

BLAST

Basic Local Alignment Search Tool

NCBI/ [BLAST](#)/ [blastp suite](#)/ [Formatting Results](#) - RWD5BE8C01N

C.merolae | query NCBI_tblastn Sell (118)

Query ID	lcl 17921	Database Name	nr
Description	C.merolae query NCBI_tblastn Sell	Description	All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+F excluding environmental samples from WGS projects
Molecule type	amino acid	Program	BLASTP 2.2.25+
Query Length	118		

Descriptions

Legend for links to other resources: [U](#) UniGene [E](#) GEO [G](#) Gene [S](#) Structure [M](#) Map Viewer [P](#) PubChem BioAssay

Accession	Description	Max score	Total score	Query coverage	E value	Links
XP_002326159.1	predicted protein [Populus trichocarpa] >gb EEE71829.1 predicted protein [Populus trichocarpa]	73.9	73.9	100%	6e-12	UG
NP_001049763.1	Os03g0284600 [Oryza sativa Japonica Group] >gb ABF95344.1 UPF0308 protein, chloroplast precursor, putative, expressed [Oryza sativa Japonica Group] >dbj BAF11677.1 Os03g0284600 [Oryza sativa Japonica Group] >gb EEE58829.1 hypothetical protein OsJ_10400 [Oryza sativa Japonica Group]	72.0	72.0	91%	3e-11	UG
EEC75002.1	hypothetical protein Osl_11064 [Oryza sativa Indica Group]	71.6	71.6	91%	3e-11	
XP_002881499.1	hypothetical protein ARALYDRAFT_482716 [Arabidopsis lyrata subsp. lyrata] >gb EFH57758.1 hypothetical protein ARALYDRAFT_482716 [Arabidopsis lyrata subsp. lyrata]	71.6	71.6	100%	4e-11	G
NP_001145525.1	hypothetical protein LOC100278941 [Zea mays] >gb ACG48189.1 hypothetical protein [Zea mays]	71.2	71.2	91%	4e-11	UG
NP_030274.1	unknown protein [Arabidopsis thaliana] >sp Q9ZUU2.2 U308_ARATH RecName: Full=UPF0308 protein At2g37240, chloroplastic; Flags: Precursor >gb AAK91362.1 At2g37240/F3G5.3 [Arabidopsis thaliana] >gb AAC98045.2 expressed protein [Arabidopsis thaliana] >gb AAM67197.1 unknown [Arabidopsis thaliana] >gb AAP21145.1 At2g37240/F3G5.3 [Arabidopsis thaliana]	70.5	70.5	100%	7e-11	UG

Accession	Description	Max score	Total score	Query coverage	E value	Links
XP_002334170.1	predicted protein [Populus trichocarpa] >gb EEF07066.1 predicted protein [Populus trichocarpa]	69.7	69.7	89%	1e-10	UG
XP_002468056.1	hypothetical protein SORBIDRAFT_01g038790 [Sorghum bicolor] >gb EER95054.1 hypothetical protein SORBIDRAFT_01g038790 [Sorghum bicolor]	69.7	69.7	91%	1e-10	UG
ACR37670.1	unknown [Zea mays]	69.3	69.3	91%	2e-10	
ACN35248.1	unknown [Zea mays]	68.6	68.6	91%	2e-10	G
XP_002965388.1	hypothetical protein SELMODRAFT_68006 [Selaginella moellendorffii] >gb EFJ34226.1 hypothetical protein SELMODRAFT_68006 [Selaginella moellendorffii]	67.8	67.8	100%	5e-10	UG
ACU20039.1	unknown [Glycine max]	67.8	67.8	100%	5e-10	
XP_002977235.1	hypothetical protein SELMODRAFT_58056 [Selaginella moellendorffii] >gb EFJ21844.1 hypothetical protein SELMODRAFT_58056 [Selaginella moellendorffii]	67.4	67.4	100%	5e-10	UG
ABK23985.1	unknown [Picea sitchensis]	65.9	65.9	100%	2e-09	
XP_002273449.1	PREDICTED: hypothetical protein [Vitis vinifera] >emb CBI32627.3 unnamed protein product [Vitis vinifera]	65.1	65.1	100%	3e-09	UGM
XP_002525198.1	conserved hypothetical protein [Ricinus communis] >gb EEF37164.1 conserved hypothetical protein [Ricinus communis]	63.5	63.5	100%	7e-09	G
XP_002866692.1	hypothetical protein ARALYDRAFT_496819 [Arabidopsis lyrata subsp. lyrata] >gb EFH42951.1 hypothetical protein ARALYDRAFT_496819 [Arabidopsis lyrata subsp. lyrata]	52.0	52.0	99%	3e-05	G
NP_201385.2	unknown protein [Arabidopsis thaliana] >gb AAM20692.1 unknown protein [Arabidopsis thaliana] >gb AAN15652.1 unknown protein [Arabidopsis thaliana]	51.6	51.6	99%	4e-05	UG
ACU20324.1	unknown [Glycine max]	50.8	50.8	100%	5e-05	
XP_002320012.1	predicted protein [Populus trichocarpa] >gb EEE98327.1 predicted protein [Populus trichocarpa]	50.1	50.1	100%	9e-05	UG
CBI33071.3	unnamed protein product [Vitis vinifera]	49.7	49.7	100%	1e-04	
ADE77692.1	unknown [Picea sitchensis]	49.3	49.3	100%	1e-04	
XP_002263922.1	PREDICTED: hypothetical protein [Vitis vinifera] >emb CAN81556.1 hypothetical protein VITISV_040398 [Vitis vinifera]	49.3	49.3	100%	2e-04	UGM
CBI33293.3	unnamed protein product [Vitis vinifera]	49.3	49.3	100%	2e-04	
ACN32054.1	unknown [Zea mays]	49.3	49.3	95%	2e-04	G
XP_003130905.1	PREDICTED: UPF0308 protein C9orf21 homolog isoform 2 [Sus scrofa]	48.1	48.1	98%	4e-04	UGM
XP_001153024.1	PREDICTED: similar to TPA_exp: C9ORF21 isoform 1 [Pan troglodytes]	48.1	48.1	98%	4e-04	UGM

Accession	Description	Max score	Total score	Query coverage	E value	Links
NP_001051153.1	Os03g0729200 [Oryza sativa Japonica Group] >gb AAO38464.1 hypothetical protein [Oryza sativa Japonica Group] >gb ABF98679.1 expressed protein [Oryza sativa Japonica Group] >dbj BAF13067.1 Os03g0729200 [Oryza sativa Japonica Group] >gb EAY91738.1 hypothetical protein OsI_13379 [Oryza sativa Indica Group] >dbj BAG94118.1 unnamed protein product [Oryza sativa Japonica Group] >gb EEE59858.1 hypothetical protein OsJ_12440 [Oryza sativa Japonica Group]	47.8	47.8	100%	4e-04	UG
ACU18761.1	unknown [Glycine max]	47.8	47.8	100%	5e-04	
XP_002928763.1	PREDICTED: UPF0308 protein C9orf21-like, partial [Ailuropoda melanoleuca]	47.8	47.8	81%	5e-04	GM
EFB20490.1	hypothetical protein PANDA_018799 [Ailuropoda melanoleuca]	47.8	47.8	81%	5e-04	
EAW92663.1	chromosome 9 open reading frame 21, isoform CRA_c [Homo sapiens]	47.8	47.8	98%	5e-04	G
ACU24130.1	unknown [Glycine max]	47.8	47.8	100%	5e-04	
XP_002666723.1	PREDICTED: UPF0308 protein C9orf21 homolog [Danio rerio]	47.8	47.8	68%	5e-04	UGM
XP_002742741.1	PREDICTED: UPF0308 protein C9orf21-like [Callithrix jacchus]	47.4	47.4	98%	7e-04	GM
XP_002463941.1	hypothetical protein SORBIDRAFT_01g009340 [Sorghum bicolor] >gb EER90939.1 hypothetical protein SORBIDRAFT_01g009340 [Sorghum bicolor]	47.0	47.0	95%	8e-04	UG
XP_002463942.1	hypothetical protein SORBIDRAFT_01g009350 [Sorghum bicolor] >gb EER90940.1 hypothetical protein SORBIDRAFT_01g009350 [Sorghum bicolor]	47.0	47.0	95%	8e-04	UG
XP_003130904.1	PREDICTED: UPF0308 protein C9orf21 homolog isoform 1 [Sus scrofa]	47.0	47.0	81%	0.001	UGM
NP_001069904.1	hypothetical protein LOC616897 [Bos taurus] >sp Q148E0.1 CI021_BOVIN RecName: Full=UPF0308 protein C9orf21 homolog >gb AAI18426.1 Chromosome 9 open reading frame 21 ortholog [Bos taurus] >gb DAA26600.1 hypothetical protein LOC616897 [Bos taurus]	47.0	47.0	81%	0.001	UGM
XP_002269002.1	PREDICTED: hypothetical protein, partial [Vitis vinifera]	46.6	46.6	100%	0.001	UGM
NP_001170360.1	hypothetical protein LOC100384338 [Zea mays] >gb ACN36745.1 unknown [Zea mays]	46.2	46.2	100%	0.001	UG
XP_001493974.2	PREDICTED: similar to UPF0308 protein C9orf21 [Equus caballus]	46.2	46.2	68%	0.001	UGM
CBI33289.3	unnamed protein product [Vitis vinifera]	46.2	46.2	100%	0.001	
ACN25853.1	unknown [Zea mays]	46.2	46.2	95%	0.001	G
XP_002820049.1	PREDICTED: UPF0308 protein C9orf21-like [Pongo abelii]	46.2	46.2	82%	0.001	GM
XP_520707.2	PREDICTED: similar to TPA_exp: C9ORF21 isoform 2 [Pan troglodytes]	46.2	46.2	82%	0.001	UGM

Accession	Description	Max score	Total score	Query coverage	E value	Links
NP_714542.1	hypothetical protein LOC195827 [Homo sapiens] >sp Q7RTV5.1 C1021_HUMAN RecName: Full=UPF0308 protein C9orf21 >tpg DAA00065.1 TPA_exp: C9ORF21 [Homo sapiens] >emb CAI40534.1 novel protein [Homo sapiens] >gb EAW92662.1 chromosome 9 open reading frame 21, isoform CRA_b [Homo sapiens] >gb AAI36504.1 Chromosome 9 open reading frame 21 [Homo sapiens]	45.8	45.8	82%	0.002	UGM
EAW92661.1	chromosome 9 open reading frame 21, isoform CRA_a [Homo sapiens]	45.4	45.4	77%	0.002	G
XP_001106503.1	PREDICTED: UPF0308 protein C9orf21-like [Macaca mulatta]	45.4	45.4	82%	0.003	UGM
XP_002191279.1	PREDICTED: hypothetical protein [Taeniopygia guttata]	45.4	45.4	100%	0.003	UGM
CAN81555.1	hypothetical protein VITISV_040397 [Vitis vinifera]	45.1	45.1	100%	0.003	
XP_002263959.1	PREDICTED: hypothetical protein [Vitis vinifera]	45.1	45.1	100%	0.003	UGM
EAY91739.1	hypothetical protein Osl_13380 [Oryza sativa Indica Group]	45.1	45.1	100%	0.003	
EEE59859.1	hypothetical protein OsJ_12441 [Oryza sativa Japonica Group]	45.1	45.1	100%	0.003	
NP_001051154.1	Os03g0729300 [Oryza sativa Japonica Group] >gb AAO38466.1 unknown protein [Oryza sativa Japonica Group] >gb ABF98681.1 UPF0308 protein, chloroplast precursor, putative, expressed [Oryza sativa Japonica Group] >dbj BAF13068.1 Os03g0729300 [Oryza sativa Japonica Group] >dbj BAG98418.1 unnamed protein product [Oryza sativa Japonica Group]	45.1	45.1	100%	0.003	UG
XP_002533575.1	conserved hypothetical protein [Ricinus communis] >gb EEF28810.1 conserved hypothetical protein [Ricinus communis]	45.1	45.1	100%	0.004	G
XP_001588758.1	hypothetical protein SS1G_10305 [Sclerotinia sclerotiorum 1980] >gb EDN94432.1 hypothetical protein SS1G_10305 [Sclerotinia sclerotiorum 1980]	44.7	44.7	74%	0.004	G
XP_001784902.1	predicted protein [Physcomitrella patens subsp. patens] >gb EDQ50291.1 predicted protein [Physcomitrella patens subsp. patens]	44.7	44.7	100%	0.004	UG
XP_848380.1	PREDICTED: similar to UPF0308 protein C9orf21 [Canis familiaris]	44.7	44.7	81%	0.004	UGM
ZP_02926219.1	hypothetical protein VspiD_06235 [Verrucomicrobium spinosum DSM 4136]	44.7	44.7	71%	0.004	
XP_001116318.1	PREDICTED: UPF0308 protein C9orf21-like [Macaca mulatta]	44.7	44.7	63%	0.004	GM
XP_002708298.1	PREDICTED: hypothetical protein [Oryctolagus cuniculus]	44.3	44.3	68%	0.005	UGM
XP_681257.1	hypothetical protein AN7988.2 [Aspergillus nidulans FGSC A4] >gb EAA58791.1 hypothetical protein AN7988.2 [Aspergillus nidulans FGSC A4] >tpe CBF73658.1 TPA:	43.5	43.5	79%	0.008	G

Accession	Description	Max score	Total score	Query coverage	E value	Links
	conserved hypothetical protein [Aspergillus nidulans FGSC A4]					
EDL84435.1	similar to UPF0308 protein C9orf21, isoform CRA_c [Rattus norvegicus]	43.1	43.1	82%	0.012	G
BAE26709.1	unnamed protein product [Mus musculus]	43.1	43.1	98%	0.013	G
ZP_01909602.1	hypothetical protein PPSIR1_24939 [Plesiocystis pacifica SIR-1] >gb EDM77490.1 hypothetical protein PPSIR1_24939 [Plesiocystis pacifica SIR-1]	42.7	42.7	100%	0.014	
XP_001829670.2	hypothetical protein CC1G_11406 [Coprinopsis cinerea okayama7#130] >gb EAU92121.2 hypothetical protein CC1G_11406 [Coprinopsis cinerea okayama7#130]	42.4	42.4	82%	0.018	G
XP_002628008.1	conserved hypothetical protein [Ajellomyces dermatitidis SLH14081] >gb EEQ72686.1 conserved hypothetical protein [Ajellomyces dermatitidis SLH14081] >gb EEQ88802.1 conserved hypothetical protein [Ajellomyces dermatitidis ER-3]	42.4	42.4	59%	0.020	G
EFQ95671.1	hypothetical protein PTT_05482 [Pyrenophora teres f. teres 0-1]	42.4	42.4	74%	0.022	
ABF98680.1	expressed protein [Oryza sativa Japonica Group] >dbj BAG90239.1 unnamed protein product [Oryza sativa Japonica Group]	42.0	42.0	76%	0.023	
XP_001547900.1	hypothetical protein BC1G_13584 [Botryotinia fuckeliana B05.10] >gb EDN19735.1 hypothetical protein BC1G_13584 [Botryotinia fuckeliana B05.10]	42.0	42.0	76%	0.025	G
NP_001102584.1	hypothetical protein LOC498685 [Rattus norvegicus] >gb EDL84433.1 similar to UPF0308 protein C9orf21, isoform CRA_a [Rattus norvegicus]	42.0	42.0	68%	0.026	UG
AAI36502.1	C9orf21 protein [Homo sapiens]	42.0	42.0	77%	0.027	G
XP_001940231.1	conserved hypothetical protein [Pyrenophora tritici-repentis Pt-1C-BFP] >gb EDU42950.1 conserved hypothetical protein [Pyrenophora tritici-repentis Pt-1C-BFP]	42.0	42.0	74%	0.028	G
CAI40533.1	novel protein [Homo sapiens]	41.6	41.6	77%	0.032	G
CBI57389.1	unnamed protein product [Sordaria macrospora]	41.2	41.2	88%	0.040	
BAB24662.1	unnamed protein product [Mus musculus]	41.2	41.2	82%	0.041	G
BAE40544.1	unnamed protein product [Mus musculus]	41.2	41.2	82%	0.043	G
NP_079646.1	hypothetical protein LOC66129 [Mus musculus] >sp Q9D1A0.1 CI021_MOUSE RecName: Full=UPF0308 protein C9orf21 homolog >dbj BAB22993.1 unnamed protein product [Mus musculus] >dbj BAE28518.1 unnamed protein product [Mus musculus] >gb AAI40306.1 RIKEN cDNA 1110018J18 gene [synthetic construct] >gb EDL16238.1 RIKEN cDNA 1110018J18, isoform CRA_b [Mus musculus] >gb AAI56632.1 RIKEN cDNA 1110018J18 gene [synthetic construct]	41.2	41.2	82%	0.049	UGM

