Supplementary Table 1. Results of BLASTn and BLASTx homology searches of the selected markers to *Trichomonas vaginalis*.

Sequence obtained by metagenomics analysis, putatively encoding for a cathepsin L-like cysteine peptidase

CAGTATGGTCCAGCTGCCATTGCCATTGATGCATCTCATTACTCATTCCAACTTTACAGCAAAGGCATTTATGATGAAAAGAGATGCTCAACATCAAACTTAGATCACGCTGTTGGTTGCGTTGGTTATGGTTCAGAGAATGGCAAAAACTACTGGATTGTTCGTAACTCATGGGGTGACTCATGGGGTGAAAAAGGCTACATCCGTATGATCAAGGGCAAGAATAACAAATGTGGTATTGCATCAATGGCACTTGTAGTTATTCCATAAAATGAATAGTATTTTACATCTTTTACACCAATTTCATCGTTTTATAATGTAAACCATTCAAAATAAATTGTGGTCAAGCAAAAGTGACAAAAATATGAAAATTAGTTTTAGTTGGTTGATTCACAGGTTGGTAAGACATGCGTGTTAACTCGCTTAACCATGAACACATTCCAGACATCCAATCCAGCTACAATTGGAGCTGCATATAAAGGTTACAAGATAGTTAC

**BLASTn**

[ref|XM\_001328347.1|](http://www.ncbi.nlm.nih.gov/nucleotide/123502828?report=genbank&log$=nuclalign&blast_rank=1&RID=9F890RGN014) [Gene info linked to XM_001328347.1](http://www.ncbi.nlm.nih.gov/gene?term=123502828%5bNUID%5d&RID=9F890RGN014&log$=genealign&blast_rank=1) Trichomonas vaginalis G3 Clan CA, family C1, cathepsin L-like cysteine peptidase (TVAG\_465470) partial mRNA

Length=918

[GENE ID: 4774166 TVAG\_465470](http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=search&term=4774166&RID=9F890RGN014&log$=geneexplicitnucl&blast_rank=1) | Clan CA, family C1, cathepsin L-like cysteine

peptidase [Trichomonas vaginalis G3] (10 or fewer PubMed links)

Score = 187 bits (206), Expect = 1e-43

Identities = 196/253 (77%), Gaps = 10/253 (4%)

Strand=Plus/Plus

Query 3 TATGGTCCAGCTGCCATTGCCATTGATGCATCTCATTACTCATTCCAACTTTA--CAGCA 60

||||||||||||||| |||| |||||||| ||| || ||||||||| |||| || |

Sbjct 649 TATGGTCCAGCTGCCGTTGCAATTGATGCTTCTGCTTGGTCATTCCAAGTTTATTCATC- 707

Query 61 AAGGCATTTATGATGAAAAGAGATGCTCAACATCAAACTTAGATCACGCTGTTGGTTGCG 120

|||||| ||||||||| |||||| | | ||| | ||||| | |||||| || |

Sbjct 708 -AGGCATCTATGATGAACCATCATGCTCCTCTTACAACCTCGATCATGGTGTTGGCTGTG 766

Query 121 TTGGTTATGGTTCAGAGAATGGC---AAAAACTACTGGATTGTTCGTAACTCATGGGGTG 177

| |||| ||| | |||| ||| || ||||| |||||||||||||| || |||||||

Sbjct 767 TAGGTTTCGGT--AAAGAA-GGCTCTAAGAACTATTGGATTGTTCGTAATTCTTGGGGTG 823

Query 178 ACTCATGGGGTGAAAAAGGCTACATCCGTATGATCAAGGGCAAGAATAACAAATGTGGTA 237

| | ||||||||||| ||||| ||||| |||||||||| ||||| ||| |||| || |

Sbjct 824 AATATTGGGGTGAAAAGGGCTATATCCGCATGATCAAGGATAAGAACAACCAATGCGGCA 883

Query 238 TTGCATCAATGGC 250

||||| |||||||

Sbjct 884 TTGCAACAATGGC 896

[ref|XM\_001327403.1|](http://www.ncbi.nlm.nih.gov/nucleotide/123498601?report=genbank&log$=nuclalign&blast_rank=2&RID=9F890RGN014) [Gene info linked to XM_001327403.1](http://www.ncbi.nlm.nih.gov/gene?term=123498601%5bNUID%5d&RID=9F890RGN014&log$=genealign&blast_rank=2) Trichomonas vaginalis G3 Clan CA, family C1, cathepsin L-like cysteine peptidase (TVAG\_202090) partial mRNA Length=1179

[GENE ID: 4773216 TVAG\_202090](http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=search&term=4773216&RID=9F890RGN014&log$=geneexplicitnucl&blast_rank=2) | Clan CA, family C1, cathepsin L-like cysteine

peptidase [Trichomonas vaginalis G3] (10 or fewer PubMed links)

Score = 179 bits (198), Expect = 1e-41

Identities = 184/239 (77%), Gaps = 6/239 (3%)

Strand=Plus/Plus

Query 1 AGTATGGTCCAGCTGCCATTGCCATTGATGCATCTCATTACTCATTCCAACTTTACAGCA 60

|||| ||||||||||| || || || ||||| || || ||||||||||| |||||| |

Sbjct 880 AGTACGGTCCAGCTGCTATCGCTATCGATGCTTCCCACTACTCATTCCAGCTTTACTCCT 939

Query 61 AAGGCATTTATGATGAAAAGAGATGCTCAACATCAAACTTAGATCACGCTGTTGGTTGCG 120

||||||||||||||| ||||| || | | | ||||||||||| || ||||

Sbjct 940 CCGGCATTTATGATGAATCCTCCTGCTCCCCAGAAGGCCTCGATCACGCTGTCGGCTGCG 999

Query 121 TTGGTTATGGTTCAGAGAATGGC---AAAAACTACTGGATTGTTCGTAACTCATGGGGTG 177

| || || ||||| ||| ||| || ||||||||||| ||||||||||| ||||| |

Sbjct 1000 TCGGCTACGGTTCTGAG---GGCTCAAAGAACTACTGGATCGTTCGTAACTCCTGGGGCG 1056

Query 178 ACTCATGGGGTGAAAAAGGCTACATCCGTATGATCAAGGGCAAGAATAACAAATGTGGT 236

|| |||||||| || ||||||||||| |||||||||| |||||| ||| | || |||

Sbjct 1057 TTTCTTGGGGTGAGAAGGGCTACATCCGCATGATCAAGGACAAGAACAACCAGTGCGGT 1115

**BLASTx**

[ref|XP\_001327438.1|](http://www.ncbi.nlm.nih.gov/protein/123498602?report=genbank&log$=protalign&blast_rank=1&RID=9F88CZ9P014) [Gene info linked to XP_001327438.1](http://www.ncbi.nlm.nih.gov/gene?term=123498602%5bPUID%5d&RID=9F88CZ9P014&log$=genealign&blast_rank=1) Clan CA, family C1, cathepsin L-like cysteine peptidase [Trichomonas vaginalis G3]

