

DNA BINDING DOMAIN

GNATHOSTOMES

COE1

<i>M. musculus</i>	COE1	RAHFQKPPS	NLRKSNFFHF	VLALYDRQGG	PVEIERTAFV	GFVEKEK---	-EANSEKTNN	GIHYRLQLLY	SNGIRTEQDF	YVRLIDSMTK	QAIYVEGQDK
<i>H. sapiens</i>	COE1	RAHFQKPPS	NLRKSNFFHF	VLALYDRQGG	PVEIERTAFV	GFVEKEK---	-EANSEKTNN	GIHYRLQLLY	SNGIRTEQDF	YVRLIDSMTK	QAIYVEGQDK
<i>M. domestica</i>	COE1	RAHFQKPPS	NLRKSNFFHF	VLALYDRQGG	PVEIERTAFV	GFVEKEK---	-EANSEKTNN	GIHYRLQLLY	SNGIRTEQDF	YVRLIDSMTK	QAIYVEGQDK
<i>O. anatinus</i>	COE1	RAHFQKPPS	NLRKSNFFHF	VLALYDRQGG	PVEIERTAFV	GFVEKEK---	-EANSEKTNN	GIHYRLQLLY	SNGXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
<i>G. gallus</i>	COE1	RAHFQKPPS	NLRKSNFFHF	VLALYDRQGG	PVEIERTAFV	GFVEKEK---	-EANSEKTNN	GIHYRLQLLY	SNGIRTEQDF	YVRLIDSMTK	QAIYVEGQDK
<i>X. tropicalis</i>	COE1	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
<i>D. rerio</i>	COE1	RAHFQKPPS	NLRKSNFFHF	VLALYDRQGG	PVEIERTSFV	GFVEKEK---	-ARIGKTTNN	GIHYRLQLLY	SNGIRTEQDF	YVRLIDSMTK	QAIYVEGQDK

COE2

<i>M. musculus</i>	COE2	RAHFQKPPS	NLRKSNFFHF	VLALYDRQGG	PVEIERTAFV	DFVENDK---	-EQGNEKTNN	GTHYKLQLLY	SNGVRTEQDL	YVRLIDSVTK	QPIAYEGQNK
<i>H. sapiens</i>	COE2	RAHFQKPPS	NLRKSNFFHF	VLALYDRQGG	PVEIERTAFV	DFVENDK---	-EQGNEKTNN	GTHYKLQLLY	SNGVRTEQDL	YVRLIDSVTK	QPIAYEGQNK
<i>O. anatinus</i>	COE2	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
<i>G. gallus</i>	COE2	RAHFQKPPS	NLRKSNFFHF	VLALYDRQGG	PVEIERTAFV	DFVENDK---	-EQGNEKTNN	GTHYKLQLLY	SNGVRTEQDL	YVRLIDSVTK	QPITYEGQNK
<i>X. tropicalis</i>	COE2	RAHFQKPPS	NLRKSNFFHF	VLALYDRQGG	PIEIERTAFV	DFVENEK---	-ELSTKTTNN	GTHYKLQLLY	SNGVRTEQDL	YVRLIDSVTK	QPIAYEGQNK
<i>D. rerio</i>	COE2	RAHFQKPPS	NLRKSNFFHF	VLALYDRNGQ	PVEVERTSFV	DFVEQDK---	---TGEKTTNN	GTHYKLQLLY	SNGVRTEQDL	YARLIDSVTK	QPISYEGQNK

COE3

<i>M. musculus</i>	COE3	RAHFQKPPS	NLRKSNFFHF	VLALYDRQGG	PVEIERTAFV	DFVEKEK---	-EPNNEKTNN	GIHYKLQLLY	SNGVRTEQDL	YVRLIDSMTK	QAIYVEGQDK
<i>H. sapiens</i>	COE3	RAHFQKPPS	NLRKSNFFHF	VLALYDRQGG	PVEIERTAFV	DFVEKEK---	-EPNNEKTNN	GIHYKLQLLY	SNGVRTEQDL	YVRLIDSMTK	QAIYVEGQDK
<i>O. anatinus</i>	COE3	RAHFQKPPS	NLRKSNFFHF	VLALYDRQGG	PVEIERTAFV	DFVEKEK---	-EPNNEKTNN	GIHYKLQLLY	SNGVRTEQDL	YVRLIDSMTK	QAIYVEGQDK
<i>G. gallus</i>	COE3	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
<i>X. tropicalis</i>	COE3	RAHFQKPPS	NLRKSNFFHF	VLALYDRQGG	PVEIERTAFV	DFVEKDK---	-EPNSEKTNN	GIHYKLQLLY	SNGVRTEQDL	YVRLIDSMTK	QAIYVEGQDK
<i>D. rerio</i>	COE3	RAHYEQPPS	NLRKSNFFHF	VLALYDRQGG	PVEIERTAFV	DFVEKEK---	-EPNSEKTNN	GIHYKLQLLY	SNGVRTEQDL	YVRLIDSMTK	QAIYVEGQDK