Accession	Description	Max score	Total score	Query coverage	E value	Links
ADO28366.1	upf0308 protein c9orf21-like protein [Ictalurus furcatus]	40.8	40.8	68%	0.059	
YP_001543559.1	hypothetical protein Haur_0783 [Herpetosiphon aurantiacus ATCC 23779] >gb ABX03431.1 hypothetical protein Haur_0783 [Herpetosiphon aurantiacus ATCC 23779]	40.4	40.4	77%	0.071	G
XP_001824981.1	hypothetical protein AOR_1_1228084 [Aspergillus oryzae RIB40] >dbj BAE63848.1 unnamed protein product [Aspergillus oryzae]	40.4	40.4	64%	0.071	G
XP_002383597.1	conserved hypothetical protein [Aspergillus flavus NRRL3357] >gb EED46061.1 conserved hypothetical protein [Aspergillus flavus NRRL3357]	40.4	40.4	64%	0.072	G
XP_001368977.1	PREDICTED: similar to C9ORF21 [Monodelphis domestica]	40.4	40.4	100%	0.077	UGM
NP_001158627.1	UPF0308 protein C9orf21 homolog [Oncorhynchus mykiss] >gb ACO08544.1 UPF0308 protein C9orf21 homolog [Oncorhynchus mykiss]	40.4	40.4	72%	0.079	UG
NP_001177306.1	selenoprotein L [Ciona intestinalis]	40.4	40.4	77%	0.087	G
YP_002606071.1	hypothetical protein HRM2_48590 [Desulfobacterium autotrophicum HRM2] >gb ACN17907.1 conserved hypothetical protein [Desulfobacterium autotrophicum HRM2]	40.0	40.0	79%	0.089	G
XP_961259.1	hypothetical protein NCU04268 [Neurospora crassa OR74A] >emb CAC28745.1 putative protein [Neurospora crassa] >gb EAA32023.1 conserved hypothetical protein [Neurospora crassa OR74A]	40.0	40.0	67%	0.10	G
ACI68733.1	C1orf93 homolog [Salmo salar]	40.0	40.0	72%	0.11	
EDL16239.1	RIKEN cDNA 1110018J18, isoform CRA_c [Mus musculus]	39.7	39.7	68%	0.14	G
XP_003050915.1	hypothetical protein NECHADRAFT_69625 [Nectria haematococca mpVI 77-13-4] >gb EEU45202.1 hypothetical protein NECHADRAFT_69625 [Nectria haematococca mpVI 77-13-4]	39.7	39.7	88%	0.15	G
YP_001543558.1	hypothetical protein Haur_0782 [Herpetosiphon aurantiacus ATCC 23779] >gb ABX03430.1 hypothetical protein Haur_0782 [Herpetosiphon aurantiacus ATCC 23779]	39.3	39.3	61%	0.15	G
ZP_01089021.1	hypothetical protein DSM3645_00440 [Blastopirellula marina DSM 3645] >gb EAQ82136.1 hypothetical protein DSM3645_00440 [Blastopirellula marina DSM 3645]	38.9	38.9	75%	0.21	
XP_001800743.1	hypothetical protein SNOG_10473 [Phaeosphaeria nodorum SN15] >gb EAT81867.1 hypothetical protein SNOG_10473 [Phaeosphaeria nodorum SN15]	38.1	38.1	76%	0.35	G

Accession	Description	Max score	Total score	Query coverage	E value	Links
YP_001547656.1	hypothetical protein Haur_4898 [Herpetosiphon aurantiacus ATCC 23779] >gb ABX07528.1 hypothetical protein Haur_4898 [Herpetosiphon aurantiacus ATCC 23779]	38.1	38.1	75%	0.35	G
ACI70082.1	C1orf93 homolog [Salmo salar]	38.1	38.1	72%	0.37	
XP_002544130.1	predicted protein [Uncinocarpus reesii 1704] >gb EEP78801.1 predicted protein [Uncinocarpus reesii 1704]	38.1	38.1	67%	0.38	G
XP_002114328.1	hypothetical protein TRIADDRAFT_58024 [Trichoplax adhaerens] >gb EDV23418.1 hypothetical protein TRIADDRAFT_58024 [Trichoplax adhaerens]	38.1	38.1	50%	0.40	G
EFW13835.1	conserved hypothetical protein [Coccidioides posadasii str. Silveira]	37.4	37.4	67%	0.62	
XP_003070048.1	hypothetical protein CPC735_032390 [Coccidioides posadasii C735 delta SOWgp] >gb EER27903.1 hypothetical protein CPC735_032390 [Coccidioides posadasii C735 delta SOWgp]	37.4	37.4	61%	0.63	G

Alignments

>ref|XP_002326159.1| [UG](#) predicted protein [Populus trichocarpa]
 gb|EEE71829.1| [G](#) predicted protein [Populus trichocarpa]
 Length=200

GENE ID: 7459323 POPTRDRAFT_589376 | hypothetical protein [Populus trichocarpa]
 (10 or fewer PubMed links)

Score = 73.9 bits (180), Expect = 6e-12, Method: Compositional matrix adjust.
 Identities = 45/124 (37%), Positives = 61/124 (50%), Gaps = 12/124 (9%)

```

Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V + RHFGCVLCRR LAA++ L++IG G+ +QA F
Sbjct 49  VVAFARHFGCVLCRRRADYLAAKKD-IMDASGVALVLIGPGSVDQAKTFSEQTKFKG--- 104

Query 61  SIPVYTDPLRRSYLALRFRGILSTFNLPALQNILDSFQKGERQRW-----DMIPPPDAF 114
          VY DP SY AL+F G+ +TF A I+ S+ +G RQ W D + +
Sbjct 105  --EVYADPSHSSYKALQFVSGVSTFTTPKAGLKIISYMEGYRQDWKLSFEGDTPVAKGGW 162

Query 115  QQGG 118
           QQGG
Sbjct 163  QQGG 166
    
```

>ref|NP_001049763.1| [UG](#) Os03g0284600 [Oryza sativa Japonica Group]
 gb|ABF95344.1| [G](#) UPF0308 protein, chloroplast precursor, putative, expressed [Oryza sativa Japonica Group]
 dbj|BAF11677.1| [G](#) Os03g0284600 [Oryza sativa Japonica Group]
 gb|EEE58829.1| [G](#) hypothetical protein OsJ_10400 [Oryza sativa Japonica Group]
 Length=251

GENE ID: 4332468 Os03g0284600 | Os03g0284600 [Oryza sativa Japonica Group]
 (10 or fewer PubMed links)

Score = 72.0 bits (175), Expect = 3e-11, Method: Compositional matrix adjust.
 Identities = 42/108 (39%), Positives = 56/108 (52%), Gaps = 6/108 (5%)

```


Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V + RHFGCVLCR+ LAA++ A + L++IG GT EQA F
Sbjct 99  IVAFARHFGCVLCRKRADLLAAKQDAM-EAAGVALVLIGPGTVEQAKAFYDQTKF----- 152
    
```


Query 61 SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNILDSFQKGERQRWDM 108
 VY DP SY AL F G+ STF A I+ + +G RQ W++
 Sbjct 153 KGEVYADPSHSSYNALEFAFGLFSTFTPSAGLKIIQLYMEGYRQDWEL 200

>gb|EEC75002.1| hypothetical protein OsI_11064 [Oryza sativa Indica Group]
 Length=239

Score = 71.6 bits (174), Expect = 3e-11, Method: Compositional matrix adjust.
 Identities = 42/108 (39%), Positives = 56/108 (52%), Gaps = 6/108 (5%)

Query 1 LVIWPRHFGCVLCRRLVKELAAREQAQFRQERKWRLMVIGCGTPEQAAKFRDATALGPLPA 60
 +V + RHFGCVLCR+ LAA++ A + L++IG GT EOA F
 Sbjct 87 IVAFARHFGCVLCRKRADLLAAKQDAM-EAAGVALVLIGPGTVEQAKAFYDQTKF----- 140
 Query 61 SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNILDSFQKGERQRWDM 108
 VY DP SY AL F G+ STF A I+ + +G RQ W++
 Sbjct 141 KGEVYADPSHSSYNALEFAFGLFSTFTPSAGLKIIQLYMEGYRQDWEL 188


>ref|XP_002881499.1|  hypothetical protein ARALYDRAFT_482716 [Arabidopsis lyrata subsp. lyrata]


gb|EFH57758.1|  hypothetical protein ARALYDRAFT_482716 [Arabidopsis lyrata subsp. lyrata]
 Length=248

GENE ID: 9315725 ARALYDRAFT_482716 | hypothetical protein
 [Arabidopsis lyrata subsp. lyrata]

Score = 71.6 bits (174), Expect = 4e-11, Method: Compositional matrix adjust.
 Identities = 44/124 (36%), Positives = 59/124 (48%), Gaps = 12/124 (9%)

Query 1 LVIWPRHFGCVLCRRLVKELAAREQAQFRQERKWRLMVIGCGTPEQAAKFRDATALGPLPA 60
 +V + RHFGCVLCR+ LA ++ L++IG G+ +QA F
 Sbjct 96 VVAFARHFGCVLCRKRRAAYLAEKKDVM-DASGVTLVLIGPGSIDQANTFMEQTKFKG--- 151
 Query 61 SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNILDSFQKGERQRW-----DMIPPPDAF 114
 VY DP SY AL F G+ TF A IL+S+ +G RQ W D + +
 Sbjct 152 --EVYADPNHASYEALFVSGVTFTPKAAMKILESYMEGYRQDWKLSFMKDTVERGGW 209
 Query 115 QQGG 118
 QQGG
 Sbjct 210 QQGG 213


>ref|NP_001145525.1|  hypothetical protein LOC100278941 [Zea mays]


gb|ACG48189.1|  hypothetical protein [Zea mays]
 Length=258





GENE ID: 100278941 LOC100278941 | hypothetical protein LOC100278941 [Zea mays]
 (10 or fewer PubMed links)

Score = 71.2 bits (173), Expect = 4e-11, Method: Compositional matrix adjust.
 Identities = 42/108 (39%), Positives = 57/108 (53%), Gaps = 7/108 (6%)

Query 1 LVIWPRHFGCVLCRRLVKELAAREQAQFRQERKWRLMVIGCGTPEQAAKFRDATALGPLPA 60
 +V + RHFGCVLCR+ LAA++ Q L++IG G+ EOA F G
 Sbjct 107 VVAFARHFGCVLCRKRADLLAAKQDDM-QAAGVALVLIGPGSVEQAKAFEQTKFKGE--- 162
 Query 61 SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNILDSFQKGERQRWDM 108
 VY DP SY AL F G+ STF A I+ +++G RQ W++
 Sbjct 163 ---VYADPTHSSYDALEFAFGLFSTFTPAAGLKIIQLYREGYRQDWEL 207

>ref|NP_030274.1|  unknown protein [Arabidopsis thaliana]

sp|Q9ZUU2.2|U308_ARATH  RecName: Full=UPF0308 protein At2g37240, chloroplastic; Flags:
 Precursor

gb|AAK91362.1|  At2g37240/F3G5.3 [Arabidopsis thaliana]
 gb|AAC98045.2|  expressed protein [Arabidopsis thaliana]
 gb|AAM67197.1|  unknown [Arabidopsis thaliana]
 gb|AAP21145.1|  At2g37240/F3G5.3 [Arabidopsis thaliana]
 Length=248

GENE ID: 818301 AT2G37240 | hypothetical protein [Arabidopsis thaliana]
 (10 or fewer PubMed links)

Score = 70.5 bits (171), Expect = 7e-11, Method: Compositional matrix adjust.
Identities = 44/124 (36%), Positives = 59/124 (48%), Gaps = 12/124 (9%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V + RHFGCVLCR+ LA ++ L++IG G+ +QA F
Sbjct 96  VVAFARHFGCVLCRKRAAYLAEKKDVM-DASGVALVLIGPGSIDQANTFVEQTKFKG--- 151

Query 61  SIPVYTDPLRRSYLALRFRGILSTFNLPALQNILDSFQKGERQRW-----DMIPPPDAF 114
          VY DP SY AL F G+ TF A IL+S+ +G RQ W D + +
Sbjct 152  --EVYADPNHASYEALFVSGVSVTFTPKAAMKILESMEGYRQDWKLSFMKDTVERGGW 209

Query 115  QGG 118
           QGG
Sbjct 210  QGG 213
```

>ref|XP_002334170.1| **UG** predicted protein [Populus trichocarpa]
gb|EEF07066.1| **G** predicted protein [Populus trichocarpa]
Length=146

GENE ID: 7497195 POPTRDRAFT_941289 | hypothetical protein [Populus trichocarpa]
(10 or fewer PubMed links)

Score = 69.7 bits (169), Expect = 1e-10, Method: Compositional matrix adjust.
Identities = 40/106 (38%), Positives = 54/106 (51%), Gaps = 6/106 (5%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V + RHFGCVLCRR LAA++ L++IG G+ +QA F
Sbjct 46  VVAFARHFGCVLCRRRADYLAAKD-IMDASGVALVLIGPGSVDAQTFSEQTKFKG--- 101

Query 61  SIPVYTDPLRRSYLALRFRGILSTFNLPALQNILDSFQKGERQRW 106
          VY DP SY AL+F G+ +TF A I+ S+ +G RQ W
Sbjct 102  --EVYADPHSSYKALQFVSGVSTFTPKAGLKI IQSYMEGYRQDW 145
```

>ref|XP_002468056.1| **UG** hypothetical protein SORBIDRAFT_01g038790 [Sorghum bicolor]
gb|EER95054.1| **G** hypothetical protein SORBIDRAFT_01g038790 [Sorghum bicolor]
Length=258

GENE ID: 8084887 SORBDRAFT_01g038790 | hypothetical protein [Sorghum bicolor]
(10 or fewer PubMed links)

Score = 69.7 bits (169), Expect = 1e-10, Method: Compositional matrix adjust.
Identities = 42/108 (39%), Positives = 56/108 (52%), Gaps = 6/108 (5%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V + RHFGCVLCR+ LAA+ Q Q L++IG G+ EQA F
Sbjct 106  VVAFARHFGCVLCRKRADLLAAK-QDVMQAAGVALVLIGPGSVEQAKAFCEQTKF----- 159

Query 61  SIPVYTDPLRRSYLALRFRGILSTFNLPALQNILDSFQKGERQRWDM 108
          VY DP SY AL F G+ STF A I+ +++G RQ W++
Sbjct 160  KGEVYADPTHSSYDALEFAFGLFSTFTPAAGLKI IQLYREGYRQDWEL 207
```

>gb|ACR37670.1| unknown [Zea mays]
Length=259

Score = 69.3 bits (168), Expect = 2e-10, Method: Compositional matrix adjust.
Identities = 41/108 (38%), Positives = 56/108 (52%), Gaps = 6/108 (5%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V + RHFGCVLCR+ LAA++ Q L++IG G+ EQA F
Sbjct 107  VVAFARHFGCVLCRKRADLLAAKQDDM-QAAGVALVLIGPGSVEQAKAFCEQTKF----- 160

Query 61  SIPVYTDPLRRSYLALRFRGILSTFNLPALQNILDSFQKGERQRWDM 108
          VY DP SY AL F G+ STF A I+ +++G RQ W++
Sbjct 161  KGEVYADPTHSSYDALEFAFGLFSTFTPAAGLKI IQLYREGYRQDWEL 208
```

>gb|ACN35248.1| **G** unknown [Zea mays]
Length=224

GENE ID: 100278941 LOC100278941 | hypothetical protein LOC100278941 [Zea mays]
(10 or fewer PubMed links)



Score = 68.6 bits (166), Expect = 2e-10, Method: Compositional matrix adjust.
Identities = 41/108 (38%), Positives = 56/108 (52%), Gaps = 6/108 (5%)

```

Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V + RHFGCVLCR+   LAA++   0   L++IG G+ EQA F
Sbjct 107 VVAFARHFGCVLCRKRADLLAAKQDDM-QAAGVALVLIGPGSVEQAKAFCEQTKF----- 160

Query 61  SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNILDSFQKGERQRWDM 108
          VY DP   SY AL F G+ STF A   I+   +++G RQ W++
Sbjct 161  KGEVYADPTHTSSYDALEFAFGLFSTFTPAAGLKIQLYREGYRQDWEL 208

```

>ref|XP_002965388.1|  hypothetical protein SELMODRAFT_68006 [Selaginella moellendorffii]
 >gb|EFJ34226.1|  hypothetical protein SELMODRAFT_68006 [Selaginella moellendorffii]
 Length=172

GENE ID: 9633660 SELMODRAFT_68006 | hypothetical protein
 [Selaginella moellendorffii]

