

APPENDIX C

SEQUENCES IN MULTIPLE ALIGNMENT

1. *Fasciola hepatica* legumain like precursor (leg1 gene)

LOCUS AJ314846 1386 bp mRNA linear INV 05-JAN-2004
 DEFINITION *Fasciola hepatica* mRNA for legumain like precursor (leg1 gene).
 ACCESSION AJ314846
 VERSION AJ314846.1 GI:40643266
 KEYWORDS leg1 gene; legumain like precursor.
 SOURCE *Fasciola hepatica* (liver fluke)
 ORGANISM *Fasciola hepatica*
 Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 Echinostomida; Echinostomata; Echinostomatoidea; Fasciolidae;
Fasciola.
 REFERENCE 1
 AUTHORS Khaznadji, E., Moire, N. and Boulard, C.
 TITLE Functional expression and characterization of a *Fasciola hepatica*
 Legumain like precursor
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1386)
 AUTHORS Khaznadji, E.
 TITLE Direct Submission
 JOURNAL Submitted (07-JUN-2001) Khaznadji E., Centre de Recherche de Tours,
 INRA, Unite de pathologie aviaire et parasitologie, 37380 Nouzilly,
 FRANCE
 REMARK revised by author 19-JUN-2001
 FEATURES Location/Qualifiers
 source 1..1386
 /organism="Fasciola hepatica"
 /mol_type="mRNA"
 /db_xref="taxon:6192"
 /clone="TriplEx-Fh.NEJ .12"
 /clone_lib="TriplEx-Fh.NEJ"
 /dev_stage="newly excysted juvenile"
 /country="France:Central region"
 gene 1..1386
 /gene="leg1"
 CDS 1..1260
 /gene="leg1"
 /function="unknown"
 /codon_start=1
 /product="legumain like precursor"
 /protein_id="CAC85636.1"
 /db_xref="GI:40643267"
 /db_xref="GOA:Q711M2"
 /db_xref="InterPro:IPR001096"
 /db_xref="UniProtKB/TrEMBL:Q711M2"
 /translation="MIFLSLFFGSTVFLGAVAGKNWAVLVAGSNGWPNYRHHADVCHA
 YQVLIKNGFAPENIVTIFYNDVAYSQNYPYRGKIFHDYKHKDVEGKIDYSGLQTSV
 NTFASVLSGDEHKKAGYKVLQTGPDDNVFVYFTDHGARRLLFPYDPLDWEYLSYLL
 SYMKEHKMSNKMLFYVDASYSASMFATLPNNISVLAMTATNEHETNYAIFCDDPEVK
 SCMADEFSYQWISDIEKNELSKRTIENHFMAVKQAVSHSHVNLFGDMEISKLPLSEFF
 SKGDKNDFHEISTDSSMQMQDESKATQAHLISLTKQLTGSNSPRQVELAHRRLNRALE
 LGKMARETMDEIVNAVTTNGPPNGKHNKHTYVECYRTAYKQYENKCHSIYQVPEVSN
 ELEKLDHLCEQGYDVKMIVQAIQFDTC"
 sig_peptide 1..54
 /gene="leg1"
 mat_peptide 55..1257
 /gene="leg1"
 /product="unnamed"
 polyA_signal 1332..1337
 /gene="leg1"
 ORIGIN
 1 atgatatttt tgagcctggt tttcggtcgc accgtcttct tgggagctgt tgcgggaaag
 61 aactgggcag tgctttagc tggatcaaat ggttggccta actaccgtca tcacgctgat
 121 gtatgtcatg cttatcagggt gctgatcaaaa aatgggtttg ctcccagaaa catcgttacc
 181 ataattgaca atgatgttcg atatagcaga caaaaacctt atcgtggcaa aatattccac
 241 gattacaaac acaaagcgt ttatgagggg gtgaaaattg attatagtg cctacaaacc
 301 tcagtgaata cgttcgctag cgtactatca ggtgatgagc attggaagaa ggcgggctac

```

361 aaagtactcc aaaccggacc ggatgacaac gtgttcgtct atttcaccga tcacggagca
421 cgtcgcctac tattattccc atatgatccg cttgactggg agtatctgag ttacctgttg
481 tcctacatga aggaacataa gatgtccaat aaaatgctat tctatgtgga cgctagttat
541 tccgcctcga tgtttgccga gaccctacca aataacatca gtgttttggc catgactgct
601 accaatgaac acgaaacaaa ctacgccatt ttctgtgatg atcctgaagt caaatcatgt
661 atggccgacg aattttcgta tcagtggata tcagatattg aaaagaatga attgagcaaa
721 cggacgattg agaatcattt tatggctgtc aaacaggcag taagccacag tcatgtcaat
781 ttgttcggtg atatggaaat atccaagctc cctctttccg agttcttctc caagggtgac
841 aaaaatgatt tccacgaaat ctctacagac agttcaatgc aaatgcagga tgaatcgaaa
901 gccactcagg cacatctgat ttcattgaca aaacaattga cgggatcaaa ctcaccgaga
961 caagtggagt tggcacatcg tcgattgaat cgcgcactgg agttgggtaa aatggccaga
1021 gaaacatgg atgagatcgt gaatgctgtg accacgaacg gaccgcccaa tggaaaacac
1081 aatgataaac atacatatgt ggaatgttat cgaacggcat acaagcaata cgagaacaag
1141 tgtcatagca tttaccaagt tccagagggt tcaaacgaat tggagaagtt ggaccactta
1201 tgtgaacagg gatacgatgt gaaaatgatt gtgcaagcaa tattcgacac ttgtgcataa
1261 ttgccaccag ttcattgtgga tgcttatacg ttgccaacta ttcttgattt aacggggaca
1321 ttgattcaac gaataaaata tcccaaacca ttgcattgaa aaaaaaaaaa aaaaaaaaaa
1381 aaaaaa