EBF4

<i>M. musculus</i>	EBF4	RAHFQKPPS	NLRKSNFFHF	VLAMYDRQGG	PVEVERTAFI	DFVEKDR---	-EPGTEKTNN	GIHYRLRLVY	NNGLRTEQDL	YVRLIDSMK	QAIYVEGQDK
<i>H. sapiens</i>	EBF4	RAHFQKPPS	NLRKSNFFHF	VLAMYDRQGG	PVEVERTAFI	DFVEKDR---	-EPGAETKNN	GIHYRLRLVY	NNGLRTEQDL	YVRLIDSMK	QAIYVEGQDK
<i>M. domestica</i>	EBF4	RAHFQKPPS	NLRKSNFFHF	VLAMYDRQGG	PVEIERTAFI	DFVEKER---	-EPSGERTNN	GIHYRLQLLY	SNGIRTEQDL	YVRLIDSMK	QAIYVEGQDK
<i>O. anatinus</i>	EBF4	RAHFQKPPA	NLRKSNFFHF	VLAMYDRHGQ	PVEVERSAFV	GFVEKGR---	-EHGGKTTNN	GIHYRLQLLY	SNGIRTEQDL	YVRLIDSVSK	QAIYVEGQDK
UNASSIGNED											
<i>D. rerio</i>	COE	RAHFQKPPA	NLRKSNFFHF	VLALYDSQGG	SVEIEQTTFV	DFVEKDK---	-EPVSEKTNN	GIHYKLQLLY	NNGVRTEQDL	YIRLIDSVTK	QAIYVEGQDK

LAMPREYS

<i>P. marinus</i>	COE-A	RAHFQKPPS	NLRKSNFFHF	VLALYDRQGG	PVEIERTAFV	DFVEKDR---	-EPNSEKTNN	GIHYRLQLLY	SNGVRTEQDL	YVRLIDSMNK	XXXXXXXXXX
<i>P. marinus</i>	COE-B	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX

NON-VERTEBRATES

<i>C. intestinalis</i>	COE	RAHFQKPPN	NLRKSNFFHF	VLALYDRQGG	PVEIERSAFV	GFVENET---	-EIAGEKTNN	GIQYRLQLLY	HSGVRTEQDV	FVRLIDSATK	QSITYEGQDK
<i>B. floridae</i>	COE	RAHFQKPPS	NLRKSNFFHF	VIALYDRQGG	PVEIERTAFI	DFVEKDRISP	QETAGENTRN	GIHYRLQLLY	SNGVRTEQDL	YVRLIDSMTK	QAIYVEGQDK
<i>S. purpuratus</i>	COE	RAHFQKPPS	NLRKSNFFHF	VIALYDRAGQ	PIEVERTSFV	DFIEKER---	-EPDATKTTNN	GIHYRIQMLF	HNGVRTEQDL	YVRLIDSMTK	QAIYVEGQDK
<i>L. gigantea</i>	COE	RAHFQKPPS	NLRKSNFFHF	VLALYDRQGG	PVEIERTAFI	DFVEKDQ---	-EHESQKTNN	GIHYRLQLLF	ANGVRQEQEI	YVRLIDSVSK	QAIYVEGQDK
<i>C. capitata</i>	COE	RAHYEQPPS	NLRKSNFFHF	VLALYDRQGG	PVEIERTAFI	DFVEKER---	-EPEG-KTTNN	GIHYRLQLLY	ANGVRQEQDL	YVRLIDSVTK	QAIYVEGQDK
<i>D. melanogaster</i>	COE	RAHFQKPPS	NLRKSNFFHF	VIALYDRAGQ	PIEVERTAFI	GFIEKDS---	-ESDATKTTNN	GIQYRLQLLY	ANGARQEQDI	FVRLIDSVTK	QAIYVEGQDK
<i>C. elegans</i>	COE	RAHFQKPPN	NLRKSNFFHF	VIALYDRNSQ	PIEVERTQFA	GFVEKEK---	-EVDGQDTRN	GIHYRLSLMF	QNGIRSEHDL	FVRLIDSSTK	QAITVEGQDK
<i>N. vectensis</i>	COE	RCHFQKPPA	NLRKSNFFHF	VLAFFDHNGQ	AVEIERASFI	DFIEH----	-LPEDKNVRN	GVLYRVLLLY	SNGIRTEQDI	FVRLIDAATK	QIVPYEGQDK
<i>A. queenslandica</i>	COE	AAQFQKPPD	NMRKSNFFHF	IISLFDGQQH	RIQVQKAVFK	DFYDTV----	-GTDEQEYRN	GLIYKLLVYV	SDGTQKEEEL	CVRLIDSNTK	QLVPYEGTNK

