

CCTTGGGGCATGTCGTCTTCTTGCTCTGCCAGTTCGTTCCGGTGTGCGCACAGCGTCATGGTTGTTCTCCG
TGCGGCTGTCCCAGCGCAGATTGTGCGAGACGATTGTTGGAACGGTCCCCGTCCTGTGGCAGCCCTCCA
TGCCGTCCGGGCAAGCCCCGACGAACGCCGTGAGCACGAGCCTGTGTACGAGCGCGTTTTCCACCAGGTA
GGCCACCTTTAGGTAGCCATTTCGGCATCACGGACTGAGACAAGACCCGCCCGCATATACCGGCGCC
CGCGCCCATCGCACTCGATCTGCCGATCCAACTCCGCACGCGACCTCCGTGCGGAGACCTCGTACAGCC
CCTCGTATCCAAC TACCGCAACCAGGCTTCAACCATGTTCCCATTTTAGTCGAAAGGTGGGCCCTGAA
CCCCATGACAAACCGGCCTCCGTAT

AACGAG

TTTCCGATCTGGTCTGAACAACAACAGCGCCCGCAGCGGCAAGCCCACCATGTTCTGATCCACACCCAA
GAAGGCGGCGGGGACGCTGCCGCCGAGAACCTGGCGAAGTGGTTCCAGAACGGCAACGGCGTCTCG
TACCACTACACGATCTCCAAGCGTCCGATGGTGGTGTGACGGTGGTTCGATTGCGTTCGACACTGACCGC
GCCGCCTGGTCTGTGGGCAACGCCAACAGCATCAGCATCAACCTGTGCTTCGCGGGGTTCGCGAGCATCC
TGGATGCGGGATCAGTGGATGAAGCAGTCCAACGCAATCGACGTCGCAGCATACTCGCGGTGCAGGAC
GCGAAGAAGTACGGCTTACCCCGCTCGTGGTGCCACCGCCGTATACGAATGGGCGACCTGGCATCTCG
GACCACCGGTGGGTGACCGACGTGTTCAAGTGGGGCACTCACACCGACGTCGGAGACTGGTTCCCGTGG
GACTACTTCGCCGAACGGGTCAACCACTGGGCCAACGGTGGCAAGACCGAGCCTGAACCGCCCAAGGTG
AAACGCTTCCCGGACGACTGGACCGACCGGAAATCCTCGTTCGAGATTCTGCGGCAACTGCGCGGATAC
AACCTCACTGGCTGGCCGCAGCTCGGCGGAAAAACCTCGTGGACGCGGTAGCAGA ACTGTTGGGCCAC
TGA

Cut out the HNH insertion and one of the duplicated copies of AACGAG:

GTGACCACGAAAGATCAAGTCGCCCAAATCACCATCGCCGAAGCCAAGGCGCGGGCTACACCCGCAGC
GAATGCCTGGCGGTCATGTCCACCTTCTACCAAGAGTCCGGCTGGAACGACACCATCTGGGACCCACC
CACACCACCTACGGCATTGCCAGCAGGACGGCTCTACCCACACCGCTTCGACGGTGCCGCAGCCAA
ATCAAAGGCTTCTTCGACAAGCTCGACGTGTGGCGCGCAAACCCGGTGCCAGCACCGATATATGGCTG
AACATCTGCTGGATGCAGCAGGCCCCCAACTGGCCCAGCGCTGACTACTGGTACGCCAACGGCCGCCG
GCCTACCTCACGAAATCAAGTCACGCATCGCCACCGTCACCCCATACCTCGACAAGTACTGGCCCACC
ACTGGAGGTACCGCCGTGCCCGACGAACCACGCCCCGACTTC