Score = 67.8 bits (164), Expect = 5e-10, Method: Compositional matrix adjust.
 Identities = 40/119 (34%), Positives = 62/119 (53%), Gaps = 7/119 (5%)

```

Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V + RHFGCVLCR+   LA++++ F   L+++G GT +QA F   A+   P
Sbjct 21  VVAFARHFGCILCRKRADVLASKKEVF-DAAGVSLVLVGPVTVDQAKAF--ASQTQFPG 76

Query 61  SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNILDSFQKGERQRWDM-IPPDAFQGG 118
          VY DP   S+ A +F G + FN A   ++ + +G RQ W +   D   0+GG
Sbjct 77  E--VYADPTHTASFDQFVSGASTIFNPKAAMRVMGAHLEGYRQDWGLSFEKDTVQRGG 133

```

>gb|ACU20039.1| unknown [Glycine max]
 Length=256

Score = 67.8 bits (164), Expect = 5e-10, Method: Compositional matrix adjust.
 Identities = 40/124 (33%), Positives = 61/124 (50%), Gaps = 12/124 (9%)



```

Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V + RHFGCVLCR+   L++++   L++IG G+ +QA F   +
Sbjct 104 VVAFARHFGCVLCRKRADYLSKSKD-IMDASGVALVLIGPGSIDQAKSFAEKSKEFEG--- 159

Query 61  SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNILDSFQKGERQRW-----DMIPPDAF 114
          +Y DP   SY AL F G+L+TF A   I+   + +G RQ W   D +   +
Sbjct 160 --EIYADPTHTSSYEALNFVSGVLTFTPNAGLKIQLYMEGYRQDWKLSFEKDTVSRGGW 217

Query 115  QGGG 118
           +GGG
Sbjct 218  KGGG 221

```

>ref|XP_002977235.1|  hypothetical protein SELMODRAFT_58056 [Selaginella moellendorffii]
 >gb|EFJ21844.1|  hypothetical protein SELMODRAFT_58056 [Selaginella moellendorffii]
 Length=172

GENE ID: 9638970 SELMODRAFT_58056 | hypothetical protein
 [Selaginella moellendorffii]

Score = 67.4 bits (163), Expect = 5e-10, Method: Compositional matrix adjust.
 Identities = 40/119 (34%), Positives = 62/119 (53%), Gaps = 7/119 (5%)

```

Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V + RHFGCVLCR+   LA++++ F   L+++G GT +QA F   A+   P
Sbjct 21  VVAFARHFGCILCRKRADVLASKKEVF-DGAGVSLVLVGPVTVDQAKAF--ASQTQFPG 76

Query 61  SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNILDSFQKGERQRWDM-IPPDAFQGG 118
          VY DP   S+ A +F G + FN A   ++ + +G RQ W +   D   0+GG
Sbjct 77  E--VYADPTHTASFEAFQFVSGASTIFNPKAAMRVMGAHLEGYRQDWGLSFEKDTVQRGG 133

```

>gb|ABK23985.1| unknown [Picea sitchensis]
 Length=261

Score = 65.9 bits (159), Expect = 2e-09, Method: Compositional matrix adjust.
 Identities = 43/124 (35%), Positives = 62/124 (50%), Gaps = 12/124 (9%)

```

Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          ++ + RHFGCVLCR+   LA+++   L++IG G EQA F   A   P
Sbjct 109 VIGFARHFGCVLCRKRADVLASQKSQM-DAAGVALVLIGPGNIEQAKAF--ADQTKFPG 164

Query 61  SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNILDSFQKGERQRW-----DMIPPDAF 114
          I Y DP   S+ AL+F G+ +TF A   I++ + +G RQ W   D +   +

```

Sbjct 165 EI--YADPNHTSFNALKFVSGVFTTFTPLAATKIIELYVEGYRQDWGLSFQKDTMNRGGW 222
 Query 115 QQGG 118
 QGGG
 Sbjct 223 QQGG 226

>ref|XP_002273449.1| **UGM** PREDICTED: hypothetical protein [Vitis vinifera]
 emb|CBI32627.3| **G** unnamed protein product [Vitis vinifera]
 Length=254

GENE ID: 100251717 LOC100251717 | hypothetical protein LOC100251717
 [Vitis vinifera] (10 or fewer PubMed links)

Score = 65.1 bits (157), Expect = 3e-09, Method: Compositional matrix adjust.
 Identities = 42/124 (34%), Positives = 58/124 (47%), Gaps = 12/124 (9%)

Query 1 LVIWPRHFGCVLCRRLVKELAAREQAQFRQERKWRLMVGCGTPEQAAKFRDATLGPLPA 60
 +V + RHFGCV CR+ LA+++ L++IG G+ +QA F
 Sbjct 102 VVAFARHFGCVFCRKRADLLASQKDRM-DASGVALVLIGPGSIDQAKAFSEQTNFKG--- 157
 Query 61 SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNILDSFQKGERQRW-----DMIPPPDAF 114
 VY DP SY L F G+LSTF A I+ + +G RQ W D + +
 Sbjct 158 --EVYADPSHSSYEVLGFVSGVLSSTFTPAQGLKIIQLYMEGYRQDWGLSFQRDVTTRGGW 215
 Query 115 QQGG 118
 QGGG
 Sbjct 216 QQGG 219

>ref|XP_002525198.1| **G** conserved hypothetical protein [Ricinus communis]
 gb|EEF37164.1| **G** conserved hypothetical protein [Ricinus communis]
 Length=249

GENE ID: 8283839 RCOM_0819880 | hypothetical protein [Ricinus communis]

Score = 63.5 bits (153), Expect = 7e-09, Method: Compositional matrix adjust.
 Identities = 40/124 (33%), Positives = 59/124 (48%), Gaps = 12/124 (9%)

Query 1 LVIWPRHFGCVLCRRLVKELAAREQAQFRQERKWRLMVGCGTPEQAAKFRDATLGPLPA 60
 +V + RHFGCVLCR+ LAA++ L++IG G+ +QA F
 Sbjct 97 VVAFARHFGCVLCRKRADYLAACKD-IMDASGVALVLIGPGSVDQAKTFSEQTKFKG--- 152
 Query 61 SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNILDSFQKGERQRW-----DMIPPPDAF 114
 VY D SY A +F G+ +TF A I++ + +G RQ W D + +
 Sbjct 153 --EVYADTSHSSYEAQFVSGVSTTFTPAKGLKIIELYMEGYRQDWKLSFEKDTVARGGW 210
 Query 115 QQGG 118
 +QGG
 Sbjct 211 RQGG 214

>ref|XP_002866692.1| **G** hypothetical protein ARALYDRAFT_496819 [Arabidopsis lyrata subsp.
 lyrata]
 gb|EFH42951.1| **G** hypothetical protein ARALYDRAFT_496819 [Arabidopsis lyrata subsp.
 lyrata]
 Length=265

GENE ID: 9302764 ARALYDRAFT_496819 | hypothetical protein
 [Arabidopsis lyrata subsp. lyrata]

Score = 52.0 bits (123), Expect = 3e-05, Method: Compositional matrix adjust.
 Identities = 40/124 (33%), Positives = 55/124 (45%), Gaps = 12/124 (9%)

Query 2 VIWPRHFGCVLCRRLVKELAAREQAQFRQERKWRLMVGCGTPEQAAKFRDATLGPLPAS 61
 V+ RHFGCV C L L + F +L+ +G GTP++A R AT P P
 Sbjct 110 VVLLRHFGCVCCWELATALKEAKPRF-DAAGVKLIAVGVGTPDKA---RILATRLPFPME 165
 Query 62 IPVYTDPLRRSYLALRFRRGILSTFNLPALQNILDSFQK---GERQRWDMIPPD---AF 114
 +Y DP R++Y L G+ TF PA + F + + P D
 Sbjct 166 C-LYADPERKAYDVLGLYGLGRTFFNPASTKVFSRFNEIREATKNYTIATPEDRSSVL 224
 Query 115 QQGG 118
 QGGG
 Sbjct 225 QQGG 228

>ref|NP_201385.2| **UG** unknown protein [Arabidopsis thaliana]
 gb|AAM20692.1| **G** unknown protein [Arabidopsis thaliana]
 gb|AAN15652.1| **G** unknown protein [Arabidopsis thaliana]
 Length=275

GENE ID: 836713 AT5G65840 | hypothetical protein [Arabidopsis thaliana]
 (10 or fewer PubMed links)

Score = 51.6 bits (122), Expect = 4e-05, Method: Compositional matrix adjust.
 Identities = 40/124 (33%), Positives = 55/124 (45%), Gaps = 12/124 (9%)

```
Query 2  VIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVICGTPEQAAKFRTDATLGPLPAS 61
          +V RHFVC C L L + F +L+ +G GTP++A R AT P P
Sbjct 120 VVLLRHFGCVCCWELATALKEAKPRF-DAAGVKLIAVGVGTPDKA---RILATRLPFPM 175

Query 62  IPVYTDPLRRSYLALRFRRGILSTFNLPALQNILDSF----QKGERQRWDMIPPD---AF 114
          +Y DP R++Y L G+ TF PA + F + + + P D
Sbjct 176 C-LYADPERKAYDVLGLYFGLGRTFFNPASTKVFSTRFSEIREATKNYTIATPEDRSSVL 234

Query 115  QQGG 118
          QQGG
Sbjct 235 QQGG 238
```

>gb|ACU20324.1| unknown [Glycine max]
 Length=251

Score = 50.8 bits (120), Expect = 5e-05, Method: Compositional matrix adjust.
 Identities = 39/125 (32%), Positives = 53/125 (43%), Gaps = 12/125 (9%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVICGTPEQAAKFRTDATLGPLPA 60
          +V RHFVC C L L + F +L+ +G GTP +A R A P P
Sbjct 95  VVAMLRHFGCICWEFASALKESKARF-DSAGIKLIAVGVGTPNKA---RILAERLPFPM 150

Query 61  SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNIL---DSFQKGERQRWDMIPPD----A 113
          +Y DP R++Y L G+ TF PA + D+ QK + PD
Sbjct 151 DC-LYADPDRKAYNVLNLYFGLGRTFLNPASAKVFSRWDALQKAAKNYTIATPDDISGV 209

Query 114  FQQGG 118
          QQGG
Sbjct 210 LQQGG 214
```

>ref|XP_002320012.1| **UG** predicted protein [Populus trichocarpa]
 gb|EEE98327.1| **G** predicted protein [Populus trichocarpa]
 Length=199

GENE ID: 7497309 PNPTDRRAFT_571999 | hypothetical protein [Populus trichocarpa]
 (10 or fewer PubMed links)

Score = 50.1 bits (118), Expect = 9e-05, Method: Compositional matrix adjust.
 Identities = 40/125 (32%), Positives = 54/125 (44%), Gaps = 12/125 (9%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVICGTPEQAAKFRTDATLGPLPA 60
          +V RHFVC C L L ++ F +L+ IG GTP +A R A P P
Sbjct 43  VVALLRHFGCPCWELASSLKESKEKF-DSSGVKLIAGVGTNKA---RLLAERLPFPM 98

Query 61  SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNI---LDSFQKGERQRWDMIPPD----A 113
          +Y DP R++Y L G+ TF PA + D+ +K + PD
Sbjct 99  DC-LYADPERKAYDVLGLYYGLGRTFFNPASAKVFSRFDALRKAVKNYTIATPDDRSV 157

Query 114  FQQGG 118
          QQGG
Sbjct 158 LQQGG 162
```

>emb|CBI33071.3| unnamed protein product [Vitis vinifera]
 Length=159

Score = 49.7 bits (117), Expect = 1e-04, Method: Compositional matrix adjust.
 Identities = 40/125 (32%), Positives = 55/125 (44%), Gaps = 12/125 (9%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVICGTPEQAAKFRTDATLGPLPA 60
          +V RHFVC C L L ++ F +L+ +G GTP++A R A P P
Sbjct 3  VVALLRHFGCPCWDLASALKESKERF-DSAGVKLIAVGVGTPDKA---RILAERLPFPL 58