DNA BINDING DOMAIN

GNATHOSTOMES

COE1											
MmCOE1	NPENCRVLLT	---HEIMCSR	CCDKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQVVVSTTVN	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
HsCOE1	NPENCRVLLT	---HEIMCSR	CCDKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQVVVSTTVN	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
MdCOE1	NPENCRVLLT	---HEIMCSR	CCDKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQVVVSTTVN	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
OaCOE1	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
GgCOE1	NPENCRVLLT	---HEIMCSR	CCDKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQVVVSTTVN	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
XtCOE1	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXRFFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQVVVSTTVN	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
DrCOE1	NPEMLTALFL	SSSPSVSISR	CCDKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQVVVSTTVS	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
COE2											
MmCOE2	NPENCRVLLT	---HEVMCSR	CCEKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKT	AGNPRDMRR-	FQVVLSTTVN	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
HsCOE2	NPENCRVLLT	---HEVMCSR	CCEKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKT	AGNPRDMRR-	FQVVLSTTVN	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
OaCOE2	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXK-	VQVTGASKVN	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
GgCOE2	NPENCRVLLT	---HEVMCSR	CCEKKSCGNR	NETPSDPVII	DSYRFFLKFF	LKCNQNCLKT	AGNPRDMRR-	FQVVLSTTVN	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
XtCOE2	NPENCRVLLT	---HEVMCSR	CCEKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKT	AGNPRDMRR-	FQVVLSTTVN	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
DrCOE2	NPENCRVLLT	---HEVMCSR	CCEKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKT	AGNPRDMRR-	FQVVLSTTVS	VDGPVLAISD	NMFVHNNSKH	GRRSRRL-DP
COE3											
MmCOE3	NPENCRVLLT	---HEIMCSR	CCDKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQVVVSTTVN	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
HsCOE3	NPENCRVLLT	---HEIMCSR	CCDKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQVVVSTTVN	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
OaCOE3	NPENCRVLLT	---HEIMCSR	CCDKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQVVVSTTVN	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
GgCOE3	XXXXXXXXXX	XXXXXXXXXX	CCDKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQVVVSTTVN	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
XtCOE3	NPENCRVLLT	---HEIMCSR	CCDKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQVVVSTTVN	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
DrCOE3	NPENCRVLLT	---HEIMCSR	CCDKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQVVVSTTVN	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
EBF4											
MmEBF4	NPENCRVLLT	---HEIMCSR	CCDRKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQVVVSTTVS	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
HsEBF4	NPENCRVLLT	---HEIMCSR	CCDRKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQVVVSTTVS	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
MdEBF4	NPENCRVLLT	---HEIMCSR	CCDKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQVVVSTTVS	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
OaEBF4	NPENCRVLLT	---HEIMCSR	CCDKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
UNASSIGNED											
DrCOE	NPENCRVLLT	---HETMCSR	CCDKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKN	AGNPRDTRR-	FQVLVSTTAS	VKAHILAISS	NIFVHNNSKH	GRRARRF-EL

LAMPREYS

Pm COE-A	XXXXXXXXXX	XXXXXXXXXS	CCDKKNCGR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQVVVSTTVH	VDGHVLAVSD	NMFVHNNSKH	GRRARRI-DP
Pm COE-B	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX

NON-VERTEBRATES

CiCOE	NPENRRVLLT	---HEIMCSR	CCDKKSCGNR	NETPSDPVII	D--RYFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQVVVSTTVH	VDGHVLAVSD	NMFVHNNSKH	GRRARRV-DP
BfCOE	NPENQRVLLT	---HEIMCSR	CCDKKSCGNR	NETPSDPVII	D--RFFLKAF	LKCNQNCLKN	AGNPRDMRR-	FQVVIATTVN	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
SpCOE	NPENCRVLLT	---HEIMCSR	CCDKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQVVIISTPN	VDGHVLAVSD	NMFVHNNSKH	GRRARRL-DP
LgCOE	NPENCRVLLT	---HEVMCSR	CCDKKSCGNR	NETPSDPVII	D--RFFLKFF	MKCNQNCLKN	AGNPRDMRR-	FQVAVSTVA	IDGPLLTISD	NMFVHNNSKH	GRRARRL-DP
CcCOE	NPENCRVLLT	---HEIMCSR	CCDKKSCGNR	NETPSDPVII	D--RFFLKFF	MKCNQNCLKN	AGNPRDMRR-	FQVSIGTSSS	VEGPLLAVSD	NMFVHNNSKH	GRRARRL-DP
DmCOE	NPENCRVLLT	---HEVMCSR	CCDKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQVVIISTQVA	VDGPLLAISS	NMFVHNNSKH	GRRARRL-DT
CeCOE	NPENCRVLLT	---HEVMCSR	CCEKKSCGNR	NETPSDPVII	D--RFFLKFF	LKCNQNCLKN	AGNPRDMRR-	FQVVLCSAR	IDGPLLAVSD	NMFVHNNSKH	GRRTRKRT-DA
NvCOE	NPENRRVLLT	---HEVMCSR	CCEKKSCGNR	NETPSDPVII	D--RYCLKFF	MKCNQNCLKN	AGNPKDMRR-	FQVHVSTVSD	PMYGMIASS	NMFVHNNSKH	GRRTRSRADG
AmCOE	NPDFRRVLLT	---HELICSR	CLERRSCGNK	NDTPSDPIIL	D--NFRCLKIF	AKCNQNCLKN	AGNPKDSRRR	FQVALYEIDK	VSGQPIASS	SMFVHNNSKH	GRRPIYR-DN

Ig-like/Plexin/Transcription Factor Domain

GNATHOSTOMES

COE1											
MmCOE1	SEGT-PSYLE	H-----	-----	-ATPCIKAIS	PSEGWTGGA	TVIIIGDNFF	DGLQVIFGTM	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
HsCOE1	SEGT-PSYLE	H-----	-----	-ATPCIKAIS	PSEGWTGGA	TVIIIGDNFF	DGLQVIFGTM	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
MdCOE1	SEGT-PSYLE	H-----	-----	-ATPCIKAIS	PSEGWTGGA	TVIIIGDNFF	DGLQVIFGTM	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
OaCOE1	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXXL	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
GgCOE1	SEGT-PSYLE	H-----	-----	-ATPCIKAIS	PSEGWTGGA	TVIIIGDNFF	DGLQVIFGTM	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
XtCOE1	SDGT-PSYLE	HEA-----	-----	-ATPCIKAIS	PSEGWTGGA	TVIVIGDNFF	DGLQVIFGTM	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
DrCOE1	SEGT-PSYLE	H-----	-----	-ATPCIKAIS	PSEGWTGGA	TVIIIGDNFF	DGLQVIFGTM	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
COE2											
MmCOE2	SE-----	-----	-----	-ATPCIKAIS	PSEGWTGGA	MVIIIGDNFF	DGLQVVFSTM	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
HsCOE2	SE-----	-----	-----	-ATPCIKAIS	PSEGWTGGA	MVIIIGDNFF	DGLQVVFSTM	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
OaCOE2	SE-----	-----	-----	-ATPCIKAIS	PSEGWTGGA	MVIIIGDNFF	DGLQVVFSTM	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
GgCOE2	SE-----	-----	-----	-ATPCIKAIS	PSEGWTGGA	MVIIIGDNFF	DGLQVVFSTM	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
XtCOE2	SE-----	-----	-----	-ATPCIKAIS	PSEGWTGGA	MVIIIGDNFF	DGLQVVFSTM	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
DrCOE2	NETV-ENNME	Y-----	-----	-ATPCIKAIS	PSEGWTGGA	MVIVIGENFF	DGLQVVFSTM	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
COE3											
MmCOE3	SE-----	-----	-----	-ATPCIKAIS	PSEGWTGGA	TVIIIGDNFF	DGLQVVFSTM	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
HsCOE3	SEGTAPSYLE	N-----	-----	-ATPCIKAIS	PSEGWTGGA	TVIIIGDNFF	DGLQVVFSTM	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
OaCOE3	SEXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
GgCOE3	SEGTAPSYLE	NA-----	-----	-ATPCIKAIS	PSEGWTGGA	TVIIIGDNFF	DGLQVVFSTM	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
XtCOE3	SEGTAPSYLE	NV-----	-----	-ATPCIKAIS	PSEGWTGGA	TVIIIGDNFF	DGLQVVFSTM	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
DrCOE3	SEGTATPYLE	NA-----	-----	-ATPCIKAIS	PSEGWTGGA	TVIIIGDNFF	DGLQVVFSTM	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
EBF4											
MmEBF4	SEA-----	-----	-----	-ATPCIKAIS	PSEGWTGGA	TVIIIGDNFF	DGLQVVFSTN	L-LWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
HsEBF4	SEA-----	-----	-----	-ATPCIKAIS	PSEGWTGGA	TVIVIGDNFF	DGLQVVFSTN	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
MdEBF4	SEA-----	-----	-----	-ATPCIKAIS	PSEGWTGGA	TVIVIGDNFF	DGLQVVFSTM	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
OaEBF4	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
UNASSIGNED											
DrCOE	IEA-----	-----	-----	-EIPCIAIS	PSEGWTSGGA	SVIIIGDQFF	SGLQVVFSTM	L-VWSE---V	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--