```

//

2. *Opisthorchis viverrini* legumain

LOCUS DQ402101 1302 bp mRNA linear INV 11-MAR-2006
 DEFINITION *Opisthorchis viverrini* legumain mRNA, complete cds.
 ACCESSION DQ402101
 VERSION DQ402101.1 GI:89213679
 KEYWORDS .
 SOURCE *Opisthorchis viverrini*
 ORGANISM *Opisthorchis viverrini*
 Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Opisthorchiida; *Opisthorchiata*; *Opisthorchioidea*; *Opisthorchiidae*;
Opisthorchis.
 REFERENCE 1 (bases 1 to 1302)
 AUTHORS Laha,T., Sripa,J., Sripa,B., Brindley,P.J. and Loukas,A.
 TITLE Isolation and characterization of specific molecule to *Opisthorchis viverrini* infection associated cholangiocarcinoma using immunoscreening of *Opisthorchis viverrini* adult worm cDNA library
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1302)
 AUTHORS Sripa,J., Sripa,B., Kewgrai,N. and Laha,T.
 TITLE Direct Submission
 JOURNAL Submitted (15-FEB-2006) Department of Parasitology, Faculty of Medicine, Khon Kaen University, 123 Mittaparb Rd., Khon Kaen 40002, Thailand
 FEATURES Location/Qualifiers
 source 1..1302
 /organism="*Opisthorchis viverrini*"
 /mol_type="mRNA"
 /db_xref="taxon:6198"
 CDS 14..1240
 /note="asparaginyl proteinase"
 /codon_start=1
 /product="legumain"
 /protein_id="ABD64147.1"
 /db_xref="GI:89213680"
 /translation="MQRSCLLLTFLLYVNAAWLGAVCVGSRLFHSDQARNWVVLVAG
 SNGWENYRHQADVHYHAYQIMKRNNISTEQIITFAYDDIANNPENPFMGKVFNDYTHKD
 VYEGVHIDYRGEDVTPDNFLRAMRGDKLEANGKKVLKSGPEDHVFVYFSDHGADGLL
 AFPEDDLLASDLNKTGLGMHENKMYKQMVLYVEACESGSMFQDILPSDIGIYVTTAAN
 SEESSWATFCRDTIIGTCLADEYSYNWLTDSHHDLSHRTLDDQFQSVKQNTKQSHVS
 RFGELPQVLHSHPSRWHLVMTVRRMMKAETEEHEHELASRKLYRALLLAQIVKETFEE
 IVTDVTFHFQPTMRMLSKSEELQCYEEVFQEFKNRCFTIRQVPEVAQYARHLRLKCKE
 GYETEALVQSVHEVCS"
 ORIGIN
 1 aaacgtgggt gagatgcaac gttcttgctt tctcctgacg ttcttggtgt acgtcaacta
 61 cgctgcatgg ttaggcgctg tttggtgctg ttcgctggtg ttccacagtg atcaggcaag
 121 aaactgggtt gttctggtcg ctggatccaa tgggtgggag aactaccgac accaagcggg
 181 tgtatatcac gcgtatcaaa tcatgaagcg caacaacatt tgcacgggagc aaataattac
 241 cttcgcctac gatgatattg caaacaaccc cgaaaatccg tttatgggca aggtgttcaa
 301 tgactacact cacaaagacg tgtacgaagc tgtgcacata gattatcgtg gagaggatgt
 361 gacaccggac aatttctctg gtgccatgag gggtgataaa gaactgaa gaactgaaa
 421 gaaggtaacta aaaagcggtc cagaagatca cgtctttgtc tacttttccg atcatgggtc
 481 agacggactt cttgcctttc cagaggatga cctcctcgcc tcggatttga acaaaacttt
 541 gggttacatg cacgaaaaca aaatgtacaa acaaatgggt ctatacgtgg aagcatgtga
 601 atccggttct atgttccagg atatcctgcc atcggatate gggatctatg tgacaaccgc
 661 ggctaacagt gaggaatcca gctgggctac tttctgtcgc gacacaatca ttggcacttg
 721 tctggcggac gaatactcgt acaactggct cacggactct gagcatcacg atctgtcgca
 781 tgcacactcg gatgatcagt tccaatcggg gaaacagaat accaagcaaa gtcacgtatc
 841 gagattcggg gaactgcctc aggtacttca tagccatccg tcacgctggg cacatttggg
 901 accatgggtc cgacgaatga tgaagccgga aaccgaggaa gaacatgaat tggcatcccg
 961 aaaactatat cgtgcacttc tgcttgccca gatcgttaaa gaaacattcg aagaatcgt
 1021 cacggatgta acaaccttcc atcagccaac catgctgatg ttgtcaaaagt cggaggaact
 1081 ccagtgtcat gaagaagtat tccaagagtt caaaaaccgg tgcttcacca ttcgacaggt
 1141 cctgagggtg gctcaatcac caagacatct gcggaagctg tgcaagaag gatacgaaac
 1201 tgaagcactt gttcaatctg ttcataagat ctgcttctag tggacgctgt caacatttga
 1261 tgaacaacc ttaattcgaa aaatgcataa gtcctgcagt gg

//

3. *Schistosoma mansoni* asparaginyl endopeptidase (C197 Sm32)

LOCUS AJ250582 1287 bp mRNA linear INV 16-FEB-2000
 DEFINITION *Schistosoma mansoni* partial mRNA for asparaginyl endopeptidase (c197 Sm32 gene).
 ACCESSION AJ250582
 VERSION AJ250582.1 GI:6851049
 KEYWORDS asparaginyl endopeptidase; c197 Sm32 gene.
 SOURCE *Schistosoma mansoni*
 ORGANISM *Schistosoma mansoni*
 Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae; *Schistosoma*.
 REFERENCE 1
 AUTHORS Caffrey,C.R., Mathieu,M.A., Gaffney,A.M., Salter,J.P., Sajid,M., Lucas,K.D., Franklin,C., Bogyo,M. and McKerrow,J.H.
 TITLE Identification of a cDNA encoding an active asparaginyl endopeptidase of *Schistosoma mansoni* and its expression in *Pichia pastoris*
 JOURNAL FEBS Lett. 466 (2-3), 244-248 (2000)
 PUBMED 10682836
 REFERENCE 2 (bases 1 to 1287)
 AUTHORS Caffrey,C.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-NOV-1999) Caffrey C.R., Pathology, Tropical Disease Research Unit, University of California San Francisco, VAMC 113-B, Ca 94121, USA
 FEATURES Location/Qualifiers
 source 1..1287
 /organism="Schistosoma mansoni"
 /mol_type="mRNA"
 /strain="Puerto Rican"
 /db_xref="taxon:6183"
 /dev_stage="6-week-old adult"
 /country="Spain:Puerto Rico"
 gene 1..>1287
 /gene="c197 Sm32"
 CDS 1..>1287
 /gene="c197 Sm32"
 /note="possesses catalytic Cys 197"
 /codon_start=1
 /product="asparaginyl endopeptidase"
 /protein_id="CAB71158.1"
 /db_xref="GI:6851050"
 /db_xref="GOA:Q9NIFY9"
 /db_xref="InterPro:IPR001096"
 /db_xref="UniProtKB/TrEMBL:Q9NIFY9"
 /translation="MMLFSLFLISILHILLVKCQLDTNYEVSDETIVSDNNKWAFLVAG
 SNGYPNYRHQADVCHAYHVLRSKGIKPEHIITMMYDDIAYNLMNPFPGKLFNDYNHKG
 WYEGVVIDYRGKKNSTFLKVLKGDKSAGGKVLKSGKNDVFIYFTDHDGAPGLIAFP
 DDELYAKQFMSTLKYLHSHKRYSKLVIIYIACESGSMFQRIPLSNLSIYATTAASPTE
 SSYGTFCDPTITTCCLADLYSYDWIVDSQTHHLTQRTLDQQYKEVKRETNLSHVQRYG
 DTRMGKLVHSEFQGSRDKSSTENDEPPMKPRHSIASRDIPLHTLHRQIMMNTNAEDKS
 FLMQILGLKLRDLIEDTMKLIKVMNNEEIPNTKATIDQTLDCTESVYEQFKSKCF
 TLQQAPEVGGHFFSTLYNYCADGYTAETINEAIKICG"
 ORIGIN
 1 atgatgctat tctctttatt tcttattagc atcttacata ttctattagt caaatgtcaa
 61 ttagatacaa attatgaagt atccgatgaa actgttagtg ataataataa atgggctgta
 121 ttagtagctg gatctaattg ttatcccaat tatagacatc aagcagatgt ctgtcatgct
 181 tatcatgtat tacgttcaaa gggataaaa cctgagcata ttatcacaat gatgatgat
 241 gatatcgctt ataatttgat gaatccattt cctgggaaac tttttaatga ttataacct
 301 aaagattggt atgaaggagt ggtgatagat tatcgcggtg aaaaagtcaa ctcgaaact
 361 tttctgaaag ttttgaaggg agataaaaagc gctggtggga aagttttgaa aagtgaaaa
 421 aatgatgatg tattcatata cttcactgat catggtgcac cgggtctaatt tgctttccct
 481 gatgatgaat tatatgctaa acaatttatg tcaacattga aatacttaca tagtcacaaa
 541 cgttattcaa aattggtgat ttatattgaa gcatgtgaat caggttccat ctcgaaactg
 601 atattaccgt cgaatctaag tatttatgcg actacagctg ctagtccaac tgaatctagt
 661 tatggtacat tttgtgatga tccaacaata actactgtgc tggctgattt ataactatac
 721 gactggattg ttgactcaca aacacatcat ttaacacaac gaacactcga tcaacagtat
 781 aaagagggta aacgtgaaac gaacttagt catgttcaga gatattggtg tacgagaatg
 841 ggtaaattac acgttagtga atttcaagga agtcgagaca agtcttcaac tgagaacgat
 901 gaacctcaa tgaaccaag acattccatc gcttcgagag atattccatt gcataactta
 961 catcgtcaaa taatgatgac caataatgca gaagacaaaa gctttctgat gcaaatctt

```
1021 ggtttgaac tgaagagaag agatctcatt gaggatacta tgaaattaat agtgaaagta
1081 atgaataatg aagaaatacc aaataccaag gcaactatcg atcaaacatt ggattgtaca
1141 gaatcagtct atgaacagtt caaaagtaaa tgtttcacac tacaacaggc tccagaagtt
1201 ggagggcact tttctactct ttacaattat tgtgcagatg gttatacagc tgaaacgatc
1261 aatgaagcaa tcataaaaat ttgcggt
```