Query 61  SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNIL---DSFQKGERQRWDMIPPD----A 113
          +Y DP R++Y L G TF PA +L ++ QK + PD
```

Sbjct 59 DC-LYADPDRKAYDVLGLYYGFGRTFFNPASAKVLLRFEALQKAVKNYTIKATPDDKSSV 117
 Query 114 FQGG 118
 00GG
 Sbjct 118 LQGG 122

>gb|ADE77692.1| unknown [Picea sitchensis]
 Length=276

Score = 49.3 bits (116), Expect = 1e-04, Method: Compositional matrix adjust.
 Identities = 42/125 (34%), Positives = 53/125 (43%), Gaps = 12/125 (9%)

Query 1 LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDATLGPLPA 60
 +V RHFGC C L F +L+ IG GTPE+A R P P
 Sbjct 120 VVALLRHFGCPCWFASTLKDVMPKF-DSAGVKLIAIGVGTPEKA---RILGERLPFPL 175
 Query 61 SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNIL---DSFQKGERQRWDMIPPD----A 113
 +Y DP R++Y AL G+ TF PA +L DS QK + P+
 Sbjct 176 D-SLYADPDRKAYDALGLYYGLGRTFFNPASAKVLRTRFDSLQKALKKNYTIATPEDRSSV 234
 Query 114 FQGG 118
 00GG
 Sbjct 235 LQGG 239

>ref|XP_002263922.1| UGM PREDICTED: hypothetical protein [Vitis vinifera]
 emb|CAN81556.1| G hypothetical protein VITISV_040398 [Vitis vinifera]
 Length=256

GFNF TD: 100256614 LOC100256614 | hypothetical protein [Vitis vinifera]
 (10 or fewer PubMed links)

Score = 49.3 bits (116), Expect = 2e-04, Method: Compositional matrix adjust.
 Identities = 40/125 (32%), Positives = 55/125 (44%), Gaps = 12/125 (9%)

Query 1 LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDATLGPLPA 60
 +V RHFGC C L L ++ F +L+ +G GTP++A R A P P
 Sbjct 100 VVALLRHFGCPCWDLASALKESKERF-DSAGVKLIAVGVGTPDKA---RILAERLPFPL 155
 Query 61 SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNIL---DSFQKGERQRWDMIPPD----A 113
 +Y DP R++Y L G TF PA +L ++ QK + PD
 Sbjct 156 DC-LYADPDRKAYDVLGLYYGFGRTFFNPASAKVLLRFEALQKAVKNYTIKATPDDKSSV 214
 Query 114 FQGG 118
 00GG
 Sbjct 215 LQGG 219

>emb|CBI33293.3| unnamed protein product [Vitis vinifera]
 Length=256

Score = 49.3 bits (116), Expect = 2e-04, Method: Compositional matrix adjust.
 Identities = 40/125 (32%), Positives = 55/125 (44%), Gaps = 12/125 (9%)

Query 1 LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDATLGPLPA 60
 +V RHFGC C L L ++ F +L+ +G GTP++A R A P P
 Sbjct 100 VVALLRHFGCPCWDLASALKESKERF-DSAGVKLIAVGVGTPDKA---RILAERLPFPL 155
 Query 61 SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNIL---DSFQKGERQRWDMIPPD----A 113
 +Y DP R++Y L G TF PA +L ++ QK + PD
 Sbjct 156 DC-LYADPDRKAYDVLGLYYGFGRTFFNPASAKVLLRFEALQKAVKNYTIKATPDDKSSV 214
 Query 114 FQGG 118
 00GG
 Sbjct 215 LQGG 219

>gb|ACN32054.1| G unknown [Zea mays]
 Length=258

GFNF TD: 100275311 LOC100275311 | hypothetical protein LOC100275311 [Zea mays]
 (10 or fewer PubMed links)

Score = 49.3 bits (116), Expect = 2e-04, Method: Compositional matrix adjust.
 Identities = 42/120 (35%), Positives = 51/120 (43%), Gaps = 12/120 (10%)

Query 6 RHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDATLGPLPASIPVY 65
 RHFGC C L L F +L+ IG GTPE+A R A P P +Y

Sbjct 105 RHF~~G~~C~~F~~C~~W~~E~~L~~A~~S~~V~~L~~K~~D~~A~~M~~A~~E~~F~~D~~S~~A~~G~~A~~-K~~L~~I~~A~~I~~G~~V~~G~~T~~P~~E~~K~~A---R~~I~~L~~A~~D~~R~~L~~P~~F~~P~~M~~D~~-S~~L~~Y 159
 Query 66 TD~~P~~L~~R~~R~~S~~Y~~L~~A~~L~~R~~F~~R~~R~~G~~I~~L~~S~~T~~F~~N~~L~~P~~A~~L~~Q~~N~~I~~---L~~D~~S~~F~~Q~~K~~G-~~E~~R~~Q~~R~~W~~M~~I~~P~~P~~D---A~~F~~Q~~Q~~G~~G~~ 118
 D~~P~~ R~~+~~Y L G~~+~~ T P~~A~~ I L~~D~~ +~~K~~ E + P~~D~~ Q~~Q~~G~~G~~
 Sbjct 160 A~~D~~P~~E~~R~~K~~A~~S~~V~~L~~G~~L~~Y~~H~~G~~L~~G~~R~~T~~L~~F~~S~~P~~A~~S~~A~~K~~I~~Y~~S~~R~~L~~D~~Y~~I~~K~~K~~A~~T~~E~~N~~Y~~T~~L~~E~~G~~T~~P~~A~~D~~L~~T~~G~~V~~L~~Q~~Q~~G~~G 219

>ref|XP_003130905.1| **UGM** PREDICTED: UPF0308 protein C9orf21 homolog isoform 2 [Sus scr
 Length=I84

GENE ID: 100524271 LOC100524271 | UPF0308 protein C9orf21 homolog [Sus scrofa]

Score = 48.1 bits (113), Expect = 4e-04, Method: Compositional matrix adjust.
 Identities = 37/118 (32%), Positives = 56/118 (48%), Gaps = 14/118 (11%)

Query 1 LVIWPRHFGCVLCRRLVKELAAREQA~~F~~R~~Q~~E~~R~~K~~W~~R~~L~~M~~V~~I~~G~~C~~G~~T~~P~~E~~Q~~A~~A~~K~~F~~R~~T~~D~~A~~T~~L~~G~~P~~L~~P~~A 60
 +V~~+~~ R~~H~~F C +C~~+~~ V~~+~~LA ++F Q~~E~~ L+V~~I~~G Q~~+~~ +
 Sbjct 61 VVVFVRHFLCYICKEYVEDLAKIPK~~S~~FLQ~~E~~ANVT~~L~~I~~V~~I~~G~~----Q~~S~~S~~Y~~H~~H~~I~~E~~P~~F~~C~~K~~L~~T~~G~~Y~~ 115
 Query 61 SIPVYTDPLRRSYLALRFRRG--ILSTFNLPALQNI~~L~~D~~S~~F~~Q~~K~~G~~E~~R~~Q~~R~~W~~D~~M~~I~~P~~P~~D~~A~~F~~Q~~Q 116
 S +Y DP R Y L +RG I S+ N NI F +R R D P ++ Q
 Sbjct 116 SHEIYVDPEREIYKRLGMKRGEEIASSGN-----NI--HF~~I~~H~~R~~D~~R~~N~~R~~L~~D~~H~~K~~P~~I~~N~~S~~V~~L~~Q 166

>ref|XP_001153024.1| **UGM** PREDICTED: similar to TPA_exp: C90RF21 isoform 1 [Pan trogl
 Length=I82

GENE ID: 465250 LOC465250 | similar to TPA_exp: C90RF21 [Pan troglodytes]

Score = 48.1 bits (113), Expect = 4e-04, Method: Compositional matrix adjust.
 Identities = 37/118 (32%), Positives = 56/118 (48%), Gaps = 14/118 (11%)

Query 1 LVIWPRHFGCVLCRRLVKELAAREQA~~F~~R~~Q~~E~~R~~K~~W~~R~~L~~M~~V~~I~~G~~C~~G~~T~~P~~E~~Q~~A~~A~~K~~F~~R~~T~~D~~A~~T~~L~~G~~P~~L~~P~~A 60
 +V~~+~~ R~~H~~F C +C~~+~~ V~~+~~LA ++F Q~~E~~ L+V~~I~~G Q~~+~~ +
 Sbjct 59 VVVFVRHFLCYICKEYVEDLAKIPK~~S~~FLQ~~E~~ANVT~~L~~I~~V~~I~~G~~----Q~~S~~S~~Y~~H~~H~~I~~E~~P~~F~~C~~K~~L~~T~~G~~Y~~ 113
 Query 61 SIPVYTDPLRRSYLALRFRRG--ILSTFNLPALQNI~~L~~D~~S~~F~~Q~~K~~G~~E~~R~~Q~~R~~W~~D~~M~~I~~P~~P~~D~~A~~F~~Q~~Q 116
 S +Y DP R Y L +RG I S+ N NI F +R R D P ++ Q
 Sbjct 114 SHEIYVDPEREIYKRLGMKRGEEIASSGN-----NI--HF~~I~~H~~R~~D~~R~~N~~R~~L~~D~~H~~K~~P~~I~~N~~S~~V~~L~~Q 164

>ref|NP_001051153.1| **UG** Os03g0729200 [Oryza sativa Japonica Group]

gb|AA038464.1| **G** hypothetical protein [Oryza sativa Japonica Group]

gb|ABF98679.1| **G** expressed protein [Oryza sativa Japonica Group]

dbj|BAF13067.1| **G** Os03g0729200 [Oryza sativa Japonica Group]

gb|EAY91738.1| **G** hypothetical protein OsI_13379 [Oryza sativa Indica Group]

dbj|BAG94118.1| **G** unnamed protein product [Oryza sativa Japonica Group]

gb|EEE59858.1| **G** hypothetical protein OsJ_12440 [Oryza sativa Japonica Group]
 Length=258

GFNF ID: 4333987 Os03g0729200 | Os03g0729200 [Oryza sativa Japonica Group]
 (10 or fewer PubMed links)

Score = 47.8 bits (112), Expect = 4e-04, Method: Compositional matrix adjust.
 Identities = 40/125 (32%), Positives = 56/125 (45%), Gaps = 12/125 (9%)

Query 1 LVIWPRHFGCVLCRRLVKELAAREQA~~F~~R~~Q~~E~~R~~K~~W~~R~~L~~M~~V~~I~~G~~C~~G~~T~~P~~E~~Q~~A~~A~~K~~F~~R~~T~~D~~A~~T~~L~~G~~P~~L~~P~~A 60
 +V R~~H~~F~~G~~C C L L + F +L+ +G G~~T~~P~~+~~A R A P P
 Sbjct 102 VVALLRHFGCPC~~C~~W~~E~~L~~A~~S~~V~~L~~K~~D~~A~~M~~A~~E~~F~~D~~S~~A~~G~~A---D~~S~~A~~G~~V~~K~~L~~I~~A~~V~~G~~V~~G~~T~~P~~D~~K~~A~~---R~~I~~L~~A~~E~~R~~L~~P~~F~~P~~L 157
 Query 61 SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNI---LDSFQKGERQ~~R~~W~~D~~M~~I~~P~~P~~D---A 113
 +Y DP R++Y L GI TF PA ++ DS ++ + PD
 Sbjct 158 DY-LYADPERKAYDLLGLYFGIGRTFFNPASASVFSR~~F~~D~~S~~L~~K~~E~~A~~V~~K~~N~~Y~~T~~I~~E~~A~~T~~P~~D~~D~~R~~A~~S~~V~~ 216
 Query 114 FQ~~Q~~G~~G~~ 118
 Q~~Q~~G~~G~~
 Sbjct 217 LQ~~Q~~G~~G~~ 221

>**gb|ACU18761.1|** unknown [Glycine max]
 Length=251

Score = 47.8 bits (112), Expect = 5e-04, Method: Compositional matrix adjust.
 Identities = 39/125 (32%), Positives = 52/125 (42%), Gaps = 12/125 (9%)

Query 1 LVIWPRHFGCVLCRRLVKELAAREQA~~F~~R~~Q~~E~~R~~K~~W~~R~~L~~M~~V~~I~~G~~C~~G~~T~~P~~E~~Q~~A~~A~~K~~F~~R~~T~~D~~A~~T~~L~~G~~P~~L~~P~~A 60
 +V R~~H~~F~~G~~C C L L + F +L+ +G G~~T~~P~~+~~A R A P P

```

Sbjct 95  VVALLRHFGCPCWELASALKESKARF-DSAGVKLIAVGIGTPNKA---RMLAERLPFPL 150
Query 61  SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNI---LDSFQKGERQRWDMIPPD----A 113
          +Y DP R++Y L G TF P+ + D+ QK + PD
Sbjct 151 DC-LYADPDRKAYHVLNLYYGFGRTFFNPSSIKVFSRFDALQKAVKNYTIATPDDRSVG 209
Query 114 FQGG 118
          QGG
Sbjct 210 LQQGG 214

```

>ref|XP_002928763.1| PREDICTED: UPF0308 protein C9orf21-like, partial [Ailuropoda melanoleuca]
Length=192

GENE ID: 100474756 LOC100474756 | UPF0308 protein C9orf21-like
[Ailuropoda melanoleuca]

Score = 47.8 bits (112), Expect = 5e-04, Method: Compositional matrix adjust.
Identities = 32/99 (33%), Positives = 50/99 (51%), Gaps = 8/99 (8%)

```

Query 1  LVIWPRHFGCVCRLVKELAAREQAFROERKWRMLVIGCGTPEQAAKFRTDNLGPLPA 60
          +V++ RHF C +C+ V++LA ++F QE L+VIG 0++ +
Sbjct 25  VVVFVRHFLCYICKEYVEDLAKIPKSFLEADVTLVIG-----QSSYHHIEPFCKLTGY 79
Query 61  SIPVYTDPLRRSYLALRFRRG--ILSTFNLPALQ-NILD 96
          S +Y DP R Y L +RG I S+ P ++ NIL
Sbjct 80  SHEIYVDPEREIYKKGMRGEEIASSGKSPHIKSNILS 118

```

>gb|EFB20490.1| hypothetical protein PANDA_018799 [Ailuropoda melanoleuca]
Length=186

Score = 47.8 bits (112), Expect = 5e-04, Method: Compositional matrix adjust.
Identities = 32/99 (33%), Positives = 50/99 (51%), Gaps = 8/99 (8%)

```

Query 1  LVIWPRHFGCVCRLVKELAAREQAFROERKWRMLVIGCGTPEQAAKFRTDNLGPLPA 60
          +V++ RHF C +C+ V++LA ++F QE L+VIG 0++ +
Sbjct 19  VVVFVRHFLCYICKEYVEDLAKIPKSFLEADVTLVIG-----QSSYHHIEPFCKLTGY 73
Query 61  SIPVYTDPLRRSYLALRFRRG--ILSTFNLPALQ-NILD 96
          S +Y DP R Y L +RG I S+ P ++ NIL
Sbjct 74  SHEIYVDPEREIYKKGMRGEEIASSGKSPHIKSNILS 112

```

>gb|EAW92663.1| chromosome 9 open reading frame 21, isoform CRA_c [Homo sapiens]
Length=182

GENE ID: 195827 C9orf21 | chromosome 9 open reading frame 21 [Homo sapiens]
(10 or fewer PubMed links)

Score = 47.8 bits (112), Expect = 5e-04, Method: Compositional matrix adjust.
Identities = 37/118 (32%), Positives = 56/118 (48%), Gaps = 14/118 (11%)

```

Query 1  LVIWPRHFGCVCRLVKELAAREQAFROERKWRMLVIGCGTPEQAAKFRTDNLGPLPA 60
          +V++ RHF C +C+ V++LA ++F QE L+VIG 0++ +
Sbjct 59  VVVFVRHFLCYICKEYVEDLAKIPKSFLEADVTLVIG-----QSSYHHIEPFCKLTGY 113
Query 61  SIPVYTDPLRRSYLALRFRRG--ILSTFNLPALQNILDSFQKGERQRWDMIPDAFQ 116
          S +Y DP R Y L +RG I S+ N NI F +R R D P ++ Q
Sbjct 114 SHEIYVDPEREIYKRLGMRGEEIASSGN-----NI--HFIHRDRNRLDHPKINSVLQ 164

```

>gb|ACU24130.1| unknown [Glycine max]
Length=251

Score = 47.8 bits (112), Expect = 5e-04, Method: Compositional matrix adjust.
Identities = 39/125 (32%), Positives = 52/125 (42%), Gaps = 12/125 (9%)

```

Query 1  LVIWPRHFGCVCRLVKELAAREQAFROERKWRMLVIGCGTPEQAAKFRTDNLGPLPA 60
          +V RHFGC C L L + F +L+ +G GTP +A R A P P
Sbjct 95  VVALLRHFGCPCWELASALKESKARF-DSAGVKLIAVGIGTPNKA---RMLAERLPFPL 150
Query 61  SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNI---LDSFQKGERQRWDMIPPD----A 113
          +Y DP R++Y L G TF P+ + D+ QK + PD
Sbjct 151 DC-LYADPDRKAYHVLNLYYGFGRTFFNPSSIKVFSRFDALQKAVKNYTIATPDDRSVG 209
Query 114 FQGG 118
          QGG
Sbjct 210 LQQGG 214

```

>ref|XP_002666723.1| **UGM** PREDICTED: UPF0308 protein C9orf21 homolog [Danio rerio]
Length=223

GENE ID: 100332791 LOC100332791 | UPF0308 protein C9orf21 homolog [Danio rerio]

Score = 47.8 bits (112), Expect = 5e-04, Method: Compositional matrix adjust.
Identities = 29/81 (36%), Positives = 39/81 (49%), Gaps = 5/81 (6%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRMLMVGCGTPEQAAKFRTDATLGPLPA 60
          +VI+ RHF C C+ V++L Q Q+ RL+VIG + F +L P
Sbjct 55  IVIFVRHFLCYTCKEYVEDLGKIPQHVLQDSNVRLVVIGQSSYSHIQGF---CSLTGFPH 111

Query 61  SIPVYTDPLRRSYLALRFRRG 81
          I Y DP R+ Y L RRG
Sbjct 112 EI--YVDPERQIYKRLGLRRG 130
```

>ref|XP_002742741.1| **GM** PREDICTED: UPF0308 protein C9orf21-like [Callithrix jacchus]
Length=186

GENE ID: 100415410 LOC100415410 | UPF0308 protein C9orf21-like
[Callithrix jacchus]

Score = 47.4 bits (111), Expect = 7e-04, Method: Compositional matrix adjust.
Identities = 36/118 (31%), Positives = 57/118 (49%), Gaps = 14/118 (11%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRMLMVGCGTPEQAAKFRTDATLGPLPA 60
          +V++ RHF C C+ V++LA ++F QE L+VIG Q++ ++
Sbjct 63  VVVFVRHFLCYICKKEYVEDLARIPKSFLEQANVTLIVIG-----QSSYHHIESFCKLTGY 117

Query 61  SIPVYTDPLRRSYLALRFRRG--ILSTFNLPALQNILDSFQKGERQRWDMIPDAFQQ 116
          S +Y DP R Y L +RG I S+ N NI F ++ R D P ++ Q
Sbjct 118 SHEIYVDPEREIYKRLGMKRGEIASSGN-----NI--HFIHRDKNRLDHPKINSVLQ 168
```

>ref|XP_002463941.1| **UG** hypothetical protein SORBIDRAFT_01g009340 [Sorghum bicolor]
gb|EER90939.1| G hypothetical protein SORBIDRAFT_01g009340 [Sorghum bicolor]
Length=259

GENE ID: 8060536 SORBDRAFT_01g009340 | hypothetical protein [Sorghum bicolor]
(10 or fewer PubMed links)

Score = 47.0 bits (110), Expect = 8e-04, Method: Compositional matrix adjust.
Identities = 40/120 (34%), Positives = 51/120 (43%), Gaps = 12/120 (10%)

```
Query 6  RHFGCVLCRRLVKELAAREQAFRQERKWRMLMVGCGTPEQAAKFRTDATLGPLPASIPVY 65
          RHFGC C L L F +L+ IG GTPE+A R A P P +Y
Sbjct 108 RHFGCFCCWELASVLKDSIAKFD SAGA-KLIAIGVGTPEKA---RILADRLPFPLD-SLY 162

Query 66  TDPLRRSYLALRFRRGILSTFNLPALQNI---LDSFQKGERQ-RWDMIPPD---AFQGG 118
          DP R++Y L G+ T PA I LD ++ + + P D QGG
Sbjct 163 ADPERKAYNVGLGLYHGLGRTLFPASAKIYSRLDYIKEATKNYTL EATPADLTGVLQGG 222
```

>ref|XP_002463942.1| **UG** hypothetical protein SORBIDRAFT_01g009350 [Sorghum bicolor]
gb|EER90940.1| G hypothetical protein SORBIDRAFT_01g009350 [Sorghum bicolor]
Length=260

GENE ID: 8059958 SORBDRAFT_01g009350 | hypothetical protein [Sorghum bicolor]
(10 or fewer PubMed links)

Score = 47.0 bits (110), Expect = 8e-04, Method: Compositional matrix adjust.
Identities = 39/120 (33%), Positives = 52/120 (44%), Gaps = 12/120 (10%)

```
Query 6  RHFGCVLCRRLVKELAAREQAFRQERKWRMLMVGCGTPEQAAKFRTDATLGPLPASIPVY 65
          RHFGC C L L ++ F +L+ +G GTP AK R A P P +Y
Sbjct 109 RHFGCPCCWELASVLRDTKEKF-DSAGVKLIAVGVGTP---AKARILAERLPPLEY-LY 163

Query 66  TDPLRRSYLALRFRRGILSTFNLPALQNI---LDSFQKGERQRWDMIPPD---AFQGG 118
          DP R++Y L G+ TF PA + DS ++ + PD QGG
Sbjct 164 ADPDRKAYNLLGLYFGVGRFFNPASAKVFSRFDLSLKEAVKNYMEATPDDRAGVLQGG 223
```

>ref|XP_003130904.1| **UGM** PREDICTED: UPF0308 protein C9orf21 homolog isoform 1 [Sus scr]
Length=228

GENE ID: 100524271 LOC100524271 | UPF0308 protein C9orf21 homolog [Sus scrofa]

Score = 47.0 bits (110), Expect = 0.001, Method: Compositional matrix adjust.
Identities = 32/99 (33%), Positives = 50/99 (51%), Gaps = 8/99 (8%)

