Intron position and phases

===> Triangle <======

Toxin

>C.ventricosus toxin

MKAGDQLKMFFIFMCLWYVTDGQKCGRNPLIPIEAIAQKT - phase 1 6403bp

GKTTVYVLQVFRQDDPAQFEVWRQKFEAYNHCVGFLETGYYKRSLMEKSETSQDQLQRRGERGRELPGALQAAVGSSNTQRMLTGKRSVAAAAAGGGGDADGAAQGKSL

Conus

>C.ventricosus

MVKREDLLKIFFVYLCLLHVSPGHPCGRNPISRFQPSP -phase 1 6341bp

SRDINSLLNLLRSSNPALYQQVQDEWRRYADCVGLVDTGYFKRSSTGTSEAPQEHLQGRGQRAMALPEAFRAANRSRRTQRLLTLMRSLAAAAAAAGDGDGDAEGAAQALRL

Annelida

>ELT98797.1 hypothetical protein CAPTEDRAFT\_191853 [Capitella teleta]

MNIAGLLFLLGALHWSTVAAQTYGEECGANPMLKAVLVARRA - phase 1 907bp

GRPLRQVLEVLRRIDPPAFKELAQKFRQYAACVALVDTGYFRKRSQASPVTRSNILVTTKTPSPSVNLGFLRKFQMKLRAWLQTKAVSKWPSIIN

>ELT98795.1 hypothetical protein CAPTEDRAFT\_191851 [Capitella teleta]

MAHSLQMAVALVLTYSCCCNATAAVYGEECGSNPMQQVLMVARRS - phase 1 113bp

GHTLTHVLEVLRKIDPPAFRTLAQKFRMYAACVGLVDTGYFKKRSSMPSLLEEPARDFHRSQSRSVLSAEDEDYLPQEFASWLKKYSSSSASENSAYGSAREFGGSVDRFGDDLLELIHDEYPKRTQPITEQEHKELLSDFSDLRPFFVWK

Mollusca

>XP\_034298906.1\_uncharacterized\_protein\_LOC105337531\_isoform\_X1\_[Crassostrea\_gigas]

MVKKKDIVKVFVIYLSLCHVCPAHPCGENPLRRLISISFKT - phase1

GHPLRSVLLLLREMDGNVYKQLQVDFQRYAGCIGIVDTGYFKRSPAPDSFSKRTPQSSTGHNDGLELFKTWTNRN - phase2

RTSDRLSLLLEAQNSLSNELFPFSRRLQRRIR

>XP\_025092781.1\_uncharacterized\_protein\_LOC112563200\_[Pomacea\_canaliculata]

MVRKETLLKVFFIYLCFCHLVPGHPCGSNPLRRFRPLSRML - phase1

GKSIDTLLDLLRESNPALYEQVQLEWRLYAECVGLVDTGYFKRSSDDTQGRVDSEAGPLARSHDGDEGLAADDLLERVLQSWMDATSD - phase2

SDRDLDSGHL

Rotifera

>CAF1000162.1 unnamed protein product [Brachionus calyciflorus]

MKIQTVTIIILFCIAIDGTLSHTHPECGESPWLKYKVLATKI - phase1

GWEWGQLIRFMRTETTMEYNLLLKQVKEYADCIALSNIPL

>RNA38265.1 hypothetical protein BpHYR1\_030566 [Brachionus plicatilis]

MVLNSKNKICVFFIMVMITVISKAKVNSLSFHPECGESPVYKYKELANAI - phase1

GWDINQLIRFMRAEKKEDYEKLIRAVKLYADCVALSQIPIK

Arthropoda

>XP\_042220147.1 uncharacterized protein LOC121864977 [Homarus americanus]

MWTSTANTTPATVM - phase0

DRVWCVLWVGVVTVAGTGGRVMPLLQELEASDGCGPNPLTVVVAAGNDL - phase1

GHLEAVEVMAVLESDPETWGEFLALIHAYDDCVRTGQL - phase 2

RFKRMAPPPLRRSASLAGPAALRLHNASKSTSSDLSAQPRSSFSQQHLLYRLLQHQDPPSRSLDHHRHIQA

>EFX71890.1 hypothetical protein DAPPUDRAFT\_308661 [Daphnia pulex]