//

4. Mouse legumain (Lgmn)

LOCUS NM_011175 1874 bp mRNA linear ROD 18-NOV-2006
 DEFINITION Mus musculus legumain (Lgmn), mRNA.
 ACCESSION NM_011175
 VERSION NM_011175.2 GI:118130389
 KEYWORDS .
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1874)
 AUTHORS Maehr,R., Hang,H.C., Mintern,J.D., Kim,Y.M., Cuvillier,A.,
 Nishimura,M., Yamada,K., Shirahama-Noda,K., Hara-Nishimura,I. and
 Ploegh,H.L.
 TITLE Asparagine endopeptidase is not essential for class II MHC antigen
 presentation but is required for processing of cathepsin L in mice
 JOURNAL J. Immunol. 174 (11), 7066-7074 (2005)
 PUBMED 15905550
 REMARK GeneRIF: Lysosomal asparaginyl endopeptidase (AEP) is essential for
 processing of cathepsin L but not for class II major
 histocompatibility complex (MHC)-restricted antigen presentation.

REFERENCE 2 (bases 1 to 1874)
 AUTHORS Watahiki,A., Waki,K., Hayatsu,N., Shiraki,T., Kondo,S.,
 Nakamura,M., Sasaki,D., Arakawa,T., Kawai,J., Harbers,M.,
 Hayashizaki,Y. and Carninci,P.
 TITLE Libraries enriched for alternatively spliced exons reveal splicing
 patterns in melanocytes and melanomas
 JOURNAL Nat. Methods 1 (3), 233-239 (2004)
 PUBMED 15782199

REFERENCE 3 (bases 1 to 1874)
 AUTHORS Zeeuwen,P.L., van Vlijmen-Willems,I.M., Olthuis,D., Johansen,H.T.,
 Hitomi,K., Hara-Nishimura,I., Powers,J.C., James,K.E., op den
 Camp,H.J., Lemmens,R. and Schalkwijk,J.
 TITLE Evidence that unrestricted legumain activity is involved in
 disturbed epidermal cornification in cystatin M/E deficient mice
 JOURNAL Hum. Mol. Genet. 13 (10), 1069-1079 (2004)
 PUBMED 15044380
 REMARK GeneRIF: unrestricted legumain activity is involved in disturbed
 epidermal cornification in cystatin M/E deficient mice.

REFERENCE 4 (bases 1 to 1874)
 AUTHORS Shirahama-Noda,K., Yamamoto,A., Sugihara,K., Hashimoto,N.,
 Asano,M., Nishimura,M. and Hara-Nishimura,I.
 TITLE Biosynthetic processing of cathepsins and lysosomal degradation are
 abolished in asparaginyl endopeptidase-deficient mice
 JOURNAL J. Biol. Chem. 278 (35), 33194-33199 (2003)
 PUBMED 12775715
 REMARK GeneRIF: AEP has a pivotal role in the endosomal/lysosomal
 degradation system

REFERENCE 5 (bases 1 to 1874)
 AUTHORS Halfon,S., Patel,S., Vega,F., Zurawski,S. and Zurawski,G.
 TITLE Autocatalytic activation of human legumain at aspartic acid
 residues
 JOURNAL FEBS Lett. 438 (1-2), 114-118 (1998)
 PUBMED 9821970

REFERENCE 6 (bases 1 to 1874)
 AUTHORS Chen,J.M., Dando,P.M., Stevens,R.A., Fortunato,M. and Barrett,A.J.
 TITLE Cloning and expression of mouse legumain, a lysosomal endopeptidase
 JOURNAL Biochem. J. 335 (PT 1), 111-117 (1998)
 PUBMED 9742219

COMMENT PROVISIONAL REFSEQ: This record has not yet been subject to final
 NCBI review. The reference sequence was derived from AC122302.4.
 On Nov 17, 2006 this sequence version replaced gi:7242186.

 Sequence Note: The RefSeq transcript and protein were derived from
 genomic sequence to make the sequence consistent with the reference
 genome assembly. The genomic coordinates used for the transcript
 record were based on transcript alignments.

FEATURES Location/Qualifiers
 source 1..1874
 /organism="Mus musculus"

```

/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/chromosome="12"
/map="12"
gene 1..1874
/gene="Lgmn"
/note="legumain; synonyms: AEP, Prsc1, AI746452, AU022324"
/db_xref="GeneID:19141"
/db_xref="MGI:1330838"
CDS 95..1402
/gene="Lgmn"
/go_component="apical part of cell; late endosome; lysosome"
/go_function="cysteine-type endopeptidase activity; cysteine-type peptidase activity; hydrolase activity; legumain activity; peptidase activity"
/go_process="negative regulation of body size; proteolysis"
/note="protease, cysteine, 1; preprolegumain"
/codon_start=1
/product="legumain"
/protein_id="NP_035305.1"
/db_xref="GI:7242187"
/db_xref="CCDS:CCDS26119.1"
/db_xref="GeneID:19141"
/db_xref="MGI:1330838"
/translation="MTWRVAVLLSLVLGAGAVPVGVDPEGGKHWVIVAGSNGWYN
YRHQADACHAYQIIHRNGIPDEQIIVMMYDDIANSEENPTPGVVINRPNGTDVYKGVL
KDYTGEDVTPENFLAVLRGDAEAVKKGSGKVLKSGPRDHVFIYFTDHGATGILVFPN
DDLHVKDLNKTIRYMYEHKMYQKMFYIEACEGSMNHLRDDINVIYATTAANPKESS
YACYDEERGTYLGDWYSVNWMEGSDVEDLTKETLHKQYHLVKSHTNTSHVMQYGNKS
ISTMKVMQFQGMKHRASSPISLPPVTHLDLTPSPDVPLTILKRKLLRNDVKEQNLII
GQIQQFLDARHVIKSVHKIVSLLAGFGETAERHLSERTMLTAHDYQEAVTHFRTHC
FNWHSVTYEHALRYLYVLANLCEAPYPIDRIEMAMDKVCLSHY"
STS 1605..1854
/gene="Lgmn"
/standard_name="AU022324"
/db_xref="UniSTS:208854"
STS 1655..1775
/gene="Lgmn"
/standard_name="AI746452"
/db_xref="UniSTS:160246"
ORIGIN
1 gctctgagtc tgcgcgacgc ccggaattcc cacggttctg cagtcaccgc ggcgatcacc
61 cgcccagtct tctgtagcgg acacgggggtg cagaatgacc tggagagtgg ctgtgcttct
121 cagcctgggt ctgggtgctg gtgccgttcc cgtcgggtg gacgatcccg aggatggagg
181 caagcactgg gtgggtgattg tggcgggctc caatggctgg tataattacc gacaccaggc
241 agacgcatgc cacgcctacc agatcatcca ccggaacggg attcctgacg agcagatcat
301 agtgatgatg tatgacgaca ttgccaaactc tgaagaaaac cctaccccag gtgttgatg
361 caaccgacct aacggcacag atgtatacaa gggagtctg aaggactaca ccggagagga
421 tgtgactcca gagaatttcc tcgcccgtgct gagaggtgac gcagaagctg tgaagggcaa
481 aggtctgga aaagtcttga agagtggccc ccgagatcat gtcttcattt acttcaccga
541 ccacggagcc accgggatcc tgggtgttcc taatgatgat ctcatgtca aggacctgaa
601 taagactatt cgctacatgt atgaacacaa aatgtaccag aagatgggtg tctacattga
661 agcttgtagt tctggctcca tgatgaacca cctgcccgcac gacatcaacy tttatgcaac
721 tactgcccgc aacccaagg agtcatctta tgcctgctac tacgacgagg agaggggac
781 ttacctgggt gactggtaca gcgtcaactg gatggaagac tccgatgtgg aggacctgac
841 caagagagcc cttcacaagc agtaccacct ggtcaagtcc cacaccaaca ccagccatgt
901 catgcaatat ggaacaaaat ctatctctac catgaaagtg atgcagtttc agggaatgaa
961 gcacagagcc agttccccc tctccctgcc tccggtcaca caccttgacc tcaccccag
1021 cctgacgtg cccctgacca tcttgaagag gaagctgctg agaaccaacy acgtgaagga
1081 atcccagaat ctcatgggc agatccagca atttctggat gccaggcacg tcattgagaa
1141 gtctgtgcac aagatcgttt ccctgctggc gggatttggg gaaactgctg agagacatct
1201 gtcagagagg accatgctca cagcacatga ctgctaaccg gaggctgtaa cccacttccg
1261 cacacactgc ttaactggc actctgtcac gtacgagcat gccttgcggt acttgatgt
1321 gctggccaat ctctgtgagg caccatatcc gattgacagg atagagatgg ccatggacaa
1381 agtgtgtctt agtcactact gaacagctcg cttcccaatg agtgagcaca gtcactgga
1441 atatgaacca accagagact gaaagggcgg agccagaggg agcaccgcg ccctggccc
1501 ccaggaatac tgcccgcaca cccagggct tgcttttga agatacctgc ttaactaagaa
1561 gccagtttgg gtgggtaaa ctctctggaa gaaggaactt tgcttcttag gagtttttt
1621 gtttgttgt tttggttgt tttgtgtcc attagctttc aagagcaaat tcccgcggc
1681 ttctctagcc agggaaggaa tcgtctgaga aattcaagc tgaaacctct tgccgtcttc