LAMPREYS

Pm COE-A	SE-----	-----	-----	-ATPCIKAIS	PSEGWTGGA	TVIIIGDNFF	DGLQVVFSTM	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
Pm COE-B	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX

NON-VERTEBRATES

CiCOE	SEGNFFFIES	-----	-----	-ASPTIKAIN	PAEGWTGGA	TVIVIGENFF	DGLQVVFSTM	V-VWSE---L	ITQHAIRVQT	PPRHLPGVVE	VTLSYKSK--
BfCOE	SEDGEMSF	DSPNINRGQK	-----	-ATPCIKAIS	PSEGWTGGA	TVIIIGDNFF	DGLQVVFSTM	L-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
SpCOE	SEG-----	-----	-----	-ATPCIKAIS	PSEGWTGGA	TVIIIGDNFF	DGLQVVFSTM	I-VWSE---L	ITPHAIRVQT	PPRHIPGVVE	VTLSYKSK--
LgCOE	TEGGDLRFSH	LRPKLSPFAA	-----	-TPCIIKAIC	PSEGWTSGGT	TVIIIGDNFF	DGLQLVFGTM	L-VWSE---L	LTSHAIRVQT	PPRHIPGVVE	VTLSYKSK--
CcCOE	GDAGFLCA--	-----	-----	-ATPTIKAIC	PSEGWTGGT	TVIIIGDNFF	DGLQVVFSTM	L-VWSE---L	ITSHAIRVQT	PPRHIPGVVE	VTLSYKSK--
DmCOE	TEGTGNTSLS	ISGHPLAPDS	TYDGLYPPLP	VATPCIKAIS	PSEGWTGGA	TVIIIGDNFF	DGLQVVFSTM	L-VWSE---L	ITSHAIRVQT	PPRHIPGVVE	VTLSYKSK--
CeCOE	SDDSEYSESA	ELPS-----	-----	-SVPVIKALF	PSEGWIQGGT	QVVLIGENFF	EGLQVAFGTA	SPNWGESVQL	ISPHAIRVT	PPKHSAGPVD	VTLQYKSK--
NvCOE	NEQDTSA---	-----	-----	-DDPCIIKAIC	PNEGWTIGGS	NVILIGDNFF	DGLQVVFSTF	I-VWSE---F	ITPHALRVQA	PPSPMPGVVK	VYLMHKEK--
AmCOE	VVGD-----	-----	-----	-GKPCIVSIY	PNEGWTGGG	RITVIGVNF	EGLDIVFGTV	P-VQSE---V	LSPNAIAVRT	PPATMTGEVD	VTLIFRASGA