AACGAG

TTTCCGATCTGGTCTGAACAACAACAGCGCCCGCAGCGGCAAGCCCACCATGTTCTGATCCACACCCAA
GAAGGCGGCGGGGACGCTGCCGCCGAGAACCTGGCGAAGTGGTTCCAGAACGGCAACGGCGTCTCG
TACCACTACACGATCTCCAAGCGTCCGATGGTGGTGTGACGGTGGTTCGATTGCGTTCGACACTGACCGC
GCCGCCTGGTCTGTGGGCAACGCCAACAGCATCAGCATCAACCTGTGCTTCGCGGGGTTCGCGAGCATCC
TGGATGCGGGATCAGTGGATGAAGCAGTCCAACGCAATCGACGTCGCAGCATACTCGCGGTGCAGGAC
GCGAAGAAGTACGGCTTACCCCGCTCGTGGTGCCACCGCCGTATACGAATGGGCGACCTGGCATCTCG
GACCACCGGTGGGTGACCGACGTGTTCAAGTGGGGCACTCACACCGACGTCGGAGACTGGTTCCCGTGG
GACTACTTCGCCGAACGGGTCAACCACTGGGCCAACGGTGGCAAGACCGAGCCTGAACCGCCCAAGGTG
AAACGCTTCCCGGACGACTGGACCGACCGGAAATCCTCGTTCGAGATTCTGCGGCAACTGCGCGGATAC

AACCTCACTGGCTGGCCGCGAGCTCGGCGGAAAAACCCTCGTGGACGCGGTAGCAGAACTGTTGGGCCAC
TGA

Is this an ORF? Yes. (submitted sequence to a DNA translation program
<https://web.expasy.org/translate/>)

5'3' Frame 1

VTTKDQVAQITIAEAKARGYTRSECLAVMSTFYQESGWNDTIWDPHTTTYGIAQQDGSYPHREFDGGAAQIKGFFDKLDV
WRAKPGASTDIWLNICWMQQAPNWPSADYWYANGRRAYLTEIKSRIATVTPYLDKYWPTTGGTAVPDEPRPDFNEFFPIW
SNNNSARSGKPTMFLIHTQEAGGGDAAAENLAKWFQNGNGVSYHYTISQASDGGVTVVDCVDTDRAAWSVGNANSISIN
LCFAGSRASWMRDQWMKQSNADVAAYLAVQDAKKYGFTPLVPPPYTNGRPGISDHRWVTDVFKWGTHTDVGDFWFPWD
YFAERVNHWANGGKTEPEPPKVKRFPDDWTDREILVEILRQLRGYNLTGWPQLGGKTLVDAVAELLGH-

How does this protein compare to others?

BLASTp = identical to lysin A from MilleniumForce and others. Blue shows the insertion point /
break point in native gene.

>MilleniumForce_31, lysin A, 384
Length = 384

Score = 824 bits (2128), Expect = 0.0
Identities = 381/384 (99%), Positives = 384/384 (100%)

Query: 1 VTTKDQVAQITIAEAKARGYTRSECLAVMSTFYQESGWNDTIWDPHTTTYGIAQQDGSYP 60
+TTKDQVAQITIAEAKARGYTRSECLA+MSTFYQESGWNDTIWDPHTTTYGIAQQDGSYP
Sbjct: 1 MTTKDQVAQITIAEAKARGYTRSECLAIMSTFYQESGWNDTIWDPHTTTYGIAQQDGSYP 60

Query: 61 HRFDGGAAQIKGFFDKLDVWRAKPGASTDIWLNICWMQQAPNWPSADYWYANGRRAYLTE 120
HRFDGGAAQIKGFFDKLDVWRAKPGASTDIWLNICWMQQAPNWPSADYWYANGRRAYLTE
Sbjct: 61 HRFDGGAAQIKGFFDKLDVWRAKPGASTDIWLNICWMQQAPNWPSADYWYANGRRAYLTE 120

Query: 121 IKSRIATVTPYLDKYWPTTGGTAVPDEPRPDFNEFFPIWSNNNSARSGKPTMFLIHTQEAG 180
IKSRIATVTPYLDKYWPTTGGTAVPDEPRPDFNEFFPIWSNNNSARSGKPTMFLIHTQEAG
Sbjct: 121 IKSRIATVTPYLDKYWPTTGGTAVPDEPRPDFNEFFPIWSNNNSARSGKPTMFLIHTQEAG 180

Query: 181 GGAAAENLAKWFQNGNGVSYHYTISQASDGGVTVVDCVDTDRAAWSVGNANSISINLCF 240
GGAAAENLAKWFQNGNGVSYHYTISQASDGGVTVVDCVDTDRAAWSVGNANSISINLCF
Sbjct: 181 GGAAAENLAKWFQNGNGVSYHYTISQASDGGVTVVDCVDTDRAAWSVGNANSISINLCF 240

