HsCXE21 1 MLSAPVWVCVACAVVSAARARGSTDAIQDGPVTQSVSGQFAGSWMTSRRGRRFQAYRGIR 60

MLSA VWVCVACAVV+AARA STD IQD PVT SVSGQF GSWMTSRRGRRFQAYRGIR

HvCXE21 1 MLSASVWVCVACAVVNAARAHASTDVIQDAPVTHSVSGQFVGSWMTSRRGRRFQAYRGIR 60

Query 61 YAEPPIGELRFQPPKPILHYENIVDASKDGPSCPRPASPGYPVDEDCLTINVYTPANNSS 120

YAEPP+GELRFQPPKPILHY+ IVDAS+DGPSCPRPASPGYPVDEDCLTINVYTPANNSS

Sbjct 61 YAEPPVGELRFQPPKPILHYDKIVDASEDGPSCPRPASPGYPVDEDCLTINVYTPANNSS 120

Query 121 KSLPVIFYIHPGGFYSMSGRSDLAGPHYLLDRDVVLVTINYRLGSLGFLALGNKLAPGNN 180

K LPVIFYIHPGGFYSMSGRS+LAGPHYL+DRDVVLVTINYRLGSLGFLALGNKLAPGNN

Sbjct 121 KRLPVIFYIHPGGFYSMSGRSNLAGPHYLMDRDVVLVTINYRLGSLGFLALGNKLAPGNN 180

Query 181 GFKDQVAALKWVQRNIVAFGGDPDLVTITGCSAGAVSVMVHMISPMSKGLFHRGISVSGS 240

GFKDQVAALKWVQRNI AFGGD +LVTITGCSAGAVSVMVHMISPMSKGLFHR ISVSGS

Sbjct 181 GFKDQVAALKWVQRNIAAFGGDSNLVTITGCSAGAVSVMVHMISPMSKGLFHRAISVSGS 240

Query 241 PTSKLPSPPDRIDLAVKQARVSHCPEDNSTALFHCLKTKSWKEIAGSMLGFFEFGYDPLS 300

PTSKLPSPPDR+DLA+KQARVSHCPEDNSTALFHCLKTKSWKEIAGSMLGFFEFGYDPLS

Sbjct 241 PTSKLPSPPDRVDLAIKQARVSHCPEDNSTALFHCLKTKSWKEIAGSMLGFFEFGYDPLS 300

Query 301 LWRPVVETDFGQERFLVEEPTVSIREGRMHSVPFIISQTTGEFFWKAFYVLNNQTLLNTM 360

LWRPVVE DFGQERFLVEEP VSIREGRMHSVPFIISQTTGEFFWKAFYVLNNQTLLNTM

Sbjct 301 LWRPVVEPDFGQERFLVEEPMVSIREGRMHSVPFIISQTTGEFFWKAFYVLNNQTLLNTM 360

Query 361 NAEWDRIAPISFILPRNESTDKLATLRKAYLGDGRLQNSTAGADGLGKLYGDSITGFPVH 420

NAEW+RIAPISFILPRNES+D+LATLRKAYLGDGRL+NSTA ADGLGKLYGDSITGFPVH

Sbjct 361 NAEWNRIAPISFILPRNESSDRLATLRKAYLGDGRLENSTASADGLGKLYGDSITGFPVH 420

Query 421 RMANLMCRHSKHSVFYSEFAYLGNRSHYEDPISKKPVRTAHHDDLIYLFTLSYRFPPIDV 480

RMANLMCRHSKH VFYSEFAYLGNRSHY+DP+SKKPVRTAHHDDLIYLFTLSYRFPPIDV

Sbjct 421 RMANLMCRHSKHPVFYSEFAYLGNRSHYKDPVSKKPVRTAHHDDLIYLFTLSYRFPPIDV 480

Query 481 ADTEDSKMVDKMTALWYNFARYGNPNPRGDTPELGSLSWPAMTPAHRHYLRVDKEFAVKQ 540

A+TEDSKMVDKMTALWYNFARYG+PNPRGDTPEL SLSWPAMTPAHRHYLRVDKEFAVKQ

Sbjct 481 AETEDSKMVDKMTALWYNFARYGDPNPRGDTPELASLSWPAMTPAHRHYLRVDKEFAVKQ 540

Query 541 NLFEHRFAVWDHLYPTQY\* 559

NLFEHRFAVWDHLYP QY\*

Sbjct 541 NLFEHRFAVWDHLYPIQY\* 559

GCSAG is unusual. GESAG is much more common.

Hs-CXE5 1 MHLRALLCALCVAVCAHAHRHTTERANVPSSTLKPSEPVTRSVSGTFRGSWMETRRGRRF 60

M+LRALLCALCVAVCAHAHRHTTERANVPSSTLKPSEPVTRSVSGTFRGSWMETRRGRRF

Hv-CXE5 1 MYLRALLCALCVAVCAHAHRHTTERANVPSSTLKPSEPVTRSVSGTFRGSWMETRRGRRF 60

Query 61 QAYRGIRYAEPPIGPLRFQPPKLKLNYESVVDASEEGPACPLPAPPTYYVDEDCLTVNVY 120

QAYRGIRYAEPP+G LRFQPPKLKLNYE VVDASEEGPACPLPAPPTYYVDEDCLTVNVY

Sbjct 61 QAYRGIRYAEPPVGALRFQPPKLKLNYEGVVDASEEGPACPLPAPPTYYVDEDCLTVNVY 120

Query 121 TPAHNTSQRLPVIFFIHPGGFYAMTGRSDLAGPHYLLDRDVVLVTINYRLGSLGFMSTGD 180

TPAHNTSQRLPVIFFIHPGGFYAMTGRSDLAGPHYLLDRDVVLVTINYRLGSLGFMSTGD

Sbjct 121 TPAHNTSQRLPVIFFIHPGGFYAMTGRSDLAGPHYLLDRDVVLVTINYRLGSLGFMSTGD 180

Query 181 ANAPGNNGMKDQVAALKWVQRNIAAFGGDPDLVTITGCSAGSISVLLHMMSPMTKGLFHR 240

ANAPGNNGMKDQVAALKWVQRNIAAFGGDPDLVTITGCSAGSISVLLHMMSPMTKGLFHR

Sbjct 181 ANAPGNNGMKDQVAALKWVQRNIAAFGGDPDLVTITGCSAGSISVLLHMMSPMTKGLFHR 240

Query 241 GIAMSASPISKEMTTLTHQRHLAVRQAEILNCPTHNSSAIVDCLFTKPWRELGDSLTKFW 300

GIAMSASPISKEMTTLTHQRHLAVRQAEILNCPT NSSAIVDCL TKPWRELGDSLTKFW

Sbjct 241 GIAMSASPISKEMTTLTHQRHLAVRQAEILNCPTQNSSAIVDCLLTKPWRELGDSLTKFW 300

Query 301 EFGPGDPVGLWGPVVEPDFGQERYLTMNPLDAVKQGKMHTVPLIISQTTDEFFWKAFQVL 360

EFGPGDPVGLWGPVVEPDFGQERYLT+NPL+AVKQGKMHTVPLIISQTTDEFFWKAFQVL

Sbjct 301 EFGPGDPVGLWGPVVEPDFGQERYLTINPLEAVKQGKMHTVPLIISQTTDEFFWKAFQVL 360

Query 361 NETLLKTMNDDWERIAPISFILPQENRAAAVRRLREVYLKGNKLVNDSESAKALGRLYG 420

NETLLKTMNDDWERIAPISFILPQENRAAAV+RLREVYLKGNKLVNDSESAKALGRLYG

Sbjct 361 QNETLLKTMNDDWERIAPISFILPQENRAAAVKRLREVYLKGNKLVNDSESAKALGRLYG 420

Query 421 DSVVGFGVHRMANLMCRHSTHPVWYYEFAYVGNNSHYEDPSGKPQGAAHHDDLLYLFTLS 480

DSV+GFGVHRMANLMCRHSTHPVWYYEFAYVGNNSHYEDPSGKP+GAAHHDDLLYLFTLS

Sbjct 421 DSVIGFGVHRMANLMCRHSTHPVWYYEFAYVGNNSHYEDPSGKPKGAAHHDDLLYLFTLS 480

Query 481 YNFPTIELSSPHSHVVDEMTAMWYNFARYGDPNPRGDTPELGKLTWPAMTPDRRQFLHRG 540

YNFPTIELSSPHSHVVDEMTA+WYNFARYGDPN RGDTPELGKLTWPAMTPDRRQFLHRG

Sbjct 481 YNFPTIELSSPHSHVVDEMTAIWYNFARYGDPNSRGDTPELGKLTWPAMTPDRRQFLHRG 540

Query 541 DQLVVRQNMFEDRFKVWEELYPIQY\* 566

DQL++RQNMFEDRF+VWEELYPIQY\*

Sbjct 541 DQLLIRQNMFEDRFRVWEELYPIQY\* 566

GCSAG is unusual. GESAG is much more common.

