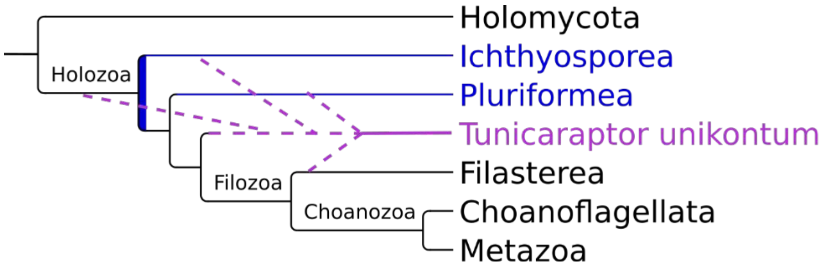
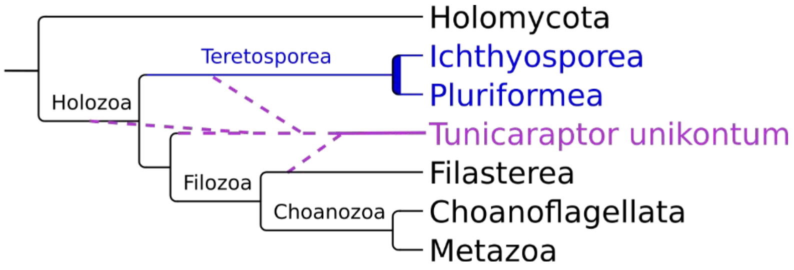
Evolution of CCA-adding Enzymes at the Origin of Metazoa

Elisa Israel
Computational EvoDevo, Leipzig University
Bled, 14.02.2023

Phylogeny

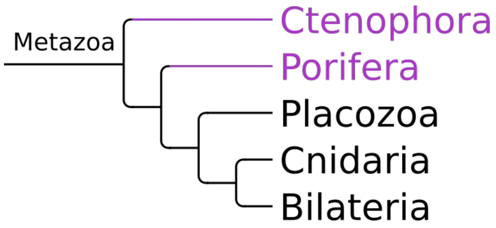
- Holozoa: Metazoa and their unicellular relatives
- holozoan phylogeny still discussed: Pluriformea as sister group to Filozoa vs. ‘Teretosporea’ hypotheses
- the unicellular species *T. unikontum* is yet to be placed in the phylogenetic tree

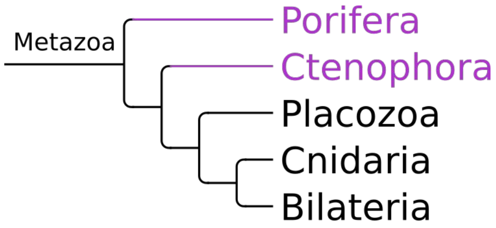




Phylogeny

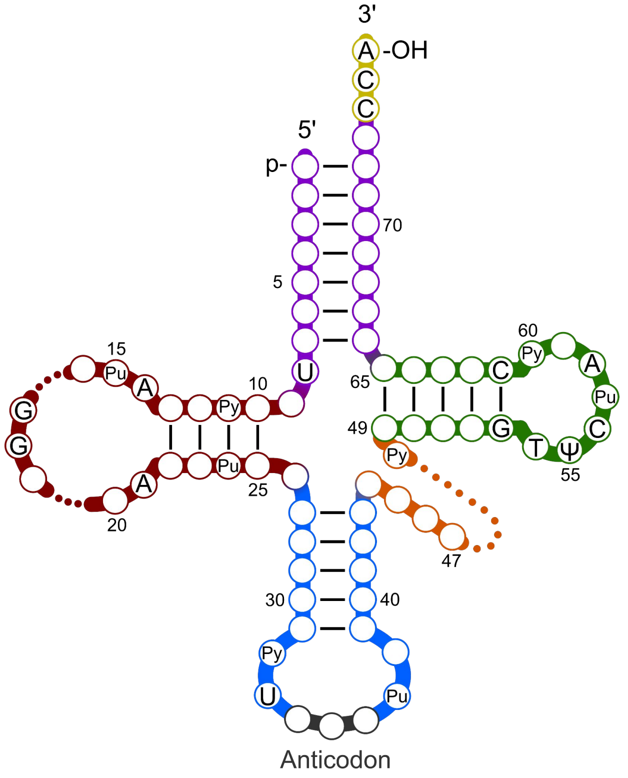
- phylogeny of early-branching Metazoa is strongly debated
- ‘Ctenophora first’ vs. ‘Porifera first’ hypotheses (comb jellies vs. sponges)
- phylogeny of unicellular Holozoa and early-branching Metazoa is based on a small set of genomes and transcriptomes
- the fast evolutionary rates of ctenophores further complicates things
- ambiguities also exist within the clades





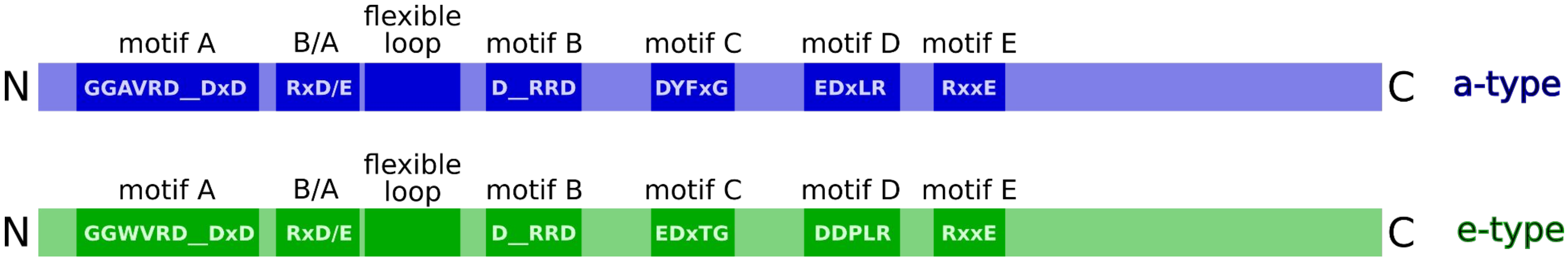
CCA-adding Enzymes

- CCA sequence is usually added post-transcriptionally by CCA-adding enzymes or CC-/A-adding enzymes
- class I: archaeal enzymes
- class II: bacterial and eukaryotic enzymes
- class II enzymes are characterized by five highly conserved catalytic core motifs



CCA-adding Enzymes

- only a few paralogs identified in eukaryotes (Choanoflagellata & Schizosaccharomyces)
- eukaryotic CCA: eCCA
- alphaproteobacterial CCA: aCCA
- differ in their conserved motifs A, C and D



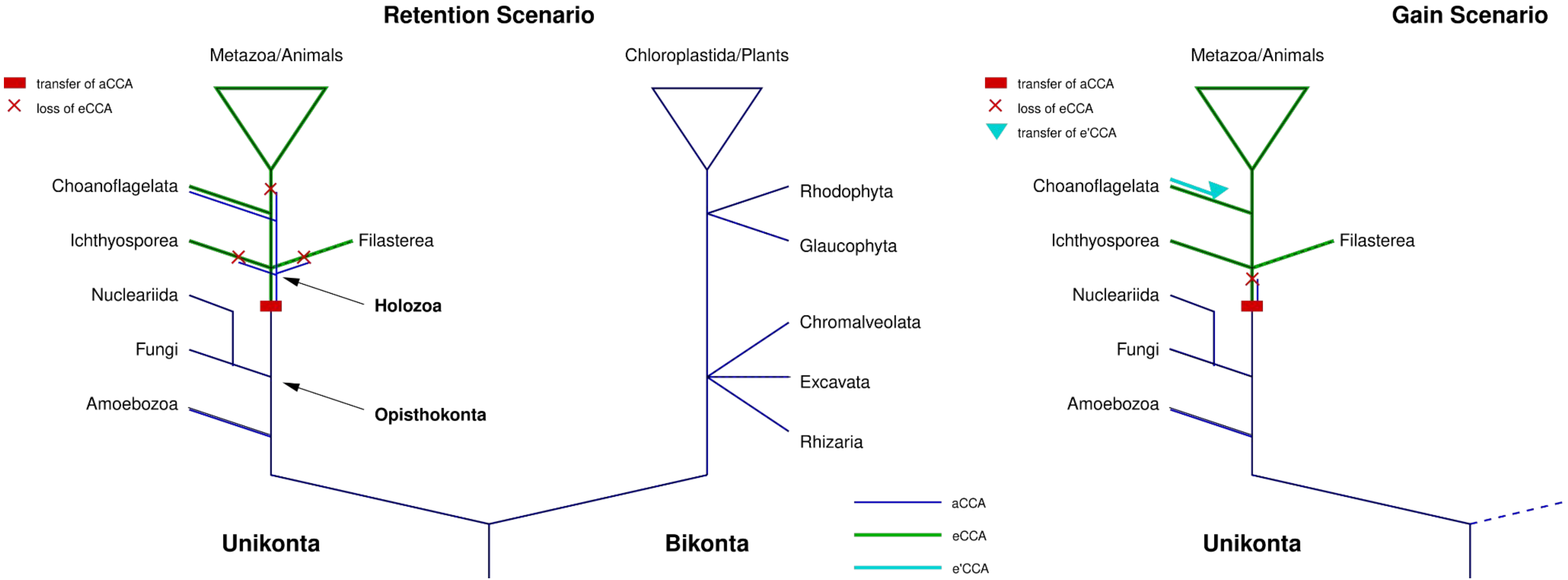
	motif A	B/A	flexible loop	motif B	motif C	motif D	motif E	
N	GGAVRD_DxD	RxD/E		D_RRD	DYFxG	EDxLR	RxxE	C a-type
N	GGWVRD_DxD	RxD/E		D_RRD	EDxTG	DDPLR	RxxE	C e-type

CCA-adding Enzymes

- Holozoa: aCCA through HGT, eCCA lost in most lineages
- Choanoflagellata: all have aCCA, some eCCA additionally
- subfunctionalization shown for *S. rosetta*
    - aCCA: CC-addition, mutations in the flexible loop
    - eCCA: A-addition