```

```
1741 acagtgattt cactgaagaa gaggtggaa agcaagccc tatgggagaa ttatTTTTag
1801 aattatataa tttttgattg cttttatatt ttattctgta ataatggatg tcttaaaca
1861 aataagtgaa gtga
```

//

5. Human legumain

LOCUS BC003061 1972 bp mRNA linear PRI 15-JUL-2006
 DEFINITION Homo sapiens legumain, mRNA (cDNA clone MGC:1395 IMAGE:3504506), complete cds.
 ACCESSION BC003061
 VERSION BC003061.2 GI:33872604
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1972)
 AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
 CONSRTM Mammalian Gene Collection Program Team
 TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 PUBMED 12477932
 REFERENCE 2 (bases 1 to 1972)
 CONSRTM NIH MGC Project
 TITLE Direct Submission
 JOURNAL Submitted (12-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Bethesda, MD 20892-2590, USA
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 COMMENT On Aug 19, 2003 this sequence version replaced gi:13111749.
 Contact: MGC help desk
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ruben Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbiology.org>
 contact: amadan@systemsbiology.org
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 7 Row: 0 Column: 5
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 56682961.

 Differences found between this sequence and the human reference genome (build 36) are described in misc_difference features below and these differences were also compared to chimpanzee genome (build 1).
 FEATURES Location/Qualifiers
 source 1..1972
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:1395 IMAGE:3504506"
 /tissue_type="Placenta, choriocarcinoma"

```

/clone_lib="NIH_MGC_21"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
gene 1..1972
/gene="LGMN"
/note="synonyms: LGMN1, AEP"
/db_xref="GeneID:5641"
/db_xref="HGNC:9472"
/db_xref="MIM:602620"
CDS 133..1434
/gene="LGMN"
/codon_start=1
/product="legumain"
/protein_id="AAH03061.1"
/db_xref="GI:13111750"
/db_xref="GeneID:5641"
/db_xref="HGNC:9472"
/db_xref="MIM:602620"
/translation="MVWKVAVFLSVALGIGAIPIDDPEDGGKHVVVIVAGSNGWYNYR
HQADACHAYQIIHRNGIPDEQIVVMYDDIAYSEDNPTPGIVINRPNGTDVYQGVPKD
YTGEDVTPQNFLAVLRGDAEAVKIGSGKVLKSGPQDHVFIYFTDHGSTGILVFPNED
LHVKDLNETIHYMYKHKMYRKMVFYIEACESGSMNHLDPDNINVYATTAANPRESSYA
CYDEKRSTYLDGWYSVNWMEDSDVEDLTKETLHKQYHLVKSHNTNTSHVMQYGNKTIIS
TMKVMQFQGMKRKASSPVLPVTHLDLTPSPDVPLTIMKRKLMNTNDLEESRQLTEE
IQRHLDARHLIEKSVRKIVSLLAASEAEVEQLLSERAPLTGHSCYPEALLHFRTHCFN
WHSPTYEYALRHLVYLVNLCEKPYPLHRKIKLSMDHVCLGHY"
misc_difference 184
/gene="LGMN"
/note="'A' in cDNA is 'G' in the human genome; amino acid
difference: 'I' in cDNA, 'V' in the human genome. The
chimpanzee genome agrees with the human genomic sequence
and not the cDNA."
misc_difference 852
/gene="LGMN"
/note="'C' in cDNA is 'T' in the human genome; no amino
acid change."
misc_difference 1383
/gene="LGMN"
/note="'A' in cDNA is 'G' in the human genome; no amino
acid change. The chimpanzee genome agrees with the human
genomic sequence and not the cDNA."
misc_difference 1945..1972
/gene="LGMN"
/note="polyA tail: 28 bases do not align to the human
genome."

ORIGIN
1 gaggctgcga gccgcccgcga gttctcacgg tcccgcgggc gccaccaccg cggctactca
61 ccgcccgcgc cgccaccact gccaccacgg tcgcctgccca cagggtgctg caattgaact
121 ccaaggtgca gaatggtttg gaaagtagct gtattcctca gtgtggccct gggcattggt
181 gccattccta tagatgatcc tgaagatgga ggcaagcact ggggtggtgat cgtggcaggt
241 tcaaatggct ggtataatta taggcaccag gcagacgcgt gccatgccta ccagatcatt
301 caccgcaatg ggattcctga cgaacagatc gttgtgatga tgtacgatga cagctcttac
361 tctgaagaca atccactcc aggaattgtg atcaacaggc ccaatggcac agatgtctat
421 cagggagtcc cgaaggacta cactggagag gatgttacc caaaaattt ccttgctgtg
481 ttgagaggcg atgcagaagc agtgaagggc ataggatccg gcaaagtccg gaagagtggc
541 ccccaggatc acgtgttcat ttacttcact gaccatggat ctactggaat actggttttt
601 cccaatgaag atcttcatgt aaaggacctg aatgagacca tccattacat gtacaaacac
661 aaaatgtacc gaaagatggt gttctacatt gaagcctgtg agtctggggtc catgatgaac
721 cacctgccgg ataacatcaa tgtttatgca actactgctg ccaaccccag agagtctgctc
781 tacgcctggt actatgatga gaagaggctc acgtacctgg gggactggta cagcgtcaac
841 tggatggaag actcggacgt ggaagatctg actaaagaga ccctgcacaa gcagtaccac
901 ctggtaaaat cgcacaccaa caccagccac gtcatgcagt atggaaacaa aacaatctcc
961 accatgaaag tgatgcagtt tcagggtatg aaacgcaaag ccagttctcc cgtcccccta
1021 cctccagtea cacaccttga cctcaccccc agccctgatg tgccctctac catcatgaaa
1081 aggaaactga tgaacaccaa tgatctggag gaggccaggc agctcacgga ggagatccag
1141 cggcatctgg atgccaggca cctcattgag aagtcagtgc gtaagatcgt ctccttgctg
1201 gcagcgtccg aggctgaggt ggagcagctc ctgtccgaga gagccccgct cacggggcac
1261 agctgctacc cagaggccct gctgcacttc cggaccact gcttcaact gcaactcccc
1321 acgtacgagt atgcggtgag acatctgtac gtgctggtca acctttgtga gaagccgtat
1381 ccacttcaca ggataaaatt gtccatggac cacgtgtgcc ttggtcacta ctgaagagct
1441 gcctcctgga agcttttcca agtgtgagcg ccccaccgac tgtgtgctga tcagagactg
1501 gagaggtgga gtgagaagtc tccgctgctc gggccctcct ggggagcccc cgcctccagg
1561 ctcgctccag gaccttcttc acaagatgac ttgctcgtg ttacctgctt ccccagctt

