Edit and Resubmit Save Search Strategies Formatting options Download

PSI blast Iteration 6

ANODAR_208657821 gi|208657821|gb|ACI30201.1|...

Skip to the first new sequence

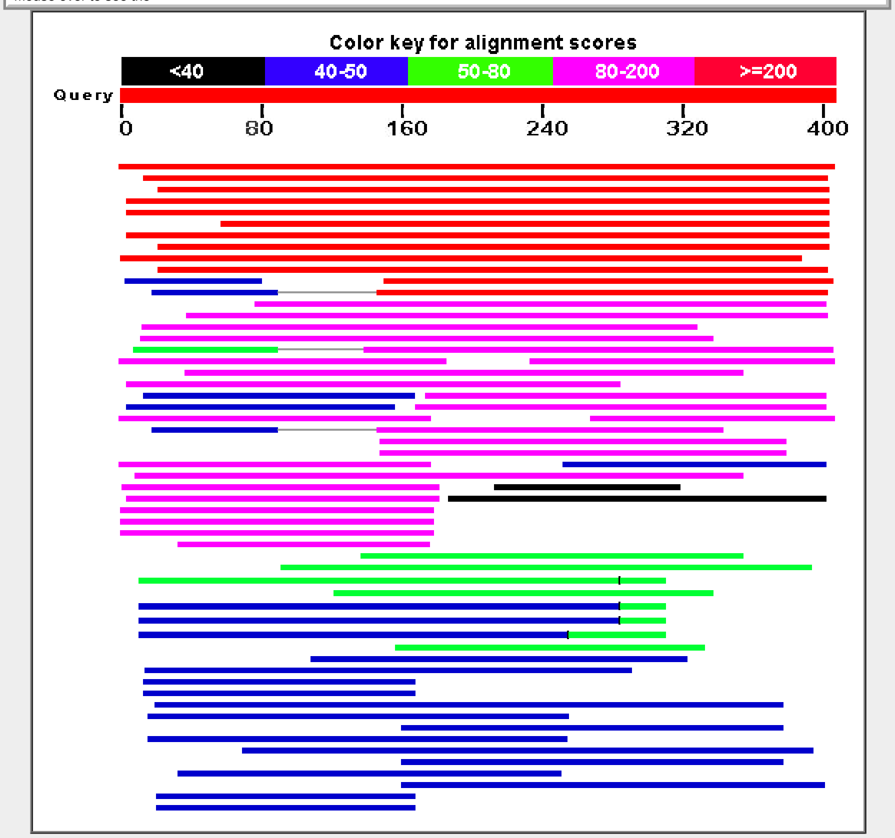
Query ID	lcl|67793	**Database Name**	nr
Description	ANODAR_208657821 gi|208657821|gb|ACI30201.1| 41 kDa family salivary secreted protein [Anopheles darlingi]	**Description**	All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
Molecule type	amino acid	**Program**	BLASTP 2.2.23+ Citation
Query Length	406		

Other reports: Search Summary [Taxonomy reports] [Distance tree of results] [Multiple alignment]

Graphic Summary

Distribution of 107 Blast Hits on the Query Sequence

Mouse over to see the



Descriptions

Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer

NEW - alignment score below the threshold on the previous iteration

● - alignment was checked on the previous iteration

Run PSI-Blast iteration 7 with max 500 Go

Sequences producing significant alignments with E-value BETTER than threshold

Accession	Description	Max score	Total score	Query coverage	E value	Links
● ☐ ACI30201.1	41 kDa family salivary secreted protein [Anopheles darlingi]	352	352	100%	4e-95	
● ☐ AAR18425.1	putative 41.9 kDa basic salivary protein [Culex quinquefasciatus]	298	298	95%	7e-79	
● ☐ ACJ64440.1	putative 41.9 kDa basic salivary protein [Culex tarsalis]	286	286	93%	4e-75	
● ☐ ABF18273.1	putative 41 kDa salivary secreted protein [Aedes aegypti]	281	281	98%	1e-73	
● ☐ XP_001648978.1	hypothetical protein AaeL_AAEL004382 [Aedes aegypti] >gb|EAT4	278	278	98%	9e-73	U G
● ☐ XP_001850512.1	41.9 kDa basic salivary protein [Culex quinquefasciatus] >gb|EDS:	277	277	84%	2e-72	U G

● ☐ ABF18024.1	putative 41 kDa salivary secreted protein [Aedes aegypti]	273	273	98%	3e-71	
● ☐ AAX55661.1	39 kDa salivary protein [Phlebotomus ariasi]	234	234	93%	1e-59	
● ☐ AAS16908.1	37.2 kDa salivary protein [Lutzomyia longipalpis]	228	228	95%	1e-57	
● ☐ ABA43063.1	46 kDa salivary protein SP19 [Phlebotomus perniciosus]	211	211	93%	2e-52	
● ☐ ACU30954.1	putative 41 kDa salivary secreted protein [Ochlerotatus triseriatus	211	211	62%	2e-52	
● ☐ XP_001648977.1	hypothetical protein AaeL_AAEL004353 [Aedes aegypti] >gb|EAT4	206	206	63%	5e-51	U G
● ☐ ACM40881.1	laminin-like secreted salivary protein [Culicoides nubeculosus]	196	196	79%	4e-48	
● ☐ ACS93519.1	46 kDa salivary protein [Phlebotomus arabicus]	196	196	89%	4e-48	
● ☐ AAV90688.1	putative 41 kDa salivary secreted basic protein [Aedes albopictus]	190	190	77%	3e-46	
● ☐ YP_078665.1	phage related protein [Bacillus licheniformis ATCC 14580] >ref|YP_	186	186	80%	6e-45	G
● ☐ XP_001850511.1	salivary mucin [Culex quinquefasciatus] >gb|EDS32122.1| salivary	176	176	65%	6e-42	U G
● ☐ ACI30188.1	putative 41.9 kDa basic salivary protein [Anopheles darlingi]	173	173	42%	4e-41	
● ☐ XP_002388385.1	hypothetical protein MPER_12601 [Moniliophthora perniciosa FA55	167	167	78%	2e-39	G
● ☐ AAX54873.1	putative 30.3 kDa secreted protein [Aedes aegypti]	162	162	68%	6e-38	
● ☐ AAU06485.1	unknown salivary protein [Culicoides sonorensis]	161	161	55%	2e-37	
● ☐ ACH56926.1	hypothetical protein [Simulium vittatum]	156	156	57%	4e-36	
● ☐ ACI30197.1	putative 41.9 kDa basic salivary protein [Anopheles darlingi]	149	149	34%	6e-34	
● ☐ ABF18165.1	putative salivary mucin [Aedes aegypti]	149	149	48%	6e-34	
● ☐ XP_771922.1	hypothetical protein CNBN1020 [Cryptococcus neoformans var. ne	136	136	56%	4e-30	U G
● ☐ XP_568542.1	vacuolar membrane protein [Cryptococcus neoformans var. neofor	136	136	56%	5e-30	G
● ☐ XP_310399.1	AGAP003841-PA [Anopheles gambiae str. PEST] >emb|CAC35525	125	125	45%	9e-27	U G
● ☐ AAR18443.1	putative threonine/serine-rich mucin [Culex quinquefasciatus]	125	125	43%	1e-26	
● ☐ XP_001864883.1	threonine/serine-rich mucin [Culex quinquefasciatus] >gb|EDS408	122	122	43%	7e-26	U G
● ☐ XP_001841204.1	hypothetical protein CC1G_10201 [Coprinopsis cinerea okayama7	121	121	84%	2e-25	G
● ☐ ABI83794.1	gSG10 salivary mucin [Anopheles funestus]	114	114	44%	3e-23	
● ☐ XP_001661479.1	hypothetical protein AaeL_AAEL011183 [Aedes aegypti] >gb|ABF1	106	106	43%	6e-21	U G
● ☐ ACI30119.1	gSG10 salivary mucin [Anopheles darlingi]	102	102	43%	8e-20	
● ☐ ACI30120.1	gSG10 salivary mucin [Anopheles darlingi]	99.0	99.0	43%	8e-19	
● ☐ ACI30122.1	gSG10 salivary mucin [Anopheles darlingi]	97.1	97.1	43%	3e-18	
● ☐ ABJ97695.1	putative threonine/serine-rich mucin [Toxorhynchites amboinensis	80.9	80.9	35%	2e-13	
NEW ☐ XP_001882977.1	predicted protein [Laccaria bicolor S238N-H82] >gb|EDR06605.1|	55.1	55.1	53%	2e-05	G
NEW ☐ XP_759967.1	hypothetical protein UM03820.1 [Ustilago maydis 521] >gb|EAK8	55.1	55.1	74%	2e-05	G
NEW ☐ ZP_06695858.1	phage tail protein [Enterococcus faecium E1636] >gb|EFF22799.1|	54.3	54.3	51%	3e-05	
NEW ☐ NP_690686.2	structural protein [Bacillus phage SPP1] >emb|CAL18687.1| unna	53.2	53.2	52%	5e-05	G
NEW ☐ ZP_00602924.1	Phage tail tape measure protein TP901, core region [Enterococcus	52.0	52.0	51%	1e-04	
NEW ☐ ZP_05713621.1	tail protein [Enterococcus faecium DO]	51.6	51.6	51%	2e-04	
NEW ☐ ZP_05661457.1	phage tail tape measure protein TP901 [Enterococcus faecium 1,2	51.2	51.2	51%	2e-04	
NEW ☐ YP_176319.1	tail length tape measure protein [Bacillus clausii KSM-K16] >dbj|	50.5	50.5	43%	4e-04	G
NEW ☐ ZP_05924009.1	phage tail tape measure protein TP901 [Enterococcus faecium TC	48.6	48.6	52%	0.001	

Run PSI-Blast iteration 7 with max 500 Go

Sequences with E-value WORSE than threshold

Accession	Description	Max score	Total score	Query coverage	E value	Links
● ☐ CAC35527.1	gSG9 protein [Anopheles gambiae]	46.2	46.2	19%	0.008	
☐ ZP_04674169.1	conserved hypothetical protein [Lactobacillus paracasei subsp. par	44.3	44.3	67%	0.031	
☐ EDL20001.1	Unc-51 like kinase 1 (C. elegans), isoform CRA_a [Mus musculus]	43.9	43.9	37%	0.031	G
☐ AAH59835.1	Ulk1 protein [Mus musculus]	43.9	43.9	37%	0.034	G
☐ BAE28476.1	unnamed protein product [Mus musculus]	43.9	43.9	37%	0.036	G
☐ ZP_04248588.1	Prophage LambdaBa01, membrane protein [Bacillus cereus Rock1-	43.9	43.9	87%	0.037	
☐ EFB27064.1	hypothetical protein PANDA_003089 [Ailuropoda melanoleuca]	43.5	43.5	58%	0.043	
☐ ZP_04223830.1	Prophage LambdaBa01, membrane protein [Bacillus cereus Rock3-	43.5	43.5	53%	0.051	
☐ AAA64834.1	ankyrin G [Homo sapiens]	43.5	43.5	58%	0.051	G
☐ ZP_05354696.1	TP901 family phage tail tape measure protein [Clostridium difficile	43.2	43.2	79%	0.059	
☐ YP_002749202.1	TMP repeat-containing protein [Bacillus cereus 03BB102] >gb|AC	43.2	43.2	53%	0.062	G
☐ ZP_04106162.1	Uncharacterized phage protein [Bacillus thuringiensis serovar berli	41.6	41.6	53%	0.20	
☐ ZP_04251409.1	Prophage LambdaBa01, membrane protein [Bacillus cereus 95/82	41.2	41.2	59%	0.21	
☐ BAE32939.1	unnamed protein product [Mus musculus]	41.2	41.2	36%	0.23	G
☐ O70405.1	RecName: Full=Serine/threonine-protein kinase ULK1; AltName: F	41.2	41.2	36%	0.23	G
☐ NP_033495.2	serine/threonine-protein kinase ULK1 [Mus musculus] >gb|AAH57	41.2	41.2	36%	0.24	U G
☐ AAC78143.1	270 kDa ankyrin G isoform [Rattus norvegicus]	40.8	40.8	59%	0.27	G
☐ NP_001029156.1	ankyrin 3, epithelial isoform 2 [Rattus norvegicus]	40.8	40.8	59%	0.27	U G
☐ XP_002619294.1	hypothetical protein CLUG_00453 [Clavispora lusitaniae ATCC 427	40.8	40.8	51%	0.28	G
☐ XP_002807495.1	PREDICTED: LOW QUALITY PROTEIN: ankyrin-3-like [Callithrix jac	40.8	40.8	58%	0.30	

☐ NP_001101811.1	serine/threonine-protein kinase ULK1 [Rattus norvegicus] >gb|ED	40.8	40.8	46%	0.31	U G
☐ ZP_05332071.1	phage protein [Clostridium difficile QCD-63q42] >ref|ZP_0539921	40.8	40.8	63%	0.34	
☐ ZP_04146935.1	Prophage LambdaBa01, membrane protein [Bacillus thuringiensis	40.5	40.5	88%	0.35	
☐ EDL97288.1	ankyrin 3, epithelial, isoform CRA_i [Rattus norvegicus] >gb|EDL9	40.5	40.5	59%	0.36	G
☐ BAC65639.1	mKIAA0722 protein [Mus musculus]	40.5	40.5	36%	0.39	G
☐ CAH73232.1	ankyrin 3, node of Ranvier (ankyrin G) [Homo sapiens] >emb|CA	40.5	40.5	58%	0.44	G
☐ EAW54200.1	ankyrin 3, node of Ranvier (ankyrin G), isoform CRA_b [Homo sap	40.5	40.5	58%	0.44	G
☐ NP_066267.2	ankyrin-3 isoform 1 [Homo sapiens] >sp|Q12955.3|ANK3_HUMAN	40.1	40.1	58%	0.46	U G
☐ BAE33746.1	unnamed protein product [Mus musculus]	40.1	40.1	36%	0.46	G
☐ XP_002612001.1	hypothetical protein BRAFLDRAFT_124780 [Branchiostoma floridae	40.1	40.1	37%	0.46	G
☐ XP_001929357.1	PREDICTED: ankyrin 3, node of Ranvier (ankyrin G) [Sus scrofa]	40.1	40.1	58%	0.50	U G
☐ YP_001987063.1	Tail tape measure protein [Lactobacillus casei BL23] >emb|CAQ66	40.1	40.1	67%	0.50	G
☐ ZP_04019473.1	conserved hypothetical protein [Finegoldia magna ATCC 53516] >	40.1	40.1	84%	0.51	
☐ EFA12920.1	hypothetical protein TcasGA2_TC006990 [Tribolium castaneum]	40.1	40.1	36%	0.53	
☐ NP_203744.1	MICAL-like protein 1 [Homo sapiens] >sp|Q8N3F8.2|MILK1_HUMA	40.1	40.1	37%	0.55	U G
☐ XP_515124.2	PREDICTED: molecule interacting with Rab13 [Pan troglodytes]	39.7	39.7	37%	0.66	U G
☐ XP_001843226.1	conserved hypothetical protein [Culex quinquefasciatus] >gb|EDS	39.3	39.3	73%	0.77	G
☐ XP_002462548.1	hypothetical protein SORBIDRAFT_02g027920 [Sorghum bicolor]	39.3	39.3	75%	0.78	U G
☐ XP_001644303.1	hypothetical protein Kpol_1066p9 [Vanderwaltozyma polyspora DS	39.3	39.3	50%	0.81	G
☐ ZP_05690681.1	phage tape measure protein [Staphylococcus aureus A8115] >ref|	39.3	39.3	45%	0.96	
☐ XP_002257750.1	SICA antigen (fragment) [Plasmodium knowlesi strain H] >emb|C	39.3	39.3	50%	0.99	G
☐ CAQ07542.1	MICAL-like 1 [Homo sapiens]	38.9	38.9	37%	1.0	
☐ BAB33338.1	KIAA1668 protein [Homo sapiens]	38.9	38.9	37%	1.0	G
☐ ZP_04105490.1	TMP repeat-containing protein [Bacillus thuringiensis serovar berli	38.9	38.9	43%	1.3	
☐ CAD38718.1	hypothetical protein [Homo sapiens]	38.5	38.5	37%	1.4	G
☐ ZP_03158061.1	Hemolysin-type calcium-binding region [Cyanothece sp. PCC 7822	38.5	38.5	72%	1.5	
☐ XP_001122565.1	PREDICTED: hypothetical protein [Apis mellifera]	38.5	38.5	25%	1.5	U G
☐ XP_001580855.1	hypothetical protein [Trichomonas vaginalis G3] >gb|EAY19869.1|	38.5	38.5	36%	1.6	G
☐ XP_001648260.1	hypothetical protein AaeL_AAEL014240 [Aedes aegypti] >gb|EAT3	38.5	38.5	66%	1.6	U G
☐ EEZ97902.1	hypothetical protein TcasGA2_TC000287 [Tribolium castaneum]	38.5	38.5	52%	1.7	
☐ ZP_05688243.1	phage tape measure protein [Staphylococcus aureus A9299] >gb|	38.2	38.2	45%	1.8	
☐ XP_002767144.1	centromere protein a, putative [Perkinsus marinus ATCC 50983] >	38.2	38.2	48%	2.1	G
☐ YP_239580.1	ORF001 [Staphylococcus phage 69] >gb|AAX90756.1| ORF001 [S	38.2	38.2	45%	2.1	G
☐ ZP_01630797.1	hypothetical protein N9414_06069 [Nodularia spumigena CCY941	38.2	38.2	49%	2.1	
☐ XP_002420533.1	RNA-binding protein, putative [Candida dubliniensis CD36] >emb|	37.8	37.8	36%	2.3	G
☐ XP_001032003.1	hypothetical protein TTHERM_00703320 [Tetrahymena thermophil	37.8	37.8	46%	2.6	G
☐ ZP_06324921.1	tape measure protein [Staphylococcus aureus subsp. aureus D139	37.8	37.8	45%	2.6	
☐ BAE38328.1	unnamed protein product [Mus musculus] >gb|EDL31272.1| RIKE	37.8	37.8	48%	2.8	G
☐ NP_766596.1	group B scavenger receptor cysteine-rich domain-containing prec	37.8	37.8	48%	2.8	U G
☐ NP_803295.1	tape measure protein [Staphylococcus phage 11] >gb|AAL82270.	37.8	37.8	44%	2.8	G
☐ YP_500528.1	phage tape measure protein [Staphylococcus aureus subsp. aureu	37.8	37.8	44%	2.9	G
☐ YP_627836.1	outer membrane protein HopI [Helicobacter pylori HPAG1] >gb|A	37.4	37.4	39%	3.0	G
☐ EEU07699.1	Sef1p [Saccharomyces cerevisiae JAY291]	37.4	37.4	49%	3.1	
☐ EDN64553.1	suppressor of essential function [Saccharomyces cerevisiae YJM78	37.4	37.4	49%	3.1	
☐ NP_009487.2	Putative transcription factor, has homolog in Kluyveromyces lactis	37.4	37.4	49%	3.2	G
☐ ZP_06327389.1	tape measure protein [Staphylococcus aureus subsp. aureus C427	37.4	37.4	45%	3.3	
☐ NP_846030.1	prophage LambdaBa01, membrane protein, putative [Bacillus anth	37.4	37.4	53%	3.3	G
☐ YP_001004316.1	tail protein [Staphylococcus phage phiETA2] >dbj|BAF43871.1| ta	37.4	37.4	45%	3.7	G
☐ YP_874001.1	phage tape measure protein [Staphylococcus phage phiNM] >ref|	37.4	37.4	45%	3.8	G
☐ ZP_00393935.1	COG5280: Phage-related minor tail protein [Bacillus anthracis str.	37.0	37.0	53%	3.8	
☐ NP_371425.1	hypothetical protein SAV0901 [Staphylococcus aureus subsp. aure	37.0	37.0	45%	4.1	G
☐ YP_001246278.1	Phage-related protein-like protein [Staphylococcus aureus subsp.	37.0	37.0	45%	4.2	G
☐ XP_001818890.1	hypothetical protein [Aspergillus oryzae RIB40] >dbj|BAE56888.1|	37.0	37.0	39%	4.2	G
☐ ZP_06327622.1	tape measure protein [Staphylococcus aureus A9765] >gb|EFB99	37.0	37.0	45%	4.3	
☐ YP_001332062.1	phage tape measure protein [Staphylococcus aureus subsp. aureu	37.0	37.0	45%	4.4	G
☐ YP_001949854.1	putative tape measure protein [Staphylococcus phage phiMR25] >	37.0	37.0	45%	4.4	G
☐ YP_239735.1	ORF001 [Staphylococcus phage 85] >gb|AAX90911.1| ORF001 [S	37.0	37.0	45%	4.7	G
☐ XP_001870945.1	conserved hypothetical protein [Culex quinquefasciatus] >gb|EDS	37.0	37.0	53%	4.7	U G
☐ YP_239657.1	ORF001 [Staphylococcus phage 53] >ref|YP_001285370.1| tape	36.6	36.6	45%	5.0	G
☐ XP_002009072.1	GI13843 [Drosophila mojavensis] >gb|EDW19548.1| GI13843 [Dr	36.6	36.6	41%	5.7	G
☐ YP_002332525.1	putative tape measure protein [Staphylococcus phage phiSauS-IP	36.6	36.6	45%	6.3	G
☐ YP_084997.1	prophage membrane protein [Bacillus cereus E33L] >gb|AAU1685	36.6	36.6	53%	6.4	G
☐ XP_002380327.1	mucin, putative [Aspergillus flavus NRRL3357] >gb|EED49946.1|	36.2	36.2	39%	6.6	G
☐ XP_001026080.1	SAM domain containing protein [Tetrahymena thermophila] >gb|E	36.2	36.2	61%	8.0	G

☐ YP_417175.1	phage-like protein [Staphylococcus aureus RF122] >emb|CAI8140	36.2	36.2	44%	8.1	G
☐ XP_002720066.1	PREDICTED: inter-alpha (globulin) inhibitor H5-like [Oryctolagus c	36.2	36.2	64%	8.4	U G
☐ ZP_04675278.1	conserved hypothetical protein [Lactobacillus jensenii 1153] >gb|E	35.8	35.8	36%	8.7	
☐ ZP_05414794.1	putative fibronectin type III domain protein [Bacteroides finegoldii	35.8	35.8	63%	9.7	
☐ ACA52029.1	surface anchor protein [Lactobacillus jensenii]	35.8	35.8	36%	9.7	