Evolutionary Events in Holozoa

- Retention vs. Gain Scenario to explain paralogs in Choanoflagellata

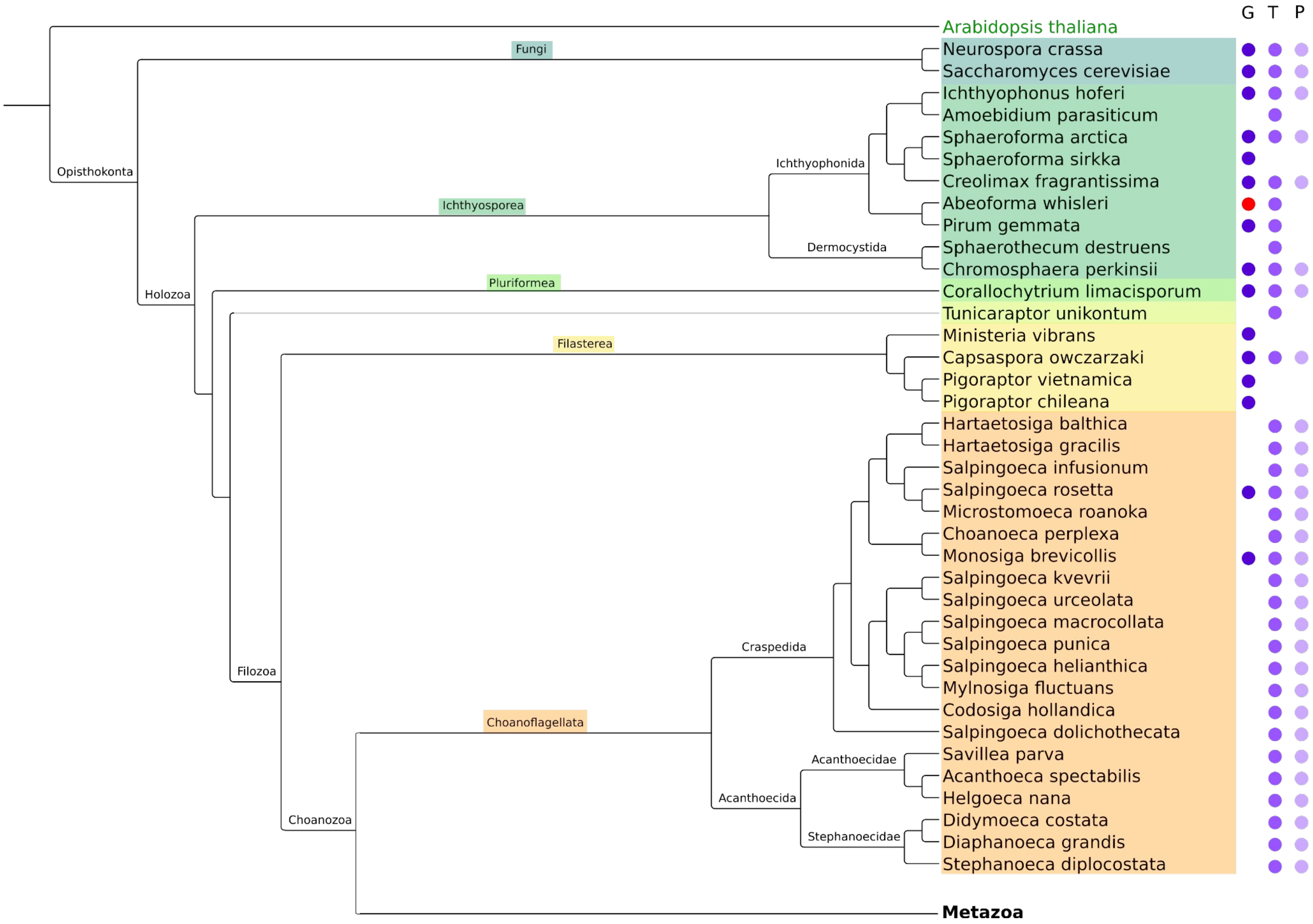


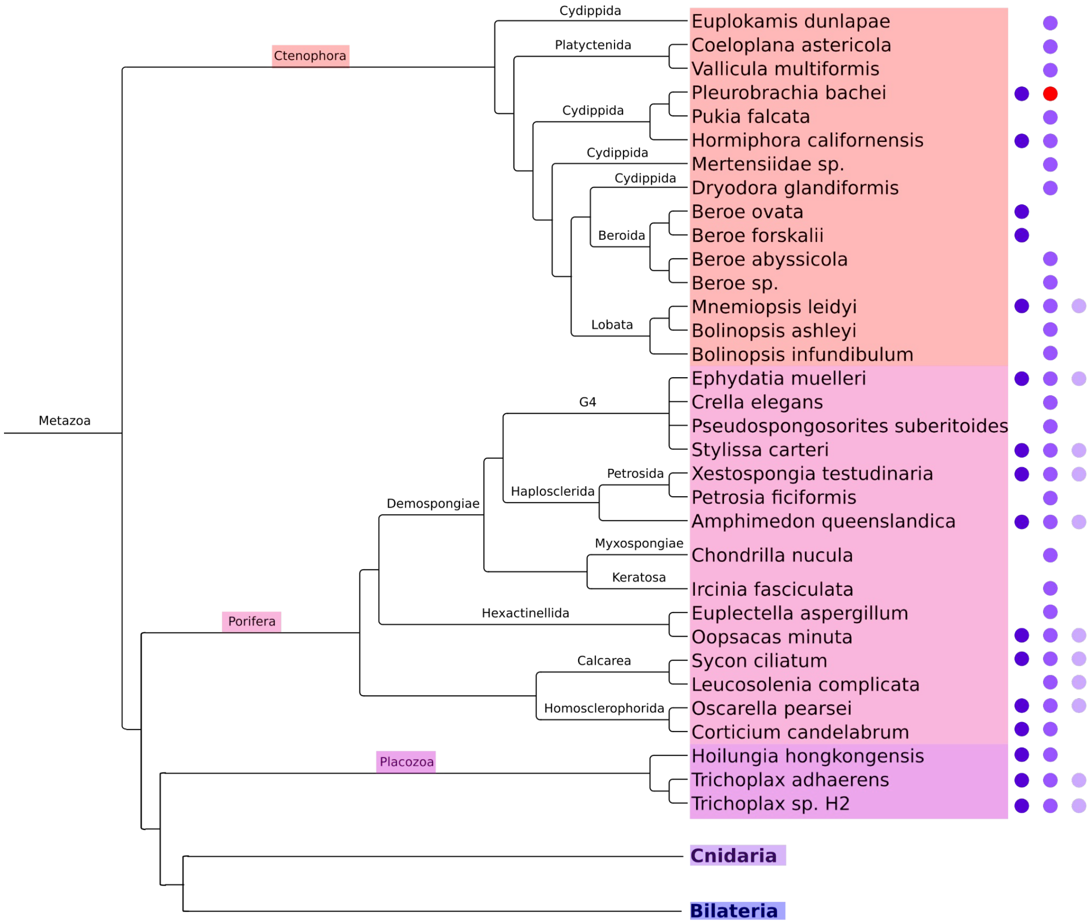
Is the occurrence of paralogs limited to our closest unicellular relatives?

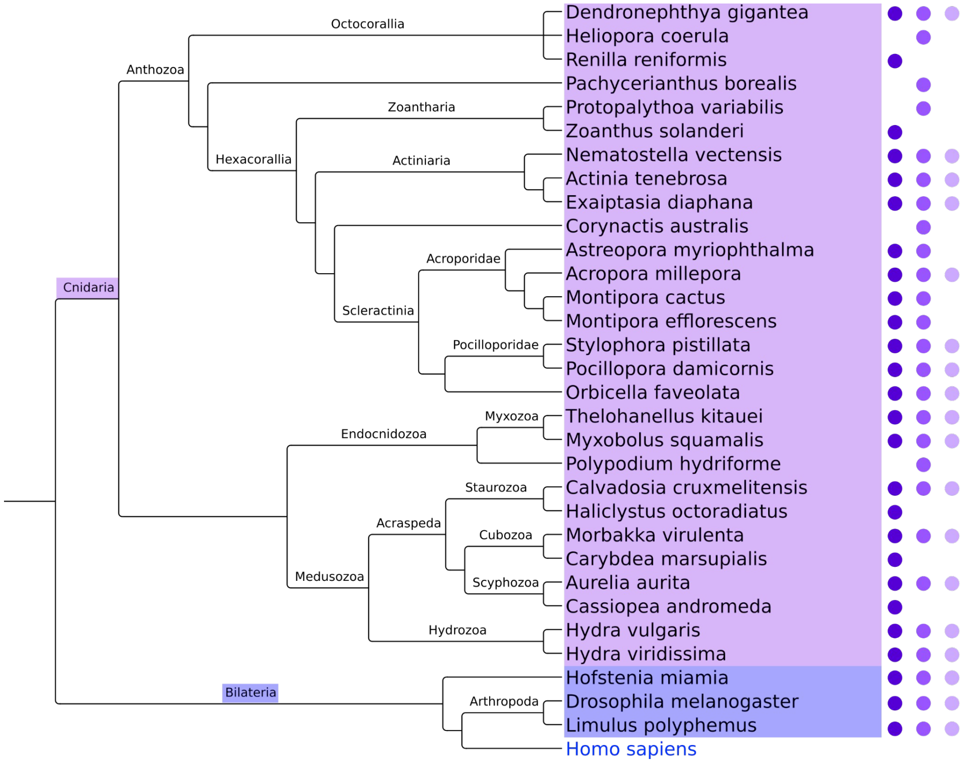
- ➢ expectation: all Metazoa possess one aCCA enzyme

Data

- proteome, transcriptome and genome data of 102 holozoan species
  - 58 genomes
  - 91 transcriptomes
  - 59 proteomes
- reference protein sequences:
  aCCA: *Homo sapiens*
  eCCA: *Arabidopsis thaliana*


G
T
P
Arabidopsis thaliana
Fungi
Neurospora crassa
Saccharomyces cerevisiae
Opisthokonta
Ichthyosporea
Ichthyophonida
Dermocystida
Ichthyophonus hoferi
Amoebidium parasiticum
Sphaeroforma arctica
Sphaeroforma sirkka
Creolimax fragrantissima
Abeoforma whisleri
Pirum gemmata
Sphaerothecum destruens
Chromosphaera perkinsii
Holozoa
Pluriformea
Corallochytrium limacisporum
Tunicaraptor unikontum
Filasterea
Ministeria vibrans
Capsaspora owczarzaki
Pigoraptor vietnamica
Pigoraptor chileana
Filozoa
Choanoflagellata
Craspedida
Hartaetosiga balthica
Hartaetosiga gracilis
Salpingoeca infusionum
Salpingoeca rosetta
Microstomoeca roanoka
Choanoeca perplexa
Monosiga brevicollis
Salpingoeca kvevrii
Salpingoeca urceolata
Salpingoeca macrocollata
Salpingoeca punica
Salpingoeca helianthica
Mylnosiga fluctuans
Codosiga hollandica
Salpingoeca dolichothecata
Acanthoecida
Acanthoecidae
Savillea parva
Acanthoeca spectabilis
Helgoeca nana
Stephanoecidae
Didymoeca costata
Diaphanoeca grandis
Stephanoeca diplocostata
Choanozoa
Metazoa


Metazoa
Ctenophora
Cydippida
Platyctenida
Cydippida
Cydippida
Cydippida
Beroida
Lobata
Euplokamis dunlapae
Coeloplana astericola
Vallicula multiformis
Pleurobrachia bachei
Pukia falcata
Hormiphora californensis
Mertensiidae sp.
Dryodora glandiformis
Beroe ovata
Beroe forskalii
Beroe abyssicola
Beroe sp.
Mnemiopsis leidyi
Bolinopsis ashleyi
Bolinopsis infundibulum
Porifera
Demospongiae
G4
Petrosida
Haplosclerida
Myxospongiae
Keratosa
Hexactinellida
Calcarea
Homoscleromorphida
Ephydatia muelleri
Crella elegans
Pseudospongosorites suberitoides
Stylissa carteri
Xestospongia testudinaria
Petrosia ficiformis
Amphimedon queenslandica
Chondrilla nucula
Ircinia fasciculata
Euplectella aspergillum
Oopsacas minuta
Sycon ciliatum
Leucosolenia complicata
Oscarella pearsei
Corticium candelabrum
Placozoa
Hoilungia hongkongensis
Trichoplax adhaerens
Trichoplax sp. H2
Cnidaria
Bilateria