	HELIX H1	Linker	HELIX H2d	HELIX H2A
GNATHOSTOMES				
COE1				
MmCOE1	QFCKGTPGRF IYT-ALNEPT IDYGFQRLQK VIPRHPGDPE RLPKEVILKR AADLVEALYG MPHNNQ--EI			ILKRAADIAE ALYSVPRNHN
HsCOE1	QFCKGTPGRF IYT-ALNEPT IDYGFQRLQK VIPRHPGDPE RLPKEVILKR AADLVEALYG MPHNNQ--EI			ILKRAADIAE ALYSVPRNHN
MdCOE1	QFCKGTPGRF IYT-ALNEPT IDYGFQRLQK VIPRHPGDPE RLPKEVILKR AADLVEALYG MPHNNQ--EI			ILKRAADIAE ALYSVPRNHN
OaCOE1	QFCKGTPGRF IYT-ALNEPT IDYGFQRLQK VIPRHPGDPE RLPKEVILKR AADLVEALYG MPHNNQ--EI			ILKRAADIAE ALYSVPRNHN
GgCOE1	QFCKGTPGRF IYT-ALNEPT IDYGFQRLQK VIPRHPGDPE RLPKEVILKR AADLVEALYG MPHNNQ--EI			ILKRAADIAE ALYSVPRNHN
XtCOE1	QFCKGTPGRF IYT-ALNEPT IDYGFQRLQK VIPRHPGDPE RLPKEVILKR AADLVEALYG MPHNNQ--EI			ILKRAADIAE ALYSVPRNHS
DrCOE1	QFCKGTPGRF IYT-ALNEPT IDYGFQRLQK VIPRHPGDPE RLPKEVILKR AADLVEALYG MPHNNQ--EI			ILKRAADIAE ALYNVPRGHN
COE2				
MmCOE2	QFCKGAPGRF IYT-ALNEPT IDYGFQRLQK VIPRHPGDPE RLAKEMLLKR AADLVEALYG TPHNNQ--DI			ILKRAADIAE ALYSVPRNPS
HsCOE2	QFCKGAPGRF IYT-ALNEPT IDYGFQRLQK VIPRHPGDPE RLAKEMLLKR AADLVEALYG TPHNNQ--DI			ILKRAADIAE ALYSVPRNPS
OaCOE2	QFCKGAPGRF IYT--LNEPT IDYGFQRLQK VIPRHPGDPE RLAKEMLLKR AADLVEALYG TPHNNQ--DI			ILKRAADIAE ALYSVPRNPS
GgCOE2	QFCKGAPGRF IYT-ALNEPT IDYGFQRLQK VIPRHPGDPE RLAKEMLLKR AADLVEALYG TPHNNQ--DI			ILKRAADIAE ALYSVPRNPA
XtCOE2	QFCKGAPGRF IYT-ALNEPT IDYGFQRLQK VIPRHPGDPE RLAKEMLLKR AADLVEALYG TPHNNQ--DI			ILKRAADIAE ALYSVPRNHN
DrCOE2	QFCKGAPGRF IYT-ALNEPT IDYGFQRLQK LIPRHPGDPD KLAKEMLLKR AADVVESLYG NTTSNQ--DM			LLKRAADIAE ALYSVPRPHS
COE3				
MmCOE3	QFCKGAPGRF VYT-ALNEPT IDYGFQRLQK VIPRHPGDPE RLPKEVLLKR AADLVEALYG MPHNNQ--EI			ILKRAADIAE ALYSVPRNHN
HsCOE3	QFCKGAPGRF VYT-ALNEPT IDYGFQRLQK VIPRHPGDPE RLPKEVLLKR AADLVEALYG MPHNNQ--EI			ILKRAADIAE ALYSVPRNHN
OaCOE3	XXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX			XXXXXXXXXX XXXXXXXXXXXX
GgCOE3	QFCKGAPGRF VYT-ALNEPT IDYGFQRLQK VIPRHPGDPE RLPKEVLLKR AADLVEALYG MPHNNQ--EI			ILKRAADIAE ALYSVPRNHN
XtCOE3	QFCKGAPGRF VYT-ALNEPT IDYGFQRLQK VIPRHPGDPE RLPKEVLLKR AADLVEALYG MPHNNQAKEI			ILKRAADIAE ALYSVPRNHN
DrCOE3	QFCKGAPGRF VYT-ALNEPT IDYGFQRLQK VIPRHPGDPE RLPKEVLLKR AADLVEALYG MPHNNQ--EI			ILKRAADIAE ALYSVPRNHN
EBF4				
MmEBF4	QFCKGAPGRF VYT-ALNEPT IDYGFQRLQK VIPRHPGDPE RLPKEVLLKR AADLAEALYG VPSSNQ--EL			LLKRAADVAE ALYSAPRAPA