[emb|CAA54435.1|](http://www.ncbi.nlm.nih.gov/protein/452292?report=genbank&log$=protalign&blast_rank=1&RID=9F88CZ9P014) cysteine proteinase, putative [Trichomonas vaginalis]

[gb|EAY15215.1|](http://www.ncbi.nlm.nih.gov/protein/121910367?report=genbank&log$=protalign&blast_rank=1&RID=9F88CZ9P014) [Gene info linked to EAY15215.1](http://www.ncbi.nlm.nih.gov/gene?term=121910367%5bPUID%5d&RID=9F88CZ9P014&log$=genealign&blast_rank=1) Clan CA, family C1, cathepsin L-like cysteine peptidase [Trichomonas vaginalis G3] Length=309

Score = 146 bits (369), Expect = 7e-40, Method: Compositional matrix adjust.

Identities = 70/85 (82%), Positives = 72/85 (85%), Gaps = 0/85 (0%)

Frame = +3

Query 3 YGPAAIAIDASHYSFQLYSKGIYDEKRCSTSNLDHAVGCVGYGSENGKNYWIVRNSWGDS 182

YGPAAIAIDASHYSFQLYS GIYDE CS LDHAVGCVGYGSE KNYWIVRNSWG S

Sbjct 220 YGPAAIAIDASHYSFQLYSSGIYDESSCSPEGLDHAVGCVGYGSEGSKNYWIVRNSWGVS 279

Query 183 WGEKGYIRMIKGKNNKCGIASMALV 257

WGEKGYIRMIK KNN+CG AS A +

Sbjct 280 WGEKGYIRMIKDKNNQCGEASAACI 304

Sequence obtained by metagenomics analysis, putatively encoding for a Laminin A

CATATGGTGTTGTTGGTGGTCTTAAATCATAACGAGAAGCAATTTGTTGAGCTGGATCGCCTTGTGAATAATTATCACGAAGGTAATTGTTATATCTCATAAATTGTTTGAATTGTTCAAAGGTCTTAATACGTGGAGCTTCTCTCTTCATGATCAAATATCTTGGACCTTCCTTATATGAACGATATGCACCATACTTACCCATTGATTTAACAAGTTCTGGATAACCAGCAAGGTTATAAAGTTCTTCGTGCCATGGACAATTGATTGATGGGAAGTACAAATCTCTGACAAGTTGGAATGTAACATCTGTCATTCTGTATGTACCTGGGAATTGTTCAATGATCCATAAAAGATCTTCTGTTGGCTTAACACCAAATTGCAATTTCTTTGAATCAACAATAACATACTGATTATTGTATGTACCTGAATTGTGCTTAATAAATTCTG

**BLASTn**

[ref|XM\_001308649.1|](http://www.ncbi.nlm.nih.gov/nucleotide/123433660?report=genbank&log$=nuclalign&blast_rank=1&RID=9PY625PE014) [Gene info linked to XM_001308649.1](http://www.ncbi.nlm.nih.gov/gene?term=123433660%5bNUID%5d&RID=9PY625PE014&log$=genealign&blast_rank=1)[Genome view with mapviewer linked to XM_001308649.1](http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?direct=on&gbgi=123433660&THE_BLAST_RID=9PY625PE014&log$=mapalign&blast_rank=1) Trichomonas vaginalis G3 Laminin A family protein (TVAG\_123110) partial mRNA Length=1930

[GENE ID: 4753482 TVAG\_123110](http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=search&term=4753482&RID=9PY625PE014&log$=geneexplicitnucl&blast_rank=1) | Laminin A family protein

[Trichomonas vaginalis G3] (10 or fewer PubMed links)

Score = 149 bits (164), Expect = 2e-32

Identities = 298/429 (69%), Gaps = 16/429 (4%)

Strand=Plus/Minus

Query 26 ATCATAACGAGAAGCAATTTGTTGAGCTGGATCGCCTTGTGAATAATTATCACGAAGGTA 85

||| || | ||||||||| ||||| ||||||||||||||||| || ||||| || |

Sbjct 1356 ATCGTATCTAGAAGCAATCTGTTGGGCTGGATCGCCTTGTGAGTAGTTATCGCGCTTCCA 1297

Query 86 ATTGTTATATCTCATAAATTGTTTGAATTGTTCAAAGGTCTTAATACGTGGAG-CTTCTC 144

| ||| || | ||| |||| ||||| | ||| || ||||| || | | || | || |

Sbjct 1296 CTGGTTGTAACGCATGAATTTCTTGAAATCTTCGAATGTCTTGATCC-TCCAGACATCGC 1238

Query 145 TCTTCATGATCAAATATCTTGG--ACCT-TCCTTATATGAACGATATGCACCATACTTAC 201

| ||| || | || | ||| || | | ||| || | || | || || | |

Sbjct 1237 GTTGCATAATGAGGTAACGTGGGGACTTGTAGTTAGAT---CTGTAGTCGCCGTAGATGC 1181

Query 202 CCATTGATTTAACAAGTTCTGGATAACCAGCAAGGTTATAAAGTTCTTCGTGCCATGGAC 261

| | || ||||| |||||||| || || ||||| ||||| |||||||| |||| |

Sbjct 1180 CGAGAGAGGCTACAAGCTCTGGATAGCCGGCGAGGTTGTAAAGGTCTTCGTGGTATGGGC 1121

Query 262 AATTGATTGATGGGAAGTA-CAAATCTCTGACAAGTTGGAATGTAACATCTGTCATTCTG 320

| |||| ||||||||||| | ||| ||||| | ||||||||| | | | |

Sbjct 1120 AGTTGACAGATGGGAAGTATCCGTGCTC-AACAAGGTAATCTGTAACATCGGAACTGCGG 1062

Query 321 TATGTACCTGGGAATTGTTCAATGATCCATAAAAGATCTTCTGT-TGGCTTAACACCAAA 379

|| ||||||||| || || |||||||| | |||||| | ||| ||||| | ||

Sbjct 1061 TAGACACCTGGGAACTGCTCGATGATCCAGATAAGATCCT-TGTCGGGCTT--CTGGAAG 1005

Query 380 TTG--CAATTTCTTTGAATCAACAATAACATACTGATTATTGTATGTACCTGAATTGTGC 437

| |||||||| |||||||||| || ||||| || || |||||||| || ||||||

Sbjct 1004 CGGGTTAATTTCTTGGAATCAACAACGACGTACTGGTTGTTATATGTACCGGAGTTGTGC 945

Query 438 TTAATAAAT 446

|| ||||||

Sbjct 944 TTGATAAAT 936

**BLASTx**

[ref|XP\_001308650.1|](http://www.ncbi.nlm.nih.gov/protein/123433661?report=genbank&log$=protalign&blast_rank=1&RID=9PY5WZB9014) [Gene info linked to XP_001308650.1](http://www.ncbi.nlm.nih.gov/gene?term=123433661%5bPUID%5d&RID=9PY5WZB9014&log$=genealign&blast_rank=1)[Genome view with mapviewer linked to XP_001308650.1](http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?direct=on&gbgi=123433661&THE_BLAST_RID=9PY5WZB9014&log$=mapalign&blast_rank=1) Laminin A family protein [Trichomonas vaginalis G3] [gb|EAX95720.1|](http://www.ncbi.nlm.nih.gov/protein/121890341?report=genbank&log$=protalign&blast_rank=1&RID=9PY5WZB9014) [Gene info linked to EAX95720.1](http://www.ncbi.nlm.nih.gov/gene?term=121890341%5bPUID%5d&RID=9PY5WZB9014&log$=genealign&blast_rank=1) Laminin A family protein [Trichomonas vaginalis G3]

Length=622

[GENE ID: 4753482 TVAG\_123110](http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=search&term=4753482&RID=9PY5WZB9014&log$=geneexplicitprot&blast_rank=1) | Laminin A family protein

[Trichomonas vaginalis G3] (10 or fewer PubMed links)

Score = 225 bits (573), Expect = 3e-67, Method: Compositional matrix adjust.

Identities = 110/149 (74%), Positives = 123/149 (83%), Gaps = 1/149 (1%)

Frame = -3

Query 445 FIKHNSGTYNNQYVIVDSKKLQFGVKPTEDLLWIIEQFPGTYRMTDVTFQLVRDLYFPSI 266

FIKHNSGTYNNQYV+VDSKKL KP +DL+WIIEQFPG YR +DVT LV YFPS+

Sbjct 313 FIKHNSGTYNNQYVVVDSKKLTRFQKPDKDLIWIIEQFPGVYRSSDVTDYLVEHGYFPSV 372

Query 265 NCPWHEELYNLAGYPELVKSMGKYGAYRSYKEGPRYLIMKREAPRIKTFEQFKQFMRYNN 86

NCP+HE+LYNLAGYPELV S+G YG YRS + PRYLIM+R+ RIKTFE FK+FMRYN

Sbjct 373 NCPYHEDLYNLAGYPELVASLGIYGDYRSNYKSPRYLIMQRDVWRIKTFEDFKKFMRYNQ 432

Query 85 YLRDNYSQGDPAQQIASRYDLRPPT-TPY 2

+ RDNYSQGDPAQQIASRYDLRP TPY

Sbjct 433 WKRDNYSQGDPAQQIASRYDLRPAEGTPY 461

Sequence obtained by metagenomics analysis, putatively encoding for a TKL family protein kinase

CCATAAACTTGATGAAACTATGAGAACCATCGCAATGTATGATATTGCTCGAGGAATGCAATTTTTACACTCAAGACAGATTATTCACAGAGACTTAAAATCATTAAATGTACTTTTGGATAAAATGGATATACAAAAATTTGCGATTTTGGATTCTCTAGAAAGGCTAATAAAGAAGATGTTTTAACACAAAATATCGGAACACCTCATTGGATGGCTCCGGAATTATTAAATGCTGAAAATGGATATGATAATAAGATTGATGTTTATGCCTATGGGATTGTTTGCTGGGAATTATTAACAGGACAATTACCTTATGCAGGTTTAGAATCAACACAGATCATTGCTATGGTTTTAATGAATGACATGAGACCACAATTTCTTCAAACGTCTCAGCTCAGTTAAAGACTTTTATTTCTGATTGTTGGGCTCGTGATCCAAAAATGAGACCAACATTTAATACAAATCGTTCGAAGATTAAAGAAAGGAGATATTTTAACT

**BLASTn**

[ref|XM\_001319512.1|](http://www.ncbi.nlm.nih.gov/nucleotide/123472709?report=genbank&log$=nuclalign&blast_rank=1&RID=9PYMZF0R01R) [Gene info linked to XM_001319512.1](http://www.ncbi.nlm.nih.gov/gene?term=123472709%5bNUID%5d&RID=9PYMZF0R01R&log$=genealign&blast_rank=1) Trichomonas vaginalis G3 TKL family protein kinase (TVAG\_139540) partial mRNA Length=2394

[GENE ID: 4765212 TVAG\_139540](http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=search&term=4765212&RID=9PYMZF0R01R&log$=geneexplicitnucl&blast_rank=1) | TKL family protein kinase

[Trichomonas vaginalis G3] (10 or fewer PubMed links)

Score = 111 bits (122), Expect = 6e-21

Identities = 245/366 (67%), Gaps = 1/366 (0%)

Strand=Plus/Plus

Query 1 CCATAAACTTGATGAAACTATGAGAACCATCGCAATGTATGATATTGCTCGAGGAATGCA 60

||| ||| | ||| || ||||| ||||| || | |||||| ||| |||| |||||

Sbjct 888 CCACAAATTGGATACGACCATGAGGACCATTGCCGCATTTGATATCGCTAGAGGTATGCA 947

Query 61 ATTTTTACACTCAAGACAGATTATTCACAGAGACTTAAAATCATTAAATGTACTTTTGGA 120

| || || | || |||||||| ||||| || |||||||||||||| | |||||

Sbjct 948 GGAACTCCATTCCAAGCATATTATTCATAGAGATTTGAAATCATTAAATGTTTTATTGGA 1007

Query 121 TAAAATGGAT-ATACAAAAATTTGCGATTTTGGATTCTCTAGAAAGGCTAATAAAGAAGA 179

||| || || | | || ||||| || ||||||||| | || | |||| |

Sbjct 1008 CAAAGACTATCATGTACACATCTGCGACTTCGGATTCTCTCGTGGAGCAGGTGAAGAGCA 1067

Query 180 TGTTTTAACACAAAATATCGGAACACCTCATTGGATGGCTCCGGAATTATTAAATGCTGA 239

| | || ||||| | ||||| || ||||||||||| || |||||| | || |

Sbjct 1068 ACTATACACTCAAAACGTTGGAACTCCCCATTGGATGGCCCCAGAATTACTTGATTCGTC 1127

Query 240 AAATGGATATGATAATAAGATTGATGTTTATGCCTATGGGATTGTTTGCTGGGAATTATT 299

| | | || || |||| | |||||||| || || || || || ||||| || |

Sbjct 1128 ACACAGTTACAATTATAAAGTCGATGTTTACGCGTACGGTATCGTACTTTGGGAGATAAT 1187

Query 300 AACAGGACAATTACCTTATGCAGGTTTAGAATCAACACAGATCATTGCTATGGTTTTAAT 359

||| | ||||| || ||| ||||||| || || || || ||||| || | ||

Sbjct 1188 GACATGCCAATTGCCGTATTCAGGTTTGGAGAGTACTCAAATTATTGCACAAGTAATGAT 1247

Query 360 GAATGA 365

||||||

Sbjct 1248 GAATGA 1253

[ref|XM\_001581078.1|](http://www.ncbi.nlm.nih.gov/nucleotide/154416210?report=genbank&log$=nuclalign&blast_rank=2&RID=9PYMZF0R01R) [Gene info linked to XM_001581078.1](http://www.ncbi.nlm.nih.gov/gene?term=154416210%5bNUID%5d&RID=9PYMZF0R01R&log$=genealign&blast_rank=2) Trichomonas vaginalis G3 TKL family protein kinase (TVAG\_020870) partial mRNA Length=2361

[GENE ID: 5465674 TVAG\_020870](http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=search&term=5465674&RID=9PYMZF0R01R&log$=geneexplicitnucl&blast_rank=2) | TKL family protein kinase

[Trichomonas vaginalis G3] (10 or fewer PubMed links)

Score = 96.9 bits (106), Expect = 1e-16

Identities = 185/268 (69%), Gaps = 6/268 (2%)

Strand=Plus/Plus

Query 29 ATCGCAATGTATGATATTGCTCGAGGAATGCAATTTTTACACTCAAGACAGATTATTCAC 88

|| ||||||| |||||||| |||||||| || |||||| ||||| |||||||||

Sbjct 913 ATTGCAATGTTCGATATTGCAAGAGGAATGAGATGTTTACATAACAGACATATTATTCAC 972

Query 89 AGAGACTTAAAATCATTAAATGTACTTTTGGATAAAATGGATATA-CAAAAATTTGCGAT 147

|||||||| ||| | | ||||| | | ||| || ||| | | ||||| |||

Sbjct 973 AGAGACTTGAAAACTCTCAATGTTTTAATTGATGCAAACAATAGAGCTAAAATCATTGAT 1032

Query 148 TTTGGATTCTCTAGAAAGGCTAATAAAGAAGATGTTTTAACACAAAATATCGGAACACCT 207

||||| || | | | || ||| ||| | || | || ||| ||| || |||||

Sbjct 1033 TTTGGCTTAAGTCGCTACGCAAAT---GAACAATTTATGTCAGAAAGTATAGGTACACCA 1089

Query 208 CATTGGATGGCTCCGGAA-TTATTAAATGCTGAAAATGGATATGATAATAAGATTGATGT 266

|||||||||||||| ||| | || || | || ||| ||| ||| | |||||

Sbjct 1090 CATTGGATGGCTCCAGAACTCCTTGGAT-CGTCCAAGCAATACGATTTGAAGGTCGATGT 1148

Query 267 TTATGCCTATGGGATTGTTTGCTGGGAA 294

||| || |||| |||||||| ||||||

Sbjct 1149 TTACGCTTATGCCATTGTTTGTTGGGAA 1176

**Blastx**

[ref|XP\_001319547.1|](http://www.ncbi.nlm.nih.gov/protein/123472710?report=genbank&log$=protalign&blast_rank=1&RID=9PYN7YSM01R) [Gene info linked to XP_001319547.1](http://www.ncbi.nlm.nih.gov/gene?term=123472710%5bPUID%5d&RID=9PYN7YSM01R&log$=genealign&blast_rank=1) TKL family protein kinase [Trichomonas vaginalis G3]

[gb|EAY07324.1|](http://www.ncbi.nlm.nih.gov/protein/121902333?report=genbank&log$=protalign&blast_rank=1&RID=9PYN7YSM01R) [Gene info linked to EAY07324.1](http://www.ncbi.nlm.nih.gov/gene?term=121902333%5bPUID%5d&RID=9PYN7YSM01R&log$=genealign&blast_rank=1) TKL family protein kinase [Trichomonas vaginalis G3]

Length=797

[GENE ID: 4765212 TVAG\_139540](http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=search&term=4765212&RID=9PYN7YSM01R&log$=geneexplicitprot&blast_rank=1) | TKL family protein kinase

[Trichomonas vaginalis G3] (10 or fewer PubMed links)

Score = 132 bits (333), Expect(3) = 3e-48, Method: Compositional matrix adjust.

Identities = 56/88 (64%), Positives = 73/88 (83%), Gaps = 0/88 (0%)

Frame = +1

Query 130 YTKICDFGFSRKANKEDVLTQNIGTPHWMAPELLNAENGYDNKIDVYAYGIVCWELLTGQ 309

+ ICDFGFSR A +E + TQN+GTPHWMAPELL++ + Y+ K+DVYAYGIV WE++T Q

Sbjct 340 HVHICDFGFSRGAGEEQLYTQNVGTPHWMAPELLDSSHSYNYKVDVYAYGIVLWEIMTCQ 399

Query 310 LPYAGLESTQIIAMVLMNDMRPQFLQTS 393

LPY+GLESTQIIA V+MND+RP +++

Sbjct 400 LPYSGLESTQIIAQVMMNDLRPSIPEST 427

Score = 73.9 bits (180), Expect(3) = 3e-48, Method: Composition-based stats.

Identities = 35/41 (85%), Positives = 37/41 (90%), Gaps = 0/41 (0%)

Frame = +2

Query 2 HKLDETMRTIAMYDIARGMQFLHSRQIIHRDLKSLNVLLDK 124

HKLD TMRTIA +DIARGMQ LHS+ IIHRDLKSLNVLLDK

Sbjct 297 HKLDTTMRTIAAFDIARGMQELHSKHIIHRDLKSLNVLLDK 337

Score = 30.8 bits (68), Expect(3) = 3e-48, Method: Composition-based stats.

Identities = 10/29 (34%), Positives = 17/29 (59%), Gaps = 0/29 (0%)

Frame = +3

Query 375 TISSNVSAQLKTFISDCWARDPKMRPTFN 461

+I + + L+ + CW R+P RPTF+

Sbjct 422 SIPESTNGPLRDLTTSCWDRNPDRRPTFD 450

Sequence obtained by metagenomics analysis, putatively encoding for a Clan MH, family M20, peptidase T-like metallopeptidase

ATCTACCTTGATGGCCATAGCGATACTGTTTTCCCACTCCGCGAGAACTGGAAGAATGTTTTAGGTGAAGGTATCGACTGCTTCGATGGCCTGAAGGATGCTGAGAAGGTTAACTTTGAAAATATGCAGCACGAATTGAAGTACATTCCACCCAAGGAAAAATGGGACCATCTCTTATTCGGTCGTGGCTCAGCTGATCAATTACAAGGTGTTGTTTCGCAAGTGTTCGCCACAAAGATTCTTCTTGAAACACGTGAGCTCGGCTCATTGAAGGGTGCAAAGATTGTTTCAATTGCCACTGTTACAGAAGAAGACAATGATGGTGGTGCTCCAATGCACATCATGAGAAAGCAAAACTTGAAGCATGAACAAGTTCCAGACTGTGTTATCCTCACAGAAGGCACTGGCGATCTTGATCTCGGTCCATGCGGCATTTACATCGGCCAACGTGGTAGATGCCAAGTTGAAGTTGAAGTTATTGGTAA

**BLASTn**

[ref|XM\_001311980.1|](http://www.ncbi.nlm.nih.gov/nucleotide/123446460?report=genbank&log$=nuclalign&blast_rank=1&RID=922J19WK016) [Gene info linked to XM_001311980.1](http://www.ncbi.nlm.nih.gov/gene?term=123446460%5bNUID%5d&RID=922J19WK016&log$=genealign&blast_rank=1)[Genome view with mapviewer linked to XM_001311980.1](http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?direct=on&gbgi=123446460&THE_BLAST_RID=922J19WK016&log$=mapalign&blast_rank=1) Trichomonas vaginalis G3 Clan MH, family M20, peptidase T-like metallopeptidase (TVAG\_433790) partial mRNA

Length=1590

[GENE ID: 4756856 TVAG\_433790](http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=search&term=4756856&RID=922J19WK016&log$=geneexplicitnucl&blast_rank=1) | Clan MH, family M20, peptidase T-like

metallopeptidase [Trichomonas vaginalis G3] (10 or fewer PubMed links)

Score = 232 bits (256), Expect = 3e-57

Identities = 346/489 (71%), Gaps = 8/489 (2%)

Strand=Plus/Plus

Query 1 ATCTACCTTGATGGCCATAGCGATACTGTTTTCCCACTCCGCGAGAACTGGAAGAATGTT 60

|||||| | |||||||||| || || |||| |||||| || || | ||| | ||

Sbjct 343 ATCTACATGGATGGCCATACTGACACAGTTTACCCACTTCGTGAAGAATGGCACAAGCGC 402

Query 61 TTAGGTGAAGGTATCGACTGCTTCGATGGCCTGAAGGATGCTGAGAAGGTTAACTTTGAA 120

| ||||| || |||||||||| | ||||| | | ||| || ||||| | |||

Sbjct 403 CTCGGTGAGGGCATCGACTGCTACAATGGCTTAACAGATAAGGATAAGGTCTCCGAGGAA 462

Query 121 AATATGCAGCACGAATTGAAGTACATTCCACCCAAGGAAAAATGGGACCATCTCTTATTC 180

| || |||| | |||| ||||| ||||||| ||||| | ||| | |||

Sbjct 463 GCCCTCAAGGCCGAACTCCGCCACATCCCACCAAAGGAAATGTGGGATCGCCTCCTCTTC 522

Query 181 GGTCGTGGCTCAGCTGATCAATTACAAGGTGTTGTTTCGCAAGTGTTCGCCACAAAGATT 240

|| |||||| |||| ||||| | ||||||||||| || || || || || ||||||||

Sbjct 523 GGCCGTGGCACAGCCGATCAGCTCCAAGGTGTTGTCTCCCAGGTCTTTGCTACAAAGATC 582

Query 241 CTTCTTGAAACACGTGAGCTCGGCTCATTGAAGGGTGCAAAGATTGTTTCAATTGCCACT 300

|| || || ||| | |||||||| | ||||| | || | || | | |||

Sbjct 583 CTCCTCGAGACAAAGCACCTCGGCTCTCTCAAGGGCTCCGTCATCATCTCCGTCGGCACA 642

Query 301 GTTACAGAAGAAGACAATGATGGTGGTGCTCCAATGCACATCATGAGAAAGCA-AAACTT 359

|| | ||||||||||| |||||||| ||||||||||||||||| | ||||| | |||

Sbjct 643 GTCGCTGAAGAAGACAACGATGGTGGCGCTCCAATGCACATCATCCGCAAGCAGGACCTT 702

Query 360 GAAGCATGAACAAGTTCCAGACTGTGTTATCCTCACAGAAGGCACTGGCGATCTTGATC- 418

| || | || || ||||| || || ||||| ||||||||||| |||||||||| ||

Sbjct 703 GCTCCA-CACCAGGTCCCAGATTGCGTCATCCTTACAGAAGGCACAGGCGATCTTGCTCC 761

Query 419 -----TCGGTCCATGCGGCATTTACATCGGCCAACGTGGTAGATGCCAAGTTGAAGTTGA 473

|| ||||||||||| ||||||||||| |||||| | ||||| || | || ||

Sbjct 762 AGGCAAGGGCCCATGCGGCATCTACATCGGCCAGCGTGGTCGTTGCCAGGTCCAGGTCGA 821

Query 474 AGTTATTGG 482

|||| ||||

Sbjct 822 AGTTGTTGG 830

**BLASTx**

[ref|XP\_001311981.1|](http://www.ncbi.nlm.nih.gov/protein/123446461?report=genbank&log$=protalign&blast_rank=1&RID=922HRGEN014) [Gene info linked to XP_001311981.1](http://www.ncbi.nlm.nih.gov/gene?term=123446461%5bPUID%5d&RID=922HRGEN014&log$=genealign&blast_rank=1)[Genome view with mapviewer linked to XP_001311981.1](http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?direct=on&gbgi=123446461&THE_BLAST_RID=922HRGEN014&log$=mapalign&blast_rank=1) Clan MH, family M20, peptidase T-like metallopeptidase [Trichomonas vaginalis G3]

[gb|EAX99051.1|](http://www.ncbi.nlm.nih.gov/protein/121893811?report=genbank&log$=protalign&blast_rank=1&RID=922HRGEN014) [Gene info linked to EAX99051.1](http://www.ncbi.nlm.nih.gov/gene?term=121893811%5bPUID%5d&RID=922HRGEN014&log$=genealign&blast_rank=1) Clan MH, family M20, peptidase T-like metallopeptidase [Trichomonas vaginalis G3]

Length=529

[GENE ID: 4756856 TVAG\_433790](http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=search&term=4756856&RID=922HRGEN014&log$=geneexplicitprot&blast_rank=1) | Clan MH, family M20, peptidase T-like

metallopeptidase [Trichomonas vaginalis G3] (10 or fewer PubMed links)

Score = 254 bits (650), Expect = 3e-79, Method: Compositional matrix adjust.

Identities = 121/164 (74%), Positives = 142/164 (87%), Gaps = 2/164 (1%)

Frame = +1

Query 1 IYLDGHSDTVFPLRENWKNVLGEGIDCFDGLKDAEKVNFENMQHELKYIPPKEKWDHLLF 180

IY+DGH+DTV+PLRE W LGEGIDC++GL D +KV+ E ++ EL++IPPKE WD LLF

Sbjct 115 IYMDGHTDTVYPLREEWHKRLGEGIDCYNGLTDKDKVSEEALKAELRHIPPKEMWDRLLF 174

Query 181 GRGSADQLQGVVSQVFATKILLETRELGSLKGAKIVSIATVTEEDNDGGAPMHIMRKQNL 360

GRG+ADQLQGVVSQVFATKILLET+ LGSLKG+ I+S+ TV EEDNDGGAPMHI+RKQ+L

Sbjct 175 GRGTADQLQGVVSQVFATKILLETKHLGSLKGSVIISVGTVAEEDNDGGAPMHIIRKQDL 234

Query 361 KHEQVPDCVILTEGTGDLDL--GPCGIYIGQRGRCQVEVEVIGK 486

QVPDCVILTEGTGDL GPCGIYIGQRGRCQV+VEV+G+

Sbjct 235 APHQVPDCVILTEGTGDLAPGKGPCGIYIGQRGRCQVQVEVVGQ 278

Sequence obtained by metagenomics analysis, putatively encoding for a Clan SC, family S9, unassigned serine peptidase

CATTATCATTTTGAGCTGGTGCTGGTGCCTTATCAACCATATTACTGAATGAAGAGAAATGAAAGTTTGATTTCTGTAATTTTCTACATTCTGAAATTTGGATACCATTTGTTGATATTTTAAATTGATCCATTGCAAACATAACTCCCAAAGTTATCCAATCTTTATTTCTTTTACTATTATGACCTCCTGGTAATTCAAGCATATATTTAAACTTTGAGTTATAATTGCTATAAAGTTCCAAACATTGTTCGAAAGGTATAAATTGATCTTTTTCAGCATGTCCAAATACAACAGGTACTTGATTCTCATCTTTAATCGAATTTAAAGGAGATACATCATCAAAATTGAAGTTGGCTGTGCCCATTACCTTCTTCTTCAAAAACCAAATAACCATGGGAACAAATACATTTGGAAGTTTCATTGATATCGCAATCGCAGAGCACATATCAGCAATTGACGTAAATGCAGAGTCAGATATCTTCCCAATAACATACGGGCTCTTGACCA

**BLASTn**

[ref|XM\_001328180.1|](http://www.ncbi.nlm.nih.gov/nucleotide/123502068?report=genbank&log$=nuclalign&blast_rank=1&RID=A1N2XUA101R) [Gene info linked to XM_001328180.1](http://www.ncbi.nlm.nih.gov/gene?term=123502068%5bNUID%5d&RID=A1N2XUA101R&log$=genealign&blast_rank=1) Trichomonas vaginalis G3 Clan SC, family S9, unassigned serine peptidase (TVAG\_262420) partial mRNA Length=1326

[GENE ID: 4773999 TVAG\_262420](http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=search&term=4773999&RID=A1N2XUA101R&log$=geneexplicitnucl&blast_rank=1) | Clan SC, family S9, unassigned serine peptidase

[Trichomonas vaginalis G3] (10 or fewer PubMed links)

Score = 68.0 bits (74), Expect = 7e-08

Identities = 170/256 (66%), Gaps = 7/256 (3%)

Strand=Plus/Minus

Query 246 CATTGTTCGAAAGGTATAAATTGATCTTTTTCAGCATGTCCAAATACAACAGGTACTTGA 305

||||||||||| || || |||| || ||| ||||||||||| || ||| || |

Sbjct 725 CATTGTTCGAATGGAATGAATTTGTCGCCTTCTGCATGTCCAAAAACGCAAGGAACGGG- 667

Query 306 TTCTCATCTTTAATCGAATTTAAAGGAGATACATCATCAAAATTGAAGTTGGCTGTGCCC 365

| |||| | ||||||| ||||||| || | | ||| | |||||

Sbjct 666 --GGCTTCTTGAGGACAATTTAATGGAGATATCGTTTCTAGGTCGAAATCGGCTGCTTGT 609

Query 366 ATTACCTTCTTCTTCAAAAACCAAATAACCATGGGAACAAATACATT--TGGAAGTTTCA 423

| || || |||||||||||||||||||| |||| | ||||| || || |||

Sbjct 608 AAAACTTTTTTCTTCAAAAACCAAATAACTGCTGGAATGA--ACATTGACGGTAGGCTCA 551

Query 424 TTGATATCGCAATCGCAGAGCACATATCAGCAATTGACGTAAATGCAGAGTCAGATATCT 483

|||| | ||||| || | ||||| || | || ||| ||||| | || || || |

Sbjct 550 TTGACTTGGCAATGGCTGCACACATGTCTGGAACTGATGTAAAGGATGAATCGCTGATTT 491

Query 484 TCCCAATAACATACGG 499

| |||||||||| ||

Sbjct 490 TAGCAATAACATATGG 475

**BLASTx**

[ref|XP\_001328215.1|](http://www.ncbi.nlm.nih.gov/protein/123502069?report=genbank&log$=protalign&blast_rank=1&RID=A1N35JBF01R) [Gene info linked to XP_001328215.1](http://www.ncbi.nlm.nih.gov/gene?term=123502069%5bPUID%5d&RID=A1N35JBF01R&log$=genealign&blast_rank=1) Clan SC, family S9, unassigned serine peptidase [Trichomonas vaginalis G3]

[gb|EAY15992.1|](http://www.ncbi.nlm.nih.gov/protein/121911155?report=genbank&log$=protalign&blast_rank=1&RID=A1N35JBF01R) [Gene info linked to EAY15992.1](http://www.ncbi.nlm.nih.gov/gene?term=121911155%5bPUID%5d&RID=A1N35JBF01R&log$=genealign&blast_rank=1) Clan SC, family S9, unassigned serine peptidase [Trichomonas

vaginalis G3]

Length=441

[GENE ID: 4773999 TVAG\_262420](http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=search&term=4773999&RID=A1N35JBF01R&log$=geneexplicitprot&blast_rank=1) | Clan SC, family S9, unassigned serine peptidase

[Trichomonas vaginalis G3] (10 or fewer PubMed links)

Score = 203 bits (517), Expect = 2e-60, Method: Compositional matrix adjust.

Identities = 93/161 (58%), Positives = 122/161 (76%), Gaps = 1/161 (1%)

Frame = -3

Query 508 VKSPYVIGKISDSAFTSIADMCSAIAISMKLPNVFVPMVIWFLKKKVMGTANFNFDDVSP 329

V +PYVI KISDS+FTS+ DMC+AIA SM LP++F+P VIWFLKKKV+ A+F+ + +SP

Sbjct 156 VDNPYVIAKISDSSFTSVPDMCAAIAKSMSLPSMFIPAVIWFLKKKVLQAADFDLETISP 215

Query 328 LNSIKDENQVPVVFGHAEKDQFIPFEQCLELYSNYNSKFKYMLELPGGHNSKRNKDWITL 149

LN + E VP VFGHAE D+FIPFEQC +LY NY + K+++ L GGHNSKR+ WITL

Sbjct 216 LNCPQ-EAPVPCVFGHAEGDKFIPFEQCRQLYDNYENPMKHIMILDGGHNSKRDLAWITL 274

Query 148 GVMFAMDQFKISTNGIQISECRKLQKSNFHFSSFSNMVDKA 26

GV FA+D + + ++ISECRKLQ +++HF SF M+ A

Sbjct 275 GVTFALDMLSVRVDDLKISECRKLQSASYHFGSFETMLSSA 315

Sequence obtained by metagenomics analysis, putatively encoding for a 14-3-3 protein

TTAGCCTTGCTAGCTGGTTCTTGCTTGTCATTTTCTGGTTTGGATTCGCAGATATAACGGTAATAATCAGCCTTAAGCTTTTCGTAGAAAACGCGTGATTCTGGTGTCTTGGCTGCTGGAAGAAGCTTTTCATCAACAAGTGCGATGAGTTCATTGCAGTAAGCTTCAAGTTCGCCGAGGATAACCTTCTTATATTCTTGAAGTTGTTCAACACGGGCAAGATTGTCACGACCAACTTCATGATCAATGATAGCAGCAATCATTCTGAGGCCATTACGACGTGATCCAACGATATTCTTGTAAGCAACTGAAAGAAGATTACGTTCGTCAGTGTTTAATTCTGGATTGAGTGAAATAACTTTCTTCATTGCATTAACCATATCTTGGTGACGTTCTGTTTGATCAAAATTTAGGGCCATGAAAAGATTAACATCGCGTTCAGATGACATTATGCG

**BLASTn**

[ref|XM\_001579544.1|](http://www.ncbi.nlm.nih.gov/nucleotide/154413125?report=genbank&log$=nuclalign&blast_rank=1&RID=9MZ73JZE014) [Gene info linked to XM_001579544.1](http://www.ncbi.nlm.nih.gov/gene?term=154413125%5bNUID%5d&RID=9MZ73JZE014&log$=genealign&blast_rank=1) Trichomonas vaginalis G3 14-3-3 protein (TVAG\_462940) mRNA, complete cds Length=749

[GENE ID: 5464121 TVAG\_462940](http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=search&term=5464121&RID=9MZ73JZE014&log$=geneexplicitnucl&blast_rank=1) | 14-3-3 protein [Trichomonas vaginalis G3]

(10 or fewer PubMed links)

Score = 239 bits (264), Expect = 2e-59

Identities = 328/453 (72%), Gaps = 4/453 (1%)

Strand=Plus/Minus

Query 1 TTAGCCTTGCTAGCTGGTTCTTGCTTGTCATTTTCTGGTTTGGATTCGCAGATATAACGG 60

|||||||| ||||||||| ||| |||| || | |||||||||||| || |||

Sbjct 457 TTAGCCTTCTGGGCTGGTTCTGTCTTCTCATCATCAGAGCGGGATTCGCAGATGTAGCGG 398

Query 61 TAATAATCAGCCTTAAGCTTTTCGTAGAAAACGCGTGATTCTGGTGTCTTGGCTGCTGGA 120

|| || |||||||| ||||| | |||||||||||| ||||| | ||||||||| |||||

Sbjct 397 TAGTAGTCAGCCTTGAGCTTCTGGTAGAAAACGCGGGATTCAGCTGTCTTGGCAGCTGGG 338

Query 121 AGAAGCTTTTCATCAACAAGTGCGATGAGTTCATTGCAGTAAGCTTCAAGTTCGCCGAGG 180

||||| ||| || || |||| || || |||||||| ||||||||||| |||

Sbjct 337 AGAAGGGAAAGATCGACGAGGCCGATAAGCTCGTTGCAGTACTTCTCAAGTTCGCCAAGG 278

Query 181 ATAACCTTCTTATATTC-TTGAAGTTGTTCAACACGGGCAAGATTGTCACGACCAACTTC 239

|||||||| ||| | |||| | | |||| || ||| |||| ||| ||||

Sbjct 277 ATAACCTTGCGATAGGCGTTGAGGGAG-GCAACGCGCTGTGTGTTGCCACGGCCATCTTC 219

Query 240 ATGATCAATGATAGCAGCAATCATTCTGAGGCCATTACGACGTGATCCAACGATATTCTT 299

||| || | |||||| || ||||| | ||| |||||||| || ||||| | ||| |||||

Sbjct 218 ATGCTCGAGGATAGCGGCGATCATGCGGAGACCATTACGGCGGGATCCGATGATGTTCTT 159

Query 300 GTAAGCAACTGAAAGAAGATTACGTTCGTCAGTGTTTAA-TTCTGGATTGAGTGAAATAA 358

||| ||||| ||||| || || || || || || | ||| || | | || |||||

Sbjct 158 GTATGCAACAGAAAGGAGGTTGCGCTCATC-GTTTGTAAGATCAGCAGAAAGCTCAATAA 100

Query 359 CTTTCTTCATTGCATTAACCATATCTTGGTGACGTTCTGTTTGATCAAAATTTAGGGCCA 418

| |||||||||| | |||||||| || || || || || || ||| | || | ||||

Sbjct 99 CCTTCTTCATTGTGTCGACCATATCCTGATGGCGGTCAGTCTGGTCATAGATTTGAGCCA 40

Query 419 TGAAAAGATTAACATCGCGTTCAGATGACATTA 451

|||| | | ||| |||||||| |||||||

Sbjct 39 TGAACAAAGCAACGTCGCGTTCGTTCGACATTA 7

**BLASTx**

[ref|XP\_001579594.1|](http://www.ncbi.nlm.nih.gov/protein/154413126?report=genbank&log$=protalign&blast_rank=1&RID=9MZ765BH014) [Gene info linked to XP_001579594.1](http://www.ncbi.nlm.nih.gov/gene?term=154413126%5bPUID%5d&RID=9MZ765BH014&log$=genealign&blast_rank=1) 14-3-3 protein [Trichomonas vaginalis G3]

[gb|AAV83996.1|](http://www.ncbi.nlm.nih.gov/protein/56182380?report=genbank&log$=protalign&blast_rank=1&RID=9MZ765BH014) 14-3-3 protein A isoform [Trichomonas vaginalis]

[gb|EAY18608.1|](http://www.ncbi.nlm.nih.gov/protein/121913802?report=genbank&log$=protalign&blast_rank=1&RID=9MZ765BH014) [Gene info linked to EAY18608.1](http://www.ncbi.nlm.nih.gov/gene?term=121913802%5bPUID%5d&RID=9MZ765BH014&log$=genealign&blast_rank=1) 14-3-3 protein [Trichomonas vaginalis G3]

Length=237

[GENE ID: 5464121 TVAG\_462940](http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=search&term=5464121&RID=9MZ765BH014&log$=geneexplicitprot&blast_rank=1) | 14-3-3 protein [Trichomonas vaginalis G3]

(10 or fewer PubMed links)

Score = 230 bits (586), Expect = 5e-73, Method: Compositional matrix adjust.

Identities = 110/149 (74%), Positives = 125/149 (84%), Gaps = 0/149 (0%)

Frame = -3

Query 449 MSSERDVNLFMALNFDQTERHQDMVNAMKKVISLNPELNTDERNLLSVAYKNIVGSRRNG 270

MS+ERDV LFMA +DQT+RHQDMV+ MKKVI L+ +L DERNLLSVAYKNI+GSRRNG

Sbjct 1 MSNERDVALFMAQIYDQTDRHQDMVDTMKKVIELSADLTNDERNLLSVAYKNIIGSRRNG 60

Query 269 LRMIAAIIDHEVGRDNLARVEQLQEYKKVILGELEAYCNELIALVDEKLLPAAKTPESRV 90

LRMIAAI++HE GR N RV L Y+KVILGELE YCNELI LVD LLPAAKT ESRV

Sbjct 61 LRMIAAILEHEDGRGNTQRVASLNAYRKVILGELEKYCNELIGLVDLSLLPAAKTAESRV 120

Query 89 FYEKLKADYYRYICESKPENDKQEPASKA 3

FY+KLKADYYRYICES+ +++K EPA KA

Sbjct 121 FYQKLKADYYRYICESRSDDEKTEPAQKA 149