```

```
1621 ttctgaaaa ctacaaatta ggggsggaaa agctctgtat tgagaagggt catatttgct
1681 ttctaggagg ttgttggtt tgctgttag ttttgaggag caggaagctc atgggggctt
1741 ctgtagcccc tctcaaaagg agtctttatt ctgagaattt gaagctgaaa cctctttaa
1801 tcttcagaat gattttattg aagagggcgc caagcccaa atggaaaact gtttttagaa
1861 aatatgatga tttttgattg cttttgtatt taattctgca ggtgttcaag tcttaaaaaa
1921 taaagattta taacagaacc caaaaaaaaa aaaaaaaaaa aaaaaaaaaa aa
```

//

6. Cow legumain (LGMN)

LOCUS NM_174101 1934 bp mRNA linear MAM 17-NOV-2006
 DEFINITION Bos taurus legumain (LGMN), mRNA.
 ACCESSION NM_174101
 VERSION NM_174101.2 GI:31342841
 KEYWORDS .
 SOURCE Bos taurus (cattle)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
 Pecora; Bovidae; Bovinae; Bos.
 REFERENCE 1 (bases 1 to 1934)
 AUTHORS Yamane,T., Takeuchi,K., Yamamoto,Y., Li,Y.H., Fujiwara,M.,
 Nishi,K., Takahashi,S. and Ohkubo,I.
 TITLE Legumain from bovine kidney: its purification, molecular cloning,
 immunohistochemical localization and degradation of annexin II and
 vitamin D-binding protein
 JOURNAL Biochim. Biophys. Acta 1596 (1), 108-120 (2002)
 PUBMED 11983426
 COMMENT PROVISIONAL REFSEQ: This record has not yet been subject to final
 NCBI review. The reference sequence was derived from AB060129.1.
 On Jun 3, 2003 this sequence version replaced gi:27806554.
 FEATURES
 source Location/Qualifiers
 1..1934
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 gene 1..1934
 /gene="LGMN"
 /note="legumain; synonym: MGC128270"
 /db_xref="GeneID:281281"
 CDS 88..1389
 /gene="LGMN"
 /codon_start=1
 /product="legumain"
 /protein_id="NP_776526.1"
 /db_xref="GI:27806555"
 /db_xref="GeneID:281281"
 /translation="MIWEFTVLLSLVLGTGAVPLEDPEDGGKHVVIVAGSNGWYNYR
 HQADACHAYQIVHRNGIPDEQIIIVMMYDDIANSEDNPTPGIVINRPNNGSDVYQGVLKD
 YTGEDVTPKNFLAVLRGDAEAVKGVGSGKVLKSGPRDHVVFVYFDHGATGILVFPNED
 LHVKDLNETIRYMYEHKMYQKMFVYIEACESGSMNHLPPDINVYATTAANPRESSYA
 CYYDEQRSTFLGDWYSVNWMESSDVEDLTKETLHKQYQLVKSHNTNTSHVMQYGNKSIS
 AMKLMQFQGLKHQASSPISLPAVSRLDLTPSPEVPLS IMKRKMLNDLQESRRLVQK
 IDRHLEARNII EKSVRKIVTLVSGSAAEVDRLLSQRAPL TEHACYQTAVSHFRSHCFN
 WHNPTYEYALRHLYVLVNLCEPYPIDRIKLSMKNKCHGYG"
 ORIGIN
 1 cggcagctca cttctttgag gctgcgacc gcgaggagt cccacggtcc cgcagcagat
 61 ttctgcagcg gaactcccgg gggcaggatg atttgggaat tcaccgtggt gctcagcttg
 121 gtctctggaa ctggtgctgt ccccctggag gaccctgagg acggaggcaa gcactgggtg
 181 gtgatcgtcg caggatcaaa cggctggtat aattacaggc accaggcaga tgcattgcat
 241 gcctaccaga tcggtcaccg aaacgggatc cctgatgagc agatcatcgt gatgatgat
 301 gacgacatcg ccaactccga agacaatccc accccgggaa ttgtgatcaa caggcccaac
 361 ggctcagacg tgtaccaggg cgtgctgaag gactacacgg gcgaggatgt caccocgaag
 421 aatttccttg ctgtgttgag aggtgatgca gaagcagtga agggcgtagg atctggaaa
 481 gtcttaaaga gtggccccag ggaccacgtg ttcgtttact tcacggatca tggagccact
 541 ggaataactg tgtttcctaa tgaagatcct catgtaaagg acctgaatga gaccatccgt
 601 tacatgtaag aacacaagat gtacaaaag atgggtgttct acattgaggc ctgtgagtct
 661 ggttccatga tgaaccacct gccaccgac atcaatgttt atgcaactac ggtgccaac
 721 cccagggagt cgtcctacgc ctgttactac gatgagcagc ggtccacgtt actgggggac
 781 tggtagacgc tcaactggat ggaagattca gatgtggagg atttgaccaa cagaccctc
 841 cacaagcagt accagctggt gaaatccac accaacacca gccacgtcat gcagtatggg
 901 aacaagtcca tctctgcat gaagctgatg cagtttcagg gtctgaaaca ccaagctagc
 961 tctccatct cgctgcccgc tgtcagtcgc cttgacctca ctccaagccc cgaggtgccg
 1021 ctacgacatc tgaaaaggaa gctgatgagc accaacgata tgcaggagtc caggcgtctt
 1081 gtccagaaga tcgacaggca tctggaggcc aggaacatca ttgagaagtc agtgcgaaag
 1141 atcgtcacct tgggtctcggg ctccgctgcc gaggtagaca gactcctgtc ccagagagcc
 1201 ccgctcacag agcacgcatg ctaccagacg gccgtgtcgc acttccgctc acactgcttc
 1261 aactggcaca accccacgta cgagtatgct ttgagacatt tgtacgtgct tgtcaacctt
 1321 tgtgaaaacc cttatccaat tgacagaata aaattgtcca tgaacaaggt gtgccatggc
 1381 tactactgaa aactgcctc cttgaagctt taccgaatgt gatcaccatc cttacaatg

```
1441 tgtgctaatc agagactgga gagctgcagc ctgcgggagg ctccgctccc ccagccccag
1501 cggctgtgtc cttcttggga tgccgtcagg ccaccggagc cacgctccag gcctctcca
1561 ggatggcttg ctcaccttta cctacttccc ccagcaccta caggtcctgt gtgatcctct
1621 gaaaagccca gagatttggg aagctctgtg ttgagaaggg tgtatttgct ttttaagagg
1681 gtgttttgcc cttgacgttt aaggagtgag aagttcactg ggacttatgt agccagtga
1741 agagtctttg ttctgaggaa tctgacgtca gaacctotta agtcttcaca tgattttctg
1801 gaagagggac ataaaccca aatggaaaac tatttttaga aaatagcata attttgatt
1861 gcttttatat tttatttggc agtaatggac aagtctaaca aataaagatt tatcactgaa
1921 cctgagtatt aaaa
```