Run PSI-Blast iteration 7 with max 500 Go

Alignments

☐ Select All   Get selected sequences   Distance tree of results   Multiple alignment

```
>☐gb|ACI30201.1|  41 kDa family salivary secreted protein [Anopheles darlingi]
Length=406

 Score =  352 bits (903),  Expect = 4e-95, Method: Composition-based stats.
 Identities = 406/406 (100%), Positives = 406/406 (100%), Gaps = 0/406 (0%)

Query  1    MNRSMSVVLALLIAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNV  60
            MNRSMSVVLALLIAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNV
Sbjct  1    MNRSMSVVLALLIAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNV  60

Query  61   TLALGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSY  120
            TLALGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSY
Sbjct  61   TLALGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSY  120

Query  121  IAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQ  180
            IAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQ
Sbjct  121  IAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQ  180

Query  181  TLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGF  240
            TLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGF
Sbjct  181  TLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGF  240

Query  241  EQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILAA  300
            EQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILAA
Sbjct  241  EQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILAA  300

Query  301  DRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLFLPVSLTI  360
            DRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLFLPVSLTI
Sbjct  301  DRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLFLPVSLTI  360

Query  361  EGGKLYAALSSLQTDVALCASEMGLEIGLVTAQVSSKIAFCQLAQL  406
            EGGKLYAALSSLQTDVALCASEMGLEIGLVTAQVSSKIAFCQLAQL
Sbjct  361  EGGKLYAALSSLQTDVALCASEMGLEIGLVTAQVSSKIAFCQLAQL  406
```

```
>☐gb|AAR18425.1|  putative 41.9 kDa basic salivary protein [Culex quinquefasciatus]
Length=412

 Score =  298 bits (762),  Expect = 7e-79, Method: Composition-based stats.
 Identities = 85/399 (21%), Positives = 164/399 (41%), Gaps = 33/399 (8%)

Query  15   AIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGN---RINTD  71
             +G      +    T   +    +            ++     + + +  G    ++N
Sbjct  36   QVGATDTGTNITGLTITLSRWSGKVSLAGSGNQIEVSSFGNDSTMIRVGYGQHLTKLNEP  95

Query  72   TALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAA--------NNPVNGSYIAI  123
            ++  +    +      +  RR+R      I    V   + +           +N +YI+I
Sbjct  96   SSSVELPAPIEPTNNTIPSRRKRETGFLPILPVTVTPFVPSVPAAPAIPIPTMNSNYISI  155

Query  124  GNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLP  183
            G ++  LP +++ ++ N     S G G  V SN             Q +Q L T  +  P
Sbjct  156  GGKLVGLPESITKLTANT----SIGTGSAVQSN----------TMQQPWQFLKTLMERRP  201

Query  184  LDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQV  243
               +D +V  L+ +Y +T K +   A L  LN+T     ++    +    DQ + G + V
Sbjct  202  SKILDPVVDFLATSYLSTGKVMVDGA-LTALNKTEAFFYNIGNSSLNYATDQVSLGLQGV  260

Query  244  IASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILAADRG  303
              SFND +   + CVG   +P + GRRV+ K T CV  +  ++  I  +   ++ A +
Sbjct  261  TRSFNDLTQMKRTCVG--GAPENAGRRVMEKATGCVRERLDEVIGIVDQFRTNVEATEGV  318

Query  304  ASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLFLPVSLTIEGG  363
             S +L  L  CN  NF     +       +A +R+CY  A+ +    ++ +P+  T
Sbjct  319  FSGWLTELQKCNAKNFPTLADAEL-----DAAQRKCYAEALVSPTGKVVDIPLRWTTLAS  373

Query  364  KLYAALSSLQTDVALCASEMGLEIGLVTAQVSSKIAFCQ  402
            ++  A+++ +  + LCA ++ LE+  +  ++ SKI  C
Sbjct  374  RISTAVTTFKAQLGLCAGKVILEVAAIANELGSKITSCA  412
```

```
>☐gb|ACJ64440.1|  putative 41.9 kDa basic salivary protein [Culex tarsalis]
Length=414

 Score =  286 bits (730),  Expect = 4e-75, Method: Composition-based stats.
 Identities = 91/392 (23%), Positives = 160/392 (40%), Gaps = 33/392 (8%)

Query  23   VDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGN---RINTDTALTDYGT  79
                  T        +              T    + + + L     ++   +   +
Sbjct  42   TRINGLTIVLRRWFGKVTVSGNGNQIDVLGTGNDSTTIRIGLKQNPAKLKAASYSLELPA  101

Query  80   NVNGLGINVNRRRRRRRDIWNIESNPVLAGLAA--------NNPVNGSYIAIGNQVFHLP  131
            +   L   +  RR+R      I  NP    + +           +NG+YI+I  ++  LP
Sbjct  102  SREILNNTIPSRRKRETGFLPILPNPPTPFVPSVPPTPAIPIPTMNGNYISIDGKLVRLP  161

Query  132  NNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIV  191
            +++SN ++N A+GSS  A                 +  Q +Q L T  +  P   +D +V
Sbjct  162  DSISNSTINAATGSSPLA--------------TAGSMQQPWQFLQTLIERRPSKILDPLV  207
```

```
Query  192  QTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSS  251
            + L++ Y +T K +    GL  L  T  +  ++    +    DQ + G + V  SFN  +
Sbjct  208  EFLANNYLSTGKMMVD-KGLAALGATEAYFYNIANTSLNYATDQVSLGLQGVTRSFNGLT  266

Query  252  ARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILAADRGASDFLGNL  311
               + CVG    P   GRRV+ + T CV  +  ++  I  +   ++ A +   S +L  +
Sbjct  267  LMKRTCVG--GVPEDAGRRVMERSTGCVRERLDEVIGIVDQFRSNVEATEGVFSGWLIEM  324

Query  312  TTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLFLPVSLTIEGGKLYAALSS  371
              CN  NF     S       +  +RQCY +A  +    L+ +P   T    + Y A+++
Sbjct  325  QKCNAKNFPNLADSEL-----DVAQRQCYSQASVSPTGKLIDIPQRWTSLALRTYTAVNT  379

Query  372  LQTDVALCASEMGLEIGLVTAQVSSKIAFCQL  403
             Q  V LC +++GLEI  ++A+  SKI  C L
Sbjct  380  FQAQVGLCVAKVGLEIAAISAEFGSKITSCAL  411


>☐gb|ABF18273.1|  putative 41 kDa salivary secreted protein [Aedes aegypti]
Length=396

 Score =  281 bits (717),  Expect = 1e-73, Method: Composition-based stats.
 Identities = 102/420 (24%), Positives = 173/420 (41%), Gaps = 57/420 (13%)

Query  5    MSVVLALLIAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLAL  64
            +SV+   LI ++   PW+        PW  +                      S VT+++
Sbjct  11   LSVMGLFLIPSLEASPWF-------WPWQSRA-------------------SPSIVTISI  44

Query  65   GNRINTDTALTDYGTNVNGLGINVNRRR---------------RRRRDIWNIESNPVLAG  109
            GNR N+++ +TD GT++  L  N++R +               R    + NI  +
Sbjct  45   GNRFNSNS-VTDEGTSIRDLTFNLSRWKGNLTFEKNGTEILITRGNDGVTNIHIDFGSNS  103

Query  110  LAANN---PVNGSYIAIGNQVFHLPNNLSNVSLNIASGSSGGAG-QTVPSNFSTFLP-MI  164
             +  N   P N +      ++F LP+ L  + +   +G+    G            P
Sbjct  104  KSQQNETVPSNATQSRKKREIFGLPS-LPAIPIPTINGTQAAIGENFFQLPNEIANPASN  162

Query  165  PDNYTQAFQKLVTQF-QTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRS  223
              N+ Q +Q L T   + LP+  +D+ V  LS  YS+T K +    G+N L ET++   S
Sbjct  163  ASNFLQPWQFLQTLITRRLPMQFLDTAVSYLSQLYSSTAKNMLDT-GVNTLAETMDTLSS  221

Query  224  VIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKW  283
            ++   +  I DQ A G   V  SF +     Q CVG    P    + VV K T CV  +W
Sbjct  222  IVNSTLDYIEDQTAAGIRGVAISFRELPQARQSCVGQ--VPDDAAQPVVTKATGCVRDRW  279

Query  284  RQLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRA  343
             +L  I  +    +  ++     +L  L  CN  NF   D +       +A +R CYV+
Sbjct  280  NELDGIVNQFLVAVSDSEEAYGGWLRALDGCNARNFAGVDEAEL-----DAAQRSCYVQN  334

Query  344  IGNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVALCASEMGLEIGLVTAQVSSKIAFCQL  403
            I    + ++ +P+       +  + +   Q  V LC + +G ++  V+A +  +I  CQ
Sbjct  335  ISKSVEKMVDIPIQWANLAARASSIVMQFQPQVGLCVAGVGAKVAAVSANIGVQIMLCQA  394


>☐ref|XP_001648978.1| UG hypothetical protein AaeL_AAEL004382 [Aedes aegypti]
 gb|EAT44234.1| G conserved hypothetical protein [Aedes aegypti]
Length=396

 GENE ID: 5564658 AaeL_AAEL004382 | hypothetical protein [Aedes aegypti]
(10 or fewer PubMed links)

 Score =  278 bits (710),  Expect = 9e-73, Method: Composition-based stats.
 Identities = 104/420 (24%), Positives = 175/420 (41%), Gaps = 57/420 (13%)

Query  5    MSVVLALLIAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLAL  64
            +SV+   LI ++   PW+        PW  +                      S VT+++
Sbjct  11   LSVMGLFLIPSLEASPWF-------WPWQSRA-------------------SPSIVTISI  44

Query  65   GNRINTDTALTDYGTNVNGLGINVNRRR---------------RRRRDIWNIESNPVLAG  109
            GNR+N++T +TD GT++  L  N++R +               R    + NI  +
Sbjct  45   GNRVNSNT-VTDEGTSIRDLTFNLSRWKGNLTFEKNGTEILITRGNDGVTNIHIDFGSNS  103

Query  110  LAANN---PVNGSYIAIGNQVFHLPNNLSNVSLNIASGSSGGAG-QTVPSNFSTFLP-MI  164
             +  N   P N +      ++F LP+ L  + +   +G+    G    P       P
Sbjct  104  KSQQNETVPSNVTQSRKKREIFGLPS-LPAIPIPTLNGTQAAIGENFFPLPNEIANPASN  162

Query  165  PDNYTQAFQKLVTQF-QTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRS  223
              N+ Q +Q L T   + LP+  +D+ V  LS  YS+T K +    G+N L ET++   S
Sbjct  163  ASNFLQPWQFLQTLITRRLPMQFLDTAVSYLSQLYSSTAKSMLDT-GVNTLAETMDTLSS  221

Query  224  VIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKW  283
            ++   +  I DQ A G   V  SF +     Q CVG    P    + VV K T CV  +W
Sbjct  222  IVNSTLDYIEDQTAAGIRGVAISFRELPQARQSCVGQ--VPDDAAQPVVTKATGCVRDRW  279

Query  284  RQLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRA  343
             +L  I  +    +  ++     +L  L  CN  NF   D +       +A +R CYV+
Sbjct  280  NELDGIVNQFLVAVSDSEEAYGGWLRALDGCNARNFAGVDEAEL-----DAAQRSCYVQN  334

Query  344  IGNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVALCASEMGLEIGLVTAQVSSKIAFCQL  403
            I    + ++ +P+       +  + +   Q  V LC + +G ++  V+A +  +I  CQ
Sbjct  335  ISKSVEKMVDIPIQWANLAARASSIVMQFQPQVGLCVAGVGAKVAAVSANIGVQIMLCQA  394


>☐ref|XP_001850512.1| UG 41.9 kDa basic salivary protein [Culex quinquefasciatus]
 gb|EDS32123.1| G 41.9 kDa basic salivary protein [Culex quinquefasciatus]
Length=338

 GENE ID: 6041220 CpipJ_CPIJ008900 | 41.9 kDa basic salivary protein
[Culex quinquefasciatus]

 Score =  277 bits (707),  Expect = 2e-72, Method: Composition-based stats.
 Identities = 82/356 (23%), Positives = 155/356 (43%), Gaps = 33/356 (9%)

Query  59   NVTLALGN---RINTDTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAA---  112
             + +  G    ++N  ++  +    +      +  RR+R      I        + +
Sbjct  1    MIRVGYGQHLTKLNEPSSSVELPAPIEPTNNTIPSRRKRETGFLPILPVTATPFVPSVPA  60

Query  113  -----NNPVNGSYIAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDN  167
                    +N +YI+IG ++  LP +++ ++ N     S G G  V SN
Sbjct  61   APAIPIPTMNSNYISIGGKLVGLPESITKLTANT----SIGTGSAVQSN---------T  106
```

```
Query  168  YTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIED  227
              Q +Q L T  +  P   +D +V  L+ +Y +T K +   A L  LN+T     ++
Sbjct  107  MQQPWQFLKTLMERRPSKILDPVVDFLATSYLSTGKVMVDGA-LTALNKTEAFFYNIGNS  165

Query  228  GIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLS  287
             +    DQ + G + V  SFND +   + CVG   +P + GRRV+ K T CV  +  ++
Sbjct  166  SLNYATDQVSLGLQGVTRSFNDLTQMKRTCVG--GAPENAGRRVMEKATGCVRERLDEVI  223

Query  288  DIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNF  347
             I  +   ++ A +   S +L  +  CN  NF     +       +A +R+CY  A+ +
Sbjct  224  GIVDQFRTNVEATEGVFSGWLIEMQKCNAKNFPTLADTEL-----DAAQRKCYAEALVSP  278

Query  348  PQSLLFLPVSLTIEGGKLYAALSSLQTDVALCASEMGLEIGLVTAQVSSKIAFCQL  403
               ++ +P+  T    ++  A+++ +  + LCA ++ LE+  +  ++ SKI  C L
Sbjct  279  TGKVVDIPLRWTTLASRISTAVTTFKAQLGLCAGKVILEVAAIANELGSKITSCAL  334


>☐gb|ABF18024.1|  putative 41 kDa salivary secreted protein [Aedes aegypti]
Length=395

 Score =  273 bits (697),  Expect = 3e-71, Method: Composition-based stats.
 Identities = 103/420 (24%), Positives = 173/420 (41%), Gaps = 58/420 (13%)

Query  5    MSVVLALLIAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLAL  64
            +SV+   LI ++   PW+        PW  +                      S VT+++
Sbjct  11   LSVMGLFLIPSLEASPWF-------WPWQSRA-------------------SPSIVTISI  44

Query  65   GNRINTDTALTDYGTNVNGLGINVNRRR---------------RRRRDIWNIESNPVLAG  109
            GNR+N++T +TD GT++  L  N++R +               R    + NI  +
Sbjct  45   GNRVNSNT-VTDEGTSIRDLTFNLSRWKGNLTFEKNGTEILITRGNDGVTNIHIDFGSNS  103

Query  110  LAANN---PVNGSYIAIGNQVFHLPNNLSNVSLNIASGSSGGAG-QTVPSNFSTFLP-MI  164
             +  N   P N +      ++F LP+ L  + +   +G+    G    P       P
Sbjct  104  KSQQNETVPSNVTQSRKKREIFGLPS-LPAIPIPTLNGTQAAIGENFFPLPNEIANPASN  162

Query  165  PDNYTQAFQKLVTQF-QTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRS  223
              N+ Q +Q L T   + LP+  +D+ V  LS  YS+T K +    G+N L ET++   S
Sbjct  163  ASNFLQPWQFLQTLITRRLPMQFLDTAVSYLSQLYSSTAKSMLDT-GVNTLAETMDTLSS  221

Query  224  VIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKW  283
            ++   +  I DQ A G   V  SF +     Q CVG    P    + VV K T CV  +W
Sbjct  222  IVNSTLDYIEDQTAAGIRGVAISFRELPQARQSCVGQ--VPDDAAQPVVTKATGCVRDRW  279

Query  284  RQLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRA  343
             +L  I  +    +  ++     +L  L  CN  NF   D +       +A +R CYV
Sbjct  280  NELDGIVNQFLVAVSDSEEAYGGWLRALDGCNARNFAGVDEAEL-----DAAQRSCYVNI  334

Query  344  IGNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVALCASEMGLEIGLVTAQVSSKIAFCQL  403
              +    ++ +P+       +  + +   Q  V LC + +G ++  V+A +  +I  CQ
Sbjct  335  SKSVE-KMVDIPIQWANLAARASSIVMQFQPQVGLCVAGVGAKVAAVSANIGVQIMLCQA  393


>☐gb|AAX55661.1|  39 kDa salivary protein [Phlebotomus ariasi]
Length=375

 Score =  234 bits (596),  Expect = 1e-59, Method: Composition-based stats.
 Identities = 43/388 (11%), Positives = 106/388 (27%), Gaps = 43/388 (11%)

Query  23   VDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGTNVN  82
                L   P + +    + +       S   P    N  ++        + +     +
Sbjct  12   ALLTLLPNPEDVRSAADVLESFTDDLKSFYPPPDDVNEEVSETESRTKRSLIEQLKESQP  71

Query  83   GLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIGN-------QVFHLPNNLS  135
               I        +     +++ P      ++N  +    +           V  L + +S
Sbjct  72   LKQIRETVAETTKYLKGFLKTKPSGNQTESSNSTSTKTQSRKRRGLTDFIPVNSLKDAIS  131

Query  136  NVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQTLS  195
              +          +     SN   FL  + D    +   +    + +  + + S+
Sbjct  132  QATSGAMKAFKPSSENKTSSNPLDFLASLSD---ISRDLVQNSIKEVSGNLVSSVALYQV  188

Query  196  DAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQ  255
            ++  +  K          ++ T +  + V+    Q+           +  S  +
Sbjct  189  NSKLDAIKQSIGI-INQEIDRTKKVQQYVMNALQQASN---------ITNSIGEQLKS-N  237

Query  256  QCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTCN  315
             C    ++P  +   V+     CV  K      I      ++  A    SD +  ++ C+
Sbjct  238  NCFAQFINPFKLFEEVIT----CVKNKIENGLKIAEETFKNLNQALSVPSDIVSEVSKCS  293

Query  316  GANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLFLPVSLTIEGGKLYAALSSLQTD  375
                                   CY+R      +  L LP+       ++    ++++ D
Sbjct  294  QNQNLNPLTKLL-----------CYLRVPLQLDEEKLLLPIEFARRIREITNYFATMRMD  342

Query  376  VALCASEMGLEIGLVTAQVSSKIAFCQL  403
            +  C              +  K+  C +
Sbjct  343  LIQCGIATIQ-------SIGDKVENCAI  363


>☐gb|AAS16908.1|  37.2 kDa salivary protein [Lutzomyia longipalpis]
Length=350

 Score =  228 bits (579),  Expect = 1e-57, Method: Composition-based stats.
 Identities = 57/390 (14%), Positives = 111/390 (28%), Gaps = 45/390 (11%)

Query  2    NRSMSVVLALLIAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVT  61
               + ++ A L+  I           F    +                   T
Sbjct  1    MTFLIILGAFLLVQIITASALGLPEQFKGLEDLPKKPLAETYYHEGLNDGKTDEMVDIFK  60

Query  62   LALGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYI  121
             +L +         D G   N    ++ +    R  + N++  P +  L +   +
Sbjct  61   -SLSDEFKFSDENLDVGEEKNYKKRDITQNSVARNFLSNVKGIPSMPSLPSMPSMPSIPS  119

Query  122  AIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNF--STFLPMIPDNYTQAFQKLVTQF  179
               +Q    P        N A          +        FL    D Y      L
Sbjct  120  LWSSQTQAAP--------NTALALPESDYSLLDMPNIVKNFLKETRDLYNDVGAFLKAIT  171

Query  180  QTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLE--HGRSVIEDGIQSIVDQGA  237
            + L   +  S + +     +N  K           NE  +    R+ +++ +Q I D  A
Sbjct  172  EALTNRSSSSQLLSSPMVSTNKTKEFIR-------NEIQKVRKVRNFVQETLQKIRDISA  224

Query  238  QGFEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDI  297
               ++V +S         +C+          + +V  G +C+  K+     I  ++  ++
Sbjct  225  AIAKKVKSS---------ECLSNLTD----IKGLVSDGINCLKEKFNDGKRIILQLYNNL  271
```

```
Query  298  LAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLFLPVS  357
            L   +  +D +  L  C+    N               R  CY        +  + LPV
Sbjct  272  LKGLKIPNDLMVELKKCDTNQNNTL------------GRIICYFLTPLQLEKEQILLPVE  319

Query  358  LTIEGGKLYAALSSLQTDVALCASEMGLEI  387
                  +L    S+++ D+  C       I
Sbjct  320  FIKRILELTHYFSTMKEDLINCGITTIASI  349


>gb|ABA43063.1|  46 kDa salivary protein SP19 [Phlebotomus perniciosus]
Length=431

 Score =  211 bits (535),  Expect = 2e-52, Method: Composition-based stats.
 Identities = 43/393 (10%), Positives = 108/393 (27%), Gaps = 58/393 (14%)

Query  23   VDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQ---------QSNVTLALGNRINTDTA  73
                    P + +    I +    S     TP           +S    +  +     +
Sbjct  71   ALLTRVPNPDDIKSAADILESFTGSLKYFQTPPDDVDQEESETKSRSKRSFTDIFKQSSP  130

Query  74   LTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIGNQVFHLPNN  133
            L + G  +  +   +    + +         P      ++N  + +       +
Sbjct  131  LKEIGERIEEIKKKLKGMLKPK------PQTPSGNQTDSSNTTSETQSRKKRALTDF---  181

Query  134  LSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYT----QAFQKLVTQFQTLPLDAMDS  189
               + ++    +    G+ +  + ++      D  +     A   +    + +  +   S
Sbjct  182  ---IPMDSLKDAISKTGEVLIPSSASANSSPLDFMSKLSDIANDLIQNSMKEISENLASS  238

Query  190  IVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFND  249
            +     ++  +  K          +++T +  + V     +  ++Q     + +
Sbjct  239  VAMYQVNSQLDAIKQSMDI-IKQEIDKTQKIQKYV-----KEALNQAKNATKSLGEKLKS  292

Query  250  SSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILAADRGASDFLG  309
            S      C    ++P  +      KG  CV  K      I      ++  A    SD
Sbjct  293  S-----NCFAQFINPFKLFE----KGITCVKNKIDNGLKIAKDTFKNLQQAMSVPSDIQS  343

Query  310  NLTTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLFLPVSLTIEGGKLYAAL  369
             ++ C+                    +  CY+R         L LP   T    ++
Sbjct  344  EVSKCSQNQQLNP-----------IAKLLCYLRTPLQLDDEKLLLPFEFTRRIREITNYF  392

Query  370  SSLQTDVALCASEMGLEIGLVTAQVSSKIAFCQ  402
            ++++ D+  C  E           +  K+  C
Sbjct  393  ATMRMDLIRCGIETIQ-------SIGDKVEDCA  418


>gb|ACU30954.1|  putative 41 kDa salivary secreted protein [Ochlerotatus triseriatus]
Length=253

 Score =  211 bits (535),  Expect = 2e-52, Method: Composition-based stats.
 Identities = 70/259 (27%), Positives = 114/259 (44%), Gaps = 12/259 (4%)

Query  151  QTVPSNFSTFLPMIPDNYT---QAFQKLVTQFQT-LPLDAMDSIVQTLSDAYSNTFKGLF  206
                +  S   P   DN     Q +Q L T     LP+  +D++V  LS  YS T K L
Sbjct  3    PLPSAISSLASPTAADNIRSLLQPWQFLQTILNRRLPMLVLDTVVSFLSQLYSATGKRLV  62

Query  207  TAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTS  266
               G+N + E  +    ++   +    +Q + G + V+ +F D +   Q CVG    P
Sbjct  63   DT-GVNAVGEITDFVSGIVNSALDLATEQASSGMQSVVGAFKDLTQTRQTCVGQ--VPDD  119

Query  267  IGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASS  326
              R+V+ K T C   +W ++  I  +    +   +     +L  L  CN  NF     S
Sbjct  120  AARQVISKATGCARERWNEVEGIANQFLDVVTDTEEAYGGWLRALDGCNARNFAGTSESR  179

Query  327  GGTGTRNALRRQCYVRAIGNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVALCASEMGLE  386
              T      +R+CYV+AIGN    ++ +P+       +   A+++ Q  V LC + +G E
Sbjct  180  KDT-----AQRECYVQAIGNSVGKIIDIPMCWANLTLRTSNAVANFQPQVGLCVAGVGAE  234

Query  387  IGLVTAQVSSKIAFCQLAQ  405
            I  V+A +  +I  CQ+ Q
Sbjct  235  IASVSANIGLRIVLCQVFQ  253


>ref|XP_001648977.1|  hypothetical protein AaeL_AAEL004353 [Aedes aegypti]
 gb|EAT44233.1|  hypothetical protein AaeL_AAEL004353 [Aedes aegypti]
Length=308

 GENE ID: 5564644 AaeL_AAEL004353 | hypothetical protein [Aedes aegypti]
(10 or fewer PubMed links)
                                                       Sort alignments for this subject sequence by:
                                                         E value  Score  Percent identity
                                                         Query start position  Subject start position

 Score =  206 bits (522),  Expect = 5e-51, Method: Composition-based stats.
 Identities = 35/259 (13%), Positives = 82/259 (31%), Gaps = 11/259 (4%)

Query  147  GGAGQTVPSNFSTFLPMIPDNYTQAFQKL--VTQFQTLPLDAMDSIVQTLSDAYSNTFKG  204
               G    +N S     +  N      +            + ++ ++ T+ + Y N  +
Sbjct  55   ISIGNRENTNASIDYGTMITNLQINLARFNGSALLGDATENRLNEVIYTVLEGYRNRSEA  114

Query  205  LFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSP  264
            +  +         L+  +   +           Q  E V  S    +  V+ C+  ++
Sbjct  115  MVDSRLAWIDEVLLDIVQYARQLRQDENNRWLKQFPEHVRTSVAQLNETVRSCLEREVVV  174

Query  265  TSIGRRVVGKG-TDCVNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAAD  323
              +   V  +  + C+  + ++L ++      ++      A D    L TC   +F+ A
Sbjct  175  EELIHAVENRSRSGCLEGRLQKLFELREVAKNNLTEFLASAGDLEDRLETCVNPDFDDAL  234

Query  324  ASSGGTGTRNALRRQCYVRAIGNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVALCASEM  383
                     N  +  C    +       L L  +++   G +  AL   +  +  CA+++
Sbjct  235  D--------NVYQEACISSILLEVEMESLKLGYTVSNLTGAMEPALGLAKAVLLECAADL  286

Query  384  GLEIGLVTAQVSSKIAFCQ  402
                   +  +   I  C
Sbjct  287  AWYAFDASMSLRQWINSCS  305


 Score = 48.2 bits (112),  Expect = 0.002, Method: Composition-based stats.
 Identities = 19/71 (26%), Positives = 35/71 (49%), Gaps = 3/71 (4%)

Query  20  PWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGT  79
```

```
           P   +AF ++  W      T       + ++ ++    S   +++GNR NT+ ++ DYGT
Sbjct  15  PTGTNAFGWS--WFFPRQTTTTTASSGTTSATSSNAVPSMYNISIGNRENTNASI-DYGT  71

Query  80  NVNGLGINVNR  90
            +  L IN+ R
Sbjct  72  MITNLQINLAR  82


>☐gb|ACM40881.1|  laminin-like secreted salivary protein [Culicoides nubeculosus]
Length=521

 Score =  196 bits (497),  Expect = 4e-48, Method: Composition-based stats.
 Identities = 67/367 (18%), Positives = 106/367 (28%), Gaps = 52/367 (14%)

Query  78   GTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAI-----GNQVFHLP-  131
            G   + L  +       +  +  + S P L  L      N +  ++        +  LP
Sbjct  157  GLPQSNLNPSSLFNSLPKMSLPGLPSLPGLPSLPGLPQQNSNPSSLFSSLPKMSLPGLPK  216

Query  132  -----NNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPLDA  186
                 +NL  + L    G           + S+    +P+               L L
Sbjct  217  TSNPFDNLPKMGLPSLPGIP-----QPNLDLSSMFKNLPNLPQIPNLFDPNTLSKLALPG  271

Query  187  MDSIVQTLSDAYSNTFKGLFTAAG-------------------------------LNGLN  215
            +  +   +S   SN FK L                                     +  N
Sbjct  272  LPDLNSLMSVLNSNKFKQLLGIPDVSSLSNSLSNGTVQSALDNLQSKTTEAAKRANDSAN  331

Query  216  ETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIG-RRVVGK  274
              L   +SV       I         QV  +    ++ +Q CV  + +P S G +
Sbjct  332  SALSGLQSVGNTISNEIDKVSKSAQTQVNNTMKGLTSDIQDCVNRNGNPGSAGVKAAQTS  391

Query  275  GTDCVNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNA  334
               CV  K  +   I      DI  A  G     GNL+ C          S     T +A
Sbjct  392  AKKCVQDKVDEAKSIIDGTRNDIKQAQLGVEYLRGNLSNCQLN----VSMSLAELPTISA  447

Query  335  LRRQCYVRAIGNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVALCASEMGLEIGLVTAQV  394
             +  C   A+ +     L +P+SL   G +    +  LQ     C + +  EI   T
Sbjct  448  SKPACLASALLSVQSETLLIPLSLASHGAQAVGLIQGLQGYAIKCMANLMEEIAKSTLST  507

Query  395  SSKIAFC  401
            S   A C
Sbjct  508  SLAAAKC  514


>☐gb|ACS93519.1|  46 kDa salivary protein [Phlebotomus arabicus]
Length=442

 Score =  196 bits (497),  Expect = 4e-48, Method: Composition-based stats.
 Identities = 39/373 (10%), Positives = 95/373 (25%), Gaps = 48/373 (12%)

Query  39   TINQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDI  98
              N       +       +S    +  +++     L      V+ +   +    + +
Sbjct  96   LTNSYVTFDDANEEASVTKSRSKRSFLDQLKESQPLKAIRERVDEVKKQLQAYLKPK---  152

Query  99   WNIESNPVLAGLAANNPVNGSYIAIGN------QVFHLPNNLSNVSLNI--ASGSSGGAG  150
                  P       N     +             V    + LS  S            +
Sbjct  153  ---PQTPSGNQTDTNTTTTETQSRRKRGLSDYIPVNSFKDALSQASAAAGKMMKPPNASL  209

Query  151  QTVPSNFSTFLPMIPDNYT-QAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAA  209
                S   +         +      + +  +   S      ++  +  K
Sbjct  210  SANASASGSNPLDFLTKISDIGRDLVQNSIKEIASNMASSAALYQVNSKLDAIKQSVNI-  268

Query  210  GLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIGR  269
                + +T +  + +I   +Q      +   ++V +S          C    ++P  +
Sbjct  269  IQQEIEKTKK-VQEIINQALQQAGAATSSFGDKVTSS---------NCFAQFINPFKLFE  318

Query  270  RVVGKGTDCVNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSGGT  329
                 G  CV  K+ +   I      +I  A    SD    ++ C+
Sbjct  319  G----GITCVKNKFDKGIKIATDTLNNISQALSVPSDIKTEVSKCSDNKELNPFTKLL--  372

Query  330  GTRNALRRQCYVRAIGNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVALCASEMGLEIGL  389
                     C+++      +  L LP+       ++    ++++ D+  C
Sbjct  373  ---------CFLKTPLQLDEEKLLLPIEFVRRIREITEYFATMRMDIIQCGIATI-----  418

Query  390  VTAQVSSKIAFCQ  402
                +  K+A C
Sbjct  419  --HSIEDKVADCA  429


>☐gb|AAV90688.1|  putative 41 kDa salivary secreted basic protein [Aedes albopictus]
Length=332

 Score =  190 bits (481),  Expect = 3e-46, Method: Composition-based stats.
 Identities = 80/343 (23%), Positives = 134/343 (39%), Gaps = 59/343 (17%)

Query  14   AAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTA  73
             ++   PW+        PW  +                      S + +++GNR+NT+T+
Sbjct  20   PSLEASPWF-------WPWQSRS-------------------APSIINISIGNRVNTNTS  53

Query  74   LTDYGTNVNGLGINVNRRR---------------RRRRDIWNIESNPVLAG---------  109
            + + GT++  L  N++R +               R    + NI  +
Sbjct  54   VDE-GTSIRDLTFNLSRWKGNLTFEKNGTEIFITRGNGGVTNIHIDFSSNSGNQRSSANR  112

Query  110  -LAANNPVNGSYIAIGNQVFHLPNNLSNVSLNIASGSSGG-AGQTVPSNFSTFLPMI-PD  166
                +N    S       +F LP+ L ++ +   +G+     G ++P   +  +P I
Sbjct  113  TTPLSNATQPSTSRRKRAIFGLPS-LPSIPIPTLNGTQAFLGGNSIPLPTAITIPAINAS  171

Query  167  NYTQAFQKLVTQF-QTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVI  225
               Q +Q L T   + LP+  +D+ V  L+  YS+T K +    G+N L ET++   SV+
Sbjct  172  TLLQPWQFLQTLITRRLPMQLLDTAVSFLAQLYSSTAKSMLDT-GVNTLGETMDTLSSVV  230

Query  226  EDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQ  285
               + +I DQ A G   V  SF   +   Q CVG    P    R VV K T CV  +W +
Sbjct  231  NSTLDTIEDQVASGIAGVAQSFKQLTRAGQSCVGQ--VPDDSARPVVTKATGCVRDRWNE  288

Query  286  LSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSGG  328
            L  I  +    +   +     +L  L  CN  NF     +  G
Sbjct  289  LDGIVNQFLVAVSDTEDAYGGWLRALDACNARNFAGLSDTGAG  331


>☐ref|YP_078665.1| G phage related protein [Bacillus licheniformis ATCC 14580]
 ref|YP_091074.1| G hypothetical protein BLi01480 [Bacillus licheniformis ATCC 14580]
```

```
 gb|AAU23027.1|  phage related protein [Bacillus licheniformis ATCC 14580]
 gb|AAU40381.1|  putative protein [Bacillus licheniformis ATCC 14580]
Length=1020

 GENE ID: 3030548 BL00969 | phage related protein
[Bacillus licheniformis ATCC 14580] (10 or fewer PubMed links)

 Score =  186 bits (470),  Expect = 6e-45, Method: Composition-based stats.
 Identities = 37/336 (11%), Positives = 83/336 (24%), Gaps = 34/336 (10%)

Query  13   IAAIGIIPWYVDAFLFTTPWN--------PQPPRTINQLPVTSATSATTPTQQSNVTLAL  64
            I  +    W      F+  W+                  P+     A      + +   L
Sbjct  556  IVELASTAWGSLTSFFSRLWSGITTTAQAAWTGFMNLVKPIWDGIVAVFGPTFNVIVTTL  615

Query  65   GNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIG  124
             N  NT ++      N     I           +  ++++  +     +   NG
Sbjct  616  SNIWNTVSSTVSSVWN----TIKTTLIGVVTSIVDGVKNHFSILSQTLSGIWNGITSIAK  671

Query  125  NQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPL  184
                 L N +             G    V            ++  Q +  +    + +
Sbjct  672  GAWQVLKNAI------------LGPVLLVIDLVQGDFKGFANHLKQIWTNISNGAKQIWN  719

Query  185  DAMDSIVQTLSDAYSNTFKGLFTAA---GLNGLNETLEHGRSVIEDGIQSIVDQGAQGFE  241
              + ++V +L     N  K  +  A        N       S+      ++V+     +
Sbjct  720  G-IKTVVSSLVKGLVNAVKNYWNTAKTVTTTIFNGIKSVLSSIWNGIKNTVVNLAKGLWN  778

Query  242  QVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILAAD  301
             V  ++N              +  ++   V       V        +I   +  +
Sbjct  779  AVKTTWNTFKTVTTTIFN---AVKTVLTTVWNAAKSVV---INAAKNIWSSVRNNFNNMK  832

Query  302  RGASDFLGNLTTCNGANFNAADASSGGTGTRNALRR  337
               +  + N+ T     +N A     G   +   +
Sbjct  833  NVVTTVMKNVKTTIQNLWNNAVKFLKGIDLKQIGKN  868


>☐ref|XP_001850511.1|  salivary mucin [Culex quinquefasciatus]
 gb|EDS32122.1|  salivary mucin [Culex quinquefasciatus]
Length=317

 GENE ID: 6041218 CpipJ_CPIJ008899 | salivary mucin [Culex quinquefasciatus]

                                                           Sort alignments for this subject sequence by:
                                                             E value  Score  Percent identity
                                                             Query start position  Subject start position

 Score =  176 bits (444),  Expect = 6e-42, Method: Composition-based stats.
 Identities = 31/271 (11%), Positives = 74/271 (27%), Gaps = 11/271 (4%)

Query  140  NIASGSSGGAGQTVPS-NFSTFLPMIPDNYTQA---FQKLVTQFQTLPLDAMDSIVQTLS  195
            N  +      G    +    T    +  N       F              +D  V +++
Sbjct  53   NTPNNYQLSIGNRQNTFANLTDYGTMISNVNINLARFNGTAASGNDAGATLLDDAVLSVA  112

Query  196  DAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQ  255
            +AY N    +         ++  +  RS      Q I D          A     S   +
Sbjct  113  EAYRNESTAVIAGRLAWFNDQIAQANRSASTVHHQHISDLVEDFARDTRAGTLGLSNATR  172