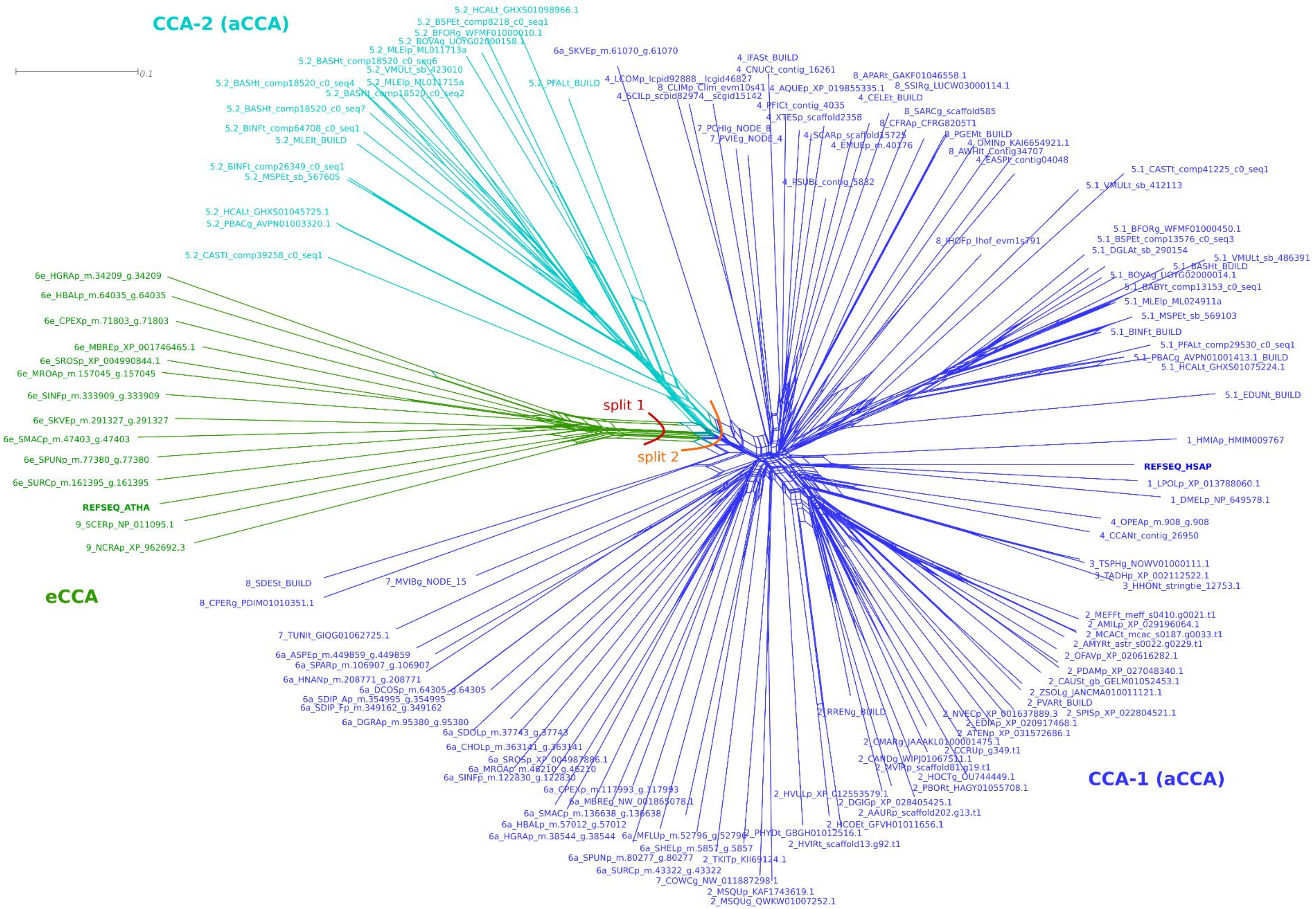

Octocorallia
Anthozoa
Zoantharia
Hexacorallia
Actiniaria
Acroporidae
Cnidaria
Scleractinia
Pocilloporidae
Myxozoa
Endocnidozoa
Staurozoa
Acraspeda
Cubozoa
Medusozoa
Scyphozoa
Hydrozoa
Bilateria
Arthropoda
Dendronephthya gigantea
Heliopora coerula
Renilla reniformis
Pachycerianthus borealis
Protopalythoa variabilis
Zoanthus solanderi
Nematostella vectensis
Actinia tenebrosa
Exaiptasia diaphana
Corynactis australis
Astreopora myriophthalma
Acropora millepora
Montipora cactus
Montipora efflorescens
Stylophora pistillata
Pocillopora damicornis
Orbicella faveolata
Thelohanellus kitauei
Myxobolus squamalis
Polypodium hydriforme
Calvadosia cruxmelitensis
Haliclystus octoradiatus
Morbakka virulenta
Carybdea marsupialis
Aurelia aurita
Cassiopea andromeda
Hydra vulgaris
Hydra viridissima
Hofstenia miamia
Drosophila melanogaster
Limulus polyphemus
Homo sapiens

Workflow

- identify candidate sequences
- transcriptome: translate candidates
- genome: splice site prediction (unicellular Holozoa) or gene prediction using clade specific parameters (early-branching Metazoa)
- check for contaminations
- iterate using clade specific protein sequence if only partial sequences/no candidates found
- align candidates, generate phylogenetic network
- classify candidates using sequence similarity, common motifs and the phylogenetic network

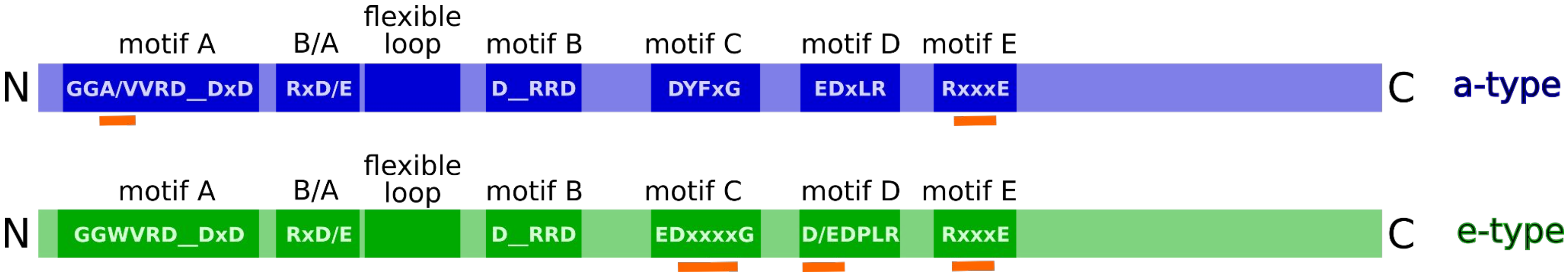
Expectations

- one aCCA enzyme in all holozoan species
- additional eCCA sequence in Choanoflagellata
- clear separation of eCCA sequences from Choanoflagellata and aCCA sequences from all holozoan lineages


CCA-2 (aCCA)
0.1
5.2_HCALt_GHXS01098966.1
5.2_BSPEt_comp8218_c0_seq1
5.2_BFORg_WFMF01000010.1
5.2_BOVAg_UOYG02000158.1
5.2_MLEIp_ML011713a
5.2_BASHt_comp18520_c0_seq6
5.2_VMULt_sb_423010
5.2_BASHt_comp18520_c0_seq4
5.2_MLEIp_ML011715a
5.2_BASHt_comp18520_c0_seq2
5.2_PFALt_BUILD
5.2_BASHt_comp18520_c0_seq7
5.2_BINFt_comp64708_c0_seq1
5.2_MLEIt_BUILD
5.2_BINFt_comp26349_c0_seq1
5.2_MSPEt_sb_567605
5.2_HCALt_GHXS01045725.1
5.2_PBACg_AVPN01003320.1
5.2_CASTt_comp39258_c0_seq1
6a_SKVEp_m.61070_g.61070
4_IFASt_BUILD
4_CNUCt_contig_16261
4_LCOMp_lcpid92888__lcgid46827
8_CLIMp_Clim_evm10s41
4_AQUEp_XP_019855335.1
4_SCILp_scpid82974__scgid15142
8_APARt_GAKF01046558.1
8_SSIRg_LUCW03000114.1
4_CELEt_BUILD
4_PFICt_contig_4035
4_XTESp_scaffold2358
8_SARCg_scaffold585
8_CFRAp_CFRG8205T1
7_PCHIg_NODE_8
7_PVIEg_NODE_4
4_SCARp_scaffold15725
4_EMUEp_m.40176
8_PGEMt_BUILD
4_OMINp_KAI6654921.1
8_AWHIt_Contig34707
4_EASPt_contig04048
4_PSUBt_contig_5832
5.1_CASTt_comp41225_c0_seq1
5.1_VMULt_sb_412113
8_IHOFp_Ihof_evm1s791
5.1_BFORg_WFMF01000450.1
5.1_BSPEt_comp13576_c0_seq3
5.1_DGLAt_sb_290154
5.1_VMULt_sb_486391
5.1_BASHt_BUILD
5.1_BOVAg_UOYG02000014.1
5.1_BABYt_comp13153_c0_seq1
5.1_MLEIp_ML024911a
5.1_MSPEt_sb_569103
5.1_BINFt_BUILD
5.1_PFALt_comp29530_c0_seq1
5.1_PBACg_AVPN01001413.1_BUILD
5.1_HCALt_GHXS01075224.1
5.1_EDUNt_BUILD
1_HMIAp_HMIM009767
REFSEQ_HSAP
1_LPOLp_XP_013788060.1
1_DMELp_NP_649578.1
4_OPEAp_m.908_g.908
4_CCANt_contig_26950
3_TSPHg_NOWV01000111.1
3_TADHp_XP_002112522.1
3_HHONt_stringtie_12753.1
2_MEFFt_meff_s0410.g0021.t1
2_AMILp_XP_029196064.1
2_MCACt_mcac_s0187.g0033.t1
2_AMYRt_astr_s0022.g0229.t1
2_OFAVp_XP_020616282.1
2_PDAMp_XP_027048340.1
2_CAUSt_gb_GELM01052453.1
2_ZSOLg_JANCMA010011121.1
2_PVARt_BUILD
2_NVECp_XP_001637889.3
2_SPISp_XP_022804521.1
2_EDIAp_XP_020917468.1
2_ATENp_XP_031572686.1
2_RRENg_BUILD
2_CMARg_JAAAKL0100001475.1
2_CCRUp_g349.t1
2_CANDg_WIPJ01067511.1
2_MVIRp_scaffold81.g19.t1
2_HOCTg_OU744449.1
2_PBORt_HAGY01055708.1
CCA-1 (aCCA)
2_HVULp_XP_012553579.1
2_DGIGp_XP_028405425.1
2_AAURp_scaffold202.g13.t1
2_HCOEt_GFVH01011656.1
2_PHYDt_GBGH01012516.1
2_HVIRt_scaffold13.g92.t1
2_TKITp_KII69124.1
7_COWCg_NW_011887298.1
2_MSQUp_KAF1743619.1
2_MSQUg_QWKW01007252.1
6a_MFLUp_m.52796_g.52796
6a_SHELp_m.5857_g.5857
6a_SPUNp_m.80277_g.80277
6a_SURCp_m.43322_g.43322
6a_HGRAp_m.38544_g.38544
6a_HBALp_m.57012_g.57012
6a_SMACp_m.136638_g.136638
6a_MBREg_NW_001865078.1
6a_CPEXp_m.117993_g.117993
6a_SINFp_m.122830_g.122830
6a_MROAp_m.46210_g.46210
6a_SROSp_XP_004987886.1
6a_CHOLp_m.363141_g.363141
6a_SDOLp_m.37743_g.37743
6a_DGRAp_m.95380_g.95380
6a_SDIP_Fp_m.349162_g.349162
6a_SDIP_Ap_m.354995_g.354995
6a_DCOSp_m.64305_g.64305
6a_HNANp_m.208771_g.208771
6a_SPARp_m.106907_g.106907
6a_ASPEp_m.449859_g.449859
7_TUNIt_GIQG01062725.1
8_CPERg_PDIM01010351.1
8_SDESt_BUILD
7_MVIBg_NODE_15
eCCA
9_NCRAp_XP_962692.3
9_SCERp_NP_011095.1
REFSEQ_ATHA
6e_SURCp_m.161395_g.161395
6e_SPUNp_m.77380_g.77380
6e_SMACp_m.47403_g.47403
6e_SKVEp_m.291327_g.291327
6e_SINFp_m.333909_g.333909
6e_MROAp_m.157045_g.157045
6e_SROSp_XP_004990844.1
6e_MBREp_XP_001746465.1
6e_CPEXp_m.71803_g.71803
6e_HBALp_m.64035_g.64035
6e_HGRAp_m.34209_g.34209
split 1
split 2