//

7. *Arabidopsis thaliana* DELTA-VPE; cysteine-type endopeptidase (DELTA-VPE)

LOCUS NM_112912 1681 bp mRNA linear PLN 09-JUN-2006
DEFINITION *Arabidopsis thaliana* DELTA-VPE; cysteine-type endopeptidase (DELTA-VPE) mRNA, complete cds.
ACCESSION NM_112912
VERSION NM_112912.2 GI:30685671
KEYWORDS .
SOURCE *Arabidopsis thaliana* (thale cress)
ORGANISM *Arabidopsis thaliana*
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
COMMENT PROVISIONAL REFSEQ: This record has not yet been subject to final NCBI review. This record has been curated by TAIR. This record is derived from an annotated genomic sequence (NC_003074). The reference sequence was derived from AT3G20210.1.
On May 13, 2003 this sequence version replaced gi:18402530.
FEATURES
Location/Qualifiers
source 1..1681
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="3"
/ecotype="Columbia"
gene 1..1681
/gene="DELTA-VPE"
/locus_tag="AT3G20210"
/note="synonym: DELTAVPE"
/function="Encodes a vacuolar processing enzyme with caspase-1-like activity that is specifically expressed in inner integument of developing seeds. Mutants display abnormal seed coat development. It is speculated to be involved in cell death of limited cell layers, the purpose of which is to form a seed coat."
/db_xref="GeneID:821565"
CDS 63..1463
/gene="DELTA-VPE"
/locus_tag="AT3G20210"
/go_component="extracellular space [pmid 15705955]"
/go_function="caspase activity [pmid 15705955]; cysteine-type endopeptidase activity"
/go_process="programmed cell death [pmid 15705955]; proteolysis; seed coat development [pmid 15705955]; vacuolar protein processing or maturation [pmid 14688293]"
/note="vacuolar processing enzyme, putative / asparaginyl endopeptidase, putative, similar to asparaginyl endopeptidase (VmPE-1) (*Vigna mungo*) GI:4589396; contains Pfam profile PF01650: Peptidase C13 family; identical to cDNA vacuolar processing enzyme delta preproprotein (At3g20210) GI:24850432"
/codon_start=1
/product="DELTA-VPE; cysteine-type endopeptidase"
/protein_id="NP_188656.1"
/db_xref="GI:15231080"
/db_xref="GeneID:821565"
/translation="MSSPLGHFQILVFLHALLIFSAESRKTQLLNDNDVSSDKSAKG TRWAVLVAGSNEYNYRHRQADICHAYQILRKGGLKDENIIVFMYDDIAFSSSENPRPGV IINKPDGEDVYKGVPKDYTKAVNVQNFYNVLLGNESGVTGGNGKVVKSGPNDNIFIY YADHGAPGLIAMPTGDEVMAKDFNEVLEKMHKRKYNKMVIYVEACESGSMFEGILKK NLNIYAVTAANSKESWGVYCPESYPPPPSEIGTCLGDTFISISWLESDLDHMSKETL EQQYHVVKRRRVGSDVPETSHVCRFGTEKMLKDYLSSYIGRNPENDNFTFTEFSFSPIS NSGLVNPRIPLLQLRKIQKAPMGSLSESKEAQKLLDDEKNHRKQIDQSITDILRLSV KQTNVLLNLTSTRTTGQPLVDDWDCFKTLVNSFKNHCGATVHYGLKYTGALANICNMG VDVKQTVSAIEQACSM"
ORIGIN
1 aatcatcatc caaaaacatt cttctcaca gaatcagatt caagatagaa gtttttcaaa
61 caatgtctag tcctccttgg cactttcaga ttcttgtttt tcttcatgct ttgcttatct
121 tctcagctga gtccccgaaa acccaattgc tgaacgataa tgatgttgaa tctagcgaca
181 agagtgcaaa aggcacacga tgggctgttt tagttgctgg atcaaatgaa tattataact
241 acagggcatca ggctgacata tgccacgcgt atcagatact ccgaaaaggc ggttttaaag

```

301 atgaaaacat cattgtgttt atgtatgatg atatcgcggt ttcctcggag aatcctaggc
361 ctggagttat cattaataaa ccagatggag aagatgttta taaaggagtt cctaaggact
421 aactaaaga agctgttaat gttcaaaact tctacaatgt gttacttggg aatgaaaagt
481 gcgtcacagg aggaaatggc aaagttgtga aaagtgggcc taatgataat atcttcatct
541 attatgctga ccatggagct cctggcttaa tagcgatgcc cactgggtgat gaagttatgg
601 caaaagatth caatgaagtc ttggagaaga tgcataagag aaaaaaatac aacaagatgg
661 tgatctatgt tgaagcatgt gaatcaggaa gtatgtttga agggatttta aagaaaaatc
721 tcaacatata cgcagtgact gctgctaatt ctaaagagag cagctgggga gtttactgtc
781 ctgagtcata tcctcctcct ccttctgaga ttggaacttg tctcggcgat acatttagca
841 tctcttggct tgaggacagt gaccttcatg acatgagcaa agagactttg gagcaacaat
901 accacgttgt aaagagaaga gtaggatctg atgtaccaga gacttctcat gtatgccggt
961 tcggaacaga gaagatgctt aaagattatc tttcctctta cattggaaga aatcctgaaa
1021 acgataactt cactttcacg gaatcctttt cctcaccaat ctctaattct ggcttgggca
1081 atccgcgcga tattcctctg ctatacctcc agagaaagat tcaaaaagct ccaatgggat
1141 cacttgaag caaagaagct cagaagaaat tgcttgacga aaagaatcat aggaacaaa
1201 tcgatcagag cattacagac attctgcggc tttcagttaa acaaaccaat gtcttaaatc
1261 tcttaacttc cacaagaaca acaggacagc ctctttaga cgattgggat tgcttcaaga
1321 ctctagttaa tagcttcaag aatcactgcg gtgcaacggg gcattacgga ttgaagtata
1381 caggagcgtc tgccaatata tgcaatatgg gagtggatgt gaagcaaac gttcagcca
1441 ttgaacaagc ttgttcgatg taatgatthg caaaacaatg tgatattcga ctttaaaaat
1501 atcaaagtta atttcaataa aactcgatgt agagatgggt gggtcatgat actactttta
1561 catgaaaagc tttttaatcg atgataacgc gaaagtcttg gtctaaattt gtgaattgga
1621 ttcattggaac aataacctcg taccaactgt acggtacgga cggctgtact ttggttgagt
1681 t

```

//

8. *Vigna mungo* asparaginyl endopeptidase (VmPE-1A)