MFAVLVSIHVVLLALAVAGLPAHHQLEARQSSDGCGPNPLVAVISYAREF - phase1

ADFSPHQLMRFIESDTRAGSQLRSLLRTYDDCVRTGDGTRYKKSGSALEMMDDVESDMGKSPFRMLAGNNNKSQQQQHSLQQANQSLLRALASHRLL

Tardigrada

>GAV04599.1 hypothetical protein RvY\_14861 [Ramazzottius varieornatus]

MLNRLVFLTLLLAGVLVAKEMADGCGMNPLMALVAVADKT -phase 1

NQSVENLMGKFSADKMAFREFVMLLGRYESCIRVGSGTRYRKRSPEQQVADLIGLL

>GAV04599.1 hypothetical protein RvY\_14861 [Ramazzottius varieornatus]

MLNRLVFLTLLLAGVLVAKEMADGCGMNPLMALVAVADKT - phase1

NQSVENLMGKFSADKMAFREFVMLLGRYESCIRVGSGTRYRKRSPEQQVADLIGLL

Platyhelmintes

>KAG5448219.1 hypothetical protein CSKR\_105721 [Clonorchis sinensis]

MQFSFLIVTTSLISCIATASANVHCGRNPMSVFKRLVFEESNG - phase1

TIDPKFALGQLEIREPRTYTRLLKSHRDWLLCIQKTETGYFKREAFGSSQRTPTIQMRIDGKFEDDQ

>XP\_009171979.1 hypothetical protein T265\_08028 [Opisthorchis viverrini]

MMQFSFLIVTAGLLSCIVNTSANVHCGRNPMGVFKRLVFEESNG - phase1

TIDPKFALGQLEIREPRTYTRLLKSHRDWLLCIQKTETGYFKRAAFDSSQRTPTSIYGTFEND

====> CHH <=====

Conus

>ventricosus\_1

MAVFGLALMTLMLLPESSYTLRHYAGK - phase 2

LPDGHDVTFYLHICSVCGEFYGGSVERNCIVDKSFETFYNCKAAILQRRRK

Molluska

>KOF87272.1 hypothetical protein OCBIM\_22017144mg [Octopus bimaculoides]

MMMSLPRVAIIALICTLMCSYGSSMTLEEIRR - phase2

SIPKGATLYSYAGTCEACGEFYGSSYTYRCLTDKTFETYHKCNIAVNEKRK

>OWF53087.1 hypothetical protein KP79\_PYT00509 [Mizuhopecten yessoensis]

MARVTCLLLLGCLLSCCMANTLESLVRR - phase2

NTPKGHTFSAFVEICESCGEYFGNSLSYRCIMDRSLTTFRQCNTAVVSVNRRR

====> Tail <=====

Toxin (no intron)

>C.ventricosus

MLSLGIIVLACLSMTRMITPMRAPPLVPSQVRTGPVRSTTGTTDTRGVTTITCTTPECVQIIQAYLKWREENGYGASNNRYG

Molluska (no introns)

>XP\_025084112.1 uncharacterized protein LOC112558117 [Pomacea canaliculata]

MCPEVAMDRVTVLLLLLAMLHAASATSTANRTAASLLQLLLADGEGDSPAAPPFQPAPDATTSGCPLHHPDCSPDGDLPAHPREGQTHDLQAPLPASEDEVLIYGIPVKDIPPHLIPAALRKERLLYLSNLPDDVPRLRDGLDAAAFCFTQDCVEMMKEFDQWQREHGYGTPGGRWG

>XP\_045210103.1 uncharacterized protein LOC123561658 [Mercenaria mercenaria]

MKLTLTMISALFCYQFTTALSLPLIEDNETQNSNDQDQQTHASQTSINTDTSLKSLIDSESEVYVDIGYSETTEKMNNSKSKLENKSNETEDDPLAGICQTPECIQQLQEYMKWRQDNGYPVPSGRWGK