Query  256  QCVGPDLSPTSIGRRVVGKGTD-CVNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTC  314
             C+   +    + R V  +    C+  K +++ ++    + ++        +    L  C
Sbjct  173  DCLARSVQVEDVIRSVENRSEGGCLRAKIQRMLELRAEASSNLTEFLESQQNVEDRLEIC  232

Query  315  NGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLFLPVSLTIEGGKLYAALSSLQT  374
                              +  +  C    + +       L ++      +    +  L+
Sbjct  233  VD------LQDDFDDDMSDFYKVACVSSILFDVQMETAKLELTAEELTAEAGVTVRQLRA  286

Query  375  DVALCASEMGLEIGLVTAQVSSKIAFCQLAQ  405
             +  C  ++       + ++   I  C  A+
Sbjct  287  GLLECVVDVANYAFDASMKLRHWINVCSAAR  317


 Score = 52.4 bits (123),  Expect = 1e-04, Method: Composition-based stats.
 Identities = 21/83 (25%), Positives = 38/83 (45%), Gaps = 1/83 (1%)

Query  9   LALLIAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTP-TQQSNVTLALGNR  67
           L + +    + P    AF +          +      +S++++       +N  L++GNR
Sbjct  6   LLVTLVISLLSPEATLAFGWNWLLGFATTASSTTSSTSSSSTSAASSNTPNNYQLSIGNR  65

Query  68  INTDTALTDYGTNVNGLGINVNR  90
            NT   LTDYGT ++ + IN+ R
Sbjct  66  QNTFANLTDYGTMISNVNINLAR  88


>☐gb|ACI30188.1|  putative 41.9 kDa basic salivary protein [Anopheles darlingi]
Length=173

 Score =  173 bits (437),  Expect = 4e-41, Method: Composition-based stats.
 Identities = 167/173 (96%), Positives = 169/173 (97%), Gaps = 0/173 (0%)

Query  234  DQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRI  293
            DQGAQGFEQVIASFNDSSARVQQCVG DLSPTSIGRRVVGKGTDCVNR+WRQLSDIGGRI
Sbjct  1    DQGAQGFEQVIASFNDSSARVQQCVGSDLSPTSIGRRVVGKGTDCVNRRWRQLSDIGGRI  60

Query  294  AGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLF  353
            AGDILA DRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLF
Sbjct  61   AGDILAVDRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLF  120

Query  354  LPVSLTIEGGKLYAALSSLQTDVALCASEMGLEIGLVTAQVSSKIAFCQLAQL  406
            LPVSLTI+GGKLYAALSSLQTD ALCASEMGLEIGLVTAQVSSKI FCQLAQL
Sbjct  121  LPVSLTIDGGKLYAALSSLQTDAALCASEMGLEIGLVTAQVSSKITFCQLAQL  173


>☐ref|XP_002388385.1|  hypothetical protein MPER_12601 [Moniliophthora perniciosa FA553]
 gb|EEB89315.1|  hypothetical protein MPER_12601 [Moniliophthora perniciosa FA553]
Length=693

 GENE ID: 7931429 MPER_12601 | hypothetical protein
[Moniliophthora perniciosa FA553] (10 or fewer PubMed links)

 Score =  167 bits (422),  Expect = 2e-39, Method: Composition-based stats.
 Identities = 35/323 (10%), Positives = 79/323 (24%), Gaps = 39/323 (12%)
```

```
Query  38   RTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGTNVNGLGINVNRRRRRRRD  97
                    + A+   T    S  T+  G+   + +   +  T   G
Sbjct  221  SLNKGSTTSQASDTATIPAPSIKTMKAGD---SSSIRDETATVKTG------------EG  265

Query  98   IWNIESNPVLAGLAANNPVNGSYIAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNF  157
              N  S       ++  P                + L+  ++  A     G       +
Sbjct  266  TINAGSVSGTPRTSSPPPT---ISTAPAPAILAASGLTATTIGTAR--PPGVRPVKRLSP  320

Query  158  STFLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNET  217
              +     D+  Q    L T  +     ++D+ V T +        GL  +      +E+
Sbjct  321  RLYFAHFVDHMEQFVVFLETVAKRRWDQSLDTDVDTKAKTNGAAENGLVDSPVDEEDSES  380

Query  218  LEHGR-SVIEDGIQSIVDQ-----GAQGFEQVIASFNDSSARVQQCVGPDLSPTSIGRRV  271
             +  + +V    ++  +        A    Q   +  D    ++     D+        +
Sbjct  381  EKKDQVAVWNTLLELYLTLPGVSATATSPAQNEETLRD--KAIRILQSKDIPYDPTHALI  438

Query  272  VGKGTDCVNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGT  331
            +     C       L  +  +    +   +     ++            A+
Sbjct  439  LCSSHGCTKG----LVLLWEK----MGMYEDVLRFWMDRDKEGV---SEASAEVVRHLRM  487

Query  332  RNALRRQCYVRAIGNFPQSLLFL  354
                R   Y   +     S   L
Sbjct  488  YGPQRPHLYPLVLRFLTSSSELL  510


>☐gb|AAX54873.1|  putative 30.3 kDa secreted protein [Aedes aegypti]
Length=295

 Score =  162 bits (409),  Expect = 6e-38, Method: Composition-based stats.
 Identities = 66/306 (21%), Positives = 109/306 (35%), Gaps = 54/306 (17%)

Query  5    MSVVLALLIAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLAL  64
            +SV+   LI ++   PW+        PW  +                      S VT+++
Sbjct  11   LSVMGLFLIPSLEASPWF-------WPWQSRA-------------------SPSIVTISI  44

Query  65   GNRINTDTALTDYGTNVNGLGINVNRRR---------------RRRRDIWNIESNPVLAG  109
            GNR N+++ +TD GT++  L  N++R +               R    + NI  +
Sbjct  45   GNRFNSNS-VTDEGTSIRDLTFNLSRWKGNLTFEKNGTEILITRGNDGVTNIHIDFGSNS  103

Query  110  LAANN---PVNGSYIAIGNQVFHLPNNLSNVSLNIASGSSGGAG-QTVPSNFSTFLP-MI  164
             +  N   P N +      ++F LP+ L  + +   +G+    G            P
Sbjct  104  KSQQNETVPSNATQSRKKREIFGLPS-LPAIPIPTINGTQAAIGENFFQLPNEIANPASN  162

Query  165  PDNYTQAFQKLVTQF-QTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRS  223
              N+ Q +Q L T   + LP+  +D+ V  LS  YS+T K +     L  L ET+ H
Sbjct  163  ASNFLQPWQFLQTLITRRLPMQFLDTAVSYLSQLYSSTAKNMLGHRELTPLAETMGHPSP  222

Query  224  VIEDGI-QSIVDQG-AQGFEQVIASFND-SSARVQQCVGPDLSPTSIGRRVVGKGT--DC  278
            V    +  +  DQ   +        F        Q CV P          +V K
Sbjct  223  VSSTPLWTTSRDQTVGRNPRGSHIPFRGTFLKLAQSCVWPRCPDECCFSPIVTKAHRDGV  282

Query  279  VNRKWR  284
               +W
Sbjct  283  TGIRWE  288


>☐gb|AAU06485.1|  unknown salivary protein [Culicoides sonorensis]
Length=243

 Score =  161 bits (406),  Expect = 2e-37, Method: Composition-based stats.
 Identities = 47/228 (20%), Positives = 77/228 (33%), Gaps = 10/228 (4%)

Query  175  LVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVD  234
            +     +L    ++S +  L    ++  K           N  L   +SV       I
Sbjct  18   VSNLGNSLSNGTVESALDNLQSKVNDAAKRANDT-----TNSALSGLQSVGNTLSNEINK  72

Query  235  QGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIG-RRVVGKGTDCVNRKWRQLSDIGGRI  293
                   QV ++    S+ ++ CV  + +P S   +        CV  K  +   I
Sbjct  73   VSKSAEAQVNSTMKGLSSDIRDCVNQNGNPGSAALKTAKTSAQKCVQDKIDEAKSIINGT  132

Query  294  AGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLF  353
              DI  A  G     GNL+ C+         +     T NA +  C   A+ +     L
Sbjct  133  RNDIQQAKLGVDYLRGNLSNCHLN----VSMNLASLPTFNASQPACLASALLSVQSETLL  188

Query  354  LPVSLTIEGGKLYAALSSLQTDVALCASEMGLEIGLVTAQVSSKIAFC  401
            +P+SL   G +    +  LQ     C + +  EI   T   S  +A C
Sbjct  189  IPLSLASHGAQAVGLVQGLQGYAMKCMANLMEEIAKATLSTSLSVAKC  236


>☐gb|ACH56926.1|  hypothetical protein [Simulium vittatum]
Length=236

 Score =  156 bits (394),  Expect = 4e-36, Method: Composition-based stats.
 Identities = 39/237 (16%), Positives = 80/237 (33%), Gaps = 12/237 (5%)

Query  169  TQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSV---I  225
            +  F       Q      + + V+ L++   +  +        NGL+        V
Sbjct  2    SNPFHTFAELGQK---GLIPNPVEILANLSMDHLENKVKDHVNNGLSNVQNGVYEVQTFG  58

Query  226  EDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPT-SIGRRVVGKGTDCVNRKWR  284
               I  I +  A   + +    N   + V  CV     P  ++ +      + CV  K
Sbjct  59   SKLINFIPNLAANAIKDIQNMIN--LSGVVSCVNKQPLPFSNVIQVAAKSASKCVAEKVD  116

Query  285  QLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRAI  344
            Q ++I    A ++  A     +   ++  C+    + +   S            CY  A+
Sbjct  117  QAANIWNNTADNLRNASNDLQNLRDSINDCSQRISSGSLGLSLSDAM---ALNTCYAAAL  173

Query  345  GNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVALCASEMGLEIGLVTAQVSSKIAFC  401
             N     + LP+++  +  +     +SL+  +  C + +  +I     + S  IA C
Sbjct  174  INVQSETILLPITIGQKIAEGVDWANSLKGYMVKCVANLAQDIATQAFKASQAIAKC  230


>☐gb|ACI30197.1|  putative 41.9 kDa basic salivary protein [Anopheles darlingi]
Length=139

 Score =  149 bits (375),  Expect = 6e-34, Method: Composition-based stats.
 Identities = 138/139 (99%), Positives = 138/139 (99%), Gaps = 0/139 (0%)

Query  268  GRRVVGKGTDCVNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSG  327
            GRRVV KGTDCVNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSG
Sbjct  1    GRRVVRKGTDCVNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSG  60
```

```
Query  328  GTGTRNALRRQCYVRAIGNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVALCASEMGLEI  387
            GTGTRNALRRQCYVRAIGNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVALCASEMGLEI
Sbjct  61   GTGTRNALRRQCYVRAIGNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVALCASEMGLEI  120

Query  388  GLVTAQVSSKIAFCQLAQL  406
            GLVTAQVSSKIAFCQLAQL
Sbjct  121  GLVTAQVSSKIAFCQLAQL  139


>☐gb|ABF18165.1|  putative salivary mucin [Aedes aegypti]
Length=248

                                                       Sort alignments for this subject sequence by:
                                                         E value  Score  Percent identity
                                                         Query start position  Subject start position

 Score =  149 bits (375),  Expect = 6e-34, Method: Composition-based stats.
 Identities = 26/200 (13%), Positives = 61/200 (30%), Gaps = 11/200 (5%)

Query  147  GGAGQTVPSNFSTFLPMIPDNYTQAFQKL--VTQFQTLPLDAMDSIVQTLSDAYSNTFKG  204
               G    +N S     +  N      +            + ++ ++ T+ + Y N  +
Sbjct  55   ISIGNRENTNASIDYGTMITNLQINLARFNGSALLGDATENRLNEVIYTVLEGYRNRSEA  114

Query  205  LFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSP  264
            +  +         L+  +   +           Q  E V  S    +  V+ C+  ++
Sbjct  115  MVDSRLAWIDEVLLDIVQYARQLRQDENNRWLKQFPEHVRTSVAQLNETVRSCLEREVVV  174

Query  265  TSIGRRVVGKG-TDCVNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAAD  323
             +    V  +  + C+  + ++L ++      ++      A D    L TC   +F+ A
Sbjct  175  EELIHAVENRSRSGCLEGRLQKLFELREVAKNNLTEFLASAGDLEDRLETCVNPDFDDAL  234

Query  324  ASSGGTGTRNALRRQCYVRA  343
                     N  +  C
Sbjct  235  D--------NVYQEACISSV  246


 Score = 48.2 bits (112),  Expect = 0.002, Method: Composition-based stats.
 Identities = 19/71 (26%), Positives = 35/71 (49%), Gaps = 3/71 (4%)

Query  20   PWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGT  79
            P   +AF ++  W      T       + ++ ++    S   +++GNR NT+ ++ DYGT
Sbjct  15   PTGTNAFGWS--WFFPRQTTTTTASSGTTSATSSNAVPSMYNISIGNRENTNASI-DYGT  71

Query  80   NVNGLGINVNR  90
             +  L IN+ R
Sbjct  72   MITNLQINLAR  82


>☐ref|XP_771922.1| UG hypothetical protein CNBN1020 [Cryptococcus neoformans var. neoformans
B-3501A]
 gb|EAL17275.1| G hypothetical protein CNBN1020 [Cryptococcus neoformans var. neoformans
B-3501A]
Length=1071

 GENE ID: 4939711 CNBN1020 | hypothetical protein
[Cryptococcus neoformans var. neoformans B-3501A]

 Score =  136 bits (342),  Expect = 4e-30, Method: Composition-based stats.
 Identities = 25/232 (10%), Positives = 56/232 (24%), Gaps = 28/232 (12%)

Query  149  AGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTA  208
                VP++   +     D+       L +    L    +D  +   S    +  K    A
Sbjct  700  TPSYVPASPRQYFAHFVDHRELFIHFLESVALNLWNQKVDPSISNSSTT--SAPKRDMDA  757

Query  209  AGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIG  268
                  + T+    +V    ++  +   A        S   + A +            +
Sbjct  758  P--PPTDPTVIDQTAVWNTLLELYLSSSA------EQSKRKALALINS---DSTPYDPMH  806

Query  269  RRVVGKGTDCVNRKWRQLS-DIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSG  327
              V+          +R+    +       +   +     ++        A      +
Sbjct  807  ALVLCSSVG-----FREGMVRLWEG----MGMYEDVLRYYMDEGQEAEDAQGITPSSKVF  857

Query  328  G-TGTRNALRRQCYVRAIGNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVAL  378
                         Y   +     S    P  LT   G+L   L+ +     +
Sbjct  858  THLEMYGPSHPHLYPLVLRYLTSS----PAILTRHKGELSKILAKIDEYQIM  905


>☐ref|XP_568542.1| G vacuolar membrane protein [Cryptococcus neoformans var. neoformans
JEC21]
 gb|AAW47025.1| G vacuolar membrane protein, putative [Cryptococcus neoformans
var. neoformans JEC21]
Length=1071

 GENE ID: 3255360 CNN01020 | vacuolar membrane protein
[Cryptococcus neoformans var. neoformans JEC21] (10 or fewer PubMed links)

 Score =  136 bits (341),  Expect = 5e-30, Method: Composition-based stats.
 Identities = 25/232 (10%), Positives = 56/232 (24%), Gaps = 28/232 (12%)

Query  149  AGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTA  208
                VP++   +     D+       L +    L    +D  +   S    +  K    A
Sbjct  700  TPSYVPASPRQYFAHFVDHRELFIHFLESVALNLWNQKVDPSISNSSTI--SAPKRDMDA  757

Query  209  AGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIG  268
                  + T+    +V    ++  +   A        S   + A +            +
Sbjct  758  P--PPTDPTVIDQTAVWNTLLELYLSSSA------EQSKRKALALINS---DSTPYDPMH  806

Query  269  RRVVGKGTDCVNRKWRQLS-DIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSG  327
              V+          +R+    +       +   +     ++        A      +
Sbjct  807  ALVLCSSVG-----FREGMVRLWEG----MGMYEDVLRYYMDEGQEAEDAQGITPSSKVF  857

Query  328  G-TGTRNALRRQCYVRAIGNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVAL  378
                         Y   +     S    P  LT   G+L   L+ +     +
Sbjct  858  THLEMYGPSHPHLYPLVLRYLTSS----PAILTRHKGELSKILAKIDEYQIM  905


>☐ref|XP_310399.1| UG AGAP003841-PA [Anopheles gambiae str. PEST]
 emb|CAC35525.1|  hypothetical protein [Anopheles gambiae]
```

```
 gb|EAA05987.2| G AGAP003841-PA [Anopheles gambiae str. PEST]
Length=195

 GENE ID: 1271577 AgaP_AGAP003841 | AGAP003841-PA [Anopheles gambiae str. PEST]
(10 or fewer PubMed links)

 Score =  125 bits (313),  Expect = 9e-27, Method: Composition-based stats.
 Identities = 50/195 (25%), Positives = 80/195 (41%), Gaps = 9/195 (4%)

Query  1    MNRSMSVVLALLIAAIGIIPWYVDAFLFTTPWNPQ--PPRTINQLPVTSATSATTPTQQS  58
            M +   + + LL AA+ + P       +  PW+     P T      +   +++  +  S
Sbjct  1    MLQMRIISVPLLSAAVLLQPLLAAPIFWFLPWSFPALSPTTTTLATTSGTAASSGASNSS  60

Query  59   NVTLALGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNG  118
            NV++A+GNR+NT T L DYGTN+  + INVNR RR   D    +    L   A     NG
Sbjct  61   NVSVAIGNRVNTSTGLDDYGTNITNVQINVNRMRRAVTDADLAKLERKLRQAADEGSTNG  120

Query  119  SYIAIGNQVFHLPNNLSNVSLNI--ASGSSGGAGQTVPSNFSTFLPM-----IPDNYTQA  171
            + I IG     LP+N+SN+++N    +G          +  +T         +
Sbjct  121  TVITIGEHTIRLPHNISNLTVNTYLINGIPENYMMLADTGSATINANETYTPLLSAIRLF  180

Query  172  FQKLVTQFQTLPLDA  186
                 + F
Sbjct  181  KNFFSSLFSKQGSQT  195


>☐gb|AAR18443.1|  putative threonine/serine-rich mucin [Culex quinquefasciatus]
Length=173

 Score =  125 bits (312),  Expect = 1e-26, Method: Composition-based stats.
 Identities = 52/180 (28%), Positives = 85/180 (47%), Gaps = 13/180 (7%)

Query  1    MNRSMSVVLALLIAAIGIIPWYVDAFLFTTPWNPQ--PPRTINQLPVTSATSATTPTQQS  58
            M R  + ++ ++++A+           +  PW+    P  T +    TS+TS  T    S
Sbjct  1    MTRVQTYLIVVIVSAL-FSTTLGSPIFWLLPWSSWLNPSTTSSTSSGTSSTSTATARTPS  59

Query  59   NVTLALGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNG  118
            NV++A+G+R NT   L D GT ++GL IN  RRRR        +  P +  +    P N
Sbjct  60   NVSIAIGDRTNTTAGLDDMGTVIDGLTINAGRRRRS------AQQLPEVDFIV---PSNE  110

Query  119  SYIAIGNQVFHLPN-NLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVT  177
            SY+ IGN V  LP+ N+SN+++N+  G        +P+      P + D+    F  +
Sbjct  111  SYVQIGNHVIRLPSRNISNITINVVDGVPYVGAGGLPAPPQDVPPNLFDHIKNFFDGVFN  170


>☐ref|XP_001864883.1| UG threonine/serine-rich mucin [Culex quinquefasciatus]
 gb|EDS40846.1| G threonine/serine-rich mucin [Culex quinquefasciatus]
Length=173

 GENE ID: 6047887 CpipJ_CPIJ014439 | threonine/serine-rich mucin
[Culex quinquefasciatus]

 Score =  122 bits (305),  Expect = 7e-26, Method: Composition-based stats.
 Identities = 48/180 (26%), Positives = 82/180 (45%), Gaps = 13/180 (7%)

Query  1    MNRSMSVVLALLIAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATT--PTQQS  58
            M R  + ++ ++++A+           +  PW+     +      +  +S +T      S
Sbjct  1    MTRVQTYLIVVIVSAL-FSTTLGSPIFWLLPWSSWLSPSTTSSTSSGTSSTSTATARTPS  59

Query  59   NVTLALGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNG  118
            NV++A+G+R NT   L D GT ++GL IN  RRRR        +  P +  +    P N
Sbjct  60   NVSIAIGDRTNTTAGLDDMGTVIDGLTINAGRRRRS------AQQPPEVDFIV---PSNE  110

Query  119  SYIAIGNQVFHLPN-NLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVT  177
            SYI IGN V  LP+ N+SN+++N+  G        +P+      P + D+    F  +
Sbjct  111  SYIQIGNHVIRLPSRNISNITINVVDGVPYVGAGGLPAPPQDVPPNLFDHIKNFFDGVFN  170


>☐ref|XP_001841204.1| G hypothetical protein CC1G_10201 [Coprinopsis cinerea okayama7#130]
 gb|EAU80634.1| G hypothetical protein CC1G_10201 [Coprinopsis cinerea okayama7#130]
Length=1003

 GENE ID: 6017881 CC1G_10201 | hypothetical protein
[Coprinopsis cinerea okayama7#130]

 Score =  121 bits (302),  Expect = 2e-25, Method: Composition-based stats.
 Identities = 35/361 (9%), Positives = 84/361 (23%), Gaps = 42/361 (11%)

Query  10   ALLIAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRIN  69
             L I               +     +       +  +   +   P   S  TL    R
Sbjct  611  PLAIKQPTPTTATTTTSYLSYLALNRSTTVPTVVSSSETQTVAAPPSPSIKTL----RAE  666

Query  70   TDTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIGNQVFH  129
              +   D   +V+G                   S   ++        V+
Sbjct  667  GASGRRD--GSVDGATP------------PPSGSAFGVSTAVVGANVDVGTALAHGAGAG  712

Query  130  LPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAMDS  189
              + L   S         G       + S +     D+  Q    L T        ++D
Sbjct  713  AVSGLPPRSSTAPPEPPAGPPPVQRLSPSLYFSHFVDHMQQFVVFLETVAVRRWGQSVDD  772

Query  190  -IVQTLSDAYSNTFKG----------LFTAAGLNGLNETLEHGR--SVIEDGIQSIVDQG  236
                     +S   +G                L+ + +  +     +V    ++  +
Sbjct  773  RDNSGEVGVWSEKGQGQGPQENENQGAEDTHHLDDIAQAQDRQDQIAVWNTLLELYLTLP  832

Query  237  AQGFEQVIASFNDSS---ARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRI  293
            A+  E   A F++S      ++     ++   +    ++             L  +  +
Sbjct  833  AKDNESDRAVFDESKMREKALRVLRSREIPYDTTHALILCSTHAFT----DGLVLLWEK-  887

Query  294  AGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLF  353
               +   +     ++    +   +  NA++                Y   +     +
Sbjct  888  ---LGMVEDVLRFWIDRHNSDPLSYPNASNKVVENLMLYGPQHPNLYPLVLRFLTSTEEL  944

Query  354  L  354
            L
Sbjct  945  L  945


>☐gb|ABI83794.1|  gSG10 salivary mucin [Anopheles funestus]
Length=190
```

```
 Score =  114 bits (283),  Expect = 3e-23, Method: Composition-based stats.
 Identities = 46/190 (24%), Positives = 75/190 (39%), Gaps = 14/190 (7%)

Query  3    RSMSVVLALLIAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSN---  59
            +   ++  LL  A+   P       +  PW+     T +    +S +   + T  ++
Sbjct  1    QIRIIIAPLLSVAVLTQPLLAAPIFWFLPWSFPNTFTTSSTTSSSTSGTASSTSGASNSS  60

Query  60   -VTLALGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNG  118
             V++A+GNR+NT T L DYGTN+  + INVNR RR   +    +    L    A    N
Sbjct  61   NVSVAIGNRVNTSTGLDDYGTNITNVQINVNRMRRAITENDLAK----LERRYARQTANE  116

Query  119  SYIAIGNQVFHLPNNLSNVSLNI--ASGSSGGAGQTVPSNFST----FLPMIPDNYTQAF  172
            S I IG     LP N+SN+++N    +G          +  +         +
Sbjct  117  STITIGEHTIRLPANISNLTVNTYLINGIPDSTMNLPDTGMTINTDEASTPLLTAVNIFK  176

Query  173  QKLVTQFQTL  182
              L + F
Sbjct  177  NFLSSFFGKQ  186


>☐ref|XP_001661479.1| UG hypothetical protein AaeL_AAEL011183 [Aedes aegypti]
 gb|ABF18096.1|  putative mucin [Aedes aegypti]
 gb|EAT36759.1| G conserved hypothetical protein [Aedes aegypti]
Length=180

 GENE ID: 5574465 AaeL_AAEL011183 | hypothetical protein [Aedes aegypti]
(10 or fewer PubMed links)

 Score =  106 bits (263),  Expect = 6e-21, Method: Composition-based stats.
 Identities = 46/179 (25%), Positives = 77/179 (43%), Gaps = 13/179 (7%)

Query  5    MSVVLALLIAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLAL  64
             S+V+AL + A+      V  F F   +      + +     S++S++T  +     +++
Sbjct  6    FSIVVALTLCAVLPSTLGVPIFWFPWSYLLPTTSSSSASGTASSSSSSTTLRPIMANISI  65

Query  65   GNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIG  124
            GNR NT   + D GT +NGL I+  RR      + N  S             + S+I +G
Sbjct  66   GNRRNTSA-MVDSGTVINGLTISRGRRDVPLETLRNDPSV-----------TSESFIQVG  113

Query  125  NQVFHL-PNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTL  182
            NQV  L P N++N+++NI  G     G T     S       +  +  F ++   F  L
Sbjct  114  NQVIRLPPGNVANLTINIIDGVPYVFGNTESDVSSGSATGSSEGMSGLFSRVRNFFGDL  172


>☐gb|ACI30119.1|  gSG10 salivary mucin [Anopheles darlingi]
Length=192

 Score =  102 bits (253),  Expect = 8e-20, Method: Composition-based stats.
 Identities = 56/187 (29%), Positives = 87/187 (46%), Gaps = 9/187 (4%)

Query  2    NRSMSVVLALLIAAIGIIPWYVDAFLFTTPWNPQPPR--TINQLPVTSATSATTPTQQSN  59
             RS  +++ LL AA+   P       +  PW+       T +    TS++++   +  SN
Sbjct  1    MRSSQILVHLLSAALLAQPLLAAPIFWFLPWSFPNFPTSTSSTSGSTSSSASNGASNSSN  60

Query  60   VTLALGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIW---NIESNPVLAGLAANNPV  116
            V++A+GNR+NT T L DYGTN+  + INVNR+RR   + +   N   +    G  A
Sbjct  61   VSVAIGNRVNTSTGLDDYGTNITNVQINVNRQRRSVANTFLNLNDLGDGGGGGDDAEADT  120

Query  117  NGSYIAIGNQVFHLPNNLSNVSLNI--ASGSSGGAGQTVPSNFSTFLPM--IPDNYTQAF  172
             GSYI+IG  +  LP N+SN++LN    +G        VPS   +      +
Sbjct  121  GGSYISIGEHIIRLPANISNLTLNTYLINGLPVDESMAVPSEMGSAGETNLLVTTIRLFR  180

Query  173  QKLVTQF  179
              L +
Sbjct  181  NFLKSFL  187


>☐gb|ACI30120.1|  gSG10 salivary mucin [Anopheles darlingi]
Length=193

 Score = 99.0 bits (244),  Expect = 8e-19, Method: Composition-based stats.
 Identities = 54/188 (28%), Positives = 85/188 (45%), Gaps = 10/188 (5%)

Query  2    NRSMSVVLALLIAAIGIIPWYVDAFLFTTPWNPQPP--RTINQLPVTSATSATTPTQQSN  59
             RS  +++ LL AA+   P       +  PW+       T +    TS+++++  +  SN
Sbjct  1    MRSSQILVHLLSAALLAQPLLAAPIFWFLPWSFPNFPTSTSSPSGSTSSSASSGASNSSN  60

Query  60   VTLALGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIW---NIESNPVLAGLAANNPV  116
            V++A+GNR+NT T L DYGTN+  + INV R+RR   + +   N   +    G  A
Sbjct  61   VSVAIGNRVNTSTGLDDYGTNITNVQINVKRQRRSVANTFLNLNDLGDGGSGGDDAEADT  120

Query  117  NGSYIAIGNQVFHLPNNLSNVSLNI--ASGSSGGAGQTVPSNFSTFLPM---IPDNYTQA  171
             GSYI+IG  +  LP N+SN++LN    +         VPS           +
Sbjct  121  GGSYISIGEHIIRLPANISNLTLNTYLVNRVPVDESMAVPSEMDGNAGETNLLVTTIRLF  180

Query  172  FQKLVTQF  179
               L +
Sbjct  181  RNFLKSFL  188


>☐gb|ACI30122.1|  gSG10 salivary mucin [Anopheles darlingi]
Length=199

 Score = 97.1 bits (239),  Expect = 3e-18, Method: Composition-based stats.
 Identities = 53/194 (27%), Positives = 82/194 (42%), Gaps = 16/194 (8%)

Query  2    NRSMSVVLALLIAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSN--  59
             RS  +++ LL AA+   P       +  PW+     T       S +S ++
Sbjct  1    MRSSQILVHLLSAALLAQPLLAAPIFWFLPWSFPNFPTSTSSTSGSTSSTSSSASSGASN  60

Query  60   ---VTLALGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIWNIESN-------PVLAG  109
               V++A+GNR+NT T L DYGTN+  + INVNR+RR   + +   ++           G
Sbjct  61   SSNVSVAIGNRVNTSTGLDDYGTNITNVQINVNRQRRSVANTFLNLNDLGDGGGGGGGGG  120

Query  110  LAANNPVNGSYIAIGNQVFHLPNNLSNVSLNI--ASGSSGGAGQTVPSNFSTFLPM--IP  165
              A     GSYI+IG  +  LP N+SN++LN    +G        VPS   +      +
Sbjct  121  DDAKADTGGSYISIGEHIIRLPANISNLTLNTYLINGLPVDESMAVPSEMGSAGETNVLV  180

Query  166  DNYTQAFQKLVTQF  179
                     L +
```

```
Sbjct  181  TTIRLFRNILKSFL  194


>☐gb|ABJ97695.1|  putative threonine/serine-rich mucin [Toxorhynchites amboinensis]
Length=172

 Score = 80.9 bits (197),  Expect = 2e-13, Method: Composition-based stats.
 Identities = 40/156 (25%), Positives = 65/156 (41%), Gaps = 14/156 (8%)

Query  34   PQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGTNVNGLGINVNRRRR  93
               P   +    TS+TS++T     N T++LGNR NT   + D GTN++GL I+ ++R
Sbjct  17   FGMPTYTSGTSSTSSTSSSTARSPVNATISLGNRENTSA-VEDLGTNIDGLTISRSKRFV  75

Query  94   R-----RRDIWNIESNPVL------AGLAANNPVNGSYIAIGNQVFHL-PNNLSNVSLNI  141
                         ++ P +      A  +     N S I +GNQ+  L P  +SN++++I
Sbjct  76   PLLIPNVTGFSFFQNLPGIFTRQKRAAESIPPLSNESVIRVGNQIIKLPPRGISNLTIHI  135

Query  142  ASGSSGGAG-QTVPSNFSTFLPMIPDNYTQAFQKLV  176
              G S         +  +       D     F  L
Sbjct  136  IDGGSYSEKFTPFQTPANGTWAPFVDQIKNFFDGLF  171


>☐ref|XP_001882977.1|  predicted protein [Laccaria bicolor S238N-H82]
 gb|EDR06605.1|  predicted protein [Laccaria bicolor S238N-H82]
Length=1039

 GENE ID: 6078472 LACBIDRAFT_251110 | hypothetical protein
[Laccaria bicolor S238N-H82] (10 or fewer PubMed links)

 Score = 55.1 bits (130),  Expect = 2e-05, Method: Composition-based stats.
 Identities = 19/221 (8%), Positives = 47/221 (21%), Gaps = 23/221 (10%)

Query  138  SLNIASGSSGGAGQTV---PSNFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQTL  194
              N        +         +   + P   DN  Q    L T        ++D    T
Sbjct  656  KSNTPFAPPPKSPSPTPHHRLSPKLYFPHFVDNVDQFLVFLETVAVRRWNQSVDPTPPTP  715

Query  195  SDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARV  254
                         A+   G+        +V    ++  +     G  +           +
Sbjct  716  GMVGVLGGVEPPDASDEVGVEMDRRDQVAVWNTLLELYLTLPLPGGAR--------KEAM  767

Query  255  QQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTC  314
            +      +    ++          +     +   +   +   +     ++
Sbjct  768  RVLRSGKIPYEPMHALILCSTHS-----FTDGLVL---LYEKMGMYEDVLRFWMD---KH  816

Query  315  NGANFNAADASSGGTGT-RNALRRQCYVRAIGNFPQSLLFL  354
            N A+   A A                Y   +     +   L
Sbjct  817  NQASSPGASAKVVEHLMHYGNQHPHLYPLVLRFLTSTPELL  857


>☐ref|XP_759967.1|  hypothetical protein UM03820.1 [Ustilago maydis 521]
 gb|EAK84746.1|  hypothetical protein UM03820.1 [Ustilago maydis 521]
Length=1372

 GENE ID: 3631904 UM03820.1 | hypothetical protein [Ustilago maydis 521]

 Score = 55.1 bits (130),  Expect = 2e-05, Method: Composition-based stats.
 Identities = 32/332 (9%), Positives = 82/332 (24%), Gaps = 49/332 (14%)

Query  93    RRRRDIWNIESNPVLAGLAANNPVNGSYIAIGNQV---FHLPNNLSNVSLNIASGSSGGA  149
             RR              G   N     S + I  +V       +      ++  +
Sbjct  901   RRNSTALVEPYTVGGEGSGKNATRPSSTVKITGEVDADASSGDAADASGVSGVNTEQDKM  960

Query  150   GQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAMD-----SIVQTLS---------  195
                   +  T+      +     + L           ++         TLS
Sbjct  961   DAYPVPSPRTYFAHFIQHRQHFTRFLEVVALARWGQVVNMDGELPADHTLSPVQAKNAPA  1020

Query  196   -------DAYSNTFKGLFTAAGLNGLNE------TLEHGRSVIEDGIQSIVDQGAQGFEQ  242
                    + Y +  +      GL+  +E       L   +S+    ++  +   +    Q
Sbjct  1021  YSEKAGGEEYMDDVERALRELGLDSNDEDAYENQELLDQKSIWNTLLELYLSSCSPSGSQ  1080

Query  243   VIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILAADR  302
                +                 P  +   V+     C    +     +   +   +
Sbjct  1081  RQRALRLLQQ-------HSTLPYDMSHAVML----CSMENFDDGLIL---LYERMGMYEE  1126

Query  303   GASDFLGN-LTTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLFLPVSLTIE  361
                  + + L     A+ +++           +L+ + Y   +     S    P+ L+
Sbjct  1127  VVRLHMDSALVEEEDASGSSSGKVIAALARYGSLQPELYDLVLRFLVSS----PILLSRH  1182

Query  362   GGKLYAALSSLQTDVALCASEMGLEIGLVTAQ  393
                L + L+ ++    +   E+   +
Sbjct  1183  TSDLLSILNVIREKSLMSTLEIVQTLSSTPHT  1214


>☐ref|ZP_06695858.1|  phage tail protein [Enterococcus faecium E1636]
 gb|EFF22799.1|  phage tail protein [Enterococcus faecium E1636]
Length=1110

                                                       Sort alignments for this subject sequence by:
                                                         E value  Score  Percent identity
                                                         Query start position  Subject start position

 Score = 54.3 bits (128),  Expect = 3e-05, Method: Composition-based stats.
 Identities = 20/212 (9%), Positives = 58/212 (27%), Gaps = 22/212 (10%)

Query  100  NIESNPVLAGLAANNPVNGSYIAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFST  159
             I +             NG + +I     ++ N++S  + +I +         V +
Sbjct  780  GIVTTFHTIFDPVVAWWNGIWTSISTTASNIWNSISATASSIWNSIKNTITSLVQAA---  836

Query  160  FLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLE  219
                        +  + +    +      +     S+ +        T++  N +N
Sbjct  837  -----ATVIQNIWSTVSSWLGGIWNSISSTA----SNIW-----NSVTSSISNAINAAKS  882

Query  220  HGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSP-TSIGRRVVGKGTDC  278
              +SV       I    +  +  +  +  +    +   ++   +         T
Sbjct  883  AIQSVWNS----ISSWISGIWNGIKNTALNLWNGITSTISSKVNDGKNAISSGWSNLTGI  938

Query  279  VNRKWRQLSDIGGRIAGDILAADRGASDFLGN  310
            V+  +  +      I   I        D++ +
Sbjct  939  VSDIFNNVKSTIANIWEGIKKTVSAPIDWIRD  970
```

```
 Score = 50.9 bits (119),  Expect = 3e-04, Method: Composition-based stats.
 Identities = 26/280 (9%), Positives = 65/280 (23%), Gaps = 30/280 (10%)

Query  12    LIAAIGIIPWYVDAFLFTTPWN--------PQPPRTINQLPVTSATSATTPTQQSNVTLA  63
             L+ A     W   +     W+                   +     A     ++++
Sbjct  747   LLGAAWQAVWTPISDFLKNTWDTMTQWISIAWNGIVTTFHTIFDPVVAWWNGIWTSISTT  806

Query  64    LGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAI  123
               N  N+ +A      N     I       +     I++             N
Sbjct  807   ASNIWNSISATASSIWN----SIKNTITSLVQAAATVIQNIWSTVSSWLGGIWNSISSTA  862

Query  124   GNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLP  183
              N    + +++SN  +N A           S   +    I    +  +  +      L
Sbjct  863   SNIWNSVTSSISNA-INAAK-----------SAIQSVWNSISSWISGIWNGIKNTALNLW  910

Query  184   LDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQV  243
                 +I      +  N  K   ++   N      +   +V                 +
Sbjct  911   NGITSTI-----SSKVNDGKNAISSGWSNLTGIVSDIFNNVKSTIANIWEGIKKTVSAPI  965

Query  244   IASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKW  283
                 D  + +   +    +    +   ++    + +  +
Sbjct  966   -DWIRDKISGIFDNLNISIPHIPLPEFIMEGSFNPLKGQI  1004
```

```
>☐ref|NP_690686.2| structural protein [Bacillus phage SPP1]
 emb|CAL18687.1| unnamed protein product [Bacillus phage SPP1]
Length=1032

 GENE ID: 955328 SPP1p035 | gene 18 [Bacillus phage SPP1]
```