Hs-CXE16 1 MLCVVLLSVALACAGADRHEQHGAGTSPASPEPLARSVSGQFAGSWMTSRRGRRFQAYRG 60

MLCVVLLS ALACA AD+HEQHGAGTSPASPEP+ARSVSGQF GSWMTSRRGRRFQAYRG

Hv-CXE16 1 MLCVVLLSAALACASADKHEQHGAGTSPASPEPMARSVSGQFVGSWMTSRRGRRFQAYRG 60

Query 61 IRYAEPPVGELRFQPPKPILNYSREVDARQEGPACPQPTYNDYPVHEDCLRLNVYTPVTA 120

IRYAEPPVGELRFQPPKPILNYSREVDAR+EGPACPQPTYNDYPVHEDCLRLNVYTP A

Sbjct 61 IRYAEPPVGELRFQPPKPILNYSREVDARKEGPACPQPTYNDYPVHEDCLRLNVYTPANA 120

Query 121 GNDKKLPVVVFMHAGGFYSVSGRSDVAGPSHLLDRDLVLVTINYRLGSLGFLSTGDALAP 180

GNDKKLPV+VFMHAGGFYSVSGRSDVAGPSHLLDRDLVLVTINYRLGSLGFLSTGDALAP

Sbjct 121 GNDKKLPVLVFMHAGGFYSVSGRSDVAGPSHLLDRDLVLVTINYRLGSLGFLSTGDALAP 180

Query 181 GNNGFKDQVMALRWVKRNIAAFGGDPDLVTISGYSAGSFSVMLHMISPMSKGLFHRAISM 240

GNNGFKDQVMALRWVKRNIAAFGGDPDLVTISGYSAGSFSVMLHMISPMSKGLFHRAISM

Sbjct 181 GNNGFKDQVMALRWVKRNIAAFGGDPDLVTISGYSAGSFSVMLHMISPMSKGLFHRAISM 240

Query 241 SGSPISQIEIPRHQRHLAERQARLLACPTDSSRAIIDCLRTKTSKEIGDSLTNMFDFGYD 300

SGSP+SQIEIPRHQRHLAERQARLLACPTDSSRAIIDCL+TKTSKEIGDSL MFDFGYD

Sbjct 241 SGSPVSQIEIPRHQRHLAERQARLLACPTDSSRAIIDCLKTKTSKEIGDSLHKMFDFGYD 300

Query 301 PVLLWVPIHEQDFGQEMFLPRQPLEALCSADLTQVPYIVSQTQDEFFWKALDVLRDPAAY 360

PVLLWVPIHEQDFGQEMFLPRQPL+ALCSADL QVPYIVSQTQDEFFWKALDVLRDPAAY

Sbjct 301 PVLLWVPIHEQDFGQEMFLPRQPLDALCSADLAQVPYIVSQTQDEFFWKALDVLRDPAAY 360

Query 361 DSWRADWPGKGRVAFYLSGDNLTATAAANRLKQAYLGGKDIANDTATGDGFGKLYSDAII 420

DSWRADWPGKGRVAFYLSGDN TATAAANRLKQAYLGGKDIANDTATGDGFGKLYSDAII

Sbjct 361 DSWRADWPGKGRVAFYLSGDNATATAAANRLKQAYLGGKDIANDTATGDGFGKLYSDAII 420

Query 421 GFGAHRLVNLMSRHLRQPLYYYEFGYIGNSSHYVDPDTKKPIAAAHHDDLLYLFNVSYSF 480

GFGAHRLVNLMSRHL QPLYYYEFGYIGNSSHYVDPDTKKPIAAAHHDDLLYLFNVSYSF

Sbjct 421 GFGAHRLVNLMSRHLHQPLYYYEFGYIGNSSHYVDPDTKKPIAAAHHDDLLYLFNVSYSF 480

Query 481 PSIPPTDSKDSQMVDKMTALVYNFARYGDPNPKADTPELSGLSWPQYKPDERKYLRVDSP 540

PSIPPTDSKDSQMVDKMTALVYNFARYGDPNPK DTPEL GLSWPQYKPDERK+LRVD P

Sbjct 481 PSIPPTDSKDSQMVDKMTALVYNFARYGDPNPKPDTPELRGLSWPQYKPDERKFLRVDEP 540

Query 541 FKVSQRLFEDRFRVWEELFPLDYQKCK\* 568

FKVSQRLFEDRFRVWEELFPLDYQKCK\*

Sbjct 541 FKVSQRLFEDRFRVWEELFPLDYQKCK\* 568

GYSAG is even more unusual. GESAG is much more common.

**Email David (27/09/2021):**

protein alignments of the three Chr28 esterases from Hs and Hv.  You can see they are quite similar.  I don’t think there are any differences that could explain one species’ protein being an esterase and the other an acetyl transferase.  
  
The catalytic triad residues are marked in red.   
  
What is slightly unusual is the sequence around the active site serine. In many esterases it is GESAG.  But CXE5 and CXE21 have GCSAG, the same in both species.  So does the H. armigera and the of homolog CXE5 in Spodoptera littoralis and litura. CXE16 has GYSAG, the same in Spodoptera litura and littoralis, H. armigera.  
  
I can find the same cluster (the 3 esterases in tandem) in the NCBI genome sequence of Spodoptera litura (but it’s separated in two contigs 🡪 It is flanked by two other genes so we know it is the same cluster)  
  
>Hs-NRK-Pep  
MVSKRNEWIVIGISGVTCGGKTSLANQLKNALSPVHVFHQDKYFFPDDDPRHVRCAALAHNNYDVLSALDMDAMLRDVTRTLAGEDRAHVTCTQRDGRYSRDGKRFLVAEGFTVLNYPPILEMCDLRYYFVLELGECVARRVLRLYDPPDIDGYFEQCVWPEHLKYRAEIEKDKRVKIIDGTRGDAFDIVMADLKAMET\*  
  
>Hv-HDphosphohydrolase-Pep  
MANNWWREPDMGVMSAPVPSGEAQGCAAMRMWRNACNTLVAPDADDAWKQVVEASPPDSLWHTLRNCVAYQAGVWQNLLNKGMDDVIQPAAFKLAIVLRHTPSELTVKLLTELLHLHPQSQDLTSYILALLTDDEAKWCGHSSDSSPPSSPASTESSTDAGETDAAEATDIDTNEAVGGPADSATSAAATSEAAEAMPAVVSSTEVDSTTANIDTNMVESNVPDSTTADDPVEIAAAASLSGDSPEPSSSSGSDSESGSAAGTPAGPAPLRDLQLFGDCELAVLAASREEYDAHAKLVRAEYSKLTHDNYVELRIKVLNQFNQIPKLFHTPEFECFESAARENIEREICTLREHLLTGRRD\*

**In H. armigera,** the homolog of CXE16 and CXE5 are on locus 110380678, and CXE21 of esterase fe4-like (LOC110380661). These locus are flanked by the homologs of Hs-NRK (LOC110380665 ) and Hv- HDphosphohydrolase (LOC110380662) 🡪 same 3 CXE cluster present

