSG-NGNGF	RAMTGLVVPP	M				
Hs	COE2	NGSPTGSPYG	IMSSSPTV--	-----GS	SSTSILPFS	SSVFP-AVKQ	KSA--FAPVI	RPQGS-PSPA	CSSG-NGNGF	RAMTGLVVPP	M				
Xt	COE2	NGSPTTSPYG	IMPSSPPV--	-----GS	SGSSILPFS	SSVFP-SIKQ	KSA--FAPVI	RPQGS-PSPA	CSSS-NSNGF	RAMTGLVVPP	M				
Dr	COE2	SASPNSPYA	IMPSSPP--	-----GS	SSSSSLLPFS	S--FPSSTKQ	KSA--FAPVL	RPQGF-PHP	SAKTSGGTSF	RAMTGLVVPP	M				
COE3															
Mm	COE3	NGSSANSPLY	-----	-----	-----	-----MKQ	KSA--FAPVV	RPQAS-PPPS	CTSA-NGNGL	QAMSGLVVPP	M				
Hs	COE3	NGSSANSPLY	IVPSSPTMAA	-SSVTLPSNC	SSTHGIFSPS	PANVISAVKQ	KSA--FAPVV	RPQAS-PPPS	CTSA-NGNGL	QAMSGLVVPP	M				
Xt	COE3	NGSSANSPLY	IVPSSPTMAA	-SSVTLPSNC	SSTHGIFSPS	PANVISAVKQ	KSA--FAPVV	RPQAS-PPPS	CTSA-NGNGL	QDMYFNPTFP	KS				
Dr	COE3	NGSSANSPLY	IVPSSPTMAA	-SSVSLSSNC	SSTHGIFSPS	PANVISAVKQ	KSA--FAPVV	RPQAS-PPPS	CTSA-NGNGL	QAMSGLVVPP	M				
EBF4															
Mm	EBF4	NGSTATSPFA	IMPSSPLAA	ASSMSLPAAA	PTTS-VFSFS	PVNMICAVKQ	RSA--FAPVL	RPPSS-PSQA	CPRA-HREGL	PAQRTGR					
Hs	EBF4	NGSTATSPFA	IMPSSPLAA	ASSMSLPAAA	PTTS-VFSFS	PVNMIASAVKQ	RSA--FAPVL	RPPSS-PPQA	CPRA-HGEGL	PDQSFEDSDK	FHSPARGLQ	LAYS			
Md	EBF4	NGSTANSPLY	IMPSSPPL-A	ASSISLPAAA	PTTS-VFSFS	PVNMIASAVKQ	KSA--FAPPA	SLLCP-PLVP	SVPL.LLDTD	PA..FEDSDK	FHSPARGLQ	LAYS			
Unassigned															
Dr	COE
LAMPREYS															
Pm	COE.A
NON-VERTEBRATES															
Ci	COE	LFTVIPSPH	NGMNLPPSSG	TTT-----	-----GIFSPS	PANMISAARK	KSA--FAPVH	RPHNS-PSPL	APSNGNIGKI	SCFVLSSI					
Bf	COE	LNGSTIPSSP	TMPPTPSSLA	TNASTP----	-----GIFFS	PANMISAARK	KSA--FAPVV	RPQ--SSPS	PAGASSNGTG	LQGGWPTIP	SAVC				
Lg	COE	PMHNIGSQGL	GSCPT-----	-----	-----GIFSPS	PANMITAVKQ	RSA--FAPVL	RGNS-PPPT	FPGIQTSGWA	HPLIHVSIIS	TSLPSNALKV	SY			
Cc	COE	PSQSPSLPPT	ASCVAP---	-----	-----GLHSFS	HANVLSAVKQ	RSA--FAPVL	RPSSSPVP	MTSIAAAGWA	HTLPTSGMTD	CHSIHA				
Dm	COE	PAPNM-----	-----	-----	-----GTLSSS	PGSVFNSTR	VSSLSFNPFA	LP---	-----	-----	--TCNTQGY	TQLVTSK			
Nv	COE	TGAFIPQSPS	LPPTNSVPP	SSNN-----	-----SIFSP	P-NMIQAVKQ	KSA--FNAVS	RGPESTAHPR	SPAVQGVMT	STYNGPFAV	VSSVNAYIKS	IQPQVKP			

Figure S1 : Alignment used in the phylogenetic analysis of metazoan COE proteins

Dashes indicate missing residues; dots at the C-terminal end of DrCOE and PmCOE-A stand for undetermined residues. Only the positions shaded in yellow were used in the phylogenetic analysis. Highly truncated (XtCOE1, PmCOE-B) or divergent (Ce COE) predicted sequences were excluded from the analysis. The 8 to 10 residues inclusion between the DBD and IPT domains in vertebrate COE proteins, due to the existence of an alternative splice donor site is underlined in MmCOE1. Same abbreviations as in Figure 1. DBD, DNA Binding Domain ; IPT, Ig-like/Plexin/transcription Factor Domain ; H1/2d/2a, Helix H1/2d/2a.