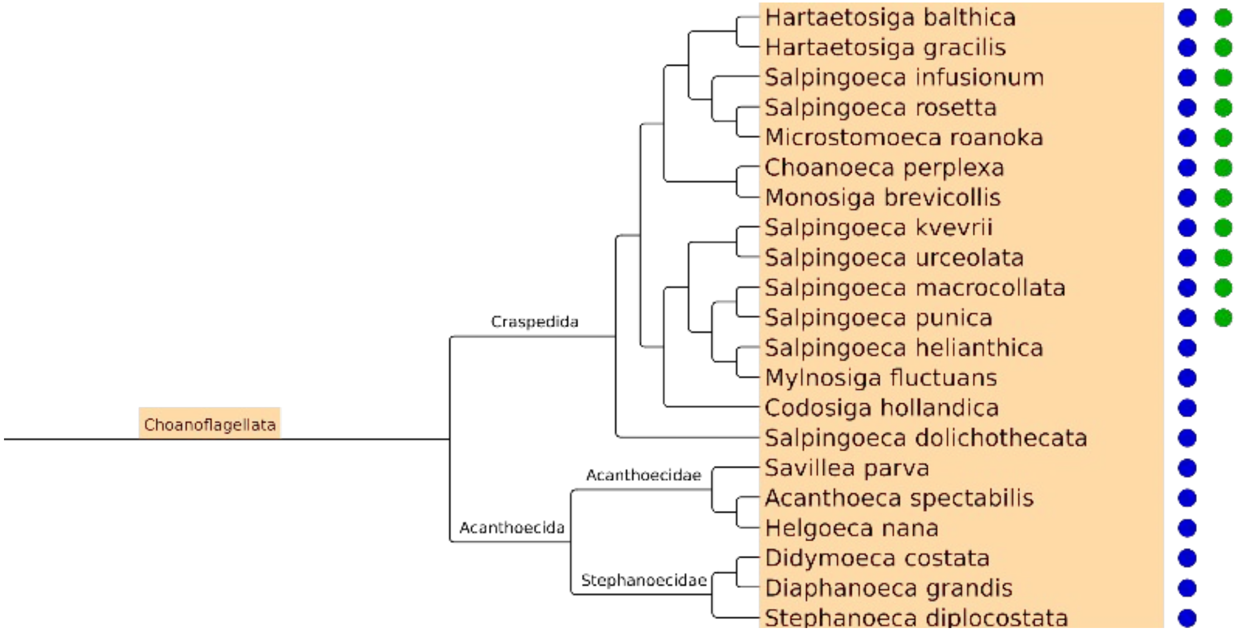
Motifs

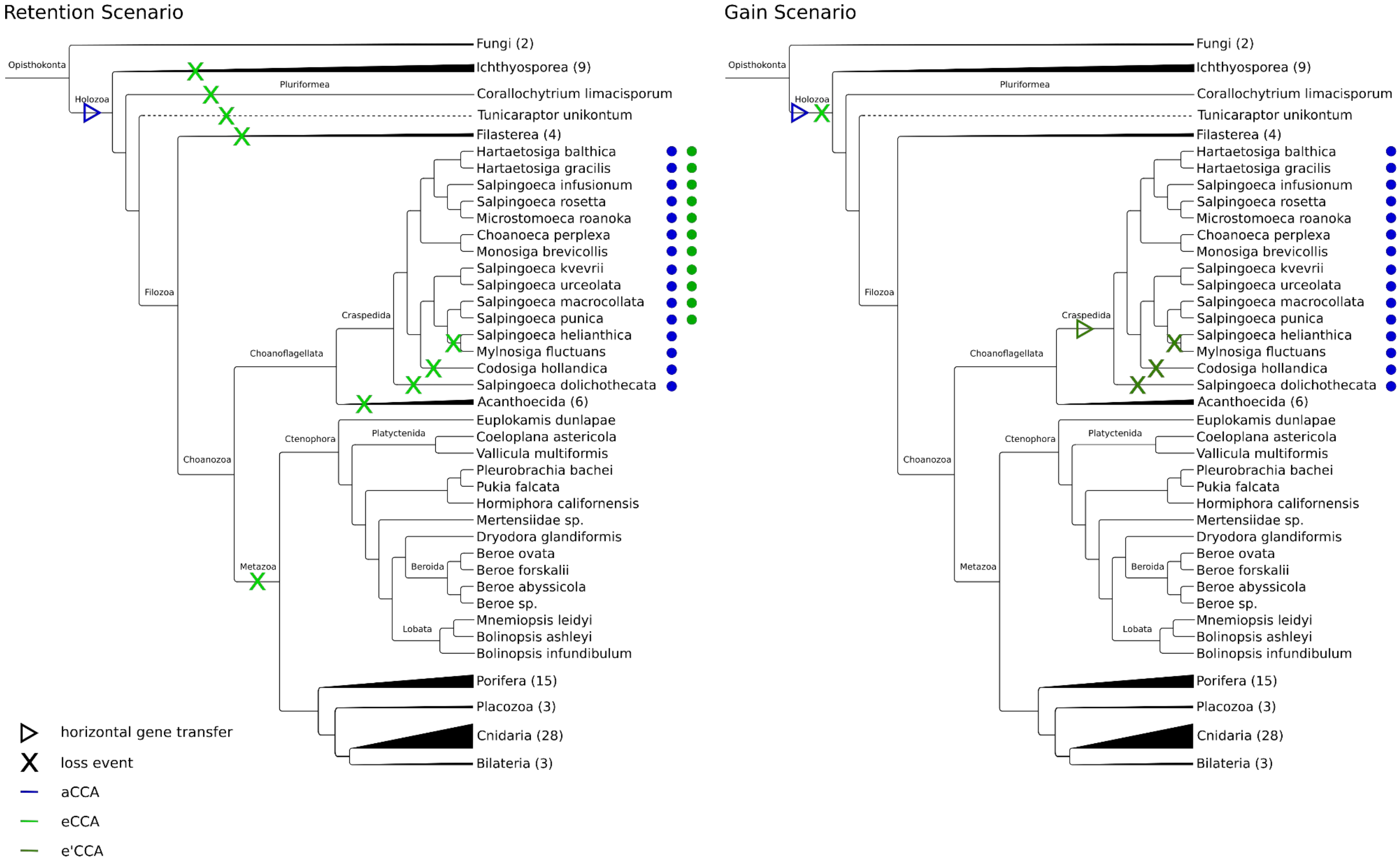
- motifs A, C, D and E need revision
- strict pattern search failed to classify eCCA sequences



Distribution of CCA-adding Enzymes

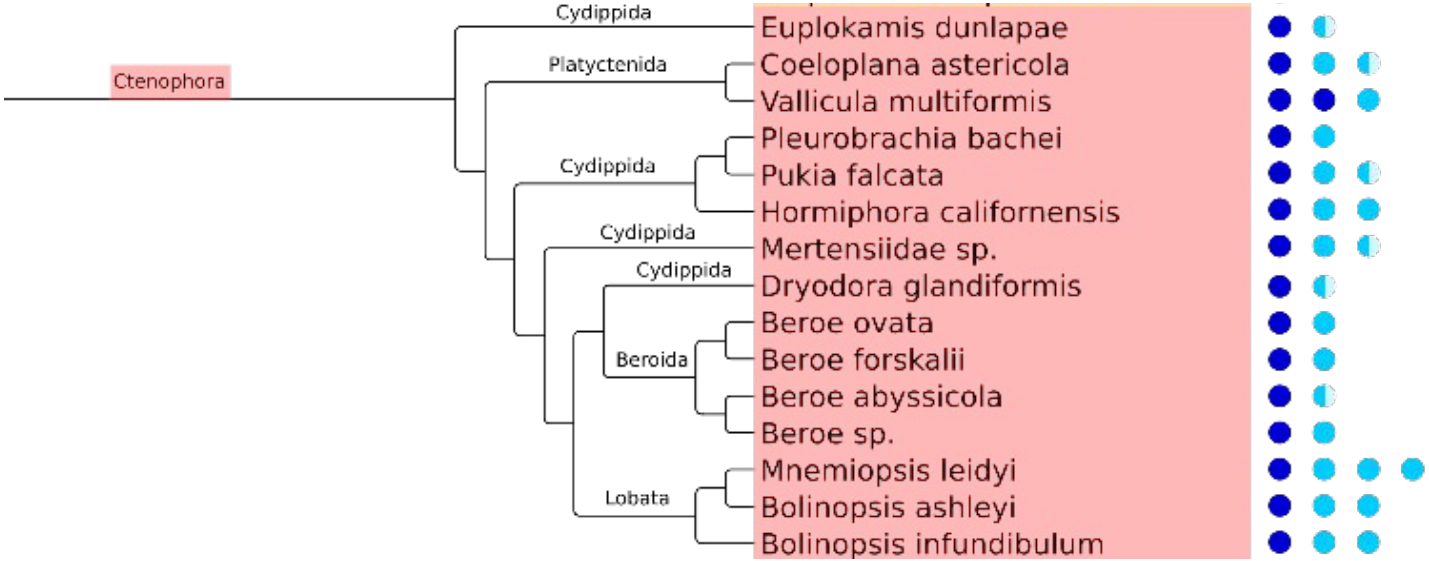
- ➢ eCCA in Craspedida clade 1 and clade 2




Retention Scenario
Opisthokonta
Holozoa
Pluriformea
Filozoa
Choanozoa
Choanoflagellata
Craspedida
Metazoa
Ctenophora
Platyctenida
Beroida
Lobata
Fungi (2)
Ichthyosporea (9)
Corallochytrium limacisporum
Tunicaraptor unikontum
Filasterea (4)
Hartaetosiga balthica
Hartaetosiga gracilis
Salpingoeca infusionum
Salpingoeca rosetta
Microstomoeca roanoka
Choanoeca perplexa
Monosiga brevicollis
Salpingoeca kvevrii
Salpingoeca urceolata
Salpingoeca macrocollata
Salpingoeca punica
Salpingoeca helianthica
Mylnosiga fluctuans
Codosiga hollandica
Salpingoeca dolichothecata
Acanthoecida (6)
Euplokamis dunlapae
Coeloplana astericola
Vallicula multiformis
Pleurobrachia bachei
Pukia falcata
Hormiphora californensis
Mertensiidae sp.
Dryodora glandiformis
Beroe ovata
Beroe forskalii
Beroe abyssicola
Beroe sp.
Mnemiopsis leidyi
Bolinopsis ashleyi
Bolinopsis infundibulum
Porifera (15)
Placozoa (3)
Cnidaria (28)
Bilateria (3)
Gain Scenario
Opisthokonta
Holozoa
Pluriformea
Filozoa
Choanozoa
Choanoflagellata
Craspedida
Metazoa
Ctenophora
Platyctenida
Beroida
Lobata
Fungi (2)
Ichthyosporea (9)
Corallochytrium limacisporum
Tunicaraptor unikontum
Filasterea (4)
Hartaetosiga balthica
Hartaetosiga gracilis
Salpingoeca infusionum
Salpingoeca rosetta
Microstomoeca roanoka
Choanoeca perplexa
Monosiga brevicollis
Salpingoeca kvevrii
Salpingoeca urceolata
Salpingoeca macrocollata
Salpingoeca punica
Salpingoeca helianthica
Mylnosiga fluctuans
Codosiga hollandica
Salpingoeca dolichothecata
Acanthoecida (6)
Euplokamis dunlapae
Coeloplana astericola
Vallicula multiformis
Pleurobrachia bachei
Pukia falcata
Hormiphora californensis
Mertensiidae sp.
Dryodora glandiformis
Beroe ovata
Beroe forskalii
Beroe abyssicola
Beroe sp.
Mnemiopsis leidyi
Bolinopsis ashleyi
Bolinopsis infundibulum
Porifera (15)
Placozoa (3)
Cnidaria (28)
Bilateria (3)
horizontal gene transfer
loss event
aCCA
eCCA
e'CCA