```
Query 1  LVIWPRHFGVCVLCRRLVKELAAREQAFROERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V++ RHF C +C+ V++LA ++F QE L+VIG Q++ +
Sbjct 61  VVVFVRHFLCYICKEYVEDLAKIPKSFLQEANVTLVIG-----QSSYHHIEPFCKLTGY 115

Query 61  SIPVYTDPLRRSYLALRFRRG--ILSTFNLPALQ-NILD 96
          S +Y DP R Y L +RG I S+ P ++ NIL
Sbjct 116 SHEIYVDPEREIYKRLGMKRGEEIASSGKSPHIKSNILS 154
```

>ref|NP_001069904.1| **UGM** hypothetical protein LOC616897 [Bos taurus]
sp|Q148E0.1|CI021_BOVIN **G** RecName: Full=UPF0308 protein C9orf21 homolog
gb|AAI18426.1| **G** Chromosome 9 open reading frame 21 ortholog [Bos taurus]
gb|DAA26600.1| **G** hypothetical protein LOC616897 [Bos taurus]
Length=228

GENE ID: 616897 C8H9orf21 | chromosome 9 open reading frame 21 ortholog
[Bos taurus] (10 or fewer PubMed links)

Score = 47.0 bits (110), Expect = 0.001, Method: Compositional matrix adjust.
Identities = 32/99 (33%), Positives = 50/99 (51%), Gaps = 8/99 (8%)

```
Query 1  LVIWPRHFGVCVLCRRLVKELAAREQAFROERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V++ RHF C +C+ V++LA ++F QE L+VIG Q++ +
Sbjct 61  IVVFVRHFLCYICKEYVEDLAKIPKSFLQEANVTLVIG-----QSSYHHIEPFCKLTGY 115

Query 61  SIPVYTDPLRRSYLALRFRRG--ILSTFNLPALQ-NILD 96
          S +Y DP R Y L +RG I S+ P ++ NIL
Sbjct 116 SHEIYVDPEREIYKRLGMKRGEEIASSGQSPHVKNILS 154
```

>ref|XP_002269002.1| **UGM** PREDICTED: hypothetical protein, partial [Vitis vinifera]
Length=223

GENE ID: 100245543 LOC100245543 | hypothetical protein LOC100245543
[Vitis vinifera]

Score = 46.6 bits (109), Expect = 0.001, Method: Compositional matrix adjust.
Identities = 38/125 (31%), Positives = 51/125 (41%), Gaps = 12/125 (9%)

```
Query 1  LVIWPRHFGVCVLCRRLVKELAAREQAFROERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V RHFGC C L L + F +L+ +G GTP +A A P P
Sbjct 67  VVALLRHFGFCFCWELASALKESKATF-DSAGVKLIAVGVGTPNKACIL--AERLPFPM 122

Query 61  SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNI---LDSFQKGERQRWDMIPPD----A 113
          +Y DP R++Y L G+ T PA + +S QK + PD
Sbjct 123 DC-LYADPDRKAYDVLGLYYGLSRTLFSASPASAKVFSRFESLQKALKNYTLEGTDDKSGV 181

Query 114 FQQGG 118
          QQQG
Sbjct 182 LQQGG 186
```

>ref|NP_001170360.1| **UG** hypothetical protein LOC100384338 [Zea mays]
gb|ACN36745.1| **G** unknown [Zea mays]
Length=162

GENE ID: 100384338 pco101707a | LOC100384338 [Zea mays]

Score = 46.2 bits (108), Expect = 0.001, Method: Compositional matrix adjust.
Identities = 40/125 (32%), Positives = 54/125 (44%), Gaps = 12/125 (9%)

```
Query 1  LVIWPRHFGVCVLCRRLVKELAAREQAFROERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V RHFGC C L L ++ F +L+ +G GTP AK R A P P
Sbjct 6  VVALLRHFGPCPCWELASVLRDTERK-DSAGVKLIAVGVGTP--AKARILAEERLPFPL 61

Query 61  SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNI---LDSFQKGERQRWDMIPPD----A 113
          +Y DP R++Y L G+ TF PA + DS ++ + PD
Sbjct 62  EY-LYADPDRKAYNLLGLYFGVGRFFNPASAKVFSRFDSLKEAVKNYTIETPDDRAGV 120

Query 114 FQQGG 118
          QQQG
```

Sbjct 121 LQQGG 125

>ref|XP_001493974.2| **UGM** PREDICTED: similar to UPF0308 protein C9orf21 [Equus caballus]
Length=232

GENE ID: 100062308 LOC100062308 | hypothetical protein LOC100062308
[Equus caballus]

Score = 46.2 bits (108), Expect = 0.001, Method: Compositional matrix adjust.
Identities = 26/81 (33%), Positives = 41/81 (51%), Gaps = 5/81 (6%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFROERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V++ RHF C +C+ V++LA ++F QE L+VIG O++ +
Sbjct 65  VVVFMRHFLCYICKEYVEDLAKIPKSFLQEANVTLVIG-----QSSYHHIEPFCKLTGY 119

Query 61  SIPVYTDPLRRSYLALRFRRG 81
          S +Y DP R Y L +RG
Sbjct 120  SHEIYVDPEREIYKRLGMKRG 140
```

>emb|CBI33289.3| unnamed protein product [Vitis vinifera]
Length=255

Score = 46.2 bits (108), Expect = 0.001, Method: Compositional matrix adjust.
Identities = 38/125 (31%), Positives = 51/125 (41%), Gaps = 12/125 (9%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFROERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V RHFGC C L L + F +L+ +G GTP +A A P P
Sbjct 99  VVALLRHFGFCFCWELASALKESKATF-DSAGVKLIAVGVGTPNKACIL--AERLPFPM 154

Query 61  SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNI---LDSFQKGERQRWDMIPPD----A 113
          +Y DP R++Y L G+ T PA + +S OK + PD
Sbjct 155  DC-LYADPDRKAYDVLGLYYGLSRTLFSASAKVFSRFESLQKALKNYTLEGTDPDDKSGV 213

Query 114  FQQGG 118
           QQQG
Sbjct 214  LQQGG 218
```

>gb|ACN25853.1| **G** unknown [Zea mays]
Length=261

GENE ID: 100384338 pco101707a | LOC100384338 [Zea mays]

Score = 46.2 bits (108), Expect = 0.001, Method: Compositional matrix adjust.
Identities = 39/120 (33%), Positives = 52/120 (44%), Gaps = 12/120 (10%)

```
Query 6  RHFVCVLCRRLVKELAAREQAFROERKWRLMVIGCGTPEQAAKFRTDNLGPLPASIPVY 65
          RHFVC C L L ++ F +L+ +G GTP AK R A P P +Y
Sbjct 110  RHFVCPCWELASVLRDTERF-DSAGVKLIAVGVGTP---AKARILAERLPPLEY-LY 164

Query 66  TDPLRRSYLALRFRRGILSTFNLPALQNI---LDSFQKGERQRWDMIPPD----AFQQGG 118
          DP R++Y L G+ TF PA + DS ++ + PD QQQG
Sbjct 165  ADPDRKAYNLLGLYFGVGRFFNPASAKVFSRFDLKEAVKNYTIEATPDDRAGVLQQGG 224
```

>ref|XP_002820049.1| **GM** PREDICTED: UPF0308 protein C9orf21-like [Pongo abelii]
Length=226

GENE ID: 100438036 LOC100438036 | UPF0308 protein C9orf21-like [Pongo abelii]

Score = 46.2 bits (108), Expect = 0.001, Method: Compositional matrix adjust.
Identities = 31/99 (32%), Positives = 50/99 (51%), Gaps = 7/99 (7%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFROERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V++ RHF C +C+ V++LA ++F QE L+VIG O++ +
Sbjct 59  VVVFVRHFLCYICKEYVEDLAKIPKSFLQEANVTLVIG-----QSSYHHIEPFCKLTGY 113

Query 61  SIPVYTDPLRRSYLALRFRRG--ILSTFNLPALQNILDS 97
          S +Y DP R Y L +RG I S+ P +++ L S
Sbjct 114  SHEIYVDPEREIYKRLGMKRGEEIASSGQSPHVKSNNLS 152
```

>ref|XP_520707.2| **UGM** PREDICTED: similar to TPA_exp: C90RF21 isoform 2 [Pan troglodyte]
Length=226

GENE ID: 465250 LOC465250 | similar to TPA_exp: C90RF21 [Pan troglodytes]

Score = 46.2 bits (108), Expect = 0.001, Method: Compositional matrix adjust.
Identities = 31/99 (32%), Positives = 50/99 (51%), Gaps = 7/99 (7%)

```
Query 1  LVIWPRHFGCVCRLVKELAAAREQAFROERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V++ RHF C +C+ V++LA ++F QE L+VIG 0++ +
Sbjct 59  VVVFVRHFLCYICKEYVEDLAKIPKSFLQEANVTLVIG-----QSSYHHIEPFCKLTGY 113

Query 61  SIPVYTDPLRRSYLALRFRRG--ILSTFNLPALQNILDS 97
          S +Y DP R Y L +RG I S+ P +++ L S
Sbjct 114 SHEIYVDPEREIYKRLGMKRGEEIASSGQSPHVKSNNLS 152
```

>ref|NP_714542.1| **UGM** hypothetical protein LOC195827 [Homo sapiens]
sp|Q7RTV5.1|CI021_HUMAN **G** RecName: Full=UPF0308 protein C9orf21
tpg|DAA00065.1| **G** TPA_exp: C9ORF21 [Homo sapiens]
emb|CAI40534.1| **G** novel protein [Homo sapiens]
gb|EAW92662.1| **G** chromosome 9 open reading frame 21, isoform CRA_b [Homo sapiens]
gb|AAI36504.1| **G** Chromosome 9 open reading frame 21 [Homo sapiens]
Length=226

GENE ID: 195827 C9orf21 | chromosome 9 open reading frame 21 [Homo sapiens]
(10 or fewer PubMed links)

Score = 45.8 bits (107), Expect = 0.002, Method: Compositional matrix adjust.
Identities = 31/99 (32%), Positives = 50/99 (51%), Gaps = 7/99 (7%)

```
Query 1  LVIWPRHFGCVCRLVKELAAAREQAFROERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V++ RHF C +C+ V++LA ++F QE L+VIG 0++ +
Sbjct 59  VVVFVRHFLCYICKEYVEDLAKIPRSFLQEANVTLVIG-----QSSYHHIEPFCKLTGY 113

Query 61  SIPVYTDPLRRSYLALRFRRG--ILSTFNLPALQNILDS 97
          S +Y DP R Y L +RG I S+ P +++ L S
Sbjct 114 SHEIYVDPEREIYKRLGMKRGEEIASSGQSPHIKSNLLS 152
```

>gb|EAW92661.1| **G** chromosome 9 open reading frame 21, isoform CRA_a [Homo sapiens]
Length=214

GENE ID: 195827 C9orf21 | chromosome 9 open reading frame 21 [Homo sapiens]
(10 or fewer PubMed links)

Score = 45.4 bits (106), Expect = 0.002, Method: Compositional matrix adjust.
Identities = 30/93 (33%), Positives = 48/93 (52%), Gaps = 7/93 (7%)

```
Query 1  LVIWPRHFGCVCRLVKELAAAREQAFROERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V++ RHF C +C+ V++LA ++F QE L+VIG 0++ +
Sbjct 59  VVVFVRHFLCYICKEYVEDLAKIPRSFLQEANVTLVIG-----QSSYHHIEPFCKLTGY 113

Query 61  SIPVYTDPLRRSYLALRFRRG--ILSTFNLPAL 91
          S +Y DP R Y L +RG I S+ +L +L
Sbjct 114 SHEIYVDPEREIYKRLGMKRGEEIASSGSLQSL 146
```

>ref|XP_001106503.1| **UGM** PREDICTED: UPF0308 protein C9orf21-like [Macaca mulatta]
Length=226

GENE ID: 710488 LOC710488 | similar to UPF0308 protein C9orf21 [Macaca mulatta]

Score = 45.4 bits (106), Expect = 0.003, Method: Compositional matrix adjust.
Identities = 31/99 (32%), Positives = 50/99 (51%), Gaps = 7/99 (7%)

```
Query 1  LVIWPRHFGCVCRLVKELAAAREQAFROERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V++ RHF C +C+ V++LA ++F QE L+VIG 0++ +
Sbjct 59  VVVFVRHFLCYICKEYVEDLAKIPKSFLQEANVTLVIG-----QSSYHHIEPFCKLTGY 113

Query 61  SIPVYTDPLRRSYLALRFRRG--ILSTFNLPALQNILDS 97
          S +Y DP R Y L +RG I S+ P +++ L S
Sbjct 114 SHEIYVDPEREIYKRLGMKRGEEIASSGQSPHVKSNNLS 152
```

>ref|XP_002191279.1| **UGM** PREDICTED: hypothetical protein [Taeniopygia guttata]
Length=222

GENE ID: 100224761 LOC100224761 | hypothetical protein LOC100224761
[Taeniopygia guttata]

Score = 45.4 bits (106), Expect = 0.003, Method: Compositional matrix adjust.
Identities = 35/125 (28%), Positives = 56/125 (45%), Gaps = 12/125 (9%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGGLPA 60
          +V++ R+F C  C+ V++LA  +AF QE  RL+VIG  +  F+
Sbjct 55  IVLFVRNFLCYTCKEYVEDLAKVPKAFLESNVRLVIVIGQSSYHHIKPFCSLTGY----- 109

Query 61  SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNILDSFQKGERQRWDMIPDAF----- 114
          + +Y DP R  Y  L  +RG  +  ++  +  ++F  R  W  +  AF
Sbjct 110  THEMVDPPREIYKILGMKRGEKNKASVRSPhVKSNTFLGSIWIRAMTGPADFQGD 169

Query 115 -OQGG 118
           OQGG
Sbjct 170 AQOQGG 174
```


>emb|CAN81555.1| hypothetical protein VITISV_040397 [Vitis vinifera]
Length=201

Score = 45.1 bits (105), Expect = 0.003, Method: Compositional matrix adjust.
Identities = 38/125 (31%), Positives = 51/125 (41%), Gaps = 12/125 (9%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGGLPA 60
          +V  RHFGC  C  L  L  +  F  +L+ +G GTP +A  A  P  P
Sbjct 45  VVALLRHFGCFCCWELASALKESKARF-DSAGVKLIAVGVGTPNKACIL---AERLPFPM 100

Query 61  SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNI---LDSFQKGERQRWDMIPD---A 113
          +Y DP R++Y  L  G+  T  PA  +  +S  QK  +  PD
Sbjct 101  DC-LYADPDRKAYDVLGLYYGLSRTLFSASAKVFSRFESLQKALKNYTLEGTPDDKSGV 159

Query 114 FQOQGG 118
           OQGG
Sbjct 160 LQOQGG 164
```

>ref|XP_002263959.1|  PREDICTED: hypothetical protein [Vitis vinifera]
Length=255

GENE ID: 100249728 LOC100249728 | hypothetical protein LOC100249728
[Vitis vinifera] (10 or fewer PubMed links)

Score = 45.1 bits (105), Expect = 0.003, Method: Compositional matrix adjust.
Identities = 38/125 (31%), Positives = 51/125 (41%), Gaps = 12/125 (9%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGGLPA 60
          +V  RHFGC  C  L  L  +  F  +L+ +G GTP +A  A  P  P
Sbjct 99  VVALLRHFGCFCCWELASALKESKARF-DSAGVKLIAVGVGTPNKACIL---AERLPFPM 154

Query 61  SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNI---LDSFQKGERQRWDMIPD---A 113
          +Y DP R++Y  L  G+  T  PA  +  +S  QK  +  PD
Sbjct 155  DC-LYADPDRKAYDVLGLYYGLSRTLFSASAKVFSRFESLQKALKNYTLEGTPDDKSGV 213

Query 114 FQOQGG 118
           OQGG
Sbjct 214 LQOQGG 218
```

>gb|EAY91739.1| hypothetical protein OsI_13380 [Oryza sativa Indica Group]
Length=259

Score = 45.1 bits (105), Expect = 0.003, Method: Compositional matrix adjust.
Identities = 40/123 (33%), Positives = 54/123 (44%), Gaps = 10/123 (8%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGGLPA 60
          +V  RHFGC  C  L  L  +  F  +L+ IG GTP++A  R  A  P  P
Sbjct 106  VVALLRHFGCFCCWELASVLKESMAKFDAAGA-KLIAIGVGVGTPDKA---RILADGLPFPV 161

Query 61  SIPVYTDPLRRSYLALRFRRGILSTFNLPA-LQNILDSFQKGERQ-RWDMIPD---AFQ 115
          +Y DP R++Y  L  G+  T  PA  +  +L+S +K  +  PD  Q
Sbjct 162  D-SLYADPERKAYDVLGLYHGLGRTLISPAMYSGLNSIKKVTKNYTLKGT PADLTGILQ 220

Query 116 OGG 118
           OGG
Sbjct 221 OGG 223
```

>gb|EEE59859.1| hypothetical protein OsJ_12441 [Oryza sativa Japonica Group]
Length=283

Score = 45.1 bits (105), Expect = 0.003, Method: Compositional matrix adjust.