Annelida

>ELT87057.1\_hypothetical\_protein\_CAPTEDRAFT\_197833\_[Capitella\_teleta]

MRPVLLFLLPLIIAAIFPPAETTNDVTEGYLDTKDEATVPTASIMPLIDGVPIDLIPDEFLRLLLLGINSELEDFDFHTNYCTTSACREILEEYEKWREENGYGKSEIR - phase2

YISVVG

====> Medial <======

Conus

>C.ventricosus

MGAPVLSRWLLLGFLLSLTLCQMPATARYTYRLQPYR - phase2

PAMTRRGCGGFPCAYNHLARLVGRFAMMRAIMRIMSDCAHDEHCSPG

Mollusca

>XP\_005095677.1\_uncharacterized\_protein\_LOC101858621\_[Aplysia\_californica]

MVRSTVLVAGSLVLVLSLVHLQVTDASDIDDVTRLLMQIDNVLA - phase2

KFRGKRGCIGYMCSYSHMSSSAGSKAVHNSLMKFLYNCAKDPHCSP - phase1

GKRKRRSVSETDTPLWSMLRNRDLETFLRRRDIKSLLSGSSSLNKAPS

>XP\_033752434.1 uncharacterized protein LOC117336157 [Pecten maximus]

MDRVMSVKVTEILTISLFLLLTLRVTLAYPFGLQNFVVPKEPAADDQSEYW - phase2

NELLEDRGPSVPVKRGCVGFPCVYTHMAERAGRASIERYIAKLIEDCMNDDHCNP - phase1

GKRKRRSLLKQINQRR

Annelida

>Alitta virens genome assembly, contig: atg000332l\_1, whole genome shotgun sequence

MNLLPGWFSLTIITLLGGLLHTYAYPTDNEESNYEIDTHLTVPDLSSLFEYEN - phase2

IPAEWPSPKTLPEKRGPCHGTACVYSHIGAKAGHSSLRRLKLTLLRECIDDPLCFAAGKTLR

====> Hairpin <======

Conus

>C.ventricosus

MVRAALSLLLLATLAVLVI - phase1 118,485bp

ANERAEAEEPQHHRAKRQDDMAAVDDYPLDDVDMMQRIFRTPLKRQWCRPGMSFNPVLGTCTLSLA - phase0 59,272 bp

ALRGRGRSFRGV

Mollusca

>XP\_005089801.1\_uncharacterized\_protein\_LOC101860216\_[Aplysia\_californica]

MGSGSRLVCLLLLSMLALAAVS - phase1

ASSLDQTRSRRSLQAADSSDDMLADDWAASVPTDDLLMRLGEGREKRGWCRPNMTFNSILGRCTYVYS - phase0

KLKGGRGKRGWCRPDMIFNPILGRCTFTYA - phase0

AIRNRGRG

>KAG5691889.1 hypothetical protein BaRGS\_033252 [Batillaria attramentaria]

MKL - phase1

GQTEAAAGRESRRVKRQGYNSFDSDLGEYPQYDVDLMQRVFKTPLKRQWCRSGMSFNPVIGTCTLSLA - phase0

FLLS

Annelida

>ELU17817.1 hypothetical protein CAPTEDRAFT\_221947 [Capitella teleta]

MCNADAMRQTT - phase1

LLGKGMLLDDEEFPPAPQMESKDMMAALLAGWSSNQDDEYQRMNPRTEGDGIQRSKRYFCRKFFVYNPVRGRCQPTPQ - phase0

RRLLRYSDWVHLRRLWQGGGVMQDWNSEWAAN - phase1

SRKVSKITGRKQVVNVEGSLSAASELIIMRSASGVDLGPLLFNSPPYFTHLMCGSVRKRMNSNCNATV

>GINI01011046.1 TSA: Alitta virens c14478\_g1\_i1, transcribed RNA sequence

MKNVIVTLFAALLVCLLLCGR - phase1

TACGVIKRDTESVPYEFIEENQALEEEEPQFGDLDLPTKRHLCRRYFIYNPVKGRCMPTLM - phase0

AFRAMRSKSLRSFPNPQAWAHFG