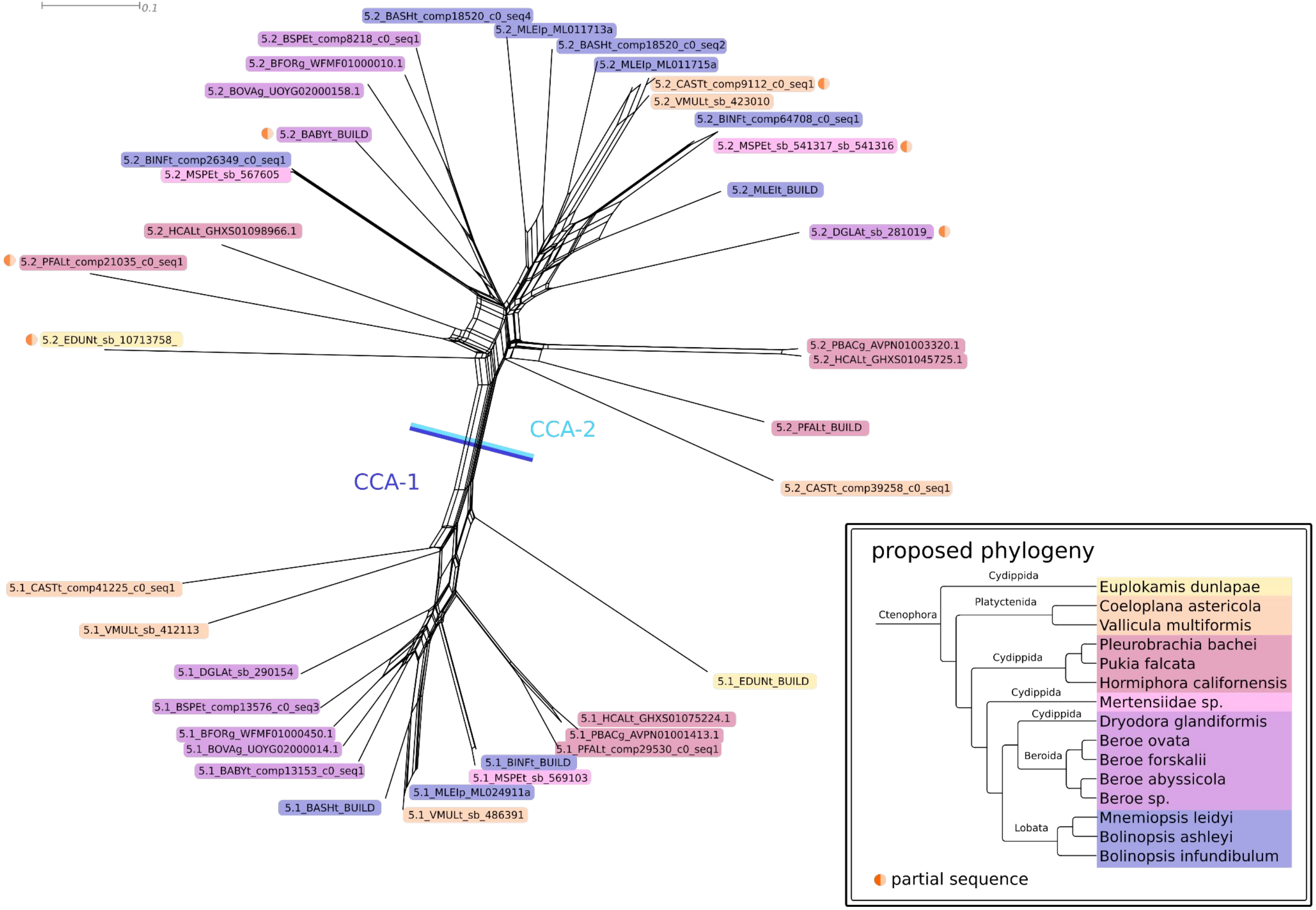
Distribution of CCA-adding Enzymes

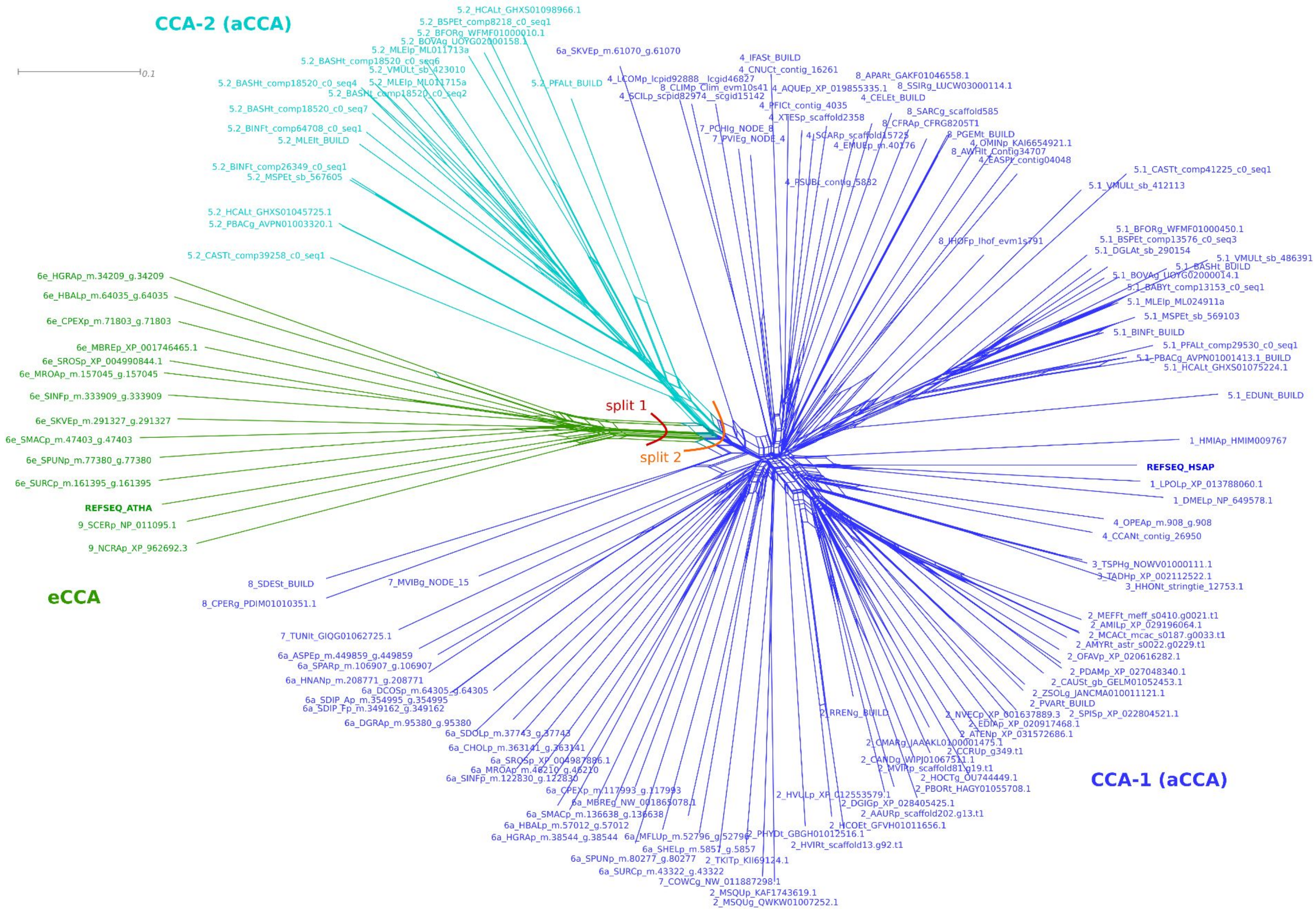
- multiple a-type enzymes in Ctenophora



Ctenophore Sequences

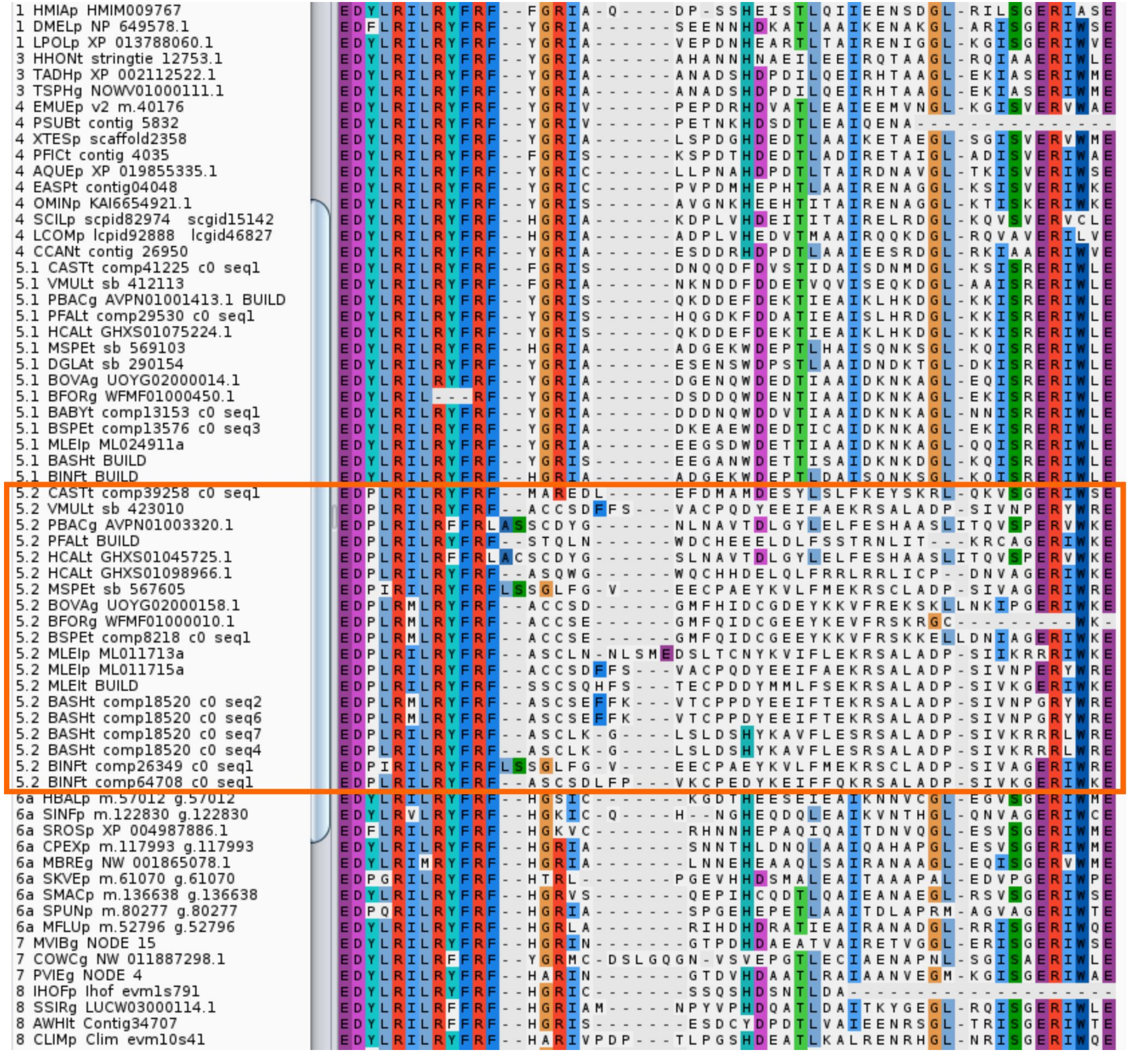
- multiple sequences found in Ctenophora, all clearly classified as aCCA
- significant differences separating them into two distinct groups
  →CCA-1 and CCA-2