LOCUS D89972 1659 bp mRNA linear PLN 06-APR-1999
DEFINITION *Vigna mungo* mRNA for asparaginyl endopeptidase (VmPE-1A), complete cds.
ACCESSION D89972
VERSION D89972.1 GI:4589397
KEYWORDS asparaginyl endopeptidase (VmPE-1A).
SOURCE *Vigna mungo* (black gram)
ORGANISM *Vigna mungo*
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; *Vigna*.
REFERENCE 1 (bases 1 to 1659)
AUTHORS Okamoto,T. and Minamikawa,T.
TITLE Molecular cloning and characterization of *Vigna mungo* processing enzyme 1 (VmPE-1), an asparaginyl endopeptidase possibly involved in post-translational processing of a vacuolar cysteine endopeptidase (SH-EP)
JOURNAL Plant Mol. Biol. 39 (1), 63-73 (1999)
PUBMED 10080709
REFERENCE 2 (bases 1 to 1659)
AUTHORS Okamoto,T.
TITLE Direct Submission
JOURNAL Submitted (19-DEC-1996) Takashi Okamoto, Tokyo Metropolitan University, Department of Biology; Minami-osawa 1-1, Hachioji, Tokyo 102-03, Japan (E-mail:okamoto-takashi@c.metro-u.ac.jp, Tel:+81-426-77-2562, Fax:+81-426-77-2559)
FEATURES
Location/Qualifiers
source 1..1659
/organism="Vigna mungo"
/mol_type="mRNA"
/db_xref="taxon:3915"
/clone="pPEB"
/tissue_type="cotyledon"
/dev_stage="seedling"
5'UTR 1..45
/experiment="experimental evidence, no additional details recorded"
CDS 46..1494
/experiment="experimental evidence, no additional details recorded"
/note="processing enzyme 1A"
/codon_start=1
/product="asparaginyl endopeptidase (VmPE-1A)"
/protein_id="BAA76745.1"
/db_xref="GI:4589398"
/translation="MATTSTLSTLFLFLATVALVAARRDHVGFRLRPSDSGNDNV
QGTRWAILFAGSNGYWNRYRHQADICHAYQILRKGLKEENIIVFMYDDIAFNWDNPRP
GVIINKPDGDDVYEGVPKDYTGEDATAHNFYSALLGDKSALTGGSGKVVSSGPDRIIF
IFYSDHGGPGVLGTPAGPYIYASDLVEVLKHKHASGTYKNLVFYLEACEAGSIFEGLL
PEDINIYATTASNAEESWGTYPGEYSPPEYSTCLGDLYSVAMMEDSDRHLRTE
SLHQYKVVKDRTLSSGWYGSVHMVQYGDVEFSKDALFLYLGTDNPANDLTFVDENSLW
SSSTAVNQRDADLVHFWHKFRKAPEGSPKNEARKQVLEVMHRMHIDDSVKLVGKLL
FGFEKAPEVLNAVRPAGSALVDDWACLKTMVRTFETHCGSLSQYGMKHMSPFANICNV
GIKKEQMAEASAQACVTPASSWSSLQRFSA"
3'UTR 1495..1659
/experiment="experimental evidence, no additional details recorded"
ORIGIN
1 gaaatcacag ctacttacct cactctttct ctctttgctc ttccaatggc caccacaact
61 tctctttcca cactctttct tctcttctc gcaaccgtcg ctcttggtgc tgctcgccgt
121 gaccacgtcg gagacttctc cgggctgccg tccgattccg gtaacgatga caacgtccag
181 gggacgcgct gggccatcct cttcgccggt tcgaacggct attggaacta cagaccag
241 gccgatattt gtcattgccta tcaaaactg aggaagggtg gcctgaaaga agaaaatc
301 atgtttttta tgtatgatga cattgcattc aattgggaca atccaaggcc tggagtcac
361 attaacaac cagacgggga tgatgtttat gaaggagttc caaaggatta cactggcgaa
421 gatgctactg ctcaactt ttactctgct ttacttggag ataagtcggc gcttaccggt
481 ggcagtgga aggttgtag cagtgggcc gatgatcgca tatttatatt ctactctgat
541 catggtggtc caggggttct cgggacgct gctggtcctt acatatatgc atctgattg
601 gttgaagtac tgaagaaaaa gcatgcttct ggaacgtata aaaacctagt attttattg

```
661 gaggcacatg aagctggaag tatcttcgaa ggtcttcttc ctgaagatat caatatttat
721 gcaacaacgg cttcgaatgc cgaagaaagt agttggggaa catattgtcc tggggagat
781 cctagtcctc ccccagaata ttcaacctgc ttgggtgacc tgtacagtgt tgcttggatg
841 gaagacagtg atagacacaa cttgcgaaca gaatctttac accaacaata taaagtggtt
901 aaggatagga ctctatctgg aggttggat ggctctcacg tgatgcagta tggatgatga
961 gagtttagca aggatgctct cttcctgtat ttgggtacag atcctgcaaa tgataatttg
1021 actttcgtgg atgaaaactc cttatggtca tcttcaactg cagtcaacca aagggatgca
1081 gatctcgtcc atttttggca taagttccgc aaagctccgg agggttctcc taagaaaaat
1141 gaagctcgga aacaagtctt ggaagtaatg tctcacagaa tgcatataga cgacagtgtg
1201 aaacttggtg gaaagctttt atttggattt gaaaaggctc ctgaagtact gaatgctgtt
1261 agaccgctg gatcggcact tgttgatgac tgggcctgcc tgaaaacat ggtaaggact
1321 tttgagacac attgtggatc cttgtctcag tatgggatga aacacatgag tccctttgca
1381 aatatctgca atgtaggaat aaagaaagag caaatggctg aagcctcagc acaagcttgt
1441 gtcactgttc ctgccagttc ctggagtctc ctgcaaagag gtttcagtgc ataattcctg
1501 taatgttcac cagtaaagat aaaggcatga tattgttaca ttatgttatg tggttgtaca
1561 tacatatctt gtcccgcctc tgtaaatgca attgggacat tactaggagt tggagaagc
1621 atccattggt ccatatattt tagtttggca taaaaaaaa
```

//

9. *Trichomonas vaginalis* legumain-like cysteine proteinase 2 (legu-2)

LOCUS AY326447 1364 bp mRNA linear INV 01-JUL-2004
DEFINITION *Trichomonas vaginalis* legumain-like cysteine proteinase 2 (legu-2)
mRNA, complete cds.
ACCESSION AY326447
VERSION AY326447.1 GI:37654531
KEYWORDS .
SOURCE *Trichomonas vaginalis*
ORGANISM *Trichomonas vaginalis*
Eukaryota; Parabasalidea; Trichomonada; Trichomonadida;
Trichomonadidae; Trichomonadinae; Trichomonas.
REFERENCE 1 (bases 1 to 1364)
AUTHORS Leon-Felix,J., Ortega-Lopez,J., Orozco-Solis,R. and Arroyo,R.
TITLE Two novel asparaginyl endopeptidase-like cysteine proteinases from
the protist *Trichomonas vaginalis*: their evolutionary relationship
within the clan CD cysteine proteinases
JOURNAL Gene 335, 25-35 (2004)
PUBMED 15194187
REFERENCE 2 (bases 1 to 1364)
AUTHORS Leon-Felix,J., Ortega-Lopez,J., Orozco-Solis,R. and Arroyo,R.
TITLE Direct Submission
JOURNAL Submitted (19-JUN-2003) Experimental Pathology, CINVESTAV-IPN, Av.
IPN # 2508, Col. San Pedro Zacatenco, Mexico, D.F. 07360, Mexico
FEATURES Location/Qualifiers
source 1..1364
/organism="Trichomonas vaginalis"
/mol_type="mRNA"
/isolate="CNCD 147"
/db_xref="taxon:5722"
gene 1..1364
/gene="legu-2"
/note="Tvlegu-2"
CDS 42..1289
/gene="legu-2"
/note="TVLEGU-2; asparagine endopeptidase/GPI
transamidase"
/codon_start=1
/product="legumain-like cysteine proteinase 2"
/protein_id="AAQ93040.1"
/db_xref="GI:37654532"
/translation="MFLVFSALSVSKQWAVLMAGSRGYNRYHQADIFHIYDIKTRG
FPKENIITLAYNDVVRHKDNPYPGKIFATADHKVYPGRENIDYTGQDANAENFFRVL
LGDTHNGRALQSTAEDDVVYDDHGAPGLLCPVHNNGPEIYADNIASVISQMKKEKK
FRNLFFVIEACYSGSVALNITEPNVFIITAASDQQPSYSAQWDSRLHTRFRSNEFTQNF
LKYILEHPDGRLLIDSANAAAERTVHSHVLSFGDMKLAKLPLSTFLNLAPEEVEVNNEDS
GDSSENSVENGASTHVAALEYLQRRLKETTSKEEANAIKGQIEHEVQRRARSDKIFDGI
TRRIVSNGLPVGTGFVNYIDYDCYRTAIEGFRTYCGEIDENELAKMNIFFHLCERTDK
KTILEDIKKECPVIQWDQEELYF"
ORIGIN
1 ggacactgac atggactgaa ggagtagaag atcaaaaaat aatgtttcct gttttttccg
61 ctctatctgt ttcaaagcaa tgggcagttt taatggccgg ttctcgtggc tataacaact
121 accgtcacca agccgatatt ttccacatct acgacatcat aaagaccctg ggattttccaa
181 aagaaaaatat tattacatta gcctacaacg atgtttgtag acataaagac aaccataacc
241 caggcaaaat cttcgccaca gcagaccata aaaacgtata cccaggccct gaaaacatcg
301 attatacagg acaagatgct aatgcagaga atttcttccg tgttttactc ggcgatacac
361 acaacggcag agcattgcaa tccacagctg aagacgacgt tttcgtttat tacgatgatc
421 atgggtgctc aggtcttctc tgcgcttctc ataacaatgg acccgaaatt tatgctgata
481 acatcgctc cgttatttcc cagatgaaaa aggaaaagaa attccgtaat ctcttcttcg
541 ttattgaagc ttgttattcc ggctccgctg cgcttaacat caccgagcca aatgttttta
601 tcattacagc ggcttcagac cagcagccat cttactcgcg tcagtgggat tcgaggctcc
661 atactttcag atctaagtag ttcacacaga acttccttaa gtacattttg gagcaccag
721 atggcagatt aatcgattcc gccaacgcag ctgccgaacg aacagtcacc agccatgttt
781 tatcatttgg cgatatgaag cttgcaaaat taccactttc caccttctta ctcaacgcga
841 caaccagaaga ggtaaacac gaggactctg gcgattccga aaactcagta gaaaacggtg
901 catcaacaca tgttgacgct cttgagatc tccagcgccg tttgaaggag acaacatcca
961 aggaagaggc caacgccatc aaaggacaga tcgaacatga agtccagcgt agagctcgct
1021 ctgataagat atttgatgga attaccgca gaattgtttc caatggctta cctgttgga
1081 caaagttcgt caactacatc gattacgact gctatagaac agctattgaa ggtttccgta
1141 cttatttggg tgaattgat gaaaatgagc ttgcaaagat gaatttttc actcaccttt
1201 gcgaacgcac agacaagaag acaattttag aagatatcaa gaaggatgc cctgttatcc
1261 agtgggatca ggaagaactt tacttctaata cataattttt ttattttaat tcttttttta

```
1321 tttaagagt ctatgcagtt ttaaaaaaaaa aaaaaaaaaa aaaa  
//
```

10. *Trichomonas vaginalis* legumain-like cysteine proteinase 1 (legu-1)

LOCUS AY326446 1225 bp mRNA linear INV 01-JUL-2004
 DEFINITION *Trichomonas vaginalis* legumain-like cysteine proteinase 1 (legu-1) mRNA, complete cds.
 ACCESSION AY326446
 VERSION AY326446.1 GI:39573849
 KEYWORDS .
 SOURCE *Trichomonas vaginalis*
 ORGANISM *Trichomonas vaginalis*
 Eukaryota; Parabasalidea; Trichomonada; Trichomonadida;
 Trichomonadidae; Trichomonadinae; Trichomonas.
 REFERENCE 1 (bases 1 to 1225)
 AUTHORS Leon-Felix,J., Ortega-Lopez,J., Orozco-Solis,R. and Arroyo,R.
 TITLE Two novel asparaginyl endopeptidase-like cysteine proteinases from the protist *Trichomonas vaginalis*: their evolutionary relationship within the clan CD cysteine proteinases
 JOURNAL *Gene* 335, 25-35 (2004)
 PUBMED 15194187
 REFERENCE 2 (bases 1 to 1225)
 AUTHORS Leon-Felix,J., Ortega-Lopez,J., Orozco-Solis,R. and Arroyo,R.
 TITLE Direct Submission
 JOURNAL Submitted (19-JUN-2003) Experimental Pathology, CINVESTAV-IPN, Av. IPN # 2508, Col. San Pedro Zacatenco, Mexico, D.F. 07360, Mexico
 FEATURES
 source Location/Qualifiers
 1..1225
 /organism="Trichomonas vaginalis"
 /mol_type="mRNA"
 /isolate="CNCD 147"
 /db_xref="taxon:5722"
 gene 1..1225
 /gene="legu-1"
 /note="Tvlegu-2"
 CDS 13..1179
 /gene="legu-1"
 /note="TVLEGU-1; asparagine endopeptidase/GPI transamidase"
 /codon_start=1
 /product="legumain-like cysteine proteinase 1"
 /protein_id="AAQ93039.1"
 /db_xref="GI:39573850"
 /translation="MFCLQLARCDRFAVLIAGSNDFYNYRHQADIFNMYQQLVKRGFDDQHITMMAYDDIALSENPFGRKVFHTLKHVNIYPGSSKINIAHNSVTADQFYTVLT TLKSTTSDNVYIYDNDHGGPGILGVPDGVPPGYIEAEPLAKAFDTMEAKGLYKGLFFG IEACYSGSVAAVFRANKMCTITAANDESSYAAVYDSTVGVYLSNEFSNYFMAYLDSN PQNTIGNLYTKVKAQTTGSHVCYYGDVNMKNLKLSDFLGTPNEVVAPKADAKIDIIPH YLATKSTLYQLAQSTDAKIAGRAKVALHEVIAAAEKLDLTLTSLAEILEPETKNVLR AKCGKITPEYFEVLHYFTEKYGVVKGDDMIKLRVLVNLALKHKVADIKAAIDAIC"
 ORIGIN
 1 agttaccaat aatggttctg ccttcttcag ctgcgaaggt gtgatagatt cgcagttctt
 61 attgctgggt ccaacgactt ctacaactat cgccaccagg cccgatattt caacatgtac
 121 cagcaacttg ttaagcgtgg attcgatgat cagcacatca caatgatggc ttatgatgat
 181 attgctttat catctgaaaa ccattcaga ggcaaggctt tccacacact taagcactgc
 241 aacatctacc caggttcctc caagatcaac tacgctcaca attccgctac agctgaccag
 301 ttctacacag tccttacaac actcaagtcc acaacaagcg ataactgcta catctactac
 361 gataaccacg gtggcccagg cattcttggg gttccagatg gtgttccagg aggcctacac
 421 gaggctgaac cacttgctaa ggccttcgac acaatggaag ctaagggact ttacggcaag
 481 ctcttctctg gtatcgaagc ttgctactcc ggctccgctg ctgctgtctt ccgtgctaag
 541 aacatgtgca caatcacagc tgctaacgat gatgaatcat cctatgctgc tgtctacgat
 601 tccacagctc gcgtatatct ttccaacgaa ttctccaact acttcatggc ctacctcgat
 661 tccaaccac agaacacaat cggcaacctc tacacaaagg tcaagggcca gacaacaggc
 721 tcccacggtt gctactacgg tgatgtcaac atgaagaacc tcaagctctc cgacttctc
 781 gccacaccaa acgaggttgt tgctccaaag gctgatgcta agatcgacat catcccacac
 841 tacctcgcta caaagctccac cctctaccag cttgctcagt caacagatgc taagatcgct
 901 gccgctgcta aggttgctct ccacgaggtt atcgtgctg ctgaaaagct ccatctcaca
 961 ctacacatca tcgctgaaa cctcgagcca gagacaaaga acgttctccg tgctaagtgc
 1021 gccaatgca caccagaata cttcaggtc ctccactact tcacagagaa gtatggcgtc
 1081 gtcaagggtg atgacatgat caagctcaga gttcttgta acctcgctct caagcacaag
 1141 gttgcagaca tcaaggctgc tattgatgcc atctgctaag tgcgggaata aaagtatcgt
 1201 gattctggaa aaaaaaaaaa aaaaa