(10 or fewer PubMed links)

```
 Score = 53.2 bits (125),  Expect = 5e-05, Method: Composition-based stats.
 Identities = 15/219 (6%), Positives = 56/219 (25%), Gaps = 9/219 (4%)

Query  123  IGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQA---FQKLVTQF  179
            I      +   +S    +  +     A     +  +               +  +  +
Sbjct  674  IKEYTVKIWGAVSKWLSDAWNKIKQAASVVWQALVTLIKKNFEMQKKIVTTVWNTIKSVS  733

Query  180  QTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQG  239
              +    + S + ++ +   N  K ++           L   +       ++     +
Sbjct  734  SKVWNG-IKSFLSSVWNGIVNAGKTVWNG-LKTFFTAWLNFQKKTWSTIWNAVKTAVTKV  791

Query  240  FEQVIASFNDSSARVQQCVGPDLS-PTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDIL  298
            ++ +++        ++      L+    +   V       V   W +L         +
Sbjct  792  WKGIVSVGKTVWNGLKSFFTSWLNGQKKLFSTVWNAVKTAVTNVWNKLKSAASSTFNAL-  850

Query  299  AADRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRR  337
                 ++ +   +    +  + +N+   +   G    +  R
Sbjct  851  --KSSVTNIMNKVKDKIKSIWNSVMSFFKGINLGSIGRN  887
```

```
>☐ref|ZP_00602924.1|  Phage tail tape measure protein TP901, core region [Enterococcus
faecium DO]
 ref|ZP_05671611.1|  phage tail tape measure protein TP901 [Enterococcus faecium 1,231,410]
 ref|ZP_05832791.1|  phage tail tape measure protein TP901 [Enterococcus faecium C68]
 ref|ZP_06676894.1|  phage-related tail protein [Enterococcus faecium E1162]
 gb|EAN10680.1|  Phage tail tape measure protein TP901, core region [Enterococcus
faecium DO]
 gb|EEV54944.1|  phage tail tape measure protein TP901 [Enterococcus faecium 1,231,410]
 gb|EEW61542.1|  phage tail tape measure protein TP901 [Enterococcus faecium C68]
 gb|EFF34992.1|  phage-related tail protein [Enterococcus faecium E1162]
Length=1143

                                                        Sort alignments for this subject sequence by:
                                                          E value  Score  Percent identity
                                                          Query start position  Subject start position

 Score = 52.0 bits (122),  Expect = 1e-04, Method: Composition-based stats.
 Identities = 20/212 (9%), Positives = 56/212 (26%), Gaps = 22/212 (10%)

Query  100   NIESNPVLAGLAANNPVNGSYIAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFST  159
              I +            NG + AI     ++ N++S  + +I +         V +
Sbjct  813   GIVTTFHTIFDPVVAWWNGIWTAISTTASNIWNSISATASSIWNSIKNTITSLVQAA---  869

Query  160   FLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLE  219
                        +  + +    +      +          +      T++  N +N
Sbjct  870   -----ATVIQNIWSTVSSWLGGIWNSISSTA---------SNIWNSVTSSISNAINAAKS  915

Query  220   HGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSP-TSIGRRVVGKGTDC  278
               +SV       I   +  +  +  +  +    +   +   ++   +         T
Sbjct  916   AIQSVWNS----ISSWISGIWNGIKNTALNLWNGITSTISSKVNDGKNAISSGWSNLTGI  971

Query  279   VNRKWRQLSDIGGRIAGDILAADRGASDFLGN  310
             V+  +  +      I   I       D++ +
Sbjct  972   VSDIFNNVKSTIANIWEGIKKTVSAPIDWIRD  1003


 Score = 48.9 bits (114),  Expect = 0.001, Method: Composition-based stats.
 Identities = 24/280 (8%), Positives = 61/280 (21%), Gaps = 30/280 (10%)

Query  12    LIAAIGIIPWYVDAFLFTTPWN--------PQPPRTINQLPVTSATSATTPTQQSNVTLA  63
             L+ A     W   +     W+                   +     A     + ++
Sbjct  780   LLGAAWQAVWTPISDFLKNTWDTMTQWISIAWNGIVTTFHTIFDPVVAWWNGIWTAISTT  839

Query  64    LGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAI  123
               N  N+ +A      N     I       +     I++             N
Sbjct  840   ASNIWNSISATASSIWN----SIKNTITSLVQAAATVIQNIWSTVSSWLGGIWNSISSTA  895

Query  124   GNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLP  183
              N    + +++SN                 S   +    I    +  +  +      L
Sbjct  896   SNIWNSVTSSISNA------------INAAKSAIQSVWNSISSWISGIWNGIKNTALNLW  943

Query  184   LDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQV  243
                 +I      +  N  K   ++   N      +   +V                 +
Sbjct  944   NGITSTI-----SSKVNDGKNAISSGWSNLTGIVSDIFNNVKSTIANIWEGIKKTVSAPI  998

Query  244   IASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKW  283
                 D  + +   +    +    +   ++    + +  +
Sbjct  999   DW-IRDKISGIFDNLNISIPHIPLPEFIMEGSFNPLKGQI  1037
```

```
>☐ref|ZP_05713621.1|  tail protein [Enterococcus faecium DO]
Length=574

                                                       Sort alignments for this subject sequence by:
                                                         E value  Score  Percent identity
                                                         Query start position  Subject start position

 Score = 51.6 bits (121),  Expect = 2e-04, Method: Composition-based stats.
 Identities = 20/212 (9%), Positives = 56/212 (26%), Gaps = 22/212 (10%)

Query  100  NIESNPVLAGLAANNPVNGSYIAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFST  159
             I +             NG + AI     ++ N++S  + +I +         V +
Sbjct  244  GIVTTFHTIFDPVVAWWNGIWTAISTTASNIWNSISATASSIWNSIKNTITSLVQAA---  300

Query  160  FLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLE  219
                        +  + +    +      +          +      T++  N +N
Sbjct  301  -----ATVIQNIWSTVSSWLGGIWNSISSTA---------SNIWNSVTSSISNAINAAKS  346

Query  220  HGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSP-TSIGRRVVGKGTDC  278
              +SV       I    +  +  +  +  +    +   +   ++   +         T
Sbjct  347  AIQSVWNS----ISSWISGIWNGIKNTALNLWNGITSTISSKVNDGKNAISSGWSNLTGI  402

Query  279  VNRKWRQLSDIGGRIAGDILAADRGASDFLGN  310
            V+  +  +      I   I        D++ +
Sbjct  403  VSDIFNNVKSTIANIWEGIKKTVSAPIDWIRD  434


 Score = 48.6 bits (113),  Expect = 0.001, Method: Composition-based stats.
 Identities = 24/280 (8%), Positives = 61/280 (21%), Gaps = 30/280 (10%)

Query  12   LIAAIGIIPWYVDAFLFTTPWN--------PQPPRTINQLPVTSATSATTPTQQSNVTLA  63
            L+ A     W   +      W+                   +     A      + ++
Sbjct  211  LLGAAWQAVWTPISDFLKNTWDTMTQWISIAWNGIVTTFHTIFDPVVAWWNGIWTAISTT  270

Query  64   LGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAI  123
              N  N+ +A      N     I        +     I++             N
Sbjct  271  ASNIWNSISATASSIWN----SIKNTITSLVQAAATVIQNIWSTVSSWLGGIWNSISSTA  326

Query  124  GNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLP  183
             N    + +++SN                  S   +    I    +  +  +      L
Sbjct  327  SNIWNSVTSSISNA------------INAAKSAIQSVWNSISSWISGIWNGIKNTALNLW  374

Query  184  LDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQV  243
                 +I      +  N  K   ++    N      +   +V                  +
Sbjct  375  NGITSTI-----SSKVNDGKNAISSGWSNLTGIVSDIFNNVKSTIANIWEGIKKTVSAPI  429

Query  244  IASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKW  283
                 D  + +   +   +    +   ++    + +  +
Sbjct  430  DW-IRDKISGIFDNLNISIPHIPLPEFIMEGSFNPLKGQI  468


>☐ref|ZP_05661457.1|  phage tail tape measure protein TP901 [Enterococcus faecium 1,231,502]
 gb|EEV44790.1|  phage tail tape measure protein TP901 [Enterococcus faecium 1,231,502]
Length=1139

                                                       Sort alignments for this subject sequence by:
                                                         E value  Score  Percent identity
                                                         Query start position  Subject start position

 Score = 51.2 bits (120),  Expect = 2e-04, Method: Composition-based stats.
 Identities = 20/212 (9%), Positives = 56/212 (26%), Gaps = 22/212 (10%)

Query  100   NIESNPVLAGLAANNPVNGSYIAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFST  159
              I +             NG + AI     ++ N++S  + +I +         V +
Sbjct  813   GIVTTFHTIFDPVVAWWNGIWTAISTTASNIWNSISATASSIWNSIKNTITSLVQAA---  869

Query  160   FLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLE  219
                         +  + +    +      +          +      T++  N +N
Sbjct  870   -----ATVIQNIWSTVSSWLGGIWNSISSTA---------SNIWNSVTSSISNAINAAKS  915

Query  220   HGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSP-TSIGRRVVGKGTDC  278
               +SV       I    +  +  +  +  +    +   +   ++   +         T
Sbjct  916   AIQSVWNS----ISSWISGIWNGIKNTALNLWNGITSTISSKVNDGKNAISSGWSNLTGI  971

Query  279   VNRKWRQLSDIGGRIAGDILAADRGASDFLGN  310
             V+  +  +      I   I        D++ +
Sbjct  972   VSDIFNNVKSTIANIWEGIKKTVSAPIDWIRD  1003


 Score = 48.9 bits (114),  Expect = 0.001, Method: Composition-based stats.
 Identities = 24/253 (9%), Positives = 64/253 (25%), Gaps = 26/253 (10%)

Query  12    LIAAIGIIPWYVDAFLFTTPWN--------PQPPRTINQLPVTSATSATTPTQQSNVTLA  63
             L+ A     W   +      W+                   +     A      + ++
Sbjct  780   LLGAAWQAVWTPISDFLKNTWDTMTQWVSIAWNGIVTTFHTIFDPVVAWWNGIWTAISTT  839

Query  64    LGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAI  123
               N  N+ +A      N     I        +     I++             N
Sbjct  840   ASNIWNSISATASSIWN----SIKNTITSLVQAAATVIQNIWSTVSSWLGGIWNSISSTA  895

Query  124   GNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLP  183
              N    + +++SN                  S   +    I    +  +  +      L
Sbjct  896   SNIWNSVTSSISNA------------INAAKSAIQSVWNSISSWISGIWNGIKNTALNLW  943

Query  184   LDAMDSIVQTLSDAYS--NTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFE  241
                  +I   ++D  +  ++     T    +  N       ++ E   +++        +
Sbjct  944   NGITSTISSKVNDGKNAISSGWSNLTGIVSDIFNNVKSTIANIWEGIKKTVSAPIDWIRD  1003

Query  242   QVIASFNDSSARV  254
             ++ + F++ +  +
Sbjct  1004  KISSIFDNLNISI  1016


>☐ref|YP_176319.1|  tail length tape measure protein [Bacillus clausii KSM-K16]
 dbj|BAD65358.1|  tail length tape measure protein [Bacillus clausii KSM-K16]
Length=1631

 GENE ID: 3202947 ABC2824 | tail length tape measure protein
[Bacillus clausii KSM-K16] (10 or fewer PubMed links)
```

```
 Score = 50.5 bits (118),  Expect = 4e-04, Method: Composition-based stats.
 Identities = 21/221 (9%), Positives = 55/221 (24%), Gaps = 46/221 (20%)

Query  158   STFLPMIPDNYTQAFQKLVTQFQTLPL---------DAMDSIVQTLSDAYSNTFKGLFTA  208
                   +   +   +      F  +             + ++V T+ +   N F  ++
Sbjct  821   GGAWTALKTAWNNIWTFTKNTFNAIWNFLKSVWDGIKTVFNVVVTIFNIVKNGFNRVWNT  880

Query  209   ---------------------AGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASF  247
                                  A    ++  + +  S       +I    +  +  +   +
Sbjct  881   TKSIFLAVWNFLKNTWNNIWGAIREFVSRIVTNISSGFNTAKNTISRIFSAVWSFIKKIW  940

Query  248   NDSSARVQQCVGPDLSP--------TSIGRRVVGKGTDCVNRKW--------RQLSDIGG  291
             +     ++  V   ++          +  + +  +    V   W         ++  I
Sbjct  941   SSIWGSIRDFVARIVNAIKTRFTESWNTIKSIWNRVWGFVKDIWGKIWGSIKDRVDRIFN  1000

Query  292   RIAGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTR  332
              I      A    S  LG++       FN     + G   R
Sbjct  1001  GIRDGFNKAKDKVSTILGDMWQAVKDKFNDIVDIAKGLPKR  1041


>☐ref|ZP_05924009.1|  phage tail tape measure protein TP901 [Enterococcus faecium TC
6]
 ref|ZP_06447221.1|  phage tail tape measure protein TP901 [Enterococcus faecium D344SRF]
 gb|EEW64142.1|  phage tail tape measure protein TP901 [Enterococcus faecium TC
6]
 gb|EFD09276.1|  phage tail tape measure protein TP901 [Enterococcus faecium D344SRF]
Length=1238

 Score = 48.6 bits (113),  Expect = 0.001, Method: Composition-based stats.
 Identities = 22/228 (9%), Positives = 57/228 (25%), Gaps = 20/228 (8%)

Query  110   LAANNPVNGSYIAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYT  169
                 N  N            + + + N+S  I  G       +V +  +     I    +
Sbjct  889   TVLTNTWNTISQTAIAIWQPIQDFIINLSHAILDGI-INIWNSVVTWLTGTWNTIASTAS  947

Query  170   QAFQKLV----TQFQTLPLDAMD--SIVQTLSDAYSNTFKGL-------FTAAGLNGLNE  216
               +  +        Q+     +   + + +      N  K          T++  N  N
Sbjct  948   TIWNSIKLIVYNLVQSTKDGIISVWTAITSWLTDKWNAIKNSASNTWNSVTSSISNATNA  1007

Query  217   TLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSP-TSIGRRVVGKG  275
                  +SV       I    +  +  +  +  +    +   +   ++   +
Sbjct  1008  ANSVIQSVWNS----ISSWISGIWNGIKNTALNLWNGITSTISSKVNDGKNAISSGWSNL  1063

Query  276   TDCVNRKWRQLSDIGGRIAGDILAADRGASDF-LGNLTTCNGANFNAA  322
             T  V+  +  +      I   I        D+    ++        +
Sbjct  1064  TGIVSDIFNNVKSTIANIWEGIKKTVSTPIDWIRDKISGIFDNLNISL  1111


>☐emb|CAC35527.1|  gSG9 protein [Anopheles gambiae]
Length=76

 Score = 46.2 bits (107),  Expect = 0.008, Method: Composition-based stats.
 Identities = 31/78 (39%), Positives = 43/78 (55%), Gaps = 10/78 (12%)

Query  4    SMSVVLALLIAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLA  63
           S  +V+ ++       P++   F +  P+  QP           A++A    Q+SNVT+A
Sbjct  9    STLLVVVVIAPRHQCSPFF---FQYNRPYLSQP-------SSQLASTAANVVQRSNVTVA  58

Query  64   LGNRINTDTALTDYGTNV  81
           LGNRINTDTAL DYGT V
Sbjct  59   LGNRINTDTALDDYGTRV  76


>☐ref|ZP_04674169.1|  conserved hypothetical protein [Lactobacillus paracasei subsp.
paracasei 8700:2]
 gb|EEQ64604.1|  conserved hypothetical protein [Lactobacillus paracasei subsp.
paracasei 8700:2]
Length=1056

 Score = 44.3 bits (102),  Expect = 0.031, Method: Composition-based stats.
 Identities = 31/281 (11%), Positives = 71/281 (25%), Gaps = 19/281 (6%)

Query  16   IGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALT  75
            + I  W          W+                       +   T+A+    N
Sbjct  661  LIIKNWGAIVTWLKGVWSTVTSFFSGMWTS----------IKQIFTIAINAITNFLKPAF  710

Query  76   DYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIGNQVFHLPNNLS  135
                NV    I    +         I+    +A  A    +          +    N +
Sbjct  711  TAAVNV-IKSIWNGIKSFFSAFWNGIKVIFTVAITAIAVIIGTYLNIWKTIITTAMNFIK  769

Query  136  NVSLNIASGSSGGAGQTVPS---NFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQ  192
             +  N+ +G     G  + S      +    I    +  F K+ +   ++  +  + +
Sbjct  770  GIITNVWNGIKSFFGPILASIGNVIRSAWNSISSVTSSVFNKVKSVISSIWNNIKNVVSN  829

Query  193  TLSDAYS--NTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDS  250
             ++   S  +      ++   N  N       +V      +I +        V +++N
Sbjct  830  VVNAVKSVVSNAWNAVSSTTSNIFNSVKSAVSNVWNSIKSTISNVVGSIRNAVSSAWNAV  889

Query  251  SARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGG  291
            S+     V   +   +     +    D V      +
Sbjct  890  SSVTSN-VWNSI--KNAISGPINTAKDIVRGAIDAIRGFFN  927


>☐gb|EDL20001.1| G Unc-51 like kinase 1 (C. elegans), isoform CRA_a [Mus musculus]
Length=1159

 GENE ID: 22241 Ulk1 | Unc-51 like kinase 1 (C. elegans) [Mus musculus]
(Over 10 PubMed links)

 Score = 43.9 bits (101),  Expect = 0.031, Method: Composition-based stats.
 Identities = 24/155 (15%), Positives = 45/155 (29%), Gaps = 11/155 (7%)

Query  15   AIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTAL  74
            A   +P        +       P+  +     S    ++  + S  T  LG R+++   L
Sbjct  609  ATLFLPVGTIPERPSWS-RVPSPQGADVRVGRSPRPGSSVPEHSPRTTGLGCRLHSAPNL  667

Query  75   TDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIG-NQVFHLPNN  133
            +D          +V R +  +     + +       +A  P  G            LP+
Sbjct  668  SD---------FHVVRPKLPKPPTDPLGATFSPPQTSAPQPCPGLQSCRPLRGSPKLPDF  718

Query  134  LSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNY  168
```

```
            L    L    GS   AG +     +     +
Sbjct  719  LQRSPLPPILGSPTKAGPSFDFPKTPSSQNLLTLL  753


>☐gb|AAH59835.1| G Ulk1 protein [Mus musculus]
Length=1057

 GENE ID: 22241 Ulk1 | Unc-51 like kinase 1 (C. elegans) [Mus musculus]
(Over 10 PubMed links)

 Score = 43.9 bits (101),  Expect = 0.034, Method: Composition-based stats.
 Identities = 24/155 (15%), Positives = 45/155 (29%), Gaps = 11/155 (7%)

Query  15   AIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTAL  74
            A   +P        +       P+  +     S    ++  + S  T  LG R+++   L
Sbjct  507  ATLFLPVGTIPERPSWS-RVPSPQGADVRVGRSPRPGSSVPEHSPRTTGLGCRLHSAPNL  565

Query  75   TDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIG-NQVFHLPNN  133
            +D          +V R +  +     + +       +A  P  G            LP+
Sbjct  566  SD---------FHVVRPKLPKPPTDPLGATFSPPQTSAPQPCPGLQSCRPLRGSPKLPDF  616

Query  134  LSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNY  168
            L    L    GS   AG +     +     +
Sbjct  617  LQRSPLPPILGSPTKAGPSFDFPKTPSSQNLLTLL  651


>☐dbj|BAE28476.1| G unnamed protein product [Mus musculus]
Length=1057

 GENE ID: 22241 Ulk1 | Unc-51 like kinase 1 (C. elegans) [Mus musculus]
(Over 10 PubMed links)

 Score = 43.9 bits (101),  Expect = 0.036, Method: Composition-based stats.
 Identities = 24/155 (15%), Positives = 45/155 (29%), Gaps = 11/155 (7%)

Query  15   AIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTAL  74
            A   +P        +       P+  +     S    ++  + S  T  LG R+++   L
Sbjct  507  ATLFLPVGTIPERPSWS-RVPSPQGADVRVGRSPRPGSSVPEHSPRTTGLGCRLHSAPNL  565

Query  75   TDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIG-NQVFHLPNN  133
            +D          +V R +  +     + +       +A  P  G            LP+
Sbjct  566  SD---------FHVVRPKLPKPPTDPLGATFSPPQTSAPQPCPGLQSCRPLRGSPKLPDF  616

Query  134  LSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNY  168
            L    L    GS   AG +     +     +
Sbjct  617  LQRSPLPPILGSPTKAGPSFDFPKTPSSQNLLTLL  651


>☐ref|ZP_04248588.1|  Prophage LambdaBa01, membrane protein [Bacillus cereus Rock1-3]
 gb|EEL19689.1|  Prophage LambdaBa01, membrane protein [Bacillus cereus Rock1-3]
Length=1216

 Score = 43.9 bits (101),  Expect = 0.037, Method: Composition-based stats.
 Identities = 32/392 (8%), Positives = 93/392 (23%), Gaps = 55/392 (14%)

Query  21    WYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGTN  80
             W          W+     +        + S +T +  +++   L    N  T       N
Sbjct  644   WNSIKEFLIGIWD---GISQWATQTWESISESTASVWNSIKEFLVELWNGITESLSETWN  700

Query  81    VNGLGINVNRRRRRRRDIWNIESNPVLAGLA---ANNPVNGSYIAIGNQVFHLPNNLSNV  137
                                  +             +   +  + AI        N L
Sbjct  701   SIVETTTETWNSIVEYLTGIWDGVVETLSEVWNGISQTTSEVWTAISEFFISTWNGLVAF  760

Query  138   SLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAM--DSIVQTLS  195
                I  G +                 I       +  +    Q +    +   + +
Sbjct  761   LTPILQGIADFFSM--------IWNGISTVIQTVWNFITQYLQAVWTAILYFATPIFESI  812

Query  196   DAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGI-------QSIVDQGAQGFEQVIAS--  246
              ++  +     + A     N  +   ++     +       +++ +     ++ + ++
Sbjct  813   KSFIVSVWDAISLAVTTVWNAIVAFLQACWNGIVSIATAVFETLRNWIVNVWDVISSTTM  872

Query  247   --FNDSSARVQQC----VGPDLSPTSIGR---------------RVVGKGTDCVNRKWRQ  285
               +N     +Q C    V          +                V       ++  W
Sbjct  873   TVWNTVKNFLQACWNGLVAIVTPIFDAIKNWIVNAWNTISSTTSAVWNTIKGFLSSLWNS  932

Query  286   LSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRAIG  345
             +      +  +I  A     +    +++ + + +N   ++                 ++
Sbjct  933   IVSTASSVFNNIKEAISTVWNM---ISSTSSSIWNGIKSTLSNIWEGIKST----ASSVW  985

Query  346   NFPQSLLFLPVSLTIEGGKLYAALSSLQTDVA  377
             N  +  +  PV        +  A   +++ V
Sbjct  986   NGLKDAIMTPVRWV--TSAVSGAFEGMKSAVL  1015


>☐gb|EFB27064.1|  hypothetical protein PANDA_003089 [Ailuropoda melanoleuca]
Length=4306

 Score = 43.5 bits (100),  Expect = 0.043, Method: Composition-based stats.
 Identities = 33/271 (12%), Positives = 85/271 (31%), Gaps = 33/271 (12%)

Query  17    GIIPWYVDAFLFTTPWNPQPPRTINQ------------------LPVTSATSATTPTQQS  58
                 +    F  T P+       I                       +   S  + +  S
Sbjct  1450  LPTTYSYKPFFSTRPYQSWTTAPITVPGPAKSGFTSLSSSSSNTPSASPLKSIWSVSTPS  1509

Query  59    NVTLALGNRINTD-TALTDYGTNVNG-LGINVNRRRRRRRDIWNIESNPVL----AGLAA  112
              +   LG    +   +++D  + +     I+   +    +  +NI+ +         +
Sbjct  1510  PIKSTLGASTTSSVKSISDVASPIRSFRTISSPIKTVVSQSPYNIQVSSGTLGRAPTVPE  1569

Query  113   NNP-----VNGSYIAIGNQVFHLPNNLSNVSLN-IASGSSGGAGQTVPS--NFSTFLPMI  164
              +P      N ++ +  + V    + L   S+      S         S   F + +
Sbjct  1570  ASPLKGLASNSTFSSRTSPVTTAGSLLERSSITMTPPASPKSNINMYSSSLPFKSIITSA  1629

Query  165   PDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSV  224
                 +   + +V+  ++  +D + S   T++ + S+  K +   A +  +N ++   +
Sbjct  1630  APLISSPLKSVVSPAKS-AVDVISSAKVTMASSLSSPVKQMPGHAEVALVNGSISPLKYP  1688

Query  225   IEDGIQSIVDQGAQGFEQVIASFNDSSARVQ  255
                 + +     A   E++  + N  S+ V
Sbjct  1689  SSSTLINGCKATATLQEKISTATNSVSSVVN  1719
```

```
>☐ref|ZP_04223830.1|  Prophage LambdaBa01, membrane protein [Bacillus cereus Rock3-42]
 gb|EEL44486.1|  Prophage LambdaBa01, membrane protein [Bacillus cereus Rock3-42]
Length=1211

 Score = 43.5 bits (100),  Expect = 0.051, Method: Composition-based stats.
 Identities = 22/219 (10%), Positives = 61/219 (27%), Gaps = 25/219 (11%)

Query  161   LPMIPDNYTQAFQKLVTQFQTLPLDAMDSI--VQTLSDAYSNTFKGLFTAAGLNGLNETL  218
                I    +  ++ +     +     +  +  +     ++  T     +   +N  N
Sbjct  814   WNTISSTTSFVWETIKNFLVSCWNGLVAFVLPIFEQIKSWIITVWNTISTTTVNVWNAIK  873

Query  219   EHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDC  278
                +S     +  +V         +I ++N  S+                  V    T+
Sbjct  874   NFLQSCWNGIVAFVVPIFNSIKGWIINTWNTISSTT--------------SAVWNAVTNY  919

Query  279   VNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQ  338
             ++  W  +  I   I  +I  A     +    +++ + + +N   ++
Sbjct  920   LSSLWNSIVSIARAIFNNIKEAISTVWNM---ISSASVSVWNGIKSTLASIWEGIKYTAA  976

Query  339   CYVRAIGNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVA  377
                  + N     +  PV        +  A   +Q+ +
Sbjct  977   SIWSGLTN----AIMTPVKWVTRA--VIGAFQGMQSAIL  1009


>☐gb|AAA64834.1| G ankyrin G [Homo sapiens]
Length=4377

 GENE ID: 288 ANK3 | ankyrin 3, node of Ranvier (ankyrin G) [Homo sapiens]
(Over 10 PubMed links)

 Score = 43.5 bits (100),  Expect = 0.051, Method: Composition-based stats.
 Identities = 31/270 (11%), Positives = 78/270 (28%), Gaps = 33/270 (12%)

Query  17    GIIPWYVDAFLFTTPWNPQPPRTINQ------------------LPVTSATSATTPTQQS  58
                 +    F  T P+       I                       +   S  + +  S
Sbjct  1488  LPTTYSYKPFFSTRPYQSWTTAPITVPGPAKSGFTSLSSSSSNTPSASPLKSIWSVSTPS  1547

Query  59    NVTLALGNRINTD-TALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPV-  116
              +   LG    +   +++D  + +  L    +  +                   A
Sbjct  1548  PIKSTLGASTTSSVKSISDVASPIRSLRTMSSPIKTVVSQSPYNIQVSSGTLARAPAVTE  1607

Query  117   ---------NGSYIAIGNQVFHLPNNLSNVSLN-IASGSSGGAGQTVPS--NFSTFLPMI  164
                      N ++ +  + V    + L   S+      S         S   F + +
Sbjct  1608  ATPLKGLASNSTFSSRTSPVTTAGSLLERSSITMTPPASPKSNINMYSSSLPFKSIITSA  1667

Query  165   PDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSV  224
                 +   + +V+  ++  +D + S   T++ + S+  K +   A +  +N ++   +
Sbjct  1668  APLISSPLKSVVSPVKSR-VDVISSAKITMASSLSSPVKQMPGHAEVALVNGSISPLKYA  1726

Query  225   IEDGIQSIVDQGAQGFEQVIASFNDSSARV  254
                 + +     A   E++ ++ N  S+ V
Sbjct  1727  SSSTLINGCKATATLQEKISSATNSVSSVV  1756


>☐ref|ZP_05354696.1|  TP901 family phage tail tape measure protein [Clostridium difficile
QCD-76w55]
 ref|ZP_05386812.1|  TP901 family phage tail tape measure protein [Clostridium difficile
QCD-97b34]
Length=1129

 Score = 43.2 bits (99),  Expect = 0.059, Method: Composition-based stats.
 Identities = 37/338 (10%), Positives = 90/338 (26%), Gaps = 20/338 (5%)

Query  71   DTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIGNQVFHL  130
             T  T +   V+ +   +N        ++N+     +    +   +NG        V  +
Sbjct  579  STLATAFKAVVDTITNVLNGLASILNGVFNL--IIGIVTGDSKKMMNGVKQIFDGGVKII  636

Query  131  PNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAMDSI  190
             + +    ++  +           S F      + D +      L    +     +  +
Sbjct  637  KS-IWKGFIDFVTAPIQAIADIADSLFKDKAKWLTDKWKDLRTFLENPIEATAKISTKTN  695

Query  191  VQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDS  250
            V       ++            GL +T+ + +    D    I        + V +
Sbjct  696  VNIPGSKAADAAGKS--TPNTGGLKQTIANLKGSFSDFKTFIKPMTDSISQGVESGKTKF  753

Query  251  SARVQQCVGPDLSPTSI-------GRRVVGKGTDCVNRKWRQLSDI-GGRIAGDILAADR  302
            +    + +        +        +    +    +  K+ +L +I  G I  +I    +
Sbjct  754  TELQNKILNDVKPALKVLEDSGVNFKDKFKESITSIKDKFNELKNILAGNIKTNIDNLVK  813

Query  303  GASDF---LGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLFLPVSLT  359
                    L NL  C      A                +   + + +  +S   + +S
Sbjct  814  AFEPLKPHLDNLKECFDKVKKAITDFFTPAKQVGDTVNK-VSQPMDSVKKSTDGIKLSFQ  872

Query  360  IEGGKLYA---ALSSLQTDVALCASEMGLEIGLVTAQV  394
                 L             ++    ++   IG+  A V
Sbjct  873  SLTQALAPVKAWFQGFMPFLSQIGQQLLPSIGIAIAGV  910


>☐ref|YP_002749202.1| G TMP repeat-containing protein [Bacillus cereus 03BB102]
 gb|ACO30669.1| G TMP repeat-containing protein [Bacillus cereus 03BB102]
Length=1211

 GENE ID: 7688922 BCA_1927 | TMP repeat-containing protein
[Bacillus cereus 03BB102]

 Score = 43.2 bits (99),  Expect = 0.062, Method: Composition-based stats.
 Identities = 22/219 (10%), Positives = 62/219 (28%), Gaps = 25/219 (11%)

Query  161   LPMIPDNYTQAFQKLVTQFQTLPLDAMDSI--VQTLSDAYSNTFKGLFTAAGLNGLNETL  218
                I    +  ++ +     +     +  +  +     ++  T     +   +N  N
Sbjct  814   WNTISSTTSFVWETIKNFLVSCWNGLVAFVLPIFEQIKSWIITVWNTISTTTVNVWNAIK  873

Query  219   EHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDC  278
                +S     +  +V         +I ++N  S+                  V    T+
Sbjct  874   NFLQSCWNGIVAFVVPIFNSIKGWIINTWNTISSTT--------------SAVWNAVTNY  919

Query  279   VNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQ  338
             ++  W  +  I   I  +I  A     +    +++ + + +N   ++
Sbjct  920   LSSLWNSIVSIARAIFNNIKEAISTVWNM---ISSASVSVWNGIKSTLASIWEGIKYTAA  976

Query  339   CYVRAIGNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVA  377
```

```
                  + N     +  PV+       +  A   +Q+ +
Sbjct  977   SIWSGLTN----AIMTPVNWVTRA--VIGAFQGMQSAIL  1009


>☐ref|ZP_04106162.1|  Uncharacterized phage protein [Bacillus thuringiensis serovar
berliner ATCC 10792]
 ref|ZP_04137112.1|  Uncharacterized phage protein [Bacillus thuringiensis serovar
thuringiensis str. T01001]
 ref|ZP_04142613.1|  Uncharacterized phage protein [Bacillus thuringiensis Bt407]
 gb|EEM25676.1|  Uncharacterized phage protein [Bacillus thuringiensis Bt407]
 gb|EEM31178.1|  Uncharacterized phage protein [Bacillus thuringiensis serovar
thuringiensis str. T01001]
 gb|EEM62141.1|  Uncharacterized phage protein [Bacillus thuringiensis serovar
berliner ATCC 10792]
Length=1251

 Score = 41.6 bits (95),  Expect = 0.20, Method: Composition-based stats.
 Identities = 31/231 (13%), Positives = 64/231 (27%), Gaps = 23/231 (9%)

Query  34    PQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGTN-----VNGLGINV  88
                  T          + T       + L +    N    +T    N     ++ +   +
Sbjct  893   AWNWITNIIREAWQWITDTVNEFLGPIKLKIQEGWNYVLNITSKAWNAVSTFISDVWNEI  952

Query  89    NRRRRRRRDIWNIESNPVLAGL--AANNPVNGSYIAIGNQVFHLPNNLSNVSLNIASGSS  146
                 R + +                 +N  N     I      + +N+SN  +N
Sbjct  953   VGYIRNKINKATDTVTSGWTSTKNTTSNKWNEITSTISRVWNTITSNISNA-VNEVKSRI  1011

Query  147   GGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQ---TLPLDAMDSIVQTLSDAYSNTFK  203
              GA   V +  S     +  N T A+ +L +          + + S    +SD   + +K
Sbjct  1012  EGAWNVVYAITSGIWNNVVSNVTGAWDRLSSGVSGTFNYISNIVRSNWSYVSDILQSVYK  1071

Query  204   GLFTAAGLNGLNETLEHGRSVIEDGI---QSIVDQGAQGFEQVIASFNDSS  251
                        +N  +     +    I     I+D   + + +     N  +
Sbjct  1072  ---------PINSVVGFFYDLWNSIIRVMNGIIDTVTRAWNKAGGILNKLN  1113


>☐ref|ZP_04251409.1|  Prophage LambdaBa01, membrane protein [Bacillus cereus 95/8201]
 gb|EEL16945.1|  Prophage LambdaBa01, membrane protein [Bacillus cereus 95/8201]
Length=1210

 Score = 41.2 bits (94),  Expect = 0.21, Method: Composition-based stats.
 Identities = 26/242 (10%), Positives = 65/242 (26%), Gaps = 25/242 (10%)

Query  161   LPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQT--LSDAYSNTFKGLFTAAGLNGLNETL  218
                 I    +  ++ +     +     +  +V       ++        +   +N  N
Sbjct  814   WNTISSTTSFVWETIKNFLVSCWNGLVAFVVPIFEQIKSWIIAVWNTISTTTVNVWNAIK  873

Query  219   EHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDC  278
                 +S     +  +V         +I ++N  S+                  V    T+
Sbjct  874   NFLQSCWNGIVAFVVPIFNSIKGWIINTWNTISSTTST--------------VWNTVTNY  919

Query  279   VNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQ  338
              ++  W  +  +   I  +I  A     + + + +        +  AS        A
Sbjct  920   LSSLWNSIVSMARAIFNNIKEAISTVWNMISSASVSVWNGIKSTLASIWDGIRYTAA---  976

Query  339   CYVRAIGNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVALCASEMGLEIGLVTAQVSSKI  398
                 +I +     +  PV+       +  A   +Q+ +    S +   I      +   I