0.1
5.2_BASHt_comp18520_c0_seq4
5.2_MLEIp_ML011713a
5.2_BSPEt_comp8218_c0_seq1
5.2_BASHt_comp18520_c0_seq2
5.2_BFORg_WFMF01000010.1
5.2_MLEIp_ML011715a
5.2_BOVAg_UOYG02000158.1
5.2_CASTt_comp9112_c0_seq1
5.2_VMULt_sb_423010
5.2_BINFt_comp64708_c0_seq1
5.2_BABYt_BUILD
5.2_MSPEt_sb_541317_sb_541316
5.2_BINFt_comp26349_c0_seq1
5.2_MSPEt_sb_567605
5.2_MLEIt_BUILD
5.2_HCALt_GHXS01098966.1
5.2_DGLAt_sb_281019_
5.2_PFALt_comp21035_c0_seq1
5.2_EDUNt_sb_10713758_
5.2_PBACg_AVPN01003320.1
5.2_HCALt_GHXS01045725.1
CCA-2
5.2_PFALt_BUILD
CCA-1
5.2_CASTt_comp39258_c0_seq1
5.1_CASTt_comp41225_c0_seq1
5.1_VMULt_sb_412113
5.1_DGLAt_sb_290154
5.1_EDUNt_BUILD
5.1_BSPEt_comp13576_c0_seq3
5.1_HCALt_GHXS01075224.1
5.1_BFORg_WFMF01000450.1
5.1_PBACg_AVPN01001413.1
5.1_BOVAg_UOYG02000014.1
5.1_PFALt_comp29530_c0_seq1
5.1_BINFt_BUILD
5.1_BABYt_comp13153_c0_seq1
5.1_MSPEt_sb_569103
5.1_MLEIp_ML024911a
5.1_BASHt_BUILD
5.1_VMULt_sb_486391
proposed phylogeny
Ctenophora
Cydippida
Platyctenida
Cydippida
Cydippida
Cydippida
Beroida
Lobata
Euplokamis dunlapae
Coeloplana astericola
Vallicula multiformis
Pleurobrachia bachei
Pukia falcata
Hormiphora californensis
Mertensiidae sp.
Dryodora glandiformis
Beroe ovata
Beroe forskalii
Beroe abyssicola
Beroe sp.
Mnemiopsis leidyi
Bolinopsis ashleyi
Bolinopsis infundibulum
◐ partial sequence


CCA-2 (aCCA)
0.1
5.2_HCALt_GHXS01098966.1
5.2_BSPEt_comp8218_c0_seq1
5.2_BFORg_WFMF01000010.1
5.2_BOVAg_UOYG02000158.1
5.2_MLEIp_ML011713a
5.2_BASHt_comp18520_c0_seq6
5.2_VMULt_sb_423010
5.2_BASHt_comp18520_c0_seq4
5.2_MLEIp_ML011715a
5.2_BASHt_comp18520_c0_seq2
5.2_PFALt_BUILD
5.2_BASHt_comp18520_c0_seq7
5.2_BINFt_comp64708_c0_seq1
5.2_MLEIt_BUILD
5.2_BINFt_comp26349_c0_seq1
5.2_MSPEt_sb_567605
5.2_HCALt_GHXS01045725.1
5.2_PBACg_AVPN01003320.1
5.2_CASTt_comp39258_c0_seq1
6a_SKVEp_m.61070_g.61070
4_IFASt_BUILD
4_CNUCt_contig_16261
4_LCOMp_lcpid92888__lcgid46827
8_CLIMp_Clim_evm10s41
4_AQUEp_XP_019855335.1
4_SCILp_scpid82974__scgid15142
8_APARt_GAKF01046558.1
8_SSIRg_LUCW03000114.1
4_CELEt_BUILD
4_PFICt_contig_4035
8_SARCg_scaffold585
4_XTESp_scaffold2358
8_CFRAp_CFRG8205T1
7_PCHIg_NODE_8
7_PVIEg_NODE_4
4_SCARp_scaffold15725
8_PGEMt_BUILD
4_OMINp_KAI6654921.1
4_EMUEp_m.40176
8_AWHIt_Contig34707
4_EASPt_contig04048
4_PSUBt_contig_5832
8_IHOFp_Ihof_evm1s791
5.1_CASTt_comp41225_c0_seq1
5.1_VMULt_sb_412113
5.1_BFORg_WFMF01000450.1
5.1_BSPEt_comp13576_c0_seq3
5.1_DGLAt_sb_290154
5.1_VMULt_sb_486391
5.1_BASHt_BUILD
5.1_BOVAg_UOYG02000014.1
5.1_BABYt_comp13153_c0_seq1
5.1_MLEIp_ML024911a
5.1_MSPEt_sb_569103
5.1_BINFt_BUILD
5.1_PFALt_comp29530_c0_seq1
5.1_PBACg_AVPN01001413.1_BUILD
5.1_HCALt_GHXS01075224.1
5.1_EDUNt_BUILD
6e_HGRAp_m.34209_g.34209
6e_HBALp_m.64035_g.64035
6e_CPEXp_m.71803_g.71803
6e_MBREp_XP_001746465.1
6e_SROSp_XP_004990844.1
6e_MROAp_m.157045_g.157045
6e_SINFp_m.333909_g.333909
split 1
6e_SKVEp_m.291327_g.291327
6e_SMACp_m.47403_g.47403
split 2
1_HMIAp_HMIM009767
6e_SPUNp_m.77380_g.77380
REFSEQ_HSAP
6e_SURCp_m.161395_g.161395
1_LPOLp_XP_013788060.1
1_DMELp_NP_649578.1
REFSEQ_ATHA
9_SCERp_NP_011095.1
4_OPEAp_m.908_g.908
4_CCANt_contig_26950
9_NCRAp_XP_962692.3
3_TSPHg_NOWV01000111.1
3_TADHp_XP_002112522.1
8_SDESt_BUILD
7_MVIBg_NODE_15
3_HHONt_stringtie_12753.1
eCCA
8_CPERg_PDIM01010351.1
2_MEFFt_meff_s0410.g0021.t1
2_AMILp_XP_029196064.1
2_MCACt_mcac_s0187.g0033.t1
7_TUNIt_GIQG01062725.1
2_AMYRt_astr_s0022.g0229.t1
6a_ASPEp_m.449859_g.449859
2_OFAVp_XP_020616282.1
6a_SPARp_m.106907_g.106907
2_PDAMp_XP_027048340.1
6a_HNANp_m.208771_g.208771
2_CAUSt_gb_GELM01052453.1
6a_DCOSp_m.64305_g.64305
2_ZSOLg_JANCMA010011121.1
6a_SDIP_Ap_m.354995_g.354995
2_PVARt_BUILD
6a_SDIP_Fp_m.349162_g.349162
2_RRENg_BUILD
2_NVECp_XP_001637889.3
2_SPISp_XP_022804521.1
6a_DGRAp_m.95380_g.95380
2_EDIAp_XP_020917468.1
6a_SDOLp_m.37743_g.37743
2_ATENp_XP_031572686.1
2_CMARg_JAAAKL010001475.1
6a_CHOLp_m.363141_g.363141
2_CCRUp_g349.t1
6a_SROSp_XP_004987886.1
2_CANDg_WIPJ01067511.1
6a_MROAp_m.46210_g.46210
2_MVIRp_scaffold81.g19.t1
6a_SINFp_m.122830_g.122830
2_HOCTg_OU744449.1
CCA-1 (aCCA)
2_PBORt_HAGY01055708.1
6a_CPEXp_m.117993_g.117993
2_HVULp_XP_012553579.1
6a_MBREg_NW_001865078.1
2_DGIGp_XP_028405425.1
6a_SMACp_m.136638_g.136638
2_AAURp_scaffold202.g13.t1
6a_HBALp_m.57012_g.57012
2_HCOEt_GFVH01011656.1
6a_HGRAp_m.38544_g.38544
6a_MFLUp_m.52796_g.52796
2_PHYDt_GBGH01012516.1
2_HVIRt_scaffold13.g92.t1
6a_SHELp_m.5857_g.5857
6a_SPUNp_m.80277_g.80277
2_TKITp_KII69124.1
6a_SURCp_m.43322_g.43322
7_COWCg_NW_011887298.1
2_MSQUp_KAF1743619.1
2_MSQUg_QWKW01007252.1