Identities = 40/123 (33%), Positives = 54/123 (44%), Gaps = 10/123 (8%)






```

Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V RHFGC C L L F +L+ IG GTP++A R A P P
Sbjct 106 VVALLRHFGCFCCWELASVLKESMAKFDAAGA-KLIAIGVGTDPKA---RILADGLPFPV 161

Query 61 SIPVYTDPLRRSYLALRFRRGILSTFNLPA-LQNILDSFQKGERQ-RWDMIPPD---AFQ 115
          +Y DP R++Y L G+ T PA + + L+S +K + P D Q
Sbjct 162 D-SLYADPERKAYDVLGLYHGLGRTLISPAKMYSGLNSIKKVTKNYTLKGT PADLTGILQ 220

Query 116 QGG 118
          QGG
Sbjct 221 QGG 223

```

>ref|NP_001051154.1|  Os03g0729300 [Oryza sativa Japonica Group]
 gb|AA038466.1|  unknown protein [Oryza sativa Japonica Group]
 gb|ABF98681.1|  UPF0308 protein, chloroplast precursor, putative, expressed [Oryza sativa Japonica Group]
 dbj|BAF13068.1|  Os03g0729300 [Oryza sativa Japonica Group]
 dbj|BAG98418.1|  unnamed protein product [Oryza sativa Japonica Group]
 Length=259

GENE ID: 4333988 Os03g0729300 | Os03g0729300 [Oryza sativa Japonica Group]
 (10 or fewer PubMed links)

Score = 45.1 bits (105), Expect = 0.003, Method: Compositional matrix adjust.
 Identities = 40/123 (33%), Positives = 54/123 (44%), Gaps = 10/123 (8%)



```

Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V RHFGC C L L F +L+ IG GTP++A R A P P
Sbjct 106 VVALLRHFGCFCCWELASVLKESMAKFDAAGA-KLIAIGVGTDPKA---RILADGLPFPV 161

Query 61 SIPVYTDPLRRSYLALRFRRGILSTFNLPA-LQNILDSFQKGERQ-RWDMIPPD---AFQ 115
          +Y DP R++Y L G+ T PA + + L+S +K + P D Q
Sbjct 162 D-SLYADPERKAYDVLGLYHGLGRTLISPAKMYSGLNSIKKVTKNYTLKGT PADLTGILQ 220

Query 116 QGG 118
          QGG
Sbjct 221 QGG 223

```

>ref|XP_002533575.1|  conserved hypothetical protein [Ricinus communis]
 gb|EEF28810.1|  conserved hypothetical protein [Ricinus communis]
 Length=255

GENE ID: 8276075 RCOM_0366250 | hypothetical protein [Ricinus communis]

Score = 45.1 bits (105), Expect = 0.004, Method: Compositional matrix adjust.
 Identities = 39/125 (32%), Positives = 50/125 (40%), Gaps = 12/125 (9%)



```

Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V RHFGC C L L + F +L+ IG G P +A R A P P
Sbjct 99  VVALLRHFGCPCCWELASVLKEAKSKF-DSAGVKLIAIGVGAPNKA---RMLADRLPFPV 154

Query 61 SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNI---LDSFQKGERQRWDMIPPD---A 113
          +Y DP R +Y L G TF PA + DS ++ + PD
Sbjct 155 DC-LYADPNREAYNVLGLYGFGRFTFFNPASAKVFSRFDLSLRQAVKNYTIETPD DRSV 213

Query 114 FQGG 118
          QGG
Sbjct 214 LQGG 218

```

>ref|XP_001588758.1|  hypothetical protein SS1G_10305 [Sclerotinia sclerotiorum 1980]
 gb|EDN94432.1|  hypothetical protein SS1G_10305 [Sclerotinia sclerotiorum 1980]
 Length=342

GENE ID: 5484650 SS1G_10305 | hypothetical protein
 [Sclerotinia sclerotiorum 1980 UF-70]

Score = 44.7 bits (104), Expect = 0.004, Method: Compositional matrix adjust.
 Identities = 29/91 (32%), Positives = 46/91 (51%), Gaps = 12/91 (13%)

```

Query 1  LVIWPRHFGCVLCRRLVKELAA---REQAFRQERKWRLMVIGCGTPEQAAKFRTDNLG 57
          L+I+ RHF C C+ ++ LAA + + + V+GCG+P ++ +AT P

```

Sbjct 87 LIIFIRHFFCGNCQEFRLTLAASITEDSLQLHTPTFIAVVGCGSPSLIPMYQ-EATKCP 145
 Query 58 LPASIPVYTDPLRRSYLALRFRRGILSTFNL 88
 P+Y DP R+ Y L G++ T NL
 Sbjct 146 F----PIYADPTRKLYDEL----GMMRTLNL 168

>ref|XP_001784902.1| **UG** predicted protein [Physcomitrella patens subsp. patens]
 gb|ED050291.1| **G** predicted protein [Physcomitrella patens subsp. patens]
 Length=187

GENE ID: 5948108 PHYPADRAFT_154541 | hvnothetical protein
 [Physcomitrella patens subsp. patens] (10 or fewer PubMed links)

Score = 44.7 bits (104), Expect = 0.004, Method: Compositional matrix adjust.
 Identities = 38/125 (31%), Positives = 56/125 (45%), Gaps = 12/125 (9%)

Query 1 LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
 +V + RHF GC C L + F ++L+ IG G P A+ ++ P PA
 Sbjct 31 IVAFLRHFGCPFCWEFAAALREAKPKF-DAAGFKLITIGVG-PSSKAQVLSEKL--PFPA 86
 Query 61 SIPVYTDPLRRSYLALRFRRGILSTFNLPALQNILDSFQK--GERQRW--DMIPPD---A 113
 +Y DP R++Y AL G+ T+ PA I K + W D++P +
 Sbjct 87 DC-LYADPDRKAYDALGLYHGVARTWLNPAQMIFTRLDKVADAVKGNRDMVDPNTAAT 145
 Query 114 FQGG 118
 QGG
 Sbjct 146 LQGG 150

>ref|XP_848380.1| **UGM** PREDICTED: similar to UPF0308 protein C9orf21 [Canis familiaris]
 Length=230

GENE ID: 606835 LOC606835 | similar to UPF0308 protein C9orf21
 [Canis lupus familiaris]

Score = 44.7 bits (104), Expect = 0.004, Method: Compositional matrix adjust.
 Identities = 31/99 (32%), Positives = 49/99 (50%), Gaps = 8/99 (8%)

Query 1 LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
 +V++ RHF C +C+ V++LA ++ OE L+VIG Q++ +
 Sbjct 63 VVVFVRHFLCYICKEYVEDLAKIPKSVLQEADITLIVIG-----QSSYHHIEPFCKLTGY 117
 Query 61 SIPVYTDPLRRSYLALRFR--GILSTFNLPALQ-NILD 96
 S +Y DP R Y L +R GI S+ P ++ NIL
 Sbjct 118 SHEIYVDPEREIYKKGKMGKREGIASSGKSPHIKSNILS 156

>ref|ZP_02926219.1| hypothetical protein VspiD_06235 [Verrucomicrobium spinosum DSM
 4136]
 Length=304

Score = 44.7 bits (104), Expect = 0.004, Method: Compositional matrix adjust.
 Identities = 30/84 (36%), Positives = 38/84 (46%), Gaps = 5/84 (5%)

Query 1 LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
 LV++ RH GC CR + ++A A R+ ++ G P A F L LPA
 Sbjct 162 LVVFLRHAGCTFCREALADIARVPAIEASGT-RIALVHMGEPAFAAFSGQYGLADLPA 220
 Query 61 SIPVYTDPLRRSYLALRFRGILS 84
 DP RR Y L RRG LS
 Sbjct 221 ----VADPSRRLYRGLGLRRGKLS 240

>ref|XP_001116318.1| **GM** PREDICTED: UPF0308 protein C9orf21-like [Macaca mulatta]
 Length=I40

GENE ID: 720662 LOC720662 | UPF0308 protein C9orf21-like [Macaca mulatta]

Score = 44.7 bits (104), Expect = 0.004, Method: Compositional matrix adjust.
 Identities = 25/75 (34%), Positives = 37/75 (50%), Gaps = 5/75 (6%)

Query 7 HFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPASIPVYT 66
 HF C +C+ V++LA ++F OE L+VIG Q++ + P S +Y
 Sbjct 65 HFLCYICKEYVEDLAKIPESFLQEANVTLIVIG-----QSSYHHIEPFCKPTGYSHEIYV 119
 Query 67 DPLRRSYLALRFRRG 81
 DP R Y L +RG
 Sbjct 120 DPEREMYKRLGMKRG 134

>ref|XP_002708298.1| **UGM** PREDICTED: hypothetical protein [Oryctolagus cuniculus]
Length=226

GENE ID: 100342762 LOC100342762 | hypothetical protein LOC100342762
[Oryctolagus cuniculus]

Score = 44.3 bits (103), Expect = 0.005, Method: Compositional matrix adjust.
Identities = 26/81 (33%), Positives = 41/81 (51%), Gaps = 5/81 (6%)

```
Query 1  LVIWPRHFGVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V++ RHF C +C+ V++LA Q+F +E L+VIG Q++ +
Sbjct 59  VVVFVRHFLCYVCKEYVEDLAKVPSFLREADVTLVIG-----QSSYHHIEPFCKLTGY 113

Query 61  SIPVYTDPLRRSYLALRFRRG 81
          S +Y DP R Y L +RG
Sbjct 114  SHEIYVDPEREIYKRLGMKRG 134
```

>ref|XP_681257.1| **G** hypothetical protein AN7988.2 [Aspergillus nidulans FGSC A4]
gb|EAA58791.1| G hypothetical protein AN7988.2 [Aspergillus nidulans FGSC A4]
tp|CBF73658.1| G TPA: conserved hypothetical protein [Aspergillus nidulans FGSC A4]
Length=226

GENE ID: 2869289 AN7988.2 | hypothetical protein [Aspergillus nidulans FGSC A4]
(10 or fewer PubMed links)

Score = 43.5 bits (101), Expect = 0.008, Method: Compositional matrix adjust.
Identities = 26/100 (26%), Positives = 48/100 (48%), Gaps = 8/100 (8%)

```
Query 1  LVIWPRHFGVLCRRLVKELAAR-----EQAFRQERKWRLMVIGCGTPEQAAKFRTDAT 54
          ++I+ RHF C+ + V+ ++ + + +LM+IGCG P + ++ T
Sbjct 49  VIIFIRHFICYDQDYVRTVSHHLTDSVLQTIKSTGGPIQLMIIGCGDPSLIVPVYVSETT 108

Query 55  LGPLPASIPVYTDPLRRSYLALRFRRGILSTFNLPALQNI 94
          G + A PVYTD P + Y L +R + + + P +
Sbjct 109  -GDI-AQFPVYTDPDGKLYEKLHMKRTLNTNIMHPPVYAQV 146
```

>gb|EDL84435.1| **G** similar to UPF0308 protein C9orf21, isoform CRA_c [Rattus norvegicus]
Length=226

GENE ID: 498685 LOC498685 | similar to UPF0308 protein C9orf21
[Rattus norvegicus]

Score = 43.1 bits (100), Expect = 0.012, Method: Compositional matrix adjust.
Identities = 30/99 (31%), Positives = 49/99 (50%), Gaps = 7/99 (7%)

```
Query 1  LVIWPRHFGVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V++ RHF C +C+ V++LA ++ QE L+VIG Q++ +
Sbjct 59  VVVFVRHFLCYVCKEYVEDLAKIPKSVLQEADVTLVIG-----QSSYHHIEPFCKLTGY 113

Query 61  SIPVYTDPLRRSYLALRFRRG--ILSTFNLPALQNILDS 97
          S +Y DP R Y L +RG I S+ P +++ L S
Sbjct 114  SHEIYVDPEREIYKRLGMKRGEEISSSGQSPHIKSNLLS 152
```

>dbj|BAE26709.1| **G** unnamed protein product [Mus musculus]
Length=182

GENE ID: 66129 1110018J18Rik | RIKEN cDNA 1110018J18 gene [Mus musculus]
(Over 10 PubMed links)

Score = 43.1 bits (100), Expect = 0.013, Method: Compositional matrix adjust.
Identities = 35/118 (30%), Positives = 55/118 (47%), Gaps = 14/118 (11%)

```
Query 1  LVIWPRHFGVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +V++ RHF C +C+ V++LA ++ +E L+VIG Q++ +
Sbjct 59  VVVFVRHFLCYVCKEYVEDLAKIPKSVLREADVTLVIG-----QSSYHHIEPFCKLTGY 113

Query 61  SIPVYTDPLRRSYLALRFRRG--ILSTFNLPALQNILDSFQKGERQRWDMIPPAFQQ 116
          S +Y DP R Y L +RG I S+ N NI F +R R D P ++ Q
Sbjct 114  SHEIYVDPEREIYKRLGMKRGEEISSSGN-----NI--HFVHRDRDRDLHDKPINSVLQ 164
```

>ref|ZP_01909602.1| hypothetical protein PPSIR1_24939 [Plesiocystis pacifica SIR-1]

gb|EDM77490.1 | hypothetical protein PPSIR1_24939 [Plesiocystis pacifica SIR-1]
Length=237

Score = 42.7 bits (99), Expect = 0.014, Method: Compositional matrix adjust.
Identities = 38/122 (32%), Positives = 53/122 (44%), Gaps = 12/122 (9%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          L+++ RHF C C+ V+EL+ R R + IG G EQ A F L
Sbjct 76  LLLFLRHYACVGCQRVEELSPRVLEL-SALGVRVAVFIGSGDAEQMAAFAQRHAL--TDK 132

Query 61  SIPVYTDPLRRSYLALRFRRGILSTFNLPA----LQNILDSFQKGERQRWDMIPPDAFQQ 116
          ++ + DP SY A RR + S A L I+D F+ G + + QO
Sbjct 133  AVVMVCDPALGSYAAADMRRSLWSAVEPRAWWYQLVAIMDGFPRGRG----LQGEVTQQ 187

Query 117  GG 118
          GG
Sbjct 188  GG 189
```

>**ref|XP_001829670.2** | **G** hypothetical protein CC1G_11406 [Coprinopsis cinerea okayama7#136]

gb|EAU92121.2 | **G** hypothetical protein CC1G_11406 [Coprinopsis cinerea okayama7#130]
Length=222

GENE ID: 6006103 CC1G_11406 | hypothetical protein
[Coprinopsis cinerea oKayama7#130] (10 or fewer PubMed links)

Score = 42.4 bits (98), Expect = 0.018, Method: Compositional matrix adjust.
Identities = 28/108 (26%), Positives = 47/108 (44%), Gaps = 16/108 (14%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          +VI+ RHF C C+ V+ +A + ++ +++V+GCG + +
Sbjct 55  VVIFIRHFFCGSCQMYVEHVAKVPASNLEQAGVQIVVVGCGDWKAIPMYHETTGF----- 109

Query 61  SIPVYTDPLRRSYLALRFRRGILSTF-----NLPALQNILDS 97
          + P+Y DP R Y AL L N+P +QN+L S
Sbjct 110  TGPIYADPNRDLYFALGMTLQNLLEMPKQPRPSYLANVPVQNVLQS 157
```

>**ref|XP_002628008.1** | **G** conserved hypothetical protein [Ajellomyces dermatitidis SLH14081]

gb|EEQ72686.1 | **G** conserved hypothetical protein [Ajellomyces dermatitidis SLH14081]

gb|EEQ88802.1 | conserved hypothetical protein [Ajellomyces dermatitidis ER-3]
Length=225

GENE ID: 8507079 BDBG_00916 | hypothetical protein
[Ajellomyces dermatitidis SLH14081]

Score = 42.4 bits (98), Expect = 0.020, Method: Compositional matrix adjust.
Identities = 25/71 (36%), Positives = 36/71 (51%), Gaps = 6/71 (8%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQ-ERKWRLMVIGCGTPEQAAKFRTDNLGPLP 59
          L+I+ RHF C C+ VK ++ A Q ++++GCG A +R D T P
Sbjct 51  LIIFIRHFFCGSCQDYVKAVSESIPAPSQLPNTNTIIIVGCGASSLIATYR-DTTKCPF- 108