HsEBF4	QFCKGCPGRF VYT-ALNEPT IDYGFQRLQK VIPRHPGDPE RLPKEVLLKR AADLAEALYG VPGSNQ--EL			LLKRAADVAE ALYSTPRAPG
MdEBF4	QFCKGAPGRF VYT-ALNEPT IDYGFQRLQK VIPRHPGDPE RLPKEVLLKR AADLVEALYG MPHSNQ--DL			ILKRAADIAE ALYSVPRTPS
OaEBF4	XXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX			XXXXXXXXXX XXXXXXXXXXXX
UNASSIGNED				
DrCOE	QFCRGAPGRF VYT-ALNEPT IDYGFQRLQK VIPHHPGDVE RLPKEVLLKR AADLIETFYG APQNNQ--EI			ILKRASDIAE ALNSIPRNPS
LAMPREYS				
Pm COE-A	QFCKGAPGRF VYT-ALNEPT IDYGFQRLQK LIPRHPGDPE RLPK-----		-----Q--EI	ILKRAADIAE ALYSVPRNHN
Pm COE-B	XXXXXXXXXX XXXXALNEPT IDYGFQRLQK VIPRHPGDPE RLPKEVILKR AADLVEALYG LPHYNQ--EI			ILKRAADIAE ALYSMRPHQN
NON-VERTEBRATES				
CiCOE	QFCSGAPGRF VYT-ALNEPT LDYGFQRLK TVPRHPGDPE RLPK-----		-----EI	ILKRAADVME AVISRQYAPP
BfCOE	QFCKGAPGRF VYT-SLNEPT IDYGFQRLAK LVPRHPGDPE RLPK-----		-----EI	ILKRAADLAE AIYSMPRNPN
SpCOE	QFCKGAPGRF VYV-SLTEPT IDYGFQRLSK LVPRHPGDPD RLPK-----		-----EI	ILKRAADLAE ALYSMRPN--
LgCOE	QFCKGAPGRF VYT-ALTEPT IDYGFQRLMK LVPRHPGDPE KLPK-----		-----EI	ILKRAADLAE ALYSMRPNHN
CcCOE	QFCKGAPGRF VYT-ALTEPT IDYGFQRLMK LVPRHPGDPE RLPK-GICEG HVGPLSSQIF L-SCPQ--EI			ILKRAADLAE ALYSMRPNPN
DmCOE	QFCKGSPGRF VYVSALNEPT IDYGFQRLQK LIPRHPGDPE KLPK-----		-----EI	ILKRAADLVE ALYSMRSPG
CeCOE	TYSRGTPLRF SYI-TLAEPG IEYGFQRLQK LLPKYPGDPE RLPK-----		-----DQ	ILKRAAEALAE ALYNRTSTES
NvCOE	QYCKYAPAKF GYT-ALVEPT IDYGFQRLSK LIPRHPGDPE RIPK-----		-----EI	VLKRAADLAE TLYQMPRTPT
AmCOE	QFCISNPGKF LYT-APDDQI FENSFCRIER VI-RQQDDPD QLPK-----		-----DL	VLQRAAEELLE SCFLSSRGA-

Figure S2 : Alignment of the DBD, IPT and HLH domains of the COE proteins. Species are indicated on the top lines with abbreviations as in Fig. 1. Additional species abbreviations : Oa, *Ornithorhynchus anatinus*; Gg, *Gallus gallus*; Sp, *Strongylocentrotus purpuratus*; Ce, *Caenorhabditis elegans*; Am, *Amphimedon queenslandica*. Highly truncated (e.g., XtCOE1, PmCOE-B, OaCOE1, OaCOE3, OaEBF4) or divergent (CeCOE, AmCOE) predicted sequences were included in this alignment. Dashes indicate missing residues. X stands for undetermined sequences. The positions correspond to the DBD, IPT, Helix 1, H2d and H2a domains are shaded in orange, yellow, green, pale blue and dark blue, respectively. The linker separating H1 and H2 is squared.