Ctenophore Sequences

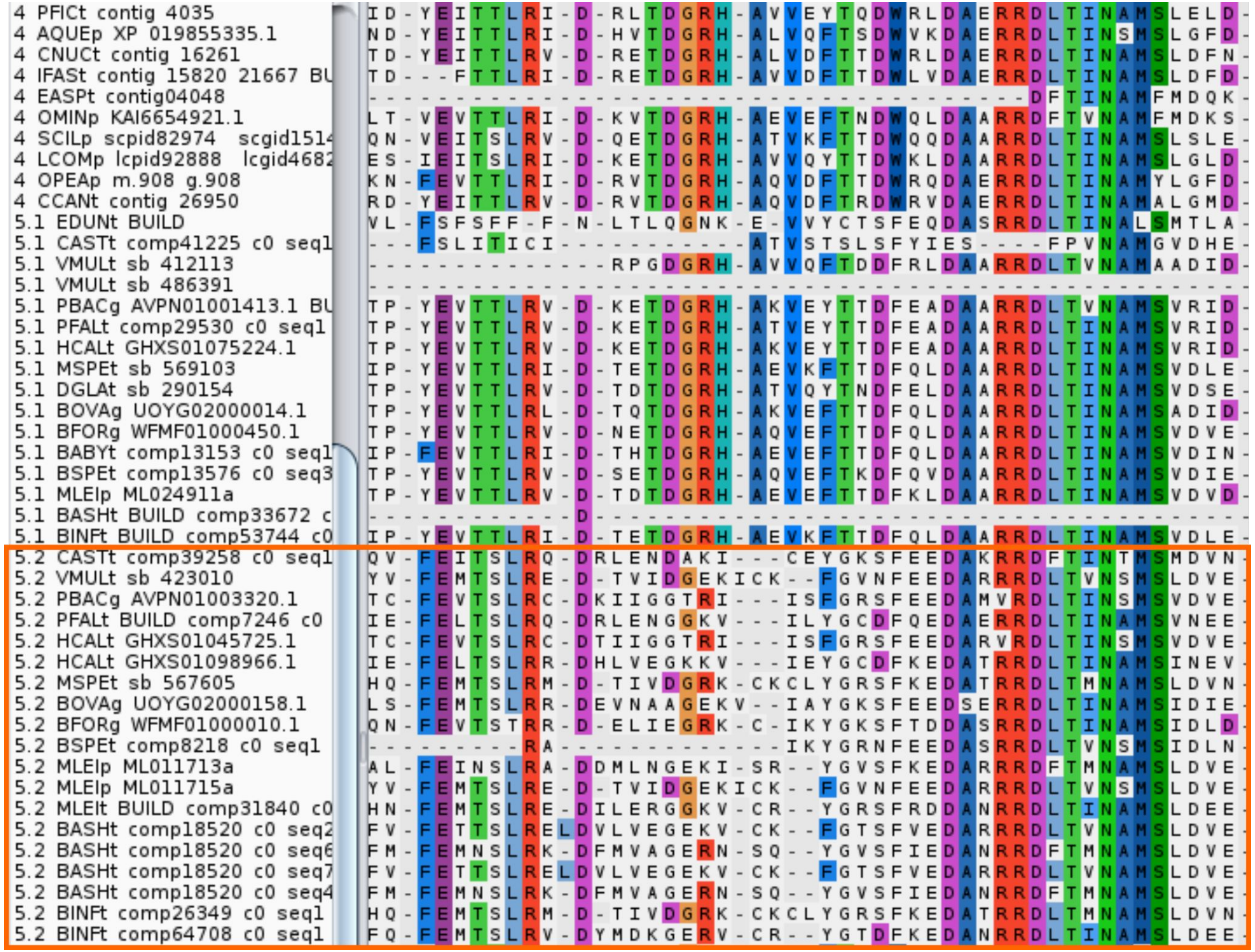
- CCA-1 sequences resemble those of all other holozoan lineages
- CCA-2 sequences differ in:
  - motif D: forms the single nucleotide binding pocket where Watson-Crick-like hydrogen bonds are formed with the incoming nucleotides
  - CCA-2 sequences show a rare proline substitution in this motif
    →indicative of structural changes



```
1 HMIAp HMIM009767                    EDYLRILRYFRF--FGRIA-Q----DP-SSHEISTLQIIEENSDGL-RILSGERIASE
1 DMELp NP 649578.1                   EDFLRILRYFRF--YGRIA------SEENNHDKATLAAIKENAKGL-ARISGERIWSE
1 LPOLp XP 013788060.1                EDYLRILRYFRF--YGRIA------VEPDNHEARTLTAIRENIGGL-KGISGERIWVE
3 HHONt stringtie 12753.1            EDYLRILRYFRF--YGRIA------AHANNHNAEILEEIRQTAAGL-RQIAAERIWLE
3 TADHp XP 002112522.1                EDYLRILRYFRF--YGRIA------ANADSHDPDILQEIRHTAAGL-EKIASERIWME
3 TSPHg NOWV01000111.1                EDYLRILRYFRF--YGRIA------ANADSHDPDILQEIRHTAAGL-EKIASERIWME
4 EMUEp v2 m.40176                   EDYLRILRYFRF--YGRIV------PEPDRHDVATLEAIEEMVNGL-KGISVERVWAE
4 PSUBt contig 5832                  EDYLRILRYFRF--YGRIV------PETNKHDSDTLEAIQENA--------------
4 XTESp scaffold2358                 EDYLRILRYFRF--YGRIA------LSPDGHDEDTLAAIKETAEGL-SGISVERVWME
4 PFICt contig 4035                  EDYLRILRYFRF--FGRIS------KSPDTHDEDTLADIRETAIGL-ADISVERIWAE
4 AQUEp XP 019855335.1               EDYLRILRYFRF--YGRIC------LLPNAHDPDTLTAIRDNAVGL-TKISVERIWSE
4 EASPt contig04048                  EDYLRILRYFRF--YGRIC------PVPDMHEPHTLAAIRENAGGL-KSISVERIWKE
4 OMINp KAI6654921.1                 EDYLRILRYFRF--YGRIS------AVGNKHEEHTITAIRENAGGL-KTISKERIWKE
4 SCILp scpid82974  scgid15142      EDYLRILRYFRF--HGRIA------KDPLVHDEITITAIRELRDGL-KQVSVERVCLE
4 LCOMp lcpid92888  lcgid46827      EDYLRILRYFRF--HGRIA------ADPLVHEDVTMAAIRQQKDGL-RQVAVERILVE
4 CCANt contig 26950                 EDYLRILRYFRF--YGRIA------ESDDRHDPDTLAAIEESRDGL-RKIAAERIWVE
5.1 CASTt comp41225 c0 seq1         EDYLRILRYFRF--FGRIS------DNQQDFDVSTIDAISDNMDGL-KSISRERIWLE
5.1 VMULt sb 412113                 EDYLRILRYFRF--FGRIA------NKNDDFDDETVQVISEQKDGL-AAISRERIWLE
5.1 PBACg AVPN01001413.1 BUILD      EDYLRILRYFRF--YGRIS------QKDDEFDEKTIEAIKLHKDGL-KKISRERIWLE
5.1 PFALt comp29530 c0 seq1         EDYLRILRYFRF--YGRIS------HQGDKFDDATIEAISLHRDGL-KKISRERIWLE
5.1 HCALt GHXS01075224.1            EDYLRILRYFRF--YGRIS------QKDDEFDEKTIEAIKLHKDGL-KKISRERIWLE
5.1 MSPEt sb 569103                 EDYLRILRYFRF--HGRIA------ADGEKWDEPTLHAISQNKSGL-KQISRERIWLE
5.1 DGLAt sb 290154                 EDYLRILRYFRF--YGRIA------ESENSWDPSTLAAIDNDKTGL-DKISRERIWLE
5.1 BOVAg UOYG02000014.1            EDYLRILRYFRF--YGRIA------DGENQWDEDTIAAIDKNKAGL-EQISRERIWLE
5.1 BFORg WFMF01000450.1            EDYLRIL---RF--YGRIA------DSDDQWDENTIAAIDKNKAGL-EKISRERIWLE
5.1 BABYt comp13153 c0 seq1         EDYLRILRYFRF--YGRIA------DDDNQWDDVTIAAIDKNKAGL-NNISRERIWLE
5.1 BSPEt comp13576 c0 seq3         EDYLRILRYFRF--YGRIA------DKEAEWDEDTICAIDKNKAGL-EKISRERIWLE
5.1 MLEIp ML024911a                 EDYLRILRYFRF--YGRIA------EEGSDWDETTIAAIDKNKAGL-QQISRERIWLE
5.1 BASHt BUILD                     EDYLRILRYFRF--YGRIS------EEGANWDETTISAIDKNKDGL-KQISRERIWLE
5.1 BINFt BUILD                     EDYLRILRYFRF--HGRIA------ADGEKWDEPTLDAISQNKSGL-KQISRERIWLE
5.2 CASTt comp39258 c0 seq1         EDPLRILRYFRF--MAREDL-----EFDMAMDESYLSLFKEYSKRL-QKVSGERIWSE
5.2 VMULt sb 423010                 EDPLRILRYFRF--ACCSDFFS---VACPQDYEEIFAEKRSALADP-SIVNPERYWRE
5.2 PBACg AVPN01003320.1            EDPLRILRFFRLASSCDYG------NLNAVTDLGYLELFESHAASLITQVSPERVWKE
5.2 PFALt BUILD                     EDPLRILRYFRF--STQLN------WDCHEEELDLFSSTRNLIT---KRCAGERIWKE
5.2 HCALt GHXS01045725.1            EDPLRILRFFRLACSCDYG------SLNAVTDLGYLELFESHAASLITQVSPERVWKE
5.2 HCALt GHXS01098966.1            EDPLRILRYFRF--ASQWG------WQCHHDELQLFRRLRRLICP--DNVAGERIWKE
5.2 MSPEt sb 567605                 EDPIRILRYFRFLSSGLFG-V----EECPAEYKVLFMEKRSCLADP-SIVAGERIWRE
5.2 BOVAg UOYG02000158.1            EDPLRMLRYFRF--ACCSD------GMFHIDCGDEYKKVFREKSKLLNKIPGERIWKE
5.2 BFORg WFMF01000010.1            EDPLRMLRYFRF--ACCSE------GMFQIDCGEEYKEVFRSKRGC---------WK-
5.2 BSPEt comp8218 c0 seq1          EDPLRMLRYFRF--ACCSE------GMFQIDCGEEYKKVFRSKKELLDNIAGERIWKE
5.2 MLEIp ML011713a                 EDPLRILRYFRF--ASCLN-NLSMEDSLTCNYKVIFLEKRSALADP-SIIKRRRIWKE
5.2 MLEIp ML011715a                 EDPLRILRYFRF--ACCSDFFS---VACPQDYEEIFAEKRSALADP-SIVNPERYWRE
5.2 MLEIt BUILD                     EDPLRILRYFRF--SSCSQHFS---TECPDDYMMLFSEKRSALADP-SIVKGERIWKE
5.2 BASHt comp18520 c0 seq2         EDPLRMLRYFRF--ASCSEFFK---VTCPPDYEEIFTEKRSALADP-SIVNPGRYWRE
5.2 BASHt comp18520 c0 seq6         EDPLRMLRYFRF--ASCSEFFK---VTCPPDYEEIFTEKRSALADP-SIVNPGRYWRE
5.2 BASHt comp18520 c0 seq7         EDPLRILRYFRF--ASCLK-G----LSLDSHYKAVFLESRSALADP-SIVKRRRLWRE
5.2 BASHt comp18520 c0 seq4         EDPLRILRYFRF--ASCLK-G----LSLDSHYKAVFLESRSALADP-SIVKRRRLWRE
5.2 BINFt comp26349 c0 seq1         EDPIRILRYFRFLSSGLFG-V----EECPAEYKVLFMEKRSCLADP-SIVAGERIWRE
5.2 BINFt comp64708 c0 seq1         EDPLRILRYFRF--ASCSDLFP---VKCPEDYKEIFFQKRSALADP-SIVKGERIWKE
6a HBALp m.57012 g.57012            EDYLRILRYFRF--HGSIC-------KGDTHEESEIEAIKNNVCGL-EGVSGERIWME
6a SINFp m.122830 g.122830          EDYLRVLRYFRF--HGKIC-Q----H--NGHEQDQLEAIKVNTHGL-QNVAGERIWCE
6a SROSp XP 004987886.1             EDFLRILRYFRF--HGKVC-------RHNNHEPAQIQAITDNVQGL-ESVSGERIWME
6a CPEXp m.117993 g.117993          EDYLRILRYFRF--HGRIA-------SNNTHLDNQLAAIQAHAPGL-ESVSGERIWME
6a MBREg NW 001865078.1             EDYLRIMRYFRF--HGRIA-------LNNEHEAAQLSAIRANAAGL-EQISGERVWME
6a SKVEp m.61070 g.61070            EDPGRILRYFRF--HTRL-------PGEVHHDSMALEAITAAAPAL-EDVPGERIWPE
6a SMACp m.136638 g.136638          EDYLRILRYFRF--HGRVS-------QEPIHCQDTLQAIEANAEGL-RSVSGERIWSE
6a SPUNp m.80277 g.80277            EDPQRILRYFRF--HGRIA-------SPGEHEPETLAAITDLAPRM-AGVAGERIWTE
6a MFLUp m.52796 g.52796            EDYLRILRYFRF--HGRLA-------RIHDHDRATIEAIRANADGL-RRISGERIWQE
7 MVIBg NODE 15                     EDYLRILRYFRF--HGRIN-------GTPDHDAEATVAIRETVGGL-ERISGERIWSE
7 COWCg NW 011887298.1              EDYLRILRFFRF--YGRMC-DSLGQGN-VSVEPGTLECIAENAPNL-SGISAERIWLE
7 PVIEg NODE 4                      EDYLRILRYFRF--HARIN-------GTDVHDAATLRAIAANVEGM-KGISGERIWAE
8 IHOFp Ihof evm1s791               EDYLRILRYFRF--HGRIC-------SSQSHDSNTLDA-------------------
8 SSIRg LUCW03000114.1              EDYLRILRFFRF--HGRIAM-----NPYVPHDQATLDAITKYGEGL-RQISGERIWLE
8 AWHIt Contig34707                 EDYLRILRFFRF--HGRIS-------ESDCYDPDTLVAIEENRSGL-TRISGERIWTE
8 CLIMp Clim evm10s41               EDYLRILRYFRF--HARIVPDP---TLPGSHDEATLKALRENRHGL-NRISGERIWQE
```