Query 60  ASIPVYTDPLR 70
          P+Y TDP R
Sbjct 109  ---PIYTDPTR 116
```

>**gb|EFQ95671.1** | hypothetical protein PTT_05482 [Pyrenophora teres f. teres 0-1]
Length=342

Score = 42.4 bits (98), Expect = 0.022, Method: Compositional matrix adjust.
Identities = 29/91 (32%), Positives = 44/91 (49%), Gaps = 12/91 (13%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAR---EQAFRQERKWRLMVIGCGTPEQAAKFRTDNLG 57
          L+I+ RHF C C+ ++ LA+ E + VIGCG P+ + +AT P
Sbjct 120  LIIFIRHFFCGNCOEYIRTLASSVKPEDLLALPTPTSITVIGCGRPDLIPMY-IEATSCP 178

Query 58  LPASIPVYTDPLRRSYLALRFRRGILSTFNL 88
          P+Y +P R+ Y L G+ T+NL
Sbjct 179  F----PIYAEPTRKLYDHL----GMTRTYNL 201
```

>**gb|ABF98680.1** | expressed protein [Oryza sativa Japonica Group]

dbj|BAG90239.1 | unnamed protein product [Oryza sativa Japonica Group]
Length=192

Score = 42.0 bits (97), Expect = 0.023, Method: Compositional matrix adjust.
Identities = 32/90 (36%), Positives = 43/90 (48%), Gaps = 5/90 (5%)

Query 1 LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
 +V RHF C L L ++ F +L+ +G GTP++A R A P P
 Sbjct 102 VVALLRHFGPCPCWELASVLRDTRKERF-DSAGVKLIAVGVGTPDKA---RILAERLPFPL 157

Query 61 SIPVYTDPLRRSYLALRFRRGILSTFNLPA 90
 +Y DP R++Y L GI TF PA
 Sbjct 158 DY-LYADPERKAYDLLGLYFGIGRTFFNPA 186

>ref|XP_001547900.1| **G** hypothetical protein BC1G_13584 [Botryotinia fuckeliana B05.10]
 gb|EDN19735.1| **G** hypothetical protein BC1G_13584 [Botryotinia fuckeliana B05.10]
 Length=364

GENE ID: 5428362 BC1G_13584 | hypothetical protein
 [Botryotinia fuckeliana B05.10]

Score = 42.0 bits (97), Expect = 0.025, Method: Composition-based stats.
 Identities = 29/93 (32%), Positives = 46/93 (50%), Gaps = 12/93 (12%)

Query 1 LVIWPRHFGCVLCRRLVKELA---AREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLG 57
 LVI+ RHF C C+ ++ LA + + + V+GCG+P ++ +AT P
 Sbjct 110 LVIFIRHFFCGNCQEYLRTLATSITEDSLLQLHTPTFIAVVGCGSPSLIPMYQ-EATNCP 168

Query 58 LPASIPVYTDPLRRSYLALRFRRGILSTFNLPA 90
 P+Y DP ++ Y L G++ T NL A
 Sbjct 169 F----PIYADPTKLYDEL----GMMRTLNLGA 193

>ref|NP_001102584.1| **UG** hypothetical protein LOC498685 [Rattus norvegicus]
 gb|EDL84433.1| **G** similar to UPF0308 protein C9orf21, isoform CRA_a [Rattus norvegicus]
 Length=147

GENE ID: 498685 LOC498685 | similar to UPF0308 protein C9orf21
 [Rattus norvegicus]

Score = 42.0 bits (97), Expect = 0.026, Method: Compositional matrix adjust.
 Identities = 25/81 (31%), Positives = 40/81 (50%), Gaps = 5/81 (6%)

Query 1 LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
 +V++ RHF C +C+ V++LA ++ QE L+VIG Q++ + S +Y
 Sbjct 59 VVVFVRHFLCYVCKEYVEDLAKIPKSVLQEADVTLIVIG-----QSSYHHIEPFCKLTGY 113

Query 61 SIPVYTDPLRRSYLALRFRRG 81
 S +Y DP R Y L +RG
 Sbjct 114 SHEIYVDPEREIYKRLGMKRG 134

>gb|AAI36502.1| **G** C9orf21 protein [Homo sapiens]
 Length=172

GENE ID: 195827 C9orf21 | chromosome 9 open reading frame 21 [Homo sapiens]
 (10 or fewer PubMed links)

Score = 42.0 bits (97), Expect = 0.027, Method: Compositional matrix adjust.
 Identities = 29/94 (31%), Positives = 46/94 (49%), Gaps = 7/94 (7%)

Query 6 RHF C +C+ V++LA ++F QE L+VIG Q++ + S +Y
 Sbjct 10 QHFLCYICKEYVEDLAKIPRSFLQEANVTIVIG-----QSSYHHIEPFCKLTGYSHEIY 64

Query 66 TDPLRRSYLALRFRRG--ILSTFNLPALQNILDS 97
 DP R Y L +RG I S+ P +++ L S
 Sbjct 65 VDPEREIYKRLGMKRGEEIASSGQSPHIKSNLLS 98

>ref|XP_001940231.1| **G** conserved hypothetical protein [Pyrenophora tritici-repentis
 Pt-1C-BFP]
 gb|EDU42950.1| **G** conserved hypothetical protein [Pyrenophora tritici-repentis
 Pt-1C-BFP]
 Length=342

GENE ID: 6348196 PTRG_09899 | hypothetical protein
 [Pyrenophora tritici-repentis Pt-1C-BFP]


Score = 42.0 bits (97), Expect = 0.028, Method: Compositional matrix adjust.
 Identities = 29/91 (32%), Positives = 44/91 (49%), Gaps = 12/91 (13%)

```

Query 1  LVIWPRHFGVCVLCRRLVKELAAR---EQAFRQERKWRLMVIGCGTPEQAAKFRTDATALGP 57
          L+I+ RHF C C+ ++ LA+ E + VIGCG P+ + +AT P
Sbjct 120 LIIFIRHFFCGNCQEYIRTLASSVKPEDLLALPTPTSITVIGCGRPDLIPMY-IEATGCP 178

Query 58 LPASIPVYTDPLRRSYLALRFRRGILSTFNL 88
          P+Y +P R+ Y L G+ T+NL
Sbjct 179 F----PIYAEPTRKLYDHL----GMTRTYNL 201

```

>emb|CAI40533.1|  novel protein [Homo sapiens]
Length=179

GENE ID: 195827 C9orf21 | chromosome 9 open reading frame 21 [Homo sapiens]
(10 or fewer PubMed links)

Score = 41.6 bits (96), Expect = 0.032, Method: Compositional matrix adjust.
Identities = 29/94 (31%), Positives = 46/94 (49%), Gaps = 7/94 (7%)

```

Query 6  RHFGVCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDATALGPLPASIPVY 65
          +HF C +C+ V++LA ++F QE L+VIG Q++ + S +Y
Sbjct 17  QHFLCYICKEYVEDLAKIPRSFLQEANVTLLVIG----QSSYHHIEPFCKLTGYSHEIY 71

Query 66 TDPLRRSYLALRFRRG--ILSTFNLPALQNILDS 97
          DP R Y L +RG I S+ P +++ L S
Sbjct 72  VDPEREIYKRLGMKRGEEIASSGQSPHIKSNLLS 105

```

>emb|CBI57389.1| unnamed protein product [Sordaria macrospora]
Length=411


Score = 41.2 bits (95), Expect = 0.040, Method: Composition-based stats.
Identities = 36/113 (32%), Positives = 51/113 (46%), Gaps = 15/113 (13%)

```

Query 1  LVIWPRHFGVCVLCRRLVKELAAR---EQAFRQERKWRLMVIGCGTPEQAAKFRTDATALGP 57
          L+I+ RHF C C+ ++ LA+ + R + VIGCG P 0 + AT P
Sbjct 124 LIIFVRHFFCGQCQEYLRTLASITPDALLRLPLSTFIAVIGCGDP-QLIDMYAQATNCP 182

Query 58 LPASIPVYTDPLRRSYLALRFRRGILSTFNLPA-----LQNILDSFQKGERO 104
          P+Y DP R+ Y L R L+ PA ++ LDS +G +Q
Sbjct 183 F----PIYADPTRKLYQELGMVR-TLALGEKPAYTSKNIVKTSLSILQGIKQ 230

```

>dbj|BAB24662.1|  unnamed protein product [Mus musculus]
Length=186

GENE ID: 66129 1110018J18Rik | RIKEN cDNA 1110018J18 gene [Mus musculus]
(Over 10 PubMed links)


Score = 41.2 bits (95), Expect = 0.041, Method: Compositional matrix adjust.
Identities = 29/99 (30%), Positives = 49/99 (50%), Gaps = 7/99 (7%)

```

Query 1  LVIWPRHFGVCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDATALGPLPA 60
          +V++ RHF C +C+ V++LA ++ +E L+VIG Q++ +
Sbjct 19  VVVFVRHFLCYVCKEYVEDLAKIPKSVLREADVTLVIG----QSSYHHIEPFCKLTGY 73

Query 61 SIPVYTDPLRRSYLALRFRRG--ILSTFNLPALQNILDS 97
          S +Y DP R Y L +RG I S+ P +++ L S
Sbjct 74  SHEIYVDPEREIYKRLGMKRGEEISSGQSPHIKSNLLS 112

```

>dbj|BAE40544.1|  unnamed protein product [Mus musculus]
Length=194

GENE ID: 66129 1110018J18Rik | RIKEN cDNA 1110018J18 gene [Mus musculus]
(Over 10 PubMed links)

Score = 41.2 bits (95), Expect = 0.043, Method: Compositional matrix adjust.
Identities = 29/99 (30%), Positives = 49/99 (50%), Gaps = 7/99 (7%)

```

Query 1  LVIWPRHFGVCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDATALGPLPA 60
          +V++ RHF C +C+ V++LA ++ +E L+VIG Q++ +
Sbjct 27  VVVFVRHFLCYVCKEYVEDLAKIPKSVLREADVTLVIG----QSSYHHIEPFCKLTGY 81

Query 61 SIPVYTDPLRRSYLALRFRRG--ILSTFNLPALQNILDS 97
          S +Y DP R Y L +RG I S+ P +++ L S
Sbjct 82  SHEIYVDPEREIYKRLGMKRGEEISSGQSPHIKSNLLS 120

```

>ref|NP_079646.1| **UGM** hypothetical protein LOC66129 [Mus musculus]
 sp|Q9D1A0.1|CI021 MOUSE **G** RecName: Full=UPF0308 protein C9orf21 homolog
 dbj|BAB22993.1| **G** unnamed protein product [Mus musculus]
 dbj|BAE28518.1| **G** unnamed protein product [Mus musculus]
 gb|AAI40306.1| **G** RIKEN cDNA 1110018J18 gene [synthetic construct]
 gb|EDL16238.1| **G** RIKEN cDNA 1110018J18, isoform CRA_b [Mus musculus]
 gb|AAI56632.1| **G** RIKEN cDNA 1110018J18 gene [synthetic construct]
 Length=226

GENE ID: 66129 1110018J18Rik | RIKEN cDNA 1110018J18 gene [Mus musculus]
 (Over 10 PubMed links)

Score = 41.2 bits (95), Expect = 0.049, Method: Compositional matrix adjust.
 Identities = 29/99 (30%), Positives = 49/99 (50%), Gaps = 7/99 (7%)

Query 1 LVIWPRHFGVCVLCRRLVKELAAREQAFRQERKWRMLMVIGCGTPEQAAKFRTDNLGPLPA 60
 +V++ RHF C +C+ V++LA ++ +E L+VIG Q++ +
 Sbjct 59 VVVFVRHFLCYVCKEYVEDLAKIPKSVLREADVTLVIG-----QSSYHHIEPFCKLTGY 113
 Query 61 SIPVYTDPLRRSYLALRFRRG--ILSTFNLPALQNILDS 97
 S +Y DP R Y L +RG I S+ P +++ L S
 Sbjct 114 SHEIYVDPEREIYKRLGMKRGEEISSSGQSPHIKSNLLS 152

>gb|AD028366.1| upf0308 protein c9orf21-like protein [Ictalurus furcatus]
 Length=223

Score = 40.8 bits (94), Expect = 0.059, Method: Compositional matrix adjust.
 Identities = 24/81 (30%), Positives = 37/81 (46%), Gaps = 5/81 (6%)

Query 1 LVIWPRHFGVCVLCRRLVKELAAREQAFRQERKWRMLMVIGCGTPEQAAKFRTDNLGPLPA 60
 ++I+ RHF C C+ V++L+ Q + RL+VIG Q+ +A
 Sbjct 55 IIIFVRHFLCFTCQEYVEDLSQIPQEILLDADVRLVIG-----QSGFSHIEAFCSLTGY 109
 Query 61 SIPVYTDPLRRSYLALRFRRG 81
 +Y DP R Y L +RG
 Sbjct 110 QHEIYVDPERHIYEKLGMRG 130

>ref|YP_001543559.1| **G** hypothetical protein Haur_0783 [Herpetosiphon aurantiacus ATCC 23779]

gb|ABX03431.1| **G** hypothetical protein Haur_0783 [Herpetosiphon aurantiacus ATCC 23779]
 Length=184

GENE ID: 5732667 Haur 0783 | hypothetical protein
 [Herpetosiphon aurantiacus ATCC 23779]

Score = 40.4 bits (93), Expect = 0.071, Method: Compositional matrix adjust.
 Identities = 31/91 (35%), Positives = 42/91 (47%), Gaps = 6/91 (6%)

Query 1 LVIWPRHFGVCVLCRRLVKELAAREQAFRQERKWRMLMVIGCGTPEQAAKFRTDNLGPLPA 60
 L++ RH GC LCR+ + L +Q F + VI G + A FR L LP
 Sbjct 32 LLMLMRHVGCGLCRQQLYRLREYQQRFHHAH-CEIAVIIMGDAKMAHGFR---QLNRLP- 86
 Query 61 SIPVYTDPLRRSYLALRFRRGILSTFNLPAL 91
 PVY+DP + Y A +G L T P +
 Sbjct 87 -FPVYSDPKQHVEAFEIGQGSGLWTVAGPHV 116

>ref|XP_001824981.1| **G** hypothetical protein AOR_1_1228084 [Aspergillus oryzae RIB40]
 dbj|BAE63848.1| unnamed protein product [Aspergillus oryzae]
 Length=250

GENE ID: 5997067 AOR_1_1228084 | hypothetical protein
 [Aspergillus oryzae RIB40] (10 or fewer PubMed links)

Score = 40.4 bits (93), Expect = 0.071, Method: Compositional matrix adjust.
 Identities = 24/79 (31%), Positives = 36/79 (46%), Gaps = 8/79 (10%)

Query 1 LVIWPRHFGVCVLCRRLVKELAAREQAFRQE---RKWRLMVIGCGTPEQAAKFRTDNLG 57
 LVI+ RHF C C+ ++ L+A E +++IGCG P + +
 Sbjct 75 LVIFVRHFFCGSCQEFRLTSLASITPKALEPLATSTSVVIIGCGDPGLIEMYEKETN--- 131
 Query 58 LPASIPVYTDPLRRSYLAL 76
 P+YTD P R+ Y L

Sbjct 132 --CQFPIYDTPTRQLYQDL 148

>ref|XP_002383597.1| **G** conserved hypothetical protein [Aspergillus flavus NRRL3357]
 gb|EED46061.1| **G** conserved hypothetical protein [Aspergillus flavus NRRL3357]
 Length=250

GENE ID: 7910923 AFLA 097240 | hypothetical protein
 [Aspergillus flavus NRRL3357]

Score = 40.4 bits (93), Expect = 0.072, Method: Compositional matrix adjust.
 Identities = 24/79 (31%), Positives = 36/79 (46%), Gaps = 8/79 (10%)

Query 1 LVIWPRHFGCVLCRRLVKELAAREQAFRQE--RKWRLMVIGCGTPEQAAKFRDATLGP 57
 LVI+ RHF C C+ ++ LA E +++IGCG P + +
 Sbjct 75 LVIFVRHFFCGSCQEFRLTSLASITPKALEPLATSTSVVIIGCGDPGLIEMYKETN--- 131
 Query 58 LPASIPVYTDPLRRSYLAL 76
 P+YTDP R+ Y L
 Sbjct 132 --CQFPIYDTPTRQLYQDL 148

>ref|XP_001368977.1| **UGM** PREDICTED: similar to C90RF21 [Monodelphis domestica]
 Length=354

GENE ID: 100014720 LOC100014720 | similar to C90RF21 [Monodelphis domestica]

Score = 40.4 bits (93), Expect = 0.077, Method: Compositional matrix adjust.
 Identities = 37/128 (29%), Positives = 57/128 (45%), Gaps = 18/128 (14%)