Query: 241 AGSRASWMRDQWMKQSNADVAAYLAVQDAKKYGFTPLVPPPYTNGRPGISDHRWVTDV 300
AGSRASWMRDQWMKQSNADVAAYLAVQDAKKYGFTPLVPPPYTNGRPGISDHRWVTDV
Sbjct: 241 AGSRASWMRDQWMKQSNADVAAYLAVQDAKKYGFTPLVPPPYTNGRPGISDHRWVTDV 300

Query: 301 FKWGTHTDVGDFWFPWDYFAERVNHWANGGKTEPEPPKVKRFPDDWTDREILVEILRQLRG 360
FKWGTHTDVGDFWFPWDYFAERVNHWANGGKTEPEPPKVKRFPDDW+DREILVEILRQLRG
Sbjct: 301 FKWGTHTDVGDFWFPWDYFAERVNHWANGGKTEPEPPKVKRFPDDWSDREILVEILRQLRG 360

Query: 361 YNLTGWPQLGGKTLVDAVAELLGH 384

YNLTGWPQLGGKTLVDAVAELLGH

Sbjct: 361 YNLTGWPQLGGKTLVDAVAELLGH 384

Where does the N-terminal domain join the C-terminal domain?

Clustal Omega with N2 and N3 versions, all have same catalytic domain:

CLUSTAL O(1.2.4) multiple sequence alignment

```
FirstPart_Coco12_lysinA  -----MTTKDQVAQITIAEAKARGYTRSECLAVMSTF----YQESGWNDTIWDF      45
Shauna1_27              -----MTTKDQVAQITIAEAKARGYTRSECLAIMSTF----YQESGWNDTIWDF      45
Phanphagia_27          -----MTTKDQVAQITIAEAKARGYTRSECLAIMSTF----YQESGWNDTIWDF      45
Tweety_30              -----MTTKDQVAQITIAEAKARGYARSECLAVMSTF----YQESGWNDTIWDF      45
GUmbie_30              -----MTTKDQVAQITIAEAKARGYARSECLAVMSTF----YQESGWNDTIWDF      45
Velveteen_26          MSFTWFRPEGPLRTRQVAREVHAVSLARGLDELATVIALMTISTEVGTGTGDDRRKWCWCP  60
Cerasum_26            MSFTWFRPEGPLRTRQVAREVHAVSLARGLDELATVIALMTISTEVGTGTGDDRRKWCWCP  60
BigPhil_26            MSFTWFRPEGPLRTRQVAREVHAVSLARGLDELATVIALMTISTEVGTGTGDDRRKWCWCP  60
Boomer_32             MSFTWFRPEGPLRTRQVAREVHAVSLARGLDELATVIALMTISTEVGTGTGDDRRKWCWCP  60
LastPart_Coco12_lysinA -----                                0

FirstPart_Coco12_lysinA  THITYGIAQQDGSYPHRFDGAAAQIKGFFDKLDVWRAKPGASTDIWLNICW-MQQAPN--    102
Shauna1_27              THITYGIAQQDGSYPHRFDGAAAQIKGFFDKLDVWRAKPGASTDIWLNICW-MQQAPN--    102
Phanphagia_27          THITYGIAQQDGSYPHRFDGAAAQIKGFFDKLDVWRAKPGASTDIWLNICW-MQQAPN--    102
Tweety_30              THITYGIAQQDGSYPHRFDGAAAQIKGFFDKLDVWRAKPGASTDIWLNICW-MQQAPN--    102
GUmbie_30              THITYGIAQQDGSYPHRFDGAAAQIKGFFDKLDVWRAKPGASTDIWLNICW-MQQAPN--    102
Velveteen_26          ANDR---VPATKNYPHDSRGGDNRSSGYFQQ-----QPGPNGEPWGTAENMMTLPQAA    111
Cerasum_26            ANDR---VPATKNYPHDSRGGDNRSSGYFQQ-----QPGPNGEPWGTAENMMTLPQAA    111
BigPhil_26            ANDR---VPATKNYPHDSRGGDNRSSGYFQQ-----QPGPNGEPWGTAENMMTLPQAA    111
Boomer_32             ANDR---VPATKNYPHDSRGGDNRSSGYFQQ-----QPGPNGEPWGTAENMMTLPQAA    111
LastPart_Coco12_lysinA -----                                0

FirstPart_Coco12_lysinA  -----WPSADYWYANGRRAYLTEIKSRIAT---VTPY---LDKYWPT-----IGG      141
Shauna1_27              -----WPSADYWYANGRRAYLTEIKSRIIT---VTPY---LDKYWPA-----DGG      141
Phanphagia_27          -----WPSADYWYANGRRAYLTEIKSRIIT---VTPY---LDKYWPA-----DGG      141
Tweety_30              -----WPSADYWYANGRRAYLTEIKSRIIT---VTPY---LDKYWPA-----DGG      141
GUmbie_30              -----WPSADYWYANGRRAYLTEIKSRIIT---VTPY---LDKYWPA-----DGG      141
Velveteen_26          NTFLERLSDDYRRAANNPRLAGEFAQRVQQSAFFDRYAEKWDEAWSVLRRLALNE---TTP    168
Cerasum_26            NTFLERLSDDYRRAANNPRLAGEFAQRVQQSAFFDRYAEKWDEAWSVLRRLALNE---TTP    168
BigPhil_26            NTFLERLSDDYRRAANNPRLAGEFAQRVQQSAFFDRYAEKWDEAWSVLRRLALNE---TTP    168
Boomer_32             NTFLERLSDDYRRAANNPRLAGEFAQRVQQSEFFDRYADKWDEAWSVLRRLALGSQPPTTP    171
LastPart_Coco12_lysinA -----                                0

FirstPart_Coco12_lysinA  TAVPDEPRPDFNEMRACIKDGRG-----                                164
Shauna1_27              TAVPDEPRPDFNEFPIWSNNNSAR---SGKPTMFLIHTQEGG-GGDAAAENLAKWFQN-G    196
Phanphagia_27          TAVPDEPRPDFNEFPIWSNNNSAR---SGKPTMFLIHTQEGG-GGDAAAENLAKWFQN-G    196
Tweety_30              TAVPDEPRPDFNEFPIWSANNSSR---SGKPTMFLIHTQEGG-GGDAAAENLAKWFQN-S    196
GUmbie_30              TAVPDEPRPDFNEFPIWSANNSSR---SGKPTMFLIHTQEGG-GGDAAAENLAKWFQN-S    196
Velveteen_26          EEPVTENRPAYNEFPIWSANNNSAR---SGKPTMFLIHTQEGG-GGDAAAENLAKWFQN-G    223
Cerasum_26            EEPVTENRPAYNEFPIWSANNNSAR---SGKPTMFLIHTQEGG-GGDAAAENLAKWFQN-G    223
BigPhil_26            EEPVTENRPAYNEFPIWSANNNSAR---SGKPTMFLIHTQEGG-GGDAAAENLAKWFQN-G    223
Boomer_32             EVVMPENRPPYNEFPIWSENHYNTKRAANDIDAFLLHTSEGFGVGRDDAAEALSLSWYQFRS    231
LastPart_Coco12_lysinA -----MTNRPPYNEFPIWSNNNSAR---SGKPTMFLIHTQEGG-GGDAAAENLAKWFQN-G    51
                        ** *: ;

FirstPart_Coco12_lysinA  -----                                164
Shauna1_27              NGVSYHYTISQASDGGVTVVDCVDIDRAAWSVGNANSISINLCFAGSRASWMRDQWMK-Q    255
```