Sbjct  977   ----SIWSGLTDAIMTPVNWVTRA--VIGAFQGMQSAILAVWSGIKSGIKAAINGLIYMI  1030

Query  399   AF  400

Sbjct  1031  NK  1032


>☐dbj|BAE32939.1| G unnamed protein product [Mus musculus]
Length=1051

 GENE ID: 22241 Ulk1 | Unc-51 like kinase 1 (C. elegans) [Mus musculus]
(Over 10 PubMed links)

 Score = 41.2 bits (94),  Expect = 0.23, Method: Composition-based stats.
 Identities = 22/148 (14%), Positives = 42/148 (28%), Gaps = 11/148 (7%)

Query  22   YVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGTNV  81
                   +       P+  +     S    ++  + S  T  LG R+++   L+D
Sbjct  508  GTIPERPSWS-RVPSPQGADVRVGRSPRPGSSVPEHSPRTTGLGCRLHSAPNLSD-----  561

Query  82   NGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIG-NQVFHLPNNLSNVSLN  140
                 +V R +  +     + +       +A  P  G            LP+ L    L
Sbjct  562  ----FHVVRPKLPKPPTDPLGATFSPPQTSAPQPCPGLQSCRPLRGSPKLPDFLQRSPLP  617

Query  141  IASGSSGGAGQTVPSNFSTFLPMIPDNY  168
               GS   AG +     +     +
Sbjct  618  PILGSPTKAGPSFDFPKTPSSQNLLTLL  645


>☐sp|O70405.1|ULK1_MOUSE G RecName: Full=Serine/threonine-protein kinase ULK1; AltName:
Full=Unc-51-like kinase 1; AltName: Full=Serine/threonine-protein
kinase Unc51.1
 gb|AAF23317.1|AF072370_1 G UNC51.1 serine/threonine kinase [Mus musculus]
 gb|AAC40118.1| G UNC-51-like kinase ULK1 [Mus musculus]
Length=1051

 GENE ID: 22241 Ulk1 | Unc-51 like kinase 1 (C. elegans) [Mus musculus]
(Over 10 PubMed links)

 Score = 41.2 bits (94),  Expect = 0.23, Method: Composition-based stats.
 Identities = 22/148 (14%), Positives = 42/148 (28%), Gaps = 11/148 (7%)

Query  22   YVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGTNV  81
                   +       P+  +     S    ++  + S  T  LG R+++   L+D
Sbjct  508  GTIPERPSWS-RVPSPQGADVRVGRSPRPGSSVPEHSPRTTGLGCRLHSAPNLSD-----  561

Query  82   NGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIG-NQVFHLPNNLSNVSLN  140
                 +V R +  +     + +       +A  P  G            LP+ L    L
Sbjct  562  ----FHVVRPKLPKPPTDPLGATFSPPQTSAPQPCPGLQSCRPLRGSPKLPDFLQRSPLP  617

Query  141  IASGSSGGAGQTVPSNFSTFLPMIPDNY  168
               GS   AG +     +     +
Sbjct  618  PILGSPTKAGPSFDFPKTPSSQNLLTLL  645
```

```
>☐ref|NP_033495.2| serine/threonine-protein kinase ULK1 [Mus musculus]
 gb|AAH57121.1| Unc-51 like kinase 1 (C. elegans) [Mus musculus]
 gb|EDL20002.1| Unc-51 like kinase 1 (C. elegans), isoform CRA_b [Mus musculus]
Length=1051

 GENE ID: 22241 Ulk1 | Unc-51 like kinase 1 (C. elegans) [Mus musculus]
(Over 10 PubMed links)

 Score = 41.2 bits (94),  Expect = 0.24, Method: Composition-based stats.
 Identities = 22/148 (14%), Positives = 42/148 (28%), Gaps = 11/148 (7%)

Query  22   YVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGTNV  81
                   +       P+  +     S    ++  + S  T  LG R+++   L+D
Sbjct  508  GTIPERPSWS-RVPSPQGADVRVGRSPRPGSSVPEHSPRTTGLGCRLHSAPNLSD-----  561

Query  82   NGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIG-NQVFHLPNNLSNVSLN  140
                 +V R +  +     + +       +A  P  G            LP+ L    L
Sbjct  562  ----FHVVRPKLPKPPTDPLGATFSPPQTSAPQPCPGLQSCRPLRGSPKLPDFLQRSPLP  617

Query  141  IASGSSGGAGQTVPSNFSTFLPMIPDNY  168
               GS   AG +     +     +
Sbjct  618  PILGSPTKAGPSFDFPKTPSSQNLLTLL  645


>☐gb|AAC78143.1| 270 kDa ankyrin G isoform [Rattus norvegicus]
Length=2622

 GENE ID: 361833 Ank3 | ankyrin 3, epithelial [Rattus norvegicus]
(Over 10 PubMed links)

 Score = 40.8 bits (93),  Expect = 0.27, Method: Composition-based stats.
 Identities = 30/271 (11%), Positives = 79/271 (29%), Gaps = 31/271 (11%)

Query  14    AAIGIIPWYVDAFLFTTPWNPQPPR----------------TINQLPVTSATSATTPTQQ  57
             A      +    F  T P+                        N    +   S  + +
Sbjct  1486  ARSLPTTYSHKPFFSTRPYQSWTTTPITVPGPAKSGSLSSSPSNTPSASPLKSIWSVSTP  1545

Query  58    SNVTLALGNRINTD-TALTDYGTNVNG-LGINVNRRRRRRRDIWNIESNPVL----AGLA  111
             S +   LG    +   +++D  + +     I+   R       +N +           +
Sbjct  1546  SPIKSTLGASTTSSVKSISDVASPIRSFRTISSPIRTVASPSPYNTQVASGTLGRVPTIT  1605

Query  112   ANNPV-----NGSYIAIGNQVFHLPNNLSNVSLN-IASGSSGG--AGQTVPSNFSTFLPM  163
                P+     N +  +  + V    + L   S+      S        +    F + +
Sbjct  1606  EATPIKGVAPNSTLSSRTSPVTTAGSLLEKSSITMTPPASPKANITMYSSSLPFKSIITS  1665

Query  164   IPDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRS  223
                  +   + +V+  ++   D + +    ++   S+  K +   A +  +N ++   +
Sbjct  1666  AAPLISSPLKSVVSPTKSAA-DVISTAKAAMASTLSSPLKQMSGHAEVALVNGSVSPLKY  1724

Query  224   VIEDGIQSIVDQGAQGFEQVIASFNDSSARV  254
                  + +     A   +++  + N  S+ V
Sbjct  1725  PSSSALINGCKATATLQDKISTATNAVSSVV  1755


>☐ref|NP_001029156.1| ankyrin 3, epithelial isoform 2 [Rattus norvegicus]
Length=2622

 GENE ID: 361833 Ank3 | ankyrin 3, epithelial [Rattus norvegicus]
(Over 10 PubMed links)

 Score = 40.8 bits (93),  Expect = 0.27, Method: Composition-based stats.
 Identities = 30/271 (11%), Positives = 79/271 (29%), Gaps = 31/271 (11%)

Query  14    AAIGIIPWYVDAFLFTTPWNPQPPR----------------TINQLPVTSATSATTPTQQ  57
             A      +    F  T P+                        N    +   S  + +
Sbjct  1486  ARSLPTTYSHKPFFSTRPYQSWTTTPITVPGPAKSGSLSSSPSNTPSASPLKSIWSVSTP  1545

Query  58    SNVTLALGNRINTD-TALTDYGTNVNG-LGINVNRRRRRRRDIWNIESNPVL----AGLA  111
             S +   LG    +   +++D  + +     I+   R       +N +           +
Sbjct  1546  SPIKSTLGASTTSSVKSISDVASPIRSFRTISSPIRTVASPSPYNTQVASGTLGRVPTIT  1605

Query  112   ANNPV-----NGSYIAIGNQVFHLPNNLSNVSLN-IASGSSGG--AGQTVPSNFSTFLPM  163
                P+     N +  +  + V    + L   S+      S        +    F + +
Sbjct  1606  EATPIKGVAPNSTLSSRTSPVTTAGSLLEKSSITMTPPASPKANITMYSSSLPFKSIITS  1665

Query  164   IPDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRS  223
                  +   + +V+  ++   D + +    ++   S+  K +   A +  +N ++   +
Sbjct  1666  AAPLISSPLKSVVSPTKSAA-DVISTAKAAMASTLSSPLKQMSGHAEVALVNGSVSPLKY  1724

Query  224   VIEDGIQSIVDQGAQGFEQVIASFNDSSARV  254
                  + +     A   +++  + N  S+ V
Sbjct  1725  PSSSALINGCKATATLQDKISTATNAVSSVV  1755


>☐ref|XP_002619294.1| hypothetical protein CLUG_00453 [Clavispora lusitaniae ATCC 42720]
 gb|EEQ36330.1| hypothetical protein CLUG_00453 [Clavispora lusitaniae ATCC 42720]
Length=1855

 GENE ID: 8499645 CLUG_00453 | hypothetical protein
[Clavispora lusitaniae ATCC 42720]

 Score = 40.8 bits (93),  Expect = 0.28, Method: Composition-based stats.
 Identities = 22/221 (9%), Positives = 58/221 (26%), Gaps = 12/221 (5%)

Query  32   WNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGTNVNGLGINVNRR  91
             + +  +  N   + S  +        N      +++ +   L      +  L   +
Sbjct  756  LDIRYNKITNVDVLGSLPNLEVLYASKNNIAGFSDKMESLRLLHFDRNPITNLEFQITLP  815

Query  92   RRRRRDIWNIESNPVLAGLAANNPVNGSYIAIGNQVFHLPNNLSNVS--------LNIAS  143
            R    D+   +   +     +  P     +   N +  LP+ LS +          N
Sbjct  816  RLTVLDLSKAKITALPGEFISKMPHIEKLVLDKNHLVTLPDELSALPRLTHLSLYANNLQ  875

Query  144  GSSGGAGQTVPSNFSTFLPMIPDNY-TQAFQKLVTQFQTLPLDAMD--SIVQTLSDAYSN  200
                  GQ V   +        +   +  +         +  + +           +  +
Sbjct  876  HLPDSIGQLVSLQYLDLHSNNIETLPSSIWNLKSLSTLNVSSNMLSSFPSPSLELASKMS  935

Query  201  TFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFE  241
```

```
            + K    +AG +  N+ +    +       S++
Sbjct  936  SPKTGNDSAGKSNQNQVMAS-SAQFNSLADSLLSLTLADNR  975


>☐ref|XP_002807495.1|  PREDICTED: LOW QUALITY PROTEIN: ankyrin-3-like [Callithrix jacchus]
Length=4392

 Score = 40.8 bits (93),  Expect = 0.30, Method: Composition-based stats.
 Identities = 31/271 (11%), Positives = 76/271 (28%), Gaps = 33/271 (12%)

Query  17    GIIPWYVDAFLFTTPWNPQPPRTINQ------------------LPVTSATSATTPTQQS  58
                 +    F  T P+       I                       +   S  + +  S
Sbjct  1488  LPTTYSYKPFFSTRPYQSWTTAPITVPGPAKSGFTSLSSSSNTPSASPLKSIWSVSTPS  1547

Query  59    NVTLALGNRINTD-TALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPV-  116
              +   LG    +   +++D  + +       +  +                   A
Sbjct  1548  PIKSTLGASTTSSVKSISDVASPIRSFRTMSSPIKTVVSQSPYNIQVSSGTLARAPAVTE  1607

Query  117   ---------NGSYIAIGNQVFHLPNNLSNVSLN-IASGSSGGAGQTVPS--NFSTFLPMI  164
                       N ++ +  + V    + L   S+      S         S   F + +
Sbjct  1608  ATPLKGLASNSTFSSRTSPVTTAGSLLERSSITMTPPASPKSNINMYSSSLPFKSIITSA  1667

Query  165   PDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSV  224
                 +   + +V+  ++  +DA+ S   T++ + S+  K +   A +  +N ++   +
Sbjct  1668  APLMSSPLKSVVSPVKS-AVDAISSAKVTMASSLSSPVKQMPGHAEVALVNGSISPLKYP  1726

Query  225   IEDGIQSIVDQGAQGFEQVIASFNDSSARVQ  255
                 + +     A   E +  + N  S+ V
Sbjct  1727  SSSTLINGCKATATLQETISTATNSVSSVVN  1757


>☐ref|NP_001101811.1| UG serine/threonine-protein kinase ULK1 [Rattus norvegicus]
 gb|EDM14029.1| G unc-51-like kinase 1 (mapped) [Rattus norvegicus]
Length=1051

 GENE ID: 360827 Ulk1 | Unc-51 like kinase 1 (C. elegans) [Rattus norvegicus]
(10 or fewer PubMed links)

 Score = 40.8 bits (93),  Expect = 0.31, Method: Composition-based stats.
 Identities = 25/191 (13%), Positives = 49/191 (25%), Gaps = 14/191 (7%)

Query  22   YVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGTNV  81
                   +       P+  +     S    ++  + S  T  LG R+++   L+D
Sbjct  508  GTIPERPSWS-RVPSPQGADVRAGRSPRPGSSVPEHSPRTTGLGCRLHSAPNLSD-----  561

Query  82   NGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIG-NQVFHLPNNLSNVSLN  140
                 +V R +  +     + +       +A  P  G            LP+ L    L
Sbjct  562  ----FHVVRPKLPKPPTDPLGATFSPPQTSAPQPCPGLQSCRPLRGSPKLPDFLQRSPLP  617

Query  141  IASGSSGGAGQTVPSNFSTFLPMIPDNY-TQAFQKLVTQFQTLPLDAMDSIVQTLSDAYS  199
               GS   AG +     +     +      Q         +   L  +
Sbjct  618  PILGSPTKAGPSFDFPKAPSSQSLLTLLARQGVVMTPP--RNRTLPDLSEAGPFQGQQLG  675

Query  200  NTFKGLFTAAG  210
            +  +      G
Sbjct  676  SGLRPAEDTRG  686


>☐ref|ZP_05332071.1|  phage protein [Clostridium difficile QCD-63q42]
 ref|ZP_05399216.1|  phage protein [Clostridium difficile QCD-37x79]
Length=1130

 Score = 40.8 bits (93),  Expect = 0.34, Method: Composition-based stats.
 Identities = 30/270 (11%), Positives = 70/270 (25%), Gaps = 17/270 (6%)

Query  139  LNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAY  198
            ++  +           S F      + D +      L    +     +  + V
Sbjct  645  IDFVTAPIQAIADIADSLFKDKAKWLTDKWRDLRTFLENPIEATAKISAKTNVNIPGSKA  704

Query  199  SNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCV  258
            ++            GL +T+ + +    D    I        + V +  +  +    + +
Sbjct  705  ADAAGKS--TPNTGGLKQTVANLKGSFSDFKTFIKPMTDSISQGVESGKSKFTELQNKIL  762

Query  259  GPDLSPTSI-------GRRVVGKGTDCVNRKWRQLSDI-GGRIAGDILAADRGASDF---  307
                    +        +    +    +  K+ +L DI  G I  +I    +
Sbjct  763  NDVKPALKVLEDSGVNFKNKFKESITSIKDKFNELKDILAGNIKTNIDNLVKAFEPLKPH  822

Query  308  LGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLFLPVSLTIEGGKLYA  367
            L NL  C      A                +   + + +  +S   + +S       L
Sbjct  823  LDNLKECFDKVKKAITDFFTPAKQVGDTVNK-VSQPMDSVKKSTDGIKLSFQSLTQALAP  881

Query  368  ---ALSSLQTDVALCASEMGLEIGLVTAQV  394
                       ++    ++   IG+  A V
Sbjct  882  VKAWFQGFMPFLSQIGQQLLPSIGIAIAGV  911


>☐ref|ZP_04146935.1|  Prophage LambdaBa01, membrane protein [Bacillus thuringiensis
serovar tochigiensis BGSC 4Y1]
 gb|EEM21327.1|  Prophage LambdaBa01, membrane protein [Bacillus thuringiensis
serovar tochigiensis BGSC 4Y1]
Length=725

 Score = 40.5 bits (92),  Expect = 0.35, Method: Composition-based stats.
 Identities = 40/393 (10%), Positives = 95/393 (24%), Gaps = 55/393 (13%)

Query  20   PWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGT  79
             W          W+             ++ S +T +  +++   L    N    L
Sbjct  152  TWNAIGEFLVGIWDV---IVQWASDTWNSISESTSSIWNSIKEYLIELWNGIKDLLTETW  208

Query  80   N--VNGLGINVNRRRRRRRDIWN-IESNPVLAGLAANNPVNGSYIAIGNQVFHLPNNLSN  136
            N  V       N        +WN +     +A  +     +  + AI        + L
Sbjct  209  NTMVENTTEIWNSIVEYLTGVWNGVVETLSVAWDSIKQTTSEVWTAISEFFISTWSGLVA  268

Query  137  VSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAM--DSIVQTL  194
                +  G +                 I       +  +    Q +    +   + +
Sbjct  269  FITPVLQGIADFFSM--------IWNGISTVIQTVWNFITQYLQAIWTAILYFATPIFES  320

Query  195  SDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGI-------QSIVDQGAQGFEQVIAS-  246
               Y       ++               S     +       + I       +  +  +
Sbjct  321  IKNYITECWNTISSTTSFVWETIKNFLVSCWNGLVAFVVPIFEQIKSWIIAVWNTISTTT  380
```

```
Query  247  ---FNDSSARVQQC----VGPDLSPTSIGR---------------RVVGKGTDCVNRKWR  284
               +N   + +Q C    V   +   +  +                V    T+ ++  W
Sbjct  381  VNVWNAIRSFLQSCWNGLVAFVVPIFNAIKDWIVNNWKTISSITSAVWTAVTNYLSSLWN  440

Query  285  QLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRAI  344
             +  +   I  +I  A     + + + +        +  AS        A        +I
Sbjct  441  SIVSMARAIFNNIKEAISTVWNMISSASVSVWNGIKSTLASIWDGIRYTAA-------SI  493

Query  345  GNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVA  377
             +     +  PV+       +  A   +Q+ +
Sbjct  494  WSGLTDAIMTPVNWVTRA--VIGAFQGMQSAIL  524


>☐gb|EDL97288.1|  ankyrin 3, epithelial, isoform CRA_i [Rattus norvegicus]
 gb|EDL97293.1|  ankyrin 3, epithelial, isoform CRA_i [Rattus norvegicus]
Length=2616

 GENE ID: 361833 Ank3 | ankyrin 3, epithelial [Rattus norvegicus]
(Over 10 PubMed links)

 Score = 40.5 bits (92),  Expect = 0.36, Method: Composition-based stats.
 Identities = 32/264 (12%), Positives = 83/264 (31%), Gaps = 23/264 (8%)

Query  14    AAIGIIPWYVDAFLFTTPWNPQPPRTINQ--------LPVTSATSATTPTQQSNVTLALG  65
             A      +    F  T P+       +          L  + + + +    +S  +++
Sbjct  1486  ARSLPTTYSHKPFFSTRPYQSWTTTPLTVPGPAKSGSLSSSPSNTPSASPLKSIWSVSTP  1545

Query  66    NRINTD--TALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVN-GSYIA  122
             + I +    + T    +++ +   +   R     I  + S        A+  +     I
Sbjct  1546  SPIKSTLGASTTSSVKSISDVASPIRSFRTISSPIRTVASPSPYNTQVASGTLGRVPTIT  1605

Query  123   IGNQVFHLPNN--LSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNY--TQAFQKLVTQ  178
                 +  L  N  LS+ +  + +  S     ++              Y  +  F+ ++T
Sbjct  1606  EATPIKGLAPNSTLSSRTSPVTTAGSLLERSSITMTPPASPKANITMYSSSLPFKSIITS  1665

Query  179   FQTLPLDAMDSIVQ--------TLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQ  230
                L    + S+V          ++   S+  K +   A +  +N ++   +      +
Sbjct  1666  AAPLISSPLKSVVSPTKSAAKAAMASTLSSPLKQMSGHAEVALVNGSVSPLKYPSSSALI  1725

Query  231   SIVDQGAQGFEQVIASFNDSSARV  254
             +     A   +++  + N  S+ V
Sbjct  1726  NGCKATATLQDKISTATNAVSSVV  1749


>☐dbj|BAC65639.1|  mKIAA0722 protein [Mus musculus]
Length=1004

 GENE ID: 22241 Ulk1 | Unc-51 like kinase 1 (C. elegans) [Mus musculus]
(Over 10 PubMed links)

 Score = 40.5 bits (92),  Expect = 0.39, Method: Composition-based stats.
 Identities = 22/148 (14%), Positives = 42/148 (28%), Gaps = 11/148 (7%)

Query  22   YVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGTNV  81
                   +       P+  +     S    ++  + S  T  LG R+++   L+D
Sbjct  461  GTIPERPSWS-RVPSPQGADVRVGRSPRPGSSVPEHSPRTTGLGCRLHSAPNLSD-----  514

Query  82   NGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIG-NQVFHLPNNLSNVSLN  140
                 +V R +  +     + +       +A  P  G            LP+ L    L
Sbjct  515  ----FHVVRPKLPKPPTDPLGATFSPPQTSAPQPCPGLQSCRPLRGSPKLPDFLQRSPLP  570

Query  141  IASGSSGGAGQTVPSNFSTFLPMIPDNY  168
               GS   AG +     +     +
Sbjct  571  PILGSPTKAGPSFDFPKTPSSQNLLTLL  598


>☐emb|CAH73232.1|  ankyrin 3, node of Ranvier (ankyrin G) [Homo sapiens]
 emb|CAI40518.1|  ankyrin 3, node of Ranvier (ankyrin G) [Homo sapiens]
 emb|CAI41373.1|  ankyrin 3, node of Ranvier (ankyrin G) [Homo sapiens]
Length=4372

 GENE ID: 288 ANK3 | ankyrin 3, node of Ranvier (ankyrin G) [Homo sapiens]
(Over 10 PubMed links)

 Score = 40.5 bits (92),  Expect = 0.44, Method: Composition-based stats.
 Identities = 30/270 (11%), Positives = 77/270 (28%), Gaps = 33/270 (12%)

Query  17    GIIPWYVDAFLFTTPWNPQPPRTINQ------------------LPVTSATSATTPTQQS  58
                 +    F  T P+       I                       +   S  + +  S
Sbjct  1483  LPTTYSYKPFFSTRPYQSWTTAPITVPGPAKSGFTSLSSSSSNTPSASPLKSIWSVSTPS  1542

Query  59    NVTLALGNRINTD-TALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPV-  116
              +   LG    +   +++D  + +       +  +                   A
Sbjct  1543  PIKSTLGASTTSSVKSISDVASPIRSFRTMSSPIKTVVSQSPYNIQVSSGTLARAPAVTE  1602

Query  117   ---------NGSYIAIGNQVFHLPNNLSNVSLN-IASGSSGGAGQTVPS--NFSTFLPMI  164
                      N ++ +  + V    + L   S+      S         S   F + +
Sbjct  1603  ATPLKGLASNSTFSSRTSPVTTAGSLLERSSITMTPPASPKSNINMYSSSLPFKSIITSA  1662

Query  165   PDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSV  224
                 +   + +V+  ++  +D + S   T++ + S+  K +   A +  +N ++   +
Sbjct  1663  APLISSPLKSVVSPVKS-AVDVISSAKITMASSLSSPVKQMPGHAEVALVNGSISPLKYP  1721

Query  225   IEDGIQSIVDQGAQGFEQVIASFNDSSARV  254
                 + +     A   E++ ++ N  S+ V
Sbjct  1722  SSSTLINGCKATATLQEKISSATNSVSSVV  1751


>☐gb|EAW54200.1|  ankyrin 3, node of Ranvier (ankyrin G), isoform CRA_b [Homo sapiens]
Length=4232

 GENE ID: 288 ANK3 | ankyrin 3, node of Ranvier (ankyrin G) [Homo sapiens]
(Over 10 PubMed links)

 Score = 40.5 bits (92),  Expect = 0.44, Method: Composition-based stats.
 Identities = 30/270 (11%), Positives = 77/270 (28%), Gaps = 33/270 (12%)

Query  17    GIIPWYVDAFLFTTPWNPQPPRTINQ------------------LPVTSATSATTPTQQS  58
```

```
                 +    F  T P+       I                       +   S  + +  S
Sbjct  1343  LPTTYSYKPFFSTRPYQSWTTAPITVPGPAKSGFTSLSSSSSNTPSASPLKSIWSVSTPS  1402

Query  59    NVTLALGNRINTD-TALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPV-  116
              +   LG    +   +++D  + +       +  +                   A
Sbjct  1403  PIKSTLGASTTSSVKSISDVASPIRSFRTMSSPIKTVVSQSPYNIQVSSGTLARAPAVTE  1462

Query  117   ---------NGSYIAIGNQVFHLPNNLSNVSLN-IASGSSGGAGQTVPS--NFSTFLPMI  164
                      N ++ +  + V    + L   S+      S         S   F + +
Sbjct  1463  ATPLKGLASNSTFSSRTSPVTTAGSLLERSSITMTPPASPKSNINMYSSSLPFKSIITSA  1522

Query  165   PDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSV  224
                  +   + +V+  ++  +D + S   T++ + S+  K +   A +  +N ++   +
Sbjct  1523  APLISSPLKSVVSPVKS-AVDVISSAKITMASSLSSPVKQMPGHAEVALVNGSISPLKYP  1581

Query  225   IEDGIQSIVDQGAQGFEQVIASFNDSSARV  254
                  + +     A   E++ ++ N  S+ V
Sbjct  1582  SSSTLINGCKATATLQEKISSATNSVSSVV  1611


>☐ref|NP_066267.2| UG ankyrin-3 isoform 1 [Homo sapiens]
 sp|Q12955.3|ANK3_HUMAN G RecName: Full=Ankyrin-3; Short=ANK-3; AltName: Full=Ankyrin-G
 gb|EAW54201.1| G ankyrin 3, node of Ranvier (ankyrin G), isoform CRA_c [Homo sapiens]
Length=4377

 GENE ID: 288 ANK3 | ankyrin 3, node of Ranvier (ankyrin G) [Homo sapiens]
(Over 10 PubMed links)

 Score = 40.1 bits (91),  Expect = 0.46, Method: Composition-based stats.
 Identities = 30/270 (11%), Positives = 77/270 (28%), Gaps = 33/270 (12%)

Query  17    GIIPWYVDAFLFTTPWNPQPPRTINQ------------------LPVTSATSATTPTQQS  58
                 +    F  T P+       I                       +   S  + +  S
Sbjct  1488  LPTTYSYKPFFSTRPYQSWTTAPITVPGPAKSGFTSLSSSSSNTPSASPLKSIWSVSTPS  1547

Query  59    NVTLALGNRINTD-TALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPV-  116
              +   LG    +   +++D  + +       +  +                   A
Sbjct  1548  PIKSTLGASTTSSVKSISDVASPIRSFRTMSSPIKTVVSQSPYNIQVSSGTLARAPAVTE  1607

Query  117   ---------NGSYIAIGNQVFHLPNNLSNVSLN-IASGSSGGAGQTVPS--NFSTFLPMI  164
                      N ++ +  + V    + L   S+      S         S   F + +
Sbjct  1608  ATPLKGLASNSTFSSRTSPVTTAGSLLERSSITMTPPASPKSNINMYSSSLPFKSIITSA  1667

Query  165   PDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSV  224
                  +   + +V+  ++  +D + S   T++ + S+  K +   A +  +N ++   +
Sbjct  1668  APLISSPLKSVVSPVKS-AVDVISSAKITMASSLSSPVKQMPGHAEVALVNGSISPLKYP  1726

Query  225   IEDGIQSIVDQGAQGFEQVIASFNDSSARV  254
                  + +     A   E++ ++ N  S+ V
Sbjct  1727  SSSTLINGCKATATLQEKISSATNSVSSVV  1756


>☐dbj|BAE33746.1| G unnamed protein product [Mus musculus]
Length=1051

 GENE ID: 22241 Ulk1 | Unc-51 like kinase 1 (C. elegans) [Mus musculus]
(Over 10 PubMed links)

 Score = 40.1 bits (91),  Expect = 0.46, Method: Composition-based stats.
 Identities = 21/148 (14%), Positives = 41/148 (27%), Gaps = 11/148 (7%)

Query  22   YVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGTNV  81
                   +        +  +     S    ++  + S  T  LG R+++   L+D
Sbjct  508  GTIPERPSWS-RVPSQQGADVRVGRSPRPGSSVPEHSPRTTGLGCRLHSAPNLSD-----  561

Query  82   NGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIG-NQVFHLPNNLSNVSLN  140
                 +V R +  +     + +       +A  P  G            LP+ L    L
Sbjct  562  ----FHVVRPKLPKPPTDPLGATFSPPQTSAPQPCPGLQSCRPLRGSPKLPDFLQRSPLP  617

Query  141  IASGSSGGAGQTVPSNFSTFLPMIPDNY  168
               GS   AG +     +     +
Sbjct  618  PILGSPTKAGPSFDFPKTPSSQNLLTLL  645


>☐ref|XP_002612001.1| G hypothetical protein BRAFLDRAFT_124780 [Branchiostoma floridae]
 gb|EEN68010.1| G hypothetical protein BRAFLDRAFT_124780 [Branchiostoma floridae]
Length=1722

 GENE ID: 7245707 BRAFLDRAFT_124780 | hypothetical protein
[Branchiostoma floridae] (10 or fewer PubMed links)

 Score = 40.1 bits (91),  Expect = 0.46, Method: Composition-based stats.
 Identities = 26/161 (16%), Positives = 48/161 (29%), Gaps = 9/161 (5%)

Query  5    MSVVLALLIAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVT--------SATSATTPTQ  56
             + V+A+L+ A+            T   N   P                  +T+ +
Sbjct  829  RTAVVAVLLCALFGQQSSSTTPSPTPATNTTSPSVTESTTNNNNATPSVTESTTNSNNAT  888

Query  57   QSNVTLALGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPV  116
             S VT +  N  N   ++T+  TN N    +V+       +     S            V
Sbjct  889  PS-VTESTTNSNNATPSVTESTTNNNNATPSVSGSSTTNNNPTPSVSGSTTTNNNPTPSV  947

Query  117  NGSYIAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNF  157
            +GS     N    +  + +  +    S S       +P+
Sbjct  948  SGSTTTNNNPTPSVSGSTTTNNNPTPSVSGSTTTNNIPTPP  988


>☐ref|XP_001929357.1| UG PREDICTED: ankyrin 3, node of Ranvier (ankyrin G) [Sus scrofa]
Length=4346

 GENE ID: 100154687 ANK3 | ankyrin 3 [Sus scrofa] (10 or fewer PubMed links)

 Score = 40.1 bits (91),  Expect = 0.50, Method: Composition-based stats.
 Identities = 30/270 (11%), Positives = 85/270 (31%), Gaps = 31/270 (11%)

Query  17    GIIPWYVDAFLFTTPWNPQPPRTINQ----------LPVTSATSATTPTQQSNVTLALGN  66
                 +    F  T P+       I            L  +S+ + +    +S  +++  +
Sbjct  1458  LPTTYSYKPFFSTRPYQSWTTAPITVPGPAKSGFTSLSSSSSNTPSASPLKSIWSVSTPS  1517
```

```
Query  67    RINT---DTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAI  123
              I +    +  +   +  +      + R         +  +P    +++   V    +
Sbjct  1518  PIKSTLGASTTSSVKSISDVASPIRSFRTISSPIKTVVSQSPYNIQVSSGTLVRAPTVTE  1577

Query  124   GNQVFHLPNN--LSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNY--TQAFQKLVTQF  179
              + +  L +N   S  +  + +  S     ++            + Y  +  F+ ++T
Sbjct  1578  ASPLKGLASNSTFSPRTSPVTTAGSLLERSSITMTPPASPKSNINMYSSSLPFKSIITSA  1637

Query  180   QTLPLDAMDSIVQ--------------TLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVI  225
               L    + S+V               T++ + S+  K +   A +  +N ++   +
Sbjct  1638  APLISSPLKSVVSSAKPTVDVVSAAKVTMASSLSSPVKQMPGPAEVVLVNGSISPLKYPS  1697

Query  226   EDGIQSIVDQGAQGFEQVIASFNDSSARVQ  255
                 + +     A   E++  + N  S+ +
Sbjct  1698  SSTLMNGCKATATFQEKISTATNSVSSVLN  1727


>☐ref|YP_001987063.1| G Tail tape measure protein [Lactobacillus casei BL23]
 emb|CAQ66205.1| G Tail tape measure protein [Lactobacillus casei BL23]
Length=1056

 GENE ID: 6404742 tmp | Tail tape measure protein [Lactobacillus casei BL23]

 Score = 40.1 bits (91),  Expect = 0.50, Method: Composition-based stats.
 Identities = 27/281 (9%), Positives = 71/281 (25%), Gaps = 19/281 (6%)

Query  16   IGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALT  75
            + I  W          W+     T     + +          + +T  L         +
Sbjct  661  LVIKNWGAIVTWLKGVWS---TVTSFFKQLWNGIKQIFTIAINAITNFLKPAFTAAANVI  717

Query  76   DYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIGNQVFHLPNNLS  135
                N          +         I+    +A  A    +          +  + N +
Sbjct  718  KSIWN--------GIKSFFSALWNGIKVIFTVAITAIAVIIGTYLNIWKTIITTVMNLIK  769

Query  136  NVSLNIASGSSGGAGQTVPS---NFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQ  192
             +   + +G     G  + S      +    I    +  F ++ +   ++  +  + +
Sbjct  770  GIITKVWNGIKSFFGPILSSIGNVIRSAWNAISSVTSSVFNRVKSVVSSIWNNIKNVVSN  829

Query  193  TLSDAYS--NTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDS  250
             ++   S  +      ++   N  N       +V      +I +        V +++N
Sbjct  830  VVNAVKSVVSNAWNAVSSTTSNIFNSVKSAVSNVWNSIKSTISNVVGSIRNAVSSAWNAV  889

Query  251  SARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGG  291
            S+     V   +   +     +    D V      +
Sbjct  890  SSVTSN-VWDSI--KNAISGPINTAKDIVRGAIDAIRGFFN  927


>☐ref|ZP_04019473.1|  conserved hypothetical protein [Finegoldia magna ATCC 53516]
 gb|EEJ67754.1|  conserved hypothetical protein [Finegoldia magna ATCC 53516]
Length=498

 Score = 40.1 bits (91),  Expect = 0.51, Method: Composition-based stats.
 Identities = 33/356 (9%), Positives = 86/356 (24%), Gaps = 28/356 (7%)

Query  21   WYVDAFLFTTPWNP--------QPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDT  72
            W   +  + + W                    +  +      +  + +     N     +
Sbjct  117  WNGLSEGWNSIWTFITTFLTESWNTFIEGAKSLWQSLGEFFTSLWTGIQTTFTNIWTAIS  176

Query  73   ALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIGNQVFHLPN  132
             +T       G  I              I +      + A    NG    +   +  +
Sbjct  177  TITTEVFTAVGEFIRTTW----EGIKTLISTVLDAIKVKAETIWNGIKTFLTTVITAIGT  232

Query  133  NLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQ  192
             +S  S N           T+ S        I    +     ++T    +  +++ S +
Sbjct  233  FIS-TSWNNIKVVIQTVLNTIKSIVMNVWNAIKSFISGVLNGIMTLVSNVW-NSIKSTIS  290

Query  193  TLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIED---GIQSIVDQGAQGFEQVIASFND  249
            +  ++  +     F +      +       S+       +  I +  +Q +       N
Sbjct  291  STVNSAKSAVTSAFNSMKSAISSSMSNILYSIRNGFNNAVNFIRNLASQAYTWGADMING  350

Query  250  SSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILAADRGASDFLG  309
             +  +   +G  +                V    R          G +   +    DF+
Sbjct  351  IARGISNAIGNVI-----------SAVSNVASTIRSYLHFSVPDVGPLTDYESWMPDFME  399

Query  310  NLTTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLFLPVSLTIEGGKL  365
             L+     +     +S     +   L        +G + +      + +  +
Sbjct  400  GLSKGIEKSRRLVQSSMKNVASDMVLNPNISAVGMGGYDKESAVNGIDIGRQISDA  455


>☐gb|EFA12920.1|  hypothetical protein TcasGA2_TC006990 [Tribolium castaneum]
Length=233

 Score = 40.1 bits (91),  Expect = 0.53, Method: Composition-based stats.
 Identities = 19/150 (12%), Positives = 40/150 (26%), Gaps = 20/150 (13%)

Query  253  RVQQCVGPDLSPTSIGRRV-VGKGTDCVNRKWRQLSDIGGRIAGDILAADRGASDFLGNL  311
             V  C        +  R +   +   C+N K ++ + +      ++            ++
Sbjct  93   DVSSCTKGQKDVINKFRELFENRVQVCINGKIQEAASLLKGSKYNVDVTMNSVHRLDHDV  152

Query  312  TTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLFLPVSLTIEGGKLYAALSS  371
              C  +                     C    +     + + LP  +  E       L
Sbjct  153  NVCGDSI-------------------LCISPILTKINLATISLPQQIKTEVNNARDLLQD  193

Query  372  LQTDVALCASEMGLEIGLVTAQVSSKIAFC  401
            L+  VA C      +       + + IA C
Sbjct  194  LKVMVAECRDNNVADFTSFVTSLVAIIANC  223


>☐ref|NP_203744.1| UG MICAL-like protein 1 [Homo sapiens]
 sp|Q8N3F8.2|MILK1_HUMAN G RecName: Full=MICAL-like protein 1; AltName: Full=Molecule interacting
with Rab13; Short=MIRab13
 emb|CAD42713.1| G molecule interacting with Rab13 [Homo sapiens]
   11 more sequence titles
 Length=863

 GENE ID: 85377 MICALL1 | MICAL-like 1 [Homo sapiens] (Over 10 PubMed links)

 Score = 40.1 bits (91),  Expect = 0.55, Method: Composition-based stats.
```

```
 Identities = 22/167 (13%), Positives = 39/167 (23%), Gaps = 16/167 (9%)

Query  13   IAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLAL-GNRINTD  71
              ++   P            +P    T    P         P   S   LAL  +R++
Sbjct  441  APSLATSPALGHPESTPKSLHPWYGITPTSSP--KTKKRPAPRAPSASPLALHASRLSHS  498

Query  72   TALTDYGTNVNGLGINVNRRR-------------RRRRDIWNIESNPVLAGLAANNPVNG  118
               +   +    +    +                  +         P   G   +   N
Sbjct  499  EPPSATPSPALSVESLSSESASQTAGAELLEPPAVPKSSSEPAVHAPGTPGNPVSLSTNS  558

Query  119  SYIAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIP  165
            S  + G  V      +   S  +A  + G +G       S   P
Sbjct  559  SLASSGELVEPRVEQMPQASPGLAPRTRGSSGPQPAKPCSGATPTPL  605
```