Query 1 LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRMLVIGCGTPEQAAKFRDATLGPLPA 60
 +V++ RHF C C+ V++LA ++F Q+ L+VIG + + F
 Sbjct 187 IVVFRHFLCYTCKEYVEDLAKIPKSFLODANVTLIVIGQSSFQHIEPFCKLTRY----- 241
 Query 61 SIPVYTDPLRRSY--LALFRFRGILSTFNLPALQNILDSFQKGEROR-WDMIPPDAF--- 114
 S +Y D R+ Y L + GI S+ P +++ L S G Q W + AF
 Sbjct 242 SHEIYVDTERKIYRKLGMNKGEGIASSEQSPHVKSNLLS---GSIQSLWRAVTPAFDFQ 298
 Query 115 ----QGG 118
 QGG
 Sbjct 299 GDPAQGG 306

>ref|NP_001158627.1| **UG** UPF0308 protein C9orf21 homolog [Oncorhynchus mykiss]
 gb|AC008544.1| **G** UPF0308 protein C9orf21 homolog [Oncorhynchus mykiss]
 Length=224

GENE ID: 100305250 ci021 | UPF0308 protein C9orf21 homolog
 [Oncorhynchus mykiss]

Score = 40.4 bits (93), Expect = 0.079, Method: Compositional matrix adjust.
 Identities = 28/85 (33%), Positives = 39/85 (46%), Gaps = 5/85 (5%)



Query 1 LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRMLVIGCGTPEQAAKFRDATLGPLPA 60
 +VI+ R+F C C+ V +L+ +E RL+VIG + F +L P
 Sbjct 55 VVIFVRNFLCHTCKEYVDDLRIAPAEILKEAGRLRVVIGQSSHHHIESF---CSLTGYPH 111
 Query 61 SIPVYTDPLRRSYLALFRFRGILST 85
 I Y DP R Y L RRG +S
 Sbjct 112 DI--YVDPERCIYKRLGMRRGEMSV 134

>ref|NP_001177306.1| **G** selenoprotein L [Ciona intestinalis]
 Length=303

GENE ID: 100175934 sell | selenoprotein L [Ciona intestinalis]
 (10 or fewer PubMed links)

Score = 40.4 bits (93), Expect = 0.087, Method: Compositional matrix adjust.
 Identities = 29/91 (32%), Positives = 43/91 (48%), Gaps = 6/91 (6%)

Query 1 LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRMLVIGCGTPEQAAKFRDATLGPLPA 60
 LVI RHF + R+ V E+ ++ F Q K ++++ G E+A D L
 Sbjct 150 LVILLRHFAULPURKHVTEIQEKQDEFSQ-LKCSIILVSFG--EEAG---ADLWLKETKF 203
 Query 61 SIPVYTDPLRRSYLALFRFRGILSTFNLPAL 91
 S P+Y D R+ Y AL R + +N AL
 Sbjct 204 SFPMYLDKQRKIYSALGLGRSVAKVYNCNAL 234



>ref|YP_002606071.1|  hypothetical protein HRM2_48590 [Desulfobacterium autotrophicum HRM2]
 >gb|ACN17907.1|  conserved hypothetical protein [Desulfobacterium autotrophicum HRM2]
 Length=134

GENE ID: 7500210 HRM2_48590 | hypothetical protein
 [Desulfobacterium autotrophicum HRM2] (10 or fewer PubMed links)

Score = 40.0 bits (92), Expect = 0.089, Method: Compositional matrix adjust.
 Identities = 24/94 (26%), Positives = 43/94 (46%), Gaps = 6/94 (6%)

```
Query 11  VLCRRLVKELAAREQAFRQERKWRMLVIGCGTPEQAAKFRDATLGPLPASIPVYTDPLR 70
          + CR+ V ELA + F+ ++ L+VIG G P +FR G +++DP
Sbjct 1   MFRCRQVAELATQANRFK-DKGAALVVIGSGDPRYFKQFREITGYGG-----QLFSDPSL 54

Query 71  RSYLALRFRRGILSTFNLPALQNILDSFQKGERQ 104
          ++ L F I ++ ++ + + G RQ
Sbjct 55  EAFSVLNFSSSITGFMSINSMFKAVALANGHRQ 88
```

>ref|XP_961259.1|  hypothetical protein NCU04268 [Neurospora crassa OR74A]
 >emb|CAC28745.1| putative protein [Neurospora crassa]
 >gb|EAA32023.1|  conserved hypothetical protein [Neurospora crassa OR74A]
 Length=411

GENE ID: 3877391 NCU04268 | hypothetical protein [Neurospora crassa OR74A]
 (10 or fewer PubMed links)

Score = 40.0 bits (92), Expect = 0.10, Method: Composition-based stats.
 Identities = 27/83 (33%), Positives = 38/83 (46%), Gaps = 8/83 (9%)

```
Query 1   LVIWPRHFGCVLCRRLVKELAAR---EQAFRQERKWRMLVIGCGTPEQAAKFRDATLGP 57
          L+I+ RHF C C+ ++ L+A + R + V+GCG P Q AT P
Sbjct 124 LIIFVRHFFCGQCQEYLRTLSASITPDALLRLPLSTFIAVVGCGNP-QLIDMYAQATNCP 182


Query 58  LPASIPVYTDPLRRSYLALRFRR 80
          P+Y DP R+ Y L R
Sbjct 183 F----PIYADPTRKLYQELGMVR 201
```

>gb|ACI68733.1| Clorf93 homolog [Salmo salar]
 Length=224

Score = 40.0 bits (92), Expect = 0.11, Method: Compositional matrix adjust.
 Identities = 28/85 (33%), Positives = 39/85 (46%), Gaps = 5/85 (5%)

```
Query 1   LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRMLVIGCGTPEQAAKFRDATLGPLPA 60
          +VI+ R+F C C+ V +L+ +E RL+VIG + F +L P
Sbjct 55  VVIFVRNFLCHTCKEYVDDLSRIPAEVLKEAGLRLVVIGQSSHHHIESF---CSLTGYPH 111

Query 61  SIPVYTDPLRRSYLALRFRRGILST 85
          I Y DP R Y L RRG +S
Sbjct 112 DI--YVDPERCIYKRLGMRRGEMSV 134
```


>gb|EDL16239.1|  RIKEN cDNA 1110018J18, isoform CRA_c [Mus musculus]
 Length=141

GENE ID: 66129 1110018J18Rik | RIKEN cDNA 1110018J18 gene [Mus musculus]
 (Over 10 PubMed links)

Score = 39.7 bits (91), Expect = 0.14, Method: Compositional matrix adjust.
 Identities = 24/81 (30%), Positives = 40/81 (50%), Gaps = 5/81 (6%)

```
Query 1   LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRMLVIGCGTPEQAAKFRDATLGPLPA 60
          +V++ RHF C +C+ V++LA ++ +E L+VIG Q++ +
Sbjct 59  VVVFVRHFLCYVCKEYVEDLAKIPKSVLREADVTLVIG-----QSSYHHIEPFCKLTGY 113

Query 61  SIPVYTDPLRRSYLALRFRRG 81
          S +Y DP R Y L +RG
Sbjct 114 SHEIYVDPEREIYKRLGMKRG 134
```

>ref|XP_003050915.1|  hypothetical protein NECHADRAFT_69625 [Nectria haematococca mpVI 77-13-4]

gb|EEU45202.1 | **G** hypothetical protein NECHADRAFT_69625 [Nectria haematococca mpVI 77-13-4]
Length=386

GENE ID: 9668148 NECHADRAFT 69625 | hypothetical protein [Nectria haematococca mpVI 77-13-4]

Score = 39.7 bits (91), Expect = 0.15, Method: Composition-based stats, Identities = 34/116 (30%), Positives = 50/116 (44%), Gaps = 21/116 (18%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAAAREQAFRQERKWRMLVIGCGTPEQAAKFRDTATLGP 57
          L+I+ RHF C C+ ++ L+ E R + VIGCG P D +
Sbjct 148 LIIFVRHFFCGNCQEFRLSLSDAVTPEALLRPLVSTFIAVIGCGDPAL-----IDMYVNE 202

Query 58 LPASIPVYTDPLRRSYLALRFRRGILSTFNLPA-----LQNILDSFQKGERQ 104
          PVYTDP R + AL G+ T + A ++I+DS +G +0
Sbjct 203 TNCRFVYTDPTRSLFHAL----GMTKTLQMGAKPAYMRKSMTRSIVDSIVQGVKQ 254
```

>**ref|YP_001543558.1** | **G** hypothetical protein Haur_0782 [Herpetosiphon aurantiacus ATCC 23779]

gb|ABX03430.1 | **G** hypothetical protein Haur_0782 [Herpetosiphon aurantiacus ATCC 23779]
Length=183

GENE ID: 5732666 Haur 0782 | hypothetical protein [Herpetosiphon aurantiacus ATCC 23779]

Score = 39.3 bits (90), Expect = 0.15, Method: Compositional matrix adjust. Identities = 20/73 (28%), Positives = 39/73 (54%), Gaps = 6/73 (8%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAAREQAFRQERKWRMLVIGCGTPEQAAKFRDTATLGPLPA 60
          L+ + R+ GC +CR+ + L +QAF+ + W+++V+ G E ++FR+ +
Sbjct 37  LIFFMRNIGCGICRQTLNLRDQYDQAFK-DAGWQIVVLMGMNVELVSRFRSMYNI----- 90

Query 61 SIPVYTDPLRRSY 73
          P+Y D + Y
Sbjct 91  PFPIYVDQSLQVY 103
```

>**ref|ZP_01089021.1** | hypothetical protein DSM3645_00440 [Blastopirellula marina DSM 3645]

gb|EAQ82136.1 | hypothetical protein DSM3645_00440 [Blastopirellula marina DSM 3645]
Length=184

Score = 38.9 bits (89), Expect = 0.21, Method: Compositional matrix adjust. Identities = 27/89 (31%), Positives = 38/89 (43%), Gaps = 6/89 (6%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAAREQAFRQERKWRMLVIGCGTPEQAAKFRDTATLGPLPA 60
          L+ + RH GC CR K+L Q+ ++ I GT E+A + + L
Sbjct 41  LIQFARHLGCTFCRDRAKQLKLDYPEI-QQHNGDVVVFITMGThERAQQLQDGMQL----- 94

Query 61 SIPVYTDPLRRSYLALRFRRGILSTFNLPA 89
          V DP R +Y A R RG +S P
Sbjct 95  PFDVLDVPHRAAYQAFRVPRGNVSQVAGP 123
```

>**ref|XP_001800743.1** | **G** hypothetical protein SNOG_10473 [Phaeosphaeria nodorum SN15]

gb|EAT81867.1 | **G** hypothetical protein SNOG_10473 [Phaeosphaeria nodorum SN15]
Length=338


GENE ID: 5977650 SNOG_10473 | hypothetical protein [Phaeosphaeria nodorum SN15]

Score = 38.1 bits (87), Expect = 0.35, Method: Compositional matrix adjust. Identities = 29/93 (32%), Positives = 42/93 (46%), Gaps = 12/93 (12%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAAR--EQAFRQERKWRMLVIGCGTPEQAAKFRDTATLGP 57
          L+I+ RHF C C+ ++ L++ E + VIGCG P + + T P
Sbjct 122 LIIFIRHFFCGNCQEYLRTLSSINA EALLALPTPTSITVIGCGKPALIPMY-AETTKCP 180

Query 58 LPASIPVYTDPLRRSYLALRFRRGILSTFNLPA 90
          P+Y DP R+ Y L G+ TF L A
Sbjct 181 F----PIYADPTRKLYDLL----GMTRTFQLGA 205
```

>**ref|YP_001547656.1** | **G** hypothetical protein Haur_4898 [Herpetosiphon aurantiacus ATCC 23779]

gb|ABX07528.1 |  hypothetical protein Haur_4898 [Herpetosiphon aurantiacus ATCC 23779]
Length=193

GENE ID: 5736734 Haur 4898 | hypothetical protein [Herpetosiphon aurantiacus ATCC 23779]

Score = 38.1 bits (87), Expect = 0.35, Method: Compositional matrix adjust.
Identities = 33/89 (38%), Positives = 45/89 (51%), Gaps = 6/89 (6%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          L+++ RH GC +CRR + +L      AFR      L VI  T  + A  +  A L  LP
Sbjct 30  LLVFLRHGCAVCRRNLLDLYEYTTAFRM-LDINLAVI---TMAEPAAAQAFARLYRLP- 84


Query 61  SIPVYTDPLRRSYLALRFRRGILSTFNLP 89
          IP+Y+DP R+ Y A  F  G L +  P
Sbjct 85  -IPIYSDPERQIYRAAGFSEGLYSVASP 112
```


>**gb|ACI70082.1** | Clorf93 homolog [Salmo salar]
Length=232

Score = 38.1 bits (87), Expect = 0.37, Method: Compositional matrix adjust.
Identities = 26/85 (31%), Positives = 39/85 (46%), Gaps = 5/85 (5%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPA 60
          ++I+ R+F C  C+  V +L+      +E  RL+VIG  +  F  +L  P
Sbjct 63  VIIFVRNFLCHTCKEYVDDLSRIPAEVLKEAGRLVIGQSSHHHIESF---CSLTGYPH 119

Query 61  SIPVYTDPLRRSYLALRFRRGILST 85
          +  Y DP R  Y  L  RRG +S
Sbjct 120  DM--YVDPERCIYKRLGMRRGEMSV 142
```

>**ref|XP_002544130.1** |  predicted protein [Uncinocarpus reesii 1704]


gb|EEP78801.1 |  predicted protein [Uncinocarpus reesii 1704]
Length=245


GENE ID: 8443780 UREG_03647 | hypothetical protein [Uncinocarpus reesii 1704]

Score = 38.1 bits (87), Expect = 0.38, Method: Compositional matrix adjust.
Identities = 23/81 (29%), Positives = 34/81 (42%), Gaps = 6/81 (7%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAR-EQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLP 59
          +V++ RHF C  C+  V+ L++      L+VIGCG  +  +
Sbjct 80  MVVFVRHFFCGSQDYVQTLSSIPSPASLPTGTSLVIGCGATSLIPMYAKTTS----- 134

Query 60  ASIPVYTDPLRRSYLALRFRR 80
          S P+YTD P  R Y  R
Sbjct 135  CSFPIYTDPTSRLYTIFGMTR 155
```

>**ref|XP_002114328.1** |  hypothetical protein TRIADDRAFT_58024 [Trichoplax adhaerens]

gb|EDV23418.1 |  hypothetical protein TRIADDRAFT_58024 [Trichoplax adhaerens]
Length=218

GENE ID: 6755541 TRIADDRAFT_58024 | hypothetical protein [Trichoplax adhaerens]
(10 or fewer PubMed links)

Score = 38.1 bits (87), Expect = 0.40, Method: Compositional matrix adjust.
Identities = 22/60 (37%), Positives = 28/60 (47%), Gaps = 5/60 (8%)

```
Query 17  VKELAAREQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLPASIPVYTDPLRRSYLAL 76
          V + + R  AF +E  RL+VIGCG P  KFR +  +Y DP R  Y  L
Sbjct 93  VYKYSKRSYAFLEKANVRLVIGCGKPNFIKFRQETGYNH----DMYCDPQRTIYSKL 147
```


>**gb|EFW13835.1** | conserved hypothetical protein [Coccidioides posadasii str. Silveira]
Length=246


Score = 37.4 bits (85), Expect = 0.62, Method: Compositional matrix adjust.
Identities = 23/81 (29%), Positives = 34/81 (42%), Gaps = 6/81 (7%)

```
Query 1  LVIWPRHFGCVLCRRLVKELAAR-EQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGPLP 59
          +V++ RHF C  C+  V  L++      L+VIGCG  +  +  +
Sbjct 81  MVVFVRHFFCGSQDYVATLSSIPSPASLPPGVNLVIGCGASSLISMYANTTS----- 135

Query 60  ASIPVYTDPLRRSYLALRFRR 80
```

Sbjct 136 S P+YTD P R Y R
CSFPIYTDPTGRLYSIFGMTR 156

>ref|XP_003070048.1|  hypothetical protein CPC735_032390 [Coccidioides posadasii C735 delta S0Wgp]

gb|EER27903.1|  hypothetical protein CPC735_032390 [Coccidioides posadasii C735 delta S0Wgp]
Length=206

GENE ID: 9695543 CPC735_032390 | hypothetical protein
[Coccidioides posadasii C735 delta S0Wgp]

Score = 37.4 bits (85), Expect = 0.63, Method: Compositional matrix adjust.
Identities = 22/74 (30%), Positives = 33/74 (45%), Gaps = 6/74 (8%)

Query	1	LVIWPRHFGCVLCRRLVKELAAR-EQAFRQERKWRLMVIGCGTPEQAAKFRTDNLGGLP	59
		+V++ RHF C C+ V L++ L+VIGCG ++ +	
Sbjct	41	MVVFVRHFFCGSCQDYVATLSSSIPSPASLPPGVNLVIGCGASSLISMYANTTS-----	95
Query	60	ASIPVYTDPLRRSY	73
		S P+YTD P R Y	
Sbjct	96	CSFPIYTDPTGRLY	109