Ctenophore Sequences

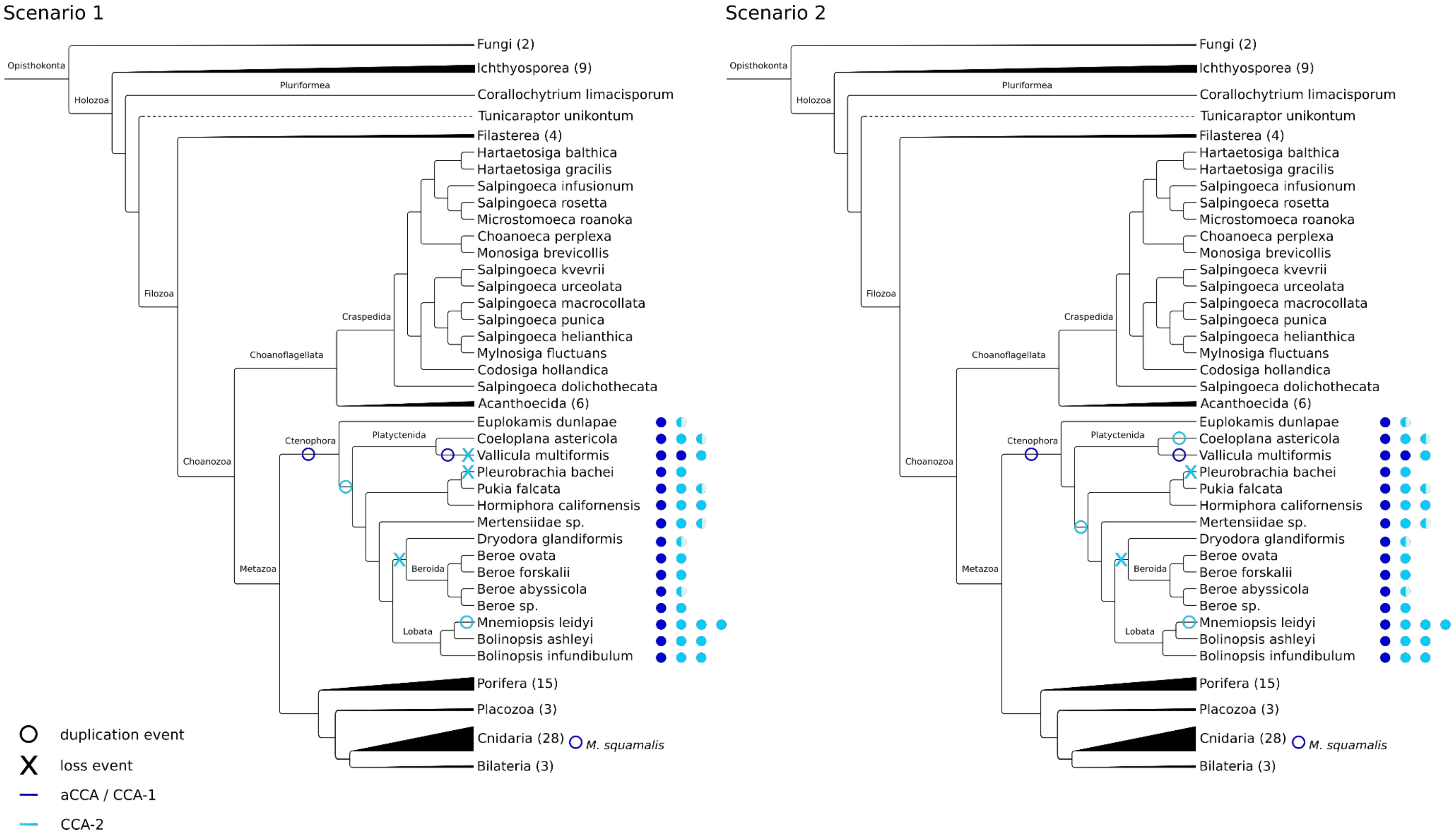
- CCA-1 sequences resemble those of all other holozoan lineages
- CCA-2 sequences differ in:
  - motif D: forms the single nucleotide binding pocket where Watson-Crick-like hydrogen bonds are formed with the incoming nucleotides
  - CCA-2 sequences show a rare proline substitution in this motif
    →indicative of structural changes
  - point mutations in the flexible loop
  - exon counts
- structural and functional differences likely, subfunctionalization possible (CCA-1 A-addition, CCA-2 CC-addition)



```
4 PFICt contig 4035              ID-YEITTLRI-D-RLTDGRH-AVVEYTQDWRLDAERRDLTINAMSLELD-
4 AQUEp XP 019855335.1           ND-YEITTLRI-D-HVTDGRH-ALVQFTSDWVKDAERRDLTINSMSLGFD-
4 CNUCt contig 16261             TD-YEITTLRV-D-RETDGRH-ALVDFTTDWRLDAERRDLTINAMSLDFN-
4 IFASt contig 15820 21667 BU    TD---FTTLRI-D-RETDGRH-AVVDFTTDWLVDAERRDLTINAMSLDFD-
4 EASPt contig04048              --------------------------------------DFTINAMFMDQK-
4 OMINp KAI6654921.1             LT-VEVTTLRI-D-KVTDGRH-AEVEFTNDWQLDAARRDFTVNAMFMDKS-
4 SCILp scpid82974  scgid1514    QN-VEITSLRV-D-QETDGRH-ATVKFTTDWQQDAARRDLTINAMSLSLE-
4 LCOMp lcpid92888  lcgid4682    ES-IEITSLRI-D-KETDGRH-AVVQYTTDWKLDAARRDLTINAMSLGLD-
4 OPEAp m.908 g.908              KN-FEVTTLRI-D-RVTDGRH-AQVDFTTDWRQDAERRDLTINAMYLGFD-
4 CCANt contig 26950             RD-YEITTLRV-D-RVTDGRH-AQVDFTRDWRVDAERRDLTINAMALGMD-
5.1 EDUNt BUILD                  VL-FSFSFF-F-N-LTLQGNK-E-VVYCTSFEQDASRRDLTINALSMTLA-
5.1 CASTt comp41225 c0 seql      ---FSLITICI-----------ATVSTSLSFYIES----FPVNAMGVDHE-
5.1 VMULt sb 412113              --------------RPGDGRH-AVVQFTDDFRLDAARRDLTVNAMAADID-
5.1 VMULt sb 486391              ---------------------------------------------------
5.1 PBACg AVPN01001413.1 BU      TP-YEVTTLRV-D-KETDGRH-AKVEYTTDFEADAARRDLTVNAMSVRID-
5.1 PFALt comp29530 c0 seql      TP-YEVTTLRV-D-KETDGRH-ATVEYTTDFEADAARRDLTINAMSVRID-
5.1 HCALt GHXS01075224.1         TP-YEVTTLRV-D-KETDGRH-AKVEYTTDFEADAARRDLTINAMSVRID-
5.1 MSPEt sb 569103              IP-YEVTTLRI-D-TETDGRH-AEVKFTTDFQLDAARRDLTINAMSVDLE-
5.1 DGLAt sb 290154              TP-YEVTTLRV-D-TDTDGRH-ATVQYTNDFELDAARRDLTINAMSVDSE-
5.1 BOVAg UOYG02000014.1         TP-YEVTTLRL-D-TQTDGRH-AKVEFTTDFQLDAARRDLTINAMSADID-
5.1 BFORg WFMF01000450.1         TP-YEVTTLRV-D-NETDGRH-AQVEFTTDFQLDAARRDLTINAMSVDVE-
5.1 BABYt comp13153 c0 seql      IP-FEVTTLRI-D-THTDGRH-AEVEFTTDFQLDAARRDLTINAMSVDIN-
5.1 BSPEt comp13576 c0 seq3      TP-YEVTTLRV-D-SETDGRH-AQVEFTKDFQVDAARRDLTINAMSVDIE-
5.1 MLEIp ML024911a              TP-YEVTTLRV-D-TDTDGRH-AEVEFTTDFKLDAARRDLTINAMSVDVD-
5.1 BASHt BUILD comp33672 c      ------------D--------------------------------------
5.1 BINFt BUILD comp53744 c0     IP-YEVTTLRI-D-TETDGRH-AEVKFTTDFQLDAARRDLTINAMSVDLE-
5.2 CASTt comp39258 c0 seql      QV-FEITSLRQ-DRLENDAKI---CEYGKSFEEDAKRRDFTINTMSMDVN-
5.2 VMULt sb 423010              YV-FEMTSLRE-D-TVIDGEKICK--FGVNFEEDARRRDLTVNSMSLDVE-
5.2 PBACg AVPN01003320.1         TC-FEVTSLRC-DKIIGGTRI---ISFGRSFEEDAMVRDLTINSMSVDVE-
5.2 PFALt BUILD comp7246 c0      IE-FELTSLRQ-DRLENGGKV---ILYGCDFQEDAERRDLTINAMSVNEE-
5.2 HCALt GHXS01045725.1         TC-FEVTSLRC-DTIIGGTRI---ISFGRSFEEDARVRDLTINSMSVDVE-
5.2 HCALt GHXS01098966.1         IE-FELTSLRR-DHLVEGKKV---IEYGCDFKEDATRRDLTINAMSINEV-
5.2 MSPEt sb 567605              HQ-FEMTSLRM-D-TIVDGRK-CKCLYGRSFKEDATRRDLTMNAMSLDVN-
5.2 BOVAg UOYG02000158.1         LS-FEMTSLRR-DEVNAAGEKV--IAYGKSFEEDSERRDLTINAMSIDIE-
5.2 BFORg WFMF01000010.1         QN-FEVTSTRR-D-ELIEGRK-C-IKYGKSFTDDASRRDLTINAMSIDLD-
5.2 BSPEt comp8218 c0 seql       ---------RA-------------IKYGRNFEEDASRRDLTVNSMSIDLN-
5.2 MLEIp ML011713a              AL-FEINSLRA-DDMLNGEKI-SR--YGVSFKEDARRRDFTMNAMSLDVE-
5.2 MLEIp ML011715a              YV-FEMTSLRE-D-TVIDGEKICK--FGVNFEEDARRRDLTVNSMSLDVE-
5.2 MLEIt BUILD comp31840 c0     HN-FEMTSLRE-DILERGGKV-CR--YGRSFRDDANRRDLTINAMSLDEE-
5.2 BASHt comp18520 c0 seq2      FV-FETTSLRELDVLVEGEKV-CK--FGTSFVEDARRRDLTVNAMSLDVE-
5.2 BASHt comp18520 c0 seq6      FM-FEMNSLRK-DFMVAGERN-SQ--YGVSFIEDANRRDFTMNAMSLDVE-
5.2 BASHt comp18520 c0 seq7      FV-FETTSLRELDVLVEGEKV-CK--FGTSFVEDARRRDLTVNAMSLDVE-
5.2 BASHt comp18520 c0 seq4      FM-FEMNSLRK-DFMVAGERN-SQ--YGVSFIEDANRRDFTMNAMSLDVE-
5.2 BINFt comp26349 c0 seql      HQ-FEMTSLRM-D-TIVDGRK-CKCLYGRSFKEDATRRDLTMNAMSLDVN-
5.2 BINFt comp64708 c0 seql      FQ-FEMTSLRV-DYMDKGERV-CR--YGTDFKEDANRRDLTINAMSLDEE-
```

Ctenophore Sequences

- multiple sequences found in Ctenophora, all clearly classified as aCCA
- significant differences separating them into two distinct groups →CCA-1 and CCA-2
- CCA-1 sequences resemble those of all other holozoan lineages
- CCA-2 sequences differ in:
  - motif D: forms the single nucleotide binding pocket where Watson-Crick-like hydrogen bonds are formed with the incoming nucleotides
  - CCA-2 sequences show a rare proline substitution in this motif →indicative of structural changes
  - point mutations in the flexible loop
  - exon counts
- structural and functional differences likely, subfunctionalization possible (CCA-1: A-addition, CCA-2: CC-addition)


Scenario 1

Scenario 2