>☐ref|XP_515124.2| UG PREDICTED: molecule interacting with Rab13 [Pan troglodytes]
Length=889

 GENE ID: 458822 MICALL1 | MICAL-like 1 [Pan troglodytes]

```
 Score = 39.7 bits (90),  Expect = 0.66, Method: Composition-based stats.
 Identities = 22/167 (13%), Positives = 39/167 (23%), Gaps = 16/167 (9%)

Query  13   IAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLAL-GNRINTD  71
              ++   P            +P    T    P         P   S   LAL  +R++
Sbjct  467  APSLATSPALGHPESTPKSLHPWYGITPTSSP--KTKKRPAPRAPSASPLALHASRLSHS  524

Query  72   TALTDYGTNVNGLGINVNRRR-------------RRRRDIWNIESNPVLAGLAANNPVNG  118
               +   +    +    +                  +         P   G   +   N
Sbjct  525  EPPSATPSPALSVESLSSESASQTAGAELLEPPAVPKSSSEPAVHAPGTPGNPVSLSTNS  584

Query  119  SYIAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIP  165
            S  + G  V      +   S  +A  + G +G       S   P
Sbjct  585  SLASSGELVEPRVEQMPQASPGLAPRTRGSSGPQPAKPCSGATPTPL  631
```

>☐ref|XP_001843226.1| G conserved hypothetical protein [Culex quinquefasciatus]
 gb|EDS31285.1| G conserved hypothetical protein [Culex quinquefasciatus]
Length=346

 GENE ID: 6032659 CpipJ_CPIJ001686 | hypothetical protein
[Culex quinquefasciatus]

```
 Score = 39.3 bits (89),  Expect = 0.77, Method: Composition-based stats.
 Identities = 33/310 (10%), Positives = 75/310 (24%), Gaps = 33/310 (10%)

Query  106  VLAGLAANNPVNGSYIAIGNQVFHLPNNLSN------VSLNIASGSSGGAGQTVPSNFST  159
             +    +   ++ S   + ++    P N+        +  +  S  S    +
Sbjct  49   GMEDGVSEPAIDSSDATVKDE----PENVPKKCVKDIIPDDSISFKSKALQKLKNLPNEV  104

Query  160  FLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLE  219
                   N  Q + ++ T FQ           + + + Y    K               +
Sbjct  105  PQAPPLPNMGQFWSQITTIFQQPTFTEKVEEAKQVFNQYRALIKAFSQEWLKFFTKNRQD  164

Query  220  HGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIGR-RVVGKGTDC  278
               +V +    +  +  +    +             +C+    +       +V      C
Sbjct  165  AIANVQKVIQDTNQEIKSFFTRKFSEFDGLCLPEANECLKSVQNSLKEYSIKVEENTAAC  224

Query  279  ---VNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNAL  335
               V+R+  Q   +       + +        +  +  C
Sbjct  225  GEFVDRQLDQHGQLVEDEQKKLESG-------MLRIDGC------------FRKKEFLGG  265

Query  336  RRQCYVRAIGNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVALCASEMGLEIGLVTAQVS  395
               C      NF               GK  A++++       C  E    +     Q++
Sbjct  266  VMACVGSVAANFFHQTSQATQKFATAMGKASASVTNRLGRFRTCVVERKRLLAGGQNQIA  325

Query  396  SKIAFCQLAQ  405
             K   C  AQ
Sbjct  326  QKATACLKAQ  335
```

>☐ref|XP_002462548.1| UG hypothetical protein SORBIDRAFT_02g027920 [Sorghum bicolor]
 gb|EER99069.1| G hypothetical protein SORBIDRAFT_02g027920 [Sorghum bicolor]
Length=542

 GENE ID: 8063762 SORBIDRAFT_02g027920 | hypothetical protein [Sorghum bicolor]
(10 or fewer PubMed links)

```
 Score = 39.3 bits (89),  Expect = 0.78, Method: Composition-based stats.
 Identities = 35/309 (11%), Positives = 84/309 (27%), Gaps = 25/309 (8%)

Query  38   RTINQLPVTSATSATTPTQQSNVTLALG---NRINTDTALTDYGTNVNGLGINVNRRRRR  94
               +    + +++A +  +    ++  G   +R  +      +G  +    +     R
Sbjct  150  SDNSDDRTSKSSTAASAPRHDFSSVGFGRERDRARSRAPGAAHGKGIRAARVRGASARVA  209

Query  95   RRDIWNIESNPVLAGLAANNPVNGSYIAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVP  154
                            A N+  +       +   H   + S V+ N         G
Sbjct  210  -----------SSPAEAENSRSSVESNLRSSNAAHARRSSSGVTSNGVHKVLFPDGHQSD  258

Query  155  SNFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGL  214
                +            +++  +    L     DS      +A +  F+   T +GL+
Sbjct  259  DEPPSADLRYATG--GFYKENGSVIGRLANCDSDSNNHIFDEASAGKFENGGTHSGLDPY  316

Query  215  NETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGK  274
             E++   +S  E     I     +  E    +  DS+    +          +      +
Sbjct  317  AESVALLQSAQEALENEI----QKFVEIREETDEDSTTHQSEIEWSSSPVEEL-----NE  367

Query  275  GTDCVNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNA  334
                +  K  + + +  +    I+  D  +     N   CN    +           +
Sbjct  368  KIKMLESKLEEATMVINKQDSKIIELDALSQVQPQNTIACNNDLLSLPSDVDQLLLEKLE  427

Query  335  LRRQCYVRA  343
               QC++
Sbjct  428  AEIQCFILT  436
```