CXE16 vs H. armigera

Query 1 MLCVVLLSVALACAGADRHEQH-----GAGTSPASPEPLARSVSGQFAGSWMTSRRGRRF 55

MLCV LLSVA+ACAGA H GAGT+PAS +P+ RSVSGQF GSWMTSRRGRRF

Sbjct 1 MLCVALLSVAVACAGAAHHHHVHDHQHGAGTNPASSDPVTRSVSGQFVGSWMTSRRGRRF 60

Query 56 QAYRGIRYAEPPVGELRFQPPKPILNYSREVDARQEGPACPQPTYNDYPVHEDCLRLNVY 115

QAYRGIRYAEPPVGELRFQPPK I NYSREVDARQEGPACPQP YNDYPVHEDCLRLNVY

Sbjct 61 QAYRGIRYAEPPVGELRFQPPKLITNYSREVDARQEGPACPQPVYNDYPVHEDCLRLNVY 120

Query 116 TPVTAGNDKKLPVVVFMHAGGFYSVSGRSDVAGPSHLLDRDLVLVTINYRLGSLGFLSTG 175

TP G KKLPV+VFMHAGGFYSVSGRSDVAGP HLLDRDLVLVTINYRLGSLGFLSTG

Sbjct 121 TPAE-GKAKKLPVLVFMHAGGFYSVSGRSDVAGPEHLLDRDLVLVTINYRLGSLGFLSTG 179

Query 176 DALAPGNNGFKDQVMALRWVKRNIAAFGGDPDLVTISGYSAGSFSVMLHMISPMSKGLFH 235

DALAPGNNGFKDQV AL+WV+RNIAAFGGDP+LVTI+GYSAGSFSVMLHMISPMS+GLFH

Sbjct 180 DALAPGNNGFKDQVAALKWVQRNIAAFGGDPELVTIAGYSAGSFSVMLHMISPMSRGLFH 239

Query 236 RAISMSGSPISQIEIPRHQRHLAERQARLLACPTDSSRAIIDCLRTKTSKEIGDSLTNMF 295

RA+SMSGSPISQIEIPRHQRHLAERQARLL+CPTDSSRAIIDCL+TK+SKEIGDSL MF

Sbjct 240 RAMSMSGSPISQIEIPRHQRHLAERQARLLSCPTDSSRAIIDCLKTKSSKEIGDSLDKMF 299

Query 296 DFGYDPVLLWVPIHEQDFGQEMFLPRQPLEALCSADLTQVPYIVSQTQDEFFWKALDVLR 355

DFGYDPVLLWVPIHEQDFGQEMFLPRQPL+ALCSADL QVPYI+SQTQDEFFWKALDVLR

Sbjct 300 DFGYDPVLLWVPIHEQDFGQEMFLPRQPLDALCSADLAQVPYIISQTQDEFFWKALDVLR 359

Query 356 DPAAYDSWRADWPGKGRVAFYLSGDNLTATAAANRLKQAYLGGKDIANDTATGDGFGKLY 415

+P+A++SWRADWPGK RVA YLSGDN TAT AANRLKQAYLGGKDIANDTATGDGFGKLY

Sbjct 360 NPSAFESWRADWPGKARVALYLSGDNATATTAANRLKQAYLGGKDIANDTATGDGFGKLY 419

Query 416 SDAIIGFGAHRLVNLMSRHLRQPLYYYEFGYIGNSSHYVDPDTKKPIAAAHHDDLLYLFN 475

SDAIIGFGAHRL+NLMSR RQPLYYYEFGYIGNSSHYVDPDTKKPIAAAHHDDLLYLFN

Sbjct 420 SDAIIGFGAHRLMNLMSRQSRQPLYYYEFGYIGNSSHYVDPDTKKPIAAAHHDDLLYLFN 479

Query 476 VSYSFPSIPPTDSKDSQMVDKMTALVYNFARYGDPNPKADTPELSGLSWPQYKPDERKYL 535

VSY+FPSIPP D+ DS MVDKMTALVYNFARYGDPNPK DTPEL+GLSWPQYKP +RKYL

Sbjct 480 VSYNFPSIPPADTNDSLMVDKMTALVYNFARYGDPNPKPDTPELAGLSWPQYKPADRKYL 539

Query 536 RVDSPFKVSQRLFEDRFRVWEELFPLDYQKCK 567

RVD+PF V ++LFEDRFRVWEELFPLDYQKCK

Sbjct 540 RVDAPFAVREKLFEDRFRVWEELFPLDYQKCK 571

CXE5 vs H. armigera

Query 1 MHLRALLCALCVAVCAHAHRHTTERANVPSSTLKPSEPVTRSVSGTFRGSWMETRRGRRF 60

M LRALLCALCVAVC HA++H TER NVPSST+ PS P+TR VSG FRGSWMETRRGRRF

Sbjct 597 MCLRALLCALCVAVCVHAYKHNTERTNVPSSTVAPSGPLTRGVSGAFRGSWMETRRGRRF 656

Query 61 QAYRGIRYAEPPIGPLRFQPPKLKLNYESVVDASEEGPACPLPAPPTYYVDEDCLTVNVY 120

QAYRGIRYAEPP+G LRFQPPKLKLNYE VDAS+EGPACPLPAPP YYVDEDCLTVNVY

Sbjct 657 QAYRGIRYAEPPVGALRFQPPKLKLNYEGEVDASKEGPACPLPAPPGYYVDEDCLTVNVY 716

Query 121 TPAHNTSQRLPVIFFIHPGGFYAMTGRSDLAGPHYLLDRDVVLVTINYRLGSLGFMSTGD 180

TPAHN S+ LPVIFFIHPGGFYA +GRSDLAGPHY+LDRDVVLVTINYRLGSLGFMSTGD

Sbjct 717 TPAHNASRPLPVIFFIHPGGFYAFSGRSDLAGPHYMLDRDVVLVTINYRLGSLGFMSTGD 776

Query 181 ANAPGNNGMKDQVAALKWVQRNIAAFGGDPDLVTITGCSAGSISVLLHMMSPMTKGLFHR 240

APGNNGMKDQVAALKWVQRNIAAFGGDP+LVTITGCSAGSISVLLHMMSPMTKGLFHR

Sbjct 777 VYAPGNNGMKDQVAALKWVQRNIAAFGGDPNLVTITGCSAGSISVLLHMMSPMTKGLFHR 836

Query 241 GIAMSASPISKEMTTLTHQRHLAVRQAEILNCPTHNSSAIVDCLFTKPWRELGDSLTKFW 300

GIAMSASPI+K MTTLTHQRHLAVRQAEILNCPTHNSSAIVDCLF KPW+ELGDSL KFW

Sbjct 837 GIAMSASPITKAMTTLTHQRHLAVRQAEILNCPTHNSSAIVDCLFKKPWKELGDSLPKFW 896

Query 301 EFGPGDPVGLWGPVVEPDFGQERYLTMNPLDAVKQGKMHTVPLIISQTTDEFFWKAFQV- 359

EFGPGDPVGLWGPVVEPDFGQERYLTMNPLDAVKQGKMHTVPLIISQTTDEFFWKA+QV

Sbjct 897 EFGPGDPVGLWGPVVEPDFGQERYLTMNPLDAVKQGKMHTVPLIISQTTDEFFWKAYQVL 956

Query 360 LNETLLKTMNDDWERIAPISFILPQENRAAAVRRLREVYLKGNKLVNDSESAKALGRLYG 419

LNETLLK+MNDDWE+IAPISF+LPQENRAAAV++L+EVY KGNK+ ND+ESAKALGRLYG

Sbjct 957 LNETLLKSMNDDWEKIAPISFMLPQENRAAAVKKLKEVYFKGNKIANDTESAKALGRLYG 1016

Query 420 DSVVGFGVHRMANLMCRHSTHPVWYYEFAYVGNNSHYEDPSGKPQGAAHHDDLLYLFTLS 479

DS+VGFGVHRMANLMCRHSTHPVWY EFAYVGNNSHYEDP+GKPQGAAHHDDLLY+FTLS

Sbjct 1017 DSIVGFGVHRMANLMCRHSTHPVWYSEFAYVGNNSHYEDPNGKPQGAAHHDDLLYVFTLS 1076

Query 480 YNFPTIELSSPHSHVVDEMTAMWYNFARYGDPNPRGDTPELGKLTWPAMTPDRRQFLHRG 539

YNFPTIE+SSPHSHVVDEMTA+WYNFARYGDPNPRGDTPELG LTWPAMTP+RR FLHRG

Sbjct 1077 YNFPTIEVSSPHSHVVDEMTAIWYNFARYGDPNPRGDTPELGNLTWPAMTPNRRLFLHRG 1136

Query 540 DQLVVRQNMFEDRFKVWEELYPIQY 564

D+L+++QNMFEDRFKVWEELYPI Y

Sbjct 1137 DKLLIQQNMFEDRFKVWEELYPIPY 1161

**In Bombyx mori,** the homolog of CXE21 and CXE5 is integument esterase 1 precursor on Chr28 while CXE16 homolog is integument esterase 2 precursor on Chr14. Esterase 1 locus (XP\_021202428.1) is flanked by the homologs of Hv-HDphosphohydrolase (XP\_037877060.1, LOC101743533) and Hs-NRK (XP\_004924360.1, LOC101743672) 🡪 Only one CXE present with a GCSAG catalytic triad (like CXE5 or CXE21)

**In Bombyx mandarina,** The homologs of Hs-NRK (LOC114246862) and Hv-HDphosphohydrolase (LOC114246860). In between there is an esterase E4-like (XP\_028035378, LOC114246859) 🡪 Only one CXE present with a GCSAG catalytic triad (like CXE5 or CXE21)

**In T. ni,** the homolog of CXE16 and CXE5 are on locus LOC113499569, and CXE21 of venom carboxylesterase-6-like isoform X2 (LOC113499325). These loci are flanked by the homologs of Hs-NRK (LOC113499192) and Hv-HDphosphohydrolase (LOC113499326) 🡪 same 3 CXE cluster present

CXE16 vs T.ni

Query 19 HEQHGAGTSPAS--PEPLARSVSGQFAGSWMTSRRGRRFQAYRGIRYAEPPVGELRFQPP 76

H G SP + EPL + +GQF GSWM SRRGRRFQAYRGIRYAEPPVGELRFQPP

Sbjct 22 HNSRGETGSPNTRTSEPLTVAPAGQFRGSWMESRRGRRFQAYRGIRYAEPPVGELRFQPP 81

Query 77 KPILNYSREVDARQEGPACPQPTYN-DYPVHEDCLRLNVYTPVTAGND-KKLPVVVFMHA 134

K I ++ EVDA Q+GPACPQP N DYPVHEDCLRLNVY P G D KLPVVVFMHA

Sbjct 82 KMITKHTGEVDATQDGPACPQPVLNKDYPVHEDCLRLNVYRP--DGKDVSKLPVVVFMHA 139

Query 135 GGFYSVSGRSDVAGPSHLLDRDLVLVTINYRLGSLGFLSTGDALAPGNNGFKDQVMALRW 194

GGFYSVSGRSDVAGP +LLDRD+VLVTINYRLG+LGFLSTGD LAPGNNGFKDQVMALRW

Sbjct 140 GGFYSVSGRSDVAGPHYLLDRDIVLVTINYRLGTLGFLSTGDKLAPGNNGFKDQVMALRW 199

Query 195 VKRNIAAFGGDPDLVTISGYSAGSFSVMLHMISPMSKGLFHRAISMSGSPISQIEIPRHQ 254

V+RNIA+FGGDP LVTI+GYSAGSFSVMLH ISPMSKGLFHRAISMSGSPISQIEIPRHQ

Sbjct 200 VQRNIASFGGDPHLVTIAGYSAGSFSVMLHTISPMSKGLFHRAISMSGSPISQIEIPRHQ 259

Query 255 RHLAERQARLLACPTDSSRAIIDCLRTKTSKEIGDSLTNMFDFGYDPVLLWVPIHEQDFG 314

RHLA +QA LL C T+SS A+IDCL+TKT KEIG+SL NMF++ YDPVLLW+P+ E DFG

Sbjct 260 RHLAVKQAELLNCTTESSEALIDCLKTKTYKEIGNSLDNMFEYNYDPVLLWMPVQELDFG 319

Query 315 QEMFLPRQPLEALCSADLTQVPYIVSQTQDEFFWKALDVLRDPAAYDSWRADWPGKGRVA 374

QE FLPRQPL ALC DL VPYIVSQTQDEFFWKA VL + A W D+P A

Sbjct 320 QERFLPRQPLAALCGGDLQPVPYIVSQTQDEFFWKAYGVLNNNGALLGWNIDFPSFANTA 379

Query 375 FYLSGDNLTATAAANRLKQAYLGGKD--IANDTATGDGFGKLYSDAIIGFGAHRLVNLMS 432

FYL D T+ A RLK AY D + ND+AT D GK+YSDAIIGFGAHRLVNL+S

Sbjct 380 FYLGDD--TSGRLATRLKGAYFPSSDSKLLNDSATVDALGKMYSDAIIGFGAHRLVNLLS 437

Query 433 RHLRQPLYYYEFGYIGNSSHYVDPDTKKPIAAAHHDDLLYLFNVSYSFPSIPPTDSKDSQ 492

RH +QP+YYYEFGYIGNSSHYVDP T +P AAAHHDDLLYLFNVS SFP+IP D+ DS+

Sbjct 438 RHSKQPVYYYEFGYIGNSSHYVDPITNEPTAAAHHDDLLYLFNVSASFPAIPAGDTHDSR 497

Query 493 MVDKMTALVYNFARYGDPNPKADTPELSGLSWPQYKPDERKYLRVDSPFKVSQRLFEDRF 552

MVDKMTALVYNFARYGDPNP +DTPEL+GL WPQYKP ER+YLR+D PF + QRLFEDRF

Sbjct 498 MVDKMTALVYNFARYGDPNPVSDTPELAGLKWPQYKPAERQYLRIDQPFSIDQRLFEDRF 557

Query 553 RVWEELFPLDYQKCK 567

+VWEELFP+DYQ CK

Sbjct 558 KVWEELFPIDYQTCK 57

CXE5 vs T.ni

Query 5 ALLCALCVAVCAHAHRHTTERANVP---SSTLKPSE--PVTRSVSGTFRGSWMETRRGRR 59

+L ALC AVC HAH+H + P +ST P+ PV RSVSG FRGSWME+RRGRR

Sbjct 596 CVLLALCAAVCVHAHQHKHQHTTEPARATSTSPPTAAGPVVRSVSGVFRGSWMESRRGRR 655

Query 60 FQAYRGIRYAEPPIGPLRFQPPKLKLNYESVVDASEEGPACPLPAPPTYYVDEDCLTVNV 119

FQAYRGIRYAEPP+G LRFQPPKLKL Y VDAS EGPACPLP PPTYYVDEDCLTVNV

Sbjct 656 FQAYRGIRYAEPPVGELRFQPPKLKLQYSGEVDASAEGPACPLPVPPTYYVDEDCLTVNV 715

Query 120 YTPAHNTSQRLPVIFFIHPGGFYAMTGRSDLAGPHYLLDRDVVLVTINYRLGSLGFMSTG 179

YTP N+S+ LPVIFFIH GGFY+MTGRSDLAGPHYLLDRDVVLVTINYR+GSLGFMSTG

Sbjct 716 YTPRTNSSKPLPVIFFIHAGGFYSMTGRSDLAGPHYLLDRDVVLVTINYRIGSLGFMSTG 775

Query 180 DANAPGNNGMKDQVAALKWVQRNIAAFGGDPDLVTITGCSAGSISVLLHMMSPMTKGLFH 239

DA APGNNGMKDQVAALKWVQRNIAAFGGDP+ VTITGCSAGSISV+LHM+SPM KGLFH

Sbjct 776 DAYAPGNNGMKDQVAALKWVQRNIAAFGGDPNSVTITGCSAGSISVMLHMISPMAKGLFH 835

Query 240 RGIAMSASPISKEMTTLTHQRHLAVRQAEILNCPTHNSSAIVDCLFTKPWRELGDSLTKF 299

R I+MSASPI+K +THQRHLAV QA+IL CPT+N+S I DCL KPWRE+GDSL +F

Sbjct 836 RAISMSASPINKG-PMVTHQRHLAVLQAQILKCPTNNTSVIYDCLMKKPWREIGDSLPQF 894

Query 300 WEFGPGDPVGLWGPVVEPDFGQERYLTMNPLDAVKQGKMHTVPLIISQTTDEFFWKAFQV 359

W FGPGDPVGLWGPVVEPDFGQERYL+ +PLD++K G+MHTVP IISQTTDEFFWKAF V

Sbjct 895 WSFGPGDPVGLWGPVVEPDFGQERYLSEDPLDSIKHGRMHTVPYIISQTTDEFFWKAFPV 954

Query 360 L-NETLLKTMNDDWERIAPISFILPQENRAAAVRRLREVYLKGNKLVNDSESAKALGRLY 418

L N TLLKTMN++WERIAPISF+LP ENR+AAV RLREVYL+ L ND ES ALGRLY

Sbjct 955 LQNATLLKTMNEEWERIAPISFLLPAENRSAAVARLREVYLRNKPLANDRESELALGRLY 1014

Query 419 GDSVVGFGVHRMANLMCRHSTHPVWYYEFAYVGNNSHYEDPSGKPQGAAHHDDLLYLFTL 478

GDSV+GFGVHR+ANLMCRHS HPVWYYEFAY+GNNSHY+DP GKP+GAAHHDDL+YLFTL

Sbjct 1015 GDSVIGFGVHRLANLMCRHSKHPVWYYEFAYIGNNSHYQDPDGKPRGAAHHDDLIYLFTL 1074

Query 479 SYNFPTIELSSPHSHVVDEMTAMWYNFARYGDPNPRGDTPELGKLTWPAMTPDRRQFLHR 538

SY FPTI LSS SHVVDEMTAMWYNFARYGDPNPRGDTPEL L WP MTP +R +L R

Sbjct 1075 SYAFPTIPLSSQDSHVVDEMTAMWYNFARYGDPNPRGDTPELAGLAWPTMTPRQRNYLRR 1134

Query 539 GDQLVVRQNMFEDRFKVWEELYPIQY 564

GDQL V QNMFEDRFKVW++LYPI+Y

Sbjct 1135 GDQLSVHQNMFEDRFKVWDDLYPIEY 1160

**In Pararge aegeria,** venom carboxylesterase-6-like, XP\_039747961.1 with a GCSAG motif and uncharacterized protein LOC120629510, XP\_039754431.2 are flanked by the homologs of Hs-NRK (XP\_039746984.1) and Hv-HDphosphohydrolase (XP\_039747974.1). Uncharacterized protein LOC120629510, XP\_039754431.2 is two esterases first part has a GCSAG motif and second part a GCSVG

Query 27 NVPSSTLKPSEPVTRSVSGTFRGSWMETRRGRRFQAYRGIRYAEPPIGPLRFQPPKLKLN 86

+VP S +PV S SG FRGS+ TRRGR+F++YRGI YAEPP+G LRFQPPKL L

Sbjct 38 DVPES----QQPVVVSPSGEFRGSYNTTRRGRQFESYRGIYYAEPPVGELRFQPPKLILE 93

Query 87 YESVVDASEEGPACPLPAPPTYYVDEDCLTVNVYTP-AHNTSQRLPVIFFIHPGGFYAMT 145

Y++ VDASE+GPACPL YY+DE+CLT+NVYTP + S+ LPVIFF+HPGGFYAMT

Sbjct 94 YKTPVDASEDGPACPLLVD-NYYMDENCLTINVYTPLKKDRSKPLPVIFFMHPGGFYAMT 152

Query 146 GRSDLAGPHYLLDRDVVLVTINYRLGSLGFMSTGDANAPGNNGMKDQVAALKWVQRNIAA 205

GRSDLAG HYLLDRD+VLVTINYRLGSLGF+STGD APGNNG KDQVAALKWVQRNIAA

Sbjct 153 GRSDLAGAHYLLDRDIVLVTINYRLGSLGFLSTGDELAPGNNGFKDQVAALKWVQRNIAA 212

Query 206 FGGDPDLVTITGCSAGSISVLLHMMSPMTKGLFHRGIAMSASPISKEMTTLTHQRHLAVR 265

FGGDP+ VTITGCSAGS SV+LHM+SPM+KGLFHRGI+MS SP++K T + HLAVR

Sbjct 213 FGGDPNSVTITGCSAGSTSVMLHMISPMSKGLFHRGISMSGSPVNKGPTP-DNLYHLAVR 271

Query 266 QAEILNCPTHNSSAIVDCLFTKPWRELGDSLTKFWEFGPGDPVGLWGPVVEPDFGQERYL 325

QAEILNCPT+NS I+DCL TK WRELG SL F+EFG DPV W P++E DFGQER+L

Sbjct 272 QAEILNCPTNNSKVIIDCLKTKHWRELGSSLPGFYEFG-NDPVSFWIPIIEKDFGQERFL 330

Query 326 TMNPLDAVKQGKMHTVPLIISQTTDEFFWKAFQVL-NETLLKTMNDDWERIAPISFILPQ 384

++P+DA+++GKMH VP IISQT DEFFW A VL N+TL MN +WE+IAPISF+L +

Sbjct 331 PIDPVDAIREGKMHAVPHIISQTEDEFFWMALTVLKNKTLTDKMNAEWEKIAPISFLLNR 390

Query 385 ENRAAAVRRLREVYLKGNKLVNDSESAKALGRLYGDSVVGFGVHRMANLMCRHSTHPVWY 444

EN A +RLR YL L ND+ESAK LG LY DS+ F VHRMANLMCRHS HPVWY

Sbjct 391 ENATYATQRLRTAYLHDKPLKNDAESAKNLGLLYQDSIESFPVHRMANLMCRHSPHPVWY 450

Query 445 YEFAYVGNNSHYEDP-SGKPQGAAHHDDLLYLFTLSYNFPTIELSSPHSHVVDEMTAMWY 503

Y+F+Y+GN+S YED + KP AAHHDDL+YLF LS+ FP I + +V+ MT +WY

Sbjct 451 YQFSYIGNHSFYEDSVTKKPVAAAHHDDLIYLFPLSFGFPGISTEGLDAVLVERMTGIWY 510

Query 504 NFARYGDPNPRGDTPELGKLTWPAMTPDRRQFLHRGDQLVVRQNMFEDRFKVWEELYPIQ 563

NFARYGDPNPRGD PEL L WPAM PD R++L G+ L + +NM E+R KVWEELYPI+

Sbjct 511 NFARYGDPNPRGDIPELEGLHWPAMKPDDRKYLRIGNDLTIHENMKEERIKVWEELYPIK 570

Query 564 Y 564

Y

Sbjct 571 Y 571

Query 3 LRALLCALCVAVCAHAHRHTTERANVPSS---TLKPSEPVTRSVSGTFRGSWMETRRGRR 59

+R C L V V A+AH + N +P +PV S G FRG + TRRGRR

Sbjct 578 MRTWFCVLLVTVGAYAHGYKHAHHNHGERHGHQAEPQQPVVVSPFGEFRGGYNVTRRGRR 637

Query 60 FQAYRGIRYAEPPIGPLRFQPPKLKLNYESVVDASEEGPACPLPAPPTYYVDEDCLTVNV 119

F+ YRGIRYAEPP+G LRFQPP+ +NY++ VDASE+GPACP P P YYVDEDCLT+NV

Sbjct 638 FETYRGIRYAEPPVGNLRFQPPQPIVNYKTPVDASEDGPACPQPTEPGYYVDEDCLTINV 697

Query 120 YTP-AHNTSQRLPVIFFIHPGGFYAMTGRSDLAGPHYLLDRDVVLVTINYRLGSLGFMST 178

YTP +N S LPVIF+IH GG+Y++TGRSDLAGPHYLLDRD+VLVTINYRL SLGF+ST

Sbjct 698 YTPLKNNRSTPLPVIFYIHAGGYYSVTGRSDLAGPHYLLDRDIVLVTINYRLSSLGFLST 757

Query 179 GDANAPGNNGMKDQVAALKWVQRNIAAFGGDPDLVTITGCSAGSISVLLHMMSPMTKGLF 238

GD APGNNG+KDQVAALKWVQRNIAAFGGDP+ VTI GCS GS+SV+LHM+SPM+KGLF

Sbjct 758 GDELAPGNNGLKDQVAALKWVQRNIAAFGGDPNSVTIAGCSVGSMSVMLHMISPMSKGLF 817

Query 239 HRGIAMSASPISKEMTTLTHQRHLAVRQAEILNCPTHNSSAIVDCLFTKPWRELGDSLTK 298

HRGI++S SP+ + + HLA +QA +LNCPT NS IVDCL TK WRELGDSL

Sbjct 818 HRGISISGSPVFA-FPSPDNLYHLAEKQARLLNCPTDNSKVIVDCLRTKSWRELGDSLDG 876

Query 299 FWEFGPGDPVGLWGPVVEPDFGQERYLTMNPLDAVKQGKMHTVPLIISQTTDEFFWKAFQ 358

F EF DP+ +W PVVE D+GQER+LTM PL+A+++GKMH VP IISQTTDEFFWKA+

Sbjct 877 FSEFA-YDPIMIWWPVVEKDYGQERFLTMEPLNAIREGKMHAVPHIISQTTDEFFWKAYT 935

Query 359 VL-NETLLKTMNDDWERIAPISFILPQENRAAAVRRLREVYLKGNKLVNDSESAKALGRL 417

VL N+TL MN +W+R+API+FI+P+EN A ++L YL + ND++SA LG L

Sbjct 936 VLENQTLTDQMNAEWDRVAPIAFIMPRENPRRATQKLFSEYLHNRPVQNDTDSATNLGLL 995

Query 418 YGDSVVGFGVHRMANLMCRHSTHPVWYYEFAYVGNNSHYEDP-SGKPQGAAHHDDLLYLF 476

Y D V F VHRMANLMCRHS HPVWYYEF+Y+G +SHY+DP + KP GAAHHDDL+YLF

Sbjct 996 YTDGVESFPVHRMANLMCRHSPHPVWYYEFSYIGAHSHYQDPVTKKPVGAAHHDDLIYLF 1055

Query 477 TLSYNFPTIELSSPHSHVVDEMTAMWYNFARYGDPNPRGDTPELGKLTWPAMTPDRRQFL 536

L FP I + +VD MTA+WYNFARYGDPNPRGD PEL L WPAM PD R++L

Sbjct 1056 ALR-AFPNIATEGRDAVLVDRMTAIWYNFARYGDPNPRGDVPELEGLEWPAMKPDERKYL 1114

Query 537 HRGDQLVVRQNMFEDRFKVWEELYPIQY 564

D L V N+ EDR KVWEELYPI+Y

Sbjct 1115 RISDDLTVHNNLKEDRIKVWEELYPIEY 1142

**In Maniola hyperantus,** Uncharacterized protein LOC117984000, XP\_034826539.1 and venom carboxylesterase-6-like, XP\_039747961.1 with a GCSVG motif are flanked by the homologs of Hs-NRK (XP\_034826285.1) and Hv-HDphosphohydrolase (XP\_034825973.1). Uncharacterized protein LOC117984000, XP\_034826539.1 is two esterases first part has a GCSAG motif and second part a GCSVG

uncharacterized protein LOC117984000, XP\_034826539.1 (10-580)

Query 3 LRALLCALCVAVCAHAHRHT---TERANVPSSTLKPSEPVTRSVSGTFRGSWMETRRGRR 59

++ C L VAV A A+ H T+ + +P +PV S +G FRG + TRRGRR

Sbjct 1 MKTCFCVLLVAVGASAYAHGSSHTQHNHGGQQGAEPQQPVVVSPAGEFRGGYNVTRRGRR 60

Query 60 FQAYRGIRYAEPPIGPLRFQPPKLKLNYESVVDASEEGPACPLPAPPTYYVDEDCLTVNV 119

F+ YRGIRYAEPP+G LRFQPPK L+Y+S VDASE+GPACP P P YY+DEDCLT+NV

Sbjct 61 FETYRGIRYAEPPVGNLRFQPPKPILSYKSAVDASEDGPACPQPTDPDYYIDEDCLTINV 120

Query 120 YTP-AHNTSQRLPVIFFIHPGGFYAMTGRSDLAGPHYLLDRDVVLVTINYRLGSLGFMST 178

YTP N S+ LPVIF+IH GG+Y++TGRSDLAGPHYLLDRD+VLVTINYRLGSLGF+ST

Sbjct 121 YTPLKKNRSKPLPVIFYIHAGGYYSVTGRSDLAGPHYLLDRDIVLVTINYRLGSLGFLST 180

Query 179 GDANAPGNNGMKDQVAALKWVQRNIAAFGGDPDLVTITGCSAGSISVLLHMMSPMTKGLF 238

GD APGNNG+KDQVAALKWVQRNIA+FGGDP VTI GCS GS+SV+LHM+SPM+KGLF

Sbjct 181 GDEVAPGNNGLKDQVAALKWVQRNIASFGGDPKSVTIAGCSVGSMSVMLHMISPMSKGLF 240

Query 239 HRGIAMSASPISKEMTTLTHQRHLAVRQAEILNCPTHNSSAIVDCLFTKPWRELGDSLTK 298

HRGI++S SPI + + HLA +QA +L+CPT NS IVDCL TK WRELGDSL

Sbjct 241 HRGISISGSPIFA-FPSPDNLYHLAEKQARLLDCPTDNSKVIVDCLKTKGWRELGDSLDG 299

Query 299 FWEFGPGDPVGLWGPVVEPDFGQERYLTMNPLDAVKQGKMHTVPLIISQTTDEFFWKAFQ 358

F EF DP+ +W PVVE D+GQER+LTM PLDA++QGKMH VP IISQTTDEFFWKAF

Sbjct 300 FSEFA-YDPIMIWWPVVEKDYGQERFLTMEPLDAIRQGKMHAVPHIISQTTDEFFWKAFT 358

Query 359 VL-NETLLKTMNDDWERIAPISFILPQENRAAAVRRLREVYLKGNKLVNDSESAKALGRL 417

+L N+TL MN +W+RIAPI+FI+P+EN A ++L YL L NDSESAK LG L

Sbjct 359 ILENQTLTDQMNAEWDRIAPITFIMPRENPRRATQKLFTEYLHNRPLQNDSESAKNLGLL 418

Query 418 YGDSVVGFGVHRMANLMCRHSTHPVWYYEFAYVGNNSHYEDP-SGKPQGAAHHDDLLYLF 476

Y D V F VHRMAN+MCRHS HPVWYYEF+Y+G +SHY DP + KP GAAHHDDL+YLF

Sbjct 419 YTDGVESFPVHRMANVMCRHSPHPVWYYEFSYIGAHSHYRDPETNKPVGAAHHDDLIYLF 478

Query 477 TLSYNFPTIELSSPHSHVVDEMTAMWYNFARYGDPNPRGDTPELGKLTWPAMTPDRRQFL 536

LS FP I + +VD MTA+WYNFARYGDPNPRGD PEL L WPAM P+ R++L

Sbjct 479 GLS-AFPNIATEGRDAILVDRMTAIWYNFARYGDPNPRGDLPELEGLQWPAMKPEDRKYL 537

Query 537 HRGDQLVVRQNMFEDRFKVWEELYPIQY 564

D L V N+ EDR K+WEELYPIQY

Sbjct 538 RIADNLTVHTNLKEDRIKIWEELYPIQY 565

uncharacterized protein LOC117984000, XP\_034826539.1 (593-1150)

Query 10 LCVAVCAHAHRHTTERANVPSSTLKPSEPVTRSVSGTFRGSWMETRRGRRFQAYRGIRYA 69

+CV VC++A HT N +++P +PV S+SG FRG + TRRGRRF+ YRGIRYA

Sbjct 593 VCVVVCSYAFGHTYTDYNNLGGSVEPQQPVVVSLSGEFRGGYNVTRRGRRFETYRGIRYA 652

Query 70 EPPIGPLRFQPPKLKLNYESVVDASEEGPACPLPAPPTYYVDEDCLTVNVYTP-AHNTSQ 128

EPP+G LRFQPP L LNYES VDASE+ PACP P P Y V+EDCLT+NVYTP +N S+

Sbjct 653 EPPVGNLRFQPPVLILNYESPVDASEDSPACPQPTKPGYNVNEDCLTINVYTPLKNNRSK 712

Query 129 RLPVIFFIHPGGFYAMTGRSDLAGPHYLLDRDVVLVTINYRLGSLGFMSTGDANAPGNNG 188

LPVIF+IH GGFY+++GRSD AGPHYLLDRD+VLVTINYRLGSLGF+STGD APGNNG

Sbjct 713 PLPVIFYIHGGGFYSLSGRSDKAGPHYLLDRDIVLVTINYRLGSLGFLSTGDELAPGNNG 772

Query 189 MKDQVAALKWVQRNIAAFGGDPDLVTITGCSAGSISVLLHMMSPMTKGLFHRGIAMSASP 248

+KDQVAALKWVQRNIA FGGDP+ VTITGCS GS SV+LHM+SPM+KGLFHR I+MS SP

Sbjct 773 LKDQVAALKWVQRNIAGFGGDPNSVTITGCSVGSKSVMLHMISPMSKGLFHRSISMSGSP 832

Query 249 ISKEMTTLTHQRHLAVRQAEILNCP-THNSSAIVDCLFTKPWRELGDSLTKFWEFGPGDP 307

+ M + + +LAV+QA +LNC T NS AI+DCL K WRELGDSL F EF +P

Sbjct 833 LYA-MPSPDNLYYLAVKQARLLNCSITDNSRAIIDCLKRKTWRELGDSLKGFNEFA-HNP 890

Query 308 VGLWGPVVEPDFGQERYLTMNPLDAVKQGKMHTVPLIISQTTDEFFWKAFQVL-NETLLK 366

+ +W PVVE D+GQER+LTM PLDA+++GKMH VP IISQTTDEFFWKAF +L N+T +

Sbjct 891 IVIWSPVVERDYGQERFLTMQPLDAIREGKMHAVPYIISQTTDEFFWKAFTILGNKTWTE 950

Query 367 TMNDDWERIAPISFILPQENRAAAVRRLREVYLKGNKLVNDSESAKALGRLYGDSVVGFG 426

MN +W+RIAPI+FI+P+EN + A ++L VYL + ND+ESAK LG L+ D V F

Sbjct 951 QMNAEWDRIAPITFIMPRENSSRATQKLFSVYLNNQPIRNDTESAKNLGLLFQDGVESFP 1010

Query 427 VHRMANLMCRHSTHPVWYYEFAYVGNNSHYEDP-SGKPQGAAHHDDLLYLFTLSYNFPTI 485

+HRMANLMCRHS +PVWYYEF Y+G +SHYEDP + KP G AHHDDL+YLF++S FP I

Sbjct 1011 IHRMANLMCRHSPYPVWYYEFNYIGTHSHYEDPVTKKPVGVAHHDDLIYLFSMS-AFPAI 1069

Query 486 ELSSPHSHVVDEMTAMWYNFARYGDPNPR--GDTPELGKLTWPAMTPDRRQFLHRGDQLV 543

+ +VD MT +WYNFARYGDPNPR G+ PEL L WPAM PD R++L D L

Sbjct 1070 STEGRDATLVDRMTGIWYNFARYGDPNPRGAGNLPELEGLQWPAMKPDDRKYLRIADDLT 1129

Query 544 VRQNMFEDRFKVWEELYPIQY 564

V N+ EDR KVWEELYPIQY

venom carboxylesterase-6-like isoform X1,XP\_034826278

Query 3 LRALLCALCVAVCA----HAHRHTTERANVPSSTLKPSEPVTRSVSGTFRGSWMETRRGR 58

++ C L VAV A H HT ++P +PV S +G FRG + TRRGR

Sbjct 1 MKTCFCVLLVAVGASAYAHGSSHTQHNHGGQQGAVEPQQPVVVSPAGEFRGGYNVTRRGR 60

Query 59 RFQAYRGIRYAEPPIGPLRFQPPKLKLNYESVVDASEEGPACPLPAPPTYYVDEDCLTVN 118

RF+ YRGIRYAEPP+G LRFQPPK L+Y+S VDASE+GPACP P P YY+DEDCLT+N

Sbjct 61 RFETYRGIRYAEPPVGNLRFQPPKPILSYKSAVDASEDGPACPQPTDPDYYIDEDCLTIN 120

Query 119 VYTP-AHNTSQRLPVIFFIHPGGFYAMTGRSDLAGPHYLLDRDVVLVTINYRLGSLGFMS 177

VYTP N S+ LPVIF+IH GG+Y++TGRSDLAGPHYLLDRD+VLVTINYRLGSLGF+S

Sbjct 121 VYTPLKKNRSKPLPVIFYIHAGGYYSVTGRSDLAGPHYLLDRDIVLVTINYRLGSLGFLS 180

Query 178 TGDANAPGNNGMKDQVAALKWVQRNIAAFGGDPDLVTITGCSAGSISVLLHMMSPMTKGL 237

TGD APGNNG+KDQVAALKWVQRNIA+FGGDP VTI GCS GS+SV+LHM+SPM+KGL

Sbjct 181 TGDEVAPGNNGLKDQVAALKWVQRNIASFGGDPKSVTIAGCSVGSMSVMLHMISPMSKGL 240

Query 238 FHRGIAMSASPISKEMTTLTHQRHLAVRQAEILNCPTHNSSAIVDCLFTKPWRELGDSLT 297

FHRGI++S SPI + + HLA +QA +L+CPT NS IVDCL TK WRELGDSL

Sbjct 241 FHRGISISGSPIFA-FPSPDNLYHLAEKQARLLDCPTDNSKVIVDCLKTKGWRELGDSLD 299

Query 298 KFWEFGPGDPVGLWGPVVEPDFGQERYLTMNPLDAVKQGKMHTVPLIISQTTDEFFWKAF 357

F EF DP+ +W PVVE D+GQER+LTM PLDA++QGKMH VP IISQTTDEFFWKAF

Sbjct 300 GFSEFA-YDPIMIWWPVVEKDYGQERFLTMEPLDAIRQGKMHAVPHIISQTTDEFFWKAF 358

Query 358 QVL-NETLLKTMNDDWERIAPISFILPQENRAAAVRRLREVYLKGNKLVNDSESAKALGR 416

+L N+TL MN +W+RIAPI+FI+P+EN A ++L YL L NDSESAK LG

Sbjct 359 TILENQTLTDQMNAEWDRIAPITFIMPRENPRRATQKLFTEYLHNRPLQNDSESAKNLGL 418

Query 417 LYGDSVVGFGVHRMANLMCRHSTHPVWYYEFAYVGNNSHYEDP-SGKPQGAAHHDDLLYL 475

LY D V F VHRMAN+MCRHS HPVWYYEF+Y+G +SHY DP + KP GAAHHDDL+YL

Sbjct 419 LYTDGVESFPVHRMANVMCRHSPHPVWYYEFSYIGAHSHYRDPETNKPVGAAHHDDLIYL 478

Query 476 FTLSYNFPTIELSSPHSHVVDEMTAMWYNFARYGDPNPRGDTPELGKLTWPAMTPDRRQF 535

F LS FP I + +VD MTA+WYNFARYGDPNPRGD PEL L WPAM P+ R++

Sbjct 479 FGLS-AFPNIATEGRDAILVDRMTAIWYNFARYGDPNPRGDLPELEGLQWPAMKPEDRKY 537

Query 536 LHRGDQLVVRQNMFEDRFKVWEELYPIQY 564

L D L V N+ EDR K+WEELYPIQY

Sbjct 538 LRIADNLTVHTNLKEDRIKIWEELYPIQY 566

**In Pieris rapae,** Uncharacterized protein LOC111001382, XP\_022126961.1 is flanked by the homologs of Hs-NRK (XP\_022126934.1) and Hv-HDphosphohydrolase (XP\_022126925.1). Uncharacterized protein LOC111001382, XP\_022126961.1 is two esterases first part has a GYSAG motif and second part a GCSAG

uncharacterized protein LOC111001382, XP\_022126961.1

Query 38 PVTRSVSGTFRGSWMETRRGRRFQAYRGIRYAEPPIGPLRFQPPKLKLNYESVVDASEEG 97

P + SG GS M +RRGR A+RGIRYAEPPIG LRF+ P L Y+ V+A+ EG

Sbjct 28 PKVVTPSGPIVGSLMTSRRGRLIHAFRGIRYAEPPIGELRFRRPILITKYKEDVNATAEG 87

Query 98 PACPLPAPPTYYVDEDCLTVNVYTPAH-NTSQRLPVIFFIHPGGFYAMTGRSDLAGPHYL 156

PACP A Y++DEDCL +NVY P + N++ LPV+ +IH GG+Y+++GRSD+AGPHYL

Sbjct 88 PACPQDADSGYFLDEDCLRLNVYKPGNRNSTSLLPVVVYIHAGGYYSVSGRSDVAGPHYL 147

Query 157 LDRDVVLVTINYRLGSLGFMSTGDANAPGNNGMKDQVAALKWVQRNIAAFGGDPDLVTIT 216

LD+D++LVTINYRLGSLGF+STGDA APGNNG KDQV AL+WV+RNI +FGGD + VT+

Sbjct 148 LDKDLLLVTINYRLGSLGFLSTGDAEAPGNNGFKDQVTALRWVKRNIQSFGGDANCVTLV 207

Query 217 GCSAGSISVLLHMMSPMTKGLFHRGIAMSASPISKEMTTLTHQRHLAVRQAEILNCPTHN 276

G SAG +SV LH +SPM+KGLFHR I+MS SP S ++ LAVRQA +++CP +

Sbjct 208 GYSAGGMSVALHTVSPMSKGLFHRAISMSGSPFS-QIPRSPDLLSLAVRQARLVDCPDSS 266

Query 277 SSAIVDCLFTKPWRELGDSLTKFWEFGPGDPVGLWGPVVEPDFGQERYLTMNPLDAVKQG 336

+ AIVDCL T PWR LG SL F +F DP+ +W PV+E DFGQER+L + +D++++G

Sbjct 267 TRAIVDCLRTVPWRRLGASLKGFRDFSI-DPIIIWRPVIERDFGQERFLAEDLMDSIREG 325

Query 337 KMHTVPLIISQTTDEFFWKAFQVLNETLLKTM-NDDWERIAPISFILPQ-ENRAAAVRRL 394

++ +VP I+SQT DEFFW+AF VLN ++ ++ N DW R++P++F+LP +RAAA+ L

Sbjct 326 RISSVPFIVSQTRDEFFWEAFLVLNNSIYTSLINSDWYRVSPMAFLLPSPSDRAAAI--L 383

Query 395 REVYLKGNKLVNDSESAKALGRLYGDSVVGFGVHRMANLMCRHSTHPVWYYEFAYVGNNS 454

RE Y+ L N+++ A+ LG LYGD+V+GFGVHR+ NL+ + + P + + F Y+GN S

Sbjct 384 REAYVGNQPLENNTKDARNLGNLYGDAVIGFGVHRLVNLLS-YQSRPFYRFVFEYIGNRS 442

Query 455 HYEDPSG-KPQGAAHHDDLLYLFTLSYNFPTIELSSPHSHV-VDEMTAMWYNFARYGDPN 512

HYEDP+ KP G AHHD+L+YLF+L F I S H V VD++T + YNFA GDPN

Sbjct 443 HYEDPATRKPVGVAHHDELIYLFSLRVAFEDIPPSPSHDSVMVDKLTTIVYNFACTGDPN 502

Query 513 PR-GDTPELGKLTWPAMTPDRRQFLHRGDQLVVRQNMFEDRFKVWEELYPIQY 564

P G +P TW TP L DQ + Q+ +++R+++WE LYPI Y

Sbjct 503 PLWGPSPP----TWTPFTPSNGTHLSISDQYTLGQDWYKERYRIWERLYPIDY 551

uncharacterized protein LOC111001382, XP\_022126961.1

Query 5 ALLCALCVAVCAHAHRHTTERAN-----------------VPSSTLKPSEPVTRSVSGTF 47

LLC V VCA AH H R + + PV RS SG F

Sbjct 582 VLLC--FVLVCAAAHDHKNNRDHELGGHGHKGHNHGHNHEHHDHDIVDGRPVVRSPSGEF 639

Query 48 RGSWMETRRGRRFQAYRGIRYAEPPIGPLRFQPPKLKLNYESVVDASEEGPACPLPAPPT 107

RG + TR+GR F+ YRGIRYAEPP+G LRFQPP+ +Y + VDA +GPACPLPA PT

Sbjct 640 RGGYNVTRKGRHFETYRGIRYAEPPVGELRFQPPRPITHYPTPVDARSDGPACPLPAAPT 699

Query 108 YYVDEDCLTVNVYTPAHNTSQRL-PVIFFIHPGGFYAMTGRSDLAGPHYLLDRDVVLVTI 166

Y VDEDCLT+NVYTP ++L PVIFFIHPGGFY MTGRSDLAGPHYLLDRD+VLVTI

Sbjct 700 YPVDEDCLTINVYTPLTTDRKKLLPVIFFIHPGGFYVMTGRSDLAGPHYLLDRDLVLVTI 759

Query 167 NYRLGSLGFMSTGDANAPGNNGMKDQVAALKWVQRNIAAFGGDPDLVTITGCSAGSISVL 226

NYRLGSLGF+STGDA APGNNG KD VAAL+WV RNI AFGGDPD VT+ GCSAGSIS +

Sbjct 760 NYRLGSLGFLSTGDAVAPGNNGFKDMVAALRWVNRNIRAFGGDPDNVTVAGCSAGSISAM 819

Query 227 LHMMSPMTKGLFHRGIAMSASPISKEMTTLTHQRHLAVRQAEILNCPTHNSSAIVDCLFT 286

L M+SPM GLFHR I+MS SP+ KE + QR+LAVRQAE+LNCPTH+S AIVDCL T

Sbjct 820 LMMVSPMATGLFHRVISMSGSPVGKEPLP-SEQRYLAVRQAELLNCPTHSSQAIVDCLKT 878

Query 287 KPWRELGDSLTKFWEFGPGDPVGLWGPVVEPDFGQERYLTMNPLDAVKQGKMHTVPLIIS 346

KPWR+LG+SL F+EFG DPVG+W V+EPDFGQER+LT P +A+++ K++ VP ++S

Sbjct 879 KPWRDLGNSLNGFYEFG-FDPVGIWTAVIEPDFGQERFLTSQPDEAIREKKLYAVPHLVS 937

Query 347 QTTDEFFWKAFQVL-NETLLKTMNDDWERIAPISFILPQENRAAAVRRLREVYLKGNKLV 405

QT DEFFW AF V N+TL + MN +WER+APISF+LP++N +R LR VYLK L

Sbjct 938 QTQDEFFWMAFTVTRNDTLREQMNSEWERVAPISFLLPRKNAEVPMRTLRRVYLKDRPLK 997

Query 406 NDSESAKALGRLYGDSVVGFGVHRMANLMCRHSTHPVWYYEFAYVGNNSHYEDP-SGKPQ 464

ND S LGRLY DSV G V+RMA LM HS PV+ YE+ Y+GN+SHYEDP + KPQ

Sbjct 998 NDETSEHDLGRLYVDSVEGLPVYRMAKLMTLHSPQPVYEYEWGYIGNHSHYEDPKTMKPQ 1057

Query 465 GAAHHDDLLYLFTLSYNFPTIELSSPHSHVVDEMTAMWYNFARYGDPNPRGDTPELGKLT 524

GAAHHD+L+YLFTL Y FP I L + +VD MTA+WYNFARYGDPNPRGDTPELG+L

Sbjct 1058 GAAHHDELIYLFTLPYRFPAIGLEGRDAVMVDRMTALWYNFARYGDPNPRGDTPELGELV 1117

Query 525 WPAMTPDRRQFLHRGDQLVVRQNMFEDRFKVWEELYPIQY 564

WP MT ++ + +R+++F+D+ VW+ELYP++Y

Sbjct 1118 WPKMTSHDMTYVRIDNDFSLRKDLFKDKVGVWDELYPMKY 1157

**In Bicyclus anynana,** Venom carboxylesterase-6-like (XP\_023940808.1) and uncharacterized protein LOC112047778, XP\_023940786.1 are flanked by the homologs of Hv-HDphosphohydrolase (XP\_023940825.1) and Hs-NRK (XP\_023940802.1). Uncharacterized protein LOC112047778, XP\_023940786.1 is two esterases first part has a GCSVG motif and second part a GTSAG

uncharacterized protein LOC112047778, XP\_023940786.1

Query 34 KPSEPVTRSVSGTFRGSWMETRRGRRFQAYRGIRYAEPPIGPLRFQPPKLKLNYESVVDA 93

+P +P+ S SG FRGS+ TRRGR+F++YRGI YAEPP+G LRFQPPKL L Y++ VDA

Sbjct 41 EPKQPIVVSPSGEFRGSYNTTRRGRQFESYRGIYYAEPPVGELRFQPPKLILEYKTPVDA 100

Query 94 SEEGPACPLPAPPTYYVDEDCLTVNVYTP-AHNTSQRLPVIFFIHPGGFYAMTGRSDLAG 152

SE+GPACPL APP Y VDE+CLT+NVYTP + S+ LPVI FIHPGGFY+MTGRSD+ G

Sbjct 101 SEDGPACPLIAPPGYPVDENCLTINVYTPLKKDRSKPLPVILFIHPGGFYSMTGRSDIPG 160

Query 153 PHYLLDRDVVLVTINYRLGSLGFMSTGDANAPGNNGMKDQVAALKWVQRNIAAFGGDPDL 212

HY+LDRDVVLVTINYRLGSLGF+STGD APGNNG KDQVAALKWVQRNIA+FGGDP+

Sbjct 161 AHYMLDRDVVLVTINYRLGSLGFLSTGDELAPGNNGFKDQVAALKWVQRNIASFGGDPNC 220

Query 213 VTITGCSAGSISVLLHMMSPMTKGLFHRGIAMSASPISKEMTTLTHQRHLAVRQAEILNC 272

VTITG SAGSISV+LHM+SPM+KGLFHRGI+MS SP K T R LAVRQAEILNC

Sbjct 221 VTITGTSAGSISVMLHMISPMSKGLFHRGISMSGSPFYKGPTPDNMYR-LAVRQAEILNC 279

Query 273 PTHNSSAIVDCLFTKPWRELGDSLTKFWEFGPGDPVGLWGPVVEPDFGQERYLTMNPLDA 332

PT NS I+DCL TKPW ELG SL F+EFG DPVG+W P++E DFGQER++ ++P+DA

Sbjct 280 PTDNSKVIIDCLKTKPWMELGTSLRGFYEFG-FDPVGIWMPIIERDFGQERFMPIDPVDA 338

Query 333 VKQGKMHTVPLIISQTTDEFFWKAFQVL-NETLLKTMNDDWERIAPISFILPQENRAAAV 391

+++GKMH VP IISQT DEFFW AF VL N+TL TMN +WER+APISF+L +EN

Sbjct 339 IREGKMHAVPHIISQTEDEFFWMAFTVLKNKTLTDTMNAEWERVAPISFMLNRENATYPA 398

Query 392 RRLREVYLKGNKLVNDSESAKALGRLYGDSVVGFGVHRMANLMCRHSTHPVWYYEFAYVG 451

RRLR YL L ND+ESAK LG LY DS+ F VHRMANLMCRHS HPVWYY F+YVG

Sbjct 399 RRLRTAYLHDKPLKNDAESAKNLGLLYQDSIESFPVHRMANLMCRHSPHPVWYYLFSYVG 458

Query 452 NNSHYEDP-SGKPQGAAHHDDLLYLFTLSYNFPTIELSSPHSHVVDEMTAMWYNFARYGD 510

++S YEDP + KP AAHHDDL+YLF LS+ FP I + +VD MT MWYNFAR+GD

Sbjct 459 DHSFYEDPVTKKPVAAAHHDDLIYLFPLSFRFPGISTEGRDAVLVDRMTGMWYNFARFGD 518

Query 511 PNPRGDTPELGKLTWPAMTPDRRQFLHRGDQLVVRQNMFEDRFKVWEELYPIQ 563

PNPRGD PEL L WPAM P+ R++L D+ V NM E+R ++WEELYPI+

Sbjct 519 PNPRGDVPELEGLQWPAMKPEDRKYLRIDDEWTVHSNMKEERMQIWEELYPIK 571

uncharacterized protein LOC112047778, XP\_023940786.1

Query 3 LRALLCALCVAVCAHAHRHTTERANVPS--STLKPSEPVTRSVSGTFRGSWMETRRGRRF 60

L ++ +CV+V A++ HT S ++ +P S SG FRG + TRRGR+F

Sbjct 582 LNYIIVLVCVSVGAYSQDHTLTHHEGQSLGEPIESHQPSVLSPSGEFRGGYNVTRRGRKF 641

Query 61 QAYRGIRYAEPPIGPLRFQPPKLKLNYESVVDASEEGPACPLPAPPTYYVDEDCLTVNVY 120

+ YRGIRYAEPPIG LRFQPPKL L Y+S VDASE+GPACP P P YYVDEDCLT+NVY

Sbjct 642 ETYRGIRYAEPPIGNLRFQPPKLILYYKSPVDASEDGPACPQPTRPGYYVDEDCLTINVY 701

Query 121 TPAH-NTSQRLPVIFFIHPGGFYAMTGRSDLAGPHYLLDRDVVLVTINYRLGSLGFMSTG 179

TP N SQ LPVIF+IH GGFY+++GRSD AGPHYLLDRD+VLVTINYRLGSLGF+STG

Sbjct 702 TPLKANRSQPLPVIFYIHGGGFYSLSGRSDRAGPHYLLDRDIVLVTINYRLGSLGFLSTG 761

Query 180 DANAPGNNGMKDQVAALKWVQRNIAAFGGDPDLVTITGCSAGSISVLLHMMSPMTKGLFH 239

D APGNNG+KDQVAALKWVQRNIAAFGGDP+ VTI GCS GS SV+LHM+SPM+KGLFH

Sbjct 762 DELAPGNNGLKDQVAALKWVQRNIAAFGGDPNSVTIAGCSVGSKSVMLHMISPMSKGLFH 821

Query 240 RGIAMSASPISKEMTTLTHQRHLAVRQAEILNCPTHNSSAIVDCLFTKPWRELGDSLTKF 299

RGI+MS SP+ + + HLA +QA +LNCP+ NS IVDCL TK RELGDSL F

Sbjct 822 RGISMSGSPLYPSPSP-GNLYHLAEKQARVLNCPSDNSKVIVDCLKTKTARELGDSLKGF 880

Query 300 WEFGPGDPVGLWGPVVEPDFGQERYLTMNPLDAVKQGKMHTVPLIISQTTDEFFWKAFQV 359

EF +P+ +W PVVE DFGQER+LTM PLDA+++GKMH VP IISQT DEFFWKAF +

Sbjct 881 NEFA-HNPIVIWSPVVEKDFGQERFLTMQPLDAIREGKMHAVPHIISQTRDEFFWKAFTI 939

Query 360 L-NETLLKTMNDDWERIAPISFILPQENRAAAVRRLREVYLKGNKLVNDSESAKALGRLY 418

L N+T + MN +W+RIAPI+FILP+EN + A ++L +YL L NDSESAK LG L+

Sbjct 940 LGNKTYMDQMNAEWDRIAPITFILPRENSSHATQKLFSMYLNNQPLRNDSESAKNLGLLF 999

Query 419 GDSVVGFGVHRMANLMCRHSTHPVWYYEFAYVGNNSHYEDP-SGKPQGAAHHDDLLYLFT 477

D V F +HRMANLMCRHS HPVWYYEF+Y+G +SHYEDP + KP G AHHDDL+YLFT

Sbjct 1000 QDGVESFPIHRMANLMCRHSPHPVWYYEFSYIGAHSHYEDPVTKKPLGVAHHDDLIYLFT 1059

Query 478 LSYNFPTIELSSPHSHVVDEMTAMWYNFARYGDPNPRGDTPELGKLTWPAMTPDRRQFLH 537

LS FP I + +VD MTA+WYNFARY DPNPRGD PEL L WPAMTPD R++L

Sbjct 1060 LS-AFPAISTKGRDATLVDRMTAIWYNFARYSDPNPRGDLPELEDLEWPAMTPDDRKYLR 1118

Query 538 RGDQLVVRQNMFEDRFKVWEELYPIQY 564

D L V N+ EDR KVWEELYPIQY

Sbjct 1119 VSDDLTVHTNLNEDRIKVWEELYPIQY 1145

**In Spodoptera litura,** the homolog of CXE21, CXE5, and CX16 are at the beginning of Chr5, two GCSAG (XP\_022837498.1 and XP\_022837432.1) one GYSAG (XP\_022837632.1 or

XP\_022837633.1). The homologs of Hs-NRK (XP\_022837637.1, LOC111364822) is close by but the homolog of Hv-HDphosphohydrolase (XP\_022837568.1, LOC111364767) is at the other end of Chr5

**In Spodoptera frugiperda,** there is three co-esterase uncharacterized protein LOC118279115, XP\_035454541.1 (divided in two with a GYSAG and GCSAG motif) and esterase E4-like isoform X2 (XP\_035454711.1) with GCSAG in between the homologs of Hs-NRK (LOC118279213) and Hv-HDphosphohydrolase (LOC118279306) on Chr27 🡪 similar CXE cluster

uncharacterized protein LOC118279115, XP\_035454541.1

Query 7 LCALCVAVCAHAHRHTTERANVPSSTLK--PSEPVTRSVSGTFRGSWMETRRGRRFQAYR 64

L +C A+ A H + +T++ P+ PV RSVSG FRGSWM +RRGR+F+AYR

Sbjct 5 LVVVCGALAAAGRHHHHVEHDHQQATVESPPAAPVVRSVSGQFRGSWMTSRRGRQFEAYR 64

Query 65 GIRYAEPPIGPLRFQPPKLKLNYESVVDASEEGPACPLPAPPTYYVDEDCLTVNVYTPAH 124

GIRYA+PP+G LRFQPP+L NY S VD S++GPACP P Y V EDCL +NVYTP H

Sbjct 65 GIRYAQPPVGELRFQPPQLIENYPSEVDGSQDGPACPQPTVNDYPVHEDCLRLNVYTPDH 124

Query 125 NTSQRLPVIFFIHPGGFYAMTGRSDLAGPHYLLDRDVVLVTINYRLGSLGFMSTGDANAP 184

+ + LPV+ F+H GGFY+++GRSD+AGP +LLDRD+VLVTINYRLGSLGF+STGD AP

Sbjct 125 QSKKPLPVVVFMHAGGFYSVSGRSDVAGPQHLLDRDLVLVTINYRLGSLGFLSTGDKYAP 184

Query 185 GNNGMKDQVAALKWVQRNIAAFGGDPDLVTITGCSAGSISVLLHMMSPMTKGLFHRGIAM 244

GNNG KDQVAAL+WVQRNIAAFGGDP+LVTI+G SAGS SV+LHM+SPM+KGLFHR I+M

Sbjct 185 GNNGFKDQVAALRWVQRNIAAFGGDPNLVTISGYSAGSFSVMLHMISPMSKGLFHRAISM 244

Query 245 SASPISKEMTTLTHQRHLAVRQAEILNCPTHNSSAIVDCLFTKPWRELGDSLTKFWEFGP 304

S SPIS ++ HQRHLA RQA++L CPT +S AI+DCL TK ++LGDSL K ++FG

Sbjct 245 SGSPIS-QIVNPRHQRHLAERQAKLLQCPTDSSKAIIDCLKTKTSKQLGDSLDKMFDFG- 302

Query 305 GDPVGLWGPVVEPDFGQERYLTMNPLDAVKQGKMHTVPLIISQTTDEFFWKAFQVL-NET 363

DPV LW P+ E DFGQE +L PL AV G + VP I+SQT DEFFWKA VL N

Sbjct 303 YDPVLLWVPIHEQDFGQEMFLPRQPLAAVCAGGL-DVPYIVSQTQDEFFWKALDVLRNPQ 361

Query 364 LLKTMNDDWERIAPISFILP----QENRAAAVRRLREVYLKGNKLVNDSESAKALGRLYG 419

+ DW +A I+ L + A RL++ YL G + ND+ + G+LY

Sbjct 362 AFEKWRADWPGLAKIALYLTGAGDNSSITTAANRLKQAYLGGKDIANDTATGDGFGKLYS 421

Query 420 DSVVGFGVHRMANLMCRHSTHPVWYYEFAYVGNNSHYEDPSGK-PQGAAHHDDLLYLFTL 478

D+++GFG HR+ NL R S PV+YYEF +VGN+SHY DP+ K P AAHHDDLLYLF +

Sbjct 422 DAIIGFGAHRLVNLAARQSARPVYYYEFGHVGNSSHYVDPATKRPIAAAHHDDLLYLFNV 481

Query 479 SYNFPTIELS-SPHSHVVDEMTAMWYNFARYGDPNPRGDTPELGKLTWPAMTPDRRQFLH 537

SY+FP+I S S S +VD+MTA++YNFAR+GDPN RGDTPEL ++WP PD R++L

Sbjct 482 SYSFPSIPPSDSNDSRMVDKMTAIFYNFARHGDPNNRGDTPELADMSWPQFKPDERKYLR 541

Query 538 RGDQLVVRQNMFEDRFKVWEELYPIQY 564

V+ N+FE+RFKVWEEL+P+ Y

Sbjct 542 VDTPFSVQSNLFEERFKVWEELFPLDY 568

uncharacterized protein LOC118279115, XP\_035454541.1

Query 3 LRALLCALCVA-VCAHAHRHTTERANVPSSTLKPSEPVTRSVSGTFRGSWMETRRGRRFQ 61

L ALLC LCV V AH H+H TE S K + P+ SVSGTFRGSWMETRRGRRFQ

Sbjct 586 LYALLCVLCVCTVHAHQHKHETE-----SVKGKSAAPLAHSVSGTFRGSWMETRRGRRFQ 640

Query 62 AYRGIRYAEPPIGPLRFQPPKLKLNYESVVDASEEGPACPLPAPPTYYVDEDCLTVNVYT 121

AYRGIRYA+PP+G LRFQPPKLKL YE VVDASEEGPACPLPAPP+YYVDEDCLTVNVYT

Sbjct 641 AYRGIRYADPPVGALRFQPPKLKLQYEGVVDASEEGPACPLPAPPSYYVDEDCLTVNVYT 700

Query 122 PAHNTSQRLPVIFFIHPGGFYAMTGRSDLAGPHYLLDRDVVLVTINYRLGSLGFMSTGDA 181

P HN++++LPVIFFIHPGGFYAMTGRSDLAGPHYLLDRDVVLVTINYRLGSLGFMSTGDA

Sbjct 701 PTHNSTKKLPVIFFIHPGGFYAMTGRSDLAGPHYLLDRDVVLVTINYRLGSLGFMSTGDA 760

Query 182 NAPGNNGMKDQVAALKWVQRNIAAFGGDPDLVTITGCSAGSISVLLHMMSPMTKGLFHRG 241

APGNNGMKDQVAALKWVQRNIAAFGGDP+LVTITGCSAGSISV+LHM+SPM KGLFHRG

Sbjct 761 LAPGNNGMKDQVAALKWVQRNIAAFGGDPNLVTITGCSAGSISVILHMISPMAKGLFHRG 820

Query 242 IAMSASPISKEMTTLTHQRHLAVRQAEILNCPTHNSSAIVDCLFTKPWRELGDSLTKFWE 301

IAMSASP+ K + +THQRHLAVRQA+ILNCPT NSSAIVDCL TKPWRELGDSL KFWE

Sbjct 821 IAMSASPVGKGVVAITHQRHLAVRQAQILNCPTDNSSAIVDCLMTKPWRELGDSLPKFWE 880

Query 302 FGPGDPVGLWGPVVEPDFGQERYLTMNPLDAVKQGKMHTVPLIISQTTDEFFWKAFQVL- 360

FG GDPVGLWGPVVEPDFGQERYL +NPLDA+K+GKMHTVPLI+SQTTDEFFWKAF VL

Sbjct 881 FGEGDPVGLWGPVVEPDFGQERYLPINPLDAIKEGKMHTVPLIVSQTTDEFFWKAFPVLQ 940

Query 361 NETLLKTMNDDWERIAPISFILPQENRAAAVRRLREVYLKGNKLVNDSESAKALGRLYGD 420

NETLLK MN++WER+APISF+LP+ENRAAAV++LREVYLKG KL ND S KALG+LYGD

Sbjct 941 NETLLKAMNEEWERVAPISFMLPKENRAAAVQKLREVYLKGKKLANDKPSEKALGQLYGD 1000

Query 421 SVVGFGVHRMANLMCRHSTHPVWYYEFAYVGNNSHYEDPSGKPQGAAHHDDLLYLFTLSY 480

SV+GFGVHR+ANLMCRHS HPVWY EFAY+GNNSHYED GKPQGAAHHDDLLYLFTLSY

Sbjct 1001 SVIGFGVHRLANLMCRHSKHPVWYSEFAYIGNNSHYEDSHGKPQGAAHHDDLLYLFTLSY 1060

Query 481 NFPTIELSSPHSHVVDEMTAMWYNFARYGDPNPRGDTPELGKLTWPAMTPDRRQFLHRGD 540

FP I S HSHVVDEMTA+WYNFARYGDPNPRGDTPELG+L WPAMT R++LHRG

Sbjct 1061 RFPVISPDSQHSHVVDEMTALWYNFARYGDPNPRGDTPELGRLRWPAMTAADRRYLHRGP 1120

Query 541 QLVVRQNMFEDRFKVWEELYPIQY 564

QL VRQN+FEDRFKVW++LYPIQY

Sbjct 1121 QLQVRQNLFEDRFKVWDDLYPIQY 1144

**In Papilio xuthus,** the homolog of CXE21, CXE5, and CX16 are are not in between the homologs of Hs-NRK (LOC106118683) is on Chr28 and Hv-HDphosphohydrolase (LOC106118721). The sequences between those two genes are uncharacterized proteins 🡪 Not sure if we can claim it’s the same cluster then

**In Manduca sexta,** the homologs of Hs-NRK (LOC115446250) and Hv-HDphosphohydrolase (LOC115450249) are both on Chr5 but very far apart from each other 🡪 I don’t think we can claim it’s the same cluster then

**In Ostrinia furnacalis** First one venom carboxylesterase 6 (XP\_028163836.1) with a GCSAG motif and then another venom carboxylesterase 6 (XP\_028163835.1) with a GVSAG motif between the homologs of Hs-NRK (XP\_028163877.1, LOC114355295) and Hv-HDphosphohydrolase (XP\_028163891.1, LOC114355305)

**In Vanessa tameamea,** no characterized esterases in between the homologs of Hs-NRK (XP\_026498129.1) is on Chr28 and Hv-HDphosphohydrolase (XP\_026498123.1).

**In Amyelois transitella,** no characterized esterases in between the homologs of Hs-NRK (XP\_013198012.1) is on Chr28 and Hv-HDphosphohydrolase (XP\_013198005.1).