FirstPart_Coco12_lysinA	-----	164
Shauna1_27	NGVSYHYTISQASDGGVTVVDCVTDRAAWSVGNANSISINLCFAGSRASWMRDQWMK-Q	255
Phanphagia_27	NGVSYHYTISQASDGGVTVVDCVTDRAAWSVGNANSISINLCFAGSRASWMRDQWMK-Q	255
Tweety_30	NGVSYHYTISQASDGGVTVVDCVTDRAAWSVGNANSISINLCFAGSRASWTRLDQWMK-Q	255
GUmbie_30	NGVSYHYTISQASDGGVTVVDCVTDRAAWSVGNANSISINLCFAGSRASWTRDQWMK-Q	255
Velveteen_26	NGVSYHYTISQASDGGVTVVDCVTDRAAWSVGNANSISINLCFAGSRASWMRDQWMK-Q	282
Cerasum_26	NGVSYHYTISQASDGGVTVVDCVTDRAAWSVGNANSISINLCFAGSRASWMRDQWMK-Q	282
BigPhil_26	NGVSYHYTISQASDGGVTVVDCVTDRAAWSVGNANSISINLCFAGSRASWMRDQWMK-Q	282
Boomer_32	RQVAYHYAVSQASDGGVTVVNDVDDYASWSALSANGRSINLCFAGTRAAWSRNEWLGKF	291
LastPart_Coco12_lysinA	NGVSYHYTISQASDGGVTVVDCVTDRAAWSVGNANSISINLCFAGSRASWMRDQWMK-Q	110
FirstPart_Coco12_lysinA	-----	164
Shauna1_27	SNAI DVAA YLAVQDAKKYGFPLVVP PPTNGRPGISDHRWVIDVFKWGHTDVGDFWFPW	315
Phanphagia_27	SNAI DVAA YLAVQDAKKYGFPLVVP PPTNGRPGISDHRWVIDVFKWGHTDVGDFWFPW	315
Tweety_30	SNAI DVAA YLAVQDAKKYGFPLVVP PPTNGRPGISDHRWVIDVFKWGHTDVGDFWFPW	315
GUmbie_30	ANAI DVAA YLAVQDAKKYGFPLVVP PPTNGRPGISDHRWVIDVFKWGHTDVGANFPW	315
Velveteen_26	SNAI DVAA YLAVQDAKKYGFPLVVP PPTNGRPGISDHRWVIDVFKWGHTDVGDFWFPW	342
Cerasum_26	SNAI DVAA YLAVQDAKKYGFPLVVP PPTNGRPGISDHRWVIDVFKWGHTDVGDFWFPW	342
BigPhil_26	SNAI DVAA YLAVQDAKKYGFPLVVP PPTNGRPGISDHRWVIDVFKWGHTDVGDFWFPW	342
Boomer_32	GNAI DVAA YLAVQDCKKYNIPTKVIAPPYTGR LPGITDHRVYTIQLKDGHTDVGDFWFPW	351
LastPart_Coco12_lysinA	SNAI DVAA YLAVQDAKKYGFPLVVP PPTNGRPGISDHRWVIDVFKWGHTDVGDFWFPW	170
FirstPart_Coco12_lysinA	-----	164
Shauna1_27	DYFAERVNHWANGGKTEPEPPKVKRFPDDWDREILVEILRQLRGYNLTGWPQLGGKTLV	375
Phanphagia_27	DYFAERVNHWANGGKTEPEPPKVKRFPDDWDREILVEILRQLRGYNLTGWPQLGGKTLV	375
Tweety_30	DYFAERVNHWANGGKTEPEPPKVKRFPDDWSDREILVEILRQLRGYNLTGWPQLGGKTLV	375
GUmbie_30	DYFTERVNHWAAGGKTEPEPPKVKHFPDDWDRELAVETLRQQRGYTLNGWPQLGGRTVV	375
Velveteen_26	DYFAERVNHWANGGKTEPEPPKVKRFPDDWDREILVEILRQLRGYNLTGWPQLGGKTLV	402
Cerasum_26	DYFAERVNHWANGGKTEPEPPKVKRFPDDWDREILVEILRQLRGYNLTGWPQLGGKTLV	402
BigPhil_26	DYFAERVNHWANGGKTEPEPPKVKRFPDDWDREILVEILRQLRGYNLTGWPQLGGKTLV	402
Boomer_32	DYFTERVNHWAAGDKTEPEPPKVKRFPDDWDRELGVETLRQQRGYTLNGWPQLGGRTVV	411
LastPart_Coco12_lysinA	DYFAERVNHWANGGKTEPEPPKVKRFPDDWDREILVEILRQLRGYNLTGWPQLGGKTLV	230
FirstPart_Coco12_lysinA	-----	164
Shauna1_27	DAVAELLGH-----	384
Phanphagia_27	DAVAELLGH-----	384
Tweety_30	DAVADLRIDIIDLQGAIEHGEITLGAQ	403
GUmbie_30	DVLGAIGAK-LGIEGCYDVKGKS-----	397
Velveteen_26	DAVAELLGH-----	411
Cerasum_26	DAVAELLGH-----	411
BigPhil_26	DAVAELLGH-----	411
Boomer_32	DVLGAIG-EKLGVEGCYDVKGKS-----	433
LastPart_Coco12_lysinA	DAVAELLGHZ-----	240

Amazing -- the transition point between domains is the same NE site that the HNH insertion is.