>☐ref|XP_001644303.1| G hypothetical protein Kpol_1066p9 [Vanderwaltozyma polyspora DSM
70294]
 gb|EDO16445.1| G hypothetical protein Kpol_1066p9 [Vanderwaltozyma polyspora DSM

```
70294]
Length=678

 GENE ID: 5544602 Kpol_1066p9 | hypothetical protein
[Vanderwaltozyma polyspora DSM 70294] (10 or fewer PubMed links)

 Score = 39.3 bits (89),  Expect = 0.81, Method: Composition-based stats.
 Identities = 25/206 (12%), Positives = 57/206 (27%), Gaps = 12/206 (5%)

Query  12   LIAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTD  71
            LI  I                N      +   P+  + S  +   +      L + +N
Sbjct  245  LILEIDTALTIQLQKYTIWTENLALNTGVTISPLEESKSMRSIAAKVTNERDLYDYLNKY  304

Query  72   TALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIGNQVFHLP  131
                             +    R       + +P +A     NP    +        +
Sbjct  305  -----------NKQSKDSILVNRNLIPVPYKKHPSMAKGQLINPTGSFTVNSSRNTNNFK  353

Query  132  NNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIV  191
            N+ ++ S+N   G    +  T  S  ++   +     + A   + +    + ++ +
Sbjct  354  NSTTSKSINTEGGFHSNSSSTYNSPANSVNNLGSTKTSLANTAIASSMSGIGMNLLSHTD  413

Query  192  QTLSDAYSNTFKGLFTAAGLNGLNET  217
             T + +  +    +F    LN  N
Sbjct  414  STGNQSPIDNS-SMFKPPSLNPGNGI  438


>☐ref|ZP_05690681.1|  phage tape measure protein [Staphylococcus aureus A8115]
 ref|ZP_06302931.1|  tape measure protein [Staphylococcus aureus A8117]
 gb|EEV76400.1|  phage tape measure protein [Staphylococcus aureus A8115]
 gb|EFC03114.1|  tape measure protein [Staphylococcus aureus A8117]
Length=1155

 Score = 39.3 bits (89),  Expect = 0.96, Method: Composition-based stats.
 Identities = 18/186 (9%), Positives = 53/186 (28%), Gaps = 5/186 (2%)

Query  122   AIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQT  181
              I N ++     L N  + +A+         + +        I + +   +  L +   T
Sbjct  861   TIWNAIWTFLQTLWNTIVTVATKVWNAITTAISTALQAAWSFISNIWNTIWSFLSSILTT  920

Query  182   LPLDAMD--SIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQG  239
             +    +   + V +      +         G+  ++       + +   +Q  V+   +
Sbjct  921   IWNKVVSIFTQVVSTISDKMSQAWNFIVTKGMQWVSTITSTLINFVNRVVQGFVNVVNKV  980

Query  240   FEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILA  299
              + +  + N   + V   V        + R ++    +       + + +
Sbjct  981   SQGMTNAVNKVKSFVDDFVSAGA---DMIRGLMRGIGNMARDLAEKAASVAKGALNAAKR  1037

Query  300   ADRGAS  305
             A    S
Sbjct  1038  ALGIHS  1043


>☐ref|XP_002257750.1| G SICA antigen (fragment) [Plasmodium knowlesi strain H]
 emb|CAQ38086.1| G SICA antigen (fragment) [Plasmodium knowlesi strain H]
Length=1059

 GENE ID: 7318622 PKH_020760 | SICA antigen (fragment)
[Plasmodium knowlesi strain H] (10 or fewer PubMed links)

 Score = 39.3 bits (89),  Expect = 0.99, Method: Composition-based stats.
 Identities = 28/213 (13%), Positives = 58/213 (27%), Gaps = 10/213 (4%)

Query  135  SNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQTL  194
             N S+    G    +                D   Q  ++       L    +D+
Sbjct  451  PNSSIGKMGGVVYDSESLCKGGGDRKSARTADEKKQQLKENTKSMAKLFGMNVDTTQSGA  510

Query  195  SDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARV  254
              +        F A   + L    E   +V++D   +   Q      ++     + +
Sbjct  511  GGSAGGNPSNPFDALIDDKLTLQQEFIENVLKDAFDASSGQV--DTNKINDVVKNITNAS  568

Query  255  QQCVGPDLSPTSIGRRVVGKGTDCVNRKWR-------QLSDIGGRIAGDILAADRGASDF  307
            QQ    +    + G   +    DC+   W           +   +  G++       S
Sbjct  569  QQVQAENCIQKNSGDDKLCSRLDCMKHLWDNKSGKTSNAENFWNKDTGEVRKLWDELSGA  628

Query  308  LGNLTTCNGANFNAADASSGGTGT-RNALRRQC  339
            +      +       D  +GG+ T  +  +R C
Sbjct  629  IKEKGKNDQNECKTMDDGTGGSRTATDPEKRAC  661


>☐emb|CAQ07542.1|  MICAL-like 1 [Homo sapiens]
Length=369

 Score = 38.9 bits (88),  Expect = 1.0, Method: Composition-based stats.
 Identities = 22/167 (13%), Positives = 39/167 (23%), Gaps = 16/167 (9%)

Query  13   IAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLAL-GNRINTD  71
              ++   P            +P    T    P         P   S   LAL  +R++
Sbjct  18   APSLATSPALGHPESTPKSLHPWYGITPTSSP--KTKKRPAPRAPSASPLALHASRLSHS  75

Query  72   TALTDYGTNVNGLGINVNRRR-------------RRRRDIWNIESNPVLAGLAANNPVNG  118
               +   +    +    +                  +         P   G   +   N
Sbjct  76   EPPSATPSPALSVESLSSESASQTAGAELLEPPAVPKSSSEPAVHAPGTPGNPVSLSTNS  135

Query  119  SYIAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIP  165
            S  + G  V      +   S  +A  + G +G       S   P
Sbjct  136  SLASSGELVEPRVEQMPQASPGLAPRTRGSSGPQPAKPCSGATPTPL  182


>☐dbj|BAB33338.1| G KIAA1668 protein [Homo sapiens]
Length=791

 GENE ID: 85377 MICALL1 | MICAL-like 1 [Homo sapiens] (Over 10 PubMed links)

 Score = 38.9 bits (88),  Expect = 1.0, Method: Composition-based stats.
 Identities = 22/167 (13%), Positives = 39/167 (23%), Gaps = 16/167 (9%)

Query  13   IAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLAL-GNRINTD  71
              ++   P            +P    T    P         P   S   LAL  +R++
Sbjct  369  APSLATSPALGHPESTPKSLHPWYGITPTSSP--KTKKRPAPRAPSASPLALHASRLSHS  426
```

```
Query  72   TALTDYGTNVNGLGINVNRRR-------------RRRRDIWNIESNPVLAGLAANNPVNG  118
               +   +    +    +                  +         P   G   +   N
Sbjct  427  EPPSATPSPALSVESLSSESASQTAGAELLEPPAVPKSSSEPAVHAPGTPGNPVSLSTNS  486

Query  119  SYIAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIP  165
            S  + G  V      +   S  +A  + G +G       S   P
Sbjct  487  SLASSGELVEPRVEQMPQASPGLAPRTRGSSGPQPAKPCSGATPTPL  533


>☐ref|ZP_04105490.1|  TMP repeat-containing protein [Bacillus thuringiensis serovar
berliner ATCC 10792]
 ref|ZP_04137053.1|  TMP repeat-containing protein [Bacillus thuringiensis serovar
thuringiensis str. T01001]
 ref|ZP_04140675.1|  TMP repeat-containing protein [Bacillus thuringiensis Bt407]
 gb|EEM27442.1|  TMP repeat-containing protein [Bacillus thuringiensis Bt407]
 gb|EEM31259.1|  TMP repeat-containing protein [Bacillus thuringiensis serovar
thuringiensis str. T01001]
 gb|EEM62827.1|  TMP repeat-containing protein [Bacillus thuringiensis serovar
berliner ATCC 10792]
Length=1208

 Score = 38.9 bits (88),  Expect = 1.3, Method: Composition-based stats.
 Identities = 20/188 (10%), Positives = 60/188 (31%), Gaps = 15/188 (7%)

Query  79   TNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIGNQVFHLPNNLSNVS  138
              ++G  + +         ++ I +   L  +A    +  + +A+G  ++   +++ N +
Sbjct  580  GVISGALMAIAPIVVSVMGVFEIGAAAALGIVAIVPIIIAAIVALGVAIYKNWDDIKNWT  639

Query  139  LNIASGSSGGAGQTVPSNFST---FLPMIPDNYTQAFQKLVTQFQTLPLDAM-------D  188
            +           +       +         D   + F  +   F  L            +
Sbjct  640  IEAWDSIKEYLVELWDGISQSCSEAWSSFLDVMHEFFDPIGQFFSDLWEGVKQACSDAWN  699

Query  189  SIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFN  248
            S V+  S+A+S+  + +      +  +   E   S+    +++     +   E     +
Sbjct  700  STVEFFSEAWSSFIEMM-----HSFFDPIGEFFSSLWSGIVETASSWWSSLVETASELWG  754

Query  249  DSSARVQQ  256
              +   Q+
Sbjct  755  TLTQAWQE  762


>☐emb|CAD38718.1| G hypothetical protein [Homo sapiens]
Length=657

 GENE ID: 85377 MICALL1 | MICAL-like 1 [Homo sapiens] (Over 10 PubMed links)

 Score = 38.5 bits (87),  Expect = 1.4, Method: Composition-based stats.
 Identities = 22/167 (13%), Positives = 39/167 (23%), Gaps = 16/167 (9%)

Query  13   IAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLAL-GNRINTD  71
              ++   P            +P    T    P         P   S   LAL  +R++
Sbjct  235  APSLATSPALGHPESTPKSLHPWYGITPTSSP--KTKKRPAPRAPSASPLALHASRLSHS  292

Query  72   TALTDYGTNVNGLGINVNRRR-------------RRRRDIWNIESNPVLAGLAANNPVNG  118
               +   +    +    +                  +         P   G   +   N
Sbjct  293  EPPSATPSPALSVESLSSESSSQTAGAELLEPPAVPKSSSEPAVHAPGTPGNPVSLSTNS  352

Query  119  SYIAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIP  165
            S  + G  V      +   S  +A  + G +G       S   P
Sbjct  353  SLASSGELVEPRVEQMPQASPGLAPRTRGSSGPQPAKPCSGATPTPL  399


>☐ref|ZP_03158061.1|  Hemolysin-type calcium-binding region [Cyanothece sp. PCC 7822]
 gb|EDX94266.1|  Hemolysin-type calcium-binding region [Cyanothece sp. PCC 7822]
Length=3976

 Score = 38.5 bits (87),  Expect = 1.5, Method: Composition-based stats.
 Identities = 30/298 (10%), Positives = 82/298 (27%), Gaps = 16/298 (5%)

Query  41    NQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIWN  100
                    +ATS       +++ ++  +  NT   +    T    + + +  +   +  +
Sbjct  1175  TADFNFTATSNIIDPTPNDLGISYSDISNTSLLVNPSYTGKGNITLPLEFKGLNQFGLET  1234

Query  101   IESNPVLAGLAANNPVNGSYIAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTF  160
                ++ ++  +A NN +  +                N  ++               + S
Sbjct  1235  PRTDLIVNFIAENNKIIKTVKPFK----------FNSPISDLEEKIKSLQNFSLDDVSNL  1284

Query  161   LPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQTL--SDAYSNTFKGLFTAAGLNGLNETL  218
                +          L  +   L   +++ ++  L       N +K    +       +T
Sbjct  1285  AEQVITFLDNKNLSLFNEKVPLVDQSINELLGFLNPIKERYNKYKESLDSLQNVLAEKTE  1344

Query  219   EHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDC  278
              +    + E  +  +    A   +Q+  +    +  +              +  +    +
Sbjct  1345  KFENLIFEPSL--LSQLPASVSQQLQNTLATLNKAIDNLPSGLDILKKGIQPELISAINQ  1402

Query  279   VNRKWRQL-SDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNAL  335
              +    + L       I  +I  A      F   L T     F   D++     + N
Sbjct  1403  IQELIKSLPQQSINTIKDEIDNAINSLP-FSDKLNTAKNNLFAVLDSADLNLESFNQA  1459


>☐ref|XP_001122565.1| UG PREDICTED: hypothetical protein [Apis mellifera]
Length=177

 GENE ID: 726847 LOC726847 | hypothetical protein LOC726847 [Apis mellifera]
(10 or fewer PubMed links)

 Score = 38.5 bits (87),  Expect = 1.5, Method: Composition-based stats.
 Identities = 10/105 (9%), Positives = 28/105 (26%), Gaps = 1/105 (0%)

Query  214  LNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVG  273
            + E  +    +  D    + +       ++          +  C   +     +  + +
Sbjct  65   VEEMKKKSERINTDARNFVNNILKTMGFELKN-LKKYLKGMSDCNAKEKELDHLFDQFID  123

Query  274  KGTDCVNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTCNGAN  318
               +C +RK  +   +   I           ++   N   C   N
Sbjct  124  STNNCTSRKLNEAIQLTVEINTKSENLLNDLNNISTNAAKCTTIN  168


>☐ref|XP_001580855.1| G hypothetical protein [Trichomonas vaginalis G3]
 gb|EAY19869.1| G hypothetical protein TVAG_129770 [Trichomonas vaginalis G3]
```

```
Length=2050

 GENE ID: 5465399 TVAG_129770 | hypothetical protein [Trichomonas vaginalis G3]
(10 or fewer PubMed links)

 Score = 38.5 bits (87),  Expect = 1.6, Method: Composition-based stats.
 Identities = 20/148 (13%), Positives = 41/148 (27%), Gaps = 3/148 (2%)

Query  18   IIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDY  77
              P +       T W        +    ++  S       SN  +   N   T    + Y
Sbjct  701  TTPAWALPSSGATSWEAWSKNVKHTENTSAQHSPAPSPVSSNNPVPSLNIPQTTVNPSAY  760

Query  78   GTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIGNQVFHLPNNLSNV  137
                +        +  +        + P +         N   I + N     P N S
Sbjct  761  SPQTSPQRAPPPNKVNQPYIPGMAPTKPPIVSSPLAPQTN-KQIFVPNNAP--PPNSSKP  817

Query  138  SLNIASGSSGGAGQTVPSNFSTFLPMIP  165
            S+ I   +      + P++ + ++P +
Sbjct  818  SIMIPGVTPPVPAPSQPTSSAPYIPGMS  845


>☐ref|XP_001648260.1| UG hypothetical protein AaeL_AAEL014240 [Aedes aegypti]
 gb|EAT33485.1| G hypothetical protein AaeL_AAEL014240 [Aedes aegypti]
Length=292

 GENE ID: 5563936 AaeL_AAEL014240 | hypothetical protein [Aedes aegypti]
(10 or fewer PubMed links)

 Score = 38.5 bits (87),  Expect = 1.6, Method: Composition-based stats.
 Identities = 39/271 (14%), Positives = 83/271 (30%), Gaps = 14/271 (5%)

Query  133  NLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQ  192
            N+S  +L  A         T   + +T       N    +   + +FQ      +  + +
Sbjct  33   NISKTTLAAAIQQLNTTINTADESVNTAWTQFTQNMASLYTLYLNRFQNYTTFDLTPLNR  92

Query  193  TLSDAYS--NTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDS  250
            T+ D     +  + +   +     ++       V      ++
Sbjct  93   TICDLQMTLSPPQPIQDWSTNQIFSDVQASALQVETALNTAVQTMSNSCAPSCANPTK--  150

Query  251  SARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILAADRGASDFLGN  310
               V  C        +     + + TDCV  +  + SDIG ++A         A ++L
Sbjct  151  ---VATCTTKFGPRLTTAPITIDRLTDCVTAELARYSDIGSKMAAQYSNVLTSAMNYLKV  207

Query  311  LTTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLFLPVSLTIEGGKLYAALS  370
            +  C+     A +  +  +    +   QC  + +         +P+++       Y
Sbjct  208  VDVCDT---PAPEVLNNPSPMYASPSTQCMTQYLQRVSN----MPINVYTVDSMRYPQTY  260

Query  371  SLQTDVALCASEMGLEIGLVTAQVSSKIAFC  401
             +Q  V  CA  + L+I     +V      C
Sbjct  261  LIQNRVQRCAKLVELDIKDRITKVLESFQNC  291


>☐gb|EEZ97902.1|  hypothetical protein TcasGA2_TC000287 [Tribolium castaneum]
Length=438

 Score = 38.5 bits (87),  Expect = 1.7, Method: Composition-based stats.
 Identities = 31/219 (14%), Positives = 62/219 (28%), Gaps = 6/219 (2%)

Query  188  DSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASF  247
            DS V T     +   +    A   N     L +     +       +  A   + +  S
Sbjct  215  DSNVSTYVTEATTILQNDAKALLQNLSELALGNILVAHDQINDLGSNLAAFSTDVIFKSE  274

Query  248  NDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILAADRGAS--  305
             D     Q       +  S+   V    ++C++    +L  +       I     G
Sbjct  275  LDIQEETQAIDEEISNFKSLADAVNKNISECLDGNEDKLKKLPETFRKQIRECVAGVQKE  334

Query  306  --DFLGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLFLPVSLTIEGG  363
                L  +        N     +   G        C    +    +++  +P+ + +E
Sbjct  335  SDSLLSKVFYNIDIMMNVVHDLNFQLGQCADDDNSCISAVVDKINEAMENIPLKIKVEIN  394

Query  364  KLYAALSSLQTDVALCA-SEMGLEIGLVTAQVSSKIAFC  401
                 +   Q  +  CA S +   +  VT  +S  +  C
Sbjct  395  SASDQVELYQAKIQQCADSGVAGYVENVTDIMSD-VTVC  432


>☐ref|ZP_05688243.1|  phage tape measure protein [Staphylococcus aureus A9299]
 gb|EEV73671.1|  phage tape measure protein [Staphylococcus aureus A9299]
Length=1154

 Score = 38.2 bits (86),  Expect = 1.8, Method: Composition-based stats.
 Identities = 20/186 (10%), Positives = 53/186 (28%), Gaps = 5/186 (2%)

Query  122   AIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQT  181
              I N ++     L N  + +A+        T+ +        I + +   +  L     T
Sbjct  860   TIWNAIWTFLQTLWNTIVTVATKVWNAITTTISTALQAAWSFISNIWNTIWSFLSGILTT  919

Query  182   LPLDAMD--SIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQG  239
             +    +   + V +      +         G+  ++       + +   IQ  V+   +
Sbjct  920   IWNKVVSIFTQVVSTISDKMSQAWNFIVTKGMQWVSTITSTLINFVNRVIQGFVNVVNKV  979

Query  240   FEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILA  299
              + +  + N   + V   V        + R ++    +       + + +
Sbjct  980   SQGMTNAVNKVKSFVDDFVSAGA---DMIRGLMRGIGNMARDLAEKAASVAKGALNAAKR  1036

Query  300   ADRGAS  305
             A    S
Sbjct  1037  ALGIHS  1042


>☐ref|XP_002767144.1| G centromere protein a, putative [Perkinsus marinus ATCC 50983]
 gb|EEQ99861.1| G centromere protein a, putative [Perkinsus marinus ATCC 50983]
Length=925

 GENE ID: 9051972 Pmar_PMAR006533 | centromere protein a, putative
[Perkinsus marinus ATCC 50983]

 Score = 38.2 bits (86),  Expect = 2.1, Method: Composition-based stats.
 Identities = 25/201 (12%), Positives = 45/201 (22%), Gaps = 6/201 (2%)

Query  14   AAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTA  73
```

```
                  P       +T       PR IN  P          +  ++  +    R +++
Sbjct  145  VQSMESPTRPSPVSWTISVTDLTPRKINNSPRQLTNRNPRKSSDNHQHVKEAPRRSSNAP  204

Query  74   LTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIG-NQVFHLPN  132
                 T       +       R              + +N P   S    G +      +
Sbjct  205  RGTSNTPRETSNTSRETSNTPRETSNTPPVANNTPRVISNTPRGTSNTPRGTSNTPRGAS  264

Query  133  NLSNVSLNIASGSSGGAGQTVPSN--FSTFLPMIPDNYTQAFQKLVTQFQTLPLDAMDSI  190
            N    + N   G+S     T  +    S       +                P    ++
Sbjct  265  NTPRGTSNTPRGTSNTPRGTSNTPRGTSNTPRGT-SNTPRGTSNTPRGTSNTP  323

Query  191  --VQTLSDAYSNTFKGLFTAA  209
                      SNT +G
Sbjct  324  RGTSNTPRGTSNTPRGTSNTP  344


>☐ref|YP_239580.1| ORF001 [Staphylococcus phage 69]
 gb|AAX90756.1| ORF001 [Staphylococcus phage 69]
Length=1155

 GENE ID: 5132759 69ORF001 | ORF001 [Staphylococcus phage 69]
(10 or fewer PubMed links)

 Score = 38.2 bits (86),  Expect = 2.1, Method: Composition-based stats.
 Identities = 18/186 (9%), Positives = 52/186 (27%), Gaps = 5/186 (2%)

Query  122   AIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQT  181
              I N ++     L N  + +A++         + +        I + +   +  L     T
Sbjct  861   TIWNAIWTFLQTLWNTIVTVATKVWNAITTAISTALQAAWSFISNIWNTIWSFLSGILTT  920

Query  182   LPLDAMD--SIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQG  239
             +    +   + V +      +         G+  ++       + +   +Q  V+   +
Sbjct  921   IWNKVVSIFTQVVSTISDKMSQAWNFIVTKGMQWVSTITSTLINFVNRVVQGFVNVVNKV  980

Query  240   FEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILA  299
              + +  + N   + V   V        + R ++    +       + + +
Sbjct  981   SQGMTNAVNKVKSFVDDFVSAGA---DMIRGLMRGIGNMARDLAEKAASVAKGALNAAKR  1037

Query  300   ADRGAS  305
             A    S
Sbjct  1038  ALGIHS  1043


>☐ref|ZP_01630797.1|  hypothetical protein N9414_06069 [Nodularia spumigena CCY9414]
 gb|EAW44609.1|  hypothetical protein N9414_06069 [Nodularia spumigena CCY9414]
Length=375

 Score = 38.2 bits (86),  Expect = 2.1, Method: Composition-based stats.
 Identities = 25/201 (12%), Positives = 55/201 (27%), Gaps = 5/201 (2%)

Query  33   NPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGTNVNGLGINVNRRR  92
            +  P  T    P             ++++ +    +N  +  +  G N        +  +
Sbjct  174  SSPPTGTPATSPT-GVNQTENSQNPNSISFSQTP-VNQASNQSLSGFNGEIANPINSLEK  231

Query  93   RRRRDIWNIESNPVLAGLAANNPVNGSYIAIGNQVFHLPNNLSNVSLNIASGSSGGAGQT  152
                 +  I SN  L           +          +PNNL     N  +  +   G
Sbjct  232  TAAGGVMQIPSNNGLPSQGVYPSTGLNAATGVQPTVGVPNNLPP---NAVNQVNTIEGLP  288

Query  153  VPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLN  212
             P   ++ +  +P       Q       ++   + D +         N    ++ +
Sbjct  289  SPVQPNSVISPLPSGVESGVQTNFAPNSSIQAPSPDIVAPMSPVVADNNGNLIWRSPAQQ  348

Query  213  GLNETLEHGRSVIEDGIQSIV  233
                +L+         G+Q I
Sbjct  349  MQPNSLDSQIPGQNRGLQDID  369


>☐ref|XP_002420533.1| RNA-binding protein, putative [Candida dubliniensis CD36]
 emb|CAX41612.1| RNA-binding protein, putative [Candida dubliniensis CD36]
Length=954

 GENE ID: 8048216 CD36_52280 | RNA-binding protein, putative
[Candida dubliniensis CD36]

 Score = 37.8 bits (85),  Expect = 2.3, Method: Composition-based stats.
 Identities = 20/163 (12%), Positives = 47/163 (28%), Gaps = 16/163 (9%)

Query  29   TTPWNPQPPRTINQLPVTSATSAT---TPTQQSNVTLALGNRINTDTALTDYGTNVNGLG  85
                              S ++      TP      + +L +  N  +   +     NG
Sbjct  14   NISLGLSTSPLNTISSGGSDSTIGHLNTPPTTRVRSGSLFSIWNEPSLSHNSNNTSNGHE  73

Query  86   INVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIGNQVFHLPNN-----------L  134
            +  +  R R         +  +   + +   + +     +    +PNN           L
Sbjct  74   MATSTSRNRSYTTTGALPSITIDSFSESLQKSSAPRVSTSPFISVPNNDMIVLDNFQTNL  133

Query  135  SNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVT  177
              +     SG+   +  T+P    +     P ++ Q F+ + +
Sbjct  134  PRLRSQTISGTPISSQNTLPHPSESLNA--PHHFPQVFETVSS  174


>☐ref|XP_001032003.1| hypothetical protein TTHERM_00703320 [Tetrahymena thermophila]
 gb|EAR84340.1| hypothetical protein TTHERM_00703320 [Tetrahymena thermophila
SB210]
Length=3324

 GENE ID: 7823579 TTHERM_00703320 | hypothetical protein
[Tetrahymena thermophila]

 Score = 37.8 bits (85),  Expect = 2.6, Method: Composition-based stats.
 Identities = 25/191 (13%), Positives = 46/191 (24%), Gaps = 11/191 (5%)

Query  85    GINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSYIAIGNQVFHLPNNLSNVSLNIASG  144
               N  +     + + +I S P      AN     S  +I  +    P N    S++
Sbjct  2759  NANSQQNPVSEQIVDSILSEPFAPQPTANFDTRSSISSIKREPISYPGNTYTASVSGIPI  2818

Query  145   SSGGA-GQTVPSNFSTFLPMIPDNY--TQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNT  201
                 +    V  N S             Q         Q      +           SN
Sbjct  2819  QPYSSWTNPVSLNSSGGTAAFLGLPVSQQNKNSQPQTNQNSANQTLFVSFGNEKILGSNP  2878
```

```
Query  202   FKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPD  261
              K           ++ L    +     +           +    SFN      Q  V
Sbjct  2879  LKSSVPG---TNKDQILSGLPTYTPSIL-----MSTSFSKLPNNSFNQGEKDKQSSVRSS  2930

Query  262   LSPTSIGRRVV  272
              + ++  ++ +
Sbjct  2931  GAVSNEYKQAL  2941


>☐ref|ZP_06324921.1|  tape measure protein [Staphylococcus aureus subsp. aureus D139]
 gb|EFB49148.1|  tape measure protein [Staphylococcus aureus subsp. aureus D139]
Length=1154

 Score = 37.8 bits (85),  Expect = 2.6, Method: Composition-based stats.
 Identities = 18/186 (9%), Positives = 53/186 (28%), Gaps = 5/186 (2%)

Query  122   AIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQT  181
              I N ++     L N  + +A+         + +        I + +   +  L     T
Sbjct  860   TIWNAIWTFLQTLWNTIVTVATKVWNAITTAISTALQAAWSFISNIWNTIWSFLSGILTT  919

Query  182   LPLDAMD--SIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQG  239
             +    +   + V +      +         G+  ++       + +   IQ  V+   +
Sbjct  920   IWNKVVSIFTQVVSTISDKMSQAWNFIVTKGMQWVSTITSTLINFVNRVIQGFVNVVNKV  979

Query  240   FEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILA  299
              + +  + N   + +   V        + R ++        +      ++  +      +
Sbjct  980   SQGMTNTVNKIKSFIGDFVSAGA---DMIRGLIRGIGQMAGQLVDAAKNVAKKALDAAKS  1036

Query  300   ADRGAS  305
             A    S
Sbjct  1037  ALGIHS  1042


>☐dbj|BAE38328.1| G unnamed protein product [Mus musculus]
 gb|EDL31272.1| G RIKEN cDNA A430110N23 [Mus musculus]
 gb|AAI41045.1| G RIKEN cDNA A430110N23 gene [Mus musculus]
Length=1371

 GENE ID: 269855 A430110N23Rik | RIKEN cDNA A430110N23 gene [Mus musculus]
(10 or fewer PubMed links)

 Score = 37.8 bits (85),  Expect = 2.8, Method: Composition-based stats.
 Identities = 25/203 (12%), Positives = 44/203 (21%), Gaps = 15/203 (7%)

Query  8     VLALLIAAIGIIPWYVDAFLFT-----TPWNPQPPRTINQLPVTSATSATT--PTQQSNV  60
             +L      +   P        T          +   T    P    T A    P    +
Sbjct  1047  LLPTRSPELSGSPTPTSPEGLTSASSMLSEVSRLSPTSELTPGPDTTPAPEIIPESSDSS  1106

Query  61    TLALGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSY  120
              L +  R  T      + T++  L          +      +        +   P N
Sbjct  1107  DLPMNTRTPTQPFTASHPTSIPQLNTTSYPTIAPQPTTNPQQPRSPHPATSPQPPTNTHP  1166

Query  121   IAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQ  180
              +               SL   +  +   +      N        P       Q L
Sbjct  1167  SSTP--------ATPTESLPSSRKTELSSPTKPRLNSELTFEEAPSTDASQTQNLELFLA  1218

Query  181   TLPLDAMDSIVQTLSDAYSNTFK  203
             +    +  S    L    ++ FK
Sbjct  1219  SESGPSSPSPASNLDPLPTDAFK  1241


>☐ref|NP_766596.1| UG group B scavenger receptor cysteine-rich domain-containing precursor
[Mus musculus]
 sp|Q8BV57.1|SRCRL_MOUSE G RecName: Full=Scavenger receptor cysteine-rich domain-containing
protein LOC284297 homolog; Flags: Precursor
 dbj|BAC37780.1| G unnamed protein product [Mus musculus]
 gb|ACF54723.1| G group B scavenger receptor cysteine-rich domain-containing protein
[Mus musculus]
Length=1371

 GENE ID: 269855 A430110N23Rik | RIKEN cDNA A430110N23 gene [Mus musculus]
(10 or fewer PubMed links)

 Score = 37.8 bits (85),  Expect = 2.8, Method: Composition-based stats.
 Identities = 25/203 (12%), Positives = 44/203 (21%), Gaps = 15/203 (7%)

Query  8     VLALLIAAIGIIPWYVDAFLFT-----TPWNPQPPRTINQLPVTSATSATT--PTQQSNV  60
             +L      +   P        T          +   T    P    T A    P    +
Sbjct  1047  LLPTRSPELSGSPTPTSPEGLTSASSMLSEVSRLSPTSELTPGPDTTPAPEIIPESSDSS  1106

Query  61    TLALGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNGSY  120
              L +  R  T      + T++  L          +      +        +   P N
Sbjct  1107  DLPMNTRTPTQPFTASHPTSIPQLNTTSYPTIAPQPTTNPQQPRSPHPATSPQPPTNTHP  1166

Query  121   IAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQ  180
              +               SL   +  +   +      N        P       Q L
Sbjct  1167  SSTP--------ATPTESLPSSRKTELSSPTKPRLNSELTFEEAPSTDASQTQNLELFLA  1218

Query  181   TLPLDAMDSIVQTLSDAYSNTFK  203
             +    +  S    L    ++ FK
Sbjct  1219  SESGPSSPSPASNLDPLPTDAFK  1241


>☐ref|NP_803295.1| G tape measure protein [Staphylococcus phage 11]
 gb|AAL82270.1| G tape measure protein [Staphylococcus phage 11]
Length=910

 GENE ID: 1258053 phi11_42 | tape measure protein [Staphylococcus phage 11]
(10 or fewer PubMed links)

 Score = 37.8 bits (85),  Expect = 2.8, Method: Composition-based stats.
 Identities = 17/183 (9%), Positives = 51/183 (27%), Gaps = 5/183 (2%)

Query  125   NQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPL  184
             N ++     L N  + +A+         + +        I + +   +  L     T+
Sbjct  619   NAIWTFLQTLWNTIVTVATKVWNAITTAISTALQAAWSFISNIWNTIWSFLSGILTTIWN  678
```

```
Query  185  DAMD--SIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQ  242
              +   + V +      +         G+  ++       + +   +Q  V+   +  +
Sbjct  679  KVVSIFTQVVSTISDKMSQAWNFIVTKGMQWVSTITSTLINFVNRVVQGFVNVVNKVSQG  738

Query  243  VIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILAADR  302
            +  + N   + V   V        + R ++    +       + + +          A
Sbjct  739  MTNAVNKVKSFVDDFVSAGA---DMIRGLMRGIGNMARDLAEKAASVAKGALNAAKRALG  795

Query  303  GAS  305
             S
Sbjct  796  IHS  798


>☐ref|YP_500528.1| phage tape measure protein [Staphylococcus aureus subsp. aureus
NCTC 8325]
 gb|ABD31087.1| phage tape measure protein [Staphylococcus aureus subsp. aureus
NCTC 8325]
Length=1155

 GENE ID: 3920485 SAOUHSC_02033 | phage tape measure protein
[Staphylococcus aureus subsp. aureus NCTC 8325]

 Score = 37.8 bits (85),  Expect = 2.9, Method: Composition-based stats.
 Identities = 17/183 (9%), Positives = 51/183 (27%), Gaps = 5/183 (2%)

Query  125   NQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPL  184
             N ++     L N  + +A++         + +        I + +   +  L     T+
Sbjct  864   NAIWTFLQTLWNTIVTVATKVWNAITTAISTALQAAWSFISNIWNTIWSFLSGILTTIWN  923

Query  185   DAMD--SIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQ  242
               +   + V +      +         G+  ++       + +   +Q  V+   +  +
Sbjct  924   KVVSIFTQVVSTISDKMSQAWNFIVTKGMQWVSTITSTLINFVNRVVQGFVNVVNKVSQG  983

Query  243   VIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILAADR  302
             +  + N   + V   V        + R ++    +       + + +          A
Sbjct  984   MTNAVNKVKSFVDDFVSAGA---DMIRGLMRGIGNMARDLAEKAASVAKGALNAAKRALG  1040

Query  303   GAS  305
              S
Sbjct  1041  IHS  1043


>☐ref|YP_627836.1| outer membrane protein HopI [Helicobacter pylori HPAG1]
 gb|ABF85162.1| outer membrane protein HopI [Helicobacter pylori HPAG1]
Length=696

 GENE ID: 4097996 HPAG1_1095 | outer membrane protein HopI
[Helicobacter pylori HPAG1] (10 or fewer PubMed links)

 Score = 37.4 bits (84),  Expect = 3.0, Method: Composition-based stats.
 Identities = 20/163 (12%), Positives = 36/163 (22%), Gaps = 11/163 (6%)

Query  95   RRDIWNIESNPVLAGL----AANNPVNGSYIAIGNQVFHLPNNLSNVSLNIASGSSGGAG  150
              D  N      +       +  N +N     IG  +   P NL N  L          G
Sbjct  115  NGDTGNNPPRGNVNATFDMQSLVNNLNKLTQLIGETLIRNPENLPNAKL-----IEVKFG  169

Query  151  QTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAG  210
                S        + +        +     TL  +   +     S    N    +
Sbjct  170  N--QSTVIALPEGLANTMNALNNDITNALTTLWYNQTLTNKSFNSGNSVNFSPQVLQHLL  227

Query  211  LNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSAR  253
             +GL     +          +  ++     +     F
Sbjct  228  QDGLATASNNQTICSTQNQCTATNEAKSIAQNAQNIFQALMQA  270


>☐gb|EEU07699.1|  Sef1p [Saccharomyces cerevisiae JAY291]
Length=1148

 Score = 37.4 bits (84),  Expect = 3.1, Method: Composition-based stats.
 Identities = 29/214 (13%), Positives = 73/214 (34%), Gaps = 16/214 (7%)

Query  47    SATSATTPTQQSNVTLALGNRINTD--TALTDYGTNVNGLGINVNRRRRRRRDIWNIE-S  103
              A++  +    S   +A     N +  +   DY +N   +  +++ +R     + +   S
Sbjct  920   DASNDISIPSNSIYPVASVPASNNNPQSTKVDYYSNGPSVIPDLSMKRSVSTPVNHFPAS  979

Query  104   NPVLAGLAANNPVNGSYIAIGNQVFHLPNNLSNVSLNIAS--------GSSGGAGQTVPS  155
              P L      N  N   + I + +    +NL NV++N  +        G S  +     +
Sbjct  980   VPGLRNHPVGNLSNNVTLGIDHPIPREHSNLQNVTMNYNNQFSNANAIGRSQSSMSHSRT  1039

Query  156   NFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAM----DSIVQTLSDAYSNTFKGLFTAAGL  211
               ++    + D ++       ++    P  ++    + I    ++   +T     T++
Sbjct  1040  PIASKSNNMTDLHSVVSDPGSSKSTAYPPLSLFSKSNDINSNKTNQRFSTGTNTVTSSNF  1099

Query  212   NGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIA  245
               ++    + ++           Q A   E   +
Sbjct  1100  QTIDN-ENNVKTPGNKLTDFFQQQSAGWIEGNSS  1132


>☐gb|EDN64553.1|  suppressor of essential function [Saccharomyces cerevisiae YJM789]
Length=1148

 Score = 37.4 bits (84),  Expect = 3.1, Method: Composition-based stats.
 Identities = 29/214 (13%), Positives = 73/214 (34%), Gaps = 16/214 (7%)

Query  47    SATSATTPTQQSNVTLALGNRINTD--TALTDYGTNVNGLGINVNRRRRRRRDIWNIE-S  103
              A++  +    S   +A     N +  +   DY +N   +  +++ +R     + +   S
Sbjct  920   DASNDMSIPSNSIYPVASVPASNNNPQSTKVDYYSNGPSVIPDLSMKRSVSTPVNHFPAS  979

Query  104   NPVLAGLAANNPVNGSYIAIGNQVFHLPNNLSNVSLNIAS--------GSSGGAGQTVPS  155
              P L      N  N   + I + +    +NL NV++N  +        G S  +     +
Sbjct  980   VPGLRNHPVGNLSNNVTLGIDHPIPREHSNLQNVTMNYNNQFSNANAIGRSQSSMSHSRT  1039

Query  156   NFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAM----DSIVQTLSDAYSNTFKGLFTAAGL  211
               ++    + D ++       ++    P  ++    + I    ++   +T     T++
Sbjct  1040  PIASKSNNMTDLHSVVSDPGSSKSTAYPPLSLFSKSNDINSNKTNQRFSTGTNTVTSSNF  1099

Query  212   NGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIA  245
               ++    + ++           Q A   E   +
Sbjct  1100  QTIDN-ENNVKTPGNKLTDFFQQQSAGWIEGNSS  1132
```

```
>☐ref|NP_009487.2|  Putative transcription factor, has homolog in Kluyveromyces lactis
[Saccharomyces cerevisiae S288c]
 sp|P34228.4|SEF1_YEAST  RecName: Full=Putative transcription factor SEF1; AltName: Full=Suppressor
protein SEF1; AltName: Full=Suppressor of essential
function protein 1
 gb|EDV12153.1|  suppressor protein SEF1 [Saccharomyces cerevisiae RM11-1a]
 gb|EDZ73921.1|  YBL066Cp-like protein [Saccharomyces cerevisiae AWRI1631]
 emb|CAY77720.1|  Sef1p [Saccharomyces cerevisiae EC1118]
 tpg|DAA07055.1|  TPA: Putative transcription factor, has homolog in Kluyveromyces
lactis [Saccharomyces cerevisiae]
Length=1148

 GENE ID: 852214 SEF1 | Putative transcription factor, has homolog in
Kluyveromyces lactis [Saccharomyces cerevisiae S288c] (Over 10 PubMed links)

 Score = 37.4 bits (84),  Expect = 3.2, Method: Composition-based stats.
 Identities = 29/214 (13%), Positives = 73/214 (34%), Gaps = 16/214 (7%)

Query  47    SATSATTPTQQSNVTLALGNRINTD--TALTDYGTNVNGLGINVNRRRRRRRDIWNIE-S  103
              A++  +    S   +A     N +  +   DY +N   +  +++ +R     + +   S
Sbjct  920   DASNDISIPSNSIYPVASVPASNNNPQSTKVDYYSNGPSVIPDLSMKRSVSTPVNHFPAS  979

Query  104   NPVLAGLAANNPVNGSYIAIGNQVFHLPNNLSNVSLNIAS--------GSSGGAGQTVPS  155
              P L      N  N   + I + +    +NL NV++N  +        G S  +     +
Sbjct  980   VPGLRNHPVGNLSNNVTLGIDHPIPREHSNLQNVTMNYNNQFSNANAIGRSQSSMSHSRT  1039

Query  156   NFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAM----DSIVQTLSDAYSNTFKGLFTAAGL  211
               ++    + D ++       ++    P  ++    + I    ++   +T     T++
Sbjct  1040  PIASKSNNMTDLHSVVSDPGSSKSTAYPPLSLFSKSNDINSNKTNQRFSTGTNTVTSSNF  1099

Query  212   NGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIA  245
               ++    + ++           Q A   E   +
Sbjct  1100  QTIDN-ENNVKTPGNKLTDFFQQQSAGWIEGNSS  1132


>☐ref|ZP_06327389.1|  tape measure protein [Staphylococcus aureus subsp. aureus C427]
 gb|EFB46672.1|  tape measure protein [Staphylococcus aureus subsp. aureus C427]
Length=1155

 Score = 37.4 bits (84),  Expect = 3.3, Method: Composition-based stats.
 Identities = 21/187 (11%), Positives = 57/187 (30%), Gaps = 7/187 (3%)

Query  122   AIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQT  181
              I N ++     L N  + +A+         + +        I + +   +  L     T
Sbjct  861   TIWNAIWTFLQTLWNTIVTVATKVWNAITTAISTALQAAWSFISNIWNTIWSFLSGILTT  920

Query  182   LPLDAMDS---IVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQ  238
             +    + +   +V T+SD   +         G+  ++       + +   IQ  V+   +
Sbjct  921   IWNKVVSTFTQVVSTISD-KMSQAWNFIVTKGMQWVSTITSTLINFVNRVIQGFVNVVNK  979

Query  239   GFEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDIL  298
               + +  + N   + +   V        + R ++        +      ++  +
Sbjct  980   VSQGMTNAVNKIKSFIGDFVSAGA---DMIRGLIRGIGQMAGQLVDAAKNVAKKALDAAK  1036

Query  299   AADRGAS  305
             +A    S
Sbjct  1037  SALGIHS  1043


>☐ref|NP_846030.1|  prophage LambdaBa01, membrane protein, putative [Bacillus anthracis
str. Ames]
 ref|YP_020410.1|  prophage LambdaBa01, membrane protein [Bacillus anthracis str.
'Ames Ancestor']
 ref|YP_029752.1|  prophage LambdaBa01, membrane protein, putative [Bacillus anthracis
str. Sterne]
 gb|AAP27516.1|  putative prophage LambdaBa01, membrane protein [Bacillus anthracis
str. Ames]
 gb|AAT32885.1|  putative prophage LambdaBa01, membrane protein [Bacillus anthracis
str. 'Ames Ancestor']
 gb|AAT55803.1|  prophage LambdaBa01, membrane protein, putative [Bacillus anthracis
str. Sterne]
Length=811

 GENE ID: 1086770 BA_3774 | prophage LambdaBa01, membrane protein, putative
[Bacillus anthracis str. Ames] (10 or fewer PubMed links)

 Score = 37.4 bits (84),  Expect = 3.3, Method: Composition-based stats.
 Identities = 17/219 (7%), Positives = 61/219 (27%), Gaps = 25/219 (11%)

Query  161  LPMIPDNYTQAFQKLVTQFQTLPLDAMDSI--VQTLSDAYSNTFKGLFTAAGLNGLNETL  218
               I    +  ++ +     +     +  +  +     ++  +     ++A +   N
Sbjct  414  WNTISSTTSFVWETIKNFLVSCWNGLVAFVMPIFEQIKSWIISVWDTISSATMTVWNAVK  473

Query  219  EHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDC  278
               +S     +  +        + +I ++N  S+                  V
Sbjct  474  NFLQSCWNGLVAFVTPIFNSIKDWIINTWNTISSTT--------------SAVWNTIKSF  519

Query  279  VNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQ  338
            ++  W  +      +  +I  A     +    +++ + + +N   ++
Sbjct  520  LSGLWNSIVSTASSVFNNIKEAISTVWNM---ISSTSSSIWNGIKSTLSNIWEGIKST--  574

Query  339  CYVRAIGNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVA  377
                ++ N  +  +  PV        +  A   +++ V
Sbjct  575  --ASSVWNGLKDAIMTPVRWV--TNAVSGAFEGMKSAVL  609


>☐ref|YP_001004316.1|  tail protein [Staphylococcus phage phiETA2]
 dbj|BAF43871.1|  tail protein [Staphylococcus phage phiETA2]
Length=1155

 GENE ID: 4712350 phiETA2_gp56 | tail protein [Staphylococcus phage phiETA2]

 Score = 37.4 bits (84),  Expect = 3.7, Method: Composition-based stats.
 Identities = 18/186 (9%), Positives = 53/186 (28%), Gaps = 5/186 (2%)

Query  122   AIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQT  181
```

```
              I N ++     L N  + +A+         + +        I + +   +  L     T
Sbjct  861   TIWNAIWTFLQTLWNTIVTVATKVWNAITTAISTALQAAWSFISNIWNTIWSFLSGILTT  920

Query  182   LPLDAMD--SIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQG  239
             +    +   + V +      +         G+  ++       + +   IQ  V+   +
Sbjct  921   IWNKVVSIFTQVVSTISDKMSQAWNFIVTKGMQWVSTITSTLINFVNRVIQGFVNVVNKV  980

Query  240   FEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILA  299
              + +  + N   + +   V        + R ++        +      ++  +      +
Sbjct  981   SQGMTNAVNKIKSFIGDFVSAGA---DMIRGLIRGIGQMAGQLVDAAKNVAKKALDAAKS  1037

Query  300   ADRGAS  305
             A    S
Sbjct  1038  ALGIHS  1043


>☐ref|YP_874001.1|  phage tape measure protein [Staphylococcus phage phiNM]
 ref|YP_001332814.1|  phage tape measure protein [Staphylococcus aureus subsp. aureus
str. Newman]
 gb|ABF73082.1|  phage tape measure protein [Staphylococcus aureus phage phiNM1]
 dbj|BAF68052.1|  phage tape measure protein [Staphylococcus aureus subsp. aureus
str. Newman]
Length=1155

 GENE ID: 4525414 SAPPV1_gp52 | phage tape measure protein
[Staphylococcus phage phiNM] (10 or fewer PubMed links)

 Score = 37.4 bits (84),  Expect = 3.8, Method: Composition-based stats.
 Identities = 18/186 (9%), Positives = 53/186 (28%), Gaps = 5/186 (2%)

Query  122   AIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQT  181
              I N ++     L N  + +A+         + +        I + +   +  L     T
Sbjct  861   TIWNAIWTFLQTLWNTIVTVATKVWNAITTAISTALQAAWSFISNIWNTIWSFLSGILTT  920

Query  182   LPLDAMD--SIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQG  239
             +    +   + V +      +         G+  ++       + +   IQ  V+   +
Sbjct  921   IWNKVVSIFTQVVSTISDKMSQAWNFIVTKGMQWVSTITSTLINFVNRVIQGFVNVVNKV  980

Query  240   FEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILA  299
              + +  + N   + +   V        + R ++        +      ++  +      +
Sbjct  981   SQGMTNAVNKIKSFIGDFVSAGA---DMIRGLIRGIGQMAGQLVDAAKNVAKKALDAAKS  1037

Query  300   ADRGAS  305
             A    S
Sbjct  1038  ALGIHS  1043


>☐ref|ZP_00393935.1|  COG5280: Phage-related minor tail protein [Bacillus anthracis
str. A2012]
 ref|ZP_02217701.1|  putative prophage LambdaBa01, membrane protein [Bacillus anthracis
str. A0488]
 ref|ZP_02394889.1|  putative prophage LambdaBa01, membrane protein [Bacillus anthracis
str. A0442]
   22 more sequence titles
 Length=794

 Score = 37.0 bits (83),  Expect = 3.8, Method: Composition-based stats.
 Identities = 17/219 (7%), Positives = 61/219 (27%), Gaps = 25/219 (11%)

Query  161  LPMIPDNYTQAFQKLVTQFQTLPLDAMDSI--VQTLSDAYSNTFKGLFTAAGLNGLNETL  218
               I    +  ++ +     +     +  +  +     ++  +     ++A +   N
Sbjct  397  WNTISSTTSFVWETIKNFLVSCWNGLVAFVMPIFEQIKSWIISVWDTISSATMTVWNAVK  456

Query  219  EHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDC  278
               +S     +  +        + +I ++N  S+                  V
Sbjct  457  NFLQSCWNGLVAFVTPIFNSIKDWIINTWNTISSTT--------------SAVWNTIKSF  502

Query  279  VNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQ  338
            ++  W  +      +  +I  A     +    +++ + + +N   ++
Sbjct  503  LSGLWNSIVSTASSVFNNIKEAISTVWNM---ISSTSSSIWNGIKSTLSNIWEGIKST--  557

Query  339  CYVRAIGNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVA  377
                ++ N  +  +  PV        +  A   +++ V
Sbjct  558  --ASSVWNGLKDAIMTPVRWV--TNAVSGAFEGMKSAVL  592


>☐ref|NP_371425.1|  hypothetical protein SAV0901 [Staphylococcus aureus subsp. aureus
Mu50]
 ref|YP_001441486.1|  hypothetical protein SAHV_0896 [Staphylococcus aureus subsp.
aureus Mu3]
 ref|ZP_05144291.2|  hypothetical protein SauraM_04455 [Staphylococcus aureus subsp.
aureus Mu50-omega]
 dbj|BAB57063.1|  hypothetical protein [Staphylococcus aureus subsp. aureus Mu50]
 dbj|BAF77779.1|  hypothetical protein [Staphylococcus aureus subsp. aureus Mu3]
Length=1154

 GENE ID: 1120876 SAV0901 | hypothetical protein
[Staphylococcus aureus subsp. aureus Mu50] (10 or fewer PubMed links)

 Score = 37.0 bits (83),  Expect = 4.1, Method: Composition-based stats.
 Identities = 18/186 (9%), Positives = 53/186 (28%), Gaps = 5/186 (2%)

Query  122   AIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQT  181
              I N ++     L N  + +A+         + +        I + +   +  L     T
Sbjct  860   TIWNAIWTFLQTLWNTIVTVATKVWNAITTAISTALQAAWSFISNIWNTIWSFLSGILTT  919

Query  182   LPLDAMD--SIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQG  239
             +    +   + V +      +         G+  ++       + +   IQ  V+   +
Sbjct  920   IWNKVVSIFTQVVSTISDKMSQAWNFIVTKGMQWVSTITSTLINFVNRVIQGFVNVVNKV  979

Query  240   FEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILA  299
              + +  + N   + +   V        + R ++        +      ++  +      +
Sbjct  980   SQGMTNAVNKIKSFIGDFVSAGA---DMIRGLIRGIGQMAGQLVDAAKNVAKKALDAAKS  1036

Query  300   ADRGAS  305
             A    S
Sbjct  1037  ALGIHS  1042
```

```
>☐ref|YP_001246278.1|  Phage-related protein-like protein [Staphylococcus aureus subsp.
aureus JH9]
 ref|YP_001316062.1|  phage tape measure protein [Staphylococcus aureus subsp. aureus
JH1]
 ref|ZP_05682384.1|  phage tail tape measure protein [Staphylococcus aureus A9763]
   12 more sequence titles
 Length=1155

 GENE ID: 5167582 SaurJH9_0901 | Phage-related protein-like protein
[Staphylococcus aureus subsp. aureus JH9]

 Score = 37.0 bits (83),  Expect = 4.2, Method: Composition-based stats.
 Identities = 19/186 (10%), Positives = 54/186 (29%), Gaps = 5/186 (2%)

Query  122   AIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQT  181
              I N ++     L N  + +A+        T+ +        I + +   +  L     T
Sbjct  861   TIWNAIWTFLQTLWNTIVTVATKVWNAITTTISTALQAAWSFISNIWNTIWSFLSGILTT  920

Query  182   LPLDAMD--SIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQG  239
             +    +   + V +      +         G+  ++       + +   IQ  V+   +
Sbjct  921   IWNKVVSIFTQVVSTISDKMSQAWNFIVTKGMQWVSTITSTLINFVNRVIQGFVNVVNKV  980

Query  240   FEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILA  299
              + +  + N   + +   V        + R ++        +      ++  +      +
Sbjct  981   SQGMTNAVNKIKSFIGDFVSAGA---DMIRGLIRGIGQMAGQLVDAAKNVAKKALDAAKS  1037

Query  300   ADRGAS  305
             A    S
Sbjct  1038  ALGIHS  1043
```

```
>☐ref|XP_001818890.1|  hypothetical protein [Aspergillus oryzae RIB40]
 dbj|BAE56888.1|  unnamed protein product [Aspergillus oryzae]
Length=630

 GENE ID: 5990861 AO090001000363 | hypothetical protein
[Aspergillus oryzae RIB40] (10 or fewer PubMed links)

 Score = 37.0 bits (83),  Expect = 4.2, Method: Composition-based stats.
 Identities = 19/163 (11%), Positives = 36/163 (22%), Gaps = 15/163 (9%)

Query  32   WNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGTNVNGLGINVNRR  91
            W+          P         P   S          +  +      T +       +
Sbjct  362  WSFPSNAPAYAPPAHLGK--EFPPMPSKKRTLDDYVEDQPSKRVAMPTTLP-----TSMS  414

Query  92   RRRRRDIWNIESNPVLAGLAANNPVNGSYIAIGNQVFHLPNNLSNVSLNIASGSSGGAGQ  151
                  +  + + P +    +    N    ++   +  LP      S NIA G
Sbjct  415  TLPSAALAGVPTLPPVLAATSAPSAN----SLSGSISRLPRPNFPPSSNIAPGIPTSVLP  470

Query  152  TVPSNFSTFLPMIPDNYTQAF--QKLVTQFQTLPLDAMDSIVQ  192
                           N +  +  Q   TQ   +     +  V
Sbjct  471  FPSVANRAM--HSVYNPSTNWVPQLPPTQVPPITNGLYNPSVS  511
```

```
>☐ref|ZP_06327622.1|  tape measure protein [Staphylococcus aureus A9765]
 gb|EFB99593.1|  tape measure protein [Staphylococcus aureus A9765]
Length=1154

 Score = 37.0 bits (83),  Expect = 4.3, Method: Composition-based stats.
 Identities = 18/186 (9%), Positives = 53/186 (28%), Gaps = 5/186 (2%)

Query  122   AIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQT  181
              I N ++     L N  + +A+         + +        I + +   +  L     T
Sbjct  860   TIWNAIWTFLQTLWNTIVTVATKVWNAITTAISTALQAAWSFISNIWNTIWSFLSGILTT  919

Query  182   LPLDAMD--SIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQG  239
             +    +   + V +      +         G+  ++       + +   IQ  V+   +
Sbjct  920   IWNKVVSIFTQVVSTISDKMSQAWNFIVTKGMQWVSTITSTLINFVNRVIQGFVNVVNKV  979

Query  240   FEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILA  299
              + +  + N   + +   V        + R ++        +      ++  +      +
Sbjct  980   SQGMTNAVNKIKSFIGDFVSAGA---DMIRGLIRGIGQMAGQLVDAAKNVAKKALDAAKS  1036

Query  300   ADRGAS  305
             A    S
Sbjct  1037  ALGIHS  1042
```

```
>☐ref|YP_001332062.1|  phage tape measure protein [Staphylococcus aureus subsp. aureus
str. Newman]
 gb|ABF73148.1|  phage tape measure protein [Staphylococcus aureus phage phiNM2]
 dbj|BAF67300.1|  phage tape measure protein [Staphylococcus aureus subsp. aureus
str. Newman]
Length=1154

 GENE ID: 5330632 NWMN_1028 | phage tape measure protein
[Staphylococcus aureus subsp. aureus str. Newman] (10 or fewer PubMed links)

 Score = 37.0 bits (83),  Expect = 4.4, Method: Composition-based stats.
 Identities = 19/186 (10%), Positives = 54/186 (29%), Gaps = 5/186 (2%)

Query  122   AIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQT  181
              I N ++     L N  + +A+        T+ +        I + +   +  L     T
Sbjct  860   TIWNAIWTFLQTLWNTIVTVATKVWNAITTTISTALQAAWSFISNIWNTIWSFLSGILTT  919

Query  182   LPLDAMD--SIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQG  239
             +    +   + V +      +         G+  ++       + +   IQ  V+   +
Sbjct  920   IWNKVVSIFTQVVSTISDKMSQAWNFIVTKGMQWVSTITSTLINFVNRVIQGFVNVVNKV  979

Query  240   FEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILA  299
              + +  + N   + +   V        + R ++        +      ++  +      +
Sbjct  980   SQGMTNAVNKIKSFIGDFVSAGA---DMIRGLIRGIGQMAGQLVDAAKNVAKKALDAAKS  1036

Query  300   ADRGAS  305
             A    S
Sbjct  1037  ALGIHS  1042
```

```
>☐ref|YP_001949854.1| putative tape measure protein [Staphylococcus phage phiMR25]
 dbj|BAG48153.1| putative tape measure protein [Staphylococcus phage phiMR25]
Length=1154

 GENE ID: 6370091 orf56 | putative tape measure protein
[Staphylococcus phage phiMR25]

 Score = 37.0 bits (83),  Expect = 4.4, Method: Composition-based stats.
 Identities = 18/186 (9%), Positives = 53/186 (28%), Gaps = 5/186 (2%)

Query  122   AIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQT  181
              I N ++     L N  + +A+         + +        I + +   +  L     T
Sbjct  860   TIWNAIWTFLQTLWNTIVTVATKVWNAITTAISTALQAAWSFISNIWNTIWSFLSGILTT  919

Query  182   LPLDAMD--SIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQG  239
             +    +   + V +      +         G+  ++       + +   IQ  V+   +
Sbjct  920   IWNKVVSIFTQVVSTISDKMSQAWNFIVTKGMQWVSTITSTLINFVNRVIQGFVNVVNKV  979

Query  240   FEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILA  299
              + +  + N   + +   V        + R ++        +      ++  +      +
Sbjct  980   SQGMTNAVNKIKSFIGDFVSAGA---DMIRGLIRGIGQMAGQLVDAAKNVAKKALDAAKS  1036

Query  300   ADRGAS  305
             A    S
Sbjct  1037  ALGIHS  1042


>☐ref|YP_239735.1| ORF001 [Staphylococcus phage 85]
 gb|AAX90911.1| ORF001 [Staphylococcus phage 85]
Length=1154

 GENE ID: 5130866 85ORF001 | ORF001 [Staphylococcus phage 85]
(10 or fewer PubMed links)

 Score = 37.0 bits (83),  Expect = 4.7, Method: Composition-based stats.
 Identities = 19/186 (10%), Positives = 54/186 (29%), Gaps = 5/186 (2%)

Query  122   AIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQT  181
              I N ++     L N  + +A+        T+ +        I + +   +  L     T
Sbjct  860   TIWNAIWTFLQTLWNTIVTVATKVWNAITTTISTALQAAWSFISNIWNTIWSFLSGILTT  919

Query  182   LPLDAMD--SIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQG  239
             +    +   + V +      +         G+  ++       + +   IQ  V+   +
Sbjct  920   IWNKVVSIFTQVVSTISDKMSQAWNFIVTKGMQWVSTITSTLINFVNRVIQGFVNVVNKV  979

Query  240   FEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILA  299
              + +  + N   + +   V        + R ++        +      ++  +      +
Sbjct  980   SQGMTNAVNKIKSFIGDFVSAGA---DMIRGLIRGIGQMAGQLVDAAKNVAKKALDAAKS  1036

Query  300   ADRGAS  305
             A    S
Sbjct  1037  ALGIHS  1042


>☐ref|XP_001870945.1| conserved hypothetical protein [Culex quinquefasciatus]
 gb|EDS34934.1| conserved hypothetical protein [Culex quinquefasciatus]
Length=286

 GENE ID: 6043347 CpipJ_CPIJ010754 | hypothetical protein
[Culex quinquefasciatus]

 Score = 37.0 bits (83),  Expect = 4.7, Method: Composition-based stats.
 Identities = 26/219 (11%), Positives = 61/219 (27%), Gaps = 25/219 (11%)

Query  184  LDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQV  243
               + ++     +    T + +        L++     +  I+  IQ+           +
Sbjct  89   SGFLSTVQYF--NQSLQTGEQMVNMQISTDLSQLTSTLQQAIDTIIQTYSSSLYTLSSSM  146

Query  244  IASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILAADRG  303
               FN     +  CV  +++        +GK   C+ ++      I   I   I
Sbjct  147  NGVFN-----MDICVAKNVTQMINIPSSIGKLGSCLQQEVTTAKAITPIIVSTIKMFKND  201

Query  304  ASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQCYVRAIGNFPQSLLFLPVSLTIEGG  363
                   L  C                        C  +   N    +  +  +L++
Sbjct  202  LKALNAQLQIC------------------QPTSNSCIDQYFMNIYSEMSTIQSALSMVQQ  243

Query  364  KLYAALSSLQTDVALCASEMGLEIGLVTAQVSSKIAFCQ  402
             + +A     T   +C+  +  +I      + S ++ C
Sbjct  244  FISSAQQDASTRNRICSDLVTGDIQDAIQNIQSAVSGCS  282


>☐ref|YP_239657.1| ORF001 [Staphylococcus phage 53]
 ref|YP_001285370.1| tape measure protein [Staphylococcus phage 80alpha]
 ref|ZP_05642383.1|  phage tail tape measure protein [Staphylococcus aureus A9781]
 ref|ZP_05684633.1|  tape measure protein [Staphylococcus aureus A9719]
 gb|AAX90832.1| ORF001 [Staphylococcus phage 53]
 gb|ABF71627.1| tape measure protein [Staphylococcus phage 80alpha]
 gb|EEV25716.1|  phage tail tape measure protein [Staphylococcus aureus A9781]
 gb|EEV66732.1|  tape measure protein [Staphylococcus aureus A9719]
Length=1154

 GENE ID: 5132829 53ORF001 | ORF001 [Staphylococcus phage 53]
(10 or fewer PubMed links)

 Score = 36.6 bits (82),  Expect = 5.0, Method: Composition-based stats.
 Identities = 18/186 (9%), Positives = 53/186 (28%), Gaps = 5/186 (2%)

Query  122   AIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQT  181
              I N ++     L N  + +A+         + +        I + +   +  L     T
Sbjct  860   TIWNAIWTFLQTLWNTIVTVATKVWNAITTAISTALQAAWSFISNIWNTIWSFLSGILTT  919

Query  182   LPLDAMD--SIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQG  239
             +    +   + V +      +         G+  ++       + +   IQ  V+   +
Sbjct  920   IWNKVVSIFTQVVSTISDKMSQAWNFIVTKGMQWVSTITSTLINFVNRVIQGFVNVVNKV  979

Query  240   FEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILA  299
              + +  + N   + +   V        + R ++        +      ++  +      +
```

```
Sbjct  980   SQGMTNAVNKIKSFIGDFVSAGA---DMIRGLIRGIGQMAGQLVDAAKNVAKKALDAAKS  1036

Query  300   ADRGAS  305
             A    S
Sbjct  1037  ALGIHS  1042


>☐ref|XP_002009072.1| GI13843 [Drosophila mojavensis]
 gb|EDW19548.1| GI13843 [Drosophila mojavensis]
Length=1694

 GENE ID: 6583404 Dmoj\GI13843 | GI13843 gene product from transcript GI13843-RA
[Drosophila mojavensis] (10 or fewer PubMed links)

 Score = 36.6 bits (82),  Expect = 5.7, Method: Composition-based stats.
 Identities = 23/174 (13%), Positives = 51/174 (29%), Gaps = 13/174 (7%)

Query  42   QLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGTNVNGLGINVNRRRRRRRDIWNI  101
                + +   +  +  ++    + N+++    L    T   G         +  +
Sbjct  187  AGHTSESNDESHSSDSNS---GVANKVDLVAELVSMETTALGPTSTSLGHSKPSQTPTPS  243

Query  102  ESNPVLAGLAANNPVNGSYIA-IGNQVFH---LPNNLSNVSL--NIASGSSGGAGQTVPS  155
                     AA    N S I+ +   +     LPN+ S  ++  N            +P+
Sbjct  244  TEQSKQTQTAAPKKTNSSNISLLSKPLNIQSPLPNSQSTGTILPNQKPKLPYPESPIMPN  303

Query  156  NFSTFLPMIPDNYTQAFQKLVTQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAA  209
                  P   ++       + +       +   + +  LS  Y+N  K    AA
Sbjct  304  VSQPTDPHTKESPQ----TVSSDKNRDFEETTITYLSNLSKMYANKAKRSIKAA  353


>☐ref|YP_002332525.1| putative tape measure protein [Staphylococcus phage phiSauS-IPLA88]
 gb|ACJ64578.1| gp50 [Staphylococcus phage phiSauS-IPLA88]
Length=1155

 GENE ID: 7057011 SauSIPLA88_gp50 | putative tape measure protein
[Staphylococcus phage phiSauS-IPLA88]

 Score = 36.6 bits (82),  Expect = 6.3, Method: Composition-based stats.
 Identities = 18/186 (9%), Positives = 53/186 (28%), Gaps = 5/186 (2%)

Query  122   AIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQT  181
              I N ++     L N  + +A+         + +        I + +   +  L     T
Sbjct  861   TIWNAIWTFLQTLWNTIVTVATTVWNAITTAISTALQAAWSFISNIWNTIWSFLSGILTT  920

Query  182   LPLDAMD--SIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQG  239
             +    +   + V +      +         G+  ++       + +   IQ  V+   +
Sbjct  921   IWNKVVSIFTQVVSTISDKMSQAWNFIVTKGMQWVSTITSTLINFVNRVIQGFVNVVNKV  980

Query  240   FEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILA  299
              + +  + N   + +   V        + R ++        +      ++  +      +
Sbjct  981   SQGMTNAVNKIKSFIGDFVSAGA---DMIRGLIRGIGQMAGQLVDAAKNVAKKALDAAKS  1037

Query  300   ADRGAS  305
             A    S
Sbjct  1038  ALGIHS  1043


>☐ref|YP_084997.1| prophage membrane protein [Bacillus cereus E33L]
 gb|AAU16851.1| possible prophage membrane protein [Bacillus cereus E33L]
Length=726

 GENE ID: 3023924 BCZK3413 | prophage membrane protein [Bacillus cereus E33L]
(10 or fewer PubMed links)

 Score = 36.6 bits (82),  Expect = 6.4, Method: Composition-based stats.
 Identities = 17/219 (7%), Positives = 61/219 (27%), Gaps = 25/219 (11%)

Query  161  LPMIPDNYTQAFQKLVTQFQTLPLDAMDSI--VQTLSDAYSNTFKGLFTAAGLNGLNETL  218
               I    +  ++ +     +     +  +  +     ++  +     ++A +   N
Sbjct  329  WNTISSTTSFVWETIKNFLVSCWNGLVAFVMPIFEQIKSWIISVWDTISSATMTVWNAVK  388

Query  219  EHGRSVIEDGIQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDC  278
               +S     +  +        + +I ++N  S+                  V
Sbjct  389  NFLQSCWNGLVAFVTPIFNSIKDWIINTWNTISSTT--------------SAVWNTIKSF  434

Query  279  VNRKWRQLSDIGGRIAGDILAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQ  338
            ++  W  +      +  +I  A     +    +++ + + +N   ++
Sbjct  435  LSGLWNSIVSTASSVFNNIKEAISTVWNM---ISSTSSSIWNGIKSTLSNIWEGIKST--  489

Query  339  CYVRAIGNFPQSLLFLPVSLTIEGGKLYAALSSLQTDVA  377
                ++ N  +  +  PV        +  A   +++ V
Sbjct  490  --ASSVWNGLKDAIMTPVRWV--TNAVSGAFEGMKSAVL  524


>☐ref|XP_002380327.1| mucin, putative [Aspergillus flavus NRRL3357]
 gb|EED49946.1| mucin, putative [Aspergillus flavus NRRL3357]
Length=565

 GENE ID: 7919175 AFLA_067680 | mucin, putative [Aspergillus flavus NRRL3357]

 Score = 36.2 bits (81),  Expect = 6.6, Method: Composition-based stats.
 Identities = 19/163 (11%), Positives = 36/163 (22%), Gaps = 15/163 (9%)

Query  32   WNPQPPRTINQLPVTSATSATTPTQQSNVTLALGNRINTDTALTDYGTNVNGLGINVNRR  91
            W+          P         P   S          +  +      T +       +
Sbjct  298  WSFPSNAPAYAPPAHLGK--EFPPMPSKKRTLDDYVEDQPSKRVAMPTTLP-----TSMS  350

Query  92   RRRRRDIWNIESNPVLAGLAANNPVNGSYIAIGNQVFHLPNNLSNVSLNIASGSSGGAGQ  151
                  +  + + P +    +    N    ++   +  LP      S NIA G
Sbjct  351  TLPSAALAGVPTLPPVLAATSAPSAN----SLSGSISRLPRPNFPPSSNIAPGIPTSVLP  406

Query  152  TVPSNFSTFLPMIPDNYTQAF--QKLVTQFQTLPLDAMDSIVQ  192
                           N +  +  Q   TQ   +     +  V
Sbjct  407  FPSVANRAM--HSVYNPSTNWVPQLPPTQVPPITNGLYNPSVS  447


>☐ref|XP_001026080.1| SAM domain containing protein [Tetrahymena thermophila]
```

```
 gb|EAS05835.1|  SAM domain containing protein [Tetrahymena thermophila SB210]
Length=694

 GENE ID: 7837189 TTHERM_00969610 | SAM domain containing protein
[Tetrahymena thermophila]

 Score = 36.2 bits (81),  Expect = 8.0, Method: Composition-based stats.
 Identities = 28/264 (10%), Positives = 76/264 (28%), Gaps = 22/264 (8%)

Query  62   LALGNRINTDT---ALTDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVNG  118
               G+R  T T    + +Y   V         ++         +       L   N
Sbjct  47   RTSGSRWKTSTTDQPIKNYANQVLDKKNLNLTQQSSNTSSSTSKPVIKPPTLKIGNYTKP  106

Query  119  SYIAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQ  178
                + +++   P+NL N               +  +  ++      +   +    L +
Sbjct  107  ETNQVQSKILKRPSNLPNSR-------PSSNQSSRSALNTSLNGKEKEKGKENENNLNSS  159

Query  179  FQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRS----VIE------DG  228
            F+     + D+ VQ  + +   +F             +T ++       V
Sbjct  160  FRKTLNQSFDNSVQQQASSSMTSFTQQTDTKEDKEEIQTDKNNSEYDFLVQNPNRKPRKV  219

Query  229  IQSIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSD  288
            I++ +++ +Q   QV A     +      +        + + V    +     +++Q  +
Sbjct  220  IETKLNELSQNTTQVNAYSQKKTQAESNLLKHLNQQKELQKEVHQFLSTVNLLQYQQAFE  279

Query  289  IGGRIAGDILAADRGASDFLGNLT  312
                   D+       ++ + ++
Sbjct  280  --DNGFDDLETILEINTEIMNSMK  301


>☐ref|YP_417175.1|  phage-like protein [Staphylococcus aureus RF122]
 emb|CAI81402.1|  hypothetical phage-related protein [Staphylococcus aureus RF122]
Length=1154

 GENE ID: 3795178 SAB1713c | phage-like protein [Staphylococcus aureus RF122]
(10 or fewer PubMed links)

 Score = 36.2 bits (81),  Expect = 8.1, Method: Composition-based stats.
 Identities = 17/183 (9%), Positives = 52/183 (28%), Gaps = 5/183 (2%)

Query  125   NQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQTLPL  184
             N ++     L N  + +A+         + +        I + +   +  L     T+
Sbjct  863   NAIWTFLQTLWNTIVTVATTVWNAITTAISTALQAAWSFISNIWNTIWSFLSGILTTIWN  922

Query  185   DAMD--SIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQ  242
               +   + V +      +         G+  ++       + +   IQ  V+   +  +
Sbjct  923   KVVSIFTQVVSTISDKMSQAWNFIVTKGMQWVSTITSTLINFVNRVIQGFVNVVNKVSQG  982

Query  243   VIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTDCVNRKWRQLSDIGGRIAGDILAADR  302
             +  + N   + +   V        + R ++        +      ++  +      +A
Sbjct  983   MTNAVNKIKSFIGDFVSAGA---DMIRGLIRGIGQMAGQLVDAAKNVAKKALDAAKSALG  1039

Query  303   GAS  305
               S
Sbjct  1040  IHS  1042


>☐ref|XP_002720066.1|  PREDICTED: inter-alpha (globulin) inhibitor H5-like [Oryctolagus
cuniculus]
Length=1320

 GENE ID: 100343166 LOC100343166 | inter-alpha (globulin) inhibitor H5-like
[Oryctolagus cuniculus]

 Score = 36.2 bits (81),  Expect = 8.4, Method: Composition-based stats.
 Identities = 35/288 (12%), Positives = 69/288 (23%), Gaps = 33/288 (11%)

Query  18    IIPWYVDAFLFTTPWNPQPPRTINQLPVTSATS---ATTPTQQSNVTLALGNRINTDTAL  74
                W             +P    +Q   +   S     TP        +   R  T
Sbjct  843   PKSWITSPKTPKITSLFKPSAPPHQSSTSLLLSKPRTLTPHTPKTPLPSRSARPRTPRLQ  902

Query  75    TDYGTNVNGLGINVNRRRRRRRDIWNIESNPVLAGLAANNPVN----------GSYIAI-  123
                           V                P    + + +P            GS
Sbjct  903   NLSTFPTTVSSPTVPSTTVTSTGFGEPLPTPFTPTVPSLSPTGRLWHQHDLLLGSQSTRK  962

Query  124   --GNQVFHLPN-NLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLVTQFQ  180
               G  V  +P  +L N S     GS       +PS+       +     +      +  +
Sbjct  963   VLGPSVAGVPTMSLPNSSRPTPEGSPPNVPVLLPSSTLPEAISLFLLPEELKLLSESMVE  1022

Query  181   TLPLDAMDSIVQTL--------SDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQ--  230
                ++++D              S  +++  + +   AG      +++  RS +E
Sbjct  1023  AKFVESLDPPAFYTFLTPDEDGSPHWTDKSEEILGGAG-----GSMKSQRSSVELAKDML  1077

Query  231   -SIVDQGAQGFEQVIASFNDSSARVQQCVGPDLSPTSIGRRVVGKGTD  277
              SI    +              +  + C   D  P  + + +
Sbjct  1078  PSIFTFSSSVDGDPHFVVRIPHSEERICFTLDGRPGDVLQLIEDPKAG  1125


>☐ref|ZP_04675278.1|  conserved hypothetical protein [Lactobacillus jensenii 1153]
 gb|EEQ69064.1|  conserved hypothetical protein [Lactobacillus jensenii 1153]
Length=2110

 Score = 35.8 bits (80),  Expect = 8.7, Method: Composition-based stats.
 Identities = 17/161 (10%), Positives = 38/161 (23%), Gaps = 12/161 (7%)

Query  189   SIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFN  248
             +    L +A ++      T    N  ++T       ++       D  A   ++V  + N
Sbjct  1623  TDKSKLDEAITDANNTKLTDKYNNASDDTKSKFDEALKKAENVKNDSNAT-QKEVDDATN  1681

Query  249   DSSARVQQCVGPDLSPTSIGRRVV-----------GKGTDCVNRKWRQLSDIGGRIAGDI  297
             +         G     + +   +               +D    K+ +       +  +
Sbjct  1682  NLKQAQNDLDGQTTDKSKLDEAITDANNTKSTDKYNNASDDTKSKFDEALKKAEEVKNNS  1741

Query  298   LAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQ  338
              A  +   D   NL              S
Sbjct  1742  NATQKEVDDATNNLKQAQNNLDGQTTDKSKLDEAITDANNT  1782


>☐ref|ZP_05414794.1|  putative fibronectin type III domain protein [Bacteroides finegoldii
DSM 17565]
```

```
 gb|EEX46343.1|  putative fibronectin type III domain protein [Bacteroides finegoldii
DSM 17565]
Length=611

 Score = 35.8 bits (80),  Expect = 9.7, Method: Composition-based stats.
 Identities = 32/261 (12%), Positives = 64/261 (24%), Gaps = 23/261 (8%)

Query  2    NRSMSVVLALLIAAIGIIPWYVDAFLFTTPWNPQPPRTINQLPVTSATSATTPTQQSNVT  61
               M  ++          P      L   P   Q   T             TP+   + T
Sbjct  221  MTVMPEMVDFTSV----TPTKTSVSLVWDPEAIQTGSTTVSHYAWCGGD-RTPSVSDHYT  275

Query  62   LALGNRINTDTALTDYGTNVNGLGINVNR--RRRRRRDIWNIESNPVLAGLAA---NNPV  116
                 +I+      D         + + R    R        +  P    L +
Sbjct  276  ALTAEQISQGQLNFDGLEPSTTYTVALMRGTYVRALTTFTTAKGIPSGYTLVSVTDITTW  335

Query  117  NGSYIAIGNQVFHLPNNLSNVSLNIASGSSGGAGQTVPSNFSTFLPMIPDNYTQAFQKLV  176
            N +    G     +P N       I           +               +   +
Sbjct  336  NEAISKTGKVAVLIPENTDLDLTGITPEIPQSITSLLIWGADIEGNPTNTKPSIKTKGFN  395

Query  177  TQFQTLPLDAMDSIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQG  236
                      + ++       YSN       +AG   +++  +   ++    I+S V
Sbjct  396  ------FNGTLGTVEFYNLHLYSN------GSAGNYIVDQ-KKADNNITNFSIESCVIDE  442

Query  237  AQGFEQVIASFNDSSARVQQC  257
            A+G  ++  +    S  ++ C
Sbjct  443  ARGLFRIRNTGVWQSITIKDC  463


>☐gb|ACA52029.1|  surface anchor protein [Lactobacillus jensenii]
Length=1903

 Score = 35.8 bits (80),  Expect = 9.7, Method: Composition-based stats.
 Identities = 17/161 (10%), Positives = 38/161 (23%), Gaps = 12/161 (7%)

Query  189   SIVQTLSDAYSNTFKGLFTAAGLNGLNETLEHGRSVIEDGIQSIVDQGAQGFEQVIASFN  248
             +    L +A ++      T    N  ++T       ++       D  A   ++V  + N
Sbjct  1416  TDKSKLDEAITDANNTKLTDKYNNASDDTKSKFDEALKKAENVKNDSNAT-QKEVDDATN  1474

Query  249   DSSARVQQCVGPDLSPTSIGRRVV-----------GKGTDCVNRKWRQLSDIGGRIAGDI  297
             +         G     + +   +               +D    K+ +       +  +
Sbjct  1475  NLKQAQNDLDGQTTDKSKLDEAITDANNTKSTDKYNNASDDTKSKFDEALKKAEEVKNNS  1534

Query  298   LAADRGASDFLGNLTTCNGANFNAADASSGGTGTRNALRRQ  338
              A  +   D   NL              S
Sbjct  1535  NATQKEVDDATNNLKQAQNNLDGQTTDKSKLDEAITDANNT  1575
```

☐ Select All   Get selected sequences   Distance tree of